The London School of Economics and Political Science

Essays in Labor Economics

Georg Graetz

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, May 2014

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Statement of conjoint work

I confirm that Chapters 1 and 2 were jointly authored with Andy Feng with equal shares in all aspects of the chapters.

Statement of inclusion of previous work

I confirm that Chapter 3 is a heavily revised version of the paper that I submitted for the Quantitative Economics Project (EC331) as part of the BSc Econometrics and Mathematical Economics degree at LSE in 2009.

Abstract

This thesis titled “Essays in Labour Economics” is comprised of three essays investigating various determinants of earnings inequality.

Chapter 1 provides a novel explanation for labor market polarization—the rise in employment shares of high and low skill jobs at the expense of middle skill jobs, and the fall in middle-skill wages. We argue that recent and historical episodes of polarization resulted from increased automation. In our theoretical model, firms deciding whether to employ machines or workers in a given task weigh the cost of using machines, which is increasing in the complexity (in an engineering sense) of the task, against the cost of employing workers, which is increasing in training time required by the task. Some tasks do not require training regardless of complexity, while in other tasks training is required and increases in complexity. In equilibrium, firms are more likely to automate a task that requires training, holding complexity constant. We assume that more-skilled workers learn faster, and thus it is middle skill workers who have a comparative advantage in tasks that are most likely to be automated when machine design costs fall. In addition to explaining job polarization, our model makes sense of observed patterns of automation and accounts for a set of novel stylized facts about occupational training requirements.

Chapter 2 establishes a novel source of wage differences among observationally similar high skill workers. We show that degree class—a coarse measure of performance in university degrees—causally affects graduates' earnings. We employ a regression discontinuity design comparing graduates who differ only by a few marks in an individual exam, and whose degree class is thus assigned randomly. A First Class is worth roughly three percent in starting wages which translates into £1,000 per annum. An Upper Second is worth more on the margin—seven percent in starting wages (£2,040). In addition to identifying a novel source of luck in the determination of earnings, our findings also show the importance of simple heuristics for hiring decisions.

Chapter 3 asks whether public policy affects the degree of intergenerational transmission of education. The chapter investigates this question in the context of secondary school transitions in Germany. During the last three decades, several German states changed the rules for admission to secondary school tracks. Combining a new data set on transition rules with micro data from the German Socioeconomic Panel (SOEP), I find that allowing free track choice raises the probability of attending the most advanced track by five percentage points. However, the effect is twice as large for children of less educated parents. The results suggest that the correlation between parents' and children's educational attainment may be reduced by more than one third when no formal restrictions to choosing a secondary school track exist.

Acknowledgments

I am indebted to my advisors, Francesco Caselli and Guy Michaels, for their many valuable insights and helpful suggestions that greatly improved this thesis. I thank them for their patience and persistence, and for being always approachable.

I thank Steve Pischke for introducing me to empirical work through supervising the project which eventually led to the third chapter of this thesis. I am grateful to Alan Manning for many insightful conversations. Luis Garicano, Gianmarco Ottaviano, Thomas Sampson, and John Van Reenen provided numerous useful suggestions.

Andy Feng co-authored the first two chapters of this thesis: I thank him for a productive and enjoyable collaboration, and for countless engaging discussions over lunch. I also benefited from many suggestions by other fellow LSE students, especially Johannes Boehm, Michael Boehm, Marta De Philippis, and Claudia Steinwender. I am grateful to Yu-Hsiang Lei and Amar Shanghavi for their companionship throughout the PhD.

Chapters two and three would not have been possible without access to specialized data. Thanks go to Lucy Burrows from LSE Careers and Tom Richey from Student Records for providing the student data used in the second chapter. Thanks also to the staff at the Archive of the Conference of the Ministers of Education in Bonn, where I collected the data on school transition rules for the third chapter, for their hospitality.

I acknowledge financial support from the Economic and Social Research Council and the Royal Economic Society.

Last but not least, I am deeply grateful to my parents and my wife for their encouragement and support.

Contents

List of Tables

List of Figures

Chapter 1

Rise of the Machines: The Effects of Labor-Saving Innovations on Jobs and Wages

1.1 Introduction

A growing empirical literature argues that recent technical change has led to polarization of labor markets in the US and Europe:[1] Employment in middle skill occupations has grown relatively more slowly than in high and low skill occupations since the 1980s, and similarly wages have grown faster at the top and bottom of the distribution than in the middle. While factors such as offshoring and changes in labor market institutions may have contributed to these developments, a consensus has emerged that identifies technical change as the main culprit. Modern information and communication technology (ICT) appears to substitute for workers in middle skill jobs, while complementing labor in high and low skill jobs, thus causing the observed reallocation of employment.[2]

Recent work has also documented *historical* instances of job polarization. Katz and Margo (2013) show that from 1850 to 1880, US manufacturing witnessed a relative decline in middle skill jobs (artisans) at the expense of high skill jobs (non-production workers) and low skill jobs (operators), concurrent with the increased adoption of steam power. Gray (2013) finds that electrification in the US during the first half of the 20th century led to a fall in demand for dexterity-intensive tasks performed by middle skill workers, relative to manual and clerical tasks performed by low and high skill workers, respectively.[3]

The innovations that preceded the three instances of job polarization have in common that they facilitated a more wide-ranging automation of tasks. The steam engine was instrumental in the

[1]Job polarization has first been documented for the US by Autor, Katz, and Kearney (2006), for the UK by Goos and Manning (2007), and for other European economies by Goos, Manning, and Salomons (2009).

[2]See Autor, Levy, and Murnane (2003), Michaels, Natraj, and Van Reenen (forthcoming), and Goos, Manning, and Salomons (2014) for evidence favoring the technological explanation.

[3]Previously, Goldin and Katz (1998) presented evidence suggesting that electrification in the US was an instance of skill-biased technical change, although their empirical work focussed on a high-vs.-low-skill dichotomy.

increased mechanization of manufacturing because it provided a more reliable power source than water, and it allowed production to be located away from water, thus lowering transportation costs (Atack, Bateman, and Weiss 1980). Electricity facilitated automation because electric motors could be arranged much more flexibly than steam engines (Boff 1967). ICT allowed for the automation of cognitive tasks as well as improved control of physical production processes. Thus, the steam engine, electricity and ICT all triggered waves of labor-saving innovations. Current advances in artificial intelligence and robotics are likely to further boost automation, raising the question how job tasks, the distribution of employment across occupations, and the wage distribution will change as a result.[4]

In this chapter, we argue that labor market polarization is inherent to labor-saving technologies. Our explanation is based on two insights. First, when technologies are available that can carry out a wide range of tasks autonomously, the allocation of workers and machines to tasks will be determined by comparative advantage (Simon 1960). Second, what is complex from an engineering point of view is not necessarily difficult to humans in the sense of requiring a lot of training (Moravec 1988). There are tasks that are easy to *any* worker but building a machine capable of performing them may be costly if not impossible; occupations such as waiters, taxi drivers, or housekeepers are intensive in the use of vision, movement, and communication, which are complex functions from an engineering point of view (Moravec 1988). On the other hand, a task like bookkeeping requires knowledge of arithmetic which takes humans years to learn, but which is trivial from an engineering perspective. We demonstrate how the principle of comparative advantage and the distinction between engineering complexity and difficulty in the sense of training time, combine to generate an equilibrium in which workers in the middle of the skill distribution are at the greatest risk of being replaced by machines.

Existing literature explains job polarization as caused by new technologies that are most suited for use in tasks initially performed by middle skill workers. Autor, Katz, and Kearney (2006) assume that ICT substitutes for routine tasks performed by middle skill workers, while complementing manual and abstract tasks that in equilibrium are carried out by low and high skill workers, respectively. Acemoglu and Autor (2011) present a model in which machines may replace any type of labor, but for the model to generate job polarization they must assume that machines replace middle skill workers. Our model imposes no such restrictions. Firms may employ labor of any type or machines to complete any given task, and will choose the factor that minimizes costs. There is no assumption about which part of the skill distribution will be most affected by increased automation—instead, this is endogenous to the attributes of factors and their interplay with the characteristics of tasks.

Labor-replacing technical change in our model refers to an exogenous fall in the cost of making machines, resulting from innovations that facilitate the automation of a wide range of tasks. Workers differ continuously by skill—higher skilled workers learn faster. Tasks are differentiated by *complexity* and by whether workers require training. Complexity refers to the intrinsic difficulty of a task, for instance the amount of information processing required, or the degrees of freedom and dexterity necessary to carry out a certain physical action. Complexity can be seen as a measure

[4] We provide a list of examples for recent progress in these areas in Section 1.2.5.

of ‘engineering difficulty’: the more complex a task, the larger an expenditure is required to build a machine capable of performing it. Over time, falling design costs make it cheaper to build machines for any task, and the complexity-cost gradient becomes flatter.

The model features a set of tasks of varying complexity for which workers do not need any training as they draw on capabilities that are innate or have been acquired at an early age—we call these *innate ability* tasks. Crushing rocks to produce gravel is an innate ability task of low complexity, while waiting tables is an innate ability task of high complexity, due to the amount of physical coordination required as well as the need for communicating in natural language. In the remaining tasks, workers cannot draw on an endowment of abilities and hence knowledge must be acquired. We call these tasks *training-intensive*, and we assume that training time increases in complexity in these tasks. Bookkeeping is a training-intensive task of low complexity: humans must acquire knowledge of arithmetic and the rules of accounting, but arithmetic and accounting rules are easily codified and require little processing power when executed by machines. Lawyering, on the other hand, is both training-intensive and highly complex: humans need to acquire knowledge of the legal system from scratch, and complexity arises from the need to apply a large set of general rules to specific cases, where rules and cases are described in different kinds of language. Both bookkeeping and lawyering are training-intensive tasks, but training takes longer for lawyers, because their task is more complex.

The distinction between *innate ability* tasks and *training-intensive* tasks is critical for explaining why middle skill workers are most affected by increasing automation. Firms deciding whether to employ machines or workers in a given task weigh the cost of using machines, which is increasing in the complexity of the task, against the cost of employing workers, which is increasing in training time required by the task. As training time increases in complexity only in training-intensive tasks, firms are more likely to automate a training-intensive task than an innate ability one, holding complexity constant. Since higher skilled workers learn faster, workers at the bottom of the skill distribution have a comparative advantage (CA) in innate ability tasks. Workers at the top have a CA in highly complex training-intensive tasks, and workers in the middle have a CA in training-intensive tasks of intermediate complexity. In addition, we assume that design costs are such that machines’ CA is in less complex tasks. Hence, the tasks most likely to be newly automated when machine design costs fall are those performed by middle-skill workers. It is in these tasks that the incentives for automation are strongest, due to the need to train workers.

Our model is able to explain job polarization as well as the hollowing out of the wage structure: given that middle-skill workers are most affected by machine replacement, they experience the strongest downward pressure on wages. The model provides a precise mechanism explaining labor market polarization, suggesting that the ICT revolution has caused labor market polarization because it has facilitated a more wide-ranging automation of tasks. This implies that the model is also suited to explain the historical instances of job polarization that followed the introduction of steam power and electricity in manufacturing.

Our model helps to explain recent patterns of automation. There are job tasks that are currently not automated despite the fact that it would be feasible to do so. For example, fast food preparation takes place in a controlled environment and involves a limited number of simple steps. It is thus not surprising that technology exists that automates this process (Melendez 2013). However, fast food jobs do not currently appear to be at risk of being automated. Moreover, in manufacturing some tasks are automated that are arguably of higher complexity (Davidson 2012). Given that fast food preparation requires very little training, in particular compared to non-trivial assembly tasks in manufacturing, this pattern of automation choices is exactly what our model would predict.

In addition to explaining labor market polarization and patterns of task automation, our theory delivers several novel predictions about trends in occupational training requirements. In the model, training requirements are higher in more complex training-intensive tasks. We measure training requirements in the US at the occupation level, using the Dictionary of Occupational Titles (DOT) combined with the 1971 April CPS, and the O*NET database combined with the 2008 ACS.[5]

We find empirical support for the model's prediction of a polarization in training requirements, i.e. an increase in the employment shares of jobs requiring minimal and very high levels of training between 1971 and 2007. Furthermore, we show that occupations that initially had intermediate training requirements experienced a fall in training requirements. The model provides a ready explanation: new technologies induced firms to automate the subset of tasks in a given occupation which required intermediate training by workers.[6] We also find that employment growth was less in occupations that experienced larger decreases in training requirements, as should be the case if automation causes training requirements to fall. Finally, we show that composition-adjusted occupational wages increased less or decreased in occupations with larger falls in training requirements, again consistent with the model.

The plan of the chapter is as follows. The following subsection discusses related literature. Section 1.2 presents and solves the model. Section 1.3 discusses comparative statics, in particular how job assignment and the wage distribution change as a response to increased automation. We also present comparative statics for a change in skill supplies. Section 1.4 presents two extensions to the model: endogenous capital accumulation and a fixed cost of technology adoption. Section 1.5 confronts the model's prediction with existing empirical evidence and takes novel implications of the model to the data. Section 3.4 concludes. All proofs are contained in Appendix 1.D.

1.1.1 Related Literature

This chapter is related to a recent literature featuring task-based models of exogenous job polarization. Autor, Levy, and Murnane (2003, henceforth ALM) categorize tasks as routine and non-routine. They call a task routine "if it can be accomplished by machines following explicit programmed rules" (ibid., p.1283). In contrast, non-routine tasks are "tasks for which rules are not sufficiently well understood to be specified in computer code and executed by machines" (ibid.). Extending the framework of ALM, Autor, Katz, and Kearney (2006) and Autor and Dorn (2013) allow non-routine tasks to be either low or high skill intensive.

We believe that our framework offers several advantages over ALM's. First, it is not context-dependent. Machine capabilities constantly expand, so we prefer to avoid a task construct that

[5]In these data, training requirements are measured as the time it takes the typical worker to become proficient in her job. This may include any occupation-specific knowledge acquired prior to entering the labor market—see Section 1.5.2.

[6]All our results on changes in training requirements are robust to controlling for changes in mean years of schooling.

depends on the current state of technology.[7] Complexity in our model is an objective, time-invariant measure of a task's intrinsic difficulty. Second, the notion that worker training may be uncorrelated with a task's complexity does not feature prominently in ALM's framework. Finally, ALM's framework implicitly leaves firms little choice to automate a given task, as routine tasks are assumed to be automated, and non-routine tasks are not. Our framework instead allows us to endogenize this choice. A task like fast food preparation might be considered "routine" in ALM's framework, so that the non-automation of this task poses a puzzle. As noted above, our model is consistent with this example.

Acemoglu and Autor (2011) allow machines to replace labor in any task in principle, but assume that machines perform tasks initially performed by middle skill workers to make the model consistent with job polarization. Because their task index does not have an empirical interpretation independently of factor assignment,[8] the model does not impose any restrictions on the data and thus could be consistent with arbitrary patterns of labor-replacing technical change, but without explaining them.[9]

We build on the literature on labor-saving innovations. Zeira (1998) presents a model in which economic development is characterized by the adoption of technologies that reduce labor requirements relative to capital requirements. Over time, an increasing number of tasks can be produced by new, more capital-intensive technologies. In an extreme example which is closely related to this chapter, new technologies only use capital, while old ones only use labor. We extend this type of setting by explicitly modeling the characteristics of tasks and thus the direction of technical change, as well as by allowing for heterogenous workers. Holmes and Mitchell (2008) present a model of firm organization where the problem of matching workers and machines to tasks is solved at the firm level. Their model admits a discrete set of worker types and they do not consider technical change.

The chapter is related to a theoretical literature that uses assignment models to investigate the effects of technical change on the role of workers in the production process and on the wage distribution. Garicano and Rossi-Hansberg (2006) analyze how hierarchical organizations are affected by declines in communication and knowledge acquisition costs as caused by the ICT revolution. They match a continuum of skills to a continuum of tasks and find that lower communication costs lead to falling inequality among one group (workers) but to rising inequality among another group (managers). Thus, their model explains wage polarization. We abstract from issues related to firm organization and focus instead on the labor-saving aspect of ICT and other innovations. Another strand of papers analyzes the matching of workers with technologies of different vintages. Wage inequality results for instance when workers must acquire vintage-specific

[7] To give an example, Levy and Murnane (2004) consider taking a left-turn on a busy road a nonroutine task unlikely to be automated in the foreseeable future. But less than a decade later, the driverless car has become a reality.

[8] The task index in their model indicates whether high skill workers have a comparative advantage over middle skill workers in a given task, whether middle skill workers have a comparative advantage over low skill workers etc.

[9] Autor (2013) argues that "capital typically takes over tasks formerly performed by labor; simultaneously, workers are typically assigned novel tasks before they are automated." One difficulty with this argument is that it does not help to explain why middle skill workers should be more affected by automation than low skill ones. A second difficulty is that there are many tasks that have existed for hundreds of years and are still not being automated (especially in low-skill services), while production of new products and services often involves tasks that workers never performed.

skills (Chari and Hopenhayn 1991) or machines are indivisible (Jovanovic 1998). Furthermore, skill or unskill bias of technical change can arise when new technologies require different learning investments than old ones, and when learning costs are a function of skill (Caselli 1999). We abstract from the issue of workers having to learn how to operate new technologies and focus instead on the problem of assigning workers and machines to tasks.

On the methodological side this chapter is in the tradition of Ricardian theories of international trade, combining aspects of Dornbusch, Fischer, and Samuelson (1977) and Costinot and Vogel (2010). While these papers characterize equilibrium allocations *given* factor endowments and productivity levels, our focus is on endogenizing productivity differences, using modeling techniques similar to those of Costinot (2009). We shed light on the *sources* of comparative advantage between differently-skilled workers and machines.

1.2 The Model

1.2.1 Overview

The model has one period that we interpret as a worker's lifetime.[10] There is a unique final good that is produced using a continuum of intermediate inputs, or *tasks*. These tasks are performed by workers of different skill levels and machines. Crucially, all factors of production are perfect substitutes at the task level. Although this may seem a strong assumption, the loss of generality is not substantial provided all tasks are essential in producing the final good, a condition that we shall maintain throughout. In fact, when tasks are imperfect substitutes in producing the final good, factors of production will appear to be imperfect substitutes in the aggregate.

Labor services as well as the economy's capital stock are supplied inelastically and all firms are perfectly competitive. Intermediate firms hire workers or capital to produce task output that is then sold to final good firms. Intermediate firms must train workers (except in innate ability tasks), and must transform generic capital into task-specific machines in order for these factors to be capable of performing tasks.

Technologies for worker training and machine design are public knowledge. Training and design expenditures are determined by task characteristics alone. We make this assumption for ease of exposition. In Appendix 1.B we present a more general model that allows firms to choose training and design expenditures, thus determining factors' productivity endogenously. That model is based on an explicit characterization of the production process following Garicano (2000). All our results apply to the more general model as well.

1.2.2 Machine Design and Worker Training

Tasks are differentiated by *complexity*, denoted by $\sigma \in [\underline{\sigma}, \overline{\sigma}]$, and by whether workers require training to complete the task, indicated by $\tau \in \{0, 1\}$. The expenditure required to convert one unit of capital into a machine capable of performing a task of complexity σ is given by $c_K\sigma$. Complexity σ is the task-specific component of design expenditure. It is a fundamental, time-invariant property

[10]We discuss a dynamic (multi-period) version of the model in Section 1.4.1.

of tasks. The design cost c_K is constant across tasks and falls over time as better technologies become available, leading to a flatter complexity-cost gradient.[11]

For workers, the complexity of a task does not necessarily affect the amount of training required. In particular, no training is required if completion of a task relies solely on functions that all workers are endowed with or have acquired at early age, regardless of the complexity of the task. This is true in the case of innate ability tasks ($\tau = 0$). Workers cannot rely on such endowments in the case of training-intensive tasks ($\tau = 1$). Training requirements do increase with complexity in training-intensive tasks: To become capable of performing the training-intensive task of complexity σ, a worker of type s requires an amount of training σ/s. Higher skilled workers face a flatter complexity-training gradient in training-intensive tasks.

1.2.3 Task Production

A machine produces A_K units of task output, regardless of the task's complexity. Hence, A_K can be viewed as task-neutral machine productivity. Since some of the hired capital is lost in machine design, a unit of capital produces an amount of output equal to $A_K(1 - c_K\sigma)$.[12]

All workers have a unit endowment of time, and produce one unit of task output if they are able to spend all their time in production. In other words, we normalize all workers' task-neutral productivity to one. Hence, workers of any type produce one unit of task output in innate ability tasks ($\tau = 0$). Taking into account training time, a worker of type s produces $1 - \sigma/s$ units of task output in training-intensive tasks ($\tau = 1$) of complexity σ.

The notation $s_K \equiv 1/c_K$, 'machine skill', will turn out to be more convenient. Let worker type range from $\underline{s} > 0$ to $\overline{s}$. To make the model interesting, we assume throughout that $s_K, \underline{s} \geq \overline{\sigma}$, so that machines and all worker types produce non-negative output in any task.

Let $k_\tau(\sigma)$ denote the amount of capital used to produce task (σ, τ) and similarly let $n_\tau(s, \sigma)$ be the amount of type-s labor. Given the task-specific productivity schedules for machines

$$\alpha^K(s_K, \sigma) = 1 - \sigma/s_K \tag{1.1}$$

and labor

$$\alpha^N(s, \sigma, \tau) = \begin{cases} 1 & \text{if } \tau = 0 \\ 1 - \sigma/s & \text{if } \tau = 1, \end{cases} \tag{1.2}$$

task output y can be written as

$$y_\tau(\sigma) = A_K \alpha^K(s_K, \sigma) k_\tau(\sigma) + \int_{\underline{s}}^{\overline{s}} \alpha^N(s, \sigma, \tau) n_\tau(s, \sigma) ds. \tag{1.3}$$

[11] Strictly speaking, c_K is the design cost per unit of capital, per unit of the complexity measure.

[12] In reality, many of the innovations that lead to a fall in c_K may also cause a rise in A_K. However, the comparative statics with respect to A_K are qualitatively the same as those with respect to c_K (a proof is available upon request).

1.2.4 Final Good Production and Market Clearing

Let Y denote the output of the final good. For tractability, we use a Cobb-Douglas production function,

$$\log Y = \frac{1}{\mu} \int_{\underline{\sigma}}^{\overline{\sigma}} \{\beta_0 \log y_0(\sigma) + \beta_1 \log y_1(\sigma)\} \, d\sigma. \tag{1.4}$$

Recall that the subscripts 0 and 1 indicate innate ability ($\tau = 0$) and training-intensive ($\tau = 1$) tasks, respectively. We impose $\sum_\tau \beta_\tau = 1$ and $\mu \equiv \overline{\sigma} - \underline{\sigma}$ to ensure constant returns to scale.

Let there be a mass K of machine capital and normalize the labor force to have unit mass. We assume a skill distribution that is continuous and without mass points. Let $V(s)$ denote the differentiable CDF, and $v(s)$ the PDF, with support $[\underline{s}, \overline{s}]$. Factor market clearing conditions are

$$v(s) = \sum_\tau \int_{\underline{\sigma}}^{\overline{\sigma}} n_\tau(s, \sigma) d\sigma \quad \textit{for all } s \in [\underline{s}, \overline{s}] \tag{1.5}$$

and

$$K = \sum_\tau \int_{\underline{\sigma}}^{\overline{\sigma}} k_\tau(\sigma) d\sigma. \tag{1.6}$$

Before characterizing the competitive equilibrium, we turn to a discussion of the assumptions underlying our task model.

1.2.5 Motivating the Assumptions of the Task Model

Researchers in artificial intelligence, robotics, and cognitive science have long been aware that some abilities that humans acquire quickly at an early age rely in fact on highly complex functions that are difficult if not impossible to reverse-engineer. In contrast, many abilities that humans must painstakingly acquire, such as mastery in arithmetic, are trivial from an engineering perspective. This observation has become known as Moravec's paradox: "[It] is comparatively easy to make computers exhibit adult-level performance in solving problems on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility" (Moravec 1988, p.15). Moravec resolves the puzzle by considering the objective or intrinsic difficulty of a task, for instance the amount of information processing required, or the degrees of freedom and dexterity necessary to carry out a certain physical action. In terms of intrinsic difficulty, arithmetic is much easier than walking or face-to-face communication. A possible indicator of the intrinsic difficulty of a task, or its complexity, is the computer power a machine requires to perform the task. A common if imperfect measure of computer power is million instructions per second (MIPS).[13] UNIVAC I, a computer built in 1951 and able to perform arithmetic operations at a much faster rate than humans, performed at 0.002 MIPS. ASIMO, a robot introduced in 2000 that walks, recognizes faces and processes natural language, requires about

[13]See Nordhaus (2007) for a discussion of MIPS as a measure of computer power.

Table 1.1: Two-Dimensional Task Framework, Examples

	Complexity $-$	$+$
Little or no training required	crushing rocks fast food preparation	customer reception waiting tables
Training required	bookkeeping weaving	lawyering management

4,000 MIPS.[14] By this measure, a set of tasks routinely performed by any three-year-old is two million times more complex than arithmetic. The reason that we are usually not aware of this fact, and why Moravec's observation at first seems puzzling, is that we rely on innate abilities[15] for functions like movement or perception, but have no such advantage when it comes to abstract tasks like arithmetic.[16]

We incorporate these insights into our framework by assuming that the amount of training a worker requires to be able to perform a task does not always depend on complexity.[17] Table 1.1 gives an overview of our task framework and contains examples. The bottom left and top right corners demonstrate the importance of a two-dimensional task space.

While we believe that our task framework is an improvement over existing literature, there are some limitations. For instance, technical change often leads to the introduction of new tasks and activities (flying airplanes, writing software). While our framework in principle allows for an endogenous task space, it does not suggest in what way technology might affect the set of tasks in the economy. Furthermore, automation does not necessarily involve machines replicating exactly the steps that humans carry out in completing a given task. Instead, a task can be made less complex by moving it to a more controlled environment.[18] Our framework does not explicitly allow for this

[14]For a comparison of various computers (including UNIVAC I) and processors by MIPS, see `http://en.wikipedia.org/wiki/Instructions_per_second`, retrieved on October 16, 2013. For technical details of ASIMO, see Sakagami, Watanabe, Aoyama, Matsunaga, Higaki, and Fujimura (2002).

[15]"Innateness" of a certain skill does not need to imply that humans are born with it; instead, the subsequent development of the skill could be genetically encoded. For a critical discussion of the concept of innateness, see Mameli and Bateson (2011).

[16]Moravec (1988, pp.15-16) provides an evolutionary explanation for this: "...survival in the fierce competition over such limited resources as space, food, or mates has often been awarded to the animal that could most quickly produce a correct action from inconclusive perceptions. Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much more powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it."

[17]Machines could be viewed as being endowed with some functions to the extent that materials have productive properties—take for instance copper with its electrical conductivity; but such endowments are usually highly specific and limited.

[18]See ALM (p.1283) and Simon (1960, pp.33-35). A recent example is the new sorting machine employed by the New York Public Library (Taylor 2010).

possibility, but our conclusions should still be broadly correct if the cost of moving a process to a more controlled environment is increasing in its complexity. Finally, technological change tends to cause organizational change, but to keep the analysis tractable and to be able to focus on a single mechanism, we omit firm organization from the model.

What we do not view as a limitation is the assumption that machines could in principle perform any task. Comparative advantage ensures that some tasks will always be performed by humans, so that the model will be consistent with the fact that some tasks are not performed by machines in reality. More importantly, recent technological progress suggests that machine capabilities might be expanding quite rapidly. Brynjolfsson and McAfee (2011, p.14) argue that machines can potentially substitute for humans in a much larger range of tasks than was thought possible not long ago, citing recent advances in pattern recognition (driverless cars), complex communication (machine translation), and combinations of the two (IBM's successful Jeopardy contestant Watson). Markoff (2012) provides an account of the increased flexibility, dexterity, and sophistication of production robots. Shein (2013) gives examples of robots being increasingly adopted in manual tasks, such as collecting items in a warehouse, or the pruning of grapes.[19] For our model to be useful as a guide to medium-term future developments in the economy, we deem it prudent to make the most conservative assumption about what tasks are safe from automation.

1.2.6 Characterizing the Competitive Equilibrium

We normalize the price of the final good to one and denote the price of task (σ, τ) by $p_\tau(\sigma)$. Profits of final good firms are given by

$$\Pi = Y - \sum_\tau \int_{\underline{\sigma}}^{\overline{\sigma}} p_\tau(\sigma) y_\tau(\sigma) d\sigma,$$

and profits of intermediate producers of task (σ, τ) are

$$\Pi_\tau(\sigma) = p_\tau(\sigma) y_\tau(\sigma) - r k_\tau(\sigma) - \int_{\underline{s}}^{\overline{s}} w(s) n_\tau(s, \sigma) ds$$

where r is the rental rate of capital and $w(s)$ is the wage paid to a worker with skill s. Design and training costs are included in intermediate producers' profits in the sense that for each unit of capital or labor hired, a fraction may be lost in design or training, as captured by (1.1), (1.2), and (1.3).

As in Costinot and Vogel (2010), a competitive equilibrium is defined as an assignment of factors to tasks such that all firms maximize profits and markets clear. Profit-maximizing task demand by final good producers is

$$y_\tau(\sigma) = \frac{\beta_\tau}{\mu} \frac{Y}{p_\tau(\sigma)}. \tag{1.7}$$

[19] An overview of recent developments in basic robotics research can be found in Nourbakhsh (2013).

Profit maximization by intermediates producers implies

$$
\begin{array}{rcll}
p_\tau(\sigma)\alpha^N(s,\sigma,\tau) & \leq & w(s) & \forall s \in [\underline{s},\overline{s}], \\
p_\tau(\sigma)\alpha^K(s_K,\sigma) & \leq & r/A_K; & \\
p_\tau(\sigma)\alpha^N(s,\sigma,\tau) & = & w(s) & \textit{if } \ n_\tau(s,\sigma) > 0, \\
p_\tau(\sigma)\alpha^K(s_K,\sigma) & = & r/A_K & \textit{if } k_\tau(\sigma) > 0.
\end{array}
\tag{1.8}
$$

Denote the economy's set of tasks by $T \equiv [\underline{\sigma},\overline{\sigma}]\times\{0,1\}$. Formally, a *competitive equilibrium* in this economy is a set of functions $y : T \to \mathbb{R}^+$ (task output); $k : T \to \mathbb{R}^+$ and $n : [\underline{s},\overline{s}] \times T \to \mathbb{R}^+$ (factor assignment); $p : T \to \mathbb{R}^+$ (task prices); $w : [\underline{s},\overline{s}] \to \mathbb{R}^+$ (wages); and a real number r (rental rate of capital) such that conditions (1.1) to (1.8) hold.

To be able to characterize the competitive equilibrium, we need to examine the properties of the productivity schedules α^K and α^N. In training-intensive tasks, workers face the same productivity schedule as machines, except for the skill parameter. Let $\breve{s} \in s_K \cup [\underline{s},\overline{s}]$ and define

$$
\alpha(\breve{s},\sigma) \equiv 1 - \sigma/\breve{s} \quad \left(\equiv \alpha^K(\breve{s},\sigma) \equiv \alpha^N(\breve{s},\sigma,\tau)\right). \tag{1.9}
$$

Note that $\alpha \in (0,1)$. Furthermore, $\alpha_\sigma < 0$ and $\alpha_{\breve{s}} > 0$. Productivity declines in complexity since a larger design or training expense is incurred. Higher skilled factors are more productive since they incur a smaller design or training expense. To characterize comparative advantage, we rely on the following result.

Lemma 1 *The productivity schedule $\alpha(\breve{s},\sigma)$ is strictly log-supermodular.*

The log-supermodularity of the productivity schedule implies that in training-intensive tasks, factors with higher skill have a comparative advantage in more complex tasks, or

$$
\breve{s}' > \breve{s}, \sigma' > \sigma \quad \Leftrightarrow \quad \frac{\alpha(\breve{s}',\sigma')}{\alpha(\breve{s},\sigma')} > \frac{\alpha(\breve{s}',\sigma)}{\alpha(\breve{s},\sigma)}.
$$

For instance, high skill workers have a comparative advantage over low skill workers in more complex tasks; all workers with $s > s_K$ have a comparative advantage over machines in more complex tasks; and so on. The result is due to the fact that for higher skilled factors, training or design expenses increase less steeply with complexity.

Comparative advantage properties regarding training intensity are straightforward. Since α is increasing in $\breve{s}$, and because all workers have productivity one in all innate ability tasks, high skill workers have a comparative advantage over low skill workers in any training-intensive task. Furthermore, because machine productivity is the same in innate ability tasks as in training-intensive tasks if complexity is held constant, it follows that machines have a comparative advantage over all workers in any training-intensive task relative to the innate ability task with the same complexity. This seemingly trivial result has profound implications for the assignment of factors to tasks, and for the reallocation of factors in response to a fall in c_K (a rise in s_K). It is at the root of the job polarization phenomenon, as we will show in Section 1.3 below.

The equilibrium assignment of factors to tasks is determined by comparative advantage, which is a consequence of the zero-profit condition (1.8).[20] Because high skill workers have a comparative advantage in training-intensive tasks (holding complexity constant), in equilibrium the labor force is divided into a group of low skill workers performing innate ability tasks, and a group of high skill workers carrying out training-intensive tasks: there exists a marginal worker with skill s^*, the least-skilled worker employed in training-intensive tasks. This is formally stated in part (a) of Lemma 2 below.

We focus on the empirically relevant case in which machines as well as workers perform both training-intensive and innate ability tasks.[21] In this case, machines are assigned to a subset of innate ability and training-intensive tasks that are relatively less complex, while low skill workers perform the remaining innate ability tasks: there is a threshold task σ_0^*, the marginal innate ability tasks, dividing the set of innate ability tasks into those performed by machines ($\sigma \leq \sigma_0^*$) and those carried out by low skill workers ($\sigma \geq \sigma_0^*$). Similarly, there is a marginal training-intensive task σ_1^* that divides the set of training-intensive tasks into those performed by machines ($\sigma \leq \sigma_1^*$) and those carried out by high skill workers ($\sigma \geq \sigma_1^*$). As in the case of the marginal worker, existence of these marginal tasks is of course a consequence of the comparative advantage properties discussed at the end of Section 1.2.3. These properties also imply $\sigma_0^* < \sigma_1^*$: the marginal training-intensive task is always more complex than the marginal innate ability task (recall that machines are relatively more productive in training-intensive tasks than workers, holding complexity constant); and $s^* > s_K$: it is always cheaper to train (though not to employ) the marginal worker than to design a machine in any task. These results are formally stated in part (b) of Lemma 2. An illustration of the equilibrium assignment is given in Figure 1.1.

Lemma 2 (a) *In a competitive equilibrium, there exists an* $s^* \in (\underline{s}, \overline{s}]$ *such that*

- $n_0(s,\sigma) > 0$ *for some* σ *if and only if* $s \leq s^*$*, and*
- $n_1(s,\sigma) > 0$ *for some* σ *if and only if* $s \geq s^*$.

(b) *If* $k_0(\sigma) > 0$ *for some* σ*, then* $s^* > s_K$*, and there exist* $\sigma_0^*, \sigma_1^* \in [\underline{\sigma}, \overline{\sigma}]$ *with* $\sigma_0^* < \sigma_1^*$ *such that*

- $k_0(\sigma) > 0$ *if and only if* $\sigma \leq \sigma_0^*$*;*
- $k_1(\sigma) > 0$ *if and only if* $\sigma \leq \sigma_1^*$*;*
- $n_0(s,\sigma) > 0$ *if and only if* $s \leq s^*$ *and* $\sigma \geq \sigma_0^*$*; and*

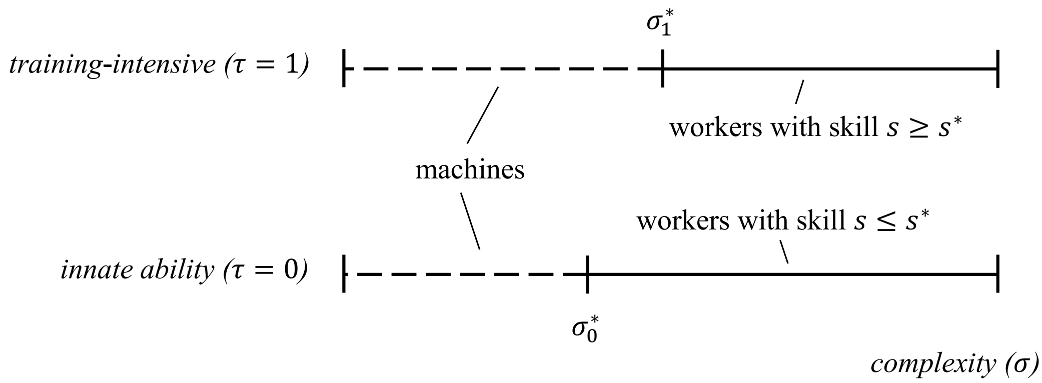
[20] To see how comparative advantage determines patterns of specialization, consider two firms, one producing training-intensive task σ, the other producing training-intensive task σ'. Suppose in equilibrium, firm σ is matched with workers of type s and firm σ' is matched with workers of type s'. Then (1.8) implies

$$\frac{\alpha(s',\sigma')}{\alpha(s,\sigma')} \geq \frac{\alpha(s',\sigma)}{\alpha(s,\sigma)},$$

which shows that type s (s') has a comparative advantage in task σ (σ'), precisely the task to which she was assumed to be matched.

[21] Sufficient conditions for the existence of such an equilibrium are derived Appendix 1.D.1. We assume throughout that these conditions are satisfied. We note however that in general, it may happen that machines do not perform any innate ability tasks, and/or that workers do not carry out any training-intensive tasks.

Figure 1.1: Assignment of Labor and Capital to Tasks.



- $n_1(s, \sigma) > 0$ *if and only if* $s \geq s^*$ *and* $\sigma \geq \sigma_1^*$.

It remains to determine the assignment of low skill workers ($s \leq s^*$) to innate ability tasks ($\tau = 0, \sigma \geq \sigma_0^*$) and that of high skill workers ($s \geq s^*$) to training-intensive tasks ($\tau = 1, \sigma \geq \sigma_1^*$). The solution to the matching problem in innate ability tasks is indeterminate as all workers are equally productive in these tasks. However, knowledge of the assignment is not necessary to pin down task output and prices, as shown below. High skill workers are assigned to training-intensive tasks according to comparative advantage, with higher skilled workers carrying out more complex tasks. Formally, we have:

Lemma 3 *In a competitive equilibrium, if* $s^* < \overline{s}$, *there exists a continuous and strictly increasing matching function* $M : [s^*, \overline{s}] \rightarrow [\sigma_1^*, \overline{\sigma}]$ *such that* $n_1(s, \sigma) > 0$ *if and only if* $M(s) = \sigma$. *Furthermore,* $M(s^*) = \sigma_1^*$ *and* $M(\overline{s}) = \overline{\sigma}$.

This result is an application of Costinot and Vogel (2010), with the added complication that domain and range of the matching function are determined by the endogenous variables s^* and σ_1^*. The matching function is characterized by a system of differential equations. Using arguments along the lines of the proof of Lemma 2 in Costinot and Vogel (2010), it can be shown that the matching function satisfies

$$M'(s) = \frac{\mu}{\beta_1} \frac{w(s)v(s)}{Y}, \tag{1.10}$$

and that the wage schedule is given by

$$\frac{d \log w(s)}{ds} = \frac{\partial \log \alpha(s, M(s))}{\partial s}. \tag{1.11}$$

The last equation is due to the fact that in equilibrium, a firm producing training-intensive task σ chooses worker skill s to minimize marginal cost $w(s)/\alpha(s, \sigma)$. Once differentiability of the matching function has been established, (1.10) can easily be derived from the market clearing condition (1.5) given Lemma 2, and using (1.7) and (1.8).[22] Figure 1.2 illustrates how the matching function assigns workers to training-intensive tasks.

In order to characterize the equilibrium more fully, and for comparative statics exercises, it is necessary to derive equations pinning down the endogenous variables σ_0^*, σ_1^*, and s^*. These equations are due to a set of no-arbitrage conditions. In particular, firms producing the marginal tasks are indifferent between hiring labor or capital, and the marginal worker is indifferent between performing innate ability tasks or the marginal training-intensive tasks. Formally, the price and wage functions must be continuous, otherwise the zero-profit condition (1.8) could not hold. This is a well-known result in the literature on comparative-advantage-based assignment models. Hence, the no-arbitrage conditions for the marginal tasks are

$$\frac{r}{A_K \alpha(s_K, \sigma_0^*)} = w(s) \quad \textit{for all } s < s^* \tag{1.13}$$

and

$$\frac{r}{A_K \alpha(s_K, \sigma_1^*)} = \frac{w(s^*)}{\alpha(s^*, \sigma_1^*)}, \tag{1.14}$$

and the no-arbitrage condition for the marginal worker is

$$w(s) = w(s^*) \quad \textit{for all } s \leq s^*. \tag{1.15}$$

The last result implies that there is a mass point at the lower end of the wage distribution. The mass point is a result of normalizing A, the amount of problems drawn, to one for all workers. To avoid the mass point, we could instead assume that $A \equiv A(s)$ with $A'(s) \geq 0$. Equilibrium assignment and comparative statics results would be qualitatively the same. We maintain the normalization to avoid additional notation.

We can now complete the characterization of a competitive equilibrium by eliminating factor prices from (1.14). A standard implication of the Cobb-Douglas production function is that the mass of capital allocated to each task is constant within innate ability tasks and within training-intensive

[22] In particular, Lemma 2 and (1.5) imply

$$\int_{s^*}^{s} v(s')ds' = \int_{\sigma_1^*}^{\sigma} n_1(M^{-1}(\sigma'), \sigma')d\sigma'.$$

Changing variables on the RHS of the last expression and differentiating with respect to s yields

$$v(s) = n_1(s, M(s))M'(s),$$

and substituting (1.3) we obtain

$$M'(s) = \frac{\alpha(s, M(s))v(s)}{y(M(s))}. \tag{1.12}$$

After eliminating task output and price using (1.7) and (1.8), (1.10) follows.

tasks. Some algebra shows[23] that machines produce task outputs

$$y_\tau(\sigma) = \frac{\beta_\tau A_K \alpha(s_K, \sigma) K}{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})} \quad \textit{for all } \sigma \in [\underline{\sigma}, \sigma_\tau^*]. \tag{1.16}$$

Using these equations to solve for the task prices in (1.7), and plugging the obtained expression into (1.8), yields

$$r = \frac{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})}{\mu} \times \frac{Y}{K}. \tag{1.17}$$

This is of course the familiar result that with a Cobb-Douglas production function, factor prices equal the factor's share in output times total output per factor unit. In this case, the factor share is endogenously given by the (weighted) share of tasks to which the factor is assigned.

We employ similar steps to solve for $w(s^*)$. Since in innate ability tasks, worker productivity does not vary across tasks nor types, all innate ability tasks with $\sigma \geq \sigma_0^*$ have the same price and all workers with $s < s^*$ earn a constant wage equal to $w(s^*)$ (as a result of the no-arbitrage condition for the marginal worker). As prices do not vary, neither does output, and so by the market clearing conditions (1.3) and (1.5),[24]

$$y_0(\sigma) = \frac{V(s^*)}{\overline{\sigma} - \sigma_0^*} \quad \textit{for all } \sigma \geq \sigma_0^*. \tag{1.18}$$

Proceeding as above when solving for r, we obtain

$$w(s^*) = \frac{\beta_0(\overline{\sigma} - \sigma_0^*)}{\mu} \times \frac{Y}{V(s^*)}. \tag{1.19}$$

With (1.17) and (1.19) in hand, we can eliminate factor prices from the marginal cost equaliza-

[23] By (1.7) and (1.8), we have

$$\frac{y_\tau(\sigma)}{y_\tau(\sigma')} = \frac{\alpha(s_K, \sigma)}{\alpha(s_K, \sigma')}, \qquad \frac{y_0(\widetilde{\sigma})}{y_1(\widetilde{\sigma}')} = \frac{\beta_0}{\beta_1} \frac{\alpha(s_K, \widetilde{\sigma})}{\alpha(s_K, \widetilde{\sigma}')}$$

for any tasks $(\sigma, \sigma', \widetilde{\sigma}, \widetilde{\sigma}')$ performed by machines. But (1.3), (1.3), and Lemma 2 imply

$$\frac{y_\tau(\sigma)}{y_\tau(\sigma')} = \frac{\alpha(s_K, \sigma) k_\tau(\sigma)}{\alpha(s_K, \sigma') k_\tau(\sigma')}, \qquad \frac{y_0(\widetilde{\sigma})}{y_1(\widetilde{\sigma}')} = \frac{\alpha(s_K, \widetilde{\sigma}) k_0(\widetilde{\sigma})}{\alpha(s_K, \widetilde{\sigma}') k_0(\widetilde{\sigma}')}.$$

The previous two equations together give $k_\tau(\sigma) = k_\tau(\sigma')$ and $k_0(\widetilde{\sigma}) = \frac{\beta_0}{\beta_1} k_1(\widetilde{\sigma}')$. By (1.6) and Lemma 2,

$$k_\tau(\sigma) = \frac{\beta_\tau K}{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})} \quad \textit{for all } \sigma \in [\underline{\sigma}, \sigma_0^*].$$

[24] Under Lemma 2, integrating (1.5) yields

$$V(s^*) = \int_{\sigma_0^*}^{\overline{\sigma}} \int_{\underline{s}}^{s^*} n_0(s, \sigma) ds d\sigma,$$

but using (1.3) and the fact that task output is a constant y_0 results in

$$V(s^*) = (\overline{\sigma} - \sigma_0^*) y_0.$$

tion condition (1.13) to obtain

$$\frac{A_K \alpha(s_K, \sigma_0^*) K}{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})} = \frac{V(s^*)}{\beta_0(\overline{\sigma} - \sigma_0^*)}. \tag{1.20}$$

Also, combining conditions (1.13) to (1.15) yields

$$\alpha(s_K, \sigma_1^*) = \alpha(s_K, \sigma_0^*)\alpha(s^*, \sigma_1^*). \tag{1.21}$$

Lastly, (1.10) and (1.19) imply

$$M'(s^*) = \frac{\beta_0(\overline{\sigma} - \sigma_0^*)}{\beta_1} \frac{v(s^*)}{V(s^*)}. \tag{1.22}$$

Equations (1.4), (1.10), (1.11), (1.20), (1.21), and (1.22) together with the boundary conditions $M(s^*) = \sigma_1^*$ and $M(\overline{s}) = \overline{\sigma}$, uniquely pin down the equilibrium objects σ_0^*, σ_1^*, s^*, w, and M. The comparative statics analysis makes extensive use of these expressions.

To conclude this section, we highlight two properties of the wage structure in our model. First, integrating (1.11) yields an expression for the wage differential between any two skill types that are both employed in training-intensive tasks,

$$\frac{w(s')}{w(s)} = \exp\left[\int_s^{s'} \frac{\partial}{\partial z} \log \alpha(z, M(z)) dz\right] \quad \textit{for all } s' \geq s \geq s^*. \tag{1.23}$$

This shows that wage inequality is fully characterized by the matching function (Sampson 2012). Second, adding (1.10) and (1.19) and integrating yields an expression for the average wage,

$$\mathrm{E}[w] = \frac{\beta_0(\overline{\sigma} - \sigma_0^*) + \beta_1(\overline{\sigma} - \sigma_1^*)}{\mu} \times Y. \tag{1.24}$$

Since the labor force is normalized to have measure one, this expression also gives the total wage bill. It follows that the labor share in the model is given by the (weighted) share of tasks performed by workers.

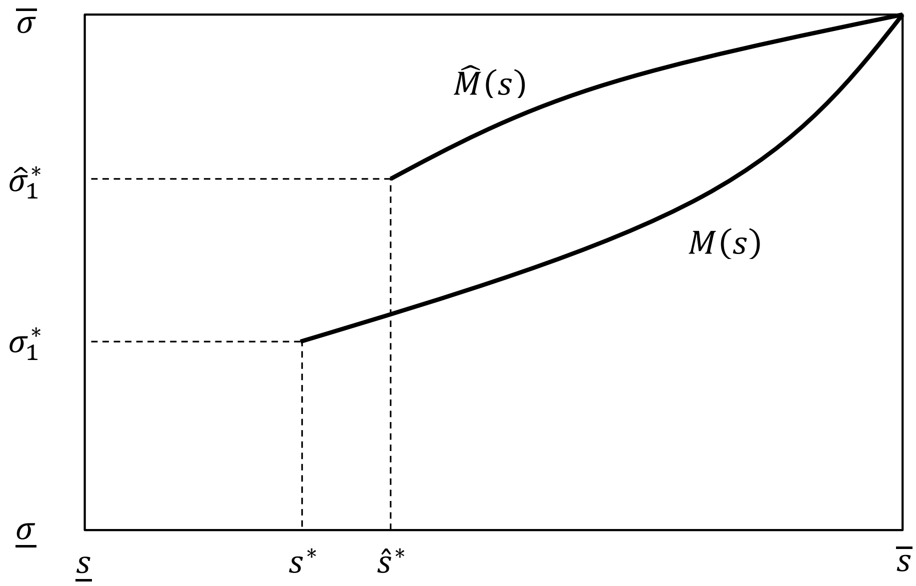
1.3 Comparative Statics

Having outlined the model and characterized its equilibrium in the previous section, we now move on to comparative statics exercises. Our main interest is in investigating the effects of a fall in the machine design cost, c_K. In addition we will analyze the effects of increased skill abundance, motivated by the large increase in relative skill endowments seen in developed countries over the previous decades.

1.3.1 Technical Change

Consider a fall in the machine design cost from c_K to $\widehat{c}_K$, so that $\widehat{s}_K > s_K$. Let M and $\widehat{M}$ be the corresponding matching functions, and similarly for σ_0^* and $\widehat{\sigma}_0^*$; σ_1^* and $\widehat{\sigma}_1^*$; and s^* and $\widehat{s}^*$. We

Figure 1.2: Assignment of Workers to Training-Intensive Tasks and the Effects of Technical Change



Complexity σ is plotted on the vertical axis, while skill level s is plotted on the horizontal axis. The upward shift of the matching function and the shift of its lower end to the northeast are brought about by a fall in the machine cost from c_K to $\widehat{c}_K$ as stated in Proposition 1.

now state the main result of the chapter.

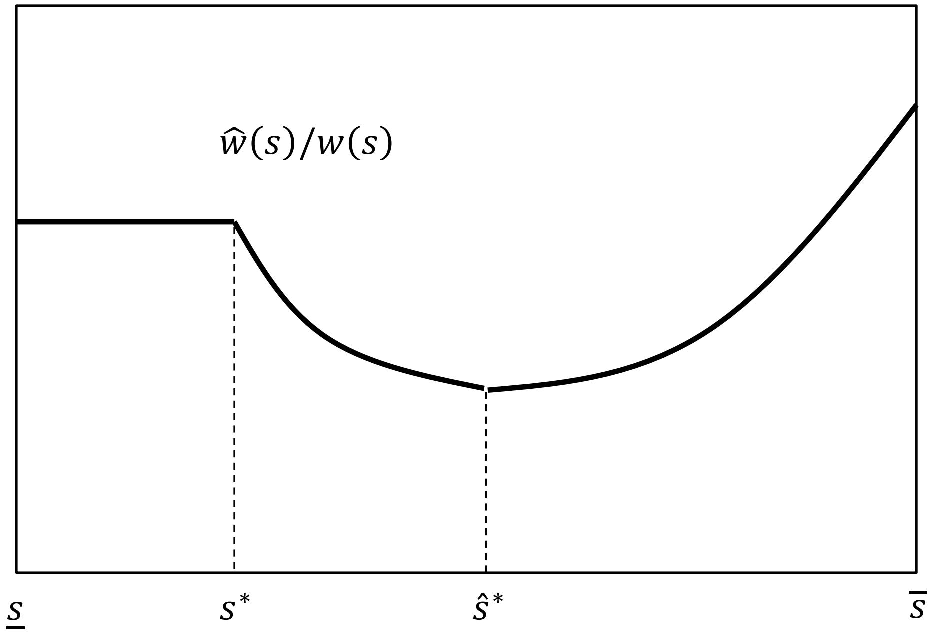
Proposition 1 *Suppose the machine design cost falls, $\widehat{c}_K < c_K$ and so $\widehat{s}_K > s_K$. Then the marginal training-intensive task becomes more complex, $\widehat{\sigma}_1^* > \sigma_1^*$, and the matching function shifts up, $\widehat{M}(s) > M(s)$ for all $s \in [\max\{s^*, \widehat{s}^*\}, \overline{s})$. If the fall in the machine design cost is such that $\widehat{s}_K \geq s^*$, then the marginal worker becomes more skilled, $\widehat{s}^* > s^*$.*

A fall in the machine design cost implies a rise in machine productivity and thus a fall in the marginal cost of employing machines in any task. Crucially, the marginal cost of employing machines in the threshold training-intensive tasks falls by more than the marginal cost in the threshold innate ability task, since $\sigma_0^* < \sigma_1^*$.[25] This means that machine employment in training-intensive tasks increases by more than in innate ability tasks. In fact, numerical simulations suggest that the effect of a fall in c_K on σ_0^* is ambiguous.

As machines are newly adopted in a subset of training-intensive tasks, the workers initially performing these tasks get replaced. Some of these workers upgrade to more complex training-intensive tasks—the matching function shifts up. Others downgrade to innate ability tasks, at least for a sufficiently large fall in the machine design cost. Thus, labor-replacing technical change causes job polarization. These effects are illustrated by Figure 1.2.

[25]Because $\sigma_0^* < \sigma_0^*$ and due to the log-supermodularity of α, the ratio $\alpha(s_K, \sigma_1^*)/\alpha(s_K, \sigma_0^*)$ is increasing in s_K.

Figure 1.3: Changes in Wages as a Result of a Fall in the Machine Design Cost from c_K to $\widehat{c}_K$.



For each skill level s, the ratio of new to old wages is plotted. Workers with $s \in [\widehat{s}^*, \overline{s}]$ remain in training-intensive tasks and experience a rise in the skill premium. Workers with $s \in [s^*, \widehat{s}^*)$ switch to innate ability tasks and experience a fall in the skill premium. See Corollary 1 for details.

While we are able to show that the threshold training-intensive task always becomes more complex and the matching function always shifts up, we are unable to rule out $\widehat{s}^* \leq s^*$ for small decreases in the machine design cost. However, we can prove that if machine design costs fall steadily over time, so that $\widehat{s}_K \geq s^*$ at some point, then the skill cutoff level must rise eventually. Thus, we limit our attention to the case in which a fall in c_K triggers a rise in s^* and hence job polarization occurs.

The matching function is a sufficient statistic for inequality (Sampson 2012), so that the shift in the matching function contains all the required information for deriving changes in relative wages among workers who remain in training-intensive tasks. Intuitively, as the upward shift implies skill downgrading by firms (but task upgrading for workers), the zero profit conditions imply that relatively low skill workers must have become relatively cheaper, or else their new employers would not be willing to absorb them. Hence the skill premium goes up for workers remaining in training-intensive tasks. Similar reasoning implies that workers who moved to innate ability tasks now earn relatively less than workers who were already performing these tasks. In sum, middle-skill workers who are displaced by machines experience downward pressure on their wages, and they end up worse off in relative terms compared to high and low skill workers who are less affected by automation (or not at all). Wage inequality rises at the top, but falls at the bottom of the distribution, as illustrated by Figure 1.3. The formal result is as follows.

Corollary 1 *Suppose $\widehat{c}_K < c_K$ and consider the case in which $\widehat{s}^* > s^*$. Wage inequality increases at the top of the distribution but decreases at the bottom. Formally,*

$$\frac{\widehat{w}(s')}{\widehat{w}(s)} > \frac{w(s')}{w(s)} \quad \text{for all } s' > s \geq \widehat{s}^*$$

and

$$\frac{\widehat{w}(s')}{\widehat{w}(s)} < \frac{w(s')}{w(s)} \quad \text{for all } s', s \text{ such that } \widehat{s}^* > s' > s \geq s^*.$$

Relative wages are affected by technical change despite the fact that all factors are perfect substitutes at the task level. This is because tasks are q-complements in the production of the final good.[26] Intuitively, firms respond in two to the fall in the design cost. First, they upgrade existing machines. Second, they adopt machines in tasks previously performed by workers. The first effect on its own would lead to a rise in wages for all workers, because the increase in machines' task output raises the marginal product of all other tasks; moreover, relative wages would remain unchanged. The second effect, however, forces some workers to move to different tasks, putting downward pressure on their wages.[27] Since middle skill workers are most likely to be displaced by increased automation, their wages relative to low skill and high skill workers will decline.[28] Thus, whether technology substitutes for or complements a worker of given skill type (in terms of relative wage effects) depends on that worker's exposure to automation, which is endogenous in our model.

To map the model's predictions for changes in wage inequality to the data, following Costinot and Vogel (2010) it is useful to distinguish between observable and unobservable skills. In particular, our continuous skill index s is unlikely to be observed by the econometrician. Instead, we assume that the labor force is partitioned according to some observable attribute e, which takes on a finite number of values and may index education or experience. Suppose further that high-s workers are disproportionately found in high-e groups. Formally, if $s' > s$ and $e' > e$, we require $v(s', e')v(s, e) \geq v(s, e')v(s', e)$. Costinot and Vogel (2010) show that an increase in wage inequality in the sense of Corollary 1 implies an increase in the premium paid to high-e workers as well as an increase in wage inequality among workers with the same e. In other words, the model predicts that if the machine design cost falls, both between and within (or residual) wage inequality will rise for the fraction of workers remaining in training-intensive tasks. On the other hand, within and between inequality falls for the set of workers below the new cutoff. This group includes stayers in as well as movers to innate ability tasks. In Section 1.5.1, we review existing evidence that is consistent with this prediction.

Although the effect on the marginal innate ability task is uncertain, the overall weighted share of tasks performed by machines increases. By (1.24), this is equivalent to a decrease in the labor share.

Corollary 2 *Suppose $\widehat{c}_K < c_K$. The labor share decreases.*

[26] This means that the price of a task increases in the output of all other task. The mechanism described in this paragraph has been highlighted by Acemoglu and Autor (2011).

[27] The effect works mainly through changes in task prices. Physical productivity actually increases for middle skill workers who get reassigned to innate ability tasks.

[28] But middle skill workers' wages will not decline absolutely if the first effect dominates.

The fact that the labor share decreases means that it is not possible to sign the effect of a fall in the design cost on wage levels. Equation (1.24) shows that the average wage is affected both by the decrease in workers' task shares and the increase in output, so that the overall change is ambiguous. Of course, wage levels may also differentially change by worker type. For instance, high and low skill workers may enjoy absolute wage gains, while middle skill workers may suffer absolute wage losses.

1.3.2 Increase in Skill Abundance

Now consider an increase in the relative supply of skills. Following Costinot and Vogel (2010), we say that $\widehat{V}$ is more skill abundant relative to V, or $\widehat{V} \succeq V$, if

$$\widehat{v}(s')v(s) \geq \widehat{v}(s)v(s') \quad \textit{for all } s' > s.$$

Such a shift in the skill distribution implies first-order stochastic dominance and hence an increase in the mean. The shift may be due, for instance, to an increase in average education levels, to the extent that general education leads to attainment of knowledge applicable to a wide range of job tasks. If so, then training costs for a given level of complexity will fall on average, exactly as occurs if $\mathrm{E}[s]$ decreases.

For simplicity, we restrict attention to distributions with common support, and we assume that $\widehat{v}(\overline{s}) > v(\overline{s})$. Characterizing comparative statics for changes in skill supplies is more challenging in our model than in the original Costinot-Vogel framework because domain and range of the matching function are endogenous. We are able to offer a partial result.

Proposition 2 *Suppose that skill becomes more abundant, $\widehat{V} \succeq V$ and $\widehat{v}(\overline{s}) > v(\overline{s})$. If this change in skill endowments induces an increase in the share of income accruing to labor, then the marginal training-intensive task becomes less complex, $\widehat{\sigma}_1^* < \sigma_1^*$; the marginal worker becomes more skilled, $\widehat{s}^* > s^*$; and the matching function shifts down, $\widehat{M}(s) < M(s)$ for all $s \in [\widehat{s}^*, \overline{s})$.*

Intuitively, such a change to the distribution of skills should raise the labor share, because the labor share equals the share of tasks performed by workers, and an increase in the average worker's productivity should induce more firms to hire labor. While the labor share always increases in our numerical simulations, we are unable to prove the general result.[29]

The implications of Proposition 2 are as follows. Firms take advantage of the increased supply of skilled workers and engage in skill upgrading, which is equivalent to task downgrading for workers. This can be seen for training-intensive tasks by the downward shift of the matching function. For innate ability tasks, skill-upgrading is equivalent to the marginal worker becoming more skilled. Skill upgrading implies that the price of skill must have declined, so that the distribution of wages becomes more equal.

[29] The labor share is given by $\int_{\underline{s}}^{\overline{s}} \frac{w(s)}{Y} dV(s)$. Because $\widehat{V}$ first-order stochastically dominates V and $w(s)/Y$ is an increasing function, we have $\int_{\underline{s}}^{\overline{s}} \frac{w(s)}{Y} d\widehat{V}(s) > \int_{\underline{s}}^{\overline{s}} \frac{w(s)}{Y} dV(s)$. Thus, for the labor share to *decrease*, there would need to be a sufficiently large decline in wage-output ratios for a subset of workers.

Corollary 3 *Suppose $\widehat{V} \succeq V$, and that the labor share increases as a result. Then wage inequality decreases globally: for all s, s' with $s' > s \geq s^*$,*

$$\frac{\widehat{w}(s)}{\widehat{w}(s')} > \frac{w(s)}{w(s')}.$$

The formal result implies a fall in both within and between inequality. This is consistent with a fall in the college premium induced by an increase in the supply of college educated workers as occurred in the US in the 1970s (Acemoglu 2002). Thus, our model features a modified version of the “Race between Education and Technology” (Goldin and Katz 2008), in the sense that education and technology have opposite effects on wage dispersion in the upper part of the distribution, but not in the lower part.

Proposition 2 says that the marginal training-intensive tasks becomes less complex, implying a decline in technology use for such tasks. In contrast, our simulations show that the marginal innate ability task becomes more complex. Thus, skill upgrading appears to coincide with technology being more (less) widely adopted in innate ability (training-intensive) tasks.

Equation (1.24) implies that changes in wages are in general not proportional to changes in GDP in this economy, unless task assignment is constant. By Corollary 2, technical change reduces the labor share, while numerical simulations suggest that an increase in skill supplies leads to an increase in the labor share. Thus, if both forces are at work, the labor share may not exhibit any trend. We return to this issue when discussing existing empirical findings in Section 1.5.1.

1.4 Extensions

1.4.1 Making the Model Dynamic

Up to this point we have treated the economy’s capital stock as exogenously given. To determine how endogenous capital accumulation would affect our comparative statics results, we assume that in the long run, the rental rate of capital is a constant pinned down by a time preference parameter[30] and that machines fully depreciate in every period. Furthermore, we assume that worker’s knowledge depreciates fully in every period, or equivalently, there is an overlapping generations structure with each generation only working for one period. Suppose that the economy starts out in a steady state with the interest rate equal to its long-run value. Now recall that a fall in the machine design cost leads to a rise in the labor share. Furthermore, because the First Welfare Theorem applies to our model economy, output must not decrease, since the economy’s resource constraint is less tight. By (1.17), we have that the interest rate increases. Thus, in the long run, the capital stock must increase to bring the interest rate back down.

It can be shown that a rise in the capital stock K has qualitatively the same effects on the marginal tasks, the matching function, and wages, as a fall in the machine design cost c_K.[31] This

[30] Alternatively, we could assume that the economy is open to world capital markets, where it is a price taker.

[31] The proof is along similar lines as the proof of Proposition 1 and is available upon request. Since task-neutral machine productivity A_K enters the relevant model equations in the same way as K, the statement also applies to an increase in A_K.

is because a higher supply of capital makes it cheaper to rent machines and thus encourages technology adoption. Thus, our predictions about the effects of a fall in c_K are not overturned with endogenous capital accumulation. In fact, the rise in the marginal training-intensive task, the upward shift of the matching function, the rise in the skill of the marginal worker, and the increase in wage inequality will be more pronounced in the long run as a result of the higher capital stock.

1.4.2 A Model with Fixed Costs

Our baseline model emphasizes that when a firm automates its production, total costs will generally be increasing in the firm's output and in the complexity of the processes required for production. While this in itself should be uncontroversial, our focus on variable costs with the implication of constant returns to scale is certainly restrictive. In particular, firms usually face large one-off expenses when installing new machinery.[32] While such expenses would generally depend on the scale at which the firm plans to operate, it is useful to consider the extreme case of a fixed setup cost.

In Appendix 1.C we modify our baseline model such that firms wanting to automate production face a fixed cost (in units of the final good) which is increasing in the complexity of the task, but does not depend on the scale of production. We derive conditions ensuring an equilibrium assignment that is qualitatively the same as the one analyzed for the baseline model (see Figure 1.1). In particular, the marginal cost of using a machine must be sufficiently small, which can be achieved by making A_K very large, a realistic assumption; and the fixed cost must increase sufficiently in complexity. The model is much less tractable than the baseline model, and we are unable to derive general comparative statics results. Intuitively, when the fixed machine design cost falls, there is an incentive for firms to adopt machines in more-complex tasks. This incentive is stronger in training-intensive tasks: as complexity increases, the marginal cost of employing labor increases in training-intensive tasks but not in innate ability tasks. Thus, we would expect to see an increase in the share of workers performing innate-ability tasks. We have solved the model numerically and verified this intuition. We present results in Appendix 1.C.

1.5 Empirical Support for the Model's Predictions

Section 1.3.1 has established that any technological advance that facilitates automation of a wide range of tasks should lead to systematic shifts in task input, job polarization, and a hollowing out of the wage distribution. In addition, the model also predicts which worker types will be replaced as more tasks are automated, and to which task a displaced worker gets reassigned. In this section we briefly review papers that document these patterns for the recent information and communication technology revolution. We then discuss two studies presenting historical evidence that we also find to be consistent with the model's prediction. Finally, we present new evidence consistent with our model's predictions about trends in worker training levels.

[32]For an example relating to recent advances in AI, consider the concept of 'machine learning', where a software requires a considerable amount of initial 'training' before becoming operational.

1.5.1 Existing Evidence

Changes in task input.—In a seminal contribution, Autor, Levy, and Murnane (2003) document a decline in the fraction of workers performing "routine tasks", and show that this decline is larger in industries that more rapidly adopted information technologies. They also find that "non-routine" interactive and analytic task inputs increased, and more so in industries with more rapid ICT adoption. Although routine-ness is conceptually distinct from complexity, ALM's empirical measures of routine-ness may in fact be correlated with it. For example, they classify routine occupations as those that require "finger dexterity" and "adaptability to situations requiring the precise attainment of set limits, tolerances or standards." It is likely that these are occupations with low complexity (though not necessarily low training intensity). The measured shift away from routine tasks is then consistent with our prediction of a reallocation towards more-complex tasks.

Job polarization.—Goos and Manning (2007) were the first to suggest that the "de-routinization" documented by ALM implies a polarization of employment since routine tasks were traditionally performed by middle-skill workers. They do find evidence of job polarization for the UK, and subsequently Autor, Katz, and Kearney (2006) showed this to be the case in the US as well. Goos, Manning, and Salomons (2014) provide evidence for job polarization in a majority of European economies, and show that much of it can be attributed to tasks shifts consistent with technical change being the driving force. Importantly, Michaels, Natraj, and Van Reenen (forthcoming) show that in a sample of several developed countries it is indeed the case that industries that invested more heavily in information and communication technologies witnessed a decline in relative middle skill employment and wage bills, confirming the link between technical change and job polarization.

Cortes (2012) uses panel data from the US and shows that worker ability is a strong determinant of the destination occupation for workers exiting from routine occupations. He shows that low (high) ability workers are more likely to switch to non-routine manual (non-routine) cognitive occupations. This is consistent with our model if we interpret non-routine manual as innate ability tasks and non-routine cognitive as training-intensive and complex tasks.

Wages.—Recall that Corollary 1 implies a fall in wage inequality at the bottom of the distribution and a rise at the top. Consistent with this, Autor, Katz, and Kearney (2006) document that in the US since the late 1980s, wages in the middle of the distribution have risen more slowly than those at the top and bottom. Dickens, Manning, and Butcher (2012) show similar evidence for the UK and argue that the compression of the lower part of the distribution is partly explained by rises in the minimum wage. We interpret this as leaving room for a technological explanation like the one proposed here.

Lemieux (2006a) shows that in the 1990s increases in within-group inequality were concentrated in the upper part of the wage distribution. For between-group wage differentials, Lindley and Machin (2011) document that in addition to a rise in the college premium, there has also been an increase in the wages of workers with a graduate degree relative to those with college only.

Firpo, Fortin, and Lemieux (2011) investigate using US data whether changes in the wage distribution can be attributed to changes in the returns to tasks that are due to technical change or offshoring. They find a prominent role of technology, while offshoring has become more important in the most recent decade. However, their identification assumptions may be viewed as restrictive from the perspective of our model, so that further research is required. Cortes (2012), in addition to providing evidence on worker movements, also shows that relative wages of those workers staying in middle-wage, routine occupations decline. Boehm (2013) uses NLSY data to estimate workers' selection into occupations based on observed comparative advantage. He finds that workers with a comparative advantage in routine occupations saw their wages decline relative to other workers, and even absolutely. Overall, the evidence on wages appears consistent with our model.

Labor share.—Recall that our comparative statics suggest the absence of a trend in the labor share when technical change and increases in skill supply occur simultaneously. However when technical change dominates, the labor share should fall. Rodriguez and Jayadev (2010) document that the labor share started to decline in most countries around 1980. Karabarbounis and Neiman (2014) find that a substantial part of the decline can be explained by advances in ICT, which may be consistent with our model, although the structural model they estimate is quite different from ours. Elsby, Hobijn, and Sahin (2013) do not find a strong role of technology in explaining the decline in the labor share. Thus, there appears to be no consensus on this question yet, and further research is required.

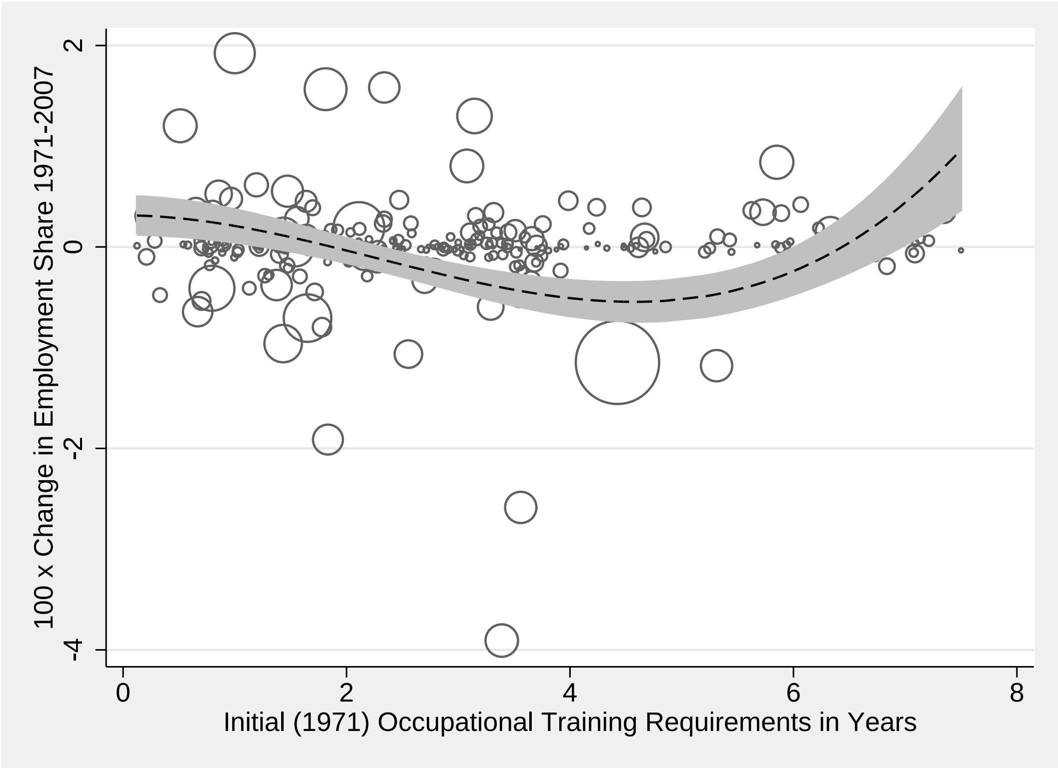
Historical evidence.—Katz and Margo (2013) find that from 1850 to 1880, US manufacturing witnessed a relative decline in middle skill jobs like artisans at the expense of high skill jobs (non-production workers) and low skill jobs (operatives), concurrent with the increased adoption of steam power. Gray (2013) shows that electrification in the US during the first half of the 20th century led to a fall in the demand for dexterity-intensive tasks performed by middle skill workers, relative to manual and clerical tasks performed by low and high skill workers, respectively. These papers provide support for the model's implication that job polarization is not a unique consequence of the ICT revolution.

Bessen (2011) provides evidence on weavers employed at a 19th century Massachusetts firm that gradually increased the degree of mechanization during the period studied. Even though some of workers' skills were no longer needed as more tasks were automated, the tasks to which workers were reassigned required substantial on-the-job learning, much like the reassignment of workers to more-complex, training-intensive tasks in our model. Crucially, worker productivity in the remaining tasks increased, supporting the assumption of q-complementarity of tasks that underlies our model. Note that we would not necessarily expect an aggregate phenomenon like job polarization to occur at the firm level.

1.5.2 Trends in Occupational Training Requirements

In the model, training levels vary systematically with task characteristics. In particular, more complex training-intensive tasks require more training. We view occupations as bundles of tasks, so that a given occupation may combine tasks from across the task space. Measures of occupational characteristics should be informative about which region of the task space features most prominently in a given occupation. Thus, occupations with low training requirements should be intensive in innate ability tasks; and occupations with very high training requirements should feature highly

Figure 1.4: Changes in Occupational Employment Shares by Initial Training Requirements



We calculate training requirements using the variable *specific vocational preparation* (SVP) from the Dictionary of Occupational Titles and the O*NET database. Observations are weighted by average occupational employment shares. Fitted curves are fractional polynomials, drawn using Stata's *fpfitci* option.

complex, training-intensive tasks.

To measure training requirements of occupations, we use the Fourth Edition Dictionary of Occupational Titles (DOT) in combination with the 1971 April Current Population Survey (CPS) (National Academy of Sciences 1981), and the US Department of Labor's O*NET database in combination with the 2008 American Community Survey (ACS). The information in the 2008 ACS refers to the previous year. Hence, our data cover the years 1971 and 2007. Since the 1971 April CPS lacks information on earnings, we also used the IPUMS 1970 census extract which contains earnings data pertaining to 1969.[33] We use David Dorn's three-digit occupation codes throughout (Dorn 2009). Our analysis is based on a sample of all employed persons aged 17 to 65. To see whether our results are driven by changes in composition, we repeated the analysis using a sample of white males only. The results, available upon request, are qualitatively identical.

Both the DOT and O*NET contain the variable *Specific Vocational Preparation* (SVP), which indicates "the amount of time required to learn the techniques, acquire the information, and develop the facility needed for average performance in a specific job-worker situation. SVP includes training

[33]Because we have to merge separate data sets at the three-digit occupation level, we prefer using the census to the much smaller 1971 March CPS for obtaining earnings data.

acquired in a school, work, military, institutional, or vocational environment, but excludes schooling without specific vocational content" (National Academy of Sciences 1981, p.21 in codebook). SVP is a bracketed variable and we use midpoints to convert it into training time measured in years. See Appendix 1.A for details. Tables 1.A.2 and 1.A.3 list the twenty most and least training-intensive occupations in 1971 and 2007, respectively.

The definition of SVP matches our concept of task-specific training more closely than years of education. This is because much of education, at least up to high school graduation, is general in nature and the skills acquired are portable across occupations. Also, the average level of education of workers in a given occupation may be affected by the supply of educated workers independently of actual training requirements—we provide evidence for this below. In professional occupations such as lawyers and physicians there is a clear mapping between years of schooling and training requirements, but in general this is not the case. In terms of our model, we think of general education as affecting the ability to acquire task-specific knowledge. Thus, years of schooling may proxy for s.

The model delivers several predictions about trends in training requirements. First, as a fall in the machine design cost triggers a reallocation of workers towards tasks of higher complexity on the one hand (the upward shift of the matching function) and towards innate ability tasks on the other (the upward shift of the skill cutoff), the model predicts a polarization of employment by initial training requirements. Figure 1.4 plots changes in an occupation's employment share against its initial (1971) training requirements in years. The figure also shows predicted values from a fractional polynomial with 95-percent confidence intervals. The U-shape of the fitted curve is consistent with the model's prediction.

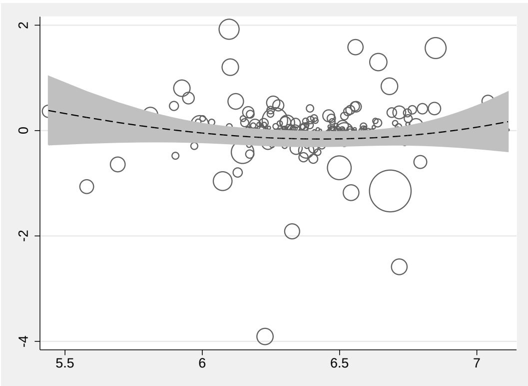
The empirical literature on job polarization usually uses 1980 or 1990 as base years, while our base year is 1971 since the micro data on training requirements is only available for that year. For comparison with the literature, we also plot changes in employment shares by 1971 adjusted mean wages. We obtain adjusted occupational mean log wages as the predicted values from a regression of log wages on occupation dummies, a quartic in potential experience, region dummies, and indicators for female and non-white, evaluated at sample means of covariates other than occupation. Panel (a) of Figure 1.5 shows some suggestive evidence of job polarization going back as far as 1971, although the U-shape is not very pronounced.

However, since wages in the model only differ because of different training requirements across tasks, a more relevant exercise is to plot changes in employment shares against initial wages that are predicted by training requirements. There is a strong positive relationship between mean log wages and training requirements in 1971 ($\beta = 0.11(0.01)$, $R^2 = 0.39$). Hence, job polarization by predicted initial (1971) wages is much more pronounced, as shown in panel (b) of Figure 1.5. In contrast, years of education do not yet predict polarization, but only an increase in employment shares in high-education occupations, as shown in panel (c).

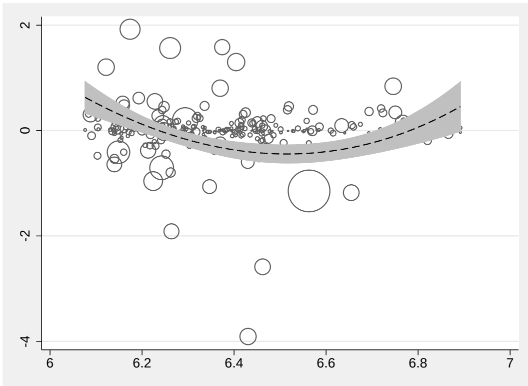
A further prediction of the model can potentially help to make sense of changes in training requirements within occupations. Let us assume that measured training requirements are indicative of the most complex training-intensive task within an occupation's bundle of tasks. Then we would expect training requirements to decrease in occupations with intermediate training requirements,

Figure 1.5: Changes in Occupational Employment Shares by Initial Wages

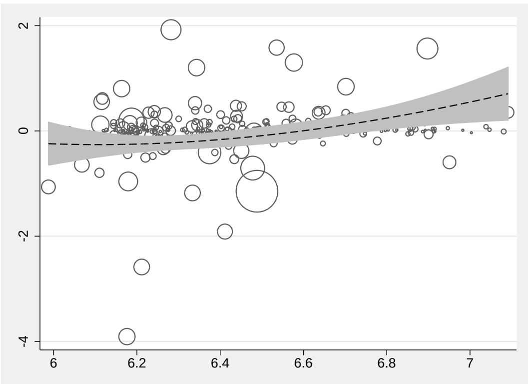
(a) Log wages



(b) Wages predicted by training



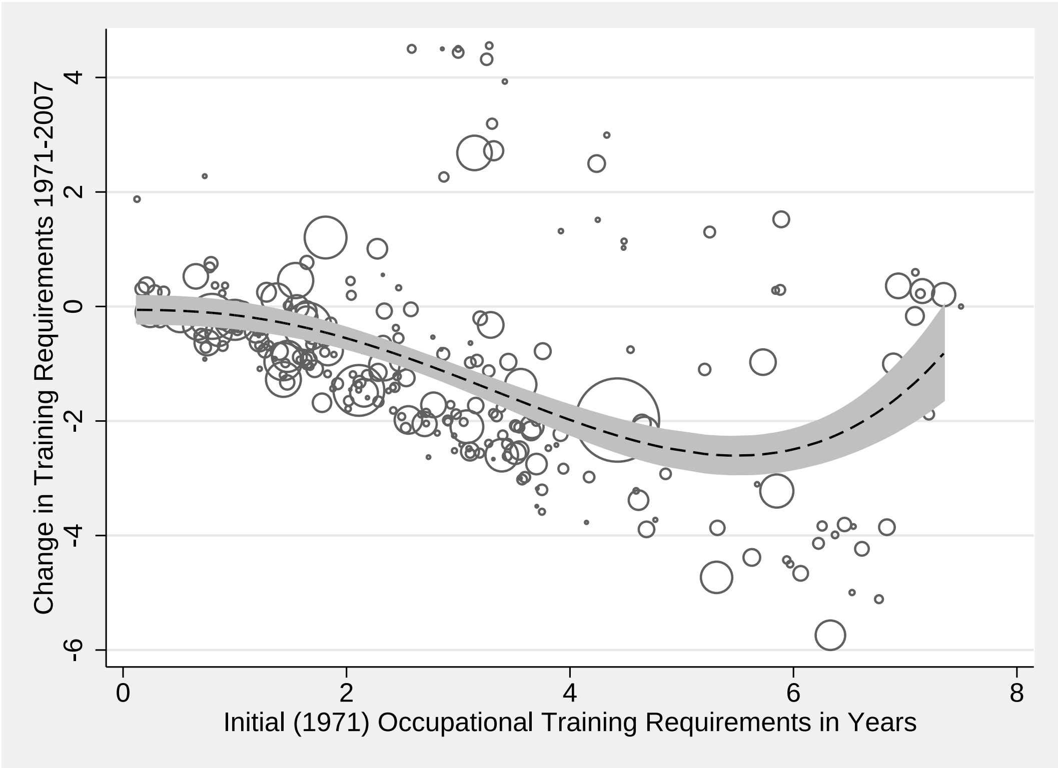
(c) Wages predicted by education



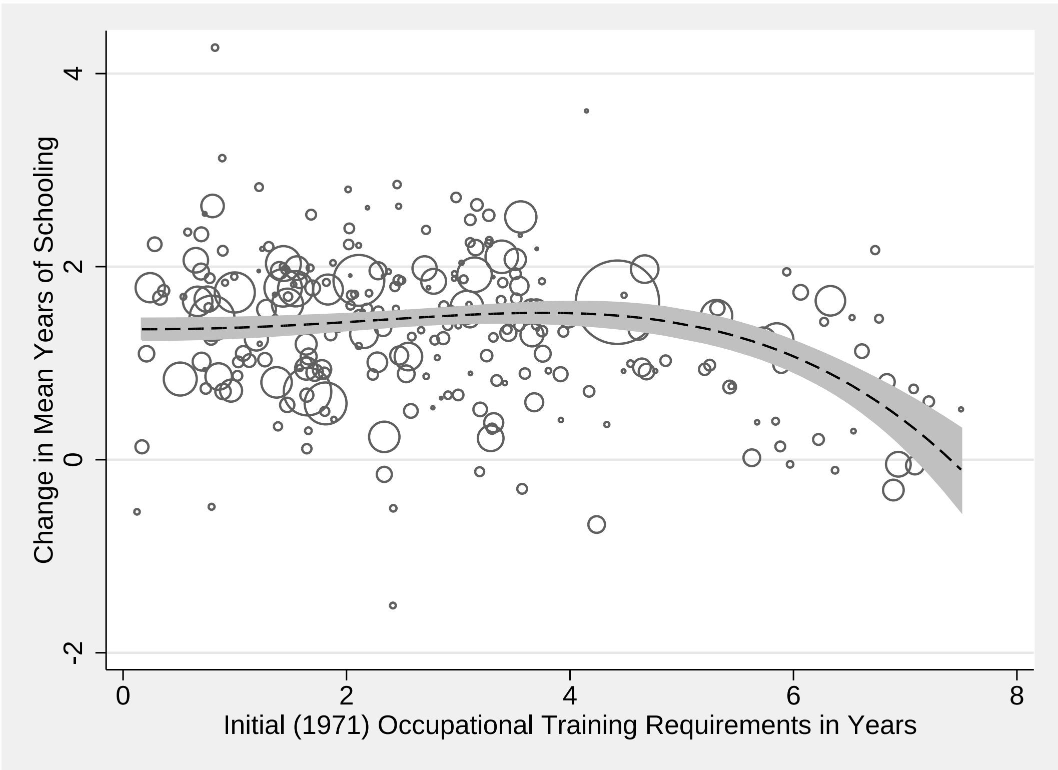
Units on the horizontal and vertical axes are log wages and changes in employment shares 1971-2007 times one hundred, respectively. Occupational mean log wages in 1971 have been adjusted for sex, race, experience, and region. This is the variable on the horizontal axis in panel (a). In the remaining panels, mean log wages are replaced by predicted values regressions, as indicated. Observations are weighted by average occupational employment shares. Fitted curves are fractional polynomials, drawn using Stata's *fpfitci* option.

Figure 1.6: Changes in Occupational Training Requirements and Average Years of Schooling

(a) Changes in occupational training requirements

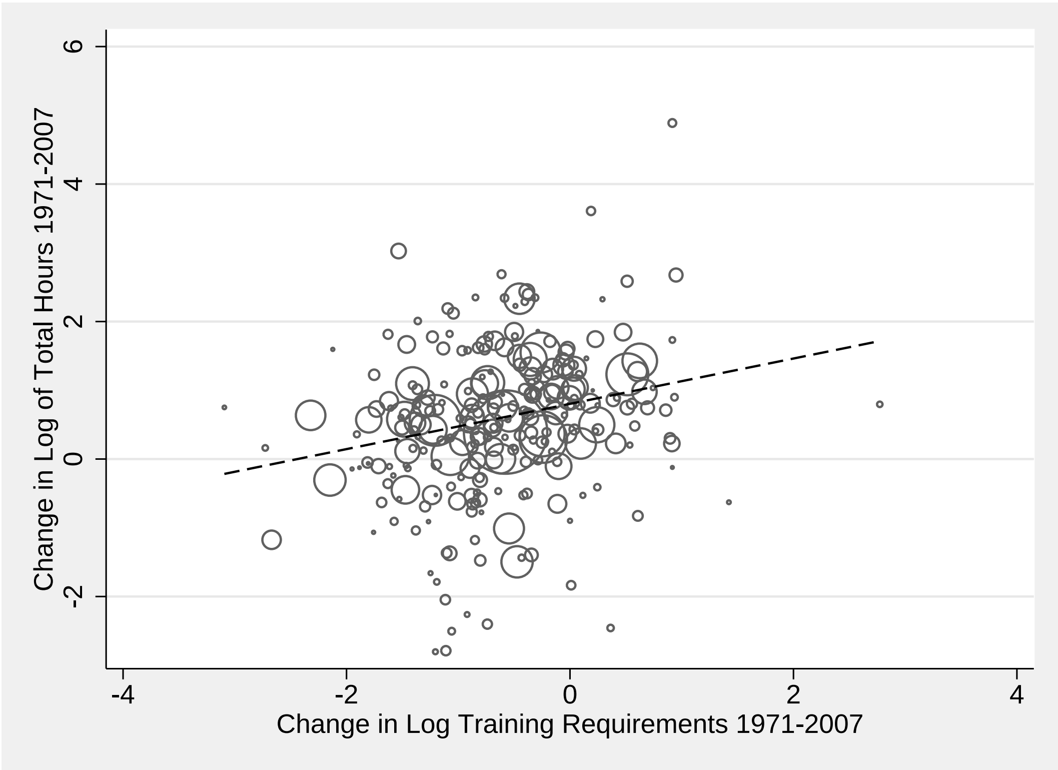


(b) Changes in occupational average years of schooling



Observations are weighted by average occupational employment shares. Fitted curves are fractional polynomials, drawn using Stata's *fpfitci* option.

Figure 1.7: Growth of Occupational Labor Input against Changes in Training Requirements



Fitted line from a regression of changes in log total hours on changes in log training requirements. The estimated coefficient is 0.33 with a robust standard error of 0.08. Observations are weighted by average occupational employment shares.

as the most complex tasks in these occupations are automated. Panel (a) of Figure 1.6 shows that indeed, occupations with intermediate initial training requirements saw the largest declines in training requirements. These occupations include air traffic controllers, precision makers, insurance adjusters, and various engineering occupations (see Table 1.A.4), which appears consistent with our automation-based explanation. Panel (b) of Figure 1.6 shows that average years of schooling increased in almost all occupations. Thus, changes in occupational schooling levels do not follow the same pattern as changes in training requirements, supporting our assertion that the two measures relate to distinct concepts.

If decreases in training requirements are due to increased automation, then employment growth should have been lower in occupations with larger decreases in training requirements. This is indeed the case. A regression of changes in log total hours on changes in log training requirements yields a coefficient of 0.33 (robust standard error 0.08). Raw data and fitted line are plotted in Figure 1.7. Including changes in log years of education on the right hand side slightly decreases the coefficient on training.[34]

[34] A positive and statistically significant relationship also exists between employment growth and changes in the level of training requirements; and between changes in occupational employment shares and changes in both the level and log of training requirements.

Figure 1.8: Changes in Occupational Mean Wages against Changes in Training Requirements

Occupational mean log wages have been adjusted for sex, race, experience, and region. Fitted line from a regression of changes in mean wages on changes in log training requirements. The estimated coefficient is 0.070 with a robust standard error of 0.026. Observations are weighted by average occupational employment shares.

Finally, we consider how changes in training requirements correlate with changes in occupational mean wages. A regression of changes in adjusted occupation log wages on changes in log training requirements yields a coefficient of 0.07 (standard error 0.026), see Figure 1.8. Including changes in log years of education on the right hand side slightly increases the coefficient on training. The finding is consistent with the model if we interpret falls in training requirements as increased automation of tasks. For concreteness, consider an occupation whose task bundle initially includes training-intensive tasks with complexities between σ_1^* and $\sigma' > \widehat{\sigma}_1^*$. Let s' be the skill level of the worker initially performing task σ'. After the fall in machine design costs, all tasks in the interval $[\sigma_1^*, \widehat{\sigma}_1^*]$ are newly automated. Workers with skill levels between $\widehat{s}^*$ and some $s'' < s'$ will remain in the occupation. Figure 1.3 shows that these workers experience wage declines relative to most other workers.

1.6 Conclusion

In this chapter we explain labor market polarization as resulting from technological advances that allow firms to more cheaply automate tasks. Our model thus explains why polarization occurred not only following the information and communication technology revolution, but also after the introduction of the steam engine in the US in the 19th century and during the electrification of US manufacturing in the early 20th century. The model explains why firms do not automate processes although it would be feasible to do so, as in the case of fast food preparation, and why many arguably more complex processes, such as in manufacturing, have already been automated. The key is that difficulty in an engineering sense is not necessarily correlated with difficulty in terms of the amount of training that a human requires to complete a task. Our model delivers novel predictions about changes in training requirements, which we find to be consistent with US data. We have not allowed for changes in the economy's task mix or changes in firm organization resulting from technical change—further research is necessary to determine whether our results are robust to these extensions.

Appendix

1.A Data Sources and Measurement of Training Requirements

Data sources.—Our 1971 training measure comes from the Fourth Edition Dictionary of Occupational Titles (DOT), which is made available in combination with the 1971 April Current Population Survey (CPS) (National Academy of Sciences 1981). We obtain contemporaneous wage data from the IPUMS 1970 census extract (the processing of this data follows the procedure of Acemoglu and Autor (2011)). Our 2007 training measure comes from the Job Zones file in the O*NET database available at `http://www.onetcenter.org/database.html?p=2`. For contemporaneous micro data we use the IPUMS 2008 American Community Survey (ACS).

Measuring training requirements.—SVP (see definition in Section 1.5.2) is measured on a nine-point scale in the DOT. In the O*NET database, Job Zones are measured on a five-point scale which maps into the nine-point SVP scale. See Table 1.A.1 for the interpretation of the SVP scale and the mapping between SVP and Job Zones. In the DOT data, we convert SVP into Job Zones. We assign midpoints to consistently measure training requirements over time. We assign a conservative value to the highest category. See the last column in Table 1.A.1 for details.

The DOT variables, including SVP, in the 1971 April CPS extract vary at the level of 4,528 distinct occupations. For the occupation-level analysis, we collapse the CPS micro data to the three-digit occupation level using David Dorn's classification of occupations (Dorn 2009), weighting by the product of sampling weights and hours worked. The Job Zones variable in the O*NET database is available for 904 distinct occupations of the Standard Occupational Classification System (SOC). In the 2008 ACS data there are 443 distinct SOC occupations. We collapse the O*NET data to these 443 occupations and then merge it to the ACS data. For the occupation-level analysis, we collapse the ACS micro data to the three-digit occupation level in the same way as the CPS data.

Table 1.A.2 lists the twenty least and most training-intensive occupations (using David Dorn's classification) in 1971. Table 1.A.3 does the same for 2007. Table 1.A.4 lists the twenty occupations experiencing the largest declines and increases in training requirements.

Table 1.A.1: Measuring Training Requirements Based on SVP and Job Zones

	SVP	*Job Zone*	*Training time*
1	short demonstration	1	1.5 months
2	up to 30 days	1	1.5 months
3	30 days to 3 months	1	1.5 months
4	3 to 6 months	2	7.5 months
5	6 months to 1 year	2	7.5 months
6	1 to 2 years	3	1.5 years
7	2 to 4 years	4	3 years
8	4 to 10 years	5	7.5 years
9	over 10 years	5	7.5 years

Table 1.A.2: Least and Most Training-Intensive Occupations, 1971

Occupation (occ1990dd grouping)	Training requirements in years (1971)
a) least training-intensive	
Public transportation attendants and inspectors	0.1
Packers and packagers by hand	0.2
Waiter/waitress	0.2
Mail carriers for postal service	0.3
Garage and service station related occupations	0.4
Bartenders	0.4
Messengers	0.4
Parking lot attendants	0.4
Cashiers	0.5
Child care workers	0.6
Misc material moving occupations	0.6
Taxi cab drivers and chauffeurs	0.7
Baggage porters	0.7
Housekeepers, maids, butlers, stewards, and lodging quarters cleaners	0.7
Typists	0.7
Mail and paper handlers	0.7
Proofreaders	0.7
Bus drivers	0.7
File clerks	0.7
Helpers, surveyors	0.8
b) most training-intensive	
Musician or composer	6.8
Mechanical engineers	6.8
Aerospace engineer	6.8
Electrical engineer	6.9
Biological scientists	6.9
Chemical engineers	7.0
Chemists	7.0
Managers in education and related fields	7.0
Petroleum, mining, and geological engineers	7.1
Architects	7.1
Subject instructors (HS/college)	7.1
Dentists	7.2
Veterinarians	7.2
Lawyers	7.2
Civil engineers	7.2
Clergy and religious workers	7.3
Psychologists	7.3
Physicians	7.3
Geologists	7.5
Physicists and astronomers	7.5

Table 1.A.3: Least and Most Training-Intensive Occupations, 2007

Occupation (occ1990dd grouping)	Training requirements in years (2007)
a) least training-intensive	
Waiter/waitress	0.1
Misc food prep workers	0.1
Ushers	0.1
Parking lot attendants	0.1
Kitchen workers	0.1
Furniture and wood finishers	0.1
Pressing machine operators (clothing)	0.1
Fishers, hunters, and kindred	0.1
Textile sewing machine operators	0.1
Graders and sorters of agricultural products	0.1
Garage and service station related occupations	0.1
Taxi cab drivers and chauffeurs	0.1
Animal caretakers, except farm	0.2
Butchers and meat cutters	0.3
Janitors	0.4
Sales demonstrators / promoters / models	0.4
Housekeepers, maids, butlers, stewards, and lodging quarters cleaners	0.4
Miners	0.4
Cashiers	0.4
Stock and inventory clerks	0.4
b) most training-intensive	
Other health and therapy	7.5
Psychologists	7.5
Physicians	7.5
Economists, market researchers, and survey researchers	7.5
Lawyers	7.5
Managers of medicine and health occupations	7.5
Physicians' assistants	7.5
Biological scientists	7.5
Medical scientists	7.5
Physical scientists, n.e.c.	7.5
Podiatrists	7.5
Veterinarians	7.5
Subject instructors (HS/college)	7.5
Dietitians and nutritionists	7.5
Urban and regional planners	7.5
Pharmacists	7.5
Librarians	7.5
Optometrists	7.5
Dentists	7.5
Physicists and astronomers	7.5

Table 1.A.4: Largest Decreases and Increases in Training Requirements, 1971-2007

Occupation (occ1990dd grouping)	Change in training requirements (years) 1971-2007	Training requirements in 1971 (years)
a) largest decreases in training requirements		
Carpenters	-5.7	6.4
Musician or composer	-5.1	6.8
Air traffic controllers	-5.0	6.5
Production supervisors or foremen	-4.7	5.4
Dental laboratory and medical appliance technicians	-4.7	5.9
Geologists	-4.5	7.5
Precision makers, repairers, and smiths	-4.4	5.9
Insurance adjusters, examiners, and investigators	-4.4	5.7
Civil engineers	-4.2	7.2
Recreation and fitness workers	-4.1	6.4
Chemical engineers	-4.0	7.0
Masons, tilers, and carpet installers	-3.9	4.7
Heating, air conditioning, and refigeration mechanics	-3.9	5.4
Electrical engineer	-3.9	6.9
Petroleum, mining, and geological engineers	-3.8	7.1
Aerospace engineer	-3.8	6.8
Mechanical engineers	-3.8	6.8
Explosives workers	-3.8	4.4
Patternmakers and model makers	-3.7	5.2
Molders, and casting machine operators	-3.6	4.2
b) largest increases in training requirements		
Primary school teachers	1.2	1.8
Operations and systems researchers and analysts	1.3	4.6
Agricultural and food scientists	1.3	4.7
Archivists and curators	1.5	4.5
Managers of medicine and health occupations	1.5	6.0
Public transportation attendants and inspectors	1.9	0.1
Therapists, n.e.c.	2.3	2.9
Proofreaders	2.3	0.7
Vocational and educational counselors	2.5	4.1
Registered nurses	2.7	3.1
Social workers	2.7	3.3
Social scientists, n.e.c.	3.0	4.2
Economists, market researchers, and survey researchers	3.2	4.3
Optometrists	3.9	3.6
Pharmacists	4.3	3.2
Librarians	4.4	3.1
Podiatrists	4.5	3.0
Physical scientists, n.e.c.	4.5	3.0
Other health and therapy	4.5	3.0
Dietitians and nutritionists	4.6	2.9

1.B An Extended Model of Task Production and Firms' Productivity Choices

Here we model the production process for tasks explicitly, following Garicano (2000). In order to produce, factors (workers, machines) must confront and solve problems. These problems are task-specific. There is a continuum of problems $Z \in [0, \infty)$ in each task, and problems are ordered by frequency. Thus, there exists a non-increasing probability density function for problems in each task.

Factors draw problems and produce if and only if they know the solution to the problem drawn. We assume that a mass A of problems is drawn, and A may vary across factors. Hence, the task-neutral productivity term introduced in Section 1.2.2 has a more precise interpretation in this context. Task output per factor unit is equal to A times the integral of the density function over the set of problems to which the factor knows the solution.

For simplicity, we will again assume that all workers draw a unit mass of problems in all tasks, or $A = 1$. Equilibrium assignment and comparative statics results are qualitatively the same if we instead assume that $A \equiv A(s)$ with $A'(s) \geq 0$.

The distribution of problems in a task with complexity σ is given by the cumulative distribution function $F(Z; \sigma)$, which we assume to be continuously differentiable in both Z and the shift parameter σ. Let $\partial F/\partial\sigma < 0$, so that σ indexes first-order stochastic dominance. In terms of the some of the examples discussed in Section 1.2.5, driving a car and grading an essay are more complex (higher σ) than driving a train or grading an MCQ test since the number of distinct problems typically encountered in the former set of tasks is higher than in the latter.

The probability density function corresponding to F is $f(Z; \sigma)$. Because F is continuously differentiable and Z indexes frequency, f is strictly decreasing in Z. We impose the following condition on the family of distributions $F(Z, \sigma)$.

Assumption 1 *$F(z, \sigma)$ is strictly log-supermodular.*

This assumption will give rise to the same comparative advantage properties as in the baseline model. One of the distributions satisfying Assumption 1 is the exponential distribution with mean σ.

Note that the distribution of problems depends only on σ and not on τ. As discussed above, training intensity is not an intrinsic property of a task, but arises from the fact that humans have evolved such that some tasks require less effort to master than others, even holding constant (objective) complexity. In this context, humans are assumed to be endowed with knowledge of the solutions to all problems in innate ability tasks.

We now characterize optimal training and design choices and derive equilibrium productivity of workers and machines. First observe that firms will equip factors with a set of knowledge $[0, z]$, since it can never be optimal not to know the solutions to the most frequent problems. As each worker is endowed with one efficiency unit of labor, after incurring learning costs $1 - z/s$ efficiency units are left for production, solving a fraction $F(z; \sigma)$ of problems drawn. Similarly, after the design cost, $1 - z/s_K$ units of capital are left, and the machine solves a fraction $F(z; \sigma)$ of problems drawn. Let the productivity level of an optimally trained worker of skill s in task $(\sigma, 1)$

be denoted by $\alpha^N(s, \sigma, 1)$, and similarly let $\alpha^K(s_K, \sigma)$ be the productivity level of an optimally designed machine. For simplicity, we omit the task-neutral productivity term A_K here, as it does not affect optimal machine design. Then we have

$$\begin{aligned} \alpha^N(s, \sigma, 1) &\equiv \max_z F(z; \sigma) \left[1 - z/s\right], \\ \alpha^K(s_K, \sigma) &\equiv \max_z F(z; \sigma) \left[1 - z/s_K\right], \end{aligned}$$

A unique interior solution to the worker training and machine design problems always exists. Unlike in the baseline model, we do not require any restrictions on s_K and $\underline{s}$ in relation to $\overline{\sigma}$ to ensure that productivity is non-negative.[35] The optimal knowledge levels $z^N(s, \sigma)$ and $z^K(s_K, \sigma)$ are pinned down by the first-order conditions

$$\begin{aligned} f(z(s, \sigma); \sigma) \left[1 - z(s, \sigma)/s\right] &= F(z(s, \sigma, \tau); \sigma)/s, \\ f(z(s_K, \sigma); \sigma) \left[1 - z(s_K, \sigma)/s_K\right] &= F(z(s_K, \sigma); \sigma)/s_K. \end{aligned} \tag{1.B.1}$$

Optimality requires that the benefit of learning the solution to an additional problem—the probability that the problem occurs times the number of efficiency units left for production, be equal to the cost of doing so—the number of efficiency units lost times the fraction of problems these efficiency units would have solved. Optimal worker and machine productivities are given by

$$\alpha^N(s, \sigma, \tau) = \begin{cases} F(z(s, \sigma, \tau); \sigma) \left[1 - z(s, \sigma, \tau)/s\right] & \textit{if } \tau = 1 \\ 1 & \textit{if } \tau = 0 \end{cases}$$

and

$$\alpha^K(s_K, \sigma) = F(z(s_K, \sigma); \sigma) \left[1 - z(s_K, \sigma)/s_K\right].$$

Let $\breve{s}$ be an element in set $\breve{S} = s_K \cup [\underline{s}, \overline{s}]$. By the above results, we have that $\alpha^N(\breve{s}, \sigma, 1) \equiv \alpha^K(\breve{s}, \sigma)$. Thus, workers and machines face the same productivity schedule in training-intensive tasks. We drop superscripts and define the function

$$\alpha(\breve{s}, \sigma) = F(z(\breve{s}, \sigma); \sigma) \left[1 - \frac{1}{\breve{s}} z(\breve{s}, \sigma)\right] \qquad \breve{s} \in \breve{S} = s_K \cup [\underline{s}, \overline{s}], \tag{1.B.2}$$

where $z(\breve{s}, \sigma)$ is implicitly given by (1.B.1).

The qualitative properties of the productivity schedule $\alpha(\breve{s}, \sigma)$ are the same as in the baseline model. First notice that $\alpha \in (0, 1)$ by (1.B.2). Furthermore, from applying the envelope theorem to (1.B.2) it follows that α is increasing in $\breve{s}$ and decreasing in σ. Higher skilled factors are more

[35] A unique interior solution to the worker training problem exists if $\tau > 0$ because first, the problem is strictly concave as f is strictly decreasing; second, the derivative of the objective at $z = 0$ is strictly positive; finally, the value of the objective function becomes negative for a sufficiently large z. The same arguments also establish the result for the machine design problem.

productive since they face a lower learning/design cost, and productivity declines in complexity since a larger cost is incurred to achieve a given level of productivity.

To characterize comparative advantage, we again rely on log-supermodularity: Under Assumption 1, the productivity schedule $\alpha(\breve{s}, \sigma)$ is strictly log-supermodular. To show this, start by observing that $\alpha(\breve{s}, \sigma)$ is strictly log-supermodular if and only if

$$\frac{\partial^2}{\partial \breve{s} \partial \sigma} \log \alpha(\breve{s}, \sigma) > 0.$$

Applying the envelope theorem to (1.B.2) yields

$$\frac{\partial}{\partial \breve{s}} \log \alpha(\breve{s}, \sigma) = \frac{z(\breve{s}, \sigma)}{(\breve{s})^2 - \breve{s} z(\breve{s}, \sigma)}.$$

The RHS is an increasing function of $z(\breve{s}, \sigma)$, and so

$$\frac{\partial^2}{\partial \breve{s} \partial \sigma} \log \alpha(\breve{s}, \sigma) > 0 \quad \Leftrightarrow \quad \frac{\partial}{\partial \sigma} z(\breve{s}, \sigma) > 0.$$

Thus, α is log-supermodular if and only if optimal knowledge levels are increasing in σ. Differentiating the FOC (1.B.1) yields

$$\frac{\partial}{\partial \sigma} z(\breve{s}, \sigma) = \frac{F_\sigma \frac{1}{\breve{s}} - f_\sigma \left[1 - \frac{1}{\breve{s}} z\right]}{f_z \left[1 - \frac{1}{\breve{s}} z\right] - 2f \frac{1}{\breve{s}}}.$$

The denominator of the RHS is negative as $f_z < 0$, and so, using the FOC we find that

$$\frac{\partial}{\partial \sigma} z(\breve{s}, \sigma) > 0 \qquad \Leftarrow \qquad F f_\sigma > F_\sigma f \quad \forall Z, \sigma > 0.$$

But this condition is equivalent to F being strictly log-supermodular.

Given that $\alpha(\breve{s}, \sigma)$ has the same qualitative properties as the productivity schedule in the baseline model, assignment and comparative statics results are qualitatively the same, as well, since none of the results for the baseline model rely on the specific functional form.

1.C A Model with Fixed Costs

1.C.1 Model Setup

Worker training technologies are as in the baseline model. However, we now assume that an upfront expense of $\varphi(\sigma)$ is required to equip the firm's stock of machines with σ units of knowledge. This cost is independent of the size of the stock, as in the case of software. We make the critical assumption $\varphi' > 0$, and for simplicity we set $\varphi(\underline{\sigma}) = 0$ and $\varphi'' > 0$. In our numerical solutions we choose $\varphi(\sigma) = c_K(\sigma - \underline{\sigma})^2$ where c_K is the parameter capturing labor-replacing technical change. Machines capable of performing a task produce A_K units of task output—that is, machine productivity is independent of complexity. As in the baseline model, worker productivity is independent of complexity and worker skill in innate ability tasks, but in training-intensive tasks worker productivity is given by $\alpha(s, \sigma) \equiv 1 - \sigma/s$.

We assume that each task is produced by a single monopolistic firm.[36] In contrast, final good firms are perfectly competitive just as in the baseline version of the model. The final good production function is now

(1.C.1)
$$Y = \left[\int_{\underline{\sigma}}^{\overline{\sigma}} \left\{ \beta_1 y_0(\sigma)^{\frac{\varepsilon-1}{\varepsilon}} + \beta_1 y_1(\sigma)^{\frac{\varepsilon-1}{\varepsilon}} \right\} d\sigma \right]^{\frac{\varepsilon}{\varepsilon-1}},$$

with $\varepsilon > 1$ and $\sum_\tau \beta_\tau = 1$ for CRS. Given profit maximization by final good firms, the CES production function yields the standard isoelastic input demand curve, inducing the well-known constant-markup pricing rule.

Standard arguments establish that equilibrium variable profits of the firm supplying task (σ, τ) are given by

(1.C.2)
$$\pi_\tau(\sigma, \chi) = \beta_\tau^\varepsilon \frac{(\varepsilon - 1)^{\varepsilon-1}}{\varepsilon^\varepsilon} \chi^{-(\varepsilon-1)} Y$$

where χ is marginal cost which depends on the characteristics of tasks and factors employed. In particular, if employing labor of type s we have

$$\chi \equiv \chi(s, \sigma, \tau) = \begin{cases} w(s) & \textit{if } \tau = 0 \\ w(s)/\alpha(s, \sigma) & \textit{if } \tau = 1, \end{cases}$$

and if employing capital,

$$\chi = \frac{r}{A_K}.$$

Furthermore, equilibrium task output is

(1.C.3)
$$y_\tau(\sigma, \chi) = \beta_\tau^\varepsilon \left(\frac{\varepsilon - 1}{\varepsilon} \right)^\varepsilon \chi^{-\varepsilon} Y.$$

[36] Holmes and Mitchell (2008) present a more complex model where labor and machines are optimally assigned to tasks *within* monopolistic firms. We suspect that our results would hold in a version of that model as well.

1.C.2 Equilibrium Assignment

We focus on equilibria in which $r/A_K < w(\overline{s})$, so that the marginal cost of using machines is less than that of employing labor of any type in any task.[37] In such an equilibrium, the assignment is qualitatively the same as in the case we analyzed for the baseline model. Task producers employ the factor that delivers the highest total profits. Note that variable profits among firms that use machines are constant across tasks. In any equilibrium with the above characteristic there are a threshold tasks σ_0^*, σ_1^* such that it is optimal for firms to use machines in all innate ability tasks with $\sigma \leq \sigma_0^*$ and in all training-intensive tasks with $\sigma \leq \sigma_1^*$.[38] As in the baseline model, there is a cutoff s^* such that workers below the cutoff perform innate ability tasks, while workers above the cutoff carry out training intensive tasks, with higher skilled workers performing more complex tasks.

1.C.3 Solving the Model

The threshold tasks are determined by no-arbitrage conditions. Profits from employing labor in these tasks must be equal to profits from using machines. This means that the difference in variable profits between machines and labor must equal the fixed cost of designing the machine,

$$\beta_0^\varepsilon \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon} Y \left[\left(\frac{A_K}{r}\right)^{\varepsilon-1} - \left(\frac{1}{w(s^*)}\right)^{\varepsilon-1} \right] = \varphi(\sigma_0^*) \tag{1.C.4}$$

and

$$\beta_1^\varepsilon \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon} Y \left[\left(\frac{A_K}{r}\right)^{\varepsilon-1} - \left(\frac{\alpha(s^*,\sigma_1^*)}{w(s^*)}\right)^{\varepsilon-1} \right] = \varphi(\sigma_1^*). \tag{1.C.5}$$

Worker assignment to training intensive tasks is assortative as in the baseline model, with the wage schedule given by

$$\frac{d \log w(s)}{ds} = \frac{\partial \log \alpha(s, M(s))}{\partial s} \tag{1.C.6}$$

and the matching function satisfying

$$M'(s) = \frac{\alpha(s, M(s))^{1-\varepsilon} v(s) w(s)^\varepsilon}{\beta_1^\varepsilon \left(\frac{\varepsilon-1}{\varepsilon}\right)^\varepsilon Y} \tag{1.C.7}$$

as well as the boundary conditions $M(s^*) = \sigma_1^*$ and $M(\overline{s}) = \overline{\sigma}$.

The model is closed by market clearing conditions and the production function. For capital markets to clear, we must have $\int_{\underline{\sigma}}^{\sigma_0^*} k_0(\sigma) d\sigma + \int_{\underline{\sigma}}^{\sigma_1^*} k_1(\sigma) d\sigma = K$. Because of our assumption

[37] We do not derive formal conditions ensuring this property, but it will feature in our numerical solutions.

[38] Notice that profits of firms employing type-s labor are flat in σ ($\tau = 0$) or a decreasing and convex function of σ approaching zero as σ gets large ($\tau = 1$). However, because fixed machine design costs are decreasing in σ, and because of the convexity of the cost function, profits of firms using machines become negative as σ gets large. Due to the properties of α, profits when employing labor are a convex function of σ, so that there exists a value of σ such that profits from employing typs-s labor are equal to profits from using machines.

of constant marginal costs of using machines regardless of a task's complexity, we have that task outputs are constant within innate ability tasks and within training-intensive tasks, and so are machine inputs. Thus, the capital market clearing condition becomes $(\sigma_0^* - \underline{\sigma})k_0 + (\sigma_0^* - \underline{\sigma})k_1 = K$ Using the task production function $y_\tau = A_K k_\tau$ and equilibrium task output (1.C.3), we obtain $k_0/k_1 = (\beta_0/\beta_1)^\varepsilon$. Together with the market clearing condition and the task production function this implies

$$y_\tau(\sigma) = \frac{\beta_\tau^\varepsilon A_K K}{\sum_\tau \beta_\tau(\sigma_\tau^* - \underline{\sigma})} \quad \textit{for all } \sigma \in [\underline{\sigma}, \sigma_\tau^*]. \tag{1.C.8}$$

When firms employ labor in innate ability tasks, the task production function is $y_0(\sigma) = \int_{\underline{s}}^{s^*} n_0(s,\sigma)ds$. The market clearing condition is $\int_{\sigma_0^*}^{\overline{\sigma}} \int_{\underline{s}}^{s^*} n_0(s,\sigma)dsd\sigma = V(s^*)$, and so

$$y_0(\sigma) = \frac{V(s^*)}{\overline{\sigma} - \sigma_0^*} \quad \textit{for all } \sigma \in [\sigma_0^*, \overline{\sigma}]. \tag{1.C.9}$$

Given (1.12), (1.C.8), and (1.C.9), final good output must satisfy

$$\begin{aligned} Y^{\frac{\varepsilon-1}{\varepsilon}} &= [\beta_0^\varepsilon(\sigma_0^* - \underline{\sigma}) + \beta_1^\varepsilon(\sigma_1^* - \underline{\sigma})]^{\frac{1}{\varepsilon}} (A_K K)^{\frac{\varepsilon-1}{\varepsilon}} \\ &\quad + \beta_0[\overline{\sigma} - \sigma_0^*]^{\frac{1}{\varepsilon}} V(s^*)^{\frac{\varepsilon-1}{\varepsilon}} + \beta_1 \int_{s^*}^{\overline{s}} M'(s)^{\frac{1}{\varepsilon}} [\alpha(s, M(s))v(s)]^{\frac{\varepsilon-1}{\varepsilon}} ds. \end{aligned} \tag{1.C.10}$$

Plugging (1.C.8) into (1.C.3) yields an expression for the rental rate,

$$\frac{r}{A_K} = \frac{\varepsilon - 1}{\varepsilon} [\beta_0^\varepsilon(\sigma_0^* - \underline{\sigma}) + \beta_1^\varepsilon(\sigma_1^* - \underline{\sigma})]^{\frac{1}{\varepsilon}} \left(\frac{Y}{A_K K}\right)^{\frac{1}{\varepsilon}}. \tag{1.C.11}$$

Similarly, plugging (1.C.9) into (1.C.3) gives an expression for the wage paid to the marginal worker,

$$w(s^*) = \frac{\varepsilon - 1}{\varepsilon} \beta_0 (\overline{\sigma} - \sigma_0^*)^{\frac{1}{\varepsilon}} \left(\frac{Y}{V(s^*)}\right)^{\frac{1}{\varepsilon}}. \tag{1.C.12}$$

We solve the model by grid search. Given a guess of (s^*, σ_1^*), we solve for the matching function and the wage distribution, obtaining $w(s^*)/Y^{1/\varepsilon}$. Using this, we calculate σ_0^* from (1.C.12) and Y from (1.C.10), thus obtaining $w(s^*)$. We then calculate r/A_K from (1.C.11). Finally, we check whether the no-arbitrage conditions (1.C.4) and (1.C.5) are satisfied.

1.C.4 Numerical Solution

We solve the model for values of c_K ranging from one to two. Parameter values are given in Table 1.C.1. The skill distribution $v(s)$ is a truncated log-normal. We construct the distribution such that the corresponding (non-truncated) normal distribution has mean $\log \underline{s} + 1/3 * \log(\overline{s}/\underline{s})$ and standard deviation $1/6 * \log(\overline{s}/\underline{s})$. This implies that the original normal distribution is truncated at four (two) standard deviations above (below) the mean.

Table 1.C.1: Parameter Values for the Model with Fixed Design Costs

β_0	$=$	$1/3$
ε	$=$	2
σ	$\in$	$[0,1]$
s	$\in$	$[1.01,2]$
$v(s)$		see text
$A_K K$	$=$	1
$\phi(\sigma)$	$=$	$c_K\sigma^2$
c_K	$\in$	$[1,2]$

As in the baseline model, firms adopt machines more widely in training-intensive tasks as design becomes cheaper—the marginal training-intensive task becomes more complex. The effect on the marginal innate ability task is ambiguous, however. (See Figure 1.C.1.)

The skill cutoff increases as design gets cheaper, so that the employment share of innate ability tasks rises. (See Figure 1.C.2.) But because σ_1^* also increases the matching function shifts up, implying a reallocation of workers to more complex training-intensive tasks. Market clearing implies a compression of the wage distribution in the lower part of the wage distribution but increasing dispersion in the upper part. Thus, the model with a fixed design cost features job and wage polarization just like the baseline model.

Figure 1.C.1: Changes in Cutoff Tasks in the Model with Fixed Design Cost as Machine Design Becomes Cheaper.

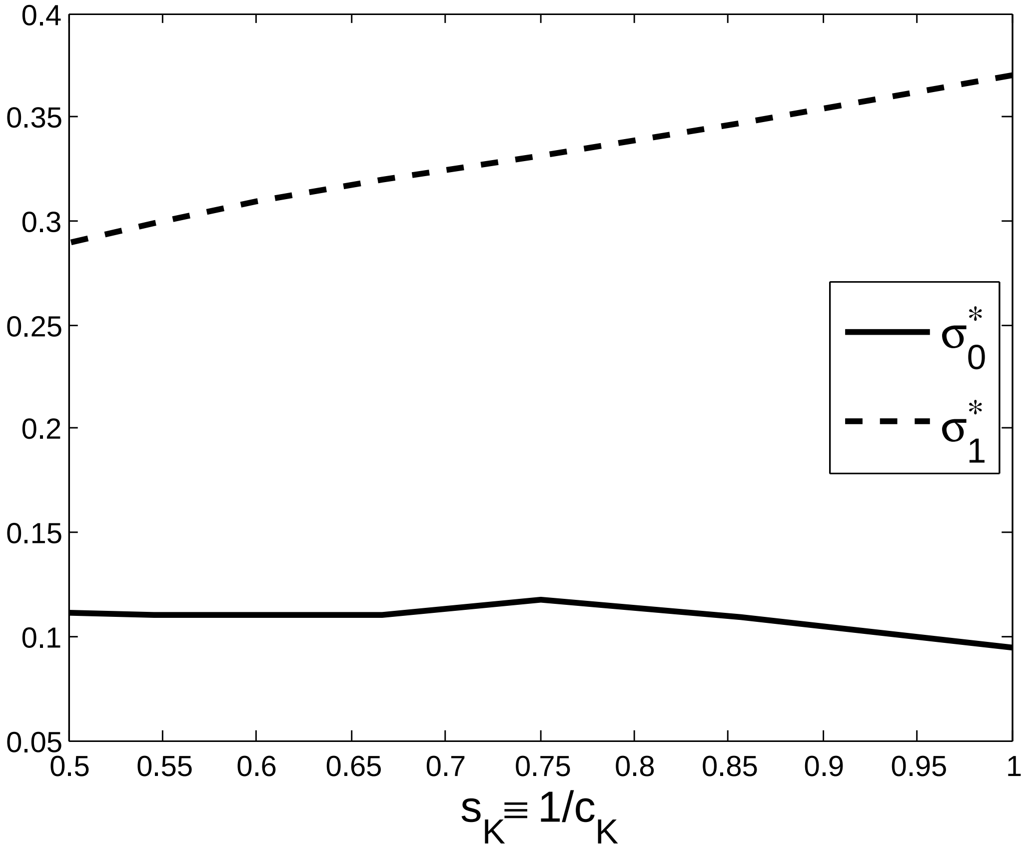
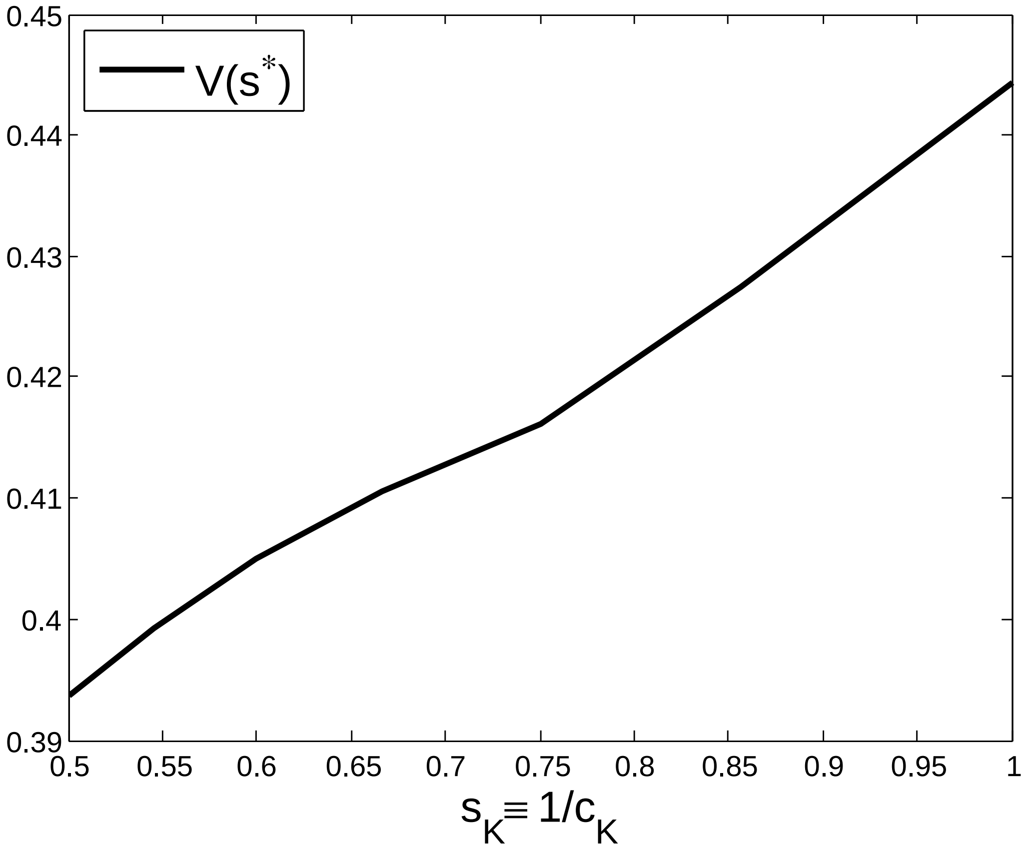


Figure 1.C.2: Changes in the Skill Cutoff in the Model with Fixed Design Cost as Machine Design Becomes Cheaper



The fraction of workers below the cutoff $V(s^*)$ is plotted.

1.D Proofs of Formal Results Stated in the Text

1.D.1 Sufficient Conditions for Existence of an Interior Equilibrium

We derive sufficient conditions ensuring that an interior equilibrium with $\sigma_0^*, \sigma_1^* \in (\underline{\sigma}, \overline{\sigma})$ and hence $s^* \in (\underline{s}, \overline{s})$ prevails. These conditions will consist of mild restrictions on the values that the economy's endowment of efficiency units of capital $A_K K$ may take, given a particular choice of values $(\overline{s}, \underline{\sigma}, \overline{\sigma})$.

In any equilibrium in which $k_0(\sigma) = 0$ for all $\sigma \in [\underline{\sigma}, \overline{\sigma}]$, we have by (1.8)

$$\begin{aligned} p_0(\underline{\sigma})\alpha(s_K, \underline{\sigma}) &\leq r/A_K \\ p_0(\underline{\sigma}) &= w(s^*), \end{aligned}$$

which yields $\alpha(s_K, \underline{\sigma}) \leq r/[A_K w(s^*)]$. Using (1.17) and (1.19) this inequality is shown to be equivalent to

$$\alpha(s_K, \underline{\sigma}) \leq \frac{\beta_1(\sigma_1^* - \underline{\sigma})}{\beta_0(\overline{\sigma} - \underline{\sigma})} \times \frac{V(s^*)}{A_K K}.$$

The RHS of the last inequality is strictly less than $\beta_1/(\beta_0 A_K K)$, hence a sufficient condition to rule out any equilibrium in which $k_0(\sigma) = 0$ for all $\sigma \in [\underline{\sigma}, \overline{\sigma}]$ is $\alpha(s_K, \underline{\sigma}) > \beta_1/(\beta_0 A_K K)$ or

$$A_K K > \frac{\beta_1}{\beta_0} \frac{1}{\alpha(s_K, \underline{\sigma})}. \tag{1.D.1}$$

And in any equilibrium in which $n_1(s, \sigma) = 0$ for all $s \in [\underline{s}, \overline{s}]$ and $\sigma \in [\underline{\sigma}, \overline{\sigma}]$ we have by (1.8)

$$\begin{aligned} p_1(\overline{\sigma})\alpha(c_K, \overline{\sigma}) &= r/A_K \\ p_1(\overline{\sigma})\alpha(\overline{s}, \overline{\sigma}) &\leq w(\overline{s}) = w(s^*), \end{aligned}$$

from which we obtain $\alpha(s_K, \overline{\sigma})/\alpha(\overline{s}, \overline{\sigma}) \geq r/[A_K w(s^*)]$. Using (1.17) and (1.19) this inequality becomes

$$\frac{\alpha(s_K, \overline{\sigma})}{\alpha(\overline{s}, \overline{\sigma})} \geq \frac{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\overline{\sigma} - \underline{\sigma})}{\beta_0(\overline{\sigma} - \sigma_0^*)} \times \frac{1}{A_K K}.$$

The RHS of the last inequality is strictly greater than $\beta_1/(\beta_0 A_K K)$, hence a sufficient condition to rule out any equilibrium in which $n_1(s, \sigma) = 0$ for all $s \in [\underline{s}, \overline{s}]$ and $\sigma \in [\underline{\sigma}, \overline{\sigma}]$ is $\alpha(s_K, \overline{\sigma})/\alpha(\overline{s}, \overline{\sigma}) < \beta_1/(\beta_0 A_K K)$ or

$$A_K K < \frac{\beta_1}{\beta_0} \frac{\alpha(\overline{s}, \overline{\sigma})}{\alpha(s_K, \overline{\sigma})}. \tag{1.D.2}$$

Combining (1.D.1) and (1.D.2), we conclude that if

$$A_K K \in S, \quad S \equiv \frac{\beta_1}{\beta_0}\left(\frac{1}{\alpha(s_K, \underline{\sigma})}, \frac{\alpha(\overline{s}, \overline{\sigma})}{\alpha(s_K, \overline{\sigma})}\right),$$

then the equilibrium is interior with $\sigma_0^*, \sigma_1^* \in (\underline{\sigma}, \overline{\sigma})$ and hence $s^* \in (\underline{s}, \overline{s})$. Existence of an interior equilibrium is ensured by choosing parameter values for $(\overline{s}, \underline{\sigma}, \overline{\sigma})$ such that S is a non-empty set.

Our claim that the restrictions on $A_K K$ are mild given a particular choice of $(\overline{s}, \underline{\sigma}, \overline{\sigma})$ is justified both for the baseline model and the more general production process in Appendix 1.B, because we can choose parameters such that $S \to \frac{\beta_1}{\beta_0}(1, \infty)$. In the case of the baseline model, we can set $\underline{\sigma} = 0$ and let $s_K \to \overline{\sigma}$. In the more general model, we can assume that $\underline{\sigma}$ is sufficiently small so that $F(Z; \underline{\sigma})$ is close to one even for very small Z; and that $\overline{\sigma}$ is sufficiently large so that $F(Z; \overline{\sigma})$ is close to zero even for very large Z, while at the same time $\overline{s}$ is large so that $\alpha(\overline{s}, \overline{\sigma})$ stays finite.

1.D.2 Proofs of Lemmas Stated in the Text

Proof of Lemma 1 Follows from the definition of strict log-supermodularity and simple differentiation. ∎

Proof of Lemma 2 (a) For any vectors (s, σ) and (s', σ') such that $n_0(s, \sigma) > 0$ and $n_1(s', \sigma') > 0$ we have by the zero-profit condition (1.8) $p_0(\sigma) = w(s)$ and $p_0(\sigma) \leq w(s')$, or $w(s) \leq w(s')$, and

$$\begin{aligned} p_1(\sigma')\alpha(s', \sigma') &= w(s'), \\ p_1(\sigma')\alpha(s, \sigma') &\leq w(s). \end{aligned}$$

Together these conditions imply $\alpha(s', \sigma')/\alpha(s, \sigma') \geq 1$. Since α is increasing in s we must have $s' \geq s$. Furthermore, it must be that $s^* > \underline{s}$, for suppose not. Then market clearing (1.3) implies that $k_0(\sigma) > 0$ for all σ (task output must be strictly positive due to the INADA properties of the Cobb-Douglas production function). By (1.8), for some (s, σ)

$$\begin{aligned} p_1(\sigma)\alpha(s, \sigma) &= w(s), \\ p_1(\sigma)\alpha(s_K, \sigma) &\leq r/A_K, \end{aligned}$$

which yields

$$\frac{w(s)}{r/A_K} \leq \frac{\alpha(s, \sigma)}{\alpha(s_K, \sigma)}.$$

Furthermore, $p_0(\sigma)\alpha(s_K, \sigma) = r/A_K$ and $p_0(\sigma) \leq w(s)$. This yields

$$\frac{w(s)}{r/A_K} \geq \frac{1}{\alpha(s_K, \sigma)}.$$

Together with the previous result this implies $\alpha(s, \sigma) \geq 1$ which is impossible given (1.9).

(b) If $k_0(\sigma) > 0$, then by the zero-profit condition (1.8)

$$\frac{w(s^*)}{r/A_K} \geq \frac{1}{\alpha(s_K, \sigma)},$$

and there is some σ' such that $n_1(s^*, \sigma') > 0$ and hence by (1.8)

$$\frac{w(s^*)}{r/A_K} \leq \frac{\alpha(s^*, \sigma')}{\alpha(s_K, \sigma')}.$$

The previous two inequalities imply

$$\frac{\alpha(s^*,\sigma')}{\alpha(s_K,\sigma')} \geq \frac{1}{\alpha(s_K,\sigma)},$$

but since $\alpha(s_K,\sigma) < 1$, we have $\alpha(s^*,\sigma')/\alpha(s_K,\sigma') > 1$ which is only possible if $s^* > s_K$.

Next, observe that for any (σ,σ') and $s \leq s^*$ such that $k_0(\sigma) > 0$ and $n_0(s,\sigma') > 0$ we have by (1.8),

$$\begin{aligned} p_0(\sigma)\alpha(s_K,\sigma) &= r/A_K \\ p_0(\sigma) &\leq w(s), \end{aligned}$$

and

$$\begin{aligned} p_0(\sigma')\alpha(s_K,\sigma') &\leq r/A_K \\ p_0(\sigma') &= w(s), \end{aligned}$$

which yields $\alpha(s_K,\sigma) \geq \alpha(s_K,\sigma')$ and so $\sigma \leq \sigma'$. Thus we have established existence of σ_0^*.

Similarly, for any (σ,σ') and $s \geq s^*$ such that $k_1(\sigma) > 0$ and $n_1(s,\sigma') > 0$, we have by (1.8),

$$\begin{aligned} p_1(\sigma)\alpha(s_K,\sigma) &= r/A_K \\ p_1(\sigma)\alpha(s,\sigma) &\leq w(s), \end{aligned}$$

and

$$\begin{aligned} p_1(\sigma')\alpha(s_K,\sigma') &\leq r/A_K \\ p_1(\sigma')\alpha(s,\sigma') &= w(s), \end{aligned}$$

which yields

$$\frac{\alpha(s_K,\sigma)}{\alpha(s,\sigma)} \geq \frac{\alpha(s_K,\sigma')}{\alpha(s,\sigma')},$$

and so $\sigma \leq \sigma'$ by the log-supermodularity of α and since $s > s_K$. This establishes existence of σ_1^*.

Now, it must be that $\sigma_0^* < \sigma_1^*$, for suppose not. If $\sigma_0^* > \sigma_1^*$, then there exist (s,σ) such that $k_0(\sigma) > 0$, $k_1(\sigma) = 0$, $n_0(s,\sigma) = 0$, and $n_1(s,\sigma) > 0$. By (1.8),

$$\begin{aligned} p_0(\sigma)\alpha(s_K,\sigma) &= r/A_K \\ p_0(\sigma) &\leq w(s), \end{aligned}$$

and

$$\begin{aligned} p_1(\sigma)\alpha(s_K,\sigma) &\leq r/A_K \\ p_1(\sigma)\alpha(s,\sigma) &= w(s). \end{aligned}$$

This yields $\alpha(s,\sigma) \geq 1$ which contradicts (1.9). If $\sigma_0^* = \sigma_1^*$, then similar arguments lead to $\alpha(s,\sigma) = 1$, which also contradicts (1.9). ∎

Proof of Lemma 3 Given Lemma 2, the problem is to match workers of skill levels $s \in [s^*,\overline{s}]$ to tasks $\sigma \in [\sigma_1^*,\overline{\sigma}]$ in a setting identical to that in Costinot and Vogel (2010). Hence, the proof of Lemma 1 from their paper applies. ∎

1.D.3 Proofs of Propositions Stated in the Text

Proof of Proposition 1 We first show that in the absence of changes to the distribution of skills, a flattening (steepening) of the matching function at the upper end implies an upward (downward) shift of the matching function everywhere. Formally, if $\widehat{M}'(\overline{s}) < M'(\overline{s})$, then $\widehat{M}(s) < M(s)$ for all $s \in [\max\{s^*, \widehat{s}^*\}, \overline{s})$. For suppose that $\widehat{M}'(\overline{s}) < M'(\overline{s})$ and that there exists some $s' \in [\max\{s^*, \widehat{s}^*\}, \overline{s})$ such that $\widehat{M}(s') \leq M(s')$. Then there exists some $s'' \in [s', \overline{s})$ such that $\widehat{M}(s'') = M(s'')$, $\widehat{M}'(s'') \geq M'(s'')$, and $\widehat{M}(s) > M(s)$ for all $s \in (s'', \overline{s})$. We will show that this leads to a contradiction.

Integrating (1.11) yields an expression for the wage premium of the most skilled worker with respect to any other skill group employed in training-intensive tasks,

$$\frac{w(\overline{s})}{w(s)} = \omega(s; M), \quad s \geq s^*$$

where

$$\omega(s; M) \equiv \exp\left[\int_s^{\overline{s}} \frac{\partial}{\partial z} \log \alpha(z, M(z)) dz\right]. \tag{1.D.3}$$

Because α is increasing in its first argument, ω is decreasing in s. Moreover, by the log-supermodularity of α, if $\widehat{M}(z) > M(z)$ for all $z \in (s, \overline{s})$ and any s that belongs to the domains of both $\widehat{M}$ and M, then $\omega(s; \widehat{M}) > \omega(s; M)$.

Plugging (1.D.3) into (1.10), we obtain

$$\frac{M'(\overline{s})}{M'(s)} = \omega(s; M) \frac{v(\overline{s})}{v(s)}. \tag{1.D.4}$$

Therefore,

$$\frac{\widehat{M}'(\overline{s})}{M'(\overline{s})} = \frac{\omega(s''; \widehat{M})}{\omega(s''; M)} \frac{\widehat{M}'(s'')}{M'(s'')}.$$

By the above arguments, the right side of the last equation is larger than one, so that we must have $\widehat{M}'(\overline{s}) > M'(\overline{s})$, a contradiction. A similar argument establishes that a steepening at the upper end leads to a downward shift everywhere.

Proof that $\widehat{\sigma}_1^* > \sigma_1^*$ First suppose $\widehat{\sigma}_1^* \leq \sigma_1^*$ and $\widehat{M}'(\overline{s}) \geq M'(\overline{s})$.

By (1.22) and (1.D.4),

$$\frac{V(s^*)}{\overline{\sigma} - \sigma_0^*} \times \frac{M'(\overline{s})}{\omega(s^*; M)} = \frac{\beta_0 v(\overline{s})}{\beta_1}. \tag{1.D.5}$$

This together with (1.20), implies

$$\frac{A_K \alpha(s_K, \sigma_0^*) K}{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})} \times \frac{M'(\overline{s})}{\omega(s^*; M)} = \frac{v(\overline{s})}{\beta_1}. \tag{1.D.6}$$

Suppose that $\widehat{s}^* \geq s^*$. Then (1.D.5) implies that $\widehat{\sigma}_0^* < \sigma_0^*$, while (1.D.6) implies $\widehat{\sigma}_0^* > \sigma_0^*$, a contradiction. So we must have $\widehat{s}^* < s^*$. If $\widehat{\sigma}_0^* \geq \sigma_0^*$, then from (1.21), $\widehat{s}^* > s^*$,[39] so it must be that $\widehat{\sigma}_0^* < \sigma_0^*$. Then by 1.21, $\alpha(\widehat{s}_K, \widehat{\sigma}_0^*) > \alpha(s_K, \sigma_0^*)$. This implies that the LHS of (1.20) increases, while the RHS decreases, a contradiction.

Next, suppose that $\widehat{\sigma}_1^* \leq \sigma_1^*$ and $\widehat{M}'(\overline{s}) < M'(\overline{s})$. We have shown that in this case the matching function shifts up, so we must have $\widehat{s}^* \leq s^*$. Then $\widehat{\sigma}_0^* < \sigma_0^*$ from (1.21). But we have just shown that it is impossible to have $\widehat{\sigma}_1^* \leq \sigma_1^*$, $\widehat{\sigma}_0^* < \sigma_0^*$, and $\widehat{s}^* \leq s^*$ at the same time. Thus we have established that $\widehat{\sigma}_1^* > \sigma_1^*$.

Proof that $\widehat{M}(s) > M(s)$ Suppose that $\widehat{M}'(\overline{s}) > M'(\overline{s})$, which we have shown implies $\widehat{M}(s) < M(s)$ and, by (1.D.4), $\widehat{M}'(s) > M'(s)$ for all s belonging to the domains of both $\widehat{M}$ and M. As we have established that $\widehat{\sigma}_1^* < \sigma_1^*$, by the properties of the matching function we must have $\widehat{s}^* > s^*$. By (1.10), the wage share of a worker who is always assigned to training-intensive tasks has increased,

$$\frac{\widehat{w}(s)}{\widehat{Y}} = \frac{\beta_1}{\mu}\frac{\widehat{M}'(s)}{v(s)} > \frac{\beta_1}{\mu}\frac{M'(s)}{v(s)} = \frac{w(s)}{Y} \quad \forall s \in [\widehat{s}^*, \overline{s}].$$

But this means that the wage shares of all remaining workers have increased, as well,

$$\frac{\widehat{w}(s)}{\widehat{Y}} = \frac{\widehat{w}(\widehat{s}^*)}{\widehat{Y}} > \frac{w(\widehat{s}^*)}{Y} > \frac{w(s)}{Y} \quad \forall s \in [\underline{s}, \widehat{s}^*),$$

where the last inequality is due to (1.23). Therefore, the total labor share has increased,

$$\frac{\int_{\underline{s}}^{\overline{s}} \widehat{w}(s)v(s)ds}{\widehat{Y}} > \frac{\int_{\underline{s}}^{\overline{s}} w(s)v(s)ds}{Y}.$$

By (1.10) and (1.19), this implies $\beta_0\widehat{\sigma}_0^* + \beta_1\widehat{\sigma}_1^* < \beta_0\sigma_0^* + \beta_1\sigma_1^*$.

Now observe that if $\widehat{M}(s) < M(s)$ then $\omega(\widehat{s}^*; \widehat{M}) < \omega(s^*; M)$ since also $\widehat{s}^* > s^*$. By (1.D.5), we must have $\widehat{\sigma}_0^* < \sigma_0^*$. But this means that (1.D.6) can only hold if also the total labor share has decreased, $\beta_0\widehat{\sigma}_0^* + \beta_1\widehat{\sigma}_1^* > \beta_0\sigma_0^* + \beta_1\sigma_1^*$, a contradiction.

Proof that if $\widehat{s}_K \geq s^*$ *then* $\widehat{s}^* > s^*$ Immediate from Lemma 2 which says that $\widehat{s}^* > \widehat{s}_K$. ∎

Proof of Proposition 2 We proceed in three steps.

1. If the labor share increases, then the marginal training-intensive task becomes less complex. Formally, if $\beta_0\widehat{\sigma}_0^* + \beta_1\widehat{\sigma}_1^* < \beta_0\sigma_0^* + \beta_1\sigma_1^*$, then $\widehat{\sigma}_1^* < \sigma_1^*$. For suppose that $\beta_0\widehat{\sigma}_0^* + \beta_1\widehat{\sigma}_1^* < \beta_0\sigma_0^* + \beta_1\sigma_1^*$, but $\widehat{\sigma}_1^* \geq \sigma_1^*$. Then $\widehat{\sigma}_0^* < \sigma_0^*$. By (1.21), $\widehat{s}^* < s^*$. But by (1.20), $\widehat{s}^* > s^*$, a contradiction.

2. If the marginal training-intensive task becomes less complex, then the marginal worker becomes more skilled. Formally, if $\widehat{\sigma}_1^* < \sigma_1^*$, then $\widehat{s}^* > s^*$. For suppose that $\widehat{\sigma}_1^* < \sigma_1^*$ but

[39]To see this, rewrite (1.21) as

$$\frac{\alpha(s_K, \sigma_1^*)}{\alpha(s_K, \sigma_0^*)\alpha(s^*, \sigma_1^*)} = 1.$$

By the log-supermodularity of α, a rise in s_K leads the ratio $\alpha(s_K, \sigma_1^*)/\alpha(s_K, \sigma_0^*)$ to rise since $\sigma_1^* > \sigma_0^*$. Again due to log-supermodularity, the fall in σ_1^* raises the ratio $\alpha(s_K, \sigma_1^*)/\alpha(s^*, \sigma_1^*)$ since $s_K < s^*$. The rise in σ_0^* raises the LHS further. Therefore, s^* must increase.

$\widehat{s}^* \leq s^*$. Then (1.21) implies $\widehat{\sigma}_0^* < \sigma_0^*$. But since $\widehat{V}(\widehat{s}^*) < V(s^*)$, (1.20) implies $\widehat{\sigma}_0^* > \sigma_0^*$, a contradiction.

3. If at one point the new matching function is flatter and does not lie below the old matching function, then it lies above the old one everywhere to the left of this point. Formally, if $\widehat{M}'(s') \leq M'(s')$ and $\widehat{M}(s') \geq M(s')$ for some $s' \in (\max\{s^*, \widehat{s}^*\}, \overline{s}]$, then $\widehat{M}(s) \geq M(s)$ for all $s \in [\max\{s^*, \widehat{s}^*\}, s']$. For suppose that $\widehat{M}'(s') \leq M'(s')$ and $\widehat{M}(s') \geq M(s')$, and that there exists some $s'' \in [\max\{s^*, \widehat{s}^*\}, s')$ such that $\widehat{M}(s'') < M(s'')$. Then there exists some $s''' \in (s'', s')$ such that $\widehat{M}(s''') = M(s''')$, $\widehat{M}'(s''') > M'(s''')$, and $\widehat{M}(s) \geq M(s)$ for all $s \in [s''', s']$. By (1.10),

$$\frac{\widehat{M}'(s''')}{M'(s''')} = \frac{\widehat{w}(s''')/\widehat{w}(s')}{w(s''')/w(s')} \times \frac{\widehat{v}(s''')/\widehat{v}(s')}{v(s''')/v(s')} \times \frac{\widehat{M}'(s')}{M'(s')}.$$

Since $\widehat{V} \succeq V$, and because the upward shift of the matching function raises inequality and thus lowers the wage of type s''' relative to that of type s', the right side of the last equation is no greater than one, so that $\widehat{M}'(s''') \leq M'(s''')$, a contradiction.

Thus, we have shown that if the increase in skill abundance results in an increase in the labor share, then the lower endpoint of the matching function moves southeast (Steps 1 and 2). This means that the matching function must shift down everywhere, for if it shifted up at one point, it would shift up everywhere (Step 3), and it would be impossible for its lower endpoint to move southeast. ∎

1.D.4 Proofs of Corollaries Stated in the Text

Proof of Corollary 1 Integrating (1.11), the first part of the result is immediate given the shift in the matching function and the log-supermodularity of α. The second part follows since $\widehat{w}(s')/\widehat{w}(s) = 1$ but $w(s')/w(s) > 1$ for all such s', s. ∎

Proof of Corollary 2 Recall that the labor share is proportional to $\beta_0(\overline{\sigma} - \sigma_0^*) + \beta_1(\overline{\sigma} - \sigma_1^*)$. As $\widehat{\sigma}_1^* > \sigma_1^*$, the result is immediate if $\widehat{\sigma}_0^* \geq \sigma_0^*$. Then consider the case $\widehat{\sigma}_0^* < \sigma_0^*$. Rewrite (1.20) as

$$A_K \alpha(s_K, \sigma_0^*) K = \frac{\beta_0(\sigma_0^* - \underline{\sigma}) + \beta_1(\sigma_1^* - \underline{\sigma})}{\frac{\beta_0(\overline{\sigma} - \sigma_0^*)}{V(s^*)}}.$$

The LHS increases. If the denominator of the RHS increases, then so must the numerator, which is proportional to the capital share. Hence the labor share decreases. If the denominator of the RHS decreases, then the wage share of all workers falls, again implying a fall in the labor share. ∎

Proof of Corollary 3 Analogous to the proof of Corollary 1. ∎

Chapter 2

A Question of Degree: The Effects of Degree Class on Labor Market Outcomes

2.1 Introduction

Observable characteristics such as years of education and experience explain only a small part of wage dispersion, particularly among high skill workers.[1] One possible explanation invokes unobserved skill differences, but a recent literature also points to the role of luck: macroeconomic conditions at the time of studying and graduation substantially affect graduates' earnings profiles through channels such as selection into firms and sectors.[2] In this chapter, we identify a source of idiosyncratic risk affecting graduates' earnings in the first job, namely, the classification of their degrees.

In the United Kingdom (UK) and other Commonwealth nations, degree class is used as a coarse measure of performance in university degrees.[3] The importance of the system is highlighted by the sizeable fraction of employers who report using the classification in hiring decisions and by universities that use degree class to screen applicants to postgraduate programmes. It is not obvious, however, that the classification system is useful because degree transcripts provide more information about applicant quality.

We investigate whether employers rely on degree class when forming beliefs about a graduate's ability. Identifying the effects of degree class is complicated by the fact that a naive comparison of, say, students who received a First Class with students who received an Upper Second could be biased by the differing ability composition of the two groups. To isolate the casual effect of degree class we need to approximate an ideal experiment and randomly assign degree class across students.

[1] See Lemieux (2006b)

[2] See Oreopoulos, von Wachter, and Heisz (2012), Oyer (2008), and Oyer (2006).

[3] Degrees are classified as First Class, Upper Second Class, Lower Second Class, Third Class, and Pass. This coarse measure of performance stands in contrast to the much more detailed GPA measure used in the US.

Using survey and administrative data from the London School of Economics and Political Science (LSE), we adopt a fuzzy regression discontinuity design (RD) which utilizes institutional rules governing the award of degree class. Undergraduates at the LSE typically take nine courses over three years. Every course is graded out of 100 marks and fixed thresholds are used to map the marks to degree class. A First (Upper Second) Class Honors degree requires at least four marks of 70 (60) or above.[4] We use the discontinuous relationship between degree class and marks received on the fourth highest mark in our RD. This amounts to comparing students who barely made and barely missed a degree class within a narrow window of the marks received. We argue that this generates quasi-experimental variation needed for clean identification of degree class effects.

We find sizeable and significant effects for Upper Second degrees and positive but smaller effects for First Class degrees on wages six months after graduation. A First Class is worth roughly three percent in starting wages which translates into £1,000 per annum. An Upper Second is worth more on the margin—seven percent in starting wages which is roughly £2,040. These results are robust to a battery of specification checks.

Our results imply that employers indeed rely on degree class when forming beliefs about graduates' abilities. This is despite the fact that a graduate's detailed exam grades are typically available to the employer as well—it appears that it would be too costly for employers to process and exploit this richer source of information. Given that employers use degree class to infer an applicant's ability, we use a simple theory of statistical discrimination to interpret results on subgroups. In the model, groups with higher average scores, higher variance in scores or lower variance in the noise associated with the degree class signal, would display stronger effects. We indeed find larger effects for men and mathematical degree programmes as predicted by the simple theory.

The chapter is related to a literature that points towards an important role of luck in determining labor market outcomes. Oreopoulos, von Wachter, and Heisz (2012) document substantial earnings losses associated with graduating during a recession for university graduates in Canada. Oyer (2008) shows that stock market conditions while MBAs students are still in school affect their decision whether to work in the finance industry. Oyer (2006) finds that macroeconomic conditions at the time of graduation affect job characteristics in the short and long run for economists. While these papers focus on the role of aggregate risk, this chapter shows how idiosyncratic risk may affect initial outcomes in the high skill labor market. The earnings differences we identify can be seen as due to luck because it is often a difference of only a few marks that determines degree class. The effect would be exacerbated if initial earnings differences persist due to path dependence in graduates' careers.

Our findings contribute to a literature that documents the importance of simple heuristics for decision making in real world settings. Anderson and Magruder (2012) find substantial effects of Yelp.com ratings on restaurant reservation availability. The ratings are rounded to the nearest half-star. While the true average score is not shown, consumers could in principle calculate the score based on the individual reviews. Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013) find

[4] In terms of letter grades, a mark of 70 or higher would correspond to an A, while a mark between 60 and 69 would be a B.

that prices for used cars drop discontinuously at 10,000-mile odometer thresholds, implying that consumers pay significantly more attention to the first digit than to subsequent ones.

The chapter is related to the literature on the effects of performance in degrees on labor market outcomes. In papers most closely related to this chapter, Di Pietro (2010), Ireland, Naylor, Smith, and Telhaj (2009) and McKnight, Naylor, and Smith (2007) examine the effects of degree classification for students in the UK. Notably Di Pietro (2010) adopts a regression discontinuity design using final year marks and finds no effect on employment. We get similar results on employment but extend the analysis by looking at wage differences. Ireland, Naylor, Smith, and Telhaj (2009) use OLS regressions and find 4 and 5 percent returns to First Class and Upper Second degrees respectively. Their sample consists of a much larger dataset of UK students across many universities and years but does not have the course history information we have to construct finer comparison groups.

This literature interprets the earnings differences associated with degree class as pure signaling effects as in Spence (1973). We hesitate to follow this interpretation because in our setting, unlike in signaling theory, employers have much information in addition to degree class, but choose not to make use of it. Even if employers had no other information than degree class, at best our results would support the presence of a pure signaling effect of degree class for those graduates who fall a few marks to the left or right of the threshold. However, we cannot rule out the possibility that on average, degree class corresponds to the amount of human capital students have accumulated: The decision of how hard to study would then be driven by the costs and benefits of accumulating human capital, rather than by signaling concerns.

The rest of the chapter is organized as follows. In Section 2.2 we discuss the institutional setting, in Section 2.3 we explore the data sources and empirical strategy, in Section 2.4 we present our results and specification checks. Section 2.5 presents a simple model of statistical discrimination and additional results on subgroups that are consistent with the model. Section 2.6 discusses our findings and Section 2.7 concludes.

2.2 Institutional Setting

2.2.1 University Description

Our data come from the London School of Economics and Political Science (LSE). LSE is a highly ranked public research university located in London, UK, specializing in the social sciences. Admission to LSE is highly competitive and it offers a range of degree programmes. In 2012, LSE students came top for employability in the UK in the Sunday Times University Guide. Thus, our results speak to the high end of the skills market.

2.2.2 UK Degree Classification

The degree classification system in the UK is a grading scheme for degrees. The highest distinction for an undergraduate is the First Class honors followed by the Upper Second, Lower Second, Third Class, Pass and Fail degrees. While all universities in the UK follow this classification scheme, each university applies its own standards and rules to determine the distribution of degrees. A similar system operations in other Commonwealth countries including Australia, Canada, India and many others. In the US, a system of Latin Honors performs the similar purpose of classifying degrees. In principle, this implies that our results apply to a broad range of countries.[5] Anecdotal evidence points to the importance of degree class in hiring decisions. One report found that 75 percent of employers in 2012 required at least an Upper Second degree as minimum entry requirement.[6]

2.2.3 LSE Degree Classification Rules

In our identification strategy, we use a unique feature of the rules governing the award of degree class. Undergraduates in the LSE typically take nine courses over three years. Every course is graded out of 100 marks and fixed thresholds are used to map the marks to degree class. As shown in Appendix Table 2.B.1, a First Class Honors degree requires 5 marks of 70 or above or 4 marks of 70 or above with aggregate marks of at least 590. This mapping from course marks to final degree class applies to all departments and years.[7]

We use the discontinuous relationship between degree class and marks received on the fourth highest mark in a fuzzy regression discontinuity design (RD). We employ a fuzzy, as opposed to a sharp, regression discontinuity because the receipt of the degree class also depends on aggregate marks, as shown in Appendix Table 2.B.1. Our strategy is intuitive and amounts to comparing otherwise similar students who differ only in a critical course mark that determines their final degree class.

To be specific, let us consider the award of a First Class degree that depends on the receipt of at least four first class marks. This suggests that the fourth highest mark for any student plays a critical role in determining the degree class. A student whose fourth highest mark is higher than 70 is about twice as likely to obtain a First Class degree as a student whose mark just missed 70, everything else equal. This is seen clearly in Figure 2.1 which plots the fraction of students who receive a First Class degree against their fourth highest mark received. There is a jump in the probability of receiving a First Class after the 70-mark threshold. A similar story is seen in the award of an Upper Second degree at the 60-mark threshold. To summarize, the fourth highest mark plays the role of the assignment variable in our RD strategy.

---

[5] In the US, the grade point average (GPA) system is also used. This is usually a scale from 0 to 4 with one decimal accuracy and is a finer measure of performance than the UK system. There have been calls to scrap the UK system in favor of a GPA system, see “Degree classifications: time for a change?”, the Guardian, July 9th 2012. More recently, a group of UK universities have decided to experiment with a more detailed letter grading scheme, see “Universities testing more detailed degree grades”, BBC News, September 25th, 2013.

[6] See “Top jobs ’restricted to graduates with first-class degrees’ ”, the Daily Telegraph, July 4th 2012 and “Most graduate recruiters now looking for at least a 2:1”, the Guardian, July 4th 2012.

[7] Four courses are taken each year, however only the average of the best three courses in the first year counts towards final classification. Undergraduate law students are an exception and follow a different set of rules. We exclude them from all analyses. Full details of the classification system is available online at the LSE website.

2.3 Data and Empirical Strategy

2.3.1 Student Characteristics and University Performance

From student records we obtain age, gender, nationality and country of domicile information. Course history includes information on degree programme, courses taken and grades awarded, and eventual degree classification. Table 2.1 reports the descriptive statistics of the variables used in our analysis. We have 5,912 students in the population from 2005-2010 of which 2,649 are included in the Destination of Leavers from Higher Education (DLHE) survey (described in detail below). Columns (1) and (4) report the mean and standard deviations of variables for surveyed and non-surveyed students, respectively, while column (5) reports whether the difference is significant. Surveyed students are less likely to be female, more likely to be UK nationals, more likely to receive an Upper Second and less likely to receive a Lower Second.

To implement our empirical strategy, we create two samples. In column (2), the "First Class sample" consists of students who received either a First Class or an Upper Second and the "Upper Second sample" in column (3) consists of students who received either an Upper Second or Lower Second.[8] This provides two discontinuities that we examine separately and narrows our comparisons to students who are on either side of each threshold. In Table 2.1 *First Class*, *Upper Second* and *Lower Second* are dummy variables for the degree classes. Among all surveyed students, the majority of 60 percent received an Upper Second with the remaining 40 percent roughly evenly split between First Class and Lower Second. $1[\textit{4th MARK} \geq 70]$ and $1[\textit{4th MARK} \geq 60]$ are dummy variables equal to one if the fourth highest mark is no less than 70 or 60 respectively.

One shortcoming of this database is that we do not have measures of a student's pre-university ability. For a typical UK student this might include her GCSE and A-level results. Although admissions to LSE programmes require A-level or equivalent results to be reported, these data are not collected centrally but are received by each department separately. To partly address this shortcoming, in all our regressions we control for department $\times$ year fixed effects.[9] Furthermore, the validity of our RD strategy does not rely on controlling for ability. As noted in Lee and Lemieux (2010) an RD design mimics a natural experiment close to the discontinuity. Hence there should be no need for additional controls except to improve precision of estimates.

2.3.2 Labor Market Outcomes

Data on labor market outcomes come from the DLHE survey which is a national survey of students who have recently graduated from a university in the UK. This survey is conducted twice a year to find out employment circumstances of students six months after graduation.[10] Due to the frequency of the survey and its statutory nature, LSE oversees the survey and reports the results to HESA (Higher Education Statistics Authority). The survey is sent by email and responded to online

[8] We dropped Third Class and below because they constituted less than 5 percent of the population. Including them among the Lower Second population did not change results.

[9] Results in McKnight, Naylor, and Smith (2007) suggest that controlling for degree programme reduces the importance of pre-university academic results.

[10] The surveys are conducted from November to March for the "January" survey, and from April to June for the "April" survey.

and includes all students including non-domiciled and non-UK nationals. Typically response rates are higher for domiciled and UK nationals.[11] The survey provides us with data from 2005-2010. Our key variables of interest are industry and employment status. Industry is coded in four digit SIC codes, although we aggregate to two digits for merging with LFS data (see Section 2.3.3). In Table 2.1, "employed" is a dummy variable equal to one if a graduate is employed in full-time work. [12]

Table 2.1 shows that 85 percent of students who responded are employed within six months of graduation. More than one-third are employed in the finance industry although this varies slightly across the degree classes (see Appendix Table 2.A.1). Given the importance of the finance industry, we construct a dummy variable for employment in finance as a separate outcome variable and look at results excluding the finance industry.

Because the survey is conducted six months after graduation, we interpret our analysis as applying to first jobs. Although we do not observe previous job experience and cannot control for this in our analysis, 98 percent of our students were younger than 21 years of age when they started their degrees. Thus, any work experience is unlikely to have been in permanent employment. Also, we cannot follow students over longer periods of employment to examine the dynamic effects of degrees.

A further concern is that employment six months after graduation may have been secured before the final degree class is known. Anecdotes suggest that students start Summer internships, work experience and job applications prior to graduation. The more common it is that students sign job contracts before graduating, the less likely we are to pick up any effects of degree class. However, anecdotal evidence suggests that many early job offers are conditional on achieving a specified degree class. If students then narrowly miss the requirement of their job offer and are forced to work in jobs with lower pay, then this would be picked up by our identification strategy.

2.3.3 Labor Force Survey

We merge wage data from the LFS into the DLHE survey at the industry $\times$ year $\times$ gender level. We calculate mean log hourly wages for each industry $\times$ year $\times$ gender cell unconditional on skills or experience. One concern with this approach is that mean wages are not representative of the earnings facing undergraduates. To address this concern we also calculate mean log wages conditional on university and three experience levels. To match the labor market prospects of undergraduates we chose 1, 3 and 5 years of potential experience.

This gives us five different measures of industry wages– overall mean, university with 1, 3 and 5 years of experience and overall mean for the sub-sample of students in non-finance industries. Our preferred measure is the overall mean because it provides a clean measure of the industry's "rank" compared to other industries. In any case the five measures are highly correlated with pairwise correlations never less than 0.8. Table 2.1 shows that the mean log wage is 2.45 which is roughly

[11] Formally, LSE is required to reach a response rate of 80 percent for UK nationals and 50 percent for others.

[12] Self-employed, freelance and voluntary work is coded as zero along with the unemployed or unable to work. An annual salary question is included but response is voluntary and less than half report it. The correlation between reported salary and industry salary is 0.39.

£11.60 per hour in 2005£. As expected, industry wages increase in years of experience.

Using industry wages implies that we do not have within-industry variation in outcomes. The lack of a more direct wage measure is an issue for other studies in the literature as well (Di Pietro 2010, McKnight, Naylor, and Smith 2007). Appendix Table 2.A.1 shows the top 15 industries ranked by total share of employment. Even accounting for the large share in finance, there is substantial distribution in employment across industries– of the 84 two-digit SIC codes, 66 are represented in our data.

2.3.4 Empirical Strategy

Our unit of observation is a student. For each student we observe her degree classification and her course grades. In particular, we observe her fourth highest mark taken over three years of the degree. As described in Section 2.2.3, institutional rules imply that the fourth highest mark is critical in determining her degree class. When the fourth highest mark crosses the 70-mark or 60-mark cutoff, there is a discontinuous jump in the probability of receiving a First Class and Upper Second respectively. We use a dummy variable for the fourth highest mark crossing these thresholds as an instrument for the degree class "treatment".

Identification in a fuzzy RD setup requires the continuity assumption (Lee and Lemieux 2010).[13] Apart from the treatment– in this case degree class– all other observables and unobservables vary continuously across the threshold. This also means that the assignment variable should not be precisely manipulated by agents. We cannot test the continuity of the unobservables directly. Instead we test the continuity of observables. Second, we employ the McCrary (2008) test to see if there is a discontinuity in the probability density of the treatment which may suggest manipulation of the assignment variable. These are discussed in Section 2.4.2.

In our benchmark specification we use a non-parametric local linear regression with a rectangular bandwidth of 5 marks above and below the cutoff (Imbens and Wooldridge 2009). This means we include the fourth mark linearly and interacted with the dummy variable as additional controls. A non-parametric approach observes that a regression discontinuity is a kernel regression at a boundary point (Imbens and Lemieux 2008). This motivates the use of local regressions with various kernels and bandwidths (Fan and Gijbels 1996, Li and Racine 2007). Although a parametric function such as a high order polynomial is parsimonious it is found to be quite sensitive to polynomial order (Angrist and Pischke 2009). In specification checks we vary the bandwidth and try polynomial functions to flexibly control for the fourth mark. As discussed in Section 2.4.4 these specification checks produce qualitatively similar results.

In theory, identification in an RD setup comes in the limit as we approach the discontinuity asymptotically (Hahn, Todd, and Van der Klaauw 2001). In practice, this requires sufficient data around the boundary points– as we get closer to the discontinuity estimates tend to get less precise

[13] Regression discontinuity was introduced by Thistlethwaite and Campbell (2001) and formalized in the language of treatment effects by Hahn, Todd, and Van der Klaauw (2001). The close connection between fuzzy RD and instrumental variables is noted in Lee and Lemieux (2010), Imbens and Lemieux (2008) and Imbens and Wooldridge (2009). Instead of the usual exclusion restrictions, however, we require the continuity assumption and non-manipulation of the assignment variable.

because we have fewer data. Furthermore, when the assignment variable is discrete by construction, there is the additional complication that we cannot approach the boundary infinitesimally.[14] In this chapter, we choose the 5 mark bandwidth as a reasonable starting point and accept that some of the identification necessarily comes from marks away from the boundary. We follow Lee and Card (2008) in correcting standard errors for the discrete structure of our assignment variable by clustering on marks throughout.

We write the first-stage equation as:

$$\begin{aligned} \text{CLASS}_i &= \delta_0 + \delta_1 1[\text{4th MARK} \geq \text{cutoff}]_i + \delta_2(\text{4th MARK}_i - \text{cutoff}) \\ &+\delta_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\delta_4 + u_i \end{aligned} \tag{2.1}$$

where $CLASS$ is either First Class or Upper Second and the cutoff is 70 or 60 respectively. $1[\textit{4th MARK} \geq \textit{cutoff}]$ is a dummy variable for the fourth mark crossing the cutoff and our instrument for the potentially endogenous degree class. X is a vector of covariates including female dummies, age and age squared, dummies for being a UK national, dummies for having resat or failed any course, 15 dummies for department, 5 year of graduation dummies and 75 dummies for department $\times$ year of graduation interactions.

We use the predicted degree class from our first-stage regression in our second-stage equation:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1\text{CLASS}_i + \beta_2(\text{4th MARK}_i - \text{cutoff}) \\ &+\beta_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\beta_4 + \epsilon_i \end{aligned} \tag{2.2}$$

where Y are various labor market outcomes including employment status, employment in finance industry and five measures of industry wages.

2.4 Results

2.4.1 First-Stage and Reduced Form Regressions

In this section we report results from the first-stage 2.1 and the reduced form regressions:

$$\begin{aligned} Y_i &= \gamma_0 + \gamma_1 1[\text{4th MARK} \geq \text{cutoff}]_i + \gamma_2(\text{4th MARK}_i - \text{cutoff}) \\ &+\gamma_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\gamma_4 + \nu_i \end{aligned} \tag{2.3}$$

where Y are the various labor market outcomes.

Table 2.2, column (1), reports the first-stage results for the First Class discontinuity (panel A) and Upper Second discontinuity (panel B). Both first-stage F-statistics are above the rule-of-thumb threshold of 10 and mitigate any concerns about weak instruments (Staiger and Stock 1997, Stock, Wright, and Yogo 2002).[15] In order to better interpret the first-stage, we look at the relationship between fourth highest mark and degree class without controlling for any covariates. This also

[14]This is also a problem facing designs where age in years or months is the assignment variable, e.g. Carpenter and Dobkin (2009).

[15]The sample size varies over outcome variables but we confirmed that the first-stage and other results are not sensitive to these sample differences.

allows us to do a simple count of the complier population in LSE (Angrist, Imbens, and Rubin 1996, Imbens and Angrist 1994). In Figure 2.2 the schematic shows the breakdown of students into compliers, always takers and never takers around the discontinuity. For instance, always takers are students who receive a First Class regardless of their fourth highest mark, while compliers are students who receiver a First Class *because* their fourth highest mark crosses the threshold. The breakdown suggests that the complier population is sizeable at 87 percent. This is expected because the institutional rules are strictly followed and supports the validity of our results to the rest of the LSE population.

Columns (2) to (3) report the reduced form regressions for the extensive margin of employment. Both First Class and Upper Second discontinuities show insignificant results. Columns (4) to (8) report the reduced form results for industry wages. In panel A, the results for the First Class discontinuity are positive but insignificant. In panel B, we find stronger and significant results for the Upper Second discontinuity.[16] The reduced form evidence is presented graphically in Figure 2.4. Anticipating the RD results, Figure 2.5 and Figure 2.6 suggest that the effect of degree class on wages is driven by males.

2.4.2 Randomization Checks and McCrary Test

As discussed in Section 2.3.4, identification in an RD setup requires continuity in the observables (and unobservables) across the threshold as well as non-manipulation of the assignment variable. To test for continuity in the observables, we regress each covariate on the treatment dummy in Table 2.3, columns (1) to (5). Apart from age in the First Class sample and gender in the Upper Second sample, the results are consistent with the lack of discontinuity in the observables. The apparent discontinuity in age and gender does not worry us because these are non-manipulable attributes (Holland 1986). In other words, there is less concern that agents could have taken actions to manipulate these attributes around the discontinuity to improve their degree class.

To test for the manipulation of the assignment variable, McCrary (2008) suggests using the frequency count as the dependent variable in the RD setup. The idea is that manipulation of the assignment variable should result in bunching of individuals at the cutoff. In the education literature, this was shown to be an important invalidation of the RD approach (see for e.g. Urquiola and Verhoogen (2009)). In our case, we should see a jump in the number of students at the threshold of 70 or 60 marks. In column (6) of Table 2.3 we perform the McCrary test and find large and (in the case of the Upper Second threshold) significant jumps in the number of students. *Prima facie*, this might suggest that students are manipulating their marks in order to receive better degrees.

We argue that this bunching is not the result of manipulation but is a consequence of institutional features. Figure 2.3 plots the histogram of the highest to the sixth highest marks. In every case there is a clear bunching of marks at 60 and 70 even for the highest mark which is not critical for eventual degree class. This is because exam graders actively avoid giving borderline marks (i.e. 59 or 69) and either round up or down.[17] One may still worry that students who received 58 or

[16] The coefficients on the slopes in the reduced form wage regression for the Upper Second discontinuity (panel B column 4) suggests that the visually negative slope in Figure 2.4 is not significant.

[17] In LSE, exams are taken anonymously and each script is graded by two internal examiners. Having graded

68 may appeal to have their script re-graded. From discussions with staff, the appeals process is arduous and rarely successful. Nonetheless we follow the literature in dealing with the potential manipulation of marks by excluding the threshold in specification checks reported in Section 2.4.4 (see for e.g. Almond and Doyle (2011), Almond, Joseph J. Doyle, Kowalski, and Williams (2010) and Barreca, Guldi, Lindo, and Waddell (2011)). Doing so does not change our results.[18]

2.4.3 Effects of Degree Class on Labor Market Outcomes

Table 2.4 reports the results for the effects of receiving a First Class degree compared with an Upper Second. In panel A, we compare average differences in outcomes without controlling for any covariates. There are no differences in employment in general or in the finance industry specifically. However, there are significant differences in industry wages. Using our preferred measure of mean industry log wages in column (3), a First Class receives seven percent higher wages. Conditional wage measures in columns (4) to (7) paint a similar picture. Panel B includes covariates to allow for closer comparisons of students. This corresponds to estimating 2.2 using OLS. The employment outcomes remain insignificant while the wage coefficients halve but remain significant. In panel C we report our benchmark RD model. We instrument for the First Class treatment using a dummy variable for the fourth highest mark crossing the 70 mark threshold. Although the difference in industry mean wages remains significant at 5 percent, the conditional experience measures are insignificant suggesting that the wage differences for a First Class are not precisely measured.

Table 2.5 reports the same specifications for the Upper Second degree. There are no significant differences in average outcomes across students without controlling for covariates in panel A. This is because of inter-departmental comparisons we are making in the absence of department fixed effects. Once we control for covariates including department by year fixed effects in panel B we observe that an Upper Second receives 4 percent higher wages than a Lower Second in column (3). Conditional wage measures in columns (4) to (7) are smaller in magnitude but show similar positive estimates. An Upper Second also has a 7 percentage point (20 percent) higher probability of working in finance. Using the dummy variable $1[\text{4th MARK} \geq 60]$ as an instrument for Upper Second, panel C reveals that the returns are significant and sizeable at 7 percent for mean wages and 12 percentage points (37 percent) for finance industry employment. Conditional wage measures in columns (4) to (7) offer a qualitatively similar picture of positive wage effects.

To interpret these results we translate the percentage differences to pounds. Using our preferred measure of wages in the specification in column (3) we find that a First Class and Upper Second are worth around $\pounds$1,000 and $\pounds$2,040 per annum respectively in current money.[19]

---

each script separately, graders convene to deliberate on the final mark. External examiners grade scripts for which no agreement could be reached.

[18] An alternative identification strategy would be to restrict the sample to students whose fourth mark is the average of the best three first-year exams, which are aggregated to count as one mark in determining degree class. This would produce a more balanced frequency count. It would also address the concern that examiners may sometimes determine a ten-mark bin based on a general impression, and only later decide the precise mark (conversations with staff suggest that this marking strategy is common in essay-based exams such as in philosophy or history). Unfortunately, the sample we obtain from this restriction is much too small to yield precise results.

[19] Assuming a 40 hour week for 52 weeks for a full time worker using 23 percent CPI inflation from 2005-2012. First Class: $\exp(2.473) \times 40 \times 52 \times 1.23 \times 0.033$. Upper Second: $\exp(2.418) \times 40 \times 52 \times 1.23 \times 0.071$.

2.4.4 Specification Checks

We conduct a battery of specification tests of our RD results. In Table 2.A.3 we report checks for the First Class degree while Table 2.A.4 reports the same for Upper Second. Each row is a different specification check and the columns are the different dependent variables. We report the coefficient and standard error on the degree class dummy and the number of observations. Row (1) reports the benchmark results for comparison.

Rows (2) to (10) report results using different bandwidth sizes (our benchmark is a 5-mark bandwidth). Rows (11) to (14) report specifications using parametric polynomial controls. In rows (15) and (16) we include controls for the sum of marks and all other marks separately to show that our results are not driven by omission of other course grades. In row (17) we address the concern that our results misrepresent students who are not domiciled in UK by looking only at domiciled students. In row (18) we deal with the worry that bunching of marks around the threshold reflects manipulation.

Employment outcomes appear to be sensitive to bandwidth choice. For the First Class some specifications even suggest a negative effect on employment, e.g. rows (3) and (4). Likewise for the Upper Second degree, employment outcomes do not display a consistent pattern across specifications. To be conservative we interpret this as suggesting that the extensive margin is not affected by degree class. This is similar to Di Pietro (2010) who did not find significant effects on employment. This may be due to the limited variation we have in employment and requires further investigation in future work. In the following sections we focus on the industry wage outcomes.

We find consistent results when we look at industry mean wages. Looking at industry means for First Class degrees, we find effects significant at 5 percent ranging from 2.5 to 6.8 percent with the benchmark result of 3.3 percent. For Upper Second, the range is 5.7 to 13 percent with the benchmark of 7.1 percent.

2.5 A Model of Statistical Discrimination and Additional Results

We have shown the effects of degree class on industry wages. We interpret these results as employers inferring applicant's ability based on degree class. To strengthen this interpretation we present a simple model of statistical discrimination and show additional results consistent with the theoretical predictions.

2.5.1 Simple Model of Statistical Discrimination

In statistical discrimination, employers differentiate across otherwise identical workers on the basis of observable group membership, for example race or gender. More recent versions of these models introduce the dynamics of employer learning (Farber and Gibbons 1996, Lange 2007, Altonji and Pierret 2001, Arcidiacono, Bayer, and Hizmo 2010). Our exposition follows Aigner and Cain (1977) and Belman and Heywood (1991) (see also Hungerford and Solon (1987) and Jaeger and Page (1996)).

Suppose employers observe a noisy signal of student ability:

$$y = q + u$$

where y is the signal, q is unobserved ability and u is a normally distributed mean zero random variable uncorrelated with q. Note that on average the signal is unbiased, $E[y] = E[q]$. Students know their own ability but employers only see y and know that q is distributed with mean $\bar{q}$ and some variance σ_q. Therefore, employers pay wages that are equal to the expected ability of students conditional on their signal. That is, employers solve a signal extraction problem:

$$wages = E[q|y] = (1 - \gamma)\bar{q} + \gamma y$$

which is a regression of q on y where linearity follows from the normality assumption. The regression coefficient is written as:

$$\gamma = \frac{\sigma_q}{\sigma_q + \sigma_u}$$

where σ_u is the variance of the noise term.

Additionally, employers observe a student's group. Suppose there are two groups, A and B, with means and variances $\bar{q}^A$, $\bar{q}^B$, σ^A and σ^B. For any observed signal y, the difference in predicted ability between groups is:

$$\begin{aligned} E[q|y, A] - E[q|y, B] =& (1 - \gamma^A)\bar{q}^A + \gamma^A y - (1 - \gamma^B)\bar{q}^B - \gamma^B y \\ =& (\bar{q}^A - \bar{q}^B)(1 - \gamma^B) + (y - \bar{q}^A)(\gamma^A - \gamma^B) \end{aligned}$$

This formula gives us three predictions. Given some signal y, the wages to group A are higher than group B, $E[q|y, A] - E[q|y, B] > 0$, if

1. $\bar{q}^A - \bar{q}^B > 0$, average signal is higher in group A than B
2. $\sigma_q^A - \sigma_q^B > 0$ and $y > \bar{q}$, ability variance is higher in group A than B for a "good" signal
3. $\sigma_u^A - \sigma_u^B < 0$ and $y > \bar{q}$, noise variance is lower in group A than B for a "good" signal.

We bring this theory to the data by interpreting y as the fourth highest mark. Fourth highest marks determine degree class and are a noisy signal of students' abilities. The total variance in marks, σ_y, is the sum of the variance in ability, σ_q, and the noise variance, σ_u. We can now re-state our theoretical predictions. At any given mark and resulting degree class, a student from group A has a higher predicted wage than an otherwise identical student from group B if:

1. group A has higher average marks than group B;
2. group A has higher variance in marks than group B;
3. group A has lower variance in the noise term than group B.

In our context, a positive signal is receipt of the higher degree class. Both First Class and Upper Second are positive signals because we are always comparing to the next lower class. Note that we do not actually observe the noise term or its variance, so we cannot exactly decompose the differences in average wages.

In the next section we define two groups in the data. First, we define groups by gender. Second, we group degree programmes by their math admissions requirements. Math admissions requirements are a measure of how mathematical the degree is. Mathematical degrees exhibit higher means and variances in marks than less mathematical degrees. This may be because less mathematical degrees have essay based courses which are more subjective in grading. We show that our estimates by groups are largely consistent with the simple theory of statistical discrimination.

2.5.2 Statistical Discrimination by Gender and Degree Programmes

The fourth highest mark is our measure of the signal from the theory described in Section 2.5.1. Appendix Table 2.A.2 presents the means and standard deviations of the fourth highest mark by the different groups. Males tend to have higher marks on average than females, and they tend to have higher variance in their marks.

Next we differentiate degree programmes. Appendix Table 2.B.2 lists the degree programmes in our sample. Using information on the math entry requirements, we distinguish between programmes which required at least A-level in maths and those which do not. As seen in Appendix Table 2.A.2, when we split degree programmes by their math requirements, mathematical degrees have higher average and variance in marks.

Table 2.6 presents our estimates by gender. We estimate our benchmark RD specification for each group separately. We find that First Class effects are significant and positive for males at 6 percent but insignificant and basically zero for females– this translates into £1,780 a year.[20] Upper Second effects are larger in magnitude for males but imprecisely estimated for both.

Table 2.7 splits the sample by degree programmes. For both First Class and Upper Second, mathematical programmes display larger and significant effects. A First Class is worth 6 percent in a mathematical degree compared with an insignificant 4 percent on a non-mathematical degree. Likewise, an Upper Second is worth 15 percent in a mathematical degree compared to zero in a non-mathematical degree.

In sum, these results by group are consistent with our simple theory of statistical discrimination.

2.6 Discussion

Our findings suggest that employers use degree class to form beliefs about a graduate's ability. But why would degree class matter if employers could obtain full transcripts of all course marks? With transcripts, employers should use course marks as finer signals of ability instead of using the cruder degree class. Our findings of effects from degree class, even after controlling for course marks, suggests that employers either do not observe transcripts or observe transcripts but do not fully use the information on them.

[20]Assuming a 40 hour week for 52 weeks for a full time worker using 23 percent CPI inflation from 2005-2012, $\exp(2.454) \times 40 \times 52 \times 1.23 \times 0.06$.

If the computational costs of understanding diverse transcripts is too high, employers could rely on degree class to form rules-of-thumb, or heuristics, in making hiring and salary decisions. As a rough gauge of the potential computational costs, Appendix Table 2.B.3 counts the number of modules taken by students across departments. In the department of government, for example, students took a total of 167 different modules. This suggests that it may be difficult for employers to compare course level marks to differentiate between candidates if transcripts are too diverse.

2.7 Conclusion

In this chapter we estimate the effects of university degree class on initial labor market outcomes using a regression discontinuity design that utilizes university rules governing the award of degrees. We find sizeable and significant effects for Upper Second degrees and positive but smaller effects for First Class degrees on wages—a First Class and Upper Second are worth around £1,000 and £2,040 per annum, respectively. However, we do not find significant effects on the extensive margin of employment. These results are robust to a battery of specification checks.

Explaining wage dispersion on the basis of observable characteristics has been challenging, especially in the case of high skill workers (Lemieux 2006b). We make progress on this issue by establishing a causal effect of degree class on graduates' earnings in the first job. We contribute to the literature that emphasizes the role of luck in labor market outcomes (Oreopoulos, von Wachter, and Heisz 2012, Oyer 2006, Oyer 2008). While this literature documents the importance of aggregate risk, we identify a source of idiosyncratic risk for the determinations of earnings. Our results also add to previous findings showing the importance of simple heuristics for assessing quality, for instance in the used car market or in the restaurant business (Busse, Lacetera, Pope, Silva-Risso, and Sydnor 2013, Anderson and Magruder 2012).

Our results inform a policy debate about the adequacy of degree class as a measure of degree performance. Recently, several universities in the UK have decided to experiment with a more detailed letter grading scale. Our results suggests that a more detailed grading scheme may lead to a better match of graduates' pay and ability (at least as measured by performance in university exams). On the other hand, our results also suggest that the more detailed the grading scheme, the higher the risk that employers use coarse heuristics rather than utilizing all of the information that is available.

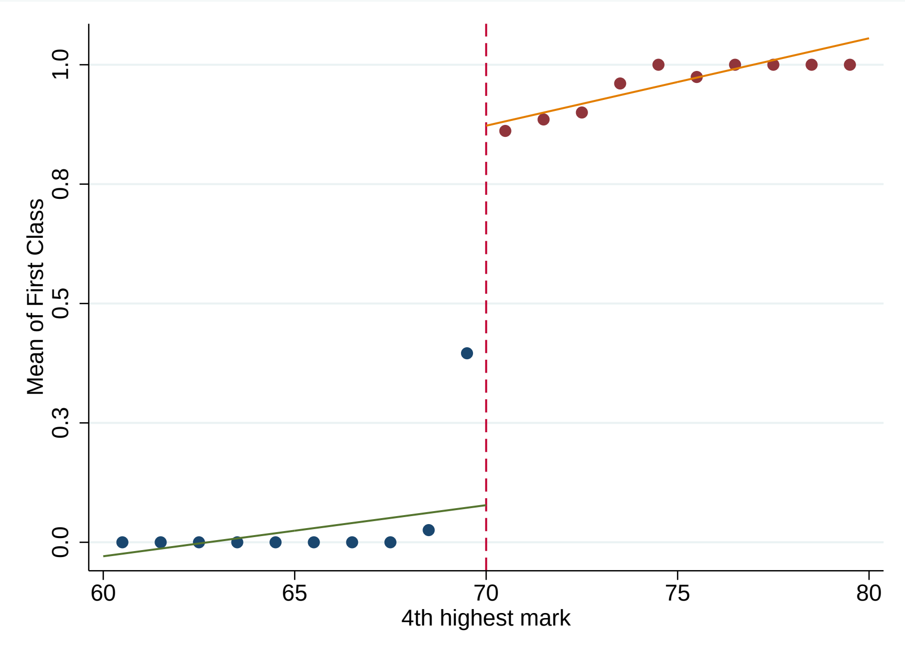
An important question that we cannot answer with our data is whether the initial differences in earnings due to degree class persist over time. Since students close to the threshold on either side have similar productivity, the effects of degree class may attenuate over time as employers learn about workers' productivities.[21] However, if initial labor market outcomes persist, we may observe earnings differences over the experience profile.

[21] The literature on employer learning argues that any signal used in initial labor market outcomes attenuates over time as employers discover more about ability (Farber and Gibbons 1996, Altonji and Pierret 2001, Lange 2007, Arcidiacono, Bayer, and Hizmo 2010). Empirically, this means that the effects of schooling attenuate over time while coefficients on hard-to-observe variables like test scores increase over time (Altonji and Pierret 2001).

Figures and Tables

Figure 2.1: Expected Degree Classification and Fourth Highest Marks

(a) Expected First Class degree, 10 marks above and below 70



(b) Expected Upper Second degree, 10 marks above and below 60

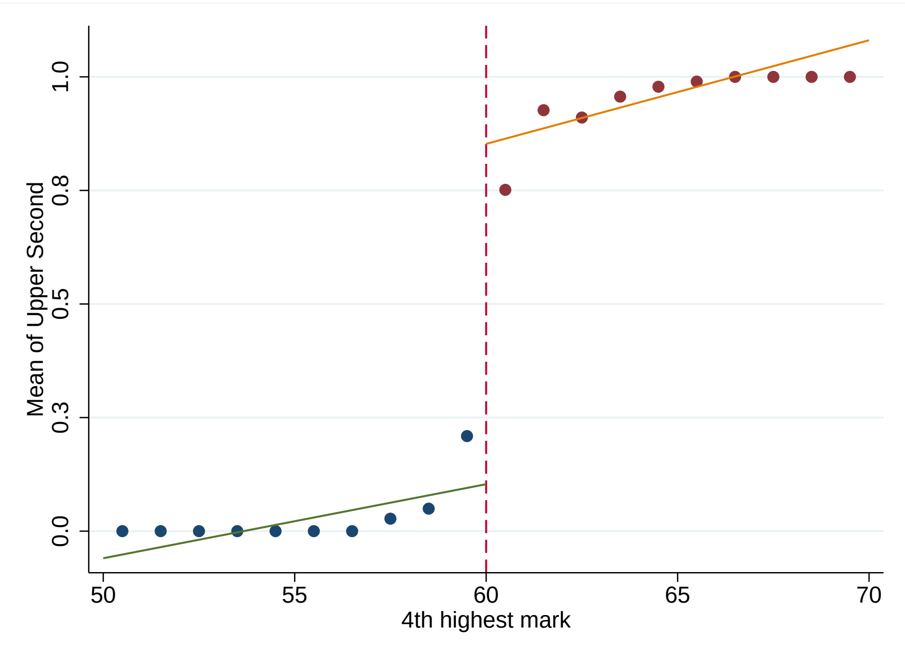


Figure 2.2: Counting Compliers

(a) Schematic

Assignment variable is above threshold

		0	**1**
Degree Class	**0**	Never takers + Compliers	Never takers
	1	Always takers	Always takers + Compliers

(b) First Class sample (N = 1,136)

4th highest mark is above 70

		0	**1**
First Class	**0**	652	44
	1	23	417

Always Takers = 3% = 23/(23+652)
Never Takers = 10% = 44/(44+417)
Compliers = 87%

(c) Upper Second sample (N = 1,406)

4th highest mark is above 60

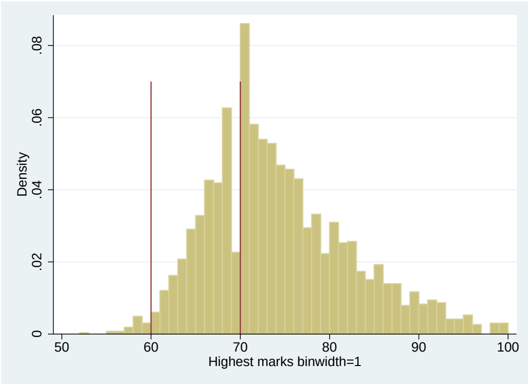
		0	**1**
Upper Second	**0**	307	87
	1	16	996

Always Takers = 5% = 16/(16+307)
Never Takers = 8% = 87/(87+996)
Compliers = 87%

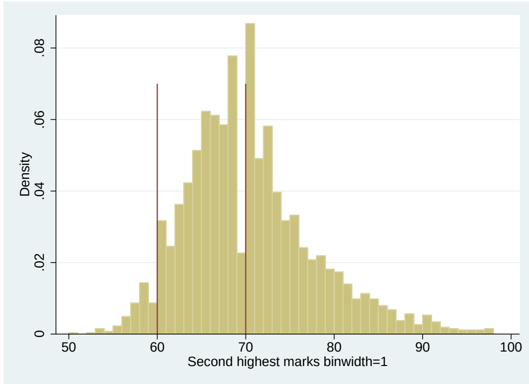
Notes: Compliers are students who received their degree class because their 4th highest marks crossed the relevant threshold.

Figure 2.3: Histogram of Marks

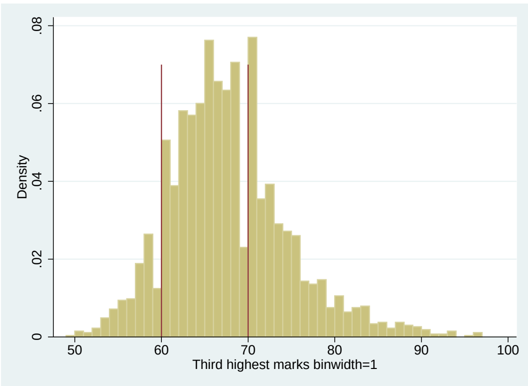
(a) Highest marks



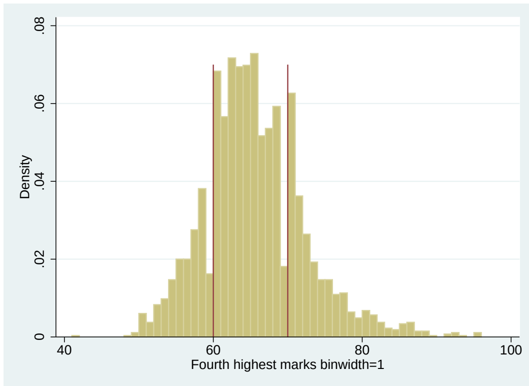
(b) Second highest marks



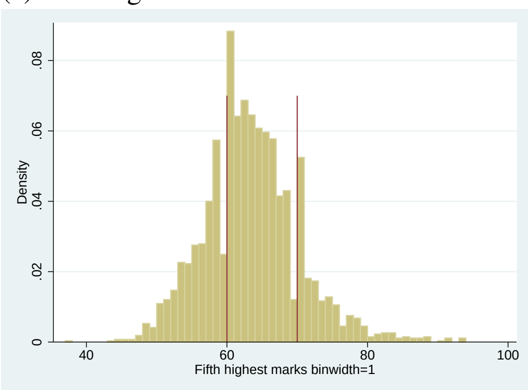
(c) Third highest marks



(d) Fourth highest marks



(e) Fifth highest marks



(f) Sixth highest marks

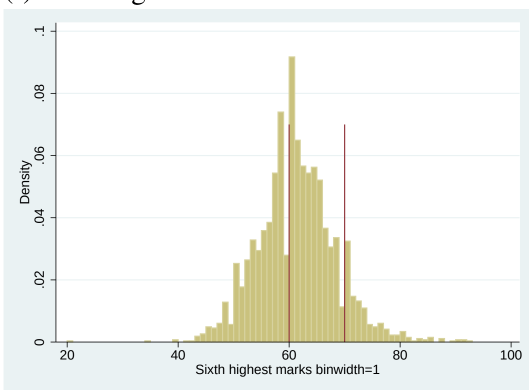
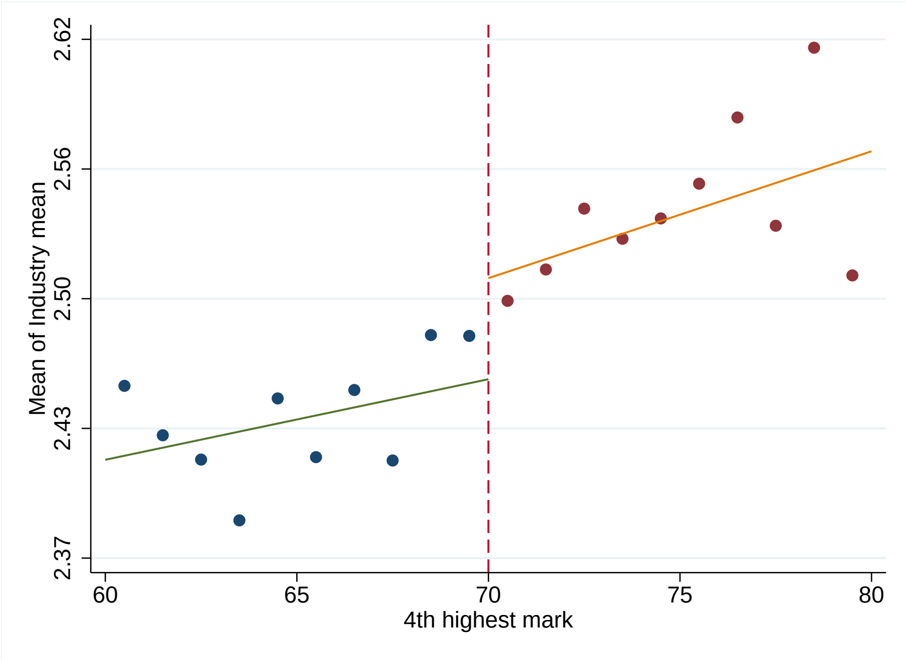


Figure 2.4: Expected Industry Mean Log Wages on Fourth Highest Marks

(a) 10 marks above and below 70



(b) 10 marks above and below 60

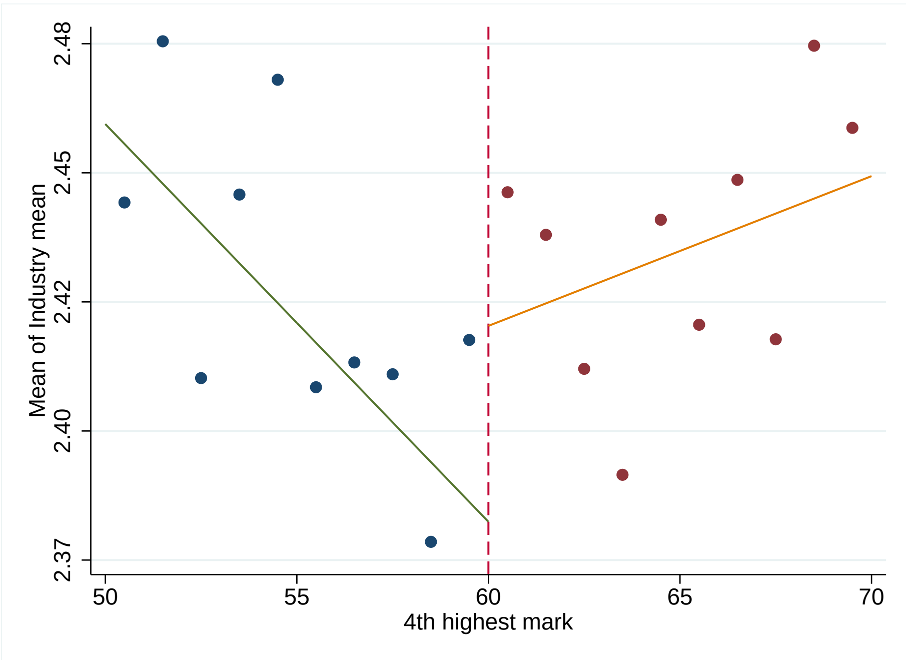
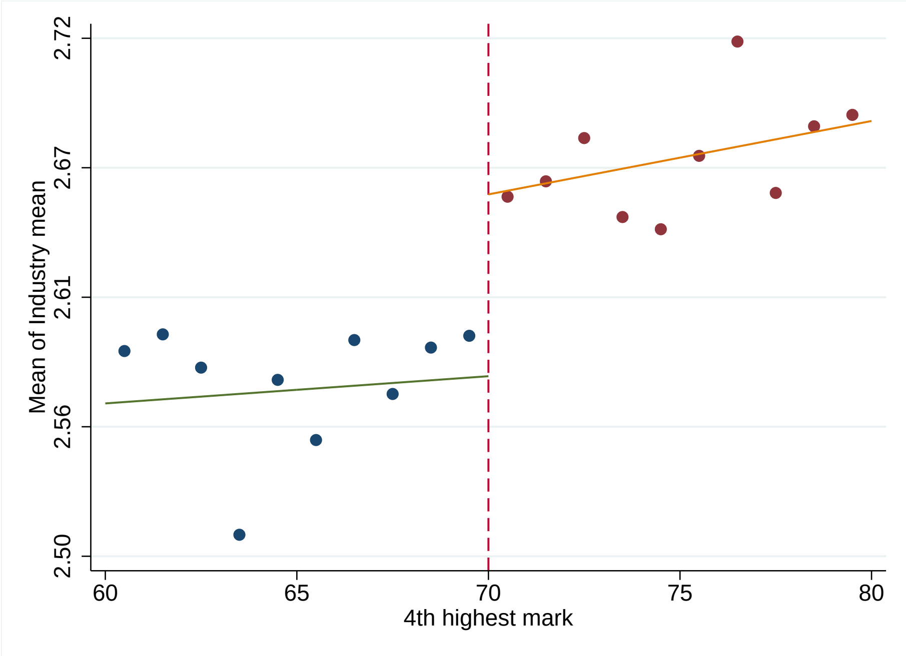


Figure 2.5: Expected Industry Mean Log Wages on Fourth Highest Marks, Males

(a) 10 marks above and below 70



(b) 10 marks above and below 60

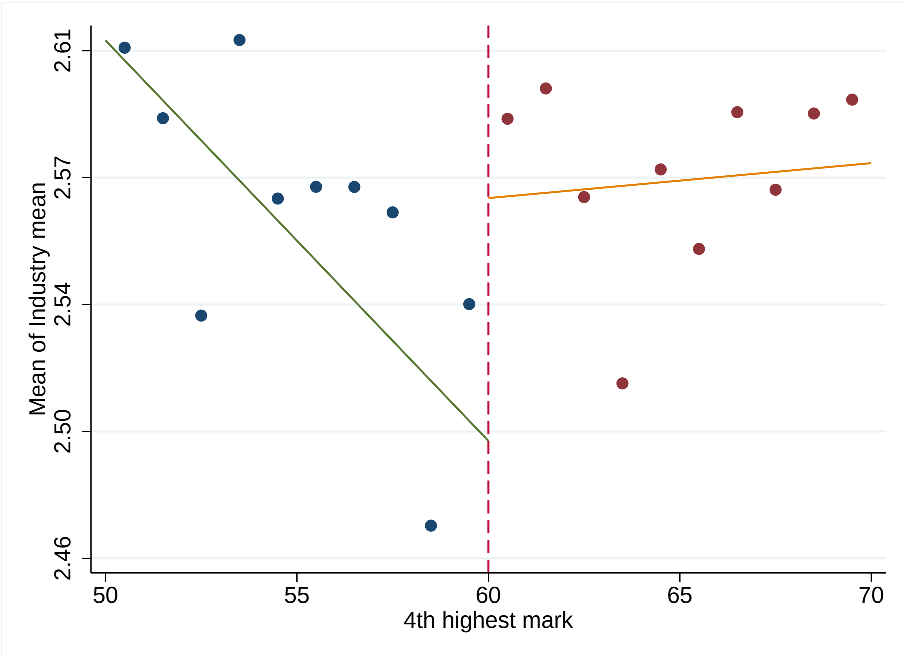
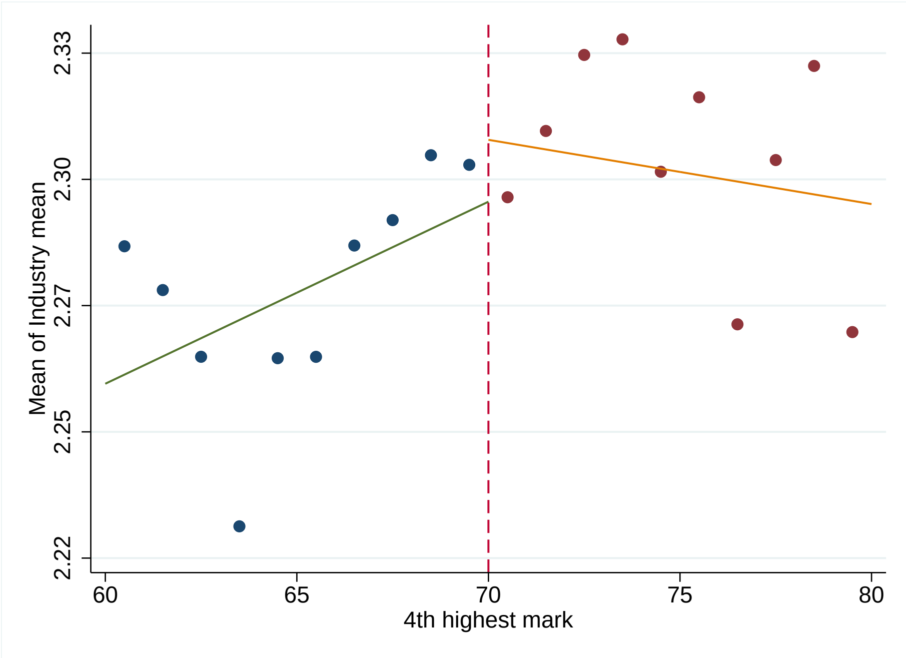


Figure 2.6: Expected Industry Mean Log Wages on Fourth Highest Marks, Females

(a) 10 marks above and below 70



(b) 10 marks above and below 60

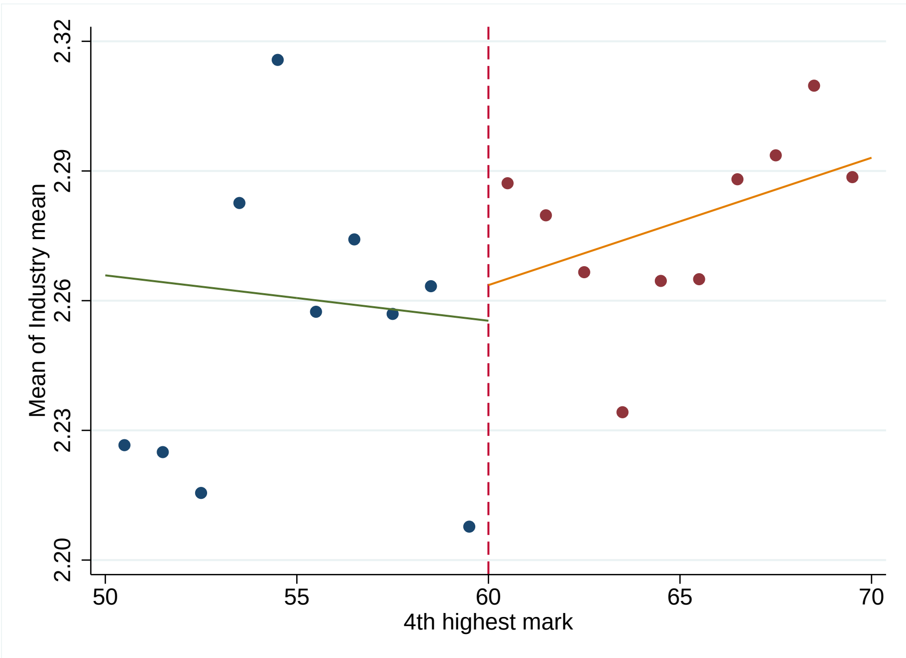


Table 2.1: Descriptive Statistics

	No. of obs	Surveyed: Total	Surveyed: First Class sample	Surveyed: Upper Second sample	Not surveyed	Difference significant (1) - (4)
		(1)	(2)	(3)	(4)	(5)
Number of observations	5912	2649	1136	1406	3263	
Female	5912	0.45	0.45	0.48	0.51	***
Age	5912	22.06	22.03	22.06	22.10	
UK national	5912	0.60	0.59	0.66	0.42	***
Resat any course	5912	0.10	0.03	0.13	0.11	
Failed any course	5912	0.06	0.02	0.08	0.06	
First Class	5912	0.23	0.39	0.00	0.25	
Upper Second	5912	0.57	0.61	0.72	0.53	***
Lower Second	5912	0.19	0.00	0.28	0.22	**
4th highest mark	5912	65.10	68.63	61.31	65.08	
1(4th mark ≥ 70)	5912	0.24	0.41	0.00	0.25	
1(4th mark ≥ 60)	5912	0.83	1.00	0.77	0.81	**
Employed	2649	0.85	0.86	0.83		
Finance industry	2244	0.38	0.42	0.32		
Industry mean log wages (2005£)						
Industry mean	2244	2.45	2.47	2.42		
		(0.24)	(0.23)	(0.25)		
College with 1 year experience	2244	2.14	2.15	2.11		
		(0.18)	(0.18)	(0.19)		
College with 3 years experience	2244	2.34	2.35	2.31		
		(0.18)	(0.18)	(0.19)		
College with 5 years experience	2244	2.48	2.50	2.45		
		(0.19)	(0.18)	(0.19)		
Industry mean excluding finance	1389	2.38	2.40	2.35		
		(0.23)	(0.22)	(0.24)		

Notes: This table shows variable means and standard deviations (in parentheses) where applicable. Surveyed students are respondents to the Destination of Leavers from Higher Education (DLHE) survey conducted six months after a student graduates. The First Class sample includes surveyed students who received either a First Class or Upper Second degree and whose fourth highest mark is within 5 marks of 70. The Upper Second sample includes surveyed students who received either an Upper Second or Lower Second degree and whose fourth highest mark is within 5 marks of 60. *First Class*, *Upper Second* and *Lower Second* are dummy variables for degree class. *4th highest mark* is the fourth highest mark received by the student among all full-unit equivalent courses taken. *1(4th mark ≥ 70)* and *1(4th mark ≥ 60)* are dummy variables for the fourth highest mark being at least 70 or 60, respectively. *Employed* is an indicator for whether a student is in employment 6 months after graduation. Self-employment, voluntary work and further studies are not considered employment. *Finance industry* is an indicator for working in the finance industry. Industry mean log wages are measures of hourly wages in two-digit SIC industry × year × gender cells. Two-digit SIC industry wage data is taken from the Labor Force Survey and rebased to 2005£. ***, **, * significant at the 1, 5 and 10 percent level.

Table 2.2: First Stage and Reduced Form Regressions for First Class and Upper Second Degrees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: First Class discontinuity								
				Industry mean log wages				
	First Class	Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
1(4th mark $\geq$ 70)	0.673***	0.007	0.007	0.022	0.014	0.009	0.012	0.035
	(0.124)	(0.034)	(0.054)	(0.014)	(0.013)	(0.012)	(0.011)	(0.023)
(4th Highest Mark - 70)	0.046	-0.006	0.017	0.007*	0.008**	0.009**	0.008**	0.006
	(0.030)	(0.012)	(0.016)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)
(4th Highest Mark - 70) * 1(4th mark $\geq$ 70)	-0.016	0.006	-0.050**	-0.011**	-0.013***	-0.013***	-0.013***	-0.005
	(0.031)	(0.015)	(0.017)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Obs	1,136	1,136	978	978	978	978	978	567
R-sq	0.803	0.205	0.255	0.606	0.437	0.405	0.466	0.496
First-stage F-stat	29.2							

(continued…)

Table 2.2: First Stage and Reduced Form Regressions for First Class and Upper Second Degrees (cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel B: Upper Second discontinuity								
				Industry mean log wages				
	Upper Second	Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
1(4th mark ≥ 60)	0.670***	-0.024	0.080	0.048**	0.036**	0.046**	0.032*	0.042*
	(0.078)	(0.030)	(0.050)	(0.020)	(0.015)	(0.016)	(0.016)	(0.019)
(4th Highest Mark - 60)	0.031	0.004	-0.013	-0.002	-0.004	-0.005	-0.003	-0.004
	(0.018)	(0.006)	(0.010)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
(4th Highest Mark - 60) *	0.006	0.006	0.015	-0.000	0.001	0.002	0.000	0.002
1(4th mark ≥ 60)	(0.022)	(0.007)	(0.015)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Obs	1,406	1,406	1,168	1,168	1,168	1,168	1,168	796
R-sq	0.722	0.103	0.203	0.484	0.353	0.321	0.368	0.405
First-stage F-stat	74.8							

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. Each cell reports a different regression. All regressions are estimated by OLS. All regressions include female dummies, age and age squared, dummies for being a UK national, dummies for having resat or failed any course, 15 dummies for department, 5 year dummies and 75 dummies for department × year interactions. Column (1) reports the first-stage regression of degree class on an indicator for marks crossing the relevant cutoff. The first stage F-stat for excluded instruments is reported in the last row of each panel. Columns (2) to (8) report reduced form regressions of labor market outcomes on the cutoff instrument.

Table 2.3: Testing the Randomization of Instruments Around the First Class and Upper Second Discontinuities

	(1) Female	(2) Age	(3) UK national	(4) Resat any course	(5) Failed any course	(6) No. of students in each mark
Panel A: First Class discontinuity						
1(4th mark ≥ 70)	-0.001	-0.158*	0.012	-0.001	-0.009	62.8
	(0.055)	(0.071)	(0.060)	(0.022)	(0.011)	(40.0)
(4th Highest Mark - 70)	-0.002	0.024	-0.007	-0.002	-0.005	-20.0**
	(0.012)	(0.025)	(0.009)	(0.006)	(0.005)	(8.90)
(4th Highest Mark - 70)	-0.016	0.013	-0.011	-0.004	0.004	-8.40
* 1(4th mark ≥ 70)	(0.012)	(0.036)	(0.016)	(0.007)	(0.005)	(10.7)
Obs.	1136	1136	1136	1136	1136	1136
Panel B: Upper Second discontinuity						
1(4th mark ≥ 60)	0.103**	0.119	-0.031	0.041	0.002	80.8**
	(0.036)	(0.383)	(0.066)	(0.054)	(0.064)	(31.9)
(4th Highest Mark - 60)	-0.033**	-0.093	0.014	-0.036**	-0.014	6.30
	(0.014)	(0.088)	(0.019)	(0.015)	(0.017)	(7.53)
(4th Highest Mark - 60)	0.025*	0.084	-0.011	0.017	0.006	-2.06
* 1(4th mark ≥ 60)	(0.012)	(0.092)	(0.020)	(0.015)	(0.017)	(7.95)
Obs.	1406	1406	1406	1406	1406	1406

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. Each cell reports a different regression. All regressions are estimated by OLS. All regressions include covariates, 15 dummies for department, 5 year dummies and 75 dummies for department × year interactions.

Table 2.4: The Effects of Obtaining a First Class Degree Compared to an Upper Second Degree on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Industry mean log wages				
	Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
Panel A: OLS without any covariates							
First Class	0.019	0.069	0.070***	0.062***	0.061***	0.062***	0.077***
	(0.023)	(0.042)	(0.015)	(0.013)	(0.012)	(0.013)	(0.020)
Obs	1136	978	978	978	978	978	567
Panel B: OLS							
First Class	-0.022	0.013	0.037***	0.033***	0.035***	0.030***	0.052***
	(0.019)	(0.035)	(0.007)	(0.007)	(0.008)	(0.007)	(0.013)
Obs	1136	978	978	978	978	978	567
Panel C: RD							
First Class	0.011	0.010	0.033**	0.021	0.014	0.018	0.054**
	(0.045)	(0.074)	(0.016)	(0.015)	(0.015)	(0.014)	(0.024)
(4th Highest Mark - 70)	-0.006	0.016	0.005	0.007**	0.008***	0.007**	0.003
	(0.012)	(0.017)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
(4th Highest Mark - 70) *	0.006	-0.050***	-0.011**	-0.013***	-0.013***	-0.013***	-0.004
1(4th mark $\geq$ 70)	(0.014)	(0.017)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)
Obs	1136	978	978	978	978	978	567

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. Each cell reports a different regression. All regressions include female dummies, age and age squared, dummies for being a UK national, dummies for having resat or failed any course, 15 dummies for department, 5 year dummies and 75 dummies for department × year interactions. See notes to Table 3.1 for descriptions of variables.

Table 2.5: The Effects of Obtaining an Upper Second Degree Compared to a Lower Second Degree on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Industry mean log wages				
	Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
Panel A: OLS without any covariates							
Upper Second	-0.004	0.029	0.020	0.001	0.001	0.005	-0.007
	(0.015)	(0.022)	(0.011)	(0.013)	(0.015)	(0.012)	(0.015)
Obs	1406	1168	1168	1168	1168	1168	796
Panel B: OLS							
Upper Second	0.027	0.069**	0.040***	0.025**	0.027**	0.028**	0.028**
	(0.015)	(0.030)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Obs	1406	1168	1168	1168	1168	1168	796
Panel C: RD							
Upper Second	-0.035	0.118**	0.071***	0.052***	0.067***	0.048**	0.063**
	(0.043)	(0.058)	(0.024)	(0.019)	(0.019)	(0.019)	(0.026)
(4th Highest Mark - 60)	0.005	-0.016*	-0.004	-0.005	-0.007*	-0.004	-0.006
	(0.006)	(0.009)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
(4th Highest Mark - 60) *	0.006	0.014	-0.001	0.001	0.001	-0.000	0.001
1(4th mark $\geq$ 60)	(0.006)	(0.012)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Obs	1406	1168	1168	1168	1168	1168	796

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. Each cell reports a different regression. All regressions include female dummies, age and age squared, dummies for being a UK national, dummies for having resat or failed any course, 15 dummies for department, 5 year dummies and 75 dummies for department × year interactions. See notes to Table 3.1 for descriptions of variables.

Table 2.6: RD Estimates by Gender

	(1) Industry mean	(2) College with 1 year experience	(3) College with 3 years experience	(4) College with 5 years experience	(5) Industry mean excl. finance
Panel A: First Class Degree					
Male					
First Class	0.059***	0.048***	0.048***	0.048***	0.054
	(0.013)	(0.013)	(0.013)	(0.013)	(0.050)
Obs	549	549	549	549	290
Female					
First Class	-0.022	-0.032	-0.032	-0.028	-0.034
	(0.029)	(0.024)	(0.023)	(0.022)	(0.057)
Obs	429	429	429	429	277
Panel B: Upper Second Degree					
Male					
Upper Second	0.084	0.081	0.089*	0.077	0.082
	(0.059)	(0.050)	(0.049)	(0.050)	(0.060)
Obs	618	618	618	618	397
Female					
Upper Second	0.052	0.034	0.036	0.029	0.062
	(0.042)	(0.041)	(0.037)	(0.038)	(0.075)
Obs	550	550	550	550	399

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks

Table 2.7: RD Estimates by Programme Admissions Math Requirements

	(1) Industry mean	(2) College with 1 year experience	(3) College with 3 years experience	(4) College with 5 years experience	(5) Industry mean excl. finance
Panel A: First Class Degree					
At least A level maths					
First Class	0.063***	0.045**	0.039**	0.039	0.124***
	(0.015)	(0.021)	(0.019)	(0.024)	(0.047)
Obs	576	576	576	576	259
No math requirement					
First Class	0.038	0.002	-0.002	0.003	0.034
	(0.036)	(0.038)	(0.041)	(0.037)	(0.031)
Obs	402	402	402	402	308
Panel B: Upper Second Degree					
At least A level maths					
Upper Second	0.146***	0.107***	0.118***	0.091***	0.171*
	(0.051)	(0.030)	(0.031)	(0.028)	(0.100)
Obs	550	550	550	550	304
No math reauirement					
Upper Second	-0.004	-0.011	0.005	-0.004	-0.007
	(0.042)	(0.032)	(0.031)	(0.036)	(0.031)
Obs	618	618	618	618	492

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks

Appendix

2.A Additional Descriptive Statistics and Results

Table 2.A.1: Top 15 Industries Ranked by Total Share of Employment

Industry (LFS, SIC two-digit)	Industry mean log wages (2005£)	Share of employment Total	First Class	Upper Second	Lower Second and below
financial ex insurance and pension	2.58	38.10	47.90	36.28	31.00
legal and accounting activities	2.52	16.22	21.21	14.43	15.15
public admin, defence, social sec	2.35	7.44	5.85	8.52	6.29
head offices; management consultancies	2.51	6.51	8.04	6.23	5.36
insurance, reinsurance and pension	2.45	4.55	4.75	3.79	6.53
education	2.36	3.88	2.01	4.97	3.03
advertising and market research	2.48	2.01	1.10	2.37	2.10
security & investigation activities	1.99	1.74	0.37	2.05	2.56
office admin, support and other	2.15	1.52	0.18	1.58	3.03
retail trade, except vehicles	1.88	1.47	0.73	1.58	2.10
auxiliary to financial and insuranc	2.55	1.34	1.46	1.50	0.70
other prof, scientific and technical	2.22	1.07	0.73	1.26	0.93
publishing activities	2.40	0.85	0.37	0.87	1.40
employment activities	2.24	0.80	0.18	1.18	0.47
human health activities	2.24	0.80	0.18	0.87	1.40

Notes: This table shows the industry mean log wages for all skills and experience groups. Industries are ranked by total share of employment.

Table 2.A.2: Summary Statistics by Groups

	First Class Sample		Upper Second	
	4th Mark mean	4th Mark S.D.	4th Mark mean	4th Mark S.D.
By gender				
Male	67.56	6.00	62.33	4.47
Female	66.60	5.40	62.32	4.32
By programme math requirements				
At least A level maths	68.74	6.57	62.33	4.75
No math requirement	65.39	4.07	62.32	4.06

Notes: This table shows summary statistics by gender and programme characteristics.

Table 2.A.3: Specification Checks for First Class Degree Specification

				Industry mean log wages				
		Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
(1)	Benchmark	0.011	0.010	0.033**	0.021	0.014	0.018	0.054**
		(0.045)	(0.074)	(0.016)	(0.015)	(0.015)	(0.014)	(0.024)
		1136	978	978	978	978	978	567
(2)	1 mark above and below threshold	0.033	0.193	0.018	0.016	0.023	0.006	-0.121
		(0.125)	(0.211)	(0.058)	(0.050)	(0.052)	(0.052)	(0.123)
		310	270	270	270	270	270	150
(3)	2 marks above and below threshold	0.146	0.732*	0.199*	0.014	0.037	0.049	-0.206
		(0.284)	(0.400)	(0.106)	(0.080)	(0.091)	(0.085)	(1.001)
		537	469	469	469	469	469	252
(4)	3 marks above and below threshold	-0.164**	0.251*	0.042**	0.010	0.014	0.006	0.009
		(0.065)	(0.139)	(0.019)	(0.017)	(0.017)	(0.021)	(0.071)
		730	629	629	629	629	629	345
(5)	4 marks above and below threshold	-0.117***	0.210***	0.068***	0.050***	0.038**	0.047***	0.046*
		(0.026)	(0.057)	(0.017)	(0.015)	(0.015)	(0.014)	(0.027)
		906	774	774	774	774	774	426
(6)	6 marks above and below threshold	-0.017	0.009	0.044***	0.031***	0.031**	0.027***	0.074***
		(0.030)	(0.053)	(0.011)	(0.011)	(0.013)	(0.010)	(0.021)
		1346	1147	1147	1147	1147	1147	671
(7)	7 marks above and below threshold	-0.012	-0.010	0.025*	0.015	0.015	0.012	0.054***
		(0.028)	(0.037)	(0.013)	(0.012)	(0.012)	(0.010)	(0.018)
		1552	1322	1322	1322	1322	1322	790
(8)	8 marks above and below threshold	-0.022	0.005	0.038***	0.032**	0.032**	0.029**	0.061***
		(0.024)	(0.037)	(0.013)	(0.013)	(0.013)	(0.013)	(0.017)
		1742	1478	1478	1478	1478	1478	884
(9)	9 marks above and below threshold	-0.025	0.038	0.051***	0.045***	0.046***	0.044***	0.071***
		(0.024)	(0.043)	(0.009)	(0.010)	(0.010)	(0.011)	(0.013)
		1894	1602	1602	1602	1602	1602	953
(10)	10 marks above and below threshold	-0.018	0.011	0.056***	0.049***	0.050***	0.047***	0.080***
		(0.025)	(0.043)	(0.007)	(0.008)	(0.009)	(0.009)	(0.015)
		2048	1735	1735	1735	1735	1735	1045
(11)	2nd order polynomial	0.009	0.054	0.043***	0.033***	0.026**	0.030**	0.058**
		(0.037)	(0.055)	(0.013)	(0.013)	(0.013)	(0.012)	(0.024)
		1136	978	978	978	978	978	567
(12)	3rd order polynomial	-0.006	0.108	0.049*	0.032	0.016	0.032	0.010
		(0.063)	(0.127)	(0.026)	(0.030)	(0.029)	(0.027)	(0.033)
		1136	978	978	978	978	978	567
(13)	4th order polynomial	-0.133***	0.205**	0.051*	0.029	0.015	0.026	0.011
		(0.029)	(0.093)	(0.029)	(0.034)	(0.033)	(0.032)	(0.037)
		1136	978	978	978	978	978	567
(14)	5th order polynomial	-0.086*	0.025	-0.002	-0.026	-0.036	-0.024	-0.007
		(0.045)	(0.144)	(0.033)	(0.047)	(0.044)	(0.040)	(0.060)
		1136	978	978	978	978	978	567

(continued…)

Table 2.A.3: Specification Checks for First Class Degree Specification (cont.)

				Industry mean log wages				
		Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
(15)	Including controls	0.010	0.010	0.032**	0.020	0.013	0.017	0.052**
	for sum of marks	(0.044)	(0.073)	(0.015)	(0.015)	(0.015)	(0.013)	(0.022)
		1136	978	978	978	978	978	567
(16)	Including controls	0.011	0.021	0.034**	0.024	0.017	0.020	0.051**
	for other marks	(0.045)	(0.073)	(0.015)	(0.015)	(0.015)	(0.014)	(0.023)
		1136	978	978	978	978	978	567
(17)	UK domicile	-0.015	0.138	0.031	0.047**	0.035*	0.039**	-0.007
	sample	(0.063)	(0.094)	(0.025)	(0.021)	(0.020)	(0.019)	(0.040)
		701	585	585	585	585	585	367
(18)	Excluding marks	-0.002	0.008	0.048***	0.035**	0.036***	0.028**	0.078***
	around disc.	(0.062)	(0.094)	(0.011)	(0.014)	(0.012)	(0.012)	(0.017)
		922	791	791	791	791	791	462

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. This table reports specification checks of the benchmark model in Table 3.4, panel C. Each cell reports a different regression where the coefficients on *First Class* are reported in the first lines, standard errors in brackets and number of observations in the third lines.

Table 2.A.4: Specification Checks for Upper Second Degree Specification

				Industry mean log wages				
		Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
(1)	Benchmark	-0.035	0.118**	0.071***	0.052***	0.067***	0.048**	0.063**
		(0.043)	(0.058)	(0.024)	(0.019)	(0.019)	(0.019)	(0.026)
		1406	1168	1168	1168	1168	1168	796
(2)	1 mark above and	-0.004	0.006	0.095**	0.046	0.063	0.042	0.192***
	below threshold	(0.103)	(0.117)	(0.047)	(0.042)	(0.042)	(0.041)	(0.053)
		374	310	310	310	310	310	211
(3)	2 marks above and	-0.144**	0.022	0.054	-0.017	0.008	-0.016	0.142
	below threshold	(0.070)	(0.088)	(0.053)	(0.037)	(0.044)	(0.034)	(0.096)
		665	546	546	546	546	546	367
(4)	3 marks above and	-0.113*	-0.014	0.082***	0.043	0.064**	0.044	0.107**
	below threshold	(0.063)	(0.079)	(0.031)	(0.028)	(0.028)	(0.029)	(0.048)
		922	759	759	759	759	759	517
(5)	4 marks above and	-0.029	0.068	0.093***	0.061**	0.075**	0.065**	0.100***
	below threshold	(0.060)	(0.074)	(0.035)	(0.031)	(0.030)	(0.031)	(0.030)
		1160	954	954	954	954	954	648
(6)	6 marks above and	-0.018	0.133**	0.080***	0.059**	0.072***	0.054**	0.067**
	below threshold	(0.038)	(0.064)	(0.030)	(0.025)	(0.025)	(0.024)	(0.028)
		1582	1310	1310	1310	1310	1310	877
(7)	7 marks above and	-0.002	0.086	0.084***	0.056***	0.066***	0.052***	0.072***
	below threshold	(0.032)	(0.060)	(0.026)	(0.021)	(0.021)	(0.020)	(0.023)
		1750	1448	1448	1448	1448	1448	962
(8)	8 marks above and	-0.030	0.114**	0.064**	0.042*	0.051**	0.038*	0.035
	below threshold	(0.035)	(0.056)	(0.028)	(0.022)	(0.023)	(0.021)	(0.039)
		1925	1602	1602	1602	1602	1602	1047
(9)	9 marks above and	-0.011	0.095*	0.057**	0.033	0.045**	0.033*	0.033
	below threshold	(0.037)	(0.054)	(0.026)	(0.021)	(0.021)	(0.020)	(0.032)
		1964	1637	1637	1637	1637	1637	1069
(10)	10 marks above	-0.014	0.055	0.047*	0.021	0.030	0.021	0.024
	and below	(0.032)	(0.058)	(0.024)	(0.021)	(0.022)	(0.020)	(0.027)
	threshold	2003	1672	1672	1672	1672	1672	1092
(11)	2nd order	-0.024	0.081	0.084***	0.061***	0.076***	0.055***	0.078***
	polynomial	(0.041)	(0.075)	(0.026)	(0.018)	(0.019)	(0.017)	(0.025)
		1406	1168	1168	1168	1168	1168	796
(12)	3rd order	0.006	-0.040	0.125***	0.090***	0.106***	0.080***	0.138***
	polynomial	(0.053)	(0.076)	(0.033)	(0.023)	(0.026)	(0.024)	(0.028)
		1406	1168	1168	1168	1168	1168	796
(13)	4th order	-0.036	-0.113	0.121***	0.071**	0.095***	0.063*	0.158***
	polynomial	(0.066)	(0.104)	(0.046)	(0.033)	(0.033)	(0.035)	(0.042)
		1406	1168	1168	1168	1168	1168	796
(14)	5th order	-0.035	-0.166	0.132***	0.069**	0.101***	0.053	0.183***
	polynomial	(0.067)	(0.103)	(0.045)	(0.033)	(0.035)	(0.033)	(0.047)
		1406	1168	1168	1168	1168	1168	796

(continued…)

Table 2.A.4: Specification Checks for Upper Second Degree Specification (cont.)

				Industry mean log wages				
		Employed	Finance industry	Industry mean	College with 1 year exp.	College with 3 years exp.	College with 5 years exp.	Industry mean excl. finance
(15)	Including controls	-0.037	0.105*	0.065**	0.047**	0.063***	0.043**	0.060**
	for sum of marks	(0.042)	(0.059)	(0.026)	(0.020)	(0.020)	(0.020)	(0.027)
		1406	1168	1168	1168	1168	1168	796
(16)	Including controls	-0.043	0.117*	0.071***	0.052***	0.067***	0.046**	0.062**
	for other marks	(0.051)	(0.060)	(0.026)	(0.020)	(0.020)	(0.020)	(0.027)
		1406	1168	1168	1168	1168	1168	796
(17)	UK domicile	-0.083*	0.033	0.091***	0.076***	0.087***	0.064***	0.102***
	sample	(0.042)	(0.059)	(0.023)	(0.021)	(0.023)	(0.022)	(0.032)
		974	792	792	792	792	792	574
(18)	Excluding marks	-0.036	0.214***	0.077***	0.055***	0.068***	0.056***	0.055*
	around disc.	(0.040)	(0.033)	(0.022)	(0.015)	(0.014)	(0.017)	(0.029)
		1182	978	978	978	978	978	654

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered by marks. This table reports specification checks of the benchmark model in Table 3.5, panel C. Each cell reports a different regression where the coefficients on Upper Second are reported in the first lines, standard errors in parentheses and number of observations in the third lines.

2.B Further Information on Institutional Background

Table 2.B.1: Mapping From Course Marks to Final Degree Class

Final degree class	Course grade requirements
First Class Honors	5 marks of 70 or above or 4 marks of 70 or above and aggregate marks of at least 590
Upper Second Class	5 marks of 60 or above or 4 marks of 60 or above and aggregate marks of at least 515
Lower Second Class	5 marks of 50 or above or 4 marks of 50 or above and aggregate marks of at least 440

Notes: Institutional rules governing award of degree class taken from http://www.lse.ac.uk/resources/calendar/academicRegulations/BA-BScDegrees.htm

Table 2.B.2: Degree Programmes

department	programme	No. of students	Math required
Accounting	BSc in Accounting and Finance	367	0
Anthropology	BA in Anthropology and Law	20	0
Anthropology	BA in Social Anthropology	26	0
Anthropology	BSc in Social Anthropology	63	0
Economic History	BSc in Economic History	72	0
Economic History	BSc in Economic History with Economics	8	1
Economic History	BSc in Economics and Economic History	30	1
Economics	BSc in Econometrics and Mathematical Economics	23	1
Economics	BSc in Economics	510	1
Economics	BSc in Economics with Economic History	11	1
Employment Relations and Organisational Behaviour	BSc in Human Resource Management and Employment Relations	32	0
Employment Relations and Organisational Behaviour	BSc in Industrial Relations and Human Resource Management	7	0
Geography & Environment	BA in Geography	65	0
Geography & Environment	BSc in Environmental Policy	12	0
Geography & Environment	BSc in Environmental Policy with Economics	12	1
Geography & Environment	BSc in Geography and Population Studies	2	0
Geography & Environment	BSc in Geography with Economics	53	1
Government	BSc in Government	68	0
Government	BSc in Government and Economics	96	1
Government	BSc in Government and History	48	0
International History	BA in History	89	0
International History	BSc in International Relations and History	60	0
International Relations	BSc in International Relations	132	0
Management Science Group	BSc in Management Sciences	78	1
Managerial Economics and Strategy Group	BSc in Management	132	1
Mathematics	BSc in Mathematics and Economics	126	1
Philosophy	BA in Philosophy	2	0

(continued…)

Table 2.B.2: Degree Programmes (cont.)

department	programme	No. of students	Math required
Philosophy	BSc in Philosophy	5	0
Philosophy	BSc in Philosophy and Economics	70	1
Philosophy	BSc in Philosophy, Logic and Scientific Method	30	0
Social Policy	BSc in Population Studies	1	0
Social Policy	BSc in Social Policy	21	0
Social Policy	BSc in Social Policy and Administration	5	0
Social Policy	BSc in Social Policy and Criminology	11	0
Social Policy	BSc in Social Policy and Economics	11	1
Social Policy	BSc in Social Policy and Government	2	0
Social Policy	BSc in Social Policy and Sociology	11	0
Social Policy	BSc in Social Policy with Government	20	0
Social Policy	BSc in Social Policy with Social Psychology	1	0
Social Policy	BSc in Social Policy, Criminal Justice and Psychology	10	0
Sociology	BSc in Sociology	77	0
Statistics	BSc in Actuarial Science	137	1
Statistics	BSc in Business Mathematics and Statistics	93	1

Notes: N=2,649. Taken from http://www2.lse.ac.uk/study/undergraduate/degreeProgrammes2013/degreeProgrammes2013.aspx. *Math required* is a dummy variable for whether the programme requires A-level maths for admissions.

Table 2.B.3: Number of Modules Taken by Students in Department

Department	No. of Modules
Accounting	100
Anthropology	90
Economic History	99
Economics	143
Employment Relations and Organisational Behaviour	76
Geography & Environment	84
Government	167
International History	125
International Relations	104
Management Science Group	46
Managerial Economics and Strategy Group	72
Mathematics	54
Philosophy	104
Social Policy	98
Sociology	86
Statistics	77

Notes: Number of different modules taken by students in the department. Students can take modules offered by other departments

Chapter 3

Do Secondary School Admission Policies Affect Intergenerational Transmission of Education? Evidence from Germany

3.1 Introduction

As income inequality has risen dramatically in most developed economies over the past three decades, there has been renewed interest in the topic of intergenerational mobility. For a given level of cross-sectional income dispersion, the "character of inequality" (Solon 1999) in an economy may differ greatly depending on how strong is the persistence of socioeconomic status (wealth, income, education etc.) across generations. It is therefore an important question whether public policy can affect intergenerational mobility. In this spirit, I investigate in this chapter whether changes in secondary school admission policies in Germany caused a decline in the degree of intergenerational transmission of educational attainment.

Parental education has historically been a strong predictor of children's educational attainment in Germany. In particular, Dustmann (2004) finds that parental education is strongly related to the secondary school track the child attends, and that this association translates into substantial earnings differences later in life. Although primary and secondary education are provided tuition free, intergenerational mobility of educational attainment may be impeded by early tracking of students according to ability. The German school system features early ability tracking, whereby students leave primary school at age ten and proceed to one of three secondary school tracks. Only the most advanced of these, henceforth referred to as the *high track*, prepares students for university attendance. There are barriers to attending the high track in many states, where a recommendation by the primary school teacher is required, often based on marks received in fourth grade. Such restrictive admission policies may lead to a greater role for parental background in determining track attendance, since academic performance in primary school is positively related to parental education.

I take advantage of changes to transition rules within states to see whether free track choice leads to a diminished association of parental background with track attendance. I have collected data on transition rules in the German states, covering the period 1983 to the present. These data show frequent policy changes. For instance, in the early 1990s the states Hessen, Nordrhein-Westfalen, and Rheinland-Pfalz (the 'treatment group') introduced free track choice while other states (the 'control group') maintained restrictive transition rules. By comparing changes in high track enrolment probabilities in the treatment group to changes in the control group, and thus employing a differences-in-differences (DD) strategy, I estimate the causal effect of free track choice on high track enrolment. Using micro data from the German Socio-Economic Panel (SOEP), I find that a child's probability of attending the high track is on average five percentage points higher under free track choice. However, the effect is twice as large for children of less educated parents (those who did not attend the high track themselves). In additional results, I find that free track choice differentially benefits females and children with an immigrant background.

Taken at face value, the results imply that the correlation between parents' and children's educational attainment is reduced by more than one third when no formal restrictions to choosing a secondary school track exist. However, my baseline estimates are for high track *attendance* at the beginning of secondary school, not for track *completion*, which is the true measure of educational attainment. Since I observe graduations for only a few cases in my data due to attrition, I cannot estimate the effect of free track choice on high track completion. Instead, I look at whether an individual attends the high track when last observed in the sample, but no earlier than four years after starting secondary school. While the effect of free track choice for children of less educated parents is reduced by half and no longer precisely estimated, I still find a strong differential effect of free track choice by parental background. Taken together, my results strongly suggest that the introduction of free track choice leads to lower persistence of educational attainment across generations.

To the best of my knowledge, this chapter is the first study to employ a DD strategy to estimate the effects of free track choice on high track enrolment and on the association of parental education with track choice.[1] The findings inform the ongoing controversy in Germany about the appropriateness of the transition rules in place, and about the extent to which parents should be free to choose a secondary school track for their child regardless of the child's performance in primary school. States continue to experiment, with Baden-Württemberg in 2012 being the most recent state that introduced free track choice. The message to policy makers of this chapter is that free track choice increases enrolment in the high track modestly on average, but by a substantial amount for children of less educated parents, thus reducing the association of parental education with track choice. The case of Germany is informative because of the frequent and decentralised policy changes that have taken place. The programme evaluation approach would not be feasible in countries with more centralised school systems such as France or Italy, or in cases where the

[1]Dollmann (2011) employs an event study approach to estimate the effects of the abolishment of free track choice in Nordrhein-Westfalen in 2006. He finds that, conditional on performance, free track choice leads to a *stronger* correlation between parental background and high track enrolment. My results concern instead the *unconditional* correlation of parental and children's education, which is arguably more relevant for policy.

stringency of the tracking system was reduced in a decentralised fashion, but in a way that allowed for extensive self-selection into school types, such as in England (Manning and Pischke 2006).

This chapter is related to a large literature on the effects of various public policies on the intergenerational correlation of outcomes such as income or education (see Black and Devereux (2011) for a survey). In particular, the chapter is related to studies that evaluate the effects of education policy on intergenerational transmission of education. Nybom and Stuhler (2014), employing a DD strategy, find that the Swedish compulsory schooling reform decreased intergenerational transmission of education by one tenth. Waldinger (2007) investigates whether early tracking raises the correlation of family background with performance in international student assessments. He confirms the finding from earlier studies that family background is more important for test scores in countries that track early, however his triple-differences strategy yields no evidence that family background has a larger effect once tracking has actually taken place, casting doubt on any causal effect of tracking on the importance of family background, at least concerning test scores.[2]

More closely related to this chapter, Checchi and Flabbi (2007) find that there is a larger correlation of parental education and track choice in Italy, where there is free track choice everywhere, than in Germany, where free track choice exists only in a minority of states. This is the opposite of what I find, but their results may mask unobserved heterogeneity between the two countries. In contrast, my results rely on within-state variation within one country.

Also related to this chapter is Dustmann, Puhani, and Schönberg (2014), who estimate the effect of attending the high track on wages, using a date of birth discontinuity that gives rise to exogenous variation due to laws regarding minimum age at school entry. Surprisingly, they find zero effect of attending the high track, and they attribute this to the fact that the German school system allows students to switch tracks at later stages. As already mentioned, track switching leads to an important caveat when interpreting my results. Nevertheless, I do find some persistence in high track attendance for those students who were enabled to attend the high track by the introduction of free track choice. These differences in results may be due to the fact that the two exogenous forces—month of birth in their paper versus transition rules here—do not act upon the same groups of students.

The remainder of the chapter is organized as follows. Section 3.2 outlines the German secondary school system, gives an overview of changes to transition rules, and describes the construction of the student-level data set. Section 3.3 presents the results. Section 3.4 discusses interpretation and implications, and concludes.

3.2 Institutional Setting, Data, and Empirical Strategy

3.2.1 The Transition to Secondary School in the German School System

While education policy in Germany is the responsibility of the states rather than of the federal government, the states' ministers of education meet regularly to agree on some common standards

[2]While my focus on track choice precludes a triple-differences strategy, there is arguably less need for this in my context as I rely on within-state variation in policies within one country, whereas Waldinger (2007) relies on cross-country variation.

later to be implemented by the state legislatures. Hence the basic structure of the school system is very similar across states, whereas there is considerable variation in the regulation of more specific issues such as the transition from primary to secondary school. I first describe those features of the school system that are shared by all states and then highlight the differences in transition rules.

Children spend four years in primary school (six years in Berlin) and then transfer into one of three secondary school tracks that I refer to as low, medium, and high (following Dustmann, Puhani, and Schönberg (2014), who provide a more detailed description of the German tracking system).

- The low track (*Hauptschule*) is the most basic track lasting until ninth grade, after which students usually start a paid apprenticeship.
- The medium track (*Realschule*) lasts until tenth grade, is slightly more advanced—in terms of breadth and depth of the academic curriculum—than the low track, but does not enable students to pursue a university education.
- The high track (*Gymnasium*) provides the most advanced secondary education, lasting until twelfth or thirteenth grade, and enabling students to attend university, provided they successfully complete the *Abitur*, a series of school-leaving examinations comparable to the British A-levels.

Some states have been experimenting with comprehensive secondary schools, which avoid early tracking of students. There are no barriers to admission to comprehensive schools. While it is possible to obtain the *Abitur* in comprehensive schools (as well as all other school-leaving certificates), for the purposes of this chapter I will regard children who attend these schools as not attending the high track. It is not clear which school leaving degree parents envisage who send their child to a comprehensive school, and the data do not contain information on later tracking within comprehensive schools.

I assembled information on the rules for admission to secondary school tracks from the relevant decrees issued by the German states, using the archive of the Conference of the Ministers of Education of the German States (*Kultusministerkonferenz*) in Bonn. For a detailed summary of current rules see Kultusministerkonferenz (2010). A short summary of the rules that were in place in 1992 is provided in Avenarius and Jeand'Heur (1992).

Students finish primary school upon graduating 4th grade (6th grade in Berlin). They receive a recommendation by their primary school teacher, indicating what the teacher considers to be the most suitable secondary school track for them. This recommendation is either based on the GPA in 4th grade, or on 'softer' criteria such as perceived academic potential, motivation, or whether a student's results show an upward trend. In the former case, a minimum GPA for enrolment in the high track is specified, usually an average of about 2.0 in German and Math.[3] The recommendation may be binding or not. When it is binding, then the only way that a child may attend the high track in the absence of a positive recommendation is to sit an admissions test or to take test lessons, which

[3]In German schools grades range from 1.0 to 6.0, where 1.0 stands for 'very good' and 4.0 is the lowest passing grade.

usually conclude with a written test as well.[4] When the recommendation is not binding, then parents may decide in which secondary school track to enrol their child. In some cases, deviating from the recommendation by enrolling the child in a higher track than specified in the recommendation may result in further compulsory consultations. Similarly, in states where the recommendation is not binding it may become binding if parents do not participate in any consultations.

Throughout the analysis I employ a binary measure of transition rules, namely a dummy indicating whether parents and their children are free to choose a secondary school track *regardless* of academic performance in primary school.[5] Thus, free track choice is absent when the teacher's recommendation is binding. In such cases it is typically necessary to pass an admissions test or test lessons to attend the high track against the recommendation, thus a minimum standard of academic performance is required. Table 3.1 shows the presence or absence of free track choice for all states and years. More detailed descriptions of the various transition rules are contained in Table 3.B.1.

As can be seen in the table, there are states that have always had free track choice (e.g. Niedersachsen, Hamburg) as well states that never had (e.g. Bayern, Baden-Württemberg).[6] The experiment of switching to a free track choice policy was undertaken in Saarland (1989), Rheinland-Pfalz (1992), Hessen (1993), and Nordrhein-Westfalen (1997). Some states went from free track choice to a more restrictive policy (e.g. Schleswig-Holstein). In sum, there is substantial within-state variation in transition rules.

3.2.2 Data

I use individual-level data from the German Socioeconomic Panel (SOEP), a longitudinal household survey that began in 1984.[7] For each student I identify the year of transition to secondary school as well as the secondary school track attended. The individuals whose transition I observe constitute the baseline sample. I am also interested in the persistence of track attendance. Thus, I record the track a student attends when last observed in the survey, though no earlier than four years after the transition to secondary school. The individuals that I observe four years after their transition or later constitute my longitudinal sample. Due to attrition it is extremely rare to observe both an individual's transition to and completion of secondary school.[8]

I further compile information on mother's and father's education, nationality, and the household's state of residence. The measure of parental education that I use throughout is a dummy indicating whether at least one parent completed the high track.[9] Given a child's year of transition

[4]It is always possible to attend a lower track than what the recommendation specifies, for instance to attend the medium track when the recommendation is for the high track.

[5]I also experimented with alternative measures, such as a numerical index or a set of dummy variables indicating the precise characteristics of transition rules. Results were qualitatively similar to the ones reported here. The drawback of the numerical index is that it is bound to be somewhat arbitrary as well as difficult to interpret, while the more detailed transition rules dummies suffered from a lack of variation along dimensions other than whether the recommendation is binding.

[6]Baden-Wuerttemberg introduced free track choice in 2012, after the end of my sample period.

[7]I use version 27 that covers years 1984-2010. For an overview of the SOEP, see Wagner, Frick, and Schupp (2007).

[8]In the SOEP, information on children is provided by the parents who are interviewed annually, until their children turn 17. If children then leave the parents' household and thus the sample, then no information about their school-leaving degrees is available.

[9]Results are very similar when using college completion, or when considering father's and mother's education

and state of residence, I assign the appropriate value of the free track choice dummy.[10]

All results reported are weighted by SOEP sampling weights adjusted for attrition. The baseline sample requires longitudinal weighting because an individual must be present in the sample for two successive periods for their transition to be observed. I adjust for attrition by multiplying the base weight by the product of the inverse of the staying probabilities supplied in the SOEP.

Table 3.2 shows descriptive statistics. At least one parent completed the high track in about one third of cases. Similarly, about one third of students attend the high track at the start of secondary school. This is consistent with figures from the Federal Statistical Office of Germany, which also show no trend in high track enrolment over the past three decades (Statistisches Bundesamt 2013). The fraction of students on the high track when last observed is slightly larger. Five percent of students exit the high track, while six percent enter it after their initial transition to secondary school. Due to attrition, the longitudinal sample is barely half the size of the baseline sample. Most students are last observed in grade 9 or 10, while completion of the high track usually comes at the end of grade 12 or 13, depending on the state.

3.2.3 Empirical Strategy

The purpose of the analysis is to determine how parental background and transition rules affect a student's probability of attending the high track.[11] Formally, we can write the probability that individual i attends the high track as

$$\text{Pr}_i(\text{high track}) \equiv G\left(x_i, u_i, \Delta_{s(i)}, \Theta_{t(i)}, FREE_{s(i),t(i)}\right) \tag{3.1}$$

where x_i is a vector of observable individual characteristics (including parental background), u_i represents unobservable individual characteristics, $\Delta_{s(i)}$ captures time-invariant factors in individual i's state of residence, and $\Theta_{t(i)}$ denotes shocks in the year of i's transition that are common across states. $FREE_{s(i),t(i)} \in \{0, 1\}$ indicates whether there was free choice of track in the state and at the time of the student's transition. G is the conditional expectation function (CEF) for the outcome variable of interest, the probability of attending the high track. I estimate a linear probability model[12] to approximate the unknown G,

$$\text{Pr}_i(\text{high track}) = x_i'\beta^i + \gamma_i \times FREE_{s(i),t(i)} + \delta_{s(i)} + \theta_{t(i)} + \varepsilon_{i,s(i),t(i)} \tag{3.2}$$

separately.

[10] In most years the SOEP combines the states Rhineland-Palatine and Saarland in one category. This is problematic because these states did not change transition rules simultaneously. I attempt to resolve this by assigning each child coded as living in either of these states a weighted average of the free track choice dummy for Rhineland-palatine and Saarland, with weights corresponding to the relative population size (0.8 and 0.2, respectively). Results are not sensitive to excluding these two states.

[11] I do not consider the choice between the low and medium tracks, since the difference between these two in terms of curriculum and their effect on future earnings is not as large as when comparing either to the demanding curriculum of the high track and the earnings profile of workers who completed the high track.

[12] I also report results from Probit regressions for selected specifications as a robustness check.

where $\delta_{s(i)}$ and $\theta_{t(i)}$ denote state and time fixed effects, respectively, and $\varepsilon_{i,s(i),t(i)}$ is an error term.[13] I am interested in the determinants of both the initial track and the determinants of staying on the high track. Thus, the dependent variable in (3.2) may be $\text{Pr}_i(\text{start on high track})$ or $\text{Pr}_i(\text{stay on high track})$. Besides approximating an arbitrary non-linear CEF (Angrist and Pischke 2009), an advantage of the linear probability model is that coefficients can be readily interpreted as fractions of high track attendance for subgroups, since all right hand side variables are dummies.

The effect of free track choice on attending the high track is denoted by γ_i, where the subscript indicates that the effect may differ across individuals (the coefficients on individual characteristics, β^i, may also be heterogenous). I attempt to capture heterogeneity in the treatment effect by interacting the dummy for free track choice with individual characteristics, as explained below. I cluster standard errors at the state-year level. Clustering at the state level would in principle be preferable, is however problematic given the small number of states (sixteen).

The vector of individual characteristics includes dummies for at least one parent having completed the high track, both parents being foreign nationals, and female.[14] Formally,

$$x_i' \equiv [HIGHTRACK_i,\ FOREIGN_i,\ FEMALE_i]$$

and so

$$\beta^i \equiv \left[\beta^i_{HIGHTRACK}, \beta^i_{FOREIGN}, \beta^i_{FEMALE}\right]'.$$

In the linear probability model, the coefficient $\beta_{HIGHTRACK}$ (if constrained to be constant across individuals) indicates how much higher is the fraction of students attending the high track among those students whose parents completed the high track. Since the variance of attending the high track does not differ much between children and parents, this coefficient approximately equals the intergenerational correlation of high track attendance. The main question of the chapter is how this correlation varies with states' transition rules. I therefore interact the dummy for free track choice with the dummy for at least one parent having completed the high track. The coefficient on the interaction term then indicates the amount by which the intergenerational correlation changes under free track choice. I also investigate whether there are differential effects for females and children of immigrants by adding the appropriate interactions.

The main concern for identification is the possible endogeneity of policy changes. For instance, if a group that would benefit from free track choice becomes more influential politically, then the probability that a state introduces free track choice may increase, and the treatment would not be randomly assigned. It is possible to check whether this may be the case at least with respect to observable variables. I regress the individual characteristics listed above on the dummy for free track choice as well as state and year fixed effects and state-level linear trends. The results, shown in Table 3.3, suggest that individual characteristics are balanced across treatment and control groups. The finding holds regardless of whether state trends are included.

[13] In most specifications I also use linear trends interacted with state dummies. This does not affect the estimates but improves precision.

[14] I do not include family income on the right hand side as it is, arguably, jointly determined with track choice and thus is a 'bad control' in the language of Angrist and Pischke (2009).

A further concern is that states which introduced free track choice may have experienced systematically different trends in the fraction of students attending the high track, violating the common-trend assumption of the differences-in-differences estimator. One way to address this concern is to include state-level trends on the right hand side. I do so throughout. Indeed, when not controlling for state-level trends the effect of the policy seems somewhat lower, suggesting that treatment states would have experienced a decline in high track enrollment in the absence of treatment. I maintain the identification assumption that treatment is exogenous conditional on state trends. A second way to check for differential pre-treatment trends is to include lags and leads of the policy change on the right hand side. As reported below, the results from this exercise do not suggest any differential pre-treatment trends. The pattern of lags and leads instead suggests that trends in treatment and control states may have started to diverge long after the policy change.

Note that I do not identify the causal effect of parental characteristics, in particular education, on track choice, since I lack exogenous variation in these variables. Thus, parental education here captures not just the 'nurture effect' due to more resources being available to more educated parents as well as differences in behaviour, but also the 'nature effect' stemming from differences in innate abilities between differently educated parents. Holmlund, Lindahl, and Plug (2011) establish that only part of the intergenerational association is causal. However, the *change* in the intergenerational association due to changes in transition rules has a causal interpretation if these policy changes are exogenous.

3.3 Results

3.3.1 Effects of Free Track Choice on Attending the High Track at the Start of Secondary School

Table 3.4 shows the main results. The first column constrains all coefficients to be constant across individuals. A free track choice policy leads to an increase in the fraction of students attending the high track of five percentage points (pp), although the effect is imprecisely estimated. Children whose parents completed the high track have a 40pp higher probability of attending the high track (the unconditional correlation between parents' and children's high track attendance is 40 percent, as expected given similar variances). Children of immigrants are less likely, while females are more likely, to attend the high track.

Column (2) introduces the interaction of free track choice with the indicator for parents having completed the high track. The coefficient on the free track choice dummy now gives the effect of the treatment on children whose parents did not complete the high track. For these students, free track choice raises the probability of attending the high track by 11pp, and this effect is precisely estimated. Children whose parents completed the high track and who do not enjoy free track choice have a 46pp higher probability of attending the high track. However, the coefficient is reduced by 17pp, more than one third, under free track choice. This is the most important result of the chapter: free track choice appears to dramatically reduce the correlation between parental education and high track attendance.

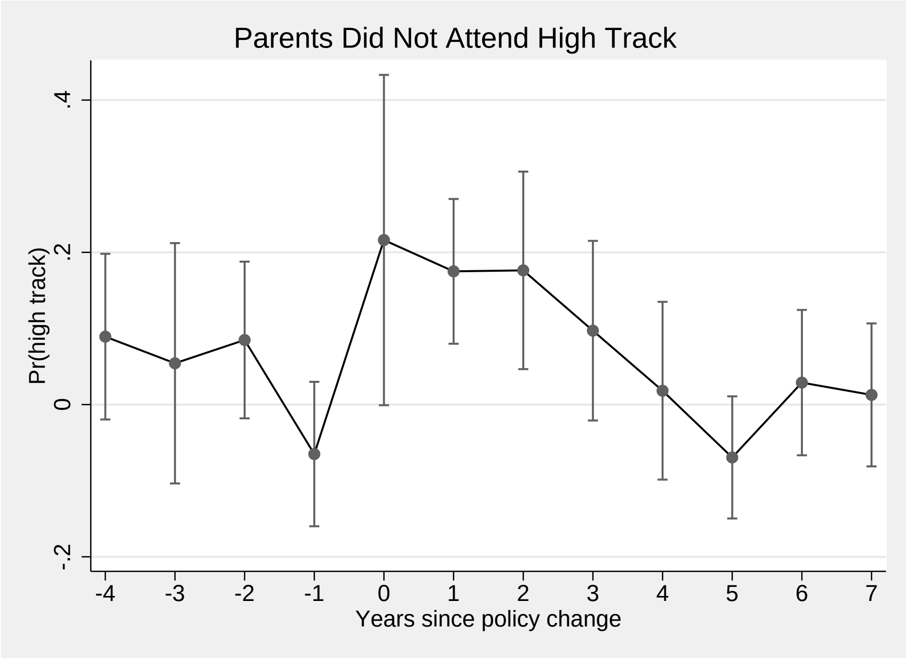
Column (3) and (4) introduce interactions of free track choice with the immigrant background and female indicators. Both groups appear to benefit more from the policy, although the effects are imprecisely estimated. Column (5) includes all three interactions simultaneously, yielding similar results. Column (6) allows for a full set of state-year fixed effects. Under this specification it is no longer possible to identify the level effect of free track choice, however the interactions can still be estimated. Again, results are similar.

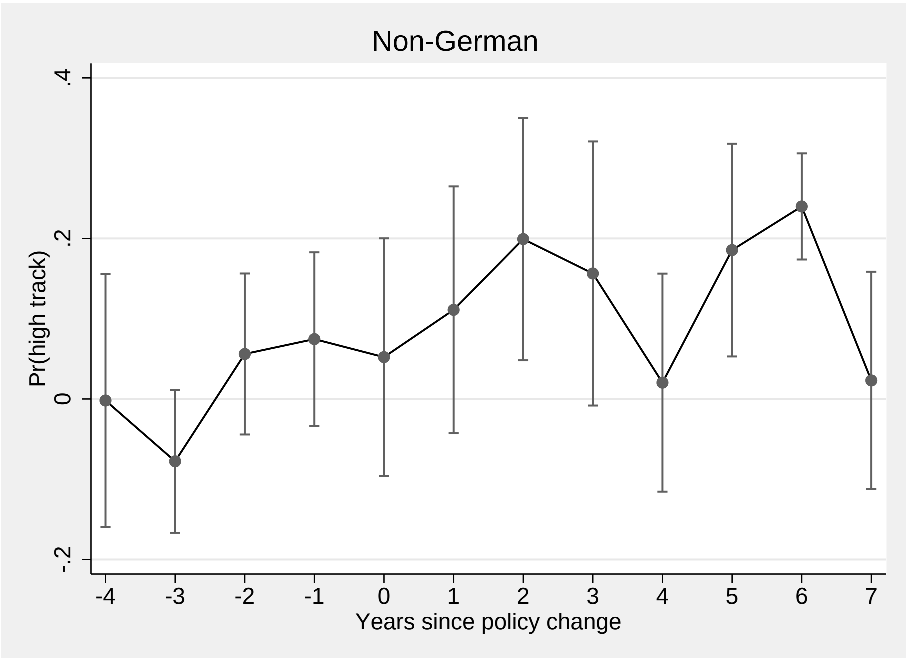
The reduction of the effect of parental high track completion by more than one third under free track choice is a finding that is highly robust across specifications. How to interpret this finding? Children of highly educated parents may have higher innate ability and these families likely have more resources at their disposal to prepare their children for the transition to secondary school. Thus, any constraints imposed by the absence of free track choice, such as GPA requirements, may not be binding for these children. If this interpretation is correct, the effect of free track choice on this group should be zero. This effect is estimated by the sum of the coefficients displayed in the first and third rows of the table. In columns (2) and (5) the effect is thus estimated to be negative. Under the null hypothesis of a zero effect such an estimate would occur in more than one out of ten cases (see the p-value at the bottom of the table), thus the evidence against the effect being zero is not very strong.

The results presented so far suggest that free track choice raises the probability of attending the high track for three subgroups, namely children whose parents did not complete the high track, children of immigrants, and females. If these estimates indeed represent causal effects, we should observe a change in enrolment rates only after a change in transition rules has taken place. To see whether this is the case, I estimate models that include four leads and seven lags of the policy change on the right hand side. I restrict the treatment group to the states Hessen and Nordrhein-Westfalen (keeping all other states that never changed transition rules as controls) and years before 2007. I thus focus on the introduction of free track choice in 1993 and 1997, respectively. Including policy changes in other states would not allow for as many lags and leads. Columns (1) and (2) of Table 3.5 show that the main results from the full sample carry over to the restricted sample. Given that Hessen and Nordrhein-Westfalen are by far the two largest states that have changed transition rules in my sample period, this should not be surprising. To avoid a large number of interaction terms, I estimate the specification including lags and leads on subgroups only. Columns (3), (4), and (5) show the plain differences-in-differences results for the three subgroups. Since the effect for females is statistically indistinguishable from zero, I show estimated lags and leads only for children whose parents did not complete the high track as well as children of immigrants.

The patterns of estimated leads and lags for the two subgroups are shown in Figure 3.1. In both cases the increase in high track attendance occurs after the introduction of free track choice, consistent with a causal interpretation. However, children of less educated parents do not appear to benefit from free track choice when their transition takes place four or more years after the policy change.

Figure 3.1: Effects of Free Track Choice on Attending the High Track Relative to Year of Introducing Free Track Choice





Note: Vertical bars mark 95 percent confidence intervals. Results are based on the introduction of free track choice in Hessen (1993) and Nordrhein-Westfalen (1997).

3.3.2 The Effects of Free Track Choice on Staying on the High Track

While children whose parents have low educational attainment are more likely to attend the high track under free track choice, they may find the curriculum too challenging and may be at risk of switching to a lower track later on. This may be because they have lower innate ability, or because of a lack of parental resources. Thus, the question is how free track choice affects the probability of track completion. Attrition prevents me from following individuals long enough to observe track completion. Instead I investigate the effect of free track choice on attending the high track when last observed in the sample, though no earlier than four years after the transition. As mentioned above, most students in this longitudinal sample are last observed when attending grades 9 or 10 (Table 3.2).

Columns (1) to (3) of Table 3.6 show that the findings regarding initial track attendance reported in Table 3.4 also hold in the longitudinal sample. Column (4) suggests that on average, free track attendance has no effect on staying on the high track four years or later after the transition. However, the estimate is imprecise and statistically indistinguishable from the coefficient on free track choice reported in column (1). Column (5) adds the interaction of free track choice with parents having completed the high track, and column (6) includes the remaining interactions. Column (7) reports results from the same specification as column (6), except that weights are multiplied by the square of the grade the individual attended when last observed. Thus, this specification gives more weight to individuals that are closer to completion.

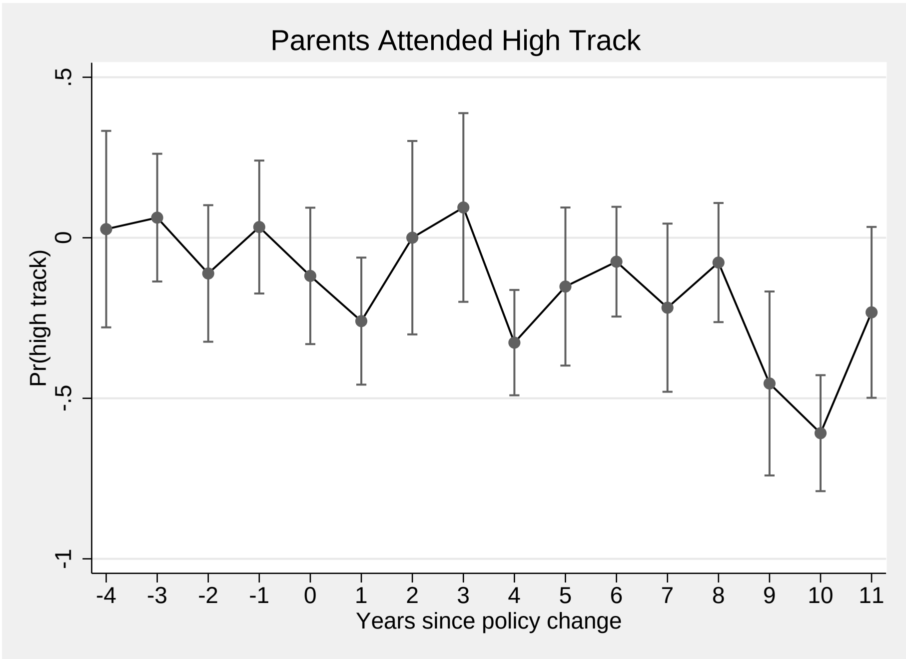
Results from columns (4) to (7) are quite similar. In all cases, free track choice dramatically reduces the effect of parents having completed the high track on staying on the high track. However, for children whose parents did not complete the high track, the effect of free track choice on staying on the high track is only 6pp and imprecisely estimated, compared to 11pp for starting on the high track. This suggests that some students who are initially helped by the policy cannot cope with the demands of the high track and later switch to a lower track. However, caution is warranted in interpreting this evidence, given that the estimates are imprecise.

3.3.3 Additional Results and Robustness Checks

I investigate whether my results change when fitting a non-linear model. Table 3.A.1 reports results for selected specifications from Probit regressions. They are qualitatively very similar to those from the linear probability models.

I also check whether the results are sensitive to not controlling for state trends. Table 3.A.2 shows that the results are quite similar, but coefficients are somewhat lower. In particular, the effect on students with highly educated parents is now negative and statistically significant, which is unexpected. However, plotting the coefficients on leads and lags for this group, using the same sample as above, yields no clear evidence for a negative effect of the policy on impact. Figure 3.2 shows instead a downward trend, largely driven by years long after the treatment. Overall, the evidence suggests that the assumption of exogenous treatment conditional on state trends is reasonable.

Figure 3.2: Effects of Free Track Choice on Attending the High Track Relative to Year of Introducing Free Track Choice, Highly Educated Parents



Note: Vertical bars mark 95 percent confidence intervals. Results are based on the introduction of free track choice in Hessen (1993) and Nordrhein-Westfalen (1997).

3.4 Discussion and Conclusion

In this chapter I have studied the effects of free choice of secondary school track on enrolment in the university-preparatory high track, as well as on the association of parental education with track choice, employing a DD strategy that takes advantage of within-state variation in transition rules. I find that free track choice increases enrolment in the high track by only five percentage points on average, but by twice this amount for children of less educated parents, thus reducing the association of parental education with track choice by more than one third. When looking at track enrolment four years or later after leaving primary school, I find qualitatively similar results, although the magnitudes are reduced. The diminished correlation between parents' and children's education does not reflect a zero-sum effect: introducing free track choice in all states would boost educational attainment of the German population.

My findings confirm that parental background predicts cost differentials in children's education. Likely due to a combination of lower ability and less available resources to support learning outside school, children of less educated parents face higher costs of achieving a given level of academic performance. These children are therefore less likely to attend the high track if academic performance is critical for admission. This interpretation is consistent with the finding by Schnabel and Schnabel (2002) that parental education is inversely related to the returns to schooling later realized by the child. The positive association of costs and benefits is exactly what the standard human capital accumulation model (Card 1999) predicts.[15]

Recent research finds that intergenerational mobility has been constant over the past few decades in the US (Chetty, Hendren, Kline, Saez, and Turner 2014), and that it seems remarkably constant across a very diverse range of countries as well as over many centuries within countries (Clark 2014). For Germany, Heineck and Riphahn (2007) find no downward trend in intergenerational transmission of education for cohorts born between 1929 and 1978 despite dramatic policy changes such as the abolishment of school fees or the introduction of a university scholarship program. In contrast, the results in this chapter suggest that policy changes that are minor in historical perspective, may have a substantial effect on intergenerational mobility. To increase confidence in this conclusion, it is important to study whether free track choice affects completion of the high track in a similar way as attendance at an early stage of secondary school. Answering this question requires more data which will only become available as more time has passed, hence this is left for future research.

[15] Admittedly, my analysis focusses rather crudely on only two groups, children who have at least one parent who completed the high track, and those who do not. There is likely to be heterogeneity within these groups. Among the students with less educated parents there may be many who do not find it optimal to attend the high track even after the introduction of free track choice.

Tables

Table 3.1: Free Access to the High Track

	85-88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
Baden-Württemberg	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Bayern	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Berlin	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
Brandenburg				N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Bremen	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hamburg	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hessen	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mecklenburg-Vorpommern				Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Niedersachsen	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Nordrhein-Westfalen	N	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
Rheinland-Pfalz	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Saarland	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	Y
Sachsen				N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Sachsen-Anhalt				N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Schleswig-Holstein	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
Thüringen				N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Y and N indicate presence and absence of free track choice, respectively. Free track choice implies that the choice of secondary school track is up to the parents, regardless of academic performance in primary school.

Table 3.2: Descriptive Statistics

	Baseline sample		Longitudinal sample	
	mean	sd	mean	sd
Parents high track	0.31	0.46	0.31	0.46
Non-German	0.10	0.30	0.08	0.27
Female	0.47	0.50	0.48	0.50
Free track choice	0.39	0.48	0.37	0.47
High track at start	0.32	0.47	0.34	0.47
Remained on high track			0.29	0.45
Transfered later to high track			0.06	0.24
Grade when last observed			9.61	0.70
Observations	4871		2523	

The baseline sample includes all individuals whose transition from primary to secondary school is observed. The longitudinal sample includes all individuals that are observed four years after the transition or later. All statistics are weighted by SOEP longitudinal weights that adjust for attrition. The baseline sample requires adjustment for attrition because individuals must be present in the sample for two successive periods for their transition to be observed.

Table 3.3: Free Track Choice and Individual Characteristics

	Pr(parents high track)		Pr(non-German)		Pr(female)	
	(1)	(2)	(3)	(4)	(5)	(6)
Free track choice	-0.015	-0.045	-0.008	-0.008	-0.009	0.004
	(0.031)	(0.035)	(0.019)	(0.025)	(0.030)	(0.036)
State trends	No	Yes	No	Yes	No	Yes
Adjusted R^2	0.027	0.029	0.025	0.025	0.009	0.008

Dependent variables are an indicator for at least one parent having attended the high track, an indicator for both parents being foreign nationals, and an indicator for female, respectively. Results from linear probability models are shown. The number of observations is 4,871. All regressions include state and year fixed effects. Regressions are weighted using SOEP longitudinal weights. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

Table 3.4: Effects of Free Track Choice on Attending the High Track

	(1)	(2)	(3)	(4)	(5)	(6)
Free track choice	0.054	0.107**	0.044	0.047	0.092*	
	(0.034)	(0.036)	(0.035)	(0.038)	(0.042)	
Parents high track	0.391***	0.459***	0.391***	0.391***	0.458***	0.463***
	(0.021)	(0.027)	(0.021)	(0.021)	(0.027)	(0.029)
Parents high × Free		-0.172***			-0.170***	-0.167***
		(0.038)			(0.039)	(0.042)
Non-German	-0.107***	-0.106***	-0.137***	-0.108***	-0.119***	-0.108***
	(0.020)	(0.020)	(0.024)	(0.020)	(0.025)	(0.027)
Non-German × Free			0.075		0.031	0.031
			(0.040)		(0.043)	(0.049)
Female	0.071***	0.073***	0.070***	0.065**	0.064**	0.072***
	(0.016)	(0.015)	(0.015)	(0.020)	(0.020)	(0.020)
Female × Free				0.015	0.023	0.012
				(0.032)	(0.031)	(0.034)
p-value: Free + Parents high × Free		0.119			0.086	
p-value: Free + Non-German × Free			0.008		0.009	
p-value: Free + Female × Free				0.089	0.004	
Adjusted R^2	0.20	0.20	0.20	0.20	0.20	0.24

The dependent variable is an indicator for attending the high track in the first year of secondary school. Results from linear probability models are shown. The number of observations is 4,871. All regressions include state and year fixed effects as well as state-level linear trends, apart from the last column which includes state-by-year fixed effects. Regressions are weighted using SOEP longitudinal weights. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

Table 3.5: Effects of Free Track Choice on Attending the High Track in Federal States Hessen and Nordrhein-Westfalen

	All		Parents no high	Non-German	Female
	(1)	(2)	(3)	(4)	(5)
Free track choice	0.021	0.073	0.091*	0.079*	0.040
	(0.039)	(0.048)	(0.040)	(0.038)	(0.042)
Parents high track	0.407***	0.470***		0.353**	0.477***
	(0.024)	(0.029)		(0.105)	(0.030)
Parents high × Free		-0.186***			
		(0.045)			
Non-German	-0.134***	-0.152***	-0.119***		-0.131***
	(0.019)	(0.022)	(0.019)		(0.032)
Non-German × Free		0.050			
		(0.044)			
Female	0.067***	0.068**	0.026	0.056	
	(0.018)	(0.022)	(0.021)	(0.029)	
Female × Free		-0.000			
		(0.035)			
Observations	3655	3655	2493	688	1743
Adjusted R^2	0.21	0.22	0.04	0.13	0.26

The dependent variable is an indicator for attending the high track in the first year of secondary school. Results from linear probability models are shown. All regressions include state and year fixed effects. Regressions are weighted using SOEP longitudinal weights. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

Table 3.6: Effects of Free Track Choice on Staying on the High Track

	Pr(start on high)			Pr(stay on high)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Free track choice	0.068	0.132*	0.119	0.001	0.059	0.062	0.066
	(0.058)	(0.058)	(0.061)	(0.058)	(0.058)	(0.058)	(0.059)
Parents high track	0.378***	0.457***	0.455***	0.374***	0.446***	0.443***	0.434***
	(0.028)	(0.034)	(0.034)	(0.030)	(0.037)	(0.037)	(0.038)
Parents high × Free		-0.218***	-0.214***		-0.200***	-0.191**	-0.187**
		(0.055)	(0.057)		(0.057)	(0.059)	(0.058)
Non-German	-0.145***	-0.141***	-0.161***	-0.126***	-0.123***	-0.155***	-0.157***
	(0.029)	(0.029)	(0.034)	(0.029)	(0.029)	(0.036)	(0.034)
Non-German × Free			0.049			0.084	0.082
			(0.061)			(0.059)	(0.058)
Female	0.083***	0.084***	0.077**	0.071**	0.072***	0.080**	0.080**
	(0.021)	(0.021)	(0.026)	(0.022)	(0.021)	(0.027)	(0.026)
Female × Free			0.018			-0.026	-0.023
			(0.042)			(0.045)	(0.046)
p-value: Free + Parents high × Free		0.233	0.194		0.044	0.066	0.089
p-value: Free + Non-German × Free			0.027			0.056	0.058
p-value: Free + Female × Free			0.039			0.609	0.543
Adjusted R^2	0.21	0.22	0.22	0.20	0.21	0.21	0.21

Results from linear probability models are shown. The number of observations is 2,523. All regressions include state and year fixed effects, as well as state-level linear trends. Regressions are weighted using SOEP longitudinal weights. For the last column, weights were multiplied by the square of the grade the individual attended when last observed in the sample. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

Appendix

3.A Additional Results

Table 3.A.1: Results from Probit Models

	Pr(start on high)				Pr(stay on high)	
	(1)	(2)	(3)	(4)	(5)	(6)
Free track choice	0.054	0.089	0.075	0.113	0.001	0.049
	(0.037)	(0.047)	(0.066)	(0.071)	(0.059)	(0.064)
Parents high track	0.405***	0.464***	0.403***	0.463***	0.376***	0.425***
	(0.021)	(0.028)	(0.030)	(0.035)	(0.031)	(0.039)
Parents high × Free		-0.151***		-0.179**		-0.121*
		(0.040)		(0.062)		(0.055)
Non-German	-0.134***	-0.151***	-0.169***	-0.195***	-0.137***	-0.169***
	(0.024)	(0.031)	(0.032)	(0.038)	(0.030)	(0.035)
Non-German × Free		0.055		0.096		0.133
		(0.063)		(0.095)		(0.095)
Female	0.082***	0.074**	0.095***	0.089**	0.075***	0.086**
	(0.017)	(0.023)	(0.024)	(0.030)	(0.022)	(0.028)
Female × Free		0.024		0.021		-0.026
		(0.035)		(0.048)		(0.047)
p-value: Free + Parents high × Free		0.175		0.404		0.276
p-value: Free + Non-German × Free		0.028		0.050		0.084
p-value: Free + Female × Free		0.008		0.065		0.754
Longitudinal sample	No	No	Yes	Yes	Yes	Yes
Observations	4871	4871	2523	2523	2523	2523
Pseudo R^2	0.17	0.18	0.19	0.19	0.20	0.20

Marginal effects from probit models are shown. All regressions include state and year fixed effects, as well as state-level linear trends. Regressions are weighted using SOEP longitudinal weights. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

Table 3.A.2: Main Results without Controlling for State Trends

	Pr(start on high)			Pr(stay on high)		
	(1)	(2)	(3)	(4)	(5)	(6)
Free track choice	0.028	0.063	0.041	0.089	-0.040	0.018
	(0.031)	(0.039)	(0.042)	(0.048)	(0.039)	(0.044)
Parents high track	0.389***	0.450***	0.377***	0.450***	0.373***	0.438***
	(0.021)	(0.027)	(0.028)	(0.034)	(0.029)	(0.037)
Parents high × Free		-0.156***		-0.201***		-0.180**
		(0.040)		(0.056)		(0.058)
Non-German	-0.110***	-0.120***	-0.144***	-0.161***	-0.126***	-0.156***
	(0.020)	(0.025)	(0.029)	(0.033)	(0.028)	(0.035)
Non-German × Free		0.030		0.051		0.084
		(0.043)		(0.061)		(0.058)
Female	0.070***	0.063**	0.082***	0.074**	0.070**	0.078**
	(0.016)	(0.020)	(0.021)	(0.026)	(0.021)	(0.027)
Female × Free		0.023		0.022		-0.022
		(0.032)		(0.042)		(0.045)
p-value: Free + Parents high × Free		0.033		0.070		0.005
p-value: Free + Non-German × Free		0.036		0.023		0.096
p-value: Free + Female × Free		0.031		0.040		0.940
Longitudinal sample	No	No	Yes	Yes	Yes	Yes
Observations	4871	4871	2523	2523	2523	2523
R^2	0.19	0.20	0.21	0.22	0.20	0.21

Results from linear probability models are shown. All regressions include state and year fixed effects. Regressions are weighted using SOEP longitudinal weights. Robust standard errors, clustered by state-year cells, are shown in parentheses. ***, **, * significant at the 1, 5 and 10 percent level.

3.B Detailed Summary of Transition Rules

Table 3.B.1 contains descriptions of transition rules for each state and year, based on the relevant decrees issued by the German states that I accessed when visiting the archive of the Conference of the Ministers of Education of the German States (*Kultusministerkonferenz*) in Bonn. The list of decrees is available upon request.

Table 3.B.1: Description of Transition Rules

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
Baden-Württemberg	1983-2011	4 years	The year-four GPA determines which track the child is recommended to attend. In the absence of a positive recommendation, attendance of the advanced track requires passing an entrance exam.	No
	since 2012	4 years	After consulting the teachers' views, the choice of track is up to the parents.	Yes
Bayern	1983-2012	4 years	The year-four GPA determines which track the child is recommended to attend. In the absence of a positive recommendation, attendance of the advanced track requires successful participation in test lessons.	No
Berlin	1983-2004	6 years	A recommendation of track is made but does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
	2005-2010	6 years	The recommendation of track depends on an explicit GPA requirement. If parents disagree with the recommendation, further consultations follow, although the choice of track is ultimately up to the parents.	No

continued on next page

Table 3.B.1 – continued from previous page

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
	since 2011	6 years	The recommendation of track depends on an explicit GPA requirement. The choice of track is up to the parents.	Yes
Brandenburg	since 1991	6 years	Both the teachers recommendation and the parents wishes determine which track the child attends. Access may be more difficult when demand for places is high.	No
Bremen	1983-2003	6 years	Choice of track is up to the parents.	Yes
	since 2004	4 years	Choice of track is up to the parents provided they attend consultations. Otherwise, the teachers choose a track for the child based on GPA, among other things.	Yes
Hamburg	1983-1997	4 years	Teachers may recommend attending the advanced track only if the child's GPA is 3.0 or better. However, the choice of track is up to the parents.	Yes
	since 1998	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes

continued on next page

Table 3.B.1 – continued from previous page

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
Hessen	1983-1993	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. In the absence of a positive recommendation, attendance of the advanced track requires successful participation in test lessons.	No
	since 1994	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. If parents disagree, further consultations follow, although the choice of track is ultimately up to the parents.	Yes
Mecklenburg-Vorpommern	since 1991	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
Niedersachsen	since 1983	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
Nordrhein-Westfalen	1983-1996	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. In the absence of a positive recommendation, attendance of the advanced track requires successful participation in test lessons.	No

continued on next page

Table 3.B.1 – continued from previous page

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
	1997-2005	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
	2006	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. If parents disagree, further consultations follow, although the choice of track is ultimately up to the parents.	Yes
	2007-2012	4 years	Same as in 1983-1996	No
	since 2012	4 years	Same as in 1997-2005	Yes
Rheinland-Pfalz	1983-1991	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. In the absence of a positive recommendation, attendance of the advanced track requires successful participation in test lessons.	No
	since 1992	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes

continued on next page

Table 3.B.1 – continued from previous page

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
Saarland	1983-1988	4 years	Teachers make a formal recommendation which depends on an explicit GPA requirement, among other things. In the absence of a positive recommendation, attendance of the advanced track requires successful participation in test lessons.	No
	1989-1999		Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
	2000-2009		Same as in 1983-1988	No
	since 2010		Same as in 1989-1999	Yes
Sachsen	since 1991	4 years	The year-four GPA determines which track the child is recommended to attend. In the absence of a positive recommendation, attendance of the advanced track requires passing an entrance exam.	No

continued on next page

Table 3.B.1 – continued from previous page

State	*Period*	*Years in Primary school*	*Description of transition rules*	*Free choice of track*
Sachsen-Anhalt	since 1991	4 years	The year-four GPA determines which track the child is recommended to attend. In the absence of a positive recommendation, attendance of the advanced track requires passing an entrance exam.	No
Schleswig-Holstein	1983-2002	4 years	Teachers make a formal recommendation which does not depend on an explicit GPA requirement. The choice of track is up to the parents.	Yes
	since 2003		Teachers make a formal recommendation which does not depend on an explicit GPA requirement. If parents disagree, further consultations follow, and a new recommendation is made, which is binding.	No
Thüringen	since 1991	4 years	The year-four GPA determines which track the child is recommended to attend. In the absence of a positive recommendation, attendance of the advanced track requires passing an entrance exam.	No

Bibliography

ACEMOGLU, D. (2002): "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 40(1), 7–72.

ACEMOGLU, D., AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4, chap. 12, pp. 1043–1171. Elsevier.

AIGNER, D. J., AND G. G. CAIN (1977): "Statistical theories of discrimination in labor markets," *Industrial and Labor Relations Review*, 30(2), 175–187.

ALMOND, D., AND J. J. DOYLE (2011): "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays," *American Economic Journal: Economic Policy*, 3(3), 1–34.

ALMOND, D., J. JOSEPH J. DOYLE, A. E. KOWALSKI, AND H. WILLIAMS (2010): "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns," *The Quarterly Journal of Economics*, 125(2), 591–634.

ALTONJI, J. G., AND C. R. PIERRET (2001): "Employer Learning And Statistical Discrimination," *The Quarterly Journal of Economics*, 116(1), 313–350.

ANDERSON, M., AND J. MAGRUDER (2012): "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, 122(563), 957–989.

ANGRIST, J., G. W. IMBENS, AND D. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91.

ANGRIST, J., AND S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

ARCIDIACONO, P., P. BAYER, AND A. HIZMO (2010): "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2(4), 76–104.

ATACK, J., F. BATEMAN, AND T. WEISS (1980): "The Regional Diffusion and Adoption of the Steam Engine in American Manufacturing," *The Journal of Economic History*, 40(02), 281–308.

AUTOR, D. H. (2013): "The task to labor markets: an overview," *Journal for Labour Market Research*, 46(3), 185–199.

AUTOR, D. H., AND D. DORN (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103(5), 1553–97.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2006): "The Polarization of the U.S. Labor Market," *American Economic Review*, 96(2), 189–194.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The Skill Content Of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), 1279–1333.

AVENARIUS, H., AND B. JEAND'HEUR (1992): *Elternwille und staatliches Bestimmungsrecht bei der Wahl der Schullaufbahn: Die gesetzlichen Grundlagen und Grenzen der Ausgestaltung von Aufnahme- bzw. Übergangsverfahren für den Besuch weiterführender Schulen*. Duncker & Humblot.

BARRECA, A. I., M. GULDI, J. M. LINDO, AND G. R. WADDELL (2011): "Saving Babies? Revisiting the effect of very low birth weight classification," *The Quarterly Journal of Economics*, 126(4), 2117–2123.

BELMAN, D., AND J. S. HEYWOOD (1991): "Sheepskin Effects in the Returns to Education: An Examination on Women and Minorities," *The Review of Economics and Statistics*, 73(4), 720–24.

BESSEN, J. (2011): "Was Mechanization De-Skilling? The Origins of Task-Biased Technical Change," Working Papers 1101, Research on Innovation.

BLACK, S. E., AND P. J. DEVEREUX (2011): "Recent Developments in Intergenerational Mobility," vol. 4 of *Handbook of Labor Economics*, chap. 16, pp. 1487–1541. Elsevier.

BOEHM, M. J. (2013): "The Wage Effects of Job Polarization: Evidence from the Allocation of Talents," mimeo, LSE.

BOFF, R. B. D. (1967): "The Introduction of Electric Power in American Manufacturing," *The Economic History Review*, 20(3), pp. 509–518.

BRYNJOLFSSON, E., AND A. MCAFEE (2011): *Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.

BUSSE, M. R., N. LACETERA, D. G. POPE, J. SILVA-RISSO, AND J. R. SYDNOR (2013): "Estimating the Effect of Salience in Wholesale and Retail Car Markets," *American Economic Review*, 103(3), 575–79.

CARD, D. (1999): "Chapter 30 The causal effect of education on earnings," vol. 3, Part A of *Handbook of Labor Economics*, pp. 1801 – 1863. Elsevier.

CARPENTER, C., AND C. DOBKIN (2009): "The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age," *American Economic Journal: Applied Economics*, 1(1), 164–82.

CASELLI, F. (1999): "Technological Revolutions," *American Economic Review*, 89(1), 78–102.

CHARI, V. V., AND H. HOPENHAYN (1991): "Vintage Human Capital, Growth, and the Diffusion of New Technology," *Journal of Political Economy*, 99(6), 1142–65.

CHECCHI, D., AND L. FLABBI (2007): "Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks," IZA Discussion Papers 2876, Institute for the Study of Labor (IZA).

CHETTY, R., N. HENDREN, P. KLINE, E. SAEZ, AND N. TURNER (2014): "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," NBER Working Papers 19844, National Bureau of Economic Research, Inc.

CLARK, G. (2014): *The Son Also Rises*. Princeton University Press.

CORTES, G. M. (2012): "Where Have the Routine Workers Gone? A Study of Polarization Using Panel Data," The School of Economics Discussion Paper Series 1224, Economics, The University of Manchester.

COSTINOT, A. (2009): "On the origins of comparative advantage," *Journal of International Economics*, 77(2), 255–264.

COSTINOT, A., AND J. VOGEL (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 747–786.

DAVIDSON, A. (2012): "Making It in America," *The Atlantic*, Jan/Feb.

DI PIETRO, G. (2010): "The Impact of Degree Class on the First Destinations of Graduates: A Regression Discontinuity Approach," IZA Discussion Papers 4836, Institute for the Study of Labor (IZA).

DICKENS, R., A. MANNING, AND T. BUTCHER (2012): "Minimum Wages and Wage Inequality: Some Theory and an Application to the UK," Working Paper Series 4512, Department of Economics, University of Sussex.

DOLLMANN, J. (2011): "Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang," *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 63(4), 595–621.

DORN, D. (2009): "Essays on Inequality, Spatial Interaction, and the Demand for Skills," Dissertation no. 3613, University of St. Gallen.

DORNBUSCH, R., S. FISCHER, AND P. A. SAMUELSON (1977): "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *American Economic Review*, 67(5), 823–39.

DUSTMANN, C. (2004): "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 56.

DUSTMANN, C., P. A. PUHANI, AND U. SCHÖNBERG (2014): "The Long-Term Effects of Early Track Choice," IZA Discussion Papers 7897, Institute for the Study of Labor (IZA).

ELSBY, M. W. L., B. HOBIJN, AND A. SAHIN (2013): "The Decline of the U.S. Labor Share," Fall 2013 Brookings Panel on Economic Activity.

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

FARBER, H. S., AND R. GIBBONS (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111(4), 1007–47.

FIRPO, S., N. M. FORTIN, AND T. LEMIEUX (2011): "Occupational Tasks and Changes in the Wage Structure," IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108(5), 874–904.

GARICANO, L., AND E. ROSSI-HANSBERG (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121(4), 1383–1435.

GOLDIN, C., AND L. F. KATZ (1998): "The Origins Of Technology-Skill Complementarity," *The Quarterly Journal of Economics*, 113(3), 693–732.

——— (2008): *The Race between Education and Technology*. Harvard University Press.

GOOS, M., AND A. MANNING (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 89(1), 118–133.

GOOS, M., A. MANNING, AND A. SALOMONS (2009): "Job Polarization in Europe," *American Economic Review*, 99(2), 58–63.

——— (2014): "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, forthcoming.

GRAY, R. (2013): "Taking technology to task: The skill content of technological change in early twentieth century United States," *Explorations in Economic History*, 50(3), 351–367.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–09.

HEINECK, G., AND R. T. RIPHAHN (2007): "Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades," Discussion paper.

HOLLAND, P. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81.

HOLMES, T. J., AND M. F. MITCHELL (2008): "A theory of factor allocation and plant size," *RAND Journal of Economics*, 39(2), 329–351.

HOLMLUND, H., M. LINDAHL, AND E. PLUG (2011): "The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods," *Journal of Economic Literature*, 49(3), 615–51.

HUNGERFORD, T., AND G. SOLON (1987): "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 69(1), 175–77.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–75.

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

IMBENS, G. W., AND J. M. WOOLDRIDGE (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

IRELAND, N., R. A. NAYLOR, J. SMITH, AND S. TELHAJ (2009): "Educational Returns, Ability Composition and Cohort Effects: Theory and Evidence for Cohorts of Early-Career UK Graduates," CEP Discussion Papers dp0939, Centre for Economic Performance, LSE.

JAEGER, D. A., AND M. E. PAGE (1996): "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 78(4), 733–40.

JOVANOVIC, B. (1998): "Vintage Capital and Inequality," *Review of Economic Dynamics*, 1(2), 497–530.

KARABARBOUNIS, L., AND B. NEIMAN (2014): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129(1), 61–103.

KATZ, L. F., AND R. A. MARGO (2013): "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective," NBER Working Papers 18752, National Bureau of Economic Research, Inc.

KULTUSMINISTERKONFERENZ (2010): "Übergang von der Grundschule in Schulen des Sekundarbereichs I," Discussion paper.

LANGE, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

LEE, D. S., AND D. CARD (2008): "Regression discontinuity inference with specification error," *Journal of Econometrics*, 142(2), 655–674.

LEE, D. S., AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

LEMIEUX, T. (2006a): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, 96(3), 461–498.

——— (2006b): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, 96(3), 461–498.

LEVY, F., AND R. J. MURNANE (2004): *The New Division of Labor: How Computers Are Creating the Next Job Market*. Princeton University Press.

LI, Q., AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

LINDLEY, J., AND S. MACHIN (2011): "Rising Wage Inequality and Postgraduate Education," IZA Discussion Papers 5981, Institute for the Study of Labor (IZA).

MAMELI, M., AND P. BATESON (2011): "An evaluation of the concept of innateness," *Phil. Trans. R. Soc. B*, 366, pp. 436–443.

MANNING, A., AND S. PISCHKE (2006): "Comprehensive versus Selective Schooling in England in Wales: What Do We Know?," *National Bureau of Economic Research Working Paper Series*, No. 12176.

MARKOFF, J. (2012): "Skilled Work, Without the Worker," *The New York Times*, Aug 18.

MCCRARY, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142(2), 698–714.

MCKNIGHT, A., R. NAYLOR, AND J. SMITH (2007): "Sheer Class? Returns to educational performance : evidence from UK graduates first destination labour market outcomes," Discussion paper.

MELENDEZ, E. D. (2013): "Minimum Wage Jobs Won't Be Replaced With This Robot Anytime Soon," *The Huffington Post*, Aug 30.

MICHAELS, G., A. NATRAJ, AND J. VAN REENEN (forthcoming): "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 Years," *Review of Economics and Statistics*, (16138).

MORAVEC, H. (1988): *Mind Children*. Harvard University Press.

NATIONAL ACADEMY OF SCIENCES (1981): "Dictionary of Occupational Titles (DOT): Part I - Current Population Survey, April 1971, Augmented With DOT Characteristics and Dictionary of Occupational Titles (DOT): Part II - Fourth Edition Dictionary of DOT Scores for 1970 Census Categories," .

NORDHAUS, W. D. (2007): "Two Centuries of Productivity Growth in Computing," *The Journal of Economic History*, 67(01), 128–159.

NOURBAKHSH, I. R. (2013): *Robot Futures*. MIT Press.

NYBOM, M., AND J. STUHLER (2014): "Interpreting Trends in Intergenerational Mobility," unpublished manuscript.

OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4(1), 1–29.

OYER, P. (2006): "Initial Labor Market Conditions and Long-Term Outcomes for Economists," *Journal of Economic Perspectives*, 20(3), 143–160.

——— (2008): "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *Journal of Finance*, 63(6), 2601–2628.

RODRIGUEZ, F., AND A. JAYADEV (2010): "The Declining Labor Share of Income," Human Development Research Papers (2009 to present) HDRP-2010-36, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

SAKAGAMI, Y., R. WATANABE, C. AOYAMA, S. MATSUNAGA, N. HIGAKI, AND K. FUJIMURA (2002): "The intelligent ASIMO: System overview and integration," Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems.

SAMPSON, T. (2012): "Selection into Trade and Wage Inequality," CEP Discussion Papers dp1152, Centre for Economic Performance, LSE.

SCHNABEL, I., AND R. SCHNABEL (2002): "Family and Gender Still Matter: The Heterogeneity of Returns to Education in Germany," *Centre for European Economic Research Discussion Paper*, No. 02-67.

SHEIN, E. (2013): "Businesses adopting robots for new tasks," *Computerworld*, Aug 1.

SIMON, H. (1960): "The corporation: Will it be managed by machines?," in *Management and corporations*, ed. by M. Anshen, and G. Bach, pp. 17–55. McGraw-Hill.

SOLON, G. (1999): "Intergenerational mobility in the labor market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 29, pp. 1761–1800. Elsevier.

SPENCE, A. M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–74.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STATISTISCHES BUNDESAMT (2013): "Allgemeinbildende Schulen," *Fachserie 11, Reihe 1*.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–29.

TAYLOR, K. (2010): "That Mighty Sorting Machine Is Certainly One for the Books," *The New York Times*, Apr 22.

THISTLETHWAITE, D., AND D. T. CAMPBELL (2001): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51(6).

URQUIOLA, M., AND E. VERHOOGEN (2009): "Class-Size Caps, Sorting, and the Regression-Discontinuity Design," *American Economic Review*, 99(1), 179–215.

WAGNER, G. G., J. R. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study (SOEP) Scope, Evolution and Enhancements," *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fr Wirtschafts- und Sozialwissenschaften*, 127(1), 139–169.

WALDINGER, F. (2007): "Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?," unpublished.

ZEIRA, J. (1998): "Workers, Machines, And Economic Growth," *The Quarterly Journal of Economics*, 113(4), 1091–1117.