The London School of Economics and Political Science

Essays on Human Capital

Andy Feng

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, September 2013

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 3 and 4 were jointly co-authored with Georg Graetz with equal shares in all aspects of the papers.

Abstract

This thesis entitled "Essays on Human Capital" is comprised of three essays on various aspects of human capital and its effects on firms and labor markets. Chapter 1 provides an overview.

In Chapter 2 we estimate the effects of human capital on firm-level management practices. We adopt an instrumental variables strategy to overcome the potential endogeneity of human capital. Starting with data on management practices from the World Management Survey, we geocode the locations of more than 6,000 manufacturing plants in 19 countries. Then, we calculate driving times to universities in the World Higher Education Database. Using distance as an instrument for human capital, we estimate that every one standard deviation increase in the share of workers with a university degree leads to 0.5 of a standard deviation improvement in management. These findings are robust to a battery of checks and a placebo instrument using distances to world heritage sites. We show that both managers' and non-managers' human capital matter.

In Chapter 3 we estimate the effects of university degree class on initial labor market outcomes. We employ a regression discontinuity design which 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 is worth roughly 3 percent in starting wages which translates into £1,000 per annum. An Upper Second is worth more-7 percent in starting wages which is roughly £2,040. We interpret these results as the signaling effects of degree class and provide evidence consistent with this.

Finally in Chapter 4 we study the labor market effects of increased automation. We build a model in which firms optimally design machines, train workers, and assign these factors to tasks. Borrowing concepts from computer science and robotics, the model features tasks which are difficult from an engineering perspective but easy for humans to carry out due to innate capacities for functions like vision, movement, and communication. In equilibrium, firms assign low-skill

workers to such tasks. High skill workers have a comparative advantage in tasks which require much training and are difficult to automate. Workers in the middle of the skill distribution perform tasks of intermediate difficulty on both dimensions. When the cost of designing machines falls, firms adopt machines mainly in tasks that were previously performed by middle-skill workers. Occupations at both the bottom and the top of the wage distribution experience employment gains. The wage distribution becomes more dispersed near the top but compressed near the bottom. As design costs fall further, only the most skilled workers enjoy rising skill premiums, and an increasing fraction of the labor force is employed in jobs that require little or no training. The model's implications are consistent with recent evidence of job polarization and a hollowing-out of the wage distribution. In addition, the model yields novel predictions about trends in occupational training requirements that are consistent with evidence we present.

Acknowledgments

A Ph.D is a humbling journey and I was fortunate to receive the guidance and support of many along the way. First and foremost, I thank my advisor John Van Reenen. He was always approachable, patient and thoughtful, making supervision sessions something I looked forward to. His comments and insights shaped my thoughts on empirical work and economics more broadly. If I learned anything at all in these four years, John played a big role in it.

I also thank Nicholas Bloom, Steve Gibbons, Guy Michaels, Steve Pischke and Raffaella Sadun for encouragement, advice and useful comments. Whether in their offices, along the corridors, via conference calls or during seminars, chatting with them always refreshed my perspectives.

Georg Graetz, Yu-Hsiang Lei and Joao Paulo Pessoa were my lunch companions for three years. I never grew tired of the many topics we discussed, even though it was often about economics! More than that, they were friends I could count on. Nitika Bagaria and Anna Valero were always ready to listen to my ideas and give a helping hand. Renata Lemos was very kind in my first summer as a research assistant. Shawn Chen, Laure De Preux, Mirko Draca, Ben Faber, Vicky Fan, Ashwini Natraj, Beyza Polat, Stephan Seiler, Amar Shanghavi, Kati Szemeredi and so many others helped brighten my days.

I thank the Singapore Ministry of Trade and Industry for financial support and giving me this opportunity. The Centre for Economic Performance accommodated me during my thesis writing and this generosity was matched by the warmth of the staff– Anita Bardhan-Roy, Jo Cantlay, Linda Cleavely, Helen Durrant, Anna Graham, Joe Joannes, Nigel Rogers, Romesh Vaitilingam and Mary Yacoob.

My parents, Mom and Dad, have been so supportive throughout my studies. It was never easy to see their only son live eight time zones away for eight years. For their sacrifices, I owe them everything.

Finally, and most importantly, I thank my wife, Cynthia. Her love and encouragement gave me the strength and confidence to pursue this ambition. Our beautiful baby, Madeleine, makes it all worthwhile.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the most important ideas in economics is that differences in earnings across people reflect differences in their human capital. Human capital is a broad idea capturing the sum of skills, abilities, talents and even social or personality traits in a person. As modeled in Mincer (1974), human capital, like physical capital, can be invested in and accumulated over time. These investments are primarily thought of as schooling although the notion encompasses non-schooling investments like health and on-the-job training (Becker 1964). In the Mincer view, differences in investments result in differences in human capital and, ultimately, differences in earnings. But why does human capital result in differences in earnings? And are differences in earnings purely the result of differences in observed human capital? At an economy-wide level, how does human capital explain the distribution of earnings? This thesis is a collection of three essays studying these aspects of human capital.

The most direct way that human capital affects earnings is when it is used in the production of goods and services. Then, human capital is an input in the production function just like physical capital or intermediate materials. Another view, that we associate with Lucas (1978), is that human capital is allied to the *organization* of production. In this view, human capital is complementary to the efficiency with which output is produced. For example, better human capital could be associated with better management practices that enable productivity improvements. This is the hypothesis explored in Chapter 2. We find evidence that higher skills improve management practices in manufacturing plants.

Not all differences in earnings can be attributed to differences in human capital. This is because we do not actually observe human capital and can only measure it indirectly. This allows for an alternative interpretation of measures of human capital. In the Spence (1973) view, differences in schooling *reflect* differences in human capital rather than cause it. Because human capital is not observable, individuals take costly actions to signal it. Schooling provides such a signal even if it does not directly change human capital. In Chapter 3 we explore this aspect in the labor market for university graduates. We find evidence that degree classification plays a signaling role in determining initial labor market outcomes.

Taking a broader view of the economy, human capital alone does not explain all the variation in earnings because labor works with capital. This interaction between human and physical capital, between man and machine, is the subject of Chapter 4. We develop a theoretical model of how workers interact with machines in the production process to explore the effects of falling machine learning costs on the distribution of earnings and employment. This theory speaks to the effects of computerization on labor markets and explains the phenomena of job polarization. We argue that as machines get better at learning how to produce, workers will be increasingly polarized in jobs that require either low or high skill.

These essays appear to be a mix of micro- and macro-economics, of empirical and theoretical studies, but they have in common a view of human capital as a useful tool for understanding earnings differences. As Becker (1993) recognized, "it becomes clear that the analysis of human capital can help explain many regularities labor markets and the economy at large".

Chapter 2

Human Capital and Management Practices: Evidence from Driving Times to Universities

Abstract. We estimate the effects of human capital on firm-level management practices. We adopt an instrumental variables strategy to overcome the potential endogeneity of human capital. Starting with data on management practices from the World Management Survey, we geocode the locations of more than 6,000 manufacturing plants in 19 countries. Then, we calculate driving times to universities in the World Higher Education Database. Using distance as an instrument for human capital, we estimate that every one standard deviation increase in the share of workers with a university degree leads to 0.5 of a standard deviation improvement in management. These findings are robust to a battery of checks and a placebo instrument using distances to world heritage sites. We show that both managers' and non-managers' human capital matter.

2.1 Introduction

Management is an important influence on productivity. This was shown in survey data (Bloom and Van Reenen 2007) and established in experimental evidence (Bloom, Eifert, Mahajan, McKenzie, and Roberts 2013). Differences in management explain some of the variation in productivity across firms and even across countries. But what determines management? If firms have some control over the management practices they adopt, it is important to understand how differences in management arise. This is useful knowledge for businesses seeking higher profits and for economists explaining the vast productivity differences across firms (Syverson 2011).

This paper finds that human capital influences management– higher skills are associated with better management practices. We start with data from the World Management Survey which provides management practice scores on more than 6,000 manufacturing plants in 19 countries. We measure human capital as the plant-level shares of workers with a university degree. The cross-sectional correlation between human capital and management scores is likely to be confounded by omitted variables so we adopt an instrumental variables strategy. We calculate driving times between plants and nearest universities as an instrument for human capital.

We find that human capital has a positive and significant effect on management. Our central IV estimates suggest that increasing the share of workers with a university degree by one standard deviation increases the management score by 0.5 of a standard deviation. Using results from Bloom, Sadun, and Van Reenen (2012a), this translates into roughly 5 percent higher total factor productivities.[1] When we look at managers and non-managers separately we find that both matter. However, managers generate larger effects than non-managers. Every 10 percent increase in the share of managers with degrees improves management by 0.11 standard deviations. A similar 10 percent increase in non-managers generates a 0.04 standard deviations improvement. These findings are robust to a battery of checks and a placebo instrument using distances to world heritage sites.

Theoretically, we explain these results as the complementarity between human capital and the organization of production. Empirically, our results rely on the plausibility of university distance as an instrument for skills. We take several steps to address concerns regarding the validity of our empirical strategy. First, we control for region fixed effects and other geographic variables to avoid confounding influences from location-specific factors. Second, we isolate universities without business departments to avoid any direct effects on plant-level management. Third, we look at universities founded after the plant was located to address concerns regarding the endogenous location of plants. Fourth, we conduct a series of specification checks to show that our results are not a statistical artifact. Finally, we use distance to UNESCO world heritage sites as a placebo instrument to show that our results are not driven by some statistical artefact.

[1] Referring to the specification in table 3 column 2 of the paper.

2.1.1 Related Literature

Lucas (1978) developed a theory of human capital where the distribution of managerial ability determines firm sizes. Higher ability managers control larger firms subject to diminishing returns in the span of control. More recent literature exploring the links between human capital and productivity include Bartel, Ichniowski, and Shaw (2007), Bresnahan, Brynjolfsson, and Hitt (2002), Garicano and Rossi-Hansberg (2006) and Gennaioli, La Porta, Lopez-De-Silanes, and Shleifer (2013). Of note is Caroli and Van Reenen (2001) who test the hypothesis of skill-biased organizational change and find strong complementarities between human capital and organization of firms.

Bloom and Van Reenen (2007) pioneered research in measuring management practices. Follow-on work in Bloom, Sadun, and Van Reenen (2012b) and Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013) identifies competition, labor market regulations, ownership, trust and informational barriers as some of the factors driving the discrepancies in management across firms, industries and countries. This variation in management in turn explains a significant share of TFP gaps– up to half of the difference between US and other countries (Bloom, Sadun, and Van Reenen 2012a). Bloom, Garicano, Sadun, and Van Reenen (2009) distinguish between the effects of information and communications technologies in driving organizational hierarchies. This highlights the fact that technologies may have competing effects on the management and organizational structure within firms.

Another literature is linked to our empirical strategy of using distance as an instrument for skills. Much of this literature in labor economics is inspired by Card (1995).[2] In contrast to the labor literature, however, we use distance as an instrument for skills in businesses as opposed to workers.

The effects of distance on firm productivity is studied in Jaffe, Tratjenberg, and Henderson (1993). They look at the effects of distance on patent citations as evidence of local spillover effects on innovative activity. Many subsequent papers have refined and explored this topic (see for e.g. Audretsch and Feldman (1996) and Lychagin, Pinkse, Slade, and Van Reenen (2010)). A strand of this literature looks specifically at the effects of distances to universities. Hausman (2012) uses the Bayh-Dole Act in the US as a policy experiment to identify the effects of university innovation on

[2] In the education literature, Frenette (2006), Kjellstrma and Regnra (1999) and Spiess and Wrohlich (2010) are some recent papers that have looked at distances to university and its effect on enrolment decisions.

nearby firms. She finds faster employment and wage increases for establishments closer in industry and geographic space to the universities. Other examples include Anselin, Varga, and Acs (1997), Belenzon and Schankerman (forthcoming) and Henderson, Jaffe, and Tratjenberg (1998).

Finally there is the literature on local labor markets and geographic variation in the price of skills. Skills may have spillover effects on workers (Moretti 2004a) and firms (Moretti 2004b). Moretti (2011) provides a good summary.

The paper is organized as follows. Section 2.2 develops a simple model of management and human capital to illustrate key ideas. Section 2.3 presents the empirical strategy and descriptive statistics. Section 2.4 examines the OLS reduced form results while Section 2.5 reports the IV regression results. Section 2.6 discusses our findings and Section 2.7 concludes.

2.2 Simple Model of Management Technology and Human Capital

In this section we outline a simple model of management and human capital to illustrate one path to our estimating equations. In a static environment we assume a neoclassical production function $Y = f(A, M, H)$ where output Y is some function of technology and human capital inputs H with $\partial Y/\partial H > 0$, $\partial^2 Y/\partial H^2 < 0$. Following Lucas (1978) we make a distinction between production technology A and management technology M.[3] Following Bresnahan, Brynjolfsson, and Hitt (2002) we model the human capital-management complementarity, $\partial^2 Y/\partial M \partial H$, as:

$$M = g(H, A, \eta) \tag{2.1}$$

where η is an idiosyncratic error term. This model captures the fact that even conditioning on the level of human capital, there is variation across plants in management due to other technological reasons or idiosyncratic factors. In this simple setup we abstract from modeling A. A dynamic model would treat technology as draws from a known distribution, see for e.g. Hopenhayn (1992) or Melitz (2003).[4] Several interpretations of our model are possible. If we interpret **??** as a production function, better management is "produced" by higher skilled workers or managers. Alternatively, a Nelson and Phelps (1966) interpretation is that higher skilled managers are able to draw and adapt random management technology from a better distribution. An interpretation closer to the Lucas

[3]See Bloom, Sadun, and Van Reenen (2012a) for a fuller description of management as a technology.

[4]We also abstract from entry and exit decisions by assuming that A is large enough to cover fixed costs.

(1978) model is that skilled managers are assigned better workers within a matching framework.

2.3 Empirical Strategy and Data Description

2.3.1 Empirical Strategy

We now move from our simple model to an empirical strategy for estimating the effects of human capital on management. Our unit of observation is a manufacturing plant. Suppose we estimated the following equation using OLS:

$$M_i = \beta_0 + \beta_1 H_i + \epsilon_i \tag{2.2}$$

where as before M is the management score, H is the level of human capital (which we measure as the share of workers with a university degree) and ϵ is an idiosyncratic error term. In the limit,

$$p \lim \beta_1^{OLS} = \beta_1 + \beta_2 \frac{cov(H, A)}{var(H)}$$

where β_2 is the effect of technology on management. Depending on the nature of the omitted technologies, their effects on management and their correlation with human capital, OLS could be biased upwards or downwards. For example, if information technologies that facilitate better management practices are positively correlated with skills, there would be an upwards bias in OLS. On the other hand, if communications technologies that facilitate better management lead to a reduction in worker skills, there would be a downwards bias in OLS.[5]

We propose an instrumental variables strategy to overcome this endogeneity bias. Our identification strategy can be described schematically as:

Distance to university $\rightarrow$ Share of workers with a university degree $\rightarrow$ Management practices

The first arrow in the diagram describes the relationship between distance and skill shares. In a frictionless world, the law of one price ensures that the price of skill is equalized across space and the distance to universities should have no effect on skill shares. Frictions and the inelastic

[5] Bloom, Garicano, Sadun, and Van Reenen (2009) investigate the effects of information and communications technologies on the hierarchy of firms. They find that improvements to information technology push decisions down leading to decentralization while improvements in communications technologies push decisions up leading to centralization.

supply of non-tradables, such as land, limit the action of price equalization (Roback 1988, Glaeser and Gottlieb 2009). In this paper we utilize within-region variation in the proximity to universities that drive variation in skill prices across firms. The empirical literature on the mobility of college graduates *within* regions after graduation is scarce.[6] Kodrzycki (2001) looks at NLSY data in the US and finds that less than one-third of college graduates migrate states after graduation. While college graduates are the most mobile education group, it appears that many migrate for the purposes of attending college and stay in the same region for work. For our distance instrument, it suffices that mobility is imperfect after graduation and plants that locate near to universities benefit from a lower cost of hiring skills.

Using the distance instrument for skills, we write our first-stage equation as:

$$H_i = \alpha_0 + \alpha_1 D_i + \nu_i. \tag{2.3}$$

We measure distance, D, as the driving time in hours between the plant and its nearest university. We expect that greater distances from universities reduces skills. This corresponds to a negative sign on α_1.

We prefer using driving times as opposed to simpler straight-line measures for two reasons.[7] First, driving times are a more refined measure of market access (Gibbons, Lyytikainen, Overman, and Sanchis-Guarner 2012, Sanchis-Guarner 2012). Second, driving times account for natural geographic features that would be missed in a straightforward measure like straight-line distance. To give an actual example in the data, a plant in Scotland has a university within 100km straight-line distance (which is predicted to correspond to a 1.7 hour driving time) but in reality this is a 7 hour driving time and a 410km driving distance. Thus the straight-line distance misses some valuable information about how isolated the plant is. Nevertheless, in robustness tests we find qualitatively similar results using straight-line or driving distances.

We have described the first-stage but a valid instrument require the exclusion restriction to be satisfied. In our schematic diagram, there should be no direct arrows from "distance to university" to "management practices" . There are at least three reasons why this could be violated. First, there could be location specific factors that both attract good managers and skilled workers and generate

[6]The literature mainly focuses on the mobility of college versus non-college workers across states from the time of birth. These estimates are biased by selection. See for instance Groen (2004), Malamud and Wozniak (2012) and Gregg, Machin, and Manning (2004). Furthermore, the literature that looks beyond the US or UK is even scarcer.

[7]The actual calculations are done using Haversine formulas which account for the curvature of the Earth.

agglomeration economies that reduce distances to universities. A second possible violation of the exclusion restriction arises when universities directly improve management. An obvious example would be a business school that trains managers or offers management consulting services. A third situation is reverse causality whereby better (or worse) managed plants endogenously choose to locate nearer to universities to tap on local skill markets.

To deal with location specific factors confounding our estimates we do several things. First, in our benchmark specification we control for region fixed effects. This eliminates omitted variables that vary at the region level, in particular regional variation in skill prices. In robustness checks we control for more demanding city-effects and find similar results. This suggests that even *within* cities, plants which are closer to universities benefit from higher skills. More directly we also include controls for the population density, longitude and latitude of the plant.

To avoid the direct effects of business schools, we look at universities without business departments. Although the majority of universities in our sample have a business department, we show in regressions below that excluding them does not change our results and suggests that our identification is not only coming from business schools.[8]

To tackle the third issue of endogenous plant location, we examine universities founded *after* the plants were located.[9] Here there is less possibility that universities choose locations close to medium-sized manufacturing plants or that the plant endogenously chose locations on the basis of future universities. We get qualitatively similar results here.

With these identification assumptions our IV estimator is consistent:

$$p\lim \beta_1^{IV} = \frac{cov(D, M)}{cov(D, H)} = \beta_1.$$

2.3.2 Data Description

We use data from two main sources (see Appendix Chapter A for full details). The World Management Survey (WMS) provides survey data on management practices and skills in a cross section of plants. The World Higher Education Database (WHED) provides location and other data on the population of universities. Our unit of analysis is the manufacturing plant. Figures 2.1 and 2.2 map the geographic distribution of plants and universities.

[8] We may still worry that non-business departments offer direct management consulting to local plants. In separate work we look more closely at the relationship between business schools and management practices.

[9] See Greenstone, Hornbeck, and Moretti (2010) for a study on how firms chose county locations.

Our outcome variable from the WMS is the standardized management score (see Appendix Chapter A for details on the survey instrument and the calculation of this score). We interpret a higher score as better management. To measure human capital we use the share of workers with a university degree. We look at the total workforce, managers and non-managers separately.[10]

We start by examining some country-level descriptives. Appendix Table B.3 shows the mean and standard deviations of management scores and degree shares across countries. Confirming the findings in Bloom, Sadun, and Van Reenen (2012a) the United States has the highest average management scores, although it is also clear that the within-country variation is substantial. For skills, Japan has the highest average skill share with 32 percent of the workforce in the average plant being university graduates.

Our instrumental variable for human capital is the distance between each plant and the nearest university. The WHED provides the addresses of universities and along with the locations of the plants. We calculate driving times between plants and nearest universities via google maps (see Appendix Chapter A for details on how this was done). Appendix Table B.3 reports the average and standard deviations of these driving times across countries. Similar to the management scores and skills, there is variation across countries in the mean distances. Although these cross-country comparisons are interesting in their own right, our focus is on finer grained analysis using within-country variation. In our main regressions we control for country effects and in our benchmark specification we control for region effects (which subsume country effects).

Appendix Table B.4 shows region level descriptive statistics. We use within-region variation in estimation so it is useful to highlight the number of regions and number of plants within regions. The region is the first-level administrative region in a country, e.g. for the United States this is the state. In this table we report the differences between the 90th and 10th percentile plant in management scores, degree share and distances. There is substantial within-region variation that we utilize in our empirical approach.

Now we move to the plant-level descriptive statistics. Table 2.1 reports descriptive statistics for the key variables that are used in our analysis. By construction, the management score is mean zero with standard deviation one. The average plant has 15 percent of its workers with a university degree. This is broken down into 58 percent of its managers and 10 percent of the non-managers

[10] Total workforce = managers + non-managers. The availability of skill measures is a unique feature of the WMS that is not readily available in other datasets like annual censuses.

with a university degree. In our regressions we also control for plant employment, firm employment, plant age (in years) and MNE status.[11]

The average distance between plants and universities is 0.4 hours (roughly 26 minutes by car). Figure 2.3 plots the histogram of driving times in 10 minute bins. We control for location features using longitude, latitude and average population density within a 100km radius of the plant. We also check the robustness of these geographic controls to various non-linear specifications.

In Appendix Table B.5 we provide summary statistics for the additional variables used in our robustness checks. 60 percent of our plants are part of multi-unit firms, 28 percent of them are in firms that are listed and 40 percent of the workforce in the average plant is a union. The universities are described by several characteristics. *Arts department*, *social sciences department* etc. are indicator variables for the presence of this department in the university. For instance, 62 percent of universities have a business department. These indicator variables are not mutually exclusive because a university may contain several departments. *University founding* is the year in which instruction first began in the university. The average university was founded in 1945 and 60 percent of the universities were founded before the manufacturing plant.[12]

Appendix Table B.5 also provides alternative distance measures that we explore in robustness checks. The average driving distance is 27 km and the average straight-line distance is 21km. As an alternative measure of access, we also counted the number of universities within a 100km and 50km radius of the plant– on average there are 34 and 19 universities for each plant respectively. In our placebo test in Section 2.5.4 we use distances to UNESCO World Heritage Sites as an instrument for skills. The average site is 1,200km away. We explore various radiuses in calculating average population densities by using 50km and nearest centroid definitions. 10 percent of plants share postal codes with their nearest university and 1.6 percent of plants did not have a university within 100km radius. For these latter plants we winsorize their driving times using the region maximum to prevent outlier bias.[13]

To round up this section we list the full set of covariates that we use in our benchmark specification. To control for plant characteristics we include plant employment, firm employment, plant age, MNE status and 21 two-digit industry effects. To control for geography we include 313

[11] Dummy variables are included in regressions where these were missing. Imputation for missing values is described in Appendix Chapter A.

[12] Where founding dates were missing we imputed this using the regional average.

[13] In robustness checks we exclude these plants and find no difference in results.

region fixed effects (which subsume country effects), average population density within 100km of the plant, longitude and latitude. Finally, to control for noise from the survey, we include survey controls which are survey wave dummies, the gender, tenure and seniority of the manager who responded, the day of the week and hour of the interview, the duration of the interview, a measure of the reliability of the information as coded by the interviewer and a full set of 106 interviewer dummies. Throughout, we cluster standard errors at the region level. This accounts for heteroscedasticity and allows for unrestricted correlation between plants in a region. In robustness checks we experiment with other corrections for the standard errors.

2.4 Reduced Form Results

2.4.1 Reduced Form and First-Stage Regressions

In this section we look at reduced form regressions of management scores and human capital on driving times to universities. That is, we estimate equations:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \eta_i \tag{2.4}$$

for $Y \in \{M, H\}$.[14] To visualize our results, Figure 2.4 plots average management scores within 10 minute driving time bins. We note two points from this simple graph. First, there is a clear negative correlation between the driving time to a university and how well managed the plant is. This negative correlation is robust to the inclusion of many other covariates we explore subsequently. Second, this negative relationship exists whether we use a simple linear specification or a non-linear specification. In robustness tests we show that the results hold when we allow for non-linearity.

Table 2.2, panel A reports results from regressing management scores on distances. Column (1) corresponds to the specification underlying Figure 2.4 and includes only survey controls and country fixed-effects. There is a significant and negative relationship between management scores and distances to universities. Every extra hour of driving time (which is roughly 2 standard deviations) leads to a 0.07 standard deviation drop in the management score. Column (2) adds region fixed-effects while column (3) adds industry fixed-effects which are both highly significant as shown by the p-values reported at the bottom of the panel. In columns (3) to (5) identifying variation is

[14] As noted in Angrist and Krueger (2001), the reduced form effects are proportional to the coefficient of interest. Thus the strength of the reduced form is an indicator of the effect of interest.

coming from within regions and industries. Column (4) adds plant and firm employment to control for plant size as well as controls for plant age and MNE status– the coefficient on distance is slightly smaller at -0.05. Finally in column (5) we control for the location specific factors that may confound the relationship between unobserved management and the distance to universities by including the average population density, longitude and latitude. These are individually insignificant and do not change the coefficient on distances. Column (5) is our benchmark specification.

Table 2.2, panel B reports the first-stage regressions of degree share on distances. Degree share is the percentage share of the total workforce in a plant with a university degree. In column (1) we see a significant and negative correlation between distances and degree shares. Every additional hour of driving reduces the degree share by 2.3 percentage points (mean degree share is 14.8 percent). The other columns in panel B are arranged as discussed previously. Moving straight to the benchmark specification in column (5), the coefficient drops in magnitude to -1.5 but is still economically and statistically significant. This is preliminary evidence that our first-stage is strong and does not suffer from a weak instruments problem (Section 2.5 explores this in more detail).

Apart from using the share of total workforce with degrees as our measure of human capital, we also look at managers and non-managers separately.[15] Table 2.2, panels C and D report the same progression of specifications using degree share of managers and non-managers as the dependent variables, respectively. Reporting the results from the benchmark specifications in column (5), every additional driving hour reduces the share of managers and non-managers with degrees by 2.5 and 1.2 percentage points respectively. The average plant has 58 percent of managers and 10 percent of non-managers with degrees so this corresponds to a 4 percent and 12 percent reduction in the managers and non-managers' skills. This is an economically significant reduction in skill.

2.4.2 Heterogeneity

In this section we explore if our reduced form relationship between management and distance to universities exhibits heterogeneity along observable university or plant characteristics. This exercise serves two purposes. First, we might be interested in modifying our empirical strategy if heterogeneities existed. For instance, if we found that our results were driven only by business departments, we may be worried that our effects stem directly from consulting services and would

[15] Although we could include both managers and non-managers in the same specification, we chose not to do so because this would require at least two instruments for the two endogenous variables. As we show in Section 2.5 our distance instruments are correlated so using any subset of them may generate a weak instruments problem.

reconsider our instrumental variables approach. Second, our interpretation of the instrumental variables regressions may be shaded by the presence of heterogeneities. In particular, strong evidence of heterogeneities would suggest that our estimates are local effects and lessen the external validity of our results

In Table 2.3 we run modifications of the benchmark regression by interacting driving times with dummies for university characteristics. That is, we estimate:

$$M_i = \phi_0 + \phi_1 D_i + \phi_2 D_i \times UNI + \phi_3 UNI + \phi_4 X_i + \eta_i \tag{2.5}$$

where UNI is a dummy variable indicating the presence of a particular department in the university. Column (1) reproduces the benchmark results for easy reference. By inspecting the coefficient on the distance variable in the first row, it appears that there is no clear evidence of heterogeneity across department types. The exceptions are the social science departments in column (3) and the science and technology departments in column (7) which appear to render the main distance effect insignificant and small in magnitude. This is surprising but does not directly affect our empirical strategy. Our main concern was that only business departments are driving the management effect and we would not be able to identify the human capital channel. Column (4) suggests that business departments do not exert any additional effects on management. In column (8) we use a dummy variable for whether a university has all the listed departments (15 percent of universities do). This is a measure of university size and the results show that our estimates are not driven only by the large universities.

Next we ask if our results vary across plant characteristics. To explore heterogenous effects for plants we estimate:

$$M_i = \psi_0 + \psi_1 D_i + \psi_2 D_i \times PLANT + \psi_3 PLANT + \psi_4 X_i + \upsilon_i \tag{2.6}$$

where $PLANT$ is a plant characteristic listed in the rows of Table 2.4. The first row of coefficients on the distance variable suggests that there is little evidence of heterogeneity across most plant characteristics. The one exception is the MNE variable in column (2). The results here suggest that the distance effect is stronger for non-MNEs which accords with the idea that MNEs who have access to larger skill markets may be less influenced by the local price of skill.

In summary, we find little evidence of heterogeneity that would affect our identification strategy.

2.5 IV Results

2.5.1 IV Regressions

So far we have estimated reduced form regressions. To identify our structural model of management practices and human capital, we now turn to instrumental variables regressions. We start by looking at the OLS results for **??** reported in column (1) of Table 2.5. In panel A we regress management on total degree share and find a positive and precisely estimated effect. Panels B and C show that this result is present when we look at managers and non-managers separately. The coefficient in panel A column (1) suggests that a one standard deviation increase in degree shares by 17 percentage points is associated with a 0.14 standard deviation increase in management scores.[16] Columns (1) in panels B and C suggest that a one standard deviation increase in manager and non-manager degree shares is associated with a 0.14 and 0.11 standard deviation increase in management scores respectively.[17] However our discussion in Section 2.3 suggests that we cannot interpret these OLS results as causal because of the endogeneity of human capital.

In column (2) we report results from the just-identified IV regression where we instrument human capital with driving times. We also report the first-stage coefficient from a regression of degree share on distances and the F-statistic for excluded instruments is given at the bottom.[18] The F-statistic of 13 is of reasonable magnitude and does not suggest a weak-instruments problem.[19] The coefficient on degree share is interpreted as the causal effect of human capital on management. A one standard deviation increase in degree share (17 percentage points) leads to a 0.5 standard deviation increase in management scores. Panels B and C report results using different measures of skill and reveal that there is a positive and significant causal effect of human capital on management. Every one standard deviation increase in manager and non-manager degree shares leads to a 0.6 and 0.7 standard deviation improvement in management scores.

Although it would be useful to include both managers and non-managers in the same specification to tease out the relative importance of the two, in practice we do not have sufficient

[16] $16.8 \times 0.008 = 0.136$

[17] $33.9 \times 0.004 = 0.135$ for managers, $16.3 \times 0.007 = 0.114$ for non-managers

[18] Note that this first-stage regression is the benchmark model we reported in column (5) of panel B in Table 2.2.

[19] Staiger and Stock (1997) suggest an F-statistic of 10 as a rule of thumb. See also Stock, Wright, and Yogo (2002)

instruments. This is seen in column (3) where we include the number of universities within a 100km radius of the plant as an additional instrument. As noted earlier, this measure is highly correlated with driving times (even though it is the least correlated compared to driving distances and straight-line distances). The coefficient on degree share does not change much from column (2). While the first-stage coefficient on the number of universities is significant at 10 percent, the first-stage F-statistic has now dropped to 9.5. A similar pattern emerges when we look at managers and non-managers separately (although for managers, the first-stage F-statistic increased slightly). We conclude that the just-identified model using only distance as instrument is our preferred model. A discussion of the magnitudes of OLS and IV estimates is left to Section 2.6.

2.5.2 Assessing Instrument Validity

As discussed in Section 2.3, there are at least three concerns about instrument validity that we need to tackle. The first concern regarding location-specific factors is addressed by controlling for region fixed-effects and geography covariates. More seriously, we may be concerned that business departments have a direct effect on management practices or that better (or worse) managed plants endogenously locate closer to universities. We address these concerns in Table 2.6.

We tackle the "business schools" problem by estimating the following second-stage and first-stage equations:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \phi_1 D_i \times BUSINESS + \phi_2 BUSINESS + \nu_i \tag{2.7}$$

$$H_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \delta_1 D_i \times BUSINESS + \delta_2 BUSINESS + \epsilon_i. \tag{2.8}$$

Here $BUSINESS$ is a dummy variable for whether a university has a business department. Thus we allow business schools to have a direct effect on management practices (both a main effect and interacted with distances).[20] We instrument for human capital using only distances to universities without business departments, i.e. D_i.

We first report the OLS results for **??** in Table 2.6, panel A, column (1). Business departments do not have a direct effect on management once distances are controlled for. This is confirmed in the IV regression reported in column (2). As before we report the first-stage coefficient of the

[20] This is a form of the over-identification test. Card (1995) employs a similar strategy allowing distance to have a direct effect and using distance interacted with family variables as the excluded instrument for college.

excluded distance instrument which is still significant with an F-statistic of 11. The coefficient on degree share is slightly smaller (0.023 compared to 0.032) and still positive and significant. Results for managers and non-managers reported in panels B and C are qualitatively similar although less precisely estimated. Business schools do not appear to have direct effects on management.[21]

We conduct a parallel exercise for universities founded *after* the plant was founded to address the concerns over endogenous location. Specifically, we estimate the following second- and first-stage equations:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \phi_1 D_i \times BEFORE + \phi_2 BEFORE + \nu_i \tag{2.9}$$

$$H_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \delta_1 D_i \times BEFORE + \delta_2 BEFORE + \epsilon_i \tag{2.10}$$

where $BEFORE$ is an indicator for whether the university was founded before the plant, i.e. pre-existing universities. The excluded instrument D is now the distance to a university that was founded *after* the plant was located.

The OLS regressions in Table 2.6, panel A, column (3) suggest that universities that were founded before the plant have little direct effect on management. This is inconsistent with the story that plants endogenously locate near to pre-existing universities on the basis of management scores but consistent with our identifying assumption that distances are exogenous. However, the IV results in column (4) reveal that the estimates are imprecise. The first-stage F-statistic is only 4.8. Although the point estimate for distance is larger in magnitude compared to the plain IV (0.062 compared to 0.032), the large standard errors means that we cannot reject that they are the same.

Results for managers and non-managers in panels B and C paint a similar picture although the estimates are imprecise. Neither business schools nor pre-existing universities have a direct effect on management scores.

2.5.3 Robustness Checks

In Table 2.7 we conduct a battery of tests on our benchmark IV specification. We report parallel robustness checks for the reduced form specification in the Appendix Table B.6 but do not detail the findings here in the interests of space. Each row reports a different specification based on the benchmark in Table 2.5, panel A, column (2). The first column of numbers is the coefficient on

[21]The corresponding reduced form regression is reported in row (2) of Table 2.3.

degree share and the second column is the standard error. The full sample of 6,406 plants is used unless otherwise stated. Row (1) repeats the results from the benchmark results for comparison.

In panel A we check the standard errors. Recall that in our benchmark specification we cluster standard errors at the region level. In row (2) we cluster the errors at the region $\times$ industry level, in row (3) we allow for two-way clustering at region and industry while in row (4) we cluster at the university level (Cameron, Gelbach, and Miller 2011). None of these corrections affect the significance of the results.[22]

Panel B allows for non-linearities in the effects of distance. First we may worry that the distribution of distance is skewed to the right (as shown in Figure 2.3). We check that taking logs does not affect the sign or significance of the results in row (5). Next in rows (6) to (8) we include various polynomials in the driving times up to a quartic (fourth-order polynomial) and use these as instruments for degree share. The coefficient on degree share continues to be significant.

Geography variables are checked in panel C. In row (9) we control for quartics in the average population density, longitude and latitude. P-values are reported in brackets and show that the polynomial in geographic controls are jointly significant but do not change the coefficient on degree share. In rows (10) and (11) we control for population density in a 50km radius and using only the nearest centroid. These do not affect the results.

Panel D checks that our results are robust to various measures of distance. Row (12) uses the driving distance from google maps. Because the driving times and driving distances are highly correlated there is little surprise that the coefficient is consistent with earlier results.[23] In row (13) we use the straight-line distance which shows a very similar coefficient that is significant at the 5 percent level. Finally in row (14), we look at the number of universities within a 100km radius of the plant as a measure of "density". Although this coefficient is correctly signed, it is imprecisely estimated because there is little variation within a region in the number of universities facing plants.

In panel E different samples of the data are used. As detailed in the Chapter A, some plants were interviewed in multiple waves and in our main analysis we used the latest interview. Row (15) checks that including all interviews does not change the results. Row (16) excludes the plants that share the same postal codes as its nearest university, which would have resulted in a zero reported driving time. Again results are robust to their exclusion. In row (17) we exclude those

[22] For the reduced form in Appendix Table B.6 we check Conley standard errors and find similar results.

[23] The pairwise correlations between driving times and driving distances, straight-line distances and number of universities are 0.61, 0.49 and -0.27 respectively.

plants for which the nearest university was greater than 100km away and thus had the driving times winsorized to the region maximum. In fact, results are stronger suggesting that geographically isolated plants are not driving our results. In rows (18) and (19) we distinguish between capital and non-capital regions.[24] The effects are positive in both capital and non-capital regions.

In panel F we experiment with various fixed effects. These specifications are very demanding on the data because we are comparing plants in smaller units of geography. Row (20) includes 2,283 region $\times$ industry fixed effects with little difference on the degree share coefficient. In row (21) we use control for 724 county fixed-effects.[25] The coefficient on degree share is now smaller and imprecisely estimated, but are still positive. Finally in row (22) we include 851 city fixed-effects thus narrowing our comparisons to plants in the same city. While the coefficient is imprecisely estimated, we observe that it is still positive and consistent with our other results.[26]

The robustness checks reveal a consistent picture of the effect of skills on management practices. Whichever way we correct the standard errors, allow for non-linearities in distance, control for geography, measure distances, select the sample, or include fixed effects, we find a positive effect of human capital on management.

2.5.4 Placebo Test Using Distance to World Heritage Sites

We may still worry that our results are capturing some statistical artefact or some other factor that by chance is correlated with distance to universities.[27] To exclude this possible explanation we use distances to UNESCO world heritage sites as a placebo instrument.[28] World heritage sites are cultural attractions and should not have any effect on the human capital or management of manufacturing plants.

Table 2.8 reports our results where we have repeated our estimation of **????** using distances to heritage sites. The reduced form result in column (1) and first stage estimates in column (2) show

[24] Capital regions correspond to the regions containing the country capital or the region with the most observations. These are Buenos Aires in Argentina, New South Wales in Australia, South in Brazil, Ontario in Canada, Region Metropolitana in Chile, Guangdong in China, Ile-de-France in France, Nordrhein-Westfalen in Germany, Attiki in Greece, Maharashtra in India, Lombardia in Italy, Chubu in Japan, Mexico in France, Auckland in New Zealand, Mazowieckie in Poland, Porto in Portugal, Vastra Gotalands lan in Sweden, South East in UK and California in US.

[25] We refer to the second regional administrative level as a county.

[26] We checked that these results were not driven by sample selection by running our benchmark specification on these smaller samples.

[27] Distances to universities may be correlated with distances to headquarters. Giroud (2013) and Kalnins and Lafontaine (2013) are two recent papers exploring how increasing distance from headquarters negatively affects plant-level productivity and survival.

[28] We chose to use world heritage sites because the geocoded list is easily available and covers all the countries in our sample. We use straight-line distances because some sites vary greatly in area and some are inaccessible by road.

that this placebo instrument has no effect on management or skills. Column (3) confirms that when we use this placebo as our instrument for skills, we find no effect of skills on management. This adds to the evidence that our instrument using distances to universities is capturing the price of skills.

2.6 Discussion

Our results suggest that human capital is important for management and point to the relevance of measuring both manager and non-manager skills. This is consistent with the theoretical complementarity between the management practices and skills but we have much less to say about the microeconomic channels through which this occurs. Explorations of heterogeneity across universities or plant characteristics in Section 2.4.2 revealed little that would shed light on this.

When we compare the OLS and IV estimates in Table 2.5, although both estimates are significant and positive, the IV figures are larger than the OLS. This may be cause for concern if this reflects violations of the exclusion restriction that are biasing our IV estimates upwards.[29] As discussed in Section 2.5.2 we have taken steps to show that our main concerns regarding the validity of our instrument appear not to affect our results. However, the exclusion restriction is ultimately not testable. Our interpretation of the differences in the OLS and IV estimates is that unobservable technologies bias the OLS estimates downwards. For instance, if communications technologies that improve management practices also reduced the need for skilled workers, the positive effects of skills on management would be attenuated.

2.7 Conclusion

In this paper we estimated the causal effect of human capital on management. We proposed using driving times to universities as an instrument for human capital. We argue that driving times are plausibly exogenous and conduct a series of checks that are consistent with this view. First, we control for region fixed effects and other geographic variables to avoid confounding effects from location-specific factors. Second, we look at universities without business departments to avoid any direct effects from universities on plant-level management. Third, we look at universities founded after the plant was founded to address concerns regarding the endogenous location of plants. In

[29] If the exclusion restriction is violated, IV estimates may be more biased than OLS (Murray 2006).

robustness checks we showed that results are qualitatively similar when we use different measures of distance, different standard error corrections, various sample selections, allow for non-linearities in distance and adopt stricter geographic controls.

Additionally, we argue that there is very little heterogeneity in these effects. First, these effects come from both manager and non-manager skills. Second, there is no heterogeneity arising from the different types of universities. Third, there is no heterogeneity when we look at plant observable characteristics.

While our results are not too surprising it does confirm the importance of human capital in an aspect of economics that is seeing increasing interest. Research on the managerial and organizational aspects of firms has been facilitated by more and larger datasets. In turn, this research is important for understanding the determinants of wage and productivity distributions.

Table 2.1: Descriptive Statistics

	Mean	S.D.	Min	Median	Max
World Management Survey plant level variables					
Management Z-Score	0	1	-2.89	0.024	2.93
Degree Share (percent)	14.8	16.8	0	9.19	100
Degree Share of Managers (percent)	58.2	33.9	0	60	100
Degree Share of Non-managers (percent)	10.5	16.3	0	5	100
Log plant employment	5.10	0.96	0	5.01	8.99
Missing log plant employment	0.017	0.13	0	0	1
Log firm employment	5.83	1.11	0	5.70	11.1
Missing log firm employment	0.0016	0.039	0	0	1
Log plant age	3.40	0.79	0	3.43	6.28
Missing log plant age	0.44	0.50	0	0	1
MNE	0.46	0.50	0	0	1
Google maps and GIS calculations					
Distance (hours)	0.45	0.54	0	0.28	7.55
Latitude	23.4	32.7	-54.8	37.9	65.7
Longitude	8.20	78.3	-127.5	0.39	176.9
Avg pop density ('000 per sqkm)	1.33	1.87	0	0.70	16.0

Notes: N=6,406. *Management Z-score* is the World Management Survey standardized score of management practices. *Degree Share, Degree Share of Managers* and *Degree Share of Non-managers* are plant-level percentages of total workforce, managers and non-managers with university degrees, respectively. *Log plant employment, Log firm employment* and *Log plant age* are employment and plant age data. Missing values of plant and firm employment are mean-coded and an indicator is included in all regressions. Missing values of plant age are imputed and an indicator is included in all regressions. *MNE* is a dummy variable indicating multinational status. Details of variable construction are provided in the Data Appendix. *Distance* is the google driving time in hours from the plant to the nearest university (full description in Data Appendix). *Longitude* and *latitude* are geographic coordinates of the plant location corresponding to its postal code. The mapping from postal codes to coordinates was done using the geopostcodes database. *Avg pop density* is the average population density within a 100km radius of the plant calculated using GIS software. Population density data is from the Gridded Population of the World, Center for International Earth Science Information Network (CIESIN).

Table 2.2: Reduced form effects of distances to universities on management and skills

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable is Management Z-score					
Distance	-0.069***	-0.073***	-0.054***	-0.050***	-0.049***
	(0.018)	(0.020)	(0.020)	(0.018)	(0.019)
Log plant employment				0.202***	0.201***
				(0.017)	(0.017)
Log firm employment				0.072***	0.072***
				(0.013)	(0.013)
Log plant age				-0.032**	-0.032**
				(0.014)	(0.014)
MNE				0.390***	0.389***
				(0.032)	(0.031)
Avg pop density					0.016
					(0.015)
Latitude					0.004
					(0.009)
Longitude					-0.011
					(0.007)
Regions p-value		< 0.01	< 0.01	< 0.01	< 0.01
Industries p-value			< 0.01	< 0.01	< 0.01
R-squared	0.345	0.380	0.404	0.492	0.493
Panel B: Dependent variable is Degree Share					
Distance	-2.267***	-2.020***	-1.565***	-1.502***	-1.533***
	(0.403)	(0.451)	(0.411)	(0.419)	(0.423)
Log plant employment				0.709**	0.696**
				(0.330)	(0.332)
Log firm employment				0.609**	0.609**
				(0.289)	(0.289)
Log plant age				-0.535*	-0.530*
				(0.284)	(0.285)
MNE				3.106***	3.105***
				(0.533)	(0.534)
Avg pop density					0.133
					(0.158)
Latitude					0.251*
					(0.152)
Longitude					-0.145
					(0.130)
Regions p-value		< 0.01	< 0.01	< 0.01	< 0.01
Industries p-value			< 0.01	< 0.01	< 0.01
R-squared	0.144	0.197	0.243	0.255	0.256
Survey controls	x	x	x	x	x
Region dummies (313)		x	x	x	x
Industry dummies (21)			x	x	x

(continued…)

Table 2.2: Reduced form effects of distances to universities on management and skills (cont.)

	(1)	(2)	(3)	(4)	(5)
Panel C: Dependent variable is Degree Share of Managers					
Distance	-3.458***	-3.302***	-2.769***	-2.607***	-2.577***
	(0.940)	(0.912)	(0.905)	(0.895)	(0.908)
Log plant employment				1.851***	1.837***
				(0.598)	(0.597)
Log firm employment				0.901*	0.903*
				(0.515)	(0.519)
Log plant age				0.596	0.602
				(0.585)	(0.585)
MNE				7.086***	7.059***
				(1.002)	(0.995)
Avg pop density					0.549***
					(0.204)
Latitude					0.281
					(0.591)
Longitude					-0.265
					(0.323)
Regions p-value		< 0.01	< 0.01	< 0.01	< 0.01
Industries p-value			< 0.01	< 0.01	< 0.01
R-squared	0.273	0.319	0.353	0.367	0.367
Panel D: Dependent variable is Degree Share of Non-managers					
Distance	-1.893***	-1.595***	-1.158***	-1.113***	-1.149***
	(0.384)	(0.440)	(0.403)	(0.412)	(0.417)
Log plant employment				1.163***	1.150***
				(0.332)	(0.334)
Log firm employment				0.653**	0.651**
				(0.295)	(0.295)
Log plant age				-0.607**	-0.603**
				(0.279)	(0.280)
MNE				2.609***	2.612***
				(0.517)	(0.518)
Avg pop density					0.085
					(0.129)
Latitude					0.207
					(0.152)
Longitude					-0.166
					(0.127)
Regions p-value		< 0.01	< 0.01	< 0.01	< 0.01
Industries p-value			< 0.01	< 0.01	< 0.01
R-squared	0.119	0.172	0.207	0.223	0.223
Survey controls	x	x	x	x	x
Region dummies (313)		x	x	x	x
Industry dummies (21)			x	x	x

Notes: ***, **, * significant at the 1, 5 and 10 percent level. N=6,406. This table shows OLS regressions of management z-scores and skill measures on the distances to nearest universities. Each panel uses a different dependent variable and each column estimates a different specification. Refer to Table 1.1 for a description of the key variables. All regressions include survey controls for the survey wave, interviewee sex, interviewee job tenure, interviewee seniority, interview reliability, interview day of week, time and duration and 106 interview analyst dummies. Dummy variables were included to indicate when these variables were missing and missings were mean-coded. Standard errors are clustered at the region level (313 clusters).

Table 2.3: Reduced form effects of distance on management with interactions on university characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable is Management Z-score								
Distance	-0.049***	-0.039	-0.001	-0.053*	-0.049**	-0.042	0.001	-0.040*
	(0.019)	(0.046)	(0.048)	(0.028)	(0.025)	(0.034)	(0.033)	(0.021)
Distance x Arts		-0.011						
		(0.046)						
Uni has arts		0.000						
		(0.034)						
Distance x Social science			-0.057					
			(0.049)					
Uni has social sciences			0.036					
			(0.026)					
Distance x Business				0.006				
				(0.031)				
Uni has business				0.001				
				(0.027)				
Distance x Law					-0.000			
					(0.034)			
Uni has law					-0.002			
					(0.028)			
Distance x Medical						-0.011		
						(0.038)		
Uni has medical						0.014		
						(0.034)		
Distance x Science							-0.071**	
							(0.036)	
Uni has science							0.024	
							(0.025)	
Distance x All depts								-0.048
								(0.045)
Uni has all listed depts								0.024
								(0.035)

Notes: ***, **, * significant at the 1, 5 and 10 percent level. N=6,406. This table shows reduced form regressions of management scores on distances, university characteristics and their interactions. Col (1) reproduces the benchmark specification from Table 1.2, panel A column 5. See notes to Table 1.2 for a full description of covariates used in the benchmark model.

Table 2.4: Reduced form effects of distance on management scores with interactions on plant characteristics

	(1)	(2)	(3)	(4)	(5)
Dependent variable is Management Z-score					
Distance	-0.049***	-0.082***	-0.055*	-0.034	-0.020
	(0.019)	(0.022)	(0.028)	(0.021)	(0.027)
Distance x MNE		0.154***			
		(0.043)			
MNE		0.347***			
		(0.034)			
Distance x Multiunit			0.011		
			(0.037)		
Multiunit			0.070**		
			(0.028)		
Distance x Listed				-0.058	
				(0.046)	
Listed				0.225***	
				(0.032)	
Distance x Union					-0.001
					(0.000)
Union (percent)					0.000
					(0.000)

Notes: ***, **, * significant at the 1, 5 and 10 percent level. N=6,406. This table shows reduced form regressions of management scores on distances, plant characteristics and their interactions. Col (1) reproduces the benchmark specification from Table 1.2, panel A column 5. See notes to Table 1.2 for a full description of covariates used in the benchmark model.

Table 2.5: Instrumental variables estimates of effects of skills on management practices

	(1)	(2)	(3)
Specification	OLS	IV	IV
Dependent variable is Management Z-score			
Panel A: Degree share is endogenous skill measure			
Degree Share	0.008***	0.032***	0.036***
	(0.001)	(0.011)	(0.011)
First stage excluded instruments			
Distance		-1.533***	-1.507***
		(0.423)	(0.422)
No. of universities within 100km			0.011*
			(0.006)
First stage F-stat		13.15	9.55
Panel B: Degree share of managers is endogenous skill measure			
Degree Share of Managers	0.004***	0.019**	0.020***
	(0.000)	(0.008)	(0.008)
First stage excluded instruments			
Distance		-2.577***	-2.500***
		(0.908)	(0.896)
No. of universities within 100km			0.032***
			(0.009)
First stage F-stat		8.06	11.85
Panel C: Degree share of non-managers is endogenous skill measure			
Degree Share of Non-managers	0.007***	0.043**	0.045***
	(0.001)	(0.017)	(0.017)
First stage excluded instruments			
Distance		-1.149***	-1.144***
		(0.417)	(0.420)
No. of universities within 100km			0.002
			(0.006)
First stage F-stat		7.61	4.31

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at the region level (313 clusters). N=6,406. The dependent variable is the management z-score. Each panel uses a different measure of skill. Each column reports a different specification. See main text for details.

Table 2.6: Extended IV regressions

	(1)	(2)	(3)	(4)
Specification	OLS	IV	OLS	IV
Dependent variable is Management Z-score				
Panel A: Degree Share is endogenous skill measure				
Degree Share	0.008***	0.023*	0.008***	0.062*
	(0.001)	(0.013)	(0.001)	(0.033)
Distance × Bus. Dept	-0.035*	-0.020		
	(0.019)	(0.022)		
Bus. Dept	0.017	0.008		
	(0.024)	(0.025)		
Distance × Before plant founded			-0.010	0.082
			(0.025)	(0.074)
University founded before plant			0.019	-0.012
			(0.024)	(0.050)
First stage excluded instruments				
Distance		-2.295***		-1.168**
		(0.686)		(0.532)
First stage F-stat		11.18		4.82
Panel B: Degree Share of Managers is endogenous skill measure				
Degree Share of Managers	0.004***	0.020	0.004***	0.051
	(0.000)	(0.013)	(0.000)	(0.045)
Distance × Bus. Dept	-0.033*	0.004		
	(0.019)	(0.037)		
Bus. Dept	0.014	-0.016		
	(0.024)	(0.041)		
Distance × Before plant founded			-0.009	0.143
			(0.025)	(0.154)
University founded before plant			0.016	-0.075
			(0.024)	(0.107)
First stage excluded instruments				
Distance		-2.653**		-1.426
		(1.066)		(1.205)
First stage F-stat		6.19		1.39

(continued…)

Table 2.6: Extended IV regressions (cont.)

	(1)	(2)	(3)	(4)
Specification	OLS	IV	OLS	IV
Dependent variable is Management Z-score				
Panel C: Degree Share of Non-managers is endogenous skill measure				
Degree Share of Non-Managers	0.007***	0.028*	0.007***	0.090
	(0.001)	(0.016)	(0.001)	(0.060)
Distance × Bus. Dept	-0.038*	-0.024		
	(0.020)	(0.022)		
Bus. Dept	0.020	0.010		
	(0.024)	(0.025)		
Distance × Before plant founded			-0.014	0.097
			(0.026)	(0.101)
University founded before plant			0.021	-0.013
			(0.024)	(0.069)
First stage excluded instruments				
Distance		-1.865***		-0.808
		(0.685)		(0.514)
First stage F-stat		7.41		2.47

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at the region level (313 clusters). N=6,406. The dependent variable is the management z-score. Each panel uses a different measure of skill. Each column reports a different specification. See main text for details.

Table 2.7: Robustness checks of benchmark IV regression

Specification	Coefficient on degree share	(S.E.)
(1) Benchmark (Table 2.5, panel A, col (2))	0.032***	(0.011)
A. Checking standard errors		
(2) Cluster at region × industry	0.032**	(0.014)
(3) 2-way cluster at region + industry	0.032***	(0.009)
(4) Cluster at university	0.032**	(0.013)
B. Non-linearities in distance		
(5) log (1 + Driving time)	0.027***	(0.010)
(6) Quadratic in driving time	0.025***	(0.009)
(7) Cubic in driving time	0.019**	(0.008)
(8) Quartic in driving time	0.019**	(0.008)
C. Checking geography controls		
(9) Including quartic in geography controls (joint p-value = 3.09e-06)	0.031**	(0.014)
(10) Average population density within 50km (joint p-value=0.752)	0.034***	(0.013)
(11) Average population density nearest centroid (joint p-value=0.649)	0.035***	(0.011)
D. Different distance measures		
(12) Driving distance ('00km)	0.029**	(0.014)
(13) Straight line distance ('00km)	0.042**	(0.020)
(14) No. of universities within 100km	0.066	(0.049)
E. Sample selection		
(15) All survey waves, N=9,586	0.024**	(0.010)
(16) Exclude same postal codes, N=5,710	0.030**	(0.013)
(17) Exclude winsorized, N=6,302	0.034***	(0.012)
(18) Capital regions, N=1,884	0.014	(0.013)
(19) Non-capital regions, N=4,498	0.035**	(0.014)

(continued…)

Table 2.7: Robustness checks of benchmark IV regression (cont.)

Specification	Coefficient on degree share	(S.E.)
F. Fixed effects		
(20) Including 2,283 region × industry fixed effects (N=6,406)	0.034**	(0.017)
(21) Including 724 county fixed effects (N=4,553)	0.019	(0.012)
(22) Including 851 city fixed effects (N=2,756)	0.011	(0.012)

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at the region level (313 clusters) except otherwise stated. N=6,406 except otherwise stated. The dependent variable is the management z-score and the endogenous skill measure is degree share. Each row presents a different robustness check of the benchmark 2SLS specification (same as Table 1.5, panel A, column 2).

Table 2.8: Placebo test using UNESCO world heritage sites

	(1)	(2)	(3)
Dependent variable	Management	Degree share	Management
Specification	OLS	OLS	IV
Distance to UNESCO world heritage site	0.010	0.003	
	(0.011)	(0.265)	
Degree share			2.817
			(206.9)

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at the region level (313 clusters). N=6,406. This table shows estimates of the reduced form, first-stage and IV regressions using the placebo instrument. *Distance to UNESCO world heritage site* is the straight line distance in '00km to nearest site. All models control for the same covariates as the benchmark specification. See main text for details.

Figure 2.1: World Management Survey plant locations, N=6,406

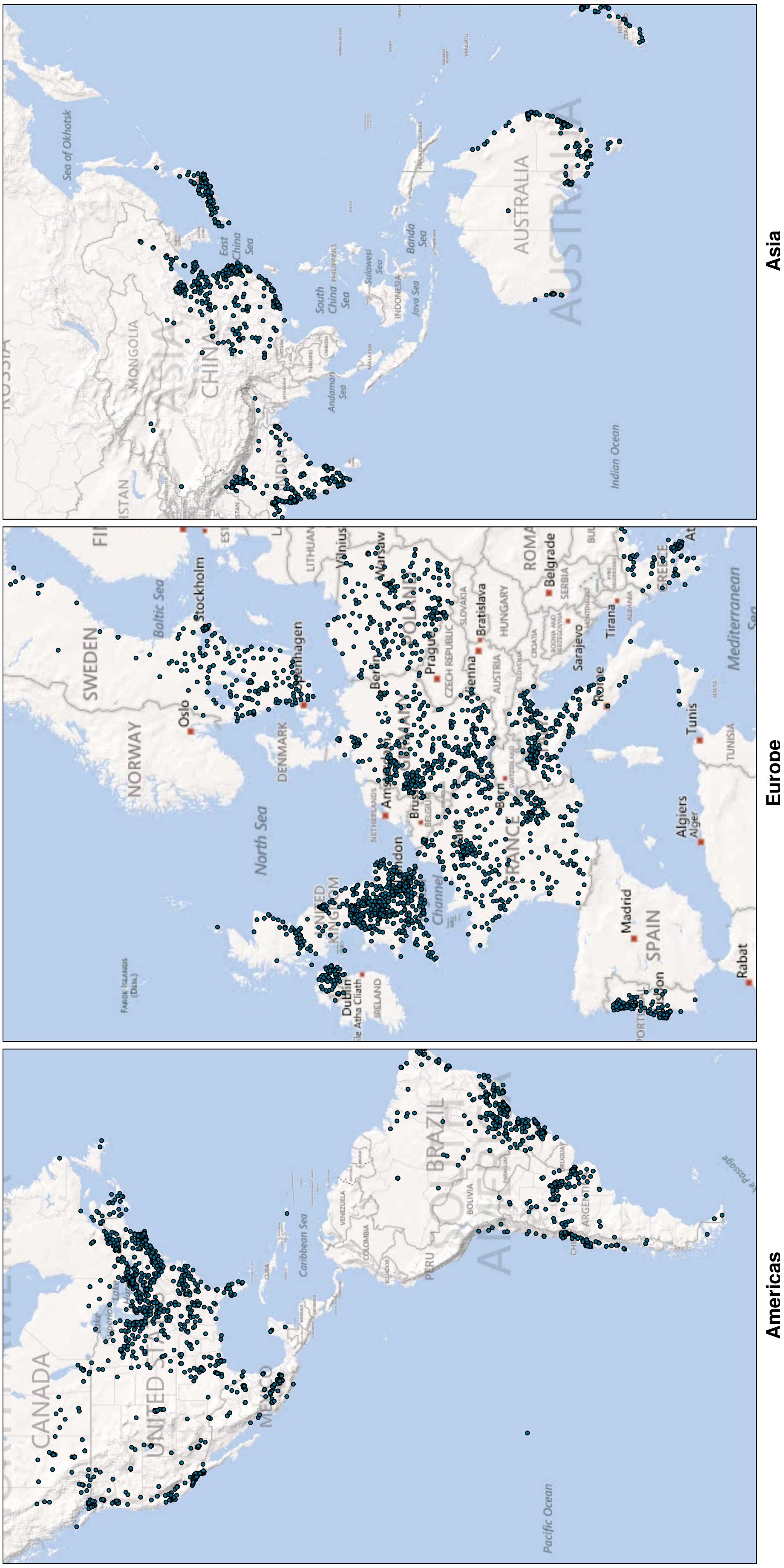


Figure 2.2: UNESCO World Higher Education Database university locations, N=8,656

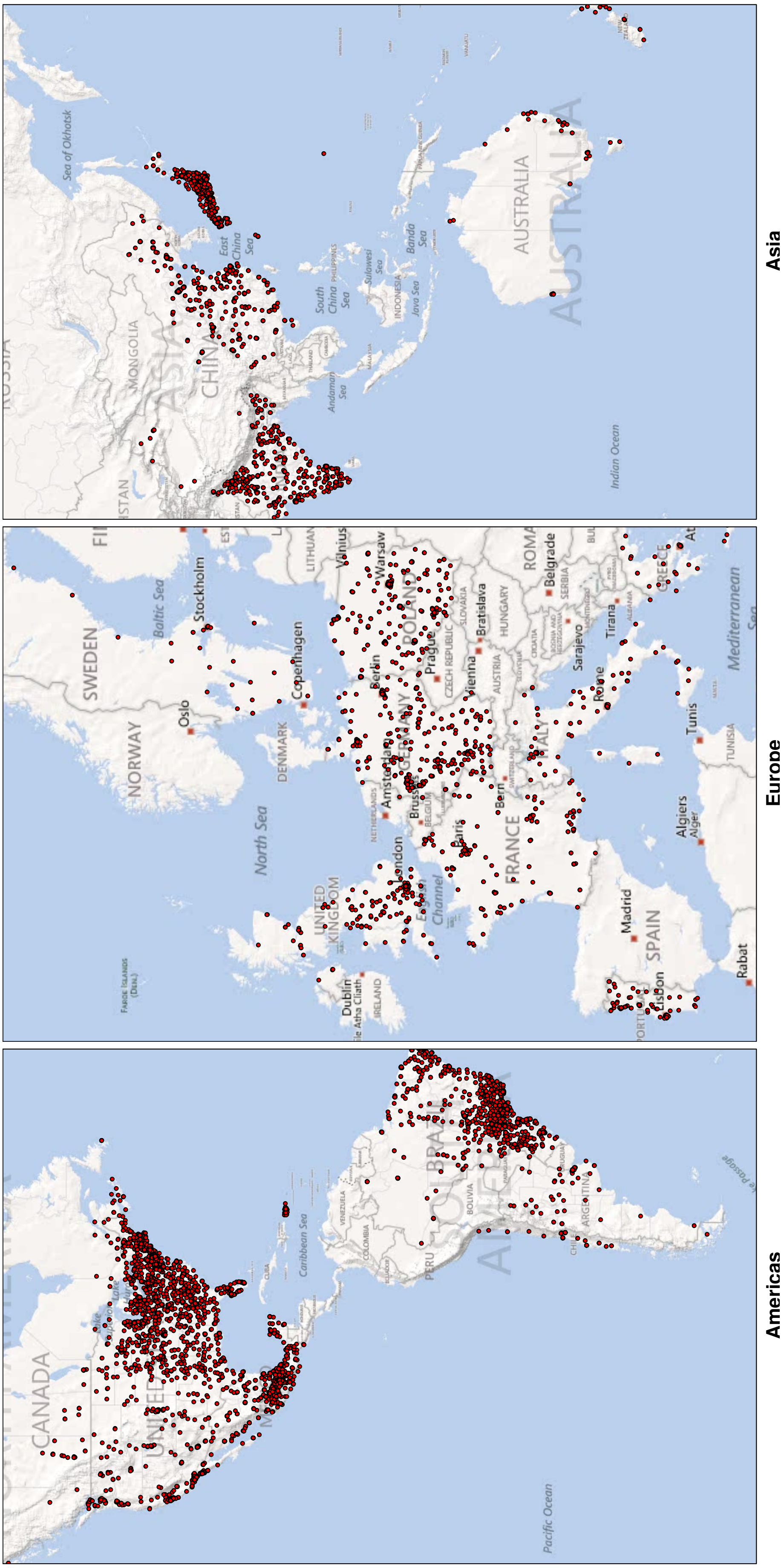


Figure 2.3: Histogram of distances between plants and nearest universities (10 minute bins)

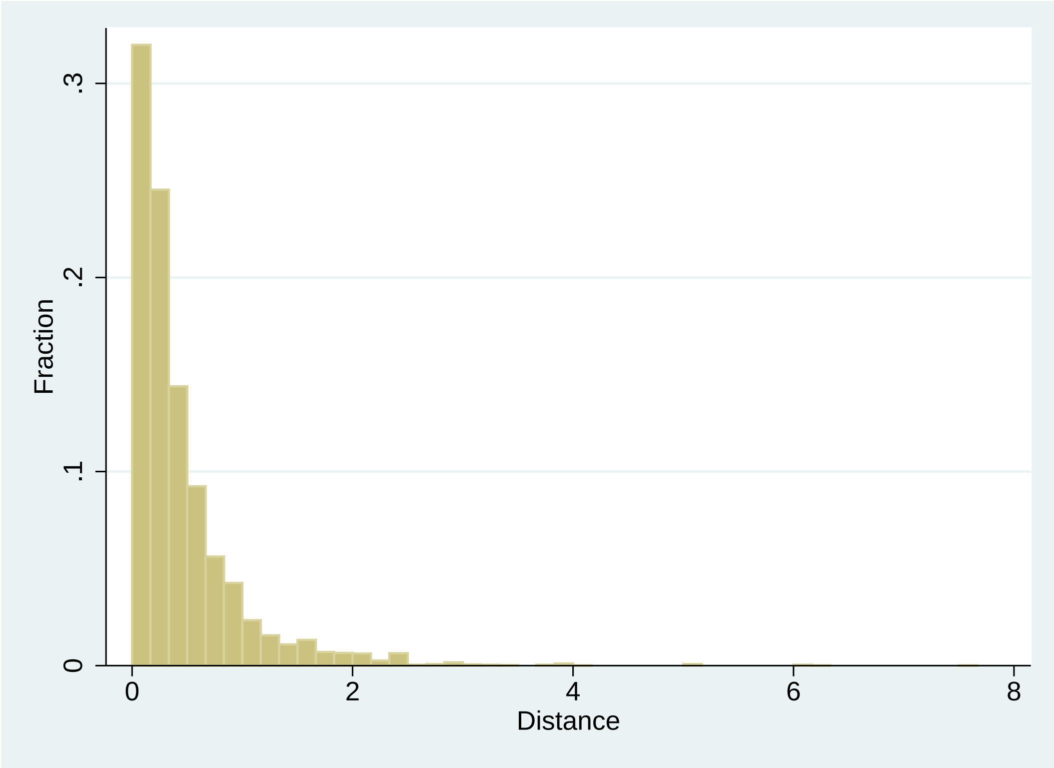
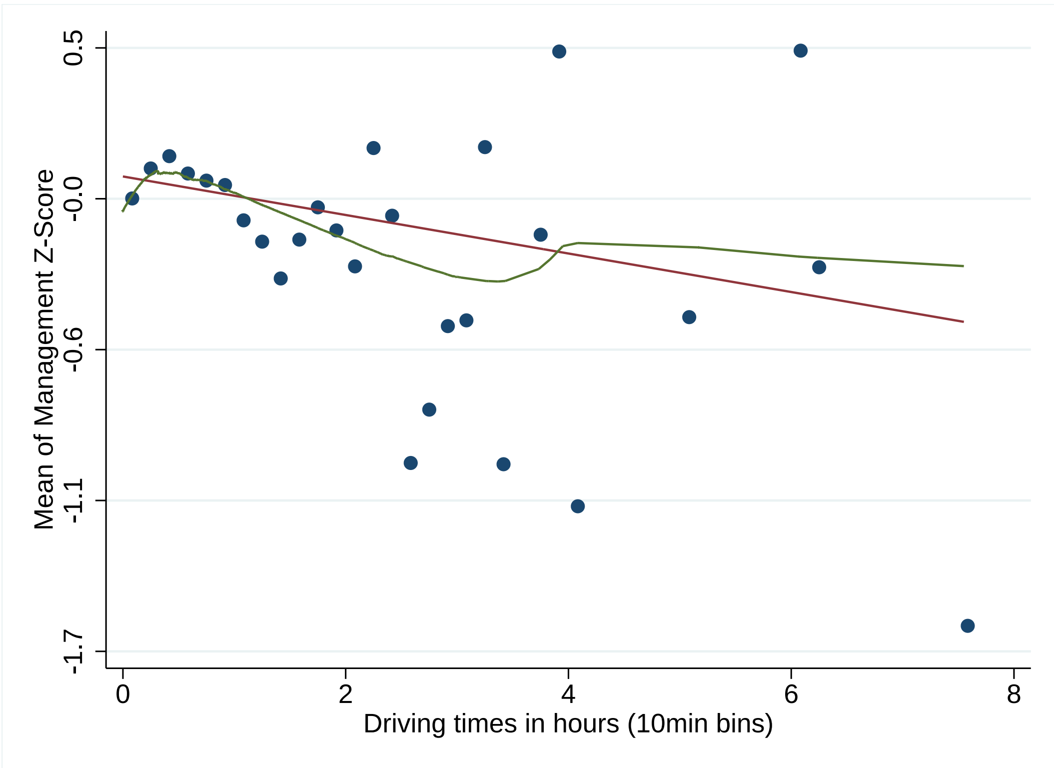


Figure 2.4: Plot of management z-scores against distances (10 minute bins)



Appendix A

Data Appendix

A.1 World Management Survey

The World Management Survey (hereafter WMS) offers unique survey data on management practices.[1] The WMS dataset gives us a cross section of management practice data in 6,406 plants across 313 regions in 19 countries in the final analysis sample. Here we describe the key aspects of this dataset and how it is used in this paper. Further details on the survey methodology is found in Bloom and Van Reenen (2007) (hereafter BVR), Bloom and Van Reenen (2010b), Bloom and Van Reenen (2010a) and Bloom, Sadun, and Van Reenen (2012a).

A.1.1 Sampling Frame

The sampling frame was based on firm-level accounting databases of the Bureau van Dijk (BVD) Amadeus dataset for Europe (France, Germany, Greece, Italy, Ireland, Poland, Portugal and the U.K.), on BVD Icarus for the US, on CMIE Firstsource dataset for India, on the BVD Oriana dataset for China and Japan, on BVD Orbis for Argentina, Brazil, Canada, Mexico, on BVD Orbis and Duns and Bradstreet for Australia and New Zealand, and on Industrial Annual Survey Sample of Firms (Encuesta Nacional Industrial Annual - ENIA) for Chile. These databases all provide sufficient information on companies to conduct a stratified telephone survey (company name, address and a size indicator). These databases also typically have accounting information on employment, sales and capital. Apart from size, accounting information was not used to form the

[1] http://worldmanagementsurvey.org/ has full details.

sampling population, however.[2]

Amadeus, Firstsource and Orbis are constructed from a range of sources, primarily the National registries of companies (such as Companies House in the UK and the Registry of Companies in India). Icarus is constructed from the Dun and Bradstreet database, which is a private database of over 5 million US trading locations built up from credit records, business telephone directories and direct research. Oriana is constructed from Huaxia credit in China and Teikoku Database in Japan, covering all public and all private firms with one of the following: 150 or more employees, 10 million US$ of sales or 20 million US$ of assets. ENIA, collected by the Chilean Statistic Agency, covers all the manufacturing plants that employ at least 10 individuals.

In every country the sampling frame for the management survey was all firms with a manufacturing primary industry code with between 50 and 5,000 employees on average over the most recent three years of data prior to the survey. In Japan and China they used all manufacturing firms with 150 to 5,000 employees since Oriana only samples firms with over 150 employees, while in Portugal they supplemented the sample with firms with 75 to 100 employees.

Because the sampling frame was based on accounting databases, one concern could be that the firms are not representative of the population. Bloom, Sadun, and Van Reenen (2012a) examine this extensively in the data appendix by comparing the size-distribution from the sample against national Census Bureau data from each of the twenty countries. The broad picture is that for most countries the coverage is comparable. In this paper we always control for country effects and this mitigates biases in cross-country comparisons.

A.1.2 Survey Method

The survey evaluation tool defines 18 key management practices in 4 broad areas and scores them from 1 (worst practice) to 5 (best practice). Table B.1 lists these 18 management dimensions and the nature of the questions asked. These practices codify the concepts of "good" and "bad" management into comparable measures across different firms, industries and countries. Together, these practices can be interpreted as a subset of a wider but unknown spectrum of management dimensions. These practices can be grouped into four areas: operations, monitoring, targets and incentives. The operations area focuses on the introduction of lean manufacturing techniques, the documentation of process improvements and the rationale behind the introduction of such im-

[2] Ireland was surveyed but excluded from our analysis because it does not use post codes.

provements. The monitoring area focuses on the tracking of performance of individuals, reviewing performance (e.g. through regular appraisals and job plans) and consequence management (e.g. making sure that plans are kept and appropriate sanctions and rewards are in place). The targets area examines the type of targets (whether goals are simply financial or operational or more holistic), the realism of the targets (stretching, unrealistic or non-binding), the transparency of targets (simple or complex) and the range and interconnection of targets (e.g. whether they are given consistently throughout the organization). Finally, the incentives area includes promotion criteria (e.g. purely tenure based or including an element linked to individual performance), pay and bonuses, and fixing or firing bad performers, where best practice is deemed the approach that gives strong rewards for those with both ability and effort.

Innovative steps were taken during the conduct of the survey to maximize the quality of the data. First, a double-blind methodology was employed to ensure unbiased responses to the survey questions. On one side of this double-blind method, interviewed managers were not told they were being scored. Open ended questions (e.g. “can you tell me how you promote your employees?”) were used as opposed to closed questions (e.g. “do you promote your employees on tenure?”). On the other side of the double-blind method, interviewers did not know anything about the firm’s financial information or performance in advance of the interview. Second, the survey was targeted at plant managers who were typically senior enough to have an overview of management practices but not senior enough to be detached from day-to-day operations of the enterprize. Third, skilled interviewers, typically MBA students, were hired to run the interviews because they generally had some business experience and training. Fourth, official government endorsements were sought to encourage responses. Fifth, interviewers were encouraged and incentivized to be persistent and in obtaining data. These steps helped to yield a high 44 percent response rate. Sixth, a series of data on the interview process itself was collected to serve as “noise controls” in regression analysis.

Following Bloom and Van Reenen (2007) we normalize each practice score to a mean zero, standard deviation one z-score then take the unweighted average of these and z-score this average again as the measure of managerial input in the plant.[3] This variable is called the “management z-score” in all regressions.

[3]This standardization is done unconditionally i.e. across regions and countries.

A.1.3 Survey Waves

Interviewers were each given a randomly selected list of firms from the sampling frame described above. This should be representative of medium sized manufacturing firms. The survey has been administered in several waves since 2004. There were three large waves in 2004, 2006 and 2009. While a few firms were re-interviewed over these waves we will not be exploiting the panel aspect of the data in this paper.[4] Instead we take the latest survey wave for the firm, although we do conduct robustness checks that include all survey waves.

The response rate was 42.2 percent, a high success rate given the voluntary nature of participation. Of the remaining firms, 14.7 percent refused to be surveyed, while the remaining 42.9 percent were in the process of being rescheduled when the survey ended. A total of 10,163 interviews were available although not all were used in the final analysis sample. See sample selection sub-section below for further details.[5]

A.1.4 Validation

BVR explore both the internal and external validity of the survey tool. To validate the survey as a consistent measure of management, a 5 percent sample was re-surveyed using a second interviewer to independently survey a second plant manager in the same firm. The ideas was that two independent management interviews on different plants within the same firms reveal the consistency of measurements. They found that the correlation between first and second interviews was 0.51 (p-value of 0.001). This is highly significant and suggests the survey tool has internal validity.

To check the external validity of the management scores, BVR correlate the scores with observable measures of firm performance including sales, profitability and survival probabilities. They conclude that the management score has important empirical content and is not merely picking up noise or “cheap talk”.

[4] Up to three interviews were carried out for some firms. A sample of 732 firms from France, Germany the UK and the US with a manufacturing primary industry code and 50 to 10,000 employees (on average between 2000 and 2003) used in Bloom and Van Reenen (2007) were re-interviewed. In 2009/10 firms interviewed from 2004 and 20066 were re-interviewed. This was a sample of 4,145 firms from China, France, Germany, Greece, India, Italy, Japan, Poland, Portugal, the UK, the US and Sweden with a manufacturing primary industry code and 100 to 5,000 employees (on average prior to the survey).

[5] We drop observations which were missing or misreported postal codes and only kept the most recent interview wave of the remaining plants.

A.1.5 Contacts Project

Plant location information was not collected in the initial surveys. A separate contacts project was conducted during 2011 to collect data on the postal code locations of the interviewed plants.[6] This project was able to yield a substantial 97.5 percent response from the sampled firms.[7] Of the 10,002 firms in the sample only 416 (4.1 percent) were either missing postal codes or had incorrectly reported information.

A.1.6 Additional Data

Apart from the management score, we have three other sets of variables from the WMS–plant-level measures of skills, plant and firm-level control variables and survey noise controls. To measure plant-level skills we use the percentage share of the total workforce, managers and non-managers with university degrees. These were collected during the survey.

Firm accounting data on sales, employment, capital, profits, shareholder equity, long-term debt, market values (for quoted firms) and wages (where available) were available from the accounting databases described above and merged into the WMS. As detailed in the paper, we use data on plant and firm employment, plant age, MNE status and two-digit SIC industry. Additionally in robustness checks we include listing status, the number of competitors perceived by the manager, the number of production sites and the percent of union members.

Information was collected about the interview process itself that we include as noise controls. These are survey wave dummies, the gender, tenure and seniority of the manager who responded, the day of the week and hour of the interview, the duration of the interview, a measure of the reliability of the information as coded by the interviewer and a full set of 106 interviewer dummies. These covariates were chosen to follow previous specifications in Bloom and Van Reenen (2010b).

Missing values for plant employment, firm employment and interview noise controls were imputed using the average of these variables. A dummy variable is included in all regressions where these were missing. For plant age we followed the following imputation strategy. We first used firm age where that was available. Otherwise we "hot-decked" plant age using regressions on plant founding dates on all other regressors for the sample that was not missing plant age.[8]

[6] I thank the project leader Daniela Scur for this information.

[7] When the initial interviewed manager was no longer at the plant, they made sure that the manager was indeed previously at the plant and obtained the postal code from another manager.

[8] The full list of covariates is the same as that used in the benchmark regressions. This includes plant employment,

We experimented with a simpler strategy of using the region average plant age and found similar results.

A.1.7 Final Analysis Sample Selection

An initial 10,163 interviews were available. Ireland was dropped because it does not use postal codes and hence we could not establish the exact location of the plants– this resulted in 10,002 interviews remaining. A further 416 interviews had missing or misreported postal codes and were dropped resulting in 9,586 interviews remaining. As mentioned previously, a few plants were interviewed multiple times either during follow up waves or during the same wave as a second interview for internal validity checks. We chose the most recent interview of the plant in our sample and this resulted in 7,191 interviews remaining. In this sample the unit of observation is a plant. Finally, we drop plants with missing observations on the degree shares which is our key explanatory variable of interest . This results in our final analysis sample of 6,406 plants. In robustness checks, we repeated the benchmark reduced form specification for the 9,586 interviews for which we had postal code information and found similar results.

A.2 World Higher Education Database

The World Higher Education Database (WHED) is a database of higher education institutions across the world compiled by the International Association of Universities, an organization associated with UNESCO.[9] The WHED can be accessed online for a fee and provides a description of the education system and credentials of over 17,000 higher education institutions in more than 180 countries. Data includes information and admission criteria for national and overseas students, quality assurance and recognition systems and contact details for national bodies. Importantly, it also contains detailed information location, brief history, funding type, academic divisions and degrees awarded. This is the most comprehensive collection of information on higher education institutions available with worldwide coverage.

In this paper we use data on the university location, founding date, funding type (public or private funded) and availability of various academic divisions (arts, business, social sciences, law,

firm employment, MNE status, industry effects, region effects and interview noise controls described before.

[9] Website here http://www.whed-online.com/.

medical and science and technology).[10] Data is available for the population of universities in all 19 countries in our WMS sample.

A.3 Geographic Data

Our empirical strategy requires measurement of the distance between plants and universities. The first step is to obtain accurate measurements of locations. For this we geocoded the plants using postal codes from the contacts project and geocoded the universities using addresses provided in the WHED database. Geocoding was done using the GeoPostcodes database described below. Driving times and distances between geocoded plants and universities were then calculated using google maps, as described below. Additional geographic information was then added using GIS software.

A.3.1 GeoPostcodes Database

The GeoPostcodes database is a commercial website providing data on the region, city, longitude and latitude of postal codes in countries.[11] We purchased country-level databases for 18 of our countries in March 2012.[12] We use this database to match postal codes to geographic coordinates and regions. In Table B.2 we show the geocoding success rates across countries for WMS plants and WHED universities. On average, the geocoding success rate is very high, yielding 96 percent match for plants and 95 percent for universities. While there is variation across countries in the success rates, we always include country effects that would mitigate biases due to the accuracy of postal code information across countries. Figure 2.1 and Figure 2.2 map the geographic distribution of plants and universities.

One point to note is that a fraction of plants and universities appear to be in the same postcode and thus have the same geographic coordinates (this affects 10 percent of the plants). This could be due to postcodes being fairly large geographies or measurement errors in the postcodes. In robustness checks, we exclude these plants and find similar results.

[10] The WHED reports the departments by name, eg "engineering". Where founding dates were missing we imputed this using the regional average.

[11] Website here http://www.geopostcodes.com/.

[12] We did not purchase the database for UK because its price was substantially higher than for other countries. Instead, we used the `geocode` command in stata to geocode UK plants and universities. Information on `geocode` stata command available here http://ideas.repec.org/c/boc/bocode/s457450.html. It uses google maps to geocode postal codes.

A.3.2 Google Driving Times

We calculated the driving times between each plant and the nearest university. This was done using the `traveltime` command in stata.[13] This command uses the geographic coordinates of plants and universities and calculates driving times (in hours) via google maps. A corresponding driving distance (in kilometers) is also calculated. To minimize computing times we limited the search of the nearest university within a 100km Euclidean radius of each plant. Where a plant did not have a university within this radius, we find the nearest university within any distance and winsorized the resulting driving times using the regional maximum. This was done to minimize outlier bias.[14]

Driving times in google maps are calculated using information from GPS-enabled devices of users. To ensure that seasonality or varying traffic conditions were not affecting our results, we calculated another set driving times several months apart. The correlation between the two sets was 0.95.

A.3.3 CIESIN Population Data

We control for the population density at the location of the plant. The Center for International Earth Science Information Network (CIESIN) provides the Gridded Population of the World (GPW) that depicts the distribution of population across the world in 2000.[15] We use GIS software to spatially intersect each plant with population density data from the CIESIN within a 100km buffer and find the average population density within that buffer.[16]

A.3.4 UNESCO World Heritage List

In our placebo regression we look at distances to UNESCO World Heritage sites. The list of sites can be found at `http://whc.unesco.org/en/list`. We use the computationally

[13] Information on `traveltime` can be found here `http://ideas.repec.org/c/boc/bocode/s457449.html`. It uses google maps to calculate driving times.

[14] This affected 1.6 percent of the sample. In robustness checks we exclude these isolated plants from the analysis and find no difference in results. It should be noted that for the fraction of plants and universities that shared postcodes, the resulting google driving time would be reported as 0. In robustness checks we exclude these plants and find no significant difference in results.

[15] Data is available here `http://sedac.ciesin.columbia.edu/data/collection/gpw-v3`. Population density is represented as centroids in a features file. These centroids correspond to the smallest geography available for the country. For example, in the US this is the Census block group.

[16] We also checked the robustness of results with varying buffer sizes including the using only the nearest centroid. See main paper for details.

easier straight-line distances between plants and heritage sites.

Appendix B

Appendix Tables

Table B.1: Management practices

Management Practice	Area	Score from 1 to 5 based on:
1) Introduction of modern manufacturing techniques	Operations	What aspects of manufacturing have been formally introduced, including just-in-time delivery from suppliers, automation, flexible manpower, support systems, attitudes, and behavior?
2) Rationale for introduction of modern manufacturing techniques	Operations	Were modern manufacturing techniques adopted just because others were using them, or are they linked to meeting business objectives like reducing costs and improving quality?
3) Process problem documentation	Operations	Are process improvements made only when problems arise, or are they actively sought out for continuous improvement as part of a normal business process?
4) Performance tracking	Monitoring	Is tracking ad hoc and incomplete, or is performance continually tracked and communicated to all staff?
5) Performance review	Monitoring	Is performance reviewed infrequently and only on a success/failure scale, or is performance reviewed continually with an expectation of continuous improvement?
6) Performance dialogue	Monitoring	In review/performance conversations, to what extent is the purpose, data, agenda, and follow-up steps (like coaching) clear to all parties?
7) Consequence management	Monitoring	To what extent does failure to achieve agreed objectives carry consequences, which can include retraining or reassignment to other jobs?
8) Target balance	Targets	Are the goals exclusively financial, or is there a balance of financial and nonfinancial targets?
9) Target interconnection	Targets	Are goals based on accounting value, or are they based on shareholder value in a way that works through business units and ultimately is connected to individual performance expectations?
10) Target time horizon	Targets	Does top management focus mainly on the short term, or does it visualize short-term targets as a "staircase" toward the main focus on long-term goals?
11) Targets are stretching	Targets	Are goals too easy to achieve, especially for some "sacred cows" areas of the firm, or are goals demanding but attainable for all parts of the firm?
12) Performance clarity	Monitoring	Are performance measures ill-defined, poorly understood, and private, or are they well-defined, clearly communicated, and made public?
13) Managing human capital	Targets	To what extent are senior managers evaluated and held accountable for attracting, retaining, and developing talent throughout the organization?
14) Rewarding high performance	Incentives	To what extent are people in the firm rewarded equally irrespective of performance level, or are rewards related to performance and effort?
15) Removing poor performers	Incentives	Are poor performers rarely removed, or are they retrained and/or moved into different roles or out of the company as soon as the weakness is identified?
16) Promoting high performers	Incentives	Are people promoted mainly on the basis of tenure, or does the firm actively identify, develop, and promote its top performers?
17) Attracting human capital	Incentives	Do competitors offer stronger reasons for talented people to join their companies, or does a firm provide a wide range of reasons to encourage talented people to join?
18) Retaining human capital	Incentives	Does the firm do relatively little to retain top talent or do whatever it takes to retain top talent when they look likely to leave?

Notes: This table is taken from Bloom and Van Reenen (2010)

Table B.2: Geocoding success rate for World Higher Education Database and World Higher Education Database

	World Management Survey		World Higher Education Database	
	No. of Plants	Geocode rate	No. of unis	Geocode rate
Argentina	249	0.95	95	0.95
Australia	452	0.95	44	1
Brazil	591	0.94	1852	0.90
Canada	419	1	146	1
Chile	372	0.89	88	1
China	763	0.92	548	0.98
France	639	0.97	281	1.00
Germany	672	0.99	339	1
Greece	272	0.96	38	0.97
India	936	0.97	559	0.99
Italy	314	0.98	93	0.94
Japan	176	0.97	696	0.92
Mexico	190	0.99	1322	0.93
New Zealand	150	0.97	23	1
Poland	364	1	408	1.00
Portugal	311	1.00	114	0.86
Sweden	404	0.98	38	1
United Kingdom	1381	0.94	174	0.99
United States of America	1347	0.95	2184	1.00
Total	10002	0.96	9081	0.95

Notes: This table shows the geocoding success rates for WMS plants and WHED universities using the GeoPostcodes database. The final analysis sample is 6,406 plants (see Data Appendix for sample selection criteria). The 9,081 universities represent the population of universities in the WHED database.

Table B.3: Country-level descriptive statistics

	Management Z-score		Degree share (percent)		Distance (hours)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Argentina	-0.27	1.06	10.18	12.15	0.53	0.91
Australia	0.07	0.86	12.02	14.13	0.53	0.60
Brazil	-0.34	1.01	10.96	12.58	0.22	0.35
Canada	0.35	0.93	11.71	13.96	0.62	0.83
Chile	-0.27	0.94	14.18	13.53	0.77	0.72
China	-0.32	0.71	10.33	13.04	0.74	0.76
France	0.12	0.80	13.61	15.56	0.63	0.46
Germany	0.47	0.84	13.91	14.74	0.36	0.22
Greece	-0.41	1.24	17.56	16.27	0.46	0.47
India	-0.56	1.04	20.62	21.61	0.35	0.47
Italy	0.14	0.90	15.06	14.79	0.59	0.32
Japan	0.44	0.85	32.00	21.61	0.10	0.25
Mexico	0.00	1.05	22.53	21.38	0.17	0.20
New Zealand	-0.14	0.83	11.37	14.44	0.40	0.43
Poland	-0.04	0.95	20.20	17.83	0.32	0.31
Portugal	-0.24	0.91	9.37	9.85	0.29	0.19
Sweden	0.45	0.79	15.37	17.34	0.61	0.48
United Kingdom	0.10	0.97	12.28	15.74	0.42	0.42
United States	0.59	0.92	19.10	18.81	0.31	0.28

Notes: This table shows the country-level means and standard deviations of key variables used in the analysis.

Table B.4: Region-level descriptive statistics

	No. of regions	No. of plants in median region	Difference between 90th and 10th percentile plant in median region: Management Z-score	Degree Share	Distance
Argentina	17	3	2.19	10.32	0.10
Australia	7	45	2.06	19.41	1.67
Brazil	5	42	2.56	22.07	0.65
Canada	10	14.5	2.38	18.39	1.25
Chile	15	5	1.74	20.64	0.52
China	28	9.5	1.62	19.16	1.28
France	21	11	2.06	25.39	0.82
Germany	15	11	2.05	30.50	0.50
Greece	10	8.5	2.83	26.10	0.67
India	23	11	2.50	39.00	0.65
Italy	14	6.5	2.41	25.05	0.66
Japan	8	10.5	1.91	45.70	0.01
Mexico	21	4	2.16	35.10	0.18
New Zealand	11	4	1.65	21.64	0.83
Poland	16	12	2.44	43.95	0.74
Portugal	13	8	2.00	18.74	0.37
Sweden	19	9	1.89	32.45	0.87
United Kingdom	13	62	2.51	28.69	0.53
United States	47	9	2.12	35.50	0.48

Notes: This table shows region-level descriptive statistics by country.

Table B.5: Additional descriptive statistics

	Mean	S.D.	Min	Median	Max
World Management Survey plant level variables					
Multiunit firm	0.60	0.49	0	1	1
Listed	0.28	0.45	0	0	1
Union (percent)	39.8	39.4	0	30	100
World Higher Education Database university characteristics					
Arts dept	0.71	0.45	0	1	1
Social sciences dept	0.72	0.45	0	1	1
Business dept	0.62	0.48	0	1	1
Law dept	0.31	0.46	0	0	1
Medical dept	0.53	0.50	0	1	1
Science and tech dept	0.62	0.49	0	1	1
All main depts	0.15	0.36	0	0	1
University founding	1941.9	98.2	1088	1968	2011
Missing founding	0.054	0.23	0	0	1
University founded before plant	0.60	0.49	0	1	1
Google maps and GIS calculations					
Driving distance ('00km)	0.27	0.55	0	0.12	15.2
Straightline distance ('00km)	0.22	0.59	0	0.093	35.7
No. of universities within 100km	34.3	55.3	0	16	441
No. of universities within 50km	19.3	38.0	0	7	316
Distance to UNESCO world heritage site ('00km)	12.9	7.07	1.95	12.7	45.4
Avg pop density within 50km radius	1.68	2.51	0	0.82	20.9
Avg pop density nearest centroid	2.99	7.27	0	0.84	84.9
Plant and university share postal code	0.11	0.31	0	0	1
Distances are winsorized	0.016	0.13	0	0	1

Notes: N=6,406. *Multiunit firm* is a dummy variable indicating more than one production plant. *Listed* is a dummy for publicly listed. *Union* is percent of workforce in unions. Details of variable construction are provided in the Data Appendix.
World Higher Education Database university characteristics describe the nearest university from the plant. *Arts dept, social sciences dept, business dept, law dept, medical dept, science and tech dept* are dummy variables for whether the university has that subject department. *All depts* is a dummy for whether the university has all depts listed. *University founding* is the foundation year of the university. Missing values or funding and founding are imputed and indicated in regressions. The imputation procedures are described in the data appendix. University founded before plant is a dummy variable for whether the university was founded before the plant.
Driving distance is the google driving distance in hundreds of kilometres to the nearest university (full description in Data Appendix). Google driving calculations were based on the locations of plants and universities. *Straightline distance* is the straight line distance in hundreds of kilometres between the plant and the nearest university. *No. universities within 100km (50km)* is the number of universities within a 100km (50km) radius of the plant. *Distance to UNESCO world heritage site* is the straightline distance between the plant and the nearest site. *Avg pop density within 50km radius* is the average population density within a 50km radius of the plant calculated using GIS software. Population density data is from the Gridded Population of the World, Center for International Earth Science Information Network (CIESIN). An indicator is created if the plant and university share a postal code. An indicator is created for plants that do not have a university within 100km radius and their distances are winsorized to the region maximum.

Table B.6: Robustness checks of reduced form regressions

	Specification	Coefficient on instrument	(S.E.)
(1)	Table 2.2, panel A, column 5	-0.049***	(0.019)
A. Checking standard errors			
(2)	Cluster at region × industry	-0.049**	(0.020)
(3)	2-way cluster at region + industry	-0.049***	(0.010)
(4)	Cluster at university	-0.049***	(0.019)
(5)	Conley standard errors, 100km	-0.049***	(0.018)
B. Non-linearities in distance			
(6)	log (1 + Driving time)	-0.102***	(0.039)
(7)	Quadratic in driving time	-0.080**	(0.035)
(8)	Cubic in driving time	-0.093*	(0.055)
(9)	Quartic in driving time	-0.113	(0.072)
C. Checking geography controls			
(10)	Including quartic in geography controls (joint p-value = 3.09e-06)	-0.041**	(0.019)
(11)	Average population density within 50km (joint p-value=0.752)	-0.050***	(0.018)
(12)	Average population density nearest centroid (joint p-value=0.649)	-0.056***	(0.018)
D. Different distance measures			
(13)	Driving distance ('00km)	-0.050*	(0.028)
(14)	Straight line distance ('00km)	-0.029	(0.018)
(15)	No. of universities within 100km	0.001	(0.001)
E. Sample selection			
(16)	All survey waves, N=9,586	-0.041**	(0.018)
(17)	Exclude same postal codes, N=5,710	-0.040**	(0.019)
(18)	Exclude winsorized, N=6,302	-0.065**	(0.026)
(19)	Capital regions, N=1,884	-0.041	(0.044)
(20)	Non-capital regions, N=4,498	-0.046**	(0.021)

(continued…)

Table B.6: Robustness checks of reduced form regressions (cont.)

Specification	Coefficient on degree share	(S.E.)
F. Fixed effects		
(21) Including 2,283 region × industry fixed effects (N=6,406)	-0.056*	(0.032)
(22) Including 724 county fixed effects (N=4,553)	-0.046	(0.031)
(23) Including 851 city fixed effects (N=2,756)	-0.207	(0.296)

Notes: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at the region level (313 clusters) except otherwise stated. N=6,406 except otherwise stated. The dependent variable is the management z-score and the instrument depends on the specification. Each row presents a different robustness check of the reduced form specification (same as Table 2.2, panel A, column 5).

Chapter 3

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

Abstract. We estimate the effects of university degree class on initial labor market outcomes. We employ a regression discontinuity design which 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 six months after graduation. A First Class is worth roughly 3 percent in starting wages which translates into £1,000 per annum. An Upper Second is worth more–7 percent in starting wages which is roughly £2,040. We interpret these results as the signaling effects of degree class and provide evidence consistent with this.

3.1 Introduction

A stable empirical fact observed across countries and over time is that individuals with more schooling earn more on average. Most theories attribute these earnings differences to variation in human capital. More contentious is the link between schooling and human capital–does schooling increase skills or reflect it? The Mincer (1974) explanation is that schooling is an investment in human capital and more schooling leads to more accumulated skills. Spence (1973) provides an alternative theory where higher skilled individuals have lower costs of learning and undertake schooling to "signal" their underlying ability. In the extreme, schooling does not lead to any improvement in human capital but serves only to reveal it.

Much empirical research has tried to distinguish between the human capital and pure signaling theories. One branch of this research focuses on credential effects– if two otherwise identical individuals have the same years of schooling but differ only because one graduated with a certificate and the other did not, any earnings differences is thought to reflect the signaling effect of the certificate (Hungerford and Solon 1987). The assumption is that the two individuals have the same amount of human capital because they have the same years of education. Econometrically, however, these empirical studies rely on OLS regressions that do not isolate the pure signaling effect of the certificate as estimates may be confounded by remaining unobserved ability differences. In an ideal experiment, one would randomly assign certification to identical individuals and observe their earnings outcomes.[1]

In this paper, we estimate the effects of university degree class on labor market outcomes. As we explain below, the degree classification is a system of categorizing performance on university degree programmes in the United Kingdom (UK) and other Commonwealth nations. 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.

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 pure signaling effects we need to approximate an ideal experiment and randomly assign degree class signals across students. We adopt a fuzzy regression discontinuity design (RD) which utilizes institutional rules governing the award of degree class on the basis of marks received on courses taken. 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

[1]In a separating equilibrium in signaling theory, differences in signals reflect differences in human capital. Thus, conditioning on human capital, we should not observe differences in signals. This complicates the identification of signaling effects because in observational data, differences in signals would be consistent with signaling theory, but could also reflect selection bias. The difference between signaling and selection is that in the former, employers and econometricians both do not observe underlying ability and make inferences on the basis of observable factors (Weiss 1995). With selection bias, employers observe characteristics that are not observed by the econometrician and thus statistical estimates are biased by these omitted factors. An ideal experiment that randomizes signals across identical individuals eliminates this selection bias. In this sense, the pure signaling effect of a certificate is a causal effect.

of degree class effects.

We use survey and administrative data from the London School of Economics and Political Science (LSE). 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 3 percent in starting wages which translates into £1,000 per annum. An Upper Second is worth more– 7 percent in starting wages which is roughly £2,040. These results are robust to a battery of specification checks.

We use a simple theory of statistical discrimination to interpret these results as evidence of the signaling effects of degree class. Under this interpretation, 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. In additional results, we show that we indeed find larger effects for men and mathematical degree programmes as predicted by the simple theory.

3.1.1 Related Literature

Our paper is related to several strands of literature. Broadly, the signaling theory of education suggests that education provides a signal of unobserved worker productivity (Spence 1973). In the simplest model there is no productive role of education in human capital acquisition although this consideration does not alter the basic predictions of the theory: high ability types choose more education to separate themselves from low ability types (Riley 1979). Both the theory of human capital investments (Mincer 1974) and signaling theories predict a positive correlation between ability and education. Complementing the signaling theories are screening models where employers take actions to separate workers into ability groups (Stiglitz 1975, Wolpin 1977). Weiss (1995) describes these classes of signaling and screening theories collectively as sorting models.

Empirical testing of signaling models has proceeded in two ways. Indirect evidence comes in the form of changes in the human capital investment decisions of one ability group from changes in the decisions made in other groups. Compulsory schooling laws for primary education that affect higher education groups (Lang and Kropp 1986) or tertiary enrolment changes that affect the high school margin (Bedard 2001) are seen as consistent with the signaling value of education. More direct evidence imagines a randomized experiment where randomly selected individuals from the same ability group get treated with an educational signal. Tyler, Murnane, and Willet (2002) mimic this experiment by using differences in passing standards for the GED diploma across US states.

Their finding of significant effects for white males stands in contrast to Clark and Martorell (2010) who find no effects for receiving the high school diploma.

More recent work has examined the dynamic effects of signaling. 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.[2]

For tertiary education the early literature looked at the credential effects associated the completion of college degrees (Layard and Psacharopoulos 1974). Hungerford and Solon (1987), Belman and Heywood (1991) and Jaeger and Page (1996) include dummy variables for college completion in Mincer (1974) regressions and interpret the coefficients as signaling effects of college certificates. In papers most closely related to ours, Di Pietro (2010), Ireland, Naylor, Smith, and Telhaj (2009) and McKnight, Naylor, and Smith (2007) examine the signaling 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.

The rest of the paper is organized as follows. In Section 3.2 we discuss the institutional setting, in Section 3.3 we explore the data sources and empirical strategy, in Section 3.4 we present our results and specification checks. Section 3.5 presents a simple model of statistical discrimination and additional results to support the signaling interpretation. Section 3.6 discusses our findings and Section 3.7 concludes.

[2]Altonji and Pierret (2001) use the AFQT test score reported in the NLSY as a proxy for ability. Test scores are not available to employers but available to econometricians.

3.2 Institutional Setting

3.2.1 University Description

Our data come from the London School of Economics and Political Science (LSE). LSE is a top 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.

3.2.2 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.[3] 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.[4]

3.2.3 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 C.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

[3] 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.

[4] 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.

class applies to all departments and years.[5]

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 C.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 much more likely to obtain a First Class degree than a student whose mark just missed 70, everything else equal. This is seen clearly in Figure 3.1 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.

3.3 Data and Empirical Strategy

3.3.1 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 3.1 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.

[5]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.

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.[6] This provides two discontinuities that we examine separately and narrows our comparisons to students who are on either side of each threshold. In Table 3.1 *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. $1[\textit{4th MARK} \geq \textit{70}]$ and $1[\textit{4th MARK} \geq \textit{60}]$ 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.[7] Furthermore, our RD strategy does not rely controlling for ability.[8]

3.3.2 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.[9] 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 and includes all students including non-domiciled and non-UK nationals. Typically response rates are higher for domiciled and UK nationals.[10] 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 3.3.3). In

[6]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.

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

[8]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.

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

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

Table 3.1, "employed" is a dummy variable equal to one if a graduate is employed in full-time work. [11]

Table 3.1 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 C.2). 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 more worrying 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. Unfortunately, we have no way of addressing these issues with the current data and leave this for future work.

3.3.3 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 3.1 shows that the mean log wage is 2.45 which is roughly

[11] 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 C.2 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.

3.3.4 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 3.2.3, 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).[12] 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 3.4.2.

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

[12]Regression discontinuity was introduced by Thistlethwaite and Campbell (1960) 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.

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 3.4.4 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 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.[13] In this paper, 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:

$$\begin{aligned}(3.1)\qquad \text{CLASS}_i =&\\ &\delta_0 + \delta_1 1[\text{4th MARK} \geq \text{cutoff}]_i + \delta_2(\text{4th MARK}_i - \text{cutoff}) +\\ &\delta_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\delta_4 + u_i\end{aligned}$$

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.

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

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

$$Y_i = \beta_0 + \beta_1 \text{CLASS}_i + \beta_2(\text{4th MARK}_i - \text{cutoff}) + \beta_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\beta_4 + \epsilon_i \tag{3.2}$$

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

3.4 Results

3.4.1 First-Stage and Reduced Form Regressions

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

$$Y_i = \gamma_0 + \gamma_1 1[\text{4th MARK} \geq \text{cutoff}]_i + \gamma_2(\text{4th MARK}_i - \text{cutoff}) + \gamma_3(\text{4th MARK}_i - \text{cutoff}) \times 1[\text{4th MARK} \geq \text{cutoff}]_i + X_i\gamma_4 + \nu_i \tag{3.3}$$

where Y are the various labor market outcomes.

Table 3.2, 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).[14] 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 allows us to do a simple count of the complier population in LSE (Angrist, Imbens, and Rubin 1996, Imbens and Angrist 1994). In Figure 3.2 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.

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

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.[15]

3.4.2 Randomization Checks and McCrary Test

As discussed in Section 3.3.4, 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 3.3, 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 3.3 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 3.3 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.[16] One may still worry that students who received 58 or 68 may appeal

[15] 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 3.4 is not significant.

[16] In LSE, exams are taken anonymously and each script is graded by one internal and one external examiner. Having graded each script separately, graders convene to deliberate on the final mark.

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 3.4.4 (see for e.g. Almond and Doyle (2011), Almond, Doyle, Kowalski, and Williams (2010) and Barreca, Guldi, Lindo, and Waddell (2011)). Doing so does not change our results.

3.4.3 Effects of Degree Class on Labor Market Outcomes

Table 3.4 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 7 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 **??** 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 3.5 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 £1,000 and £2,040 per annum respectively in current money.[17]

3.4.4 Specification Checks

We conduct a battery of specification tests of our RD results. In Table 3.6 we report checks for the First Class degree while Table 3.7 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.

[17] 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: $\exp(2.418) \times 40 \times 52 \times 1.23 \times 0.071$.

3.5 Signaling Interpretation and Additional Results

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

3.5.1 Simple Model of Statistical Discrimination

Statistical discrimination is closely related to signaling and screening theories of education (Phelps 1972, Arrow 1973, Aigner and Cain 1977). 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.

3.5.2 Statistical Discrimination by Gender and Degree Programmes

The fourth highest mark is our measure of the signal from the theory described in Section 3.5.1. Appendix Table C.3 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 C.4 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 C.3, when we split degree programmes by their math requirements, mathematical degrees have higher average and variance in marks.

Table 3.8 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.[18] Upper Second effects are larger in magnitude for males but imprecisely estimated for both.

Table 3.9 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.

These results by group are consistent with our simple theory of statistical discrimination and suggestive of the signaling effects of degree class.

3.6 Discussion

The findings of positive effects of degree class and differences in effects across groups are consistent with a signaling interpretation. The signaling effect is the causal effect in an experiment where degree class is randomly assigned across individuals. We approximated this experiment using an RD design where randomness on a critical course mark effectively assigned similar students different degree classes.

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

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.

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 C.5 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.

On the other hand, heuristics by themselves cannot explain our findings. Suppose employers thought that degree class was randomly assigned independently of course marks, then they would no longer use degree class to differentiate candidates. It is the informational content in degree class coupled with the computational burden of understanding diverse transcripts that could lead employers to potential use it as an heuristic. This interaction between the signaling effects of degree class and the use of rules-of-thumb by employers is an interesting avenue for future research.

3.7 Conclusion

In this paper we estimated 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– we find that 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.

We interpret these findings using a simple theory of statistical discrimination. Under this interpretation, 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. In additional results, we show that we indeed find larger effects for men and mathematical degree programmes.

Interesting questions remain for future research. It would be interesting to know if these initial differences persist over time. If the degree class were a pure signal, its effects would attenuate over time as employers learn about workers' productivities. However, if initial labor market outcomes persist, we may observe earnings differences over the experience profile.

If employers have access to full transcripts, they should use course marks as finer signals of ability than degree class. Our findings suggest that employers do not use the full information available on transcripts and may employ the degree classification to form rules-of-thumb in hiring and salary decisions. We speculate that this could be due to the computational costs of understanding heterogenous transcripts across job applicants. We leave these issues for future research.

Table 3.1: 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 $\geq$ 70)	5912	0.24	0.41	0.00	0.25	
1(4th mark $\geq$ 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 3.2: 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 ≥ 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) *	-0.016	0.006	-0.050**	-0.011**	-0.013***	-0.013***	-0.013***	-0.005
1(4th mark ≥ 70)	(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 3.2: 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 3.3: 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 3.4: 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 3.5: 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 3.6: 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 3.6: 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 for sum of marks	0.010	0.010	0.032**	0.020	0.013	0.017	0.052**
		(0.044)	(0.073)	(0.015)	(0.015)	(0.015)	(0.013)	(0.022)
		1136	978	978	978	978	978	567
(16)	Including controls for other marks	0.011	0.021	0.034**	0.024	0.017	0.020	0.051**
		(0.045)	(0.073)	(0.015)	(0.015)	(0.015)	(0.014)	(0.023)
		1136	978	978	978	978	978	567
(17)	UK domicile sample	-0.015	0.138	0.031	0.047**	0.035*	0.039**	-0.007
		(0.063)	(0.094)	(0.025)	(0.021)	(0.020)	(0.019)	(0.040)
		701	585	585	585	585	585	367
(18)	Excluding marks around disc.	-0.002	0.008	0.048***	0.035**	0.036***	0.028**	0.078***
		(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 3.7: 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 below threshold	-0.004	0.006	0.095**	0.046	0.063	0.042	0.192***
		(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 below threshold	-0.144**	0.022	0.054	-0.017	0.008	-0.016	0.142
		(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 below threshold	-0.113*	-0.014	0.082***	0.043	0.064**	0.044	0.107**
		(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 below threshold	-0.029	0.068	0.093***	0.061**	0.075**	0.065**	0.100***
		(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 below threshold	-0.018	0.133**	0.080***	0.059**	0.072***	0.054**	0.067**
		(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 below threshold	-0.002	0.086	0.084***	0.056***	0.066***	0.052***	0.072***
		(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 below threshold	-0.030	0.114**	0.064**	0.042*	0.051**	0.038*	0.035
		(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 below threshold	-0.011	0.095*	0.057**	0.033	0.045**	0.033*	0.033
		(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 and below threshold	-0.014	0.055	0.047*	0.021	0.030	0.021	0.024
		(0.032)	(0.058)	(0.024)	(0.021)	(0.022)	(0.020)	(0.027)
		2003	1672	1672	1672	1672	1672	1092
(11)	2nd order polynomial	-0.024	0.081	0.084***	0.061***	0.076***	0.055***	0.078***
		(0.041)	(0.075)	(0.026)	(0.018)	(0.019)	(0.017)	(0.025)
		1406	1168	1168	1168	1168	1168	796
(12)	3rd order polynomial	0.006	-0.040	0.125***	0.090***	0.106***	0.080***	0.138***
		(0.053)	(0.076)	(0.033)	(0.023)	(0.026)	(0.024)	(0.028)
		1406	1168	1168	1168	1168	1168	796
(13)	4th order polynomial	-0.036	-0.113	0.121***	0.071**	0.095***	0.063*	0.158***
		(0.066)	(0.104)	(0.046)	(0.033)	(0.033)	(0.035)	(0.042)
		1406	1168	1168	1168	1168	1168	796
(14)	5th order polynomial	-0.035	-0.166	0.132***	0.069**	0.101***	0.053	0.183***
		(0.067)	(0.103)	(0.045)	(0.033)	(0.035)	(0.033)	(0.047)
		1406	1168	1168	1168	1168	1168	796

(continued…)

Table 3.7: Specification Checks for Upper Second Degree Specification (cont.)

	Employed	Finance industry	Industry mean log wages: 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.

Table 3.8: 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

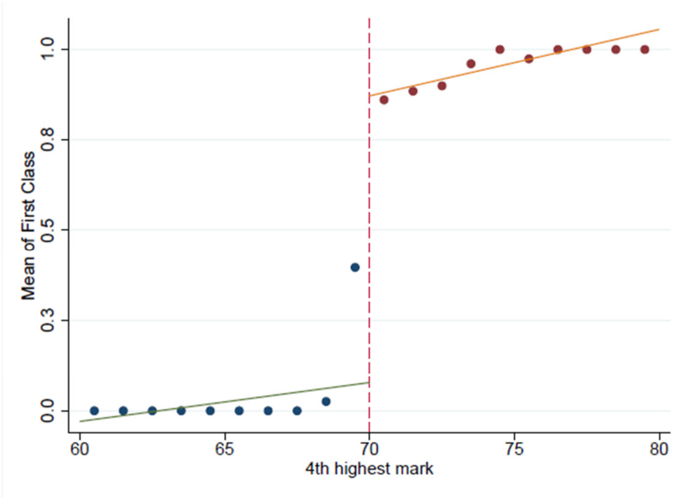
Table 3.9: RD Estimates by Programme Admissions Math Requirements

	(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

Figure 3.1: 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

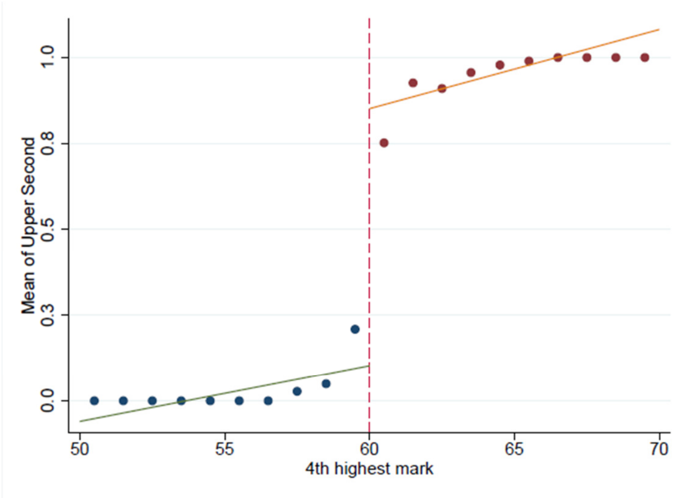


Figure 3.2: Counting Compliers

(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)

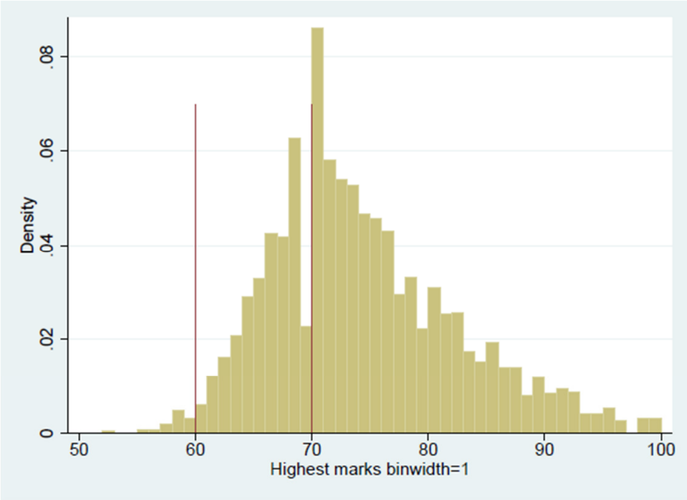
4th highest mark is above 60

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

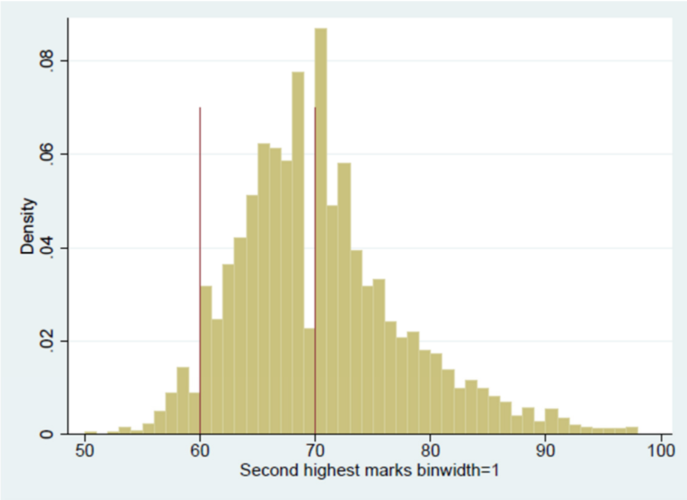
Always Takers = 5% = 16/(16+307)
Never Takers = 8% = 87/(87+996)
Compliers = 87%

Figure 3.3: Histogram of Marks

(a) Highest marks



(b) Second highest marks



(c) Third highest marks

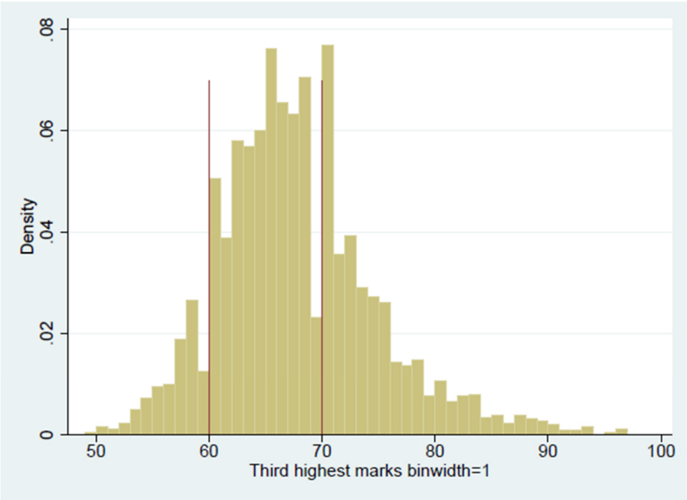
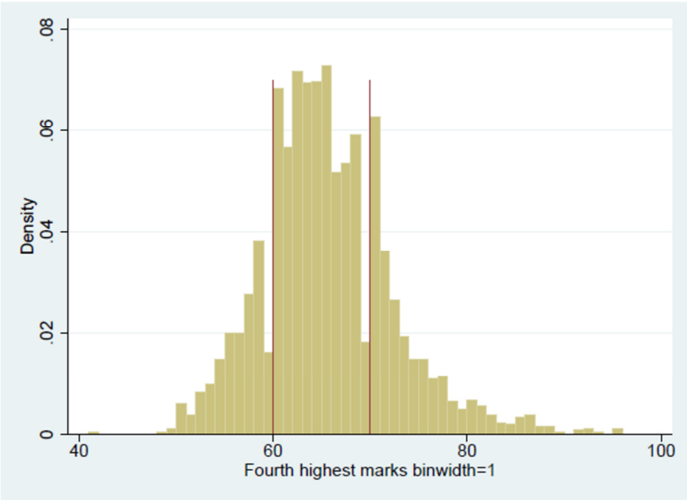
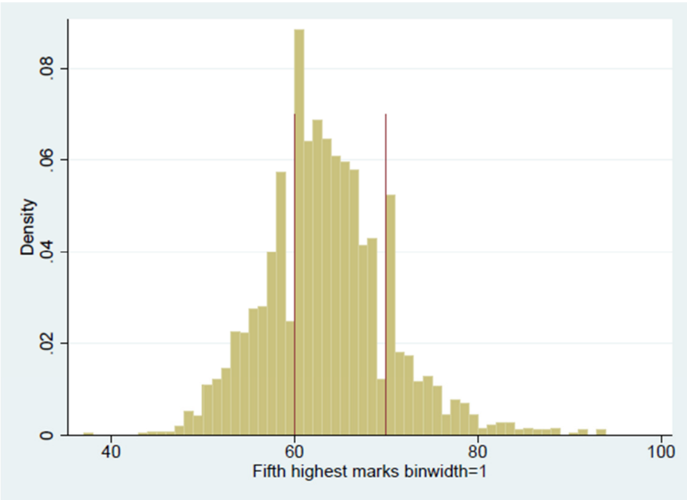


Figure 3.3: Histogram of Marks (cont.)

(d) Fourth highest marks



(e) Fifth highest marks



(f) Sixth highest marks

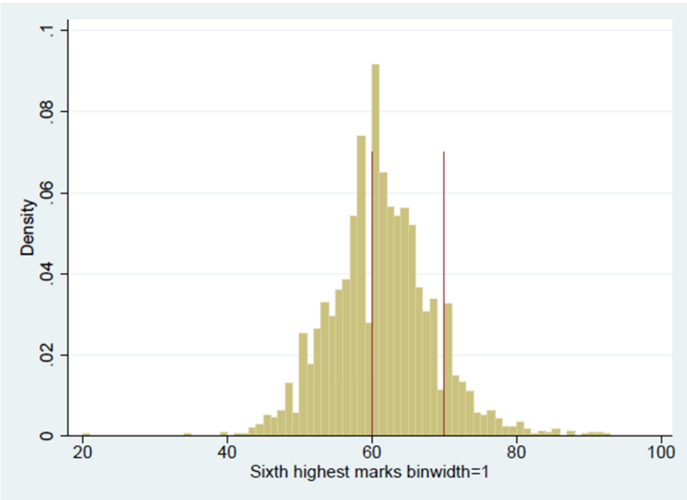
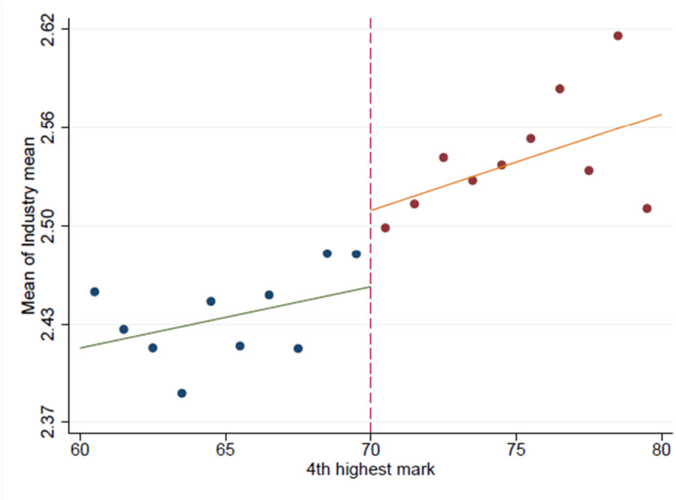
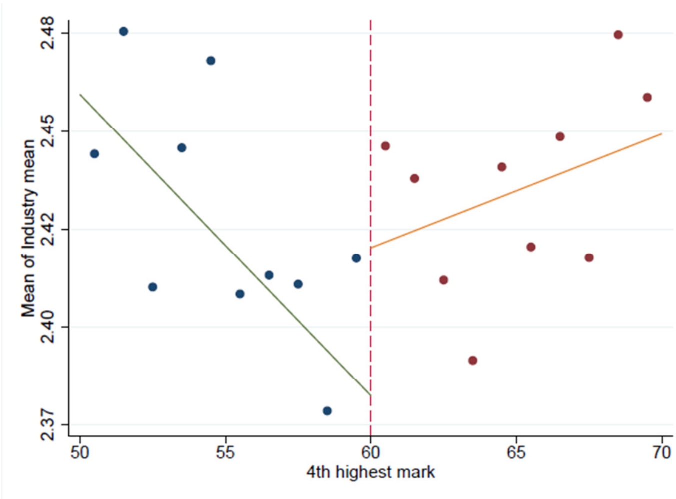


Figure 3.4: Expected Industry Mean Log Wages on Fourth Highest Marks

(a) 10 marks above and below 70



(b) 10 marks above and below 60



Appendix C

Appendix Tables

Table C.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 C.2: 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 C.3: 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 C.4: 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 C.4: 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 C.5: 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 4

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

Abstract. We study the labor market effects of increased automation. We build a model in which firms optimally design machines, train workers, and assign these factors to tasks. Borrowing concepts from computer science and robotics, the model features tasks which are difficult from an engineering perspective but easy for humans to carry out due to innate capacities for functions like vision, movement, and communication. In equilibrium, firms assign low-skill workers to such tasks. High skill workers have a comparative advantage in tasks which require much training and are difficult to automate. Workers in the middle of the skill distribution perform tasks of intermediate difficulty on both dimensions. When the cost of designing machines falls, firms adopt machines mainly in tasks that were previously performed by middle-skill workers. Occupations at both the bottom and the top of the wage distribution experience employment gains. The wage distribution becomes more dispersed near the top but compressed near the bottom. As design costs fall further, only the most skilled workers enjoy rising skill premiums, and an increasing fraction of the labor force is employed in jobs that require little or no training. The model's implications are consistent with recent evidence of job polarization and a hollowing-out of the wage distribution. In addition, the model yields novel predictions about trends in occupational training requirements that are consistent with evidence we present.

4.1 Introduction

How does labor-replacing technical change affect the allocation of workers to jobs, and what are its effects on the wage distribution? To answer these questions, we build a model guided by 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.[1] Second, there are tasks that seem 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. The two insights 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.

We model labor-replacing technical change as an exogenous fall in the cost of making machines, resulting from innovations that facilitate the automation of a wide range of tasks. Examples include the electrification of manufacturing,[2] the information and communication technology (ICT) revolution, and recent advances in robotics and artificial intelligence.[3] Responding to the fall in machine design costs, firms adopt machines in tasks that were previously performed by middle skill workers. Low skill workers' jobs might also be subject to automation, but to a lesser degree. The reallocation of workers causes occupations at both the bottom and the top of the wage distribution to experience employment gains—in short, job polarization. The wage distribution becomes more dispersed near the top but compressed near the bottom. As machine design costs drop further, only the most skilled workers enjoy rising skill premiums, and an increasing fraction of the labor force is employed in jobs that require little or no training.

We borrow from organizational economics in modeling the production process. Following Garicano (2000), we assume that production requires knowledge that must be possessed by workers or embodied in machines. The *knowledge intensity* of a task indicates the amount of knowledge required to attain a given level of productivity. The cost of building a machine capable of performing a task is determined by knowledge intensity alone. For workers however, the amount of training required may differ even across two tasks of equal knowledge intensity: in some cases people draw on innate capabilities, as when driving a car safely through traffic, but in other cases knowledge must be acquired, as when solving differential equations. The *training intensity* of a task indicates the amount of training required for a worker to perform it, holding constant the task's knowledge intensity and worker skill.

The distinction between *knowledge intensity* and *training intensity* is critical for explaining why middle skill workers are most affected by increasing automation. Skill in our model refers to the ease with which workers acquire task-specific knowledge. As workers at the bottom of the skill distribution have high learning costs, their comparative advantage is in tasks of low training intensity. These tasks may nevertheless be highly knowledge-intensive, as in the case of communication

[1] See Simon (1960, pp.23-24).

[2] Electrification facilitated automation because electric motors could be arranged much more flexibly than steam engines (Boff 1967, p.513).

[3] We provide a list of examples for recent progress in these areas in Section 4.2.

in natural language. Therefore, low skill workers face little competition from machines. High skill workers' comparative advantage is in highly training- and knowledge-intensive tasks, where automation is not impossible but too expensive. Middle skill workers perform tasks that are training intensive and of intermediate knowledge intensity. It is precisely in these tasks that a fall in machine design costs increases the incentives for automation the most, inducing firms to substitute machines for middle skill workers.

Our model features a continuum of worker types as well as a continuous task space, building on Costinot and Vogel (2010). This allows us to characterize the effects of labor-saving innovations on the entire wage distribution, and we are able to derive predictions about changes in both between- and within-group wage inequality.[4] Existing task-based models in the wage inequality literature either assume a small number of worker types and a continuum of tasks, or a continuum of types and a small number of tasks. The disadvantage of either approach is that by construction, relative wages within large sub-groups of workers are unaffected by technical change.[5] Our assumptions allow us to characterize the effects of labor-saving innovations on the entire wage distribution, and we are able to derive predictions about changes in both between- and within-group wage inequality.

At the task level, all factors are perfect substitutes. However, we can still talk about the extent to which technology complements a given skill type, because tasks are q-complements in the production of the final good.[6] The mechanism works as follows. When it gets cheaper to make machines, firms respond in two ways. 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. 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.[7] Thus, whether technology substitutes for or complements a worker of given skill type (in terms of relative wage effects) will depend on that worker's exposure to automation, which is endogenous in our model.

The model's implications are consistent with a growing empirical literature arguing that recent technical change has led to polarization of labor markets in the US and Europe.[8] Modern ICT appears to substitute for workers in middle wage jobs, while complementing labor in high and low wage jobs, thus causing the observed reallocation of employment and the hollowing-out of the wage distribution.[9] Our model provides a precise mechanism explaining these findings. In particular,

[4] In the wage inequality literature, between-group inequality refers to differences in mean wages across groups defined by observable characteristics such as education and experience. Within-group inequality refers to wage dispersion within such groups.

[5] To see this for the case of a continuum of workers and a discrete set of tasks, consider two distinct workers who are both assigned to the same task and remain so after a change in technology. The two workers' relative wage will stay constant as they both face the exact same change in the price of the task they perform.

[6] This means that the price of a task increases in the output of all other task.

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

[8] 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 European economies by Goos, Manning, and Salomons (2009).

[9] See Autor, Levy, and Murnane (2003), Michaels, Natraj, and Van Reenen (forthcoming), and Goos, Manning, and Salomons (2011) for evidence favoring the technological explanation.

the model suggests that the ICT revolution has caused job polarization because it has facilitated a more wide-ranging automation of tasks. A corollary is that job polarization should not be a unique consequence of the recent ICT revolution. Indeed, Gray (2011) finds that electrification in the US during the first half of the 20th century led to a fall in the relative demand for middle skill workers.

Our theory delivers several novel predictions about trends in occupational training requirements. In the model we distinguish between general skill and task-specific knowledge. The former refers to the ease with which a worker acquires the latter. We gauge the amount of task-specific knowledge required in an occupation using measures of training intensity from the Dictionary of Occupational Titles (DOT) and the O*NET database. This allows us to measure training requirements in the US at two points in time, 1971 and 2007.

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. Furthermore, we show that occupations that initially had intermediate training intensities 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. We also find that almost all occupations experienced an increase in mean years of schooling, irrespective of changes in training requirements. This is in line with the model's prediction about an increase in skill supply. We 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 changes in occupational wage premia are positively correlated with changes in training requirements, again consistent with the model.

The paper's main contributions may be summarized as follows. First, we present the first model of labor-saving technical change that allows firms to choose which tasks to automate, as well as featuring endogenous machine design and worker training choices. Second, to the best of our knowledge our model is the first to generate job polarization endogenously. Existing models[10] usually assume that technology substitutes for middle skill workers while complementing high and low skill ones—this is instead a result in our paper. Third, we provide comparative static results for the entire wage distribution, for instance we derive predictions about the effects of automation on wage inequality among high skill workers. Finally, we derive and test novel predictions about trends in occupational training requirements. The connection between technical change and training seems to have been neglected in the empirical literature (Handel 2000),[11] but our model suggests that the two topics are intimately linked.

The plan of the paper is as follows. The next subsection reviews related literature. Section 4.2 motivates the conceptual framework which underlies our modeling of tasks, and relates our framework to the one used by Autor, Levy, and Murnane (2003). Section 4.3 presents and solves the model. Section 4.4 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 4.5 presents two extensions to the model: endogenous capital

[10] See e.g. Autor, Levy, and Murnane (2003), Autor, Katz, and Kearney (2006), Acemoglu and Autor (2011), Autor and Dorn (2013), and Cortes (2012).

[11] Not so in the theoretical literature on wage inequality—see Section 4.1.1.

accumulation and a fixed cost of technology adoption. Section 4.6 confronts the model's prediction with existing empirical evidence and takes novel implications of the model to the data. Section 4.7 concludes. All proofs are contained in the appendix.

4.1.1 Related literature

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 our paper, 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 paper is related to a wider 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. One strand of papers analyzes the matching of workers with technologies of different vintages. Wage inequality results for instance when workers must acquire vintage-specific 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, following a recent literature that has emphasized a task-based approach to labor markets (Autor 2013). The interaction of workers and machines is nevertheless present in our model: since tasks are assumed to be q-complements, the efficiency of machines affects the marginal products of all workers in the economy.

We adopt the model of task production developed by Garicano (2000) in his theory of firm organization and knowledge hierarchies. Garicano and Rossi-Hansberg (2006) use this model to analyze how hierarchical organizations are affected by a decline in communication and knowledge acquisition costs, another consequence of the ICT revolution. Our focus is instead on labor-saving innovations, and we keep the model simple by not allowing hierarchies of multiple layers.

Finally, on the methodological side our paper is in the tradition of Ricardian theories of international trade, combining aspects of Dornbusch, Fischer, and Samuelson (1977) and Costinot and Vogel (2010).[12] 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.

[12] Acemoglu and Autor (2011) adopt the model of Dornbusch, Fischer, and Samuelson (1977) to characterize the wage effects of exogenous job polarization, assuming three distinct skill types.

4.2 Motivating the Model's Assumptions

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. Steven Pinker notes that "[the] mental abilities of a four-year-old that we take for granted—recognizing a face, lifting a pencil, walking across a room, answering a question—in fact solve some of the hardest engineering problems ever conceived" (Pinker 1994, p.192). In contrast, many abilities that humans must painstakingly acquire, such as mastery in arithmetic, are trivial from an engineering perspective. This insight 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 paradox 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. While the average human will find it somewhat challenging to divide 105 by 14 in his head, he has no trouble crossing a crowded public square on foot without constantly bumping into people. However, in terms of intrinsic difficulty the latter task is much harder than the former.[13] The reason that we are usually not aware of this fact is that we rely on innate abilities[14] for functions like movement or perception, but have no such advantage when it comes to abstract tasks like arithmetic.[15]

In our framework, a task's intrinsic difficulty is measured by its *knowledge intensity*.[16] Formally, more-knowledge-intensive tasks require a larger amount of knowledge for a given level of productivity. Solving the division exercise mentioned above is a task with low knowledge intensity, because the required procedure can easily be codified. Crossing the crowded public square, in contrast, requires a vast amount of knowledge about movement and coordination, not to mention the ability to correctly anticipate the actions of the people around.

Because machines are made of inanimate matter which is initially devoid of knowledge,[17] it is

[13] On the challenge of making walking robots, to say nothing of visual perception, Spear (2001, p.336) comments that "[in] practice this is very difficult to achieve as the leg position requires continuous sensing to ensure safe positioning and large amounts of real time computing to ensure that the robot moves without overbalancing—something the human brain achieves with ease (when sober anyway!)."

[14] "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).

[15] 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."

[16] While in reality the intrinsic difficulty of a task would have to be assessed on multiple dimensions, we adopt a one-dimensional concept for simplicity.

[17] Of course, many materials have productive properties—take for instance copper with its electrical conductivity; but

knowledge intensity alone that determines the difficulty of building a machine capable of performing a given task. However, the amount of training a human worker requires may differ even across two tasks of equal knowledge intensity. This is because she can draw on a vast endowment of knowledge providing her with certain innate capabilities, although for the most part this knowledge may be unconscious or *tacit*. The presence of such knowledge endowments (either innate or acquired early) applicable to a wide range of tasks suggests introducing a second dimension into our task framework, which we call *training intensity*: more-training-intensive tasks require more resources for equipping a human worker with a given level of knowledge specific to the task. In contrast to knowledge intensity, which refers to an objective understanding of knowledge requirements, the training intensity of a task is an attribute that only arises in the context of a worker performing a task.

Table 4.1 gives an overview of our task framework and contains examples. Here we discuss a subset of these. First, compare the task of driving a train with that of driving a car. The former takes place in a well-controlled environment, unlike the latter, which has therefore higher knowledge intensity.[18] But to humans, the two tasks may not seem all that different in terms of 'difficulty'—the uncertainties of navigating through road traffic do not pose an extraordinary challenge since many of the key functions they require, such as vision, are innate.

Second, contrast the task of grading an exam consisting of multiple choice questions (MCQs) with that of marking an essay-based test. MCQs allow only for a limited set of possible answers, and the recipe for grading them is trivial (but the task is still somewhat training intensive as it requires the ability to read and add up marks). In contrast, grading an essay may involve assessing a large variety of approaches to the questions posed. Clearly, the latter is more knowledge-intensive than the former. But in this case, it is also more training-intensive: most humans will find grading an essay the more difficult task, perhaps even impossible to complete in the absence of subject-specific training. Driverless trains and machine-grading of MCQs have been around much longer than driverless cars and automatic grading of essays, both appearing only recently (Markoff 2010, Shermis and Hamner 2012). We will show the model to be consistent with this fact.

We are not the first to employ a multi-dimensional task space to analyze the impact of technical change on jobs and wages. In particular, Autor, Levy, and Murnane (2003, henceforth ALM) categorize tasks as routine and non-routine on one dimension, and as analytic, interactive and manual on another. 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.). The terms analytic, interactive and manual are used to characterize both routine and non-routine tasks in more detail.

While ALM's framework addresses many of the issues that we have discussed here, we believe that our own framework offers several advantages. First, it is more general, as it avoids specific attributes such as interactive and manual. Second, it is not context-dependent. Machine capabilities

---

the 'knowledge' contained in materials is usually highly specific and limited.

[18] We consider only the process of driving the train, not the engineering knowledge and familiarity with railway infrastructure that train drivers posses in practice.

Table 4.1: Two-Dimensional Task Framework, Examples

		Knowledge intensity	
		−	+
Training intensity	−	assembly driving a train	driving a car language waiting tables
	+	arithmetic bookkeeping grading MCQs	grading essays research strategic decision making

constantly expand, so we prefer to avoid a task construct that depends on the current state of technology.[19] Thus, knowledge intensity is an objective, time-invariant measure of the information required to do a particular task, irrespective of whether a machine or a human does it. Third, the concept of training intensity is absent in ALM. 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.

Notwithstanding these differences, it is still possible to interpret ALM's empirical results in light of our framework. For instance, their measure of routine-ness might in practice be inversely related to knowledge intensity. We will return to this issue when discussing how our model matches up to empirical findings in Section 4.6.

While we believe that our task framework is an improvement over existing literature and that it generates useful and novel insights, 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 knowledge-intensive by moving it to a more controlled environment.[20] Our framework does not explicitly allow for this 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 knowledge intensity. 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. There are three reasons. First, comparative advantage ensures that some tasks will always be performed by humans, so that the model will be consistent with the fact that some

[19]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.

[20]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).

tasks are not performed by machines in reality. Second, we can parameterize the model such that machine productivity levels in some tasks are vanishingly small. Third, and most 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.[21] 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.

4.3 The Model

4.3.1 Overview

The model has one period that we interpret as a worker's lifetime.[22] 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. Factors' productivity is not a given: intermediate firms must train workers, 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 levels and machine quality are choices faced by the intermediate firms which, unlike the decision of what factor to hire, are made independently of factor prices and task prices. This is because training and design costs are assumed to be in units of factor inputs and not in units of the final good. Optimal training and design choices, and hence productivity, result instead from the properties of tasks and their interplay with attributes of the factors of production. Characterizing these choices is subject of the Section 4.3.5. The result is a productivity schedule that determines comparative advantage between factors and across tasks. This then allows us to apply standard results to solve for the equilibrium assignment of factors to tasks in Section 4.3.6. Thus, we proceed by a kind of 'backward induction': first, we solve for factors' productivity conditional on firms' hiring these factors; and second, we characterize hiring choices, using the results of the first step.

[21] An overview of recent developments in robotics research can be found in Nourbakhsh (2013).
[22] We discuss a dynamic (multi-period) version of the model in Section 4.5.1.

4.3.2 The Task Space

Tasks differ along two dimensions, knowledge intensity, denoted by $\sigma \in \Sigma$, and training intensity, denoted by $\tau \in T$. The higher is a task's σ, the more knowledge is required for a *worker* or a *machine* to attain a given level of productivity. The higher is a task's τ, the more resources are required to equip a *worker* with a given level of knowledge. Recall that the concept of knowledge intensity refers to an objective understanding of knowledge requirements, for instance, the amount of information processing required to perform a given task. In contrast, the training intensity of a task is an attribute that only arises in the context of a worker performing a task.

Completion of tasks results in intermediate outputs that are used to produce the final good. Let Y denote the output of the unique final good, and let task output be denoted by $y(\sigma, \tau)$. For tractability, we use a Cobb-Douglas production function,

$$\log Y = \int_{\Sigma \times T} [\log y(\sigma, \tau)] \, dB(\sigma, \tau).$$

The weighting function $B(\sigma, \tau)$ determines the relative importance of each task in final good production. To ensure constant returns to scale we assume $\int_{\Sigma \times T} dB(\sigma, \tau) = 1$.

Throughout most of our analysis we make the following, simplifying assumption about the domains of the parameters τ and σ.

Assumption 1 $\quad \tau \in T = \{0, 1\}, \quad \sigma \in \Sigma = [\underline{\sigma}, \overline{\sigma}], \underline{\sigma} > 0$

Under this assumption, there is a set of tasks for which $\tau = 0$, so that knowledge acquisition costs are zero, or equivalently, all workers have an innate ability to perform these tasks. We will call these tasks 'innate ability tasks'. We will refer to the tasks with $\tau = 1$ as 'training-intensive tasks'. Within both these sets of tasks, knowledge intensity varies continuously. We will state explicitly when Assumption 1 is imposed.

4.3.3 Worker Training, Machine Design, and Technical Change

The technologies for training workers and designing machines are as follows. Intermediate firms must pay τ/s efficiency units of labor to equip a worker of skill s with a unit measure of knowledge. Higher skilled workers have lower learning costs. Higher values of τ imply a larger learning cost, holding knowledge and skill constant.

Similarly, to transform one unit of capital into a machine equipped with a unit measure of knowledge, intermediate firms must pay $c_K \equiv 1/s_K$ units of capital. We will refer to c_K as the machine design cost, which is the main exogenous driving force in our model. As a matter of notation, it will be more convenient to work with s_K, 'machine skill', instead of c_K. Notice that a tasks's τ does not affect design costs, by definition.

Workers' and machines' productivity depends on their task-specific knowledge as well as a task-neutral productivity term, which shifts a factor's productivity proportionately in all tasks. Let task-neutral productivity of machines be denoted by A_K.

Our model admits exogenous technical change in the form of a decrease in c_K or an increase in A_K, although we will mainly be concerned with the former. A fall in c_K represents any technological advance that lowers the cost of automation of a wide range of tasks, typically a combination of improved software (programming languages, algorithms) and improved hardware (CPU speed, robotics). A rise in A_K represents improved efficiency of existing machinery. In reality, the forces affecting the two parameters may not always be mutually exclusive. This does not impair the model's ability to generate sharp predictions, however, since both parameters give rise to the same comparative statics.

4.3.4 A Simple Example

To illustrate how task characteristics and factor attributes affect productivity differences across factors and tasks, we present a simple example. We impose Assumption 1. Let us assume for the moment that worker training and machine design are exogenously determined by task characteristics. In particular, suppose that factors are either made capable of performing a task or not, so that there is no intensive margin for task-specific productivity. Let knowledge intensity σ be the amount of knowledge required for a factor to be able to perform a given task. A worker with learning cost $1/s$ will produce $A(1-\sigma/s)$ units of task output in training-intensive tasks ($\tau = 1$), where A is the worker's task-neutral productivity. The same worker will produce A units in any innate ability task ($\tau = 0$). A machine will produce $A_K(1-\sigma/s_K)$ units regardless of training intensity.[23]

Now consider two workers with skill levels s, s' such that $s' > s$, and two tasks with equal training intensity $\tau = 1$ but different knowledge intensities σ, σ' such that $\sigma' > \sigma$. (How task-neutral productivities A and A' compare is irrelevant for what follows.) Simple algebra establishes that the higher skilled worker is relatively more productive in task σ', i.e. she has a comparative advantage in the more knowledge-intensive task. Machines' comparative advantage will depend on the level of design costs $c_K \equiv 1/s_K$. For instance, if $s_K < s$, then the machine has a comparative advantage over both workers in the less knowledge-intensive task.

Next, take an innate ability task and a training-intensive task both with equal knowledge intensity σ. Machines are equally productive in both tasks but workers are more productive in the innate ability task. Therefore, machines have a comparative advantage in the training-intensive tasks. This is why some training-intensive tasks will always be performed by machines, even if machine design cost exceed the training cost of the least-skilled worker.

Finally, consider again two workers with skill levels s, s' such that $s' > s$ and take an innate ability task and a training-intensive task both with equal knowledge intensity σ. Because the higher-skilled worker has a higher *task-specific* productivity in the training-intensive task, she has a comparative advantage in that task. This is why workers at the bottom of the skill distribution will generally perform innate ability tasks, and why middle skill workers will compete with machines in training-intensive tasks of intermediate knowledge intensity.

The simple example illustrates the main forces driving our results about the effects of increased automation on job assignment and the wage distribution. In fact, the simple model presented here

[23] We assume parameter values are such that factor productivity is always strictly positive.

generates an equilibrium assignment and comparative static results that are qualitatively the same as in the model with endogenous worker training and machine design. However, the simple model does not explicitly describe the production process, so that it is not clear what precisely drives the results. Moreover, it does not allow us to assess if the results are robust to allowing firms a productivity choice (via worker training and machine design). We address these limitations in the following section.

4.3.5 The Production Process for Tasks and Firms' Productivity Choices

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 4.3.3 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 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 knowledge intensity σ is given by the cumulative density 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 examples discussed in Section 4.3.2, driving a car and grading an essay are more knowledge-intensive (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. Let $\varepsilon_{F,\sigma}(Z, \sigma)$ denote the elasticity of F with respect to σ holding Z constant, and similarly for $\varepsilon_{f,\sigma}(Z, \sigma)$. We impose the following condition on the family of distributions $F(Z; \sigma)$.

Assumption 2 $\quad \varepsilon_{F,\sigma}(Z, \sigma) < \varepsilon_{f,\sigma}(Z, \sigma) \quad$ *for all* $Z, \sigma > 0$

This assumption will give rise to a set of intuitive comparative advantage properties, for instance high skill workers will have a comparative advantage in knowledge-intensive tasks. One of the distributions satisfying Assumption 2 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) knowledge intensity.

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. Assume that each worker is endowed with one efficiency unit of labor. After incurring learning costs, $1 - \tau z/s$ 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 (σ, τ) be denoted by $\alpha^N(s,\sigma,\tau)$, 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

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

A unique interior solution to the worker training problem exists provided $\tau > 0$, while the machine design problem always admits a unique interior solution.[24] The optimal knowledge levels $z^N(s,\sigma,\tau)$ and $z^K(s_K,\sigma)$ are pinned down by the first-order conditions

$$\begin{aligned}f(z(s,\sigma,\tau);\sigma)\left[1 - \tfrac{\tau}{s} z(s,\sigma,\tau)\right] &= \tfrac{\tau}{s} F(z(s,\sigma,\tau);\sigma),\\ f(z(s_K,\sigma);\sigma)\left[1 - \tfrac{1}{s_K} z(s_K,\sigma)\right] &= \tfrac{1}{s_K} F(z(s_K,\sigma);\sigma).\end{aligned} \tag{4.1}$$

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.

We will formalize the concept of innateness by assuming that some tasks feature $\tau = 0$. It is immediate that in such innate ability tasks, $\alpha^N(s,\sigma,0) = 1$. Thus, optimal worker and machine productivities are given by

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

and

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

We impose Assumption 1 for the remainder of the paper. Let the set of worker skills be given

[24] 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.

by $S = [\underline{s}, \overline{s}]$ and let $\breve{s}$ be an element in set $\breve{S} = s_K \cup S$. By the above equations, 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

$$\alpha(\breve{s}, \sigma) = F(z(\breve{s}, \sigma); \sigma)\left[1 - \frac{1}{\breve{s}} z(\breve{s}, \sigma)\right] \qquad \breve{s} \in \breve{S} = s_K \cup [\underline{s}, \overline{s}], \tag{4.2}$$

where $z(\breve{s}, \sigma)$ is implicitly given by (4.1) when $\tau = 1$.

We now turn to the properties of the productivity schedule $\alpha(\breve{s}, \sigma)$. First notice that $\alpha \in (0, 1)$ by (4.2). Furthermore, from applying the envelope theorem to (4.2) it follows that α is increasing in $\breve{s}$ and decreasing in σ. Higher skilled factors are more productive since they face a lower learning/design cost, and productivity declines in knowledge intensity since a larger cost is incurred to achieve a given level of productivity. To characterize comparative advantage, we rely on the following result.

Lemma 1 *The productivity schedule $\alpha(\breve{s}, \sigma)$ is strictly log-supermodular if Assumption 2 holds.*

The log-supermodularity of the productivity schedule implies that in training-intensive tasks, factors with higher skill have a comparative advantage in more knowledge-intensive 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 knowledge-intensive tasks; all workers with $s > s_K$ have a comparative advantage over machines in more knowledge-intensive tasks; and so on. As the proof of Lemma 1 establishes, these comparative advantage properties hold if and only if optimal knowledge $z(\breve{s}, \sigma)$ is increasing in σ. Thus, high skill factors have a comparative advantage in more knowledge-intensive tasks because these tasks induce a higher level of knowledge, and to high skill factors this comes at a lower cost.

The effect of σ on the optimal knowledge level is in principle ambiguous. A higher σ implies a lower opportunity cost of learning an additional problem since factors are less productive, ceteris paribus. However, the marginal benefit may increase or decrease depending on the problem distribution. Assumption 2 ensures that the fall in marginal costs outweighs any effect on the marginal benefit.

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 knowledge-intensity 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 knowledge intensity. 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 4.4 below.

4.3.6 Competitive Equilibrium

To complete the setup of the model, 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, both with support $S = [\underline{s}, \overline{s}]$. Let the share of innate ability tasks ($\tau = 0$) in final good production be β. The production function can now be written as

$$\log Y = \frac{1}{\mu} \int_{\underline{\sigma}}^{\overline{\sigma}} \{\beta \log y_0(\sigma) + (1-\beta) \log y_1(\sigma)\} \, d\sigma, \tag{4.3}$$

where the term $\mu \equiv \overline{\sigma} - \underline{\sigma}$ ensures constant returns to scale. The subscripts 0 and 1 indicate innate ability ($\tau = 0$) and training-intensive ($\tau = 1$) tasks, respectively.

We have established in Section 4.3.5 that in innate ability tasks, machine productivity is given by $\alpha(s_K, \sigma)$, while worker productivity equals one. Hence, output of the innate ability task with knowledge intensity σ is given by

$$y_0(\sigma) = A_K \alpha(s_K, \sigma) k_0(\sigma) + \int_{\underline{s}}^{\overline{s}} n_0(s, \sigma) d\sigma, \tag{4.4}$$

where $k_0(\sigma)$ and $n_0(c, \sigma)$ are the masses of machine capital and of worker type s, respectively, allocated to innate ability task σ. In training-intensive tasks, as we have seen, both machine and worker productivity depends on the function $\alpha(\breve{s}, \sigma)$. Hence we can write task output of the training-intensive task σ as

$$y_1(\sigma) = A_K \alpha(s_K, \sigma) k_1(\sigma) + \int_{\underline{s}}^{\overline{s}} \alpha(s, \sigma) n_1(s, \sigma) d\sigma. \tag{4.5}$$

There is a large number of perfectly competitive firms producing the final good, and buying task output from perfectly competitive intermediates producers. We normalize the price of the final good to one and denote the price of task σ in 'sector' $\tau \in \{0, 1\}$ 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 in sector j and with knowledge intensity σ 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. Recall that design and learning costs are already included in the $\alpha(\breve{s}, \sigma)$ terms which enter intermediate producer's profits through the task production functions (4.4) and (4.5).

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

$$y_0(\sigma) = \frac{\beta}{\mu}\frac{Y}{p_0(\sigma)}, \qquad y_1(\sigma) = \frac{1-\beta}{\mu}\frac{Y}{p_1(\sigma)}. \tag{4.6}$$

Profit maximization by intermediates producers implies

$$\begin{array}{rcll} p_0(\sigma) & \le & w(s) & \forall s \in [\underline{s}, \overline{s}], \\ p_1(\sigma)\alpha(s,\sigma) & \le & w(s) & \forall s \in [\underline{s}, \overline{s}], \\ p_\tau(\sigma)\alpha(s_K,\sigma) & \le & r/A_K & \forall \tau \in \{0,1\}; \\ p_0(\sigma) & = & w(s) & \text{if } n_0(s,\sigma) > 0, \\ p_1(\sigma)\alpha(s,\sigma) & = & w(s) & \text{if } n_1(s,\sigma) > 0, \\ p_\tau(\sigma)\alpha(s_K,\sigma) & = & r/A_K & \text{if } k_\tau(\sigma) > 0. \end{array} \tag{4.7}$$

Factor market clearing conditions are

$$v(s) = \sum_\tau \int_{\underline{\sigma}}^{\overline{\sigma}} n_\tau(s,\sigma)d\sigma \quad \text{for all } s \in [\underline{s}, \overline{s}] \tag{4.8}$$

and

$$K = \sum_\tau \int_{\underline{\sigma}}^{\overline{\sigma}} k_\tau(\sigma)d\sigma. \tag{4.9}$$

A *competitive equilibrium* in this economy is a set of functions $y : \Sigma \times T \to \mathbb{R}^+$ (task output); $k : \Sigma \times T \to \mathbb{R}^+$ and $n : S \times \Sigma \times T \to \mathbb{R}^+$ (factor assignment); $p : \Sigma \times T \to \mathbb{R}^+$ (task prices); $w : S \to \mathbb{R}^+$ (wages); and a real number r (rental rate of capital) such that conditions (4.1), (4.2), and (4.4) to (4.9) hold.

The equilibrium assignment of factors to tasks is determined by comparative advantage, which is a consequence of the zero-profit condition (4.7).[25] Because high skill workers have a comparative advantage in training-intensive tasks (holding knowledge intensity 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.[26] In this case, machines are assigned to a subset of innate

---

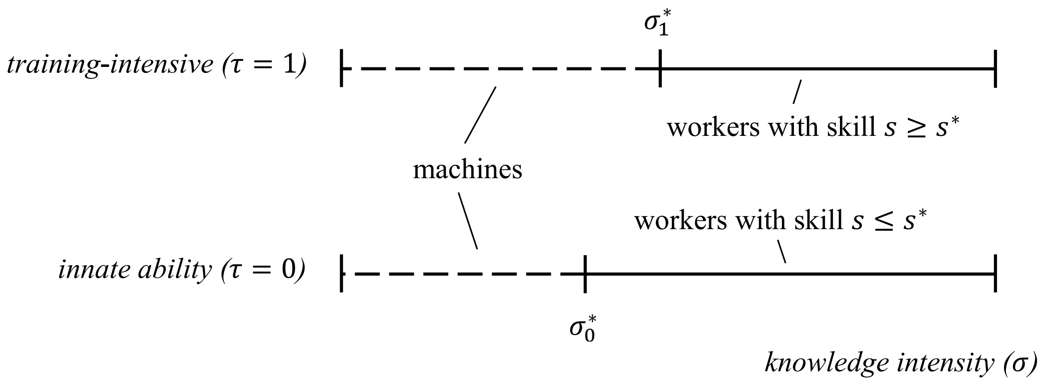
[25]To see how comparative advantage determines patterns of specialization, consider two firms, one producing training-intensive task σ, the other producing training-intensive task σ'. Suppose in equilibrium, firm σ is matched with workers of type s and firm σ' is matched with workers of type s'. Then (4.7) implies

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

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.

[26]Sufficient conditions for the existence of such an equilibrium are derived Appendix D.1.1. We assume throughout

Figure 4.1: Assignment of labor and capital to tasks.



ability and training-intensive tasks that are relatively less knowledge-intensive, 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 4.3.5. These properties also imply $\sigma_0^* < \sigma_1^*$: the marginal training-intensive task is always more knowledge-intensive than the marginal innate ability task (recall that machines are relatively more productive in training-intensive tasks than workers, holding knowledge intensity 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 4.1.

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 $\sigma_0^*, \sigma_1^* \in \Sigma$ 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*
- $n_1(s,\sigma) > 0$ *if and only if $s \geq s^*$ and $\sigma \geq \sigma_1^*$.*

---

that these conditions are satisfied. We note however that in general, no innate ability tasks may be performed by machines, and/or no training-intensive tasks may be performed by workers.

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 knowledge-intensive tasks. Formally, we have:

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

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

$$M'(s) = \frac{\mu}{1-\beta} \frac{w(s)v(s)}{Y}, \tag{4.10}$$

and that the wage schedule is given by

$$\frac{d \log w(s)}{ds} = \frac{\partial \log \alpha(s, M(s))}{\partial s}. \tag{4.11}$$

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, (4.10) can easily be derived from the market clearing condition (4.8) given Lemma 2, and using (4.6) and (4.7). In particular, Lemma 2 and (4.8) 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 (4.5) we obtain

$$M'(s) = \frac{\alpha(s, M(s))v(s)}{y(M(s))}. \tag{4.12}$$

After eliminating task output and price using (4.6) and (4.7), (4.10) follows. Figure 4.2 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 (4.7) 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

$$\frac{r}{A_K\alpha(s_K,\sigma_0^*)} = w(s) \quad \textit{for all } s \leq s^* \tag{4.13}$$

and

$$\frac{r}{A_K\alpha(s_K,\sigma_1^*)} = \frac{w(s^*)}{\alpha(s^*,\sigma_1^*)}, \tag{4.14}$$

and the no-arbitrage condition for the marginal worker is

$$w(s) = w(s^*) \quad \textit{for all } s \leq s^*. \tag{4.15}$$

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 (4.14). 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 tasks (but not across the two sectors unless $\beta = 0.5$). Some algebra shows[27] that machines produce

[27] By (4.6) and (4.7), we have

$$\frac{y_\tau(\sigma)}{y_\tau(\sigma')} = \frac{\alpha(s_K,\sigma)}{\alpha(s_K,\sigma')}, \qquad \frac{y_0(\widetilde{\sigma})}{y_1(\widetilde{\sigma}')} = \frac{\beta}{1-\beta}\frac{\alpha(s_K,\widetilde{\sigma})}{\alpha(s_K,\widetilde{\sigma}')}$$

for any tasks $(\sigma,\sigma',\widetilde{\sigma},\widetilde{\sigma}')$ performed by machines. But (4.4), (4.5), 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 $k_0(\widetilde{\sigma}) = \frac{\beta}{1-\beta}k_1(\widetilde{\sigma}')$. By (4.9) and Lemma 2,

$$k_0(\sigma) = \frac{\beta K}{\beta(\sigma_0^*-\underline{\sigma}) + (1-\beta)(\sigma_1^*-\underline{\sigma})} \quad \text{for all } \sigma \in [\underline{\sigma},\sigma_0^*]$$

and

$$k_1(\sigma) = \frac{(1-\beta)K}{\beta(\sigma_0^*-\underline{\sigma}) + (1-\beta)(\sigma_1^*-\underline{\sigma})} \quad \text{for all } \sigma \in [\underline{\sigma},\sigma_1^*].$$

task outputs

$$(4.16)\qquad \begin{aligned} y_0(\sigma) &= \frac{\beta A_K \alpha(s_K,\sigma) K}{\beta(\sigma_0^* - \underline{\sigma}) + (1-\beta)(\sigma_1^* - \underline{\sigma})} \quad \text{for all } \sigma \in [\underline{\sigma}, \sigma_0^*], \\ y_1(\sigma) &= \frac{(1-\beta) A_K \alpha(s_K,\sigma) K}{\beta(\sigma_0^* - \underline{\sigma}) + (1-\beta)(\sigma_1^* - \underline{\sigma})} \quad \text{for all } \sigma \in [\underline{\sigma}, \sigma_0^*]. \end{aligned}$$

Using these equations to solve for the task prices in (4.6), and plugging the obtained expression into (4.7), yields

$$(4.17)\qquad r = \frac{\beta(\sigma_0^* - \underline{\sigma}) + (1-\beta)(\sigma_1^* - \underline{\sigma})}{\mu} \times \frac{Y}{K}.$$

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 (4.4) and (4.8),[28]

$$(4.18)\qquad y_0(\sigma) = \frac{V(s^*)}{\overline{\sigma} - \sigma_0^*} \quad \text{for all } \sigma \geq \sigma_0^*.$$

Proceeding as above when solving for r, we obtain

$$(4.19)\qquad w(s^*) = \frac{\beta(\overline{\sigma} - \sigma_0^*)}{\mu} \times \frac{Y}{V(s^*)}.$$

With (4.17) and (4.19) in hand, we can eliminate factor prices from the marginal cost equalization condition (4.13) to obtain

$$(4.20)\qquad \frac{A_K \alpha(s_K, \sigma_0^*) K}{\beta(\sigma_0^* - \underline{\sigma}) + (1-\beta)(\sigma_1^* - \underline{\sigma})} = \frac{V(s^*)}{\beta(\overline{\sigma} - \sigma_0^*)}.$$

Also, combining conditions (4.13) to (4.15) yields

$$(4.21)\qquad \alpha(s_K, \sigma_1^*) = \alpha(s_K, \sigma_0^*)\alpha(s^*, \sigma_1^*).$$

[28] Under Lemma 2, integrating (4.8) yields

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

but using (4.4) and the fact that task output is a constant y_0 results in

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

Lastly, (4.10) and (4.19) imply

$$M'(s^*) = \frac{\beta(\overline{\sigma} - \sigma_0^*)}{1-\beta} \frac{v(s^*)}{V(s^*)}. \tag{4.22}$$

Equations (4.3), (4.10), (4.11), (4.20), (4.21), and (4.22) 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 (4.11) yields an expression for the wage differential between any two skill types that are both employed in training-intensive tasks,

$$\frac{w(s')}{w(s)} = \exp\left[\int_s^{s'} \frac{\partial}{\partial z} \log \alpha(z, M(z)) dz\right] \quad \text{for all } s' \geq s \geq s^*. \tag{4.23}$$

This shows that wage inequality is fully characterized by the matching function (Sampson 2012). Second, adding (4.10) and (4.19) and integrating yields an expression for the average wage,

$$Ew = \frac{\beta(\overline{\sigma} - \sigma_0^*) + (1-\beta)(\overline{\sigma} - \sigma_1^*)}{\mu} \times Y. \tag{4.24}$$

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.

4.4 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.

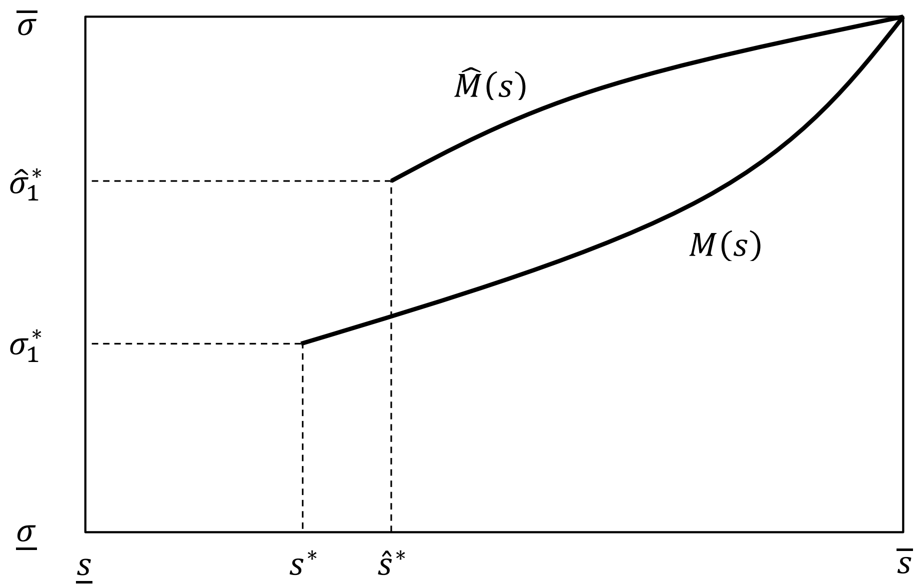
4.4.1 Technical Change

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 now state the main result of the paper.

Proposition 1 *Suppose $\widehat{c}_K < c_K$ and so $\widehat{s}_K > s_K$. Then $\widehat{\sigma}_1^* > \sigma_1^*$ and $\widehat{M}(s) > M(s)$ for all $s \in [\max\{s^*, \widehat{s}^*\}, \overline{s})$. If $\widehat{s}_K \geq s^*$, then $\widehat{s}^* > s^*$.*

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

Figure 4.2: Assignment of workers to training-intensive tasks and the effects of technical change



Notes: Knowledge intensity 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.

threshold innate ability task, since $\sigma_0^* < \sigma_1^*$.[29] 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.

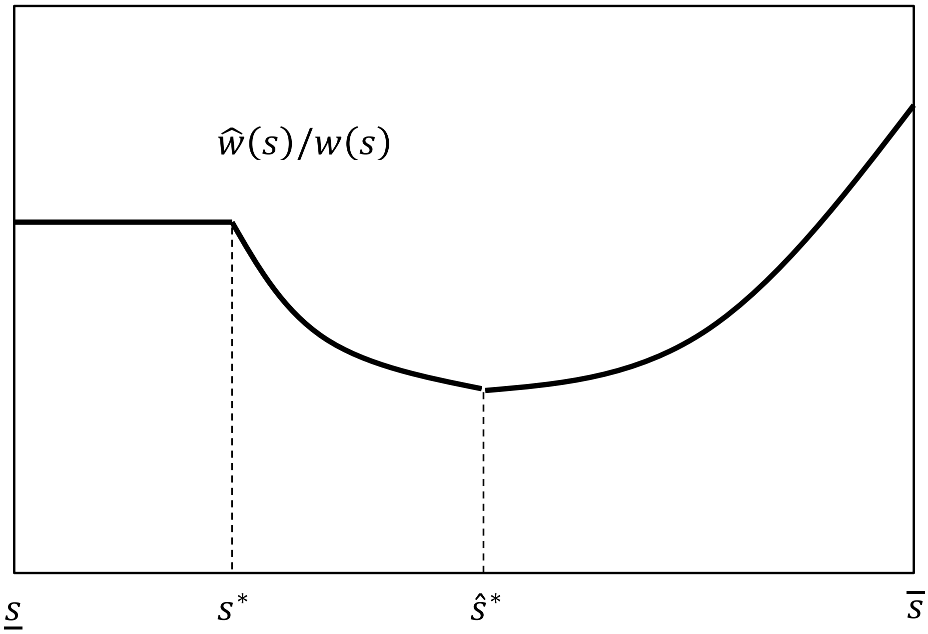
Proposition 1 says that a sufficiently large fall in the machine design cost leads to the marginal worker becoming more skilled, $\widehat{s}^* > s^*$. We are unable to rule out $\widehat{s}^* \leq s^*$ for small decreases in the machine design cost. However, if machine design costs fall steadily over time, then the skill cutoff level must rise eventually. Thus, we limit our attention to the case where a fall in c_K triggers a rise in s^*. This implies a reassignment of some workers to innate ability tasks. Importantly, workers remaining in training-intensive also experience displacement, as they are reassigned to tasks of higher knowledge intensity due to the upward shift of the matching function.[30] In sum, employment in tasks previously performed by low skill workers $s \leq s^*$ increases; employment in tasks previously carried out by middle skill workers $s \in (s^*, \widehat{s}^*)$ decreases; and employment in tasks formerly performed by high skill workers $s \geq \widehat{s}^*$ increases. Thus, a fall in the machine design cost causes job polarization. These effects are illustrated by Figure 4.2.

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. Intuitively, since 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 they would have worked for their new employers even before the

[29] 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.

[30] This will always be the case regardless of the magnitude of the decrease in the design cost.

Figure 4.3: Changes in wages as a result of a fall in the machine design cost from c_K to $\widehat{c}_K$



Notes: 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.

shift. 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. Thus, wage inequality rises at the top, but falls at the bottom of the distribution. This is illustrated by Figure 4.3. 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,*

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

and

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

Although the effect on the marginal innate ability task is uncertain, the overall weighted share of tasks performed by machines increases. By (4.24), this is equivalent to a decrease in the labor share.

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

4.4.2 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

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

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 $\widehat{V} \succeq V$ 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 $\widehat{\sigma}_1^* < \sigma_1^*$, $\widehat{s}^* > s^*$ and $\widehat{M}(s) < M(s)$ 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 in our model 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.[31]

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.

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

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

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

[31] 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.

4.5 Extensions

4.5.1 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[32] 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 (4.17), 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.[33] This 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.

4.5.2 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.[34] 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 D.2 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 (knowledge intensity) 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 4.1). 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

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

[33] 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.

[34] 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.

sufficiently in knowledge intensity. 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-knowledge-intensive tasks. This incentive is stronger in training-intensive tasks: as knowledge-intensity 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 are currently working on a numerical solution to verify the intuition.

4.6 Empirical Support for the Model's Predictions

Section 4.4.1 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.

4.6.1 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 knowledge intensity, 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 knowledge intensity (though not necessarily low training intensity). The measured shift away from routine tasks is then consistent with our prediction of a reallocation towards more-knowledge-intensive 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 (2009) 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 high training- and knowledge-intensive tasks.

Wages.—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 assigned to training-intensive tasks.

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 and Dorn (2013) document that in the US over the past three decades, 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 along the lines of our model.

Lemieux (2006) 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. Similarly, Angrist, Chernozhukov, and Fernndez-Val (2006) document a more pronounced rise in within-group inequality for college graduates than for high school graduates, and an increase in the effect of an additional year of schooling on the upper tail of the conditional wage distribution, relative to the effect on lower tail and median. Thus, the evidence on within- and between-group inequality appears consistent with our model.

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.

Historical evidence.—Gray (2011) shows that electrification in the US during the first half of the 20th century led to a fall in relative demand for tasks performed by middle skill workers, providing support for the model's prediction that job polarization is not a unique consequence of the IT 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-knowledge-intensive, 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.

4.6.2 Trends in Occupational Training Requirements

In the model, training levels (knowledge) vary systematically with task characteristics. In particular, tasks with higher knowledge intensity require more training in equilibrium, provided $\tau > 0$. And holding knowledge intensity constant, tasks with lower training intensity induce a lower training investment. In the extreme case of our innate ability tasks, the training investment is zero.

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 knowledge-intensive, 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.[35] 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

[35] 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 D.3 for details. Tables D.2 and D.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 knowledge 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 knowledge intensity on the one hand (the upward shift of the matching function) and towards innate ability tasks on the other, the model predicts a polarization of job training requirements. Figure 4.4 plots fitted values from a locally weighted regression of changes in an occupation's employment share on its percentile rank in the 1971 distribution of occupational mean wages.[36] The pattern is consistent with the model's prediction of polarization of training requirements.

Second, the model can potentially help to make sense of changes in training requirements within occupations. If an occupation consists of a large fraction of tasks with intermediate knowledge intensity, then we would expect training requirements to decrease as these tasks are automated. Panel a) of Figure 4.5 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 D.4), which appears consistent with our automation-based explanation.

Third, our model predicts that an increase in the supply of general skill s should result in skill upgrading across tasks. Indeed, average years of schooling increased in almost all occupations, as shown in panel b) of Figure 4.5. Furthermore, changes in occupation average years of schooling do not follow the same pattern as changes in training requirements, supporting our assertion that the two measures relate to distinct concepts.

Fourth, 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 4.6. Including changes in log years of education on the right hand side slightly

[36] We employ the same estimation method as Acemoglu and Autor (2011) and Autor and Dorn (2013) to facilitate comparison with their plots of employment share changes against initial occupational mean wages.

decreases the coefficient on training.[37]

Finally, we consider how changes in training requirements correlate with changes in occupational 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. A regression of changes in occupation log wages on changes in log training requirements yields a coefficient of 0.07 (standard error 0.026), see Figure 4.7. Including changes in log years of education on the right hand side slightly increases the coefficient on training.

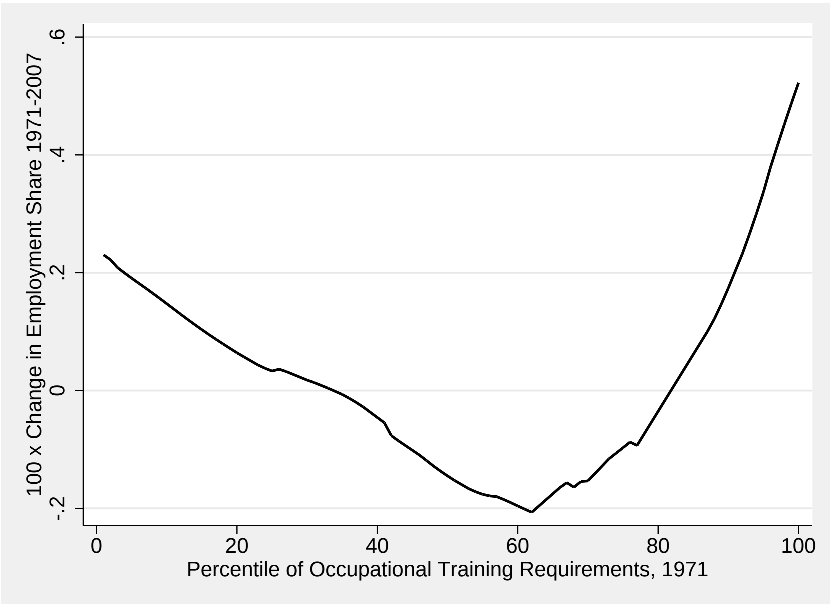
The finding is consistent with the model if we interpret falls in training requirements as increased automation of tasks. For concreteness, consider an occupation whose task bundle initially includes training-intensive tasks with knowledge intensities between σ_1^* and $\sigma' > \widehat{\sigma}_1^*$. 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 4.3 shows that these workers experience wage declines relative to most other workers.

4.7 Conclusion

In this paper we make four main contributions. First, we present a model of labor-saving technical change that endogenizes firms' decisions about what tasks to automate, as well as choices of machine design and worker training. Second, we generate job polarization endogenously. We show that job polarization and a hollowing out of the wage distribution result from any technological advance that facilitates automating a broad range of tasks, and is thus not specific to the recent information technology revolution. Third, our model allows us to investigate the effects of job polarization on wage inequality near the top of the distribution, and it generates predictions about how high skill workers might be affected by further advances in AI and robotics. Fourth, the model predicts changes in occupational training requirements that are consistent with novel evidence we present. Our model does not allow 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.

[37] 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.

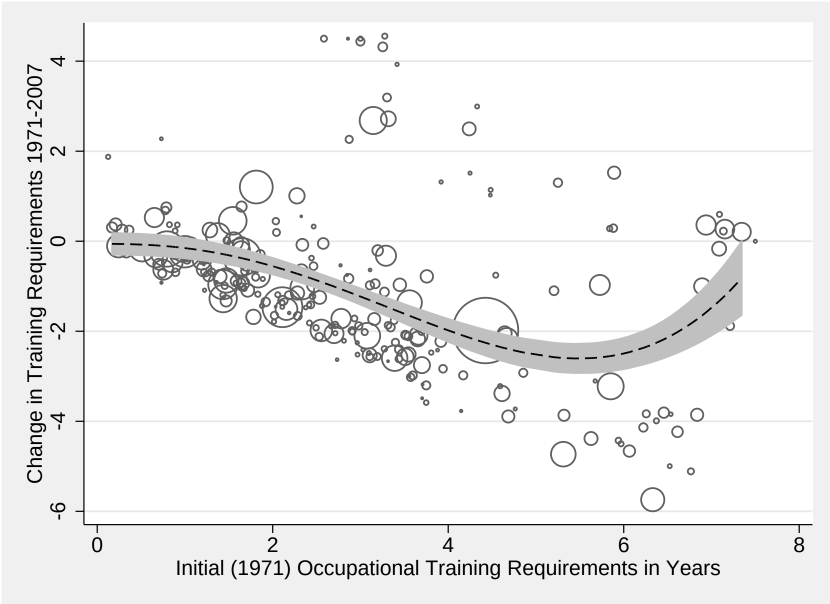
Figure 4.4: Changes in occupational shares



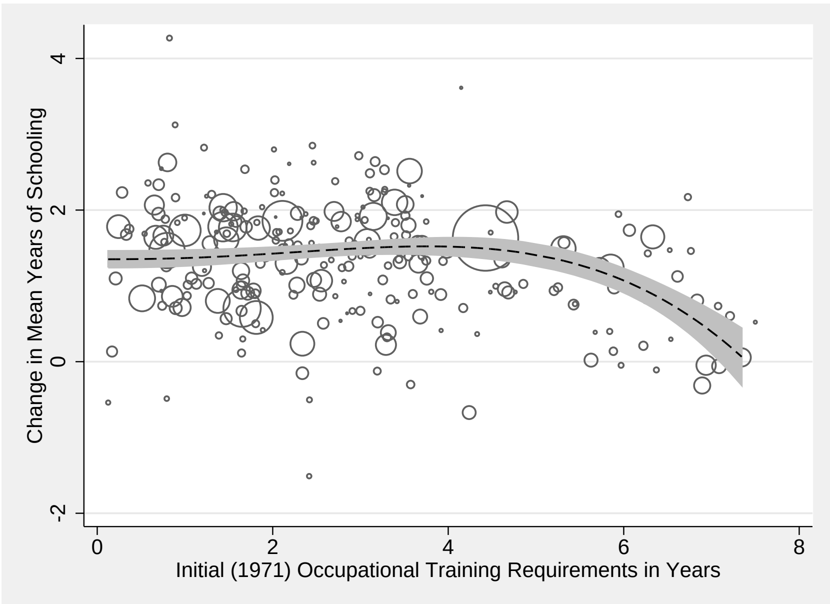
Notes: Occupations are ordered by percentile rank of the average 1980 occupational SVP-score. Fitted values from a locally weighted regression using Stata's *lowess* command.

Figure 4.5: Changes in occupational training requirements and average years of schooling

a) Changes in occupational training requirements

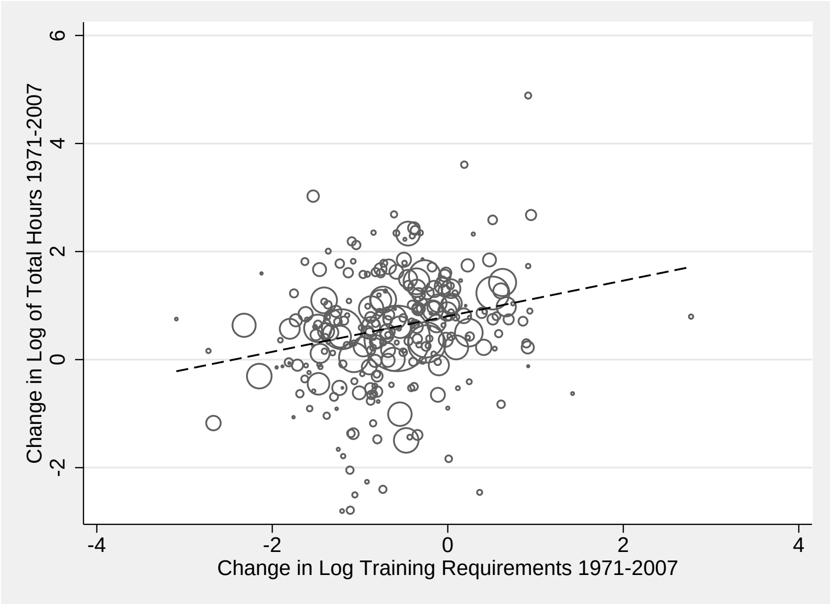


b) Changes in occupational average years of schooling



Notes: Training requirements are calculated based on 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.

Figure 4.6: Growth of occupational labor input against changes in training requirements



Notes: 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.

Figure 4.7: Changes in occupational mean wages against changes in training requirements

Notes: Occupational mean 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.

Appendix D

Appendix

D.1 Proofs of Formal Results Stated in the Text

D.1.1 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 (4.7)

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

which yields $\alpha(s_K, \underline{\sigma}) \leq r/[A_K w(s^*)]$. Using (4.17) and (4.19) this inequality is shown to be equivalent to

$$\alpha(s_K, \underline{\sigma}) \leq \frac{(1-\beta)(\sigma_1^* - \underline{\sigma})}{\beta(\overline{\sigma} - \underline{\sigma})} \times \frac{V(s^*)}{A_K K}.$$

The RHS of the last inequality is strictly less than $(1-\beta)/(\beta A_K K)$, hence a sufficient condition to rule out any equilibrium in which $k_0(\sigma) = 0$ for $\sigma \in [\underline{\sigma}, \overline{\sigma}]$ is $\alpha(s_K, \underline{\sigma}) > (1-\beta)/(\beta A_K K)$ or

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

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 (4.7)

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

from which we obtain $\alpha(s_K, \overline{\sigma})/\alpha(\overline{s}, \overline{\sigma}) \geq r/[A_K w(s^*)]$. Using (4.17) and (4.19) this inequality becomes

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

The RHS of the last inequality is strictly greater than $(1-\beta)/(\beta A_K K)$, 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 $\alpha(s_K,\overline{\sigma})/\alpha(\overline{s},\overline{\sigma}) < (1-\beta)/(\beta A_K K)$ or

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

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

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

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 if we 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 sufficiently large so that $\alpha(\overline{s},\overline{\sigma})$ stays finite. If so, then $S \to \frac{1-\beta}{\beta}(1,\infty)$.

D.1.2 Proofs of Lemmas Stated in the Text

Proof of Lemma 1 The productivity schedule α 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 (4.2) 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 (4.1) 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

$$\frac{\partial}{\partial\sigma}z(\breve{s},\sigma) > 0 \quad \Leftarrow \quad \varepsilon_{F,\sigma} < \varepsilon_{f,\sigma} \text{ for all } Z,\sigma > 0. \quad \blacksquare$$

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 (4.7) $p_0(\sigma) = w(s)$ 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 (4.4) 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 (4.7), 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 (4.2).

(b) If $k_0(\sigma) > 0$, then by the zero-profit condition (4.7)

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

and there is some σ' such that $n_1(s^*, \sigma') > 0$ and hence by (4.7)

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

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 (4.7),

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

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 (4.7),

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

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 (4.7),

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

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 (4.2). If $\sigma_0^* = \sigma_1^*$, then similar arguments lead to $\alpha(s, \sigma) = 1$, which also contradicts (4.2). ∎

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

D.1.3 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 $\widehat{M}'(\bar{s}) < M'(\bar{s})$, then $\widehat{M}(s) < M(s)$ for all $s \in [\max\{s^*, \hat{s}^*\}, \bar{s})$. For suppose that $\widehat{M}'(\bar{s}) < M'(\bar{s})$ and that there exists some $s' \in [\max\{s^*, \hat{s}^*\}, \bar{s})$ such that $\widehat{M}(s') \leq M(s')$. Then there exists some $s'' \in [s', \bar{s})$ such that $\widehat{M}(s'') = M(s'')$, $\widehat{M}'(s'') \geq M'(s'')$, and $\widehat{M}(s) > M(s)$ for all $s \in (s'', \bar{s})$. We will show that this leads to a contradiction.

Integrating (4.11) yields an expression for the wage premium of the most skilled worker with respect to any other skill group employed in training-intensive tasks,

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

where

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

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 (D.3) into (4.10), we obtain

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

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 $\widehat{\sigma}_1^* \le \sigma_1^*$ and $\widehat{M}'(\overline{s}) \ge M'(\overline{s})$.

By (4.22) and (D.4),

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

This together with (4.20), implies

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

Suppose that $\widehat{s}^* \ge s^*$. Then (D.5) implies that $\widehat{\sigma}_0^* < \sigma_0^*$, while (D.6) implies $\widehat{\sigma}_0^* > \sigma_0^*$, a contradiction. So we must have $\widehat{s}^* < s^*$. If $\widehat{\sigma}_0^* \ge \sigma_0^*$, then from (4.21), $\widehat{s}^* > s^*$,[1] so it must be that $\widehat{\sigma}_0^* < \sigma_0^*$. Then by 4.21, $\alpha(\widehat{s}_K,\widehat{\sigma}_0^*) > \alpha(s_K,\sigma_0^*)$. This implies that the LHS of (4.20) increases, while the RHS decreases, a contradiction.

Next, suppose that $\widehat{\sigma}_1^* \le \sigma_1^*$ and $\widehat{M}'(\overline{s}) < M'(\overline{s})$. We have shown that in this case the matching function shifts up, so we must have $\widehat{s}^* \le s^*$. Then $\widehat{\sigma}_0^* < \sigma_0^*$ from (4.21). But we have just shown that it is impossible to have $\widehat{\sigma}_1^* \le \sigma_1^*$, $\widehat{\sigma}_0^* < \sigma_0^*$, and $\widehat{s}^* \le s^*$ 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 (D.4), $\widehat{M}'(s) > M'(s)$ for all s belonging to the domains of both $\widehat{M}$ and M. As we

[1] To see this, rewrite (4.21) 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.

have established that $\widehat{\sigma}_1^* < \sigma_1^*$, by the properties of the matching function we must have $\widehat{s}^* > s^*$. By (4.10), the wage share of a worker who is always assigned to training-intensive tasks has increased,

$$\frac{\widehat{w}(s)}{\widehat{Y}} = \frac{1-\beta}{\mu}\frac{\widehat{M}'(s)}{v(s)} > \frac{1-\beta}{\mu}\frac{M'(s)}{v(s)} = \frac{w(s)}{Y} \quad \forall s \in [\widehat{s}^*, \overline{s}].$$

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 (4.23). 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 (4.10) and (4.19), this implies $\beta\widehat{\sigma}_0^* + (1-\beta)\widehat{\sigma}_1^* < \beta\sigma_0^* + (1-\beta)\sigma_1^*$.

Now observe that if $\widehat{M}(s) < M(s)$ then $\omega(\widehat{s}^*; \widehat{M}) < \omega(s^*; M)$ since also $\widehat{s}^* > s^*$. By (D.5), we must have $\widehat{\sigma}_0^* < \sigma_0^*$. But this means that (D.6) can only hold if also the total labor share has decreased, $\beta\widehat{\sigma}_0^* + (1-\beta)\widehat{\sigma}_1^* > \beta\sigma_0^* + (1-\beta)\sigma_1^*$, 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 knowledge-intensive. Formally, if $\beta\widehat{\sigma}_0^* + (1-\beta)\widehat{\sigma}_1^* < \beta\sigma_0^* + (1-\beta)\sigma_1^*$, then $\widehat{\sigma}_1^* < \sigma_1^*$. For suppose that $\beta\widehat{\sigma}_0^* + (1-\beta)\widehat{\sigma}_1^* < \beta\sigma_0^* + (1-\beta)\sigma_1^*$, but $\widehat{\sigma}_1^* \geq \sigma_1^*$. Then $\widehat{\sigma}_0^* < \sigma_0^*$. By (4.21), $\widehat{s}^* < s^*$. But by (4.20), $\widehat{s}^* > s^*$, a contradiction.

2. If the marginal training-intensive task becomes less knowledge-intensive, 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 $\widehat{s}^* \leq s^*$. Then (4.21) implies $\widehat{\sigma}_0^* < \sigma_0^*$. But since $\widehat{V}(\widehat{s}^*) < V(s^*)$, (4.20) 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 (4.10),

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

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. ∎

D.1.4 Proofs of Corollaries Stated in the Text

Proof of Corollary 1 Integrating (4.11), 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 $\beta(\overline{\sigma} - \sigma_0^*) + (1-\beta)(\overline{\sigma} - \sigma_1^*)$. 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 (4.20) as

$$A_K \alpha(s_K, \sigma_0^*) K = \frac{\beta(\sigma_0^* - \underline{\sigma}) + (1-\beta)(\sigma_1^* - \underline{\sigma})}{\frac{\beta(\overline{\sigma} - \sigma_0^*)}{V(s^*)}}.$$

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. ∎

D.2 A Model with Fixed Costs

We begin by simplifying the modeling of the task production process. Assume that the set of potential problems encountered in each task is given by $[0, \sigma]$. Moreover, suppose that machines and workers can only be employed in a given task if they can solve *all* problems in this interval. Thus, we abstract from training and design choices. Nevertheless, the concept of knowledge intensity is still present in the model and is captured by the parameter σ. The technologies for training workers and designing machines in the modified model are as follows. Intermediate firms must pay σ/s units of the final good to train a worker in training-intensive task σ, but face no learning cost in innate ability tasks. Maintaining the normalization that task-neutral productivity of workers equals one, we have that the marginal cost of employing labor is given by $w(s) + \sigma/s$.

To design a machine in a task with knowledge intensity σ, be it a training-intensive or an innate ability task, firms pay a one-off cost $\varphi\sigma$ and a variable cost $c_K\sigma$. Thus, the marginal cost of employing machines is $r/A_K + c_K\sigma/A_K$, where r is the rental rate of capital and A_K is task-neutral productivity of machines.

We assume that each task is produced by a single monopolistic firm.[2] In contrast, final good firms are perfectly competitive just as in the baseline version of the model. The final good production function is now

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

with $\varepsilon > 1$. 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 the profits of the firm that supplies training-intensive task σ are given by

$$\pi_1(\sigma, s|N) = a_1(\varepsilon)\left[w(s) + \sigma/s\right]^{-(\varepsilon-1)}$$

if employing workers of type s, and

$$\pi_1(\sigma|K) = a_1(\varepsilon)\left[r/A_K + c_K\sigma/A_K\right]^{-(\varepsilon-1)} - \varphi\sigma$$

if employing machines, where $a_1(\varepsilon) \equiv \varepsilon^{-\varepsilon}(\varepsilon-1)^{\varepsilon-1}(1-\beta)^{\varepsilon}Y$. In innate ability tasks, the corresponding expressions are

$$\pi_0(\sigma, s|N) = a_0(\varepsilon)w(s)^{-(\varepsilon-1)}$$

and

$$\pi_0(\sigma|K) = a_0(\varepsilon)\left[r/A_K + c_K\sigma/A_K\right]^{-(\varepsilon-1)} - \varphi\sigma$$

with $a_0(\varepsilon) \equiv \varepsilon^{-\varepsilon}(\varepsilon-1)^{\varepsilon-1}\beta^{\varepsilon}Y$. Unlike in the baseline model, incentives for employing machines depend both on knowledge intensity and training intensity. This is because of a market size effect that is present whenever the share of innate ability tasks β is different from one half.

[2] 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.

The equilibrium assignment of machines and labor to intermediate firms is qualitatively the same as in the baseline model if the marginal costs of employing machines are lower than those of employing workers. In particular, if for all s, $w(s) > r/A_k$ and $1/s > c_K/A_K$, and if $\underline{\sigma}$ is close to zero, then $\pi_0(\underline{\sigma}, s|N) < \pi_0(\underline{\sigma}|K)$ and $\pi_1(\underline{\sigma}, s|N) < \pi_1(\underline{\sigma}|K)$ for all s. Thus, the least knowledge-intensive innate ability and training-intensive tasks are performed by machines. Now observe that profits of firms employing labor approach zero, but stay strictly positive, as σ goes to infinity. In contrast, profits of firms employing machines will be negative for sufficiently large σ due to the fixed cost. Therefore, if $\overline{\sigma}$ is large, then there exist σ_0^* such that innate ability tasks with $\sigma \leq \sigma_0^*$ ($\sigma > \sigma_0^*$) are performed by machines (workers). Similarly, there exists such a marginal training-intensive task σ_1^*. If β is not too large, then $\sigma_1^* > \sigma_0^*$, so that machines are more widely adopted in training-intensive tasks. We have thus established the conditions under which technology adoption in the model with fixed costs follows the same patterns as in the baseline model (Lemma 2, part *b*).

Now consider the assignment of skill types to training-intensive tasks. If the firm supplying training-intensive task σ employs type s in equilibrium, then its profits are equal to $\pi(\sigma, s|N) = a_1(\varepsilon)\left[w(s) + \sigma/s\right]^{-(\varepsilon-1)}$. For this to be optimal, the first-order-condition

$$w'(s) - \sigma/s^2 = 0$$

and the second-order condition

$$w''(s) + 2\sigma/s^3 > 0$$

must hold. For firms supplying more-knowledge-intensive tasks to hire more highly skilled workers, it must be that $ds/d\sigma > 0$. It is easy to check that this condition is satisfied under the first- and second-order conditions above. Thus, the matching function is increasing and there is positive assortative matching as in the baseline model. Since the wage function is increasing, there must exist an s^* such that all workers with $s < s^*$ ($s \geq s^*$) are assigned to innate ability tasks (training-intensive tasks). The assignment of skill types to tasks is thus equivalent to that in the baseline model (Lemma 2, part *a*, and Lemma 3).

To solve for the matching function, follow similar steps as in the derivation of (4.10) to obtain the differential equation

$$M'(s) = \left(\frac{\varepsilon}{(\varepsilon-1)(1-\beta)}\right)^{\varepsilon} \times \frac{v(s)\left[w(s) + M(s)/s\right]^{\varepsilon}}{Y}.$$

Together with the FOC (setting $\sigma = M(s)$) and the boundary conditions $M(s^*) = \sigma_1^*$ and $M(\overline{s}) = \overline{\sigma}$, one can solve for the matching function, given a guess for s^*, σ_1^*, and Y. The model is closed by the usual market clearing conditions and the no-arbitrage equations $\pi(\sigma_0^*, s^*|N) = \pi(\sigma_0^*|K)$, $\pi(\sigma_1^*, s^*|N) = \pi(\sigma_1^*|K)$, and $w(s) = w(s^*)$ for all $s < s^*$.

Table D.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

D.3 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 4.6.2) 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 D.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 D.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 D.2 lists the twenty least and most training-intensive occupations (using David Dorn's classification) in 1971. Table D.3 does the same for 2007. Table D.4 lists the twenty occupations experiencing the largest declines and increases in training requirements.

Table D.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 D.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 D.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

Bibliography

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. Doyle (2011): "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays," *American Economic Journal: Applied Economics*, 3(3), 1–34.

Almond, D., J. 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., and C. Pierret (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116(1), 313–350.

Angrist, J., V. Chernozhukov, and I. Fernndez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74(2), 539–563.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444–472.

Angrist, J., and A. Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4), 69–85.

Angrist, J., and J.-S. Pischke (2009): *Mostly Harmless Econometrics*. Princeton University Press.

Anselin, L., A. Varga, and Z. Acs (1997): "Local geographic spillovers between university research and high technology innovations," *Journal of Urban Economics*, 42(3), 422–448.

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.

ARROW, K. (1973): "The Theory of Discrimination," *Discrimination in labor markets*, 3(10).

AUDRETSCH, D., AND M. FELDMAN (1996): "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86(3), 630–640.

AUTOR, D. (2013): "The Task Approach to Labor Markets: An Overview," IZA Discussion Papers 7178, Institute for the Study of Labor (IZA).

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.

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.

BARTEL, A., C. ICHNIOWSKI, AND K. SHAW (2007): "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills," *The Quarterly Journal of Economics*, 122(4), 1721–1758.

BECKER, G. (1964): *Human Capital*. The University of Chicago Press.

BECKER, G. S. (1993): "Nobel Lecture: The Economic Way of Looking at Behavior," *Journal of Political Economy*, 101(3), pp. 385–409.

BEDARD, K. (2001): "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, 109(4), 749–775.

BELENZON, S., AND M. SCHANKERMAN (forthcoming): "Spreading the Word: Geography, Policy and University Knowledge Diffusion," *The Review of Economics and Statistics*.

BELMAN, D., AND J. HEYWOOD (1991): "Sheepskin Effects in the Returns to Education: An Examination of 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.

BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2013): "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 128(1), 1–51.

BLOOM, N., L. GARICANO, R. SADUN, AND J. VAN REENEN (2009): "The distinct effects of Information Technology and Communication Technology on firm organization," Working Paper 14975, National Bureau of Economic Research.

BLOOM, N., R. SADUN, AND J. VAN REENEN (2012a): "Management as a Technology," *CEP Discussion Paper*.

——— (2012b): "The Organization of Firms Across Countries," *The Quarterly Journal of Economics*, 127(4), 1663–1705.

BLOOM, N., AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices across Firms and Countries," *The Quarterly Journal of Economics*, 122(4), 1351–1408.

——— (2010a): "New Approaches to Surveying Organizations," *American Economic Review: Papers and Proceedings*, 100(2), 105–109.

——— (2010b): "Why Do Management Practices Differ Across Firms and Countries," *Journal of Economic Perspectives*, 24(1), 203–224.

BOEHM, M. J. (2013): "Has Job Polarization Squeezed the Middle Class? 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.

BRESNAHAN, T., E. BRYNJOLFSSON, AND L. HITT (2002): "Information Technology, Workplace Organization and the Demand for Skilled Labor: Firm Level Evidence," *The Quarterly Journal of Economics*, 117(1), 339–376.

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.

CAMERON, A. C., J. GELBACH, AND D. MILLER (2011): "Robust Inference with Clustered Data," *Journal of Business and Economic Statistics*, 29(2), 238–249.

CARD, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *Aspects of Labour Economics: Essays in Honour of John Vanderkamp*.

CAROLI, E., AND J. VAN REENEN (2001): "Skill-Biased Organizational Change? Evidence from a Panel of British and French Establishments," *The Quarterly Journal of Economics*, 116(4), 1449–1492.

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.

CLARK, D., AND P. MARTORELL (2010): "The Signalling Value of a High School Diploma," *Princeton University Industrial Relations Section Working Paper 557*.

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.

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

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.

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.

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

FARBER, H., AND R. GIBBONS (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111(4), 1007–1047.

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).

FRENETTE, M. (2006): "Too Far to Go On? Distance to School and University Participation," *Education Economics*, 14(1), 31–58.

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.

GENNAIOLI, N., R. LA PORTA, F. LOPEZ-DE-SILANES, AND A. SHLEIFER (2013): "Human Capital and Regional Development," *The Quarterly Journal of Economics*, 128(1), 105–164.

GIBBONS, S., T. LYYTIKAINEN, H. OVERMAN, AND R. SANCHIS-GUARNER (2012): "New Road Infrastructure: the Effects on Firms," *SERC Discussion Paper 117*.

GIROUD, X. (2013): "Proximity and Investment: Evidence from Plant-Level Data," *The Quarterly Journal of Economics*, 128(2), 861–915.

GLAESER, E. L., AND J. D. GOTTLIEB (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983–1028.

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.

——— (2011): "Explaining job polarization: the roles of technology, offshoring and institutions," Open Access publications from Katholieke Universiteit Leuven urn:hdl:123456789/331184, Katholieke Universiteit Leuven.

GRAY, R. (2011): "Taking Technology to Task: The Skill Content of Technological Change in Early Twentieth Century United States," Working Papers 0009, European Historical Economics Society (EHES).

GREENSTONE, M., R. HORNBECK, AND E. MORETTI (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118(3), 536–598.

GREGG, P., S. MACHIN, AND A. MANNING (2004): *Mobility and Joblessness*pp. 371–410. University of Chicago Press.

GROEN, J. A. (2004): "The effect of college location on migration of college-educated labor," *Journal of Econometrics*, 121(1-2), 125–142.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 20109.

HANDEL, M. J. (2000): "Trends in Direct Measures of Job Skill Requirements," Economics Working Paper Archive 301, Levy Economics Institute, The.

HAUSMAN, N. (2012): "University Innovation, Local Economic Growth, and Entrepreneurship," *US Census Bureau Center for Economic Studies*, CES-WP-12-10.

HENDERSON, R., A. JAFFE, AND M. TRATJENBERG (1998): "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 19651988," *The Review of Economics and Statistics*, 80(1), 119–127.

HOLLAND, P. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396), 945–960.

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

HOPENHAYN, H. A. (1992): "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), pp. 1127–1150.

HUNGERFORD, T., AND G. SOLON (1987): "Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 69(1), 175–177.

IMBENS, G., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 61(2), 467–476.

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

IMBENS, G., 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," *Centre for Economic Performance Discussion Papers*.

JAEGER, D., AND M. PAGE (1996): "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, 78(4), 733–740.

JAFFE, A., M. TRATJENBERG, AND R. HENDERSON (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 108(3), 577–598.

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

KALNINS, A., AND F. LAFONTAINE (2013): "Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance," *American Economic Journal: Microeconomics*, 5(3), 157–79.

KJELLSTRMA, C., AND H. REGNRA (1999): "The Effects of Geographical Distance on the Decision to Enrol in University Education," *Scandinavian Journal of Educational Research*, 43(4).

KODRZYCKI, Y. (2001): "Migration of Recent College Graduates: Evidence from the National Longitudinal Survey of Youth," *New England Economic Review*, pp. 13–34.

LANG, K., AND D. KROPP (1986): "Human Capital versus Sorting: The Effects of Compulsory Attendance Laws," *The Quarterly Journal of Economics*, 101(3), 609–624.

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

LAYARD, R., AND G. PSACHAROPOULOS (1974): "The Screening Hypothesis and the Returns to Education," *Journal of Political Economy*, 82(5), 985–98.

LEE, D., AND D. CARD (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142(2), 655–674.

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

LEMIEUX, T. (2006): "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. RACINE (2007): *Nonparametric Econometrics*. 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).

LUCAS, R. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2), 508–523.

LYCHAGIN, S., J. PINKSE, M. SLADE, AND J. VAN REENEN (2010): "Spillovers in Space: Does Geography Matter?," *NBER Working Paper No. 16188*.

MALAMUD, O., AND A. WOZNIAK (2012): "The Impact of College on Migration: Evidence from the Vietnam Generation," *Journal of Human Resources*, 47(4), 913–950.

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

MARKOFF, J. (2010): "Google Cars Drive Themselves, in Traffic," *The New York Times*, Oct 9.

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," *The Warwick Economics Research Paper Series*.

MELITZ, M. (2003): "The Impact of Trade on Intra-industry Reallocations and Aggregate Productivity Growth," *Econometrica*, 71(6), 1695–1725.

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

MINCER, J. (1974): *Schooling, Experience, and Earnings*. Columbia University Press.

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

MORETTI, E. (2004a): "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-sectional Data," *Journal of Econometrics*, 121(1), 175–212.

——— (2004b): "Workers' Education, Spillovers, and Productivity: Evidence from Plant-level Production Functions," *American Economic Review*, 94(3), 656–690.

——— (2011): "Local Labor Markets," *Handbook of Labor Economics*, 4b, 1238–1313.

MURRAY, M. P. (2006): "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, 20(4), 111–132.

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," .

NELSON, R., AND E. PHELPS (1966): "Investment in Humans, Technological Diffusion and Economic Growth," *American Economic Review*, 56(1), 69–75.

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

PHELPS, E. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62(4), 659–661.

PINKER, S. (1994): *The Language Instinct: the New Science of Language and Mind*. Penguin.

RILEY, J. (1979): "Testing the Educational Screening Hypothesis," *Journal of Political Economy*, 87(5), 227–252.

ROBACK, J. (1988): "Wages, Rents and Amenities: Differences Among Workers and Regions," *Economic Inquiry*, 26(1), 23–41.

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

SANCHIS-GUARNER, R. (2012): "Driving Up Wages: The Effects of Road Construction in Great Britain," *SERC Discussion Paper 120*.

SHERMIS, M. D., AND B. HAMNER (2012): "Contrasting State-of-the-Art Automated Scoring of Essays: Analysis," Discussion paper.

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.

SPEAR, B. (2001): "Robots and patents," *World Patent Information*, 23(4), 333–338.

SPENCE, M. (1973): "Job Market Signalling," *The Quarterly Journal of Economics*, 87(3), 355–374.

SPIESS, K., AND K. WROHLICH (2010): "Does distance determine who attends a university in Germany?," *Economics of Education Review*, 29(3), 470–479.

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

STIGLITZ, J. (1975): "The Theory of Screening, Education, and the Distribution of Income," *American Economic Review*, 65(3), 283–300.

STOCK, J., J. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

SYVERSON, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–365.

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 (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51(6), 30917.

TYLER, J., R. MURNANE, AND J. WILLET (2002): "Estimating the Labor Market Signaling value of the GED," *The Quarterly Journal of Economics*, 115(2), 431–468.

URQUIOLA, M., AND E. A. VERHOOGEN (2009): "Class-Size Caps, Sorting, and the Regression-Discontinuity Design," *American Economic Review*, 99(1), 179215.

WEISS, A. (1995): "Human Capital and Sorting Models," *Journal of Economic Perspectives*, 9(4), 133–154.

WOLPIN, K. (1977): "Education and Screening," *American Economic Review*, 67(5), 949–958.

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