Essays in Applied Microeconomics

Stephanie Christina De Mel

Faculty of Economics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College May 2019

“There will be time, there will be time
To prepare a face to meet the faces that you meet;
There will be time to murder and create,
And time for all the works and days of hands
That lift and drop a question on your plate;
Time for you and time for me,
And time yet for a hundred indecisions,
And for a hundred visions and revisions,
Before the taking of a toast and tea.

In the room the women come and go
Talking of Michelangelo.

And indeed there will be time
To wonder, “Do I dare?” and, “Do I dare?”
Time to turn back and descend the stair,
With a bald spot in the middle of my hair –
(They will say: “How his hair is growing thin!”)
My morning coat, my collar mounting firmly to the chin,
My necktie rich and modest, but asserted by a simple pin –
(They will say: “But how his arms and legs are thin!”)
Do I dare
Disturb the universe?
In a minute there is time
For decisions and revisions which a minute will reverse.

For I have known them all already, known them all:
Have known the evenings, mornings, afternoons,
I have measured out my life with coffee spoons;
I know the voices dying with a dying fall
Beneath the music from a farther room.
So how should I presume?”

— ”The Love Song of J. Alfred Prufrock”, T.S. Eliot

Declaration

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed limit of 60,000 words.

Stephanie De Mel

May 2019

Acknowledgements

I feel privileged and deeply grateful to have had Kaivan Munshi and Hamish Low as my supervisor and research advisor, respectively. They have been magnanimous with time, support and guidance, and, equally, have given me the space and freedom to carve out my own path as an independent researcher. I have learned skills from them that I know will be invaluable in my future work. I could not have asked for better mentors.

Soenje Reiche has been consistently kind, supportive and a wonderful person to work with; I have learned a great deal from her and will always be grateful for her support of me.

I would also like to thank my other co-author, Hamid Sabourian, for giving me the opportunity to work with him on a really interesting project, and to acknowledge advice and encouragement from Pramila Krishnan, Kai Liu, Coen Teulings, Sriya Iyer, Gabriella Santangelo, Weilong Zhang and Noriko Amano-Patino.

I gratefully acknowledge financial support from the Cambridge Trust and the Keynes Fund, as well as IT and administrative support from Louise Cross, Silvana Dean, Ryan Hilton, Jake Dyer and Craig Peacock.

I have been blessed with wonderful friends (both within and without the Economics faculty) who have shared quality time, boardgames, film nights, cups of tea and travel, and have propped me up with an overwhelming amount of love through the more difficult parts of this journey. I am so fortunate to have them in my life.

Lastly, and most importantly, I'd like to thank my family – Mum, Dad, Monkey, Grandy and Grandma – for their unconditional love, support, and belief in me. Without them, I doubt I'd have made it to the start of this PhD, let alone its end. Words fail to suffice in this respect, but I hope they know how important they are to me.

Contents

List of Figures

List of Tables

Preface

This thesis contains three chapters, each of which employs both reduced-form and structural tools of applied microeconomics to answer questions collectively in the fields of education, labour and health economics.

The first chapter considers the effect of a recent expansion of private tertiary education in a set of six developing countries on the equity and efficiency of labour-market outcomes. While the presence of private universities in developing countries has increased significantly in recent decades, their growth experience varies widely across individual countries. This paper seeks to explain such variation and, consequently, to consider the range of impact private universities have on the efficiency of human capital allocation, aggregate output, and the dynamic evolution of income inequality. To that end, I build and estimate a structural model of tertiary education choice using data for a group of six developing countries: Armenia, Bolivia, Colombia, Georgia, Ghana and Kenya. I show that differences in underlying parameter values across countries have important implications for the composition of the graduate workforce, the growth of private universities, output per worker and income inequality. I find that, over time, when private universities are more productive than public ones, economies tend to move away from the most socially efficient of these scenarios, while the least efficient scenario is highly persistent.

The second chapter employs a structural model of endogenous education and occupational choice to demonstrate that youth unemployment in Ghana increases in parental wealth, and to consider the consequences of such a relationship for wage inequality, educational attainment, and aggregate productivity. I argue that, in the absence of unemployment insurance, only workers with a sufficiently high stock of parental wealth can afford to remain unemployed, and do so in order to search for scarce, high-productivity jobs. This leads to high income inequality and low match efficiency among workers of heterogeneous ability. I use the estimation results to compare the effectiveness of two alternative policy interventions: an education subsidy and unemployment insurance. I find that the education subsidy is most effective at increasing aggregate productivity, but comes at the cost of increasing income inequality, while unemployment insurance has a smaller effect on aggregate productivity but also decreases income inequality.

The third chapter, which is co-authored with Prof. Kaivan Munshi, Dr. Soenje Reiche and

Prof. Hamid Sabourian, examines the consequences for efficiency of information cascades among UK hospitals in organ transplant queues. While demand significantly outstrips organ availability, we observe that 40% of livers and 22% of kidneys failed to be utilised over the last decade. We build a structural model in which organs are sequentially assessed by centres of heterogeneous ability, and show that herding behaviour plays an important role in such wastage: once an organ is rejected by one or more centres, subsequent centres emulate their behaviour, ignoring their own assessment of the organ's quality. We employ unique administrative data from the NHS Blood and Transplant (NHSBT) that covers the universe of abdominal organs donated in the UK during 2006-2016 to provide reduced-form evidence of herding behaviour among transplant centres. Further, we undertake a set of counter-factual analyses to demonstrate that, while herding behaviour is common among UK transplant centres, the resulting increase in discard rates is not substantially higher than that of the full-information benchmark. Equally, it performs the important function of preventing centres from accepting organs of poor quality such that, overall, the benefits derived from observing predecessors' decisions outweigh the costs of herding traditionally emphasised in the theoretical literature. In contrast with this literature, therefore, we find that, in this context, herding is efficiency-enhancing overall.

Chapter 1

Private or Public? Efficiency and Equity Implications of Tertiary Education Choice

Abstract

While the presence of private universities in developing countries has increased significantly in recent decades, their growth experience varies widely across individual countries. This paper seeks to explain such variation and, consequently, to consider the range of impact private universities have on the efficiency of human capital allocation, aggregate output, and the dynamic evolution of income inequality. To that end, I build and estimate a structural model of tertiary education choice using data for a group of six developing countries: Armenia, Bolivia, Colombia, Georgia, Ghana and Kenya. I show that underlying parameter values position countries in one of three potential scenarios, each of which has different implications for the composition of the graduate workforce, the growth of private universities, output per worker and income inequality. I find that, over time, when private universities are more productive than public ones, economies tend to move away from the most socially efficient of these scenarios, while the least efficient scenario is highly persistent.

Keywords: Higher Education Choice, Inequality, Occupational Choice, Credit Constraints

JEL codes: J24, I24, J21

1.1 Introduction

Private tertiary education has expanded rapidly across developing countries in recent decades. In Sub-Saharan Africa, for instance, the number of private universities grew from an estimated 24 to 468 over the period 1990-2007 (Yusuf et al. (2009)), while the share of private universities in Latin America and Asia reached 40 per cent and 35 per cent of total tertiary enrolments in 2009, respectively (Praphamontripong (2012)). As private universities in these regions tend to be small and highly numerous, their share in the total number of institutions is estimated to be even larger. Nevertheless, the growth experience of private universities varies widely across different countries. Data from the World Bank's "Skills Toward Employment and Productivity" (STEP) survey demonstrates these differences quite clearly: the share of private graduates in the total tertiary-educated sample varies from less than 3 per cent in Armenia and China's Yunnan Province; to 10-15 per cent in Ghana, Kenya and Georgia; to 61 per cent in Colombia.

In this paper, I build a structural model to explain such variation. I then consider the manner in which it underlies differences in two main outcomes: firstly, the allocation of tertiary education across high- and low-ability workers and its effects on output per worker; and secondly, the linkage between tertiary education choice and the evolution of income inequality. Accordingly, I estimate the parameters of the structural model for a set of six developing countries - Armenia, Bolivia, Colombia, Georgia, Ghana and Kenya - using data from the World Bank's STEP survey, and demonstrate that the proportion of private university graduates in the population may grow; remain at a low, constant level; or collapse to zero over time. The observed heterogeneity of outcomes across countries may be explained by variation in underlying parameter values, which position a country in any one of three different scenarios.

In the first, wealthy, high-ability workers always attend private university; in the second, the same group always attends public university; and, in the third, non-zero fractions of this group attend private and public university, respectively. This sorting behaviour determines whether private universities in a given country are able to attract high-ability students, whose subsequent contribution to aggregate output in the workforce leads to higher productivity, or whether they are populated solely by wealthy students of inferior ability, who attend private university only because they are unable to obtain places in highly-competitive public universities. In the latter case, the contribution of private universities to productivity is likely to be much smaller. I show that, under the assumptions of my model framework, the first case is relatively efficient but not persistent over time, while the remaining two cases are both relatively inefficient and persistent. This suggests a role for government intervention in certain situations, either to subsidise private university education or to adopt its technology through greater investment in public education.

This paper draws on two related but largely unconnected strands of literature: a theoretical literature on the relationship between human capital and the inter-generational perpetuation of income inequality; and a small, primarily empirical, literature that attempts to measure (in a variety of contexts) the wage premium to private education.

In the literature on human capital and inter-generational income inequality, a few theoretical papers are close to the spirit of this paper. Galor and Zeira (1993) build a theoretical framework to demonstrate that, with imperfect credit markets and indivisibility in human capital, the initial wealth distribution significantly affects aggregate growth through its influence on the human capital investment decision. Loury (1981) models the dynamics of the income distribution across generations when credit constraints limit families' ability to invest in human capital. Glomm and Ravikumar (1992) present an OLG model in which the educational regime is determined by majority voting; they show that income inequality declines faster under public education than under private education. Durlauf (1996) analyses the dynamics of income inequality through neighbourhood choice and its effect on public funding for education. The theoretical framework in this paper also draws motivation from Epple and Romano (1998), which derives the optimal admissions and pricing policies of private universities (albeit under the assumption of peer effects within cohorts, which are absent from my model) and outlines a framework in which the best students sort into high-quality private institutions, while the single public university admits the lowest-ability cohort.

Research on the wage premium to private tertiary education is largely concentrated in developed-country contexts, and generally finds a sizeable wage premium for private graduates. Using US data, Brewer et al. (1999) find evidence of a large wage premium to attending an elite private institution, and a smaller premium for attending a mid-level private university, compared to attending a low-level public university. Dale and Krueger (2002) find no evidence that students who attended more selective colleges in the US earned more than students of comparable ability who attended less selective schools, but do find a substantial premium to attending a more costly university. Chevalier et al. (2003) consider the effect of university choice on graduate wages in the UK; they estimate a fee differential of \$2950-7250 between the graduates of prestigious and less prestigious universities, under different scenarios, albeit without distinguishing between private and public institutions.[1]

Research on this question in developing countries is far more sparse, particularly at the level

[1] Indeed, such a distinction is not possible as all UK universities are (either government-financed or independent) private instituitions.

of tertiary education. To the best of my knowledge, the only paper which considers the wage premium to attending private university in a developing country is Calónico and Ñopo (2007), who use the National Living Standards Measurement Household Survey for Peru in 1997 and 2000 to measure the wage differential between private and public education at primary, secondary and tertiary levels. They find evidence of higher returns to education for those who attended private university, but also of higher variation in wages for this group, suggesting a large dispersion of quality among private universities, which is consistent with general accounts in the development literature.

Research on lower levels of education provides a similar narrative. Bedi and Garg (2000) find that graduates of private secondary schools in Indonesia perform better in the labour market. This, interestingly, is despite a general consensus that Indonesian public schools are of a higher quality than private ones. Finally, Alderman et al. (2001) identify factors that affect demand by low-income households in Pakistan for private schooling. They find "that even the poorest households use private schools extensively, and that utilization increases with income".

The dearth of research concerning the (individual and aggregate) effects of private universities in developing countries is thus a notable gap in the literature, particularly in light of their rapid proliferation across Africa, Asia and Latin America in recent decades. Prior to the 1990s, private universities accounted for a negligible fraction of tertiary enrolments in most developing countries; instead, an effective state monopoly over the sector was characterised by low tuition fees and, due to low funding, heavily restricted supply. The latter resulted in high entry requirements, such that only the best students could attend university (Varghese (2006)). The growth of private universities in these countries during the late post-colonial period was a market-based response to the shortage of public university places, particularly at a time of rapidly-growing demand for tertiary education and government funding cuts for public education. A change in student tastes, particularly a growth in demand for courses such as IT, accounting and finance, which public universities were ill-equipped to provide, also underlies their popularity.

While it is beyond the scope of this research to provide a comprehensive account of the features of private universities in developing countries, the following two stylised facts are particularly significant in motivating this paper. Firstly, the majority of private universities in Africa, Asia and Latin America are profit-making institutions with nominal entry requirements (Praphamontripong (2012)). This has the following important implication - that, while entry into public universities continues to be determined by ability, entry into these new universities is largely predicated on wealth. In contexts with ill-functioning credit markets, this constraint will be particularly binding.

Secondly, despite their large variation in quality, private universities are often favoured over their public counterparts by private-sector employers as a source of new hires. A World Bank survey across 10 developing countries in 2012-2014 revealed that, while 68 per cent of employed private university graduates are placed in the private sector, only 38 per cent of employed public university graduates are located here (World Bank (2012/2014)). This is despite the fact that average graduate wages are higher in the private sector than in the public, so the outcome is unlikely to be driven by an underlying preference for public-sector jobs.[2]

A potential explanation is that private university graduates possess more and better non-curricular skills of the form particularly sought by private-sector employers. World Bank employer surveys across several developing countries reveal problem-solving, learning, ICT, communication and social skills to be skills valued by private-sector employers (Banerji (2010)). Summary statistics from the STEP data further reveal that private university graduates have higher average job-relevant skills than their public university counterparts. While this cannot be taken to indicate a causal relation, these summary statistics are, nevertheless, broadly in line with the stylised fact described above.

These two stylised facts collectively motivate the research questions underlying this paper: how does the introduction of private universities influence the allocation of heterogeneous workers in the labour-force; what are the implications of this allocation for aggregate productivity; and, finally, what does this imply for income inequality over successive generations? The remainder of this paper is structured as follows. Section 2 presents the baseline model. Section 3 presents a model extension that allows for asymmetric information about worker ability. Section 4 introduces the data and provides some summary statistics. Section 5 analyses the results from a structural estimation of the model parameters. Section 6 concludes.

1.2 Baseline Model

This section builds a theoretical framework that explains the variation across developing countries in private universities' share of total tertiary enrolments, as well as in the composition of high- and low-ability workers in the private university cohort. I begin with a discrete-choice model of tertiary education with full information and no credit markets. I

[2] Average graduate daily wages in the sample are US$16 for private-sector workers and US$12 for public-sector workers

consider an overlapping generations framework in which each generation of workers lives for two periods. Each individual in a given generation is connected to a single individual in the previous generation (a parent) from whom he receives a bequest, and to a single individual in the subsequent generation (a child) to whom he leaves a bequest.

1.2.1 Workers

The model economy consists of a working-age population of unit size. All workers have completed secondary schooling up to a high-school exit exam, and are risk-neutral.

Endowments:

Workers are heterogeneous in initial endowments of wealth and ability, which, for the first generation, are both independent random variables. Innate ability of worker i, denoted h_i, is drawn from a uniform distribution over the interval $[0, 1]$ with cdf $F(h)$. A worker i's wealth, y_i, may be high or low, denoted y^H and y^L, respectively. In subsequent generations, workers' initial wealth is equal to the bequest received from parents, but ability remains an independent random draw. The fraction of high-wealth types in the population is θ_t. A high-wealth worker is one whose initial wealth endowment is greater than or equal to the highest tuition fee (x_3^*) charged by a private university in any equilibrium. Conversely, a low-wealth type has wealth less than the private university's marginal cost (c). As the private university never sets a fee lower than c, agents with wealth y_L can never attend private university.

$$y^H \equiv y_i \geq max\{x_3^*\} \tag{1.1}$$

$$y^L \equiv y_i < c \tag{1.2}$$

Choices:

At time $t = 1$, workers must choose amongst three mutually exclusive and exhaustive alternatives: work in the non-graduate sector (i.e. do not attend university), attend public university, or attend private university. Let $d_k = 1$ if alternative k is chosen, and zero otherwise, where $k = \{1, 2, 3\}$ corresponds to choices of non-graduate work, public university and private university, respectively.

Let $s \in S$ be the individual's information set at the time of decision-making, and $s = \{y_i, h_i, h^*, w_3, w_2, w_1, x_3\}$, where w_k denotes wage payments for workers conditional on education choice, h^* denotes the public university's entry requirement, and x_3 denotes the private university tuition fee. Finally, let $R_k(s)$ denote the expected pay-off to alternative k. Now, workers seek the optimal strategy $d(s)$ to maximise the discounted sum of future pay-offs:

$$V(s) = max\left(\sum_{t=1}^{2}\beta^{t-1}\sum_{k=1}^{3}[d_k(s)R_k \mid s]\right) \tag{1.3}$$

Thus,

$$V(s) = max\{V_1(s), V_2(s), V_3(s)\} \tag{1.4}$$

where $V_k(s)$ is the present value if action k is chosen. The optimal strategy is then:

$$d_k = \begin{cases} 1 & \text{iff } V_k(s) = max\{V_1(s), V_2(s), V_3(s)\} \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

Nevertheless, not all workers face the same set of alternatives; the feasible choice set for worker i, denoted D_i, is determined by her initial endowment values, h_i and y_i. Workers are thus divided into four sub-groups, as follows:

$$D_i = \begin{cases} \{1\} & \text{if } y_i = y^L \text{ and } h_i < h^* \\ \{1,2\} & \text{if } y_i = y^L \text{ and } h_i \geq h^* \\ \{1,3\} & \text{if } y_i = y^H \text{ and } h_i < h^* \\ \{1,2,3\} & \text{if } y_i = y^H \text{ and } h_i \geq h^* \end{cases} \tag{1.6}$$

At time t=2, workers must choose an optimal allocation of their income between second-period consumption and a bequest for their child. There is no consumption in the first period. The derivation of optimal bequests for non-graduates, public university graduates and private university graduates may be found in Appendix A.1.

1.2.2 Universities

Rather than simply providing a signalling function as in Spence (1973), university education in this model augments workers' productivity by a factor γ_k. Thus:

$$\gamma_k \begin{cases} = 1 & \text{if } k = 1 \\ > 1 & \text{if } k = \{2,3\} \end{cases} \tag{1.7}$$

such that remaining a non-graduate leaves productivity unchanged, while both private and public university education strictly increase worker productivity.

Assumption 1:

$$Q = \gamma_k h_i \tag{1.8}$$

By Assumption 1, the production function for human capital displays complementarity in tertiary education and innate ability such that, ceteris paribus, a worker with relatively high ability benefits more from university education than does a worker with relatively low ability.

Assumption 2:

$$\gamma_3 > \gamma_2 \tag{1.9}$$

Here I assume that private universities augment productivity by a larger factor than do public universities. This captures the stylised fact that private university graduates have more or better job-relevant skills than their public university counterparts. I later test this assumption using the STEP data and find evidence in support of it for five out of the six countries in my sample. The productivity differential thus represents the job-relevant skills differential between private and public university graduates.

Finally, both private and public universities are homogeneous in their cost function, which is given below, and where n_k denotes the size of the student cohort in a given university type and c is the constant marginal cost.

$$C(n_k) = cn_k \tag{1.10}$$

1.2.2.1 Public Universities

Public universities are assumed to be homogenous in quality, so I treat them as a single, large institution. This institution is supported entirely by government funds, and is assumed to charge no tuition fee. The cost of education is financed via a lump-sum tax, τ, on all workers.

Assumption 3:

$$x_2 = 0 \tag{1.11}$$

Nevertheless, this university has a fixed number of seats, denoted $\bar{U}$. Thus, eligibility to enter public university is determined by an entry requirement, $h_i \geq h^*$. In equilibrium, the value of h^* is endogenously determined such that the public university exactly meets its enrolment target, given its information set, which allows it perfectly to anticipate the education choices of individuals at each level of wealth and ability.

1.2.2.2 Private Universities

Like their public counterparts, private universities are homogeneous in quality, so I treat them as a single, monopolistic firm. Price discrimination is not permitted here, so the firm must charge a single price to all applicants. In equilibrium, therefore, the private university sets a tuition fee above marginal cost and earns positive profits. Thus:

$$x_3^* > c \tag{1.12}$$

There is no entry requirement in terms of innate ability, but the value of the tuition fee creates an effective cut-off rule in terms of wealth - only workers endowed with initial wealth y^H are eligible to attend.

1.2.3 Employers

There is a perfectly competitive labour market, in which workers of all education types are paid their marginal product. The market consists of two types of employer: a graduate employer and a non-graduate employer.

1.2.3.1 Non-Graduate Employer

The non-graduate employer hires workers for an unskilled task, for which neither tertiary education nor innate ability contributes to productivity. The production function is given below, where n denotes the fraction of the labour force currently working in the non-graduate labour market.

Assumption 4:

$$Y_{NG} = n_1 \tag{1.13}$$

Non-graduate wages are thus given by:

$$w_1 = 1 \tag{1.14}$$

1.2.3.2 Graduate Employer

I assume that the graduate sector employer uses public and private university graduates as perfectly substitutable inputs in a constant elasticity of substitution production function with no capital. There are thus no complementarities between the two types of graduate worker in the production process.[3]

Assumption 5:

$$Y_G = \gamma_3 h_i n_3 + \gamma_2 h_i n_2 \tag{1.15}$$

where n_3 and n_2 denote the private and public university graduating cohorts, respectively. As tertiary education and innate ability are complements in the production function, the graduate employer never hires non-graduates. Wages for private and public university graduates are given as follows:

$$w_3 = \gamma_3 h_i \tag{1.16}$$

$$w_2 = \gamma_2 h_i \tag{1.17}$$

Finally, the private-public graduate wage premium is given by:

$$\omega = \frac{\gamma_3}{\gamma_2} \tag{1.18}$$

1.2.4 Timing of the Model

The timing of the model for a given generation is as follows. At time $t = 0$, the values of workers' initial endowments (y_i, h_i) are realised. At time $t = 1$, employers make wage offers (w_1, w_2, w_3), the private university makes a tuition fee offer (x_3^*) and the public university posts an entry requirement (h^*). Workers observe these offers and choose a type of tertiary education, or work in the non-graduate sector. At time $t = 2$, all workers are in the labourforce, with employment type and wages determined by their education choices at time $t = 1$. They allocate their income optimally between second-period consumption and a bequest for their children. All workers of the generation die at the end of this period.

1.2.5 Characterising the Equilibrium

1.2.5.1 Workers' Decision Rules

A worker with ability h_i and wealth y_i chooses d_k^* to maximise discounted lifetime earnings (see Appendix A.2 for decision rules). Thus, I can write the education choice rule for each of the four sub-groups of worker defined by variation in feasible choice sets, as follows:

Workers with $y_i = y^L$, $h_i < h^*$ are always non-graduates as their only feasible choice is

[3] This assumption has been previously made in the literature and allows me to keep the framework simple and tractable for my purposes, by avoiding the dependence of marginal products on relative labour supplies

$d_k^* = d_1$.

Workers with $y_i = y^L, h_i \geq h^*$ may be non-graduates or public university graduates. Their optimal strategy is

$$d_k^* = \begin{cases} d_2 & \text{if } h_i > \frac{w_1(1+\beta)}{\beta\gamma_2} \\ d_1 & \text{otherwise} \end{cases} \tag{1.19}$$

where β is the discount factor between periods.

Workers with $y_i = y^H, h_i < h^*$ may be non-graduates or private university graduates. Their optimal strategy is:

$$d_k^* = \begin{cases} d_3 & \text{if } h_i > \frac{w_1(1+\beta)+x_3}{\beta\gamma_3} \\ d_1 & \text{otherwise} \end{cases} \tag{1.20}$$

Finally, workers with $y_i = y^H, h_i \geq h^*$s is the only group for whom all three choices are feasible. Their optimal strategy is:

$$d_k^* = \begin{cases} d_3 & \text{if } h_i > \frac{w_1(1+\beta)+x_3}{\beta\gamma_3} \text{ and } h_i > \frac{x_3}{(\gamma_3-\gamma_2)\beta} \\ d_2 & \text{if } h_i < \frac{x_3}{(\gamma_3-\gamma_2)\beta} \text{ and } h_i > \frac{w_1(1+\beta)}{\gamma_2\beta} \\ d_1 & \text{if } h_i < \frac{w_1(1+\beta)+x_3}{\gamma_3\beta} \text{ and } h_i < \frac{w_1(1+\beta)}{\gamma_2\beta} \end{cases} \tag{1.21}$$

1.2.5.2 Equilibrium Worker Sorting

I now describe workers' sorting behaviour in equilibrium. The feasible choice sets and decision rules described above result in a set of three distinct cut-off values of ability:

i $h_i = \frac{x_3}{\beta(\gamma_3-\gamma_2)}$

ii $h_i = \frac{(1+\beta)w_1+x_3}{\beta\gamma_3}$

iii $h_i = \frac{(1+\beta)w_1}{\beta\gamma_2}$

The relative ordering of these cut-offs with respect to each other, to the public university's entry requirement $h_i \geq h^*$ and to the upper bound of the ability distribution, $\bar{h} = 1$, fully determines equilibrium sorting across the three education alternatives by the four worker sub-groups. Imposing some structure on the public university capacity constraint allows me to restrict the set of equilibria to three "reasonable" outcomes (see Appendix A.3). The rationale for this is as follows: the public university's physical space constraint $\bar{U}$ requires a high value of h^*. Thus, it is reasonable to assume that this capacity constraint is binding, such that there exists a sub-group of workers with ability $h_i < h^*$ who finds it optimal to attend public university but does not meet the entry requirement. Similarly, as the private university imposes no entry requirement and h^* is high, there is likely to exist a sub-group of workers with $h_i < h^*$ who prefers to attend private university than to remain non-graduates. Formally, this assumption amounts to a public university capacity constraint that satisfies the following condition (see Appendix A.4 for derivation):

Assumption 6

$$\bar{U} \leq \frac{(1-\theta)[\beta\gamma_3 + (1+\beta)]}{\beta\gamma_2} \tag{1.22}$$

All but the sub-group of workers with $h_i \geq h^*$ and $y_i = y^H$ maintain the same optimal behaviour across all three cases, namely that all workers with $h_i < h^*$ and $y_i = y^L$ remain non-graduates, all workers with $h_i \geq h^*$ and $y_i = y^L$ attend public university and, in the sub-group of workers with $h_i < h^*$ and $y_i = y^H$, the sub-group $\theta\left(\frac{(1+\beta)w_1+x_3}{\beta\gamma_3}\right)$ remain non-graduates, while the remaining sub-group $\theta\left(h^* - \frac{(1+\beta)w_1+x_3}{\beta\gamma_3}\right)$ attends private university. The sorting behaviour of the remaining sub-group varies across the three different cases, and is summarised as follows:

Case 1: In this case, all workers with $h_i \geq h^*$ and $y_i = y^H$ attend private university.

Case 2: In this case, the fraction $\theta\left(\frac{x_3}{\beta(\gamma_3-\gamma_2)} - h^*\right)$ attends public university and the remaining fraction $\theta\left(1 - \frac{x_3}{\beta(\gamma_3-\gamma_2)}\right)$ attends private university.

Case 3: In this case, the private university tuition fee is so high that no worker with $h_i \geq h^*$ and $y_i = y^H$ finds it optimal to attend private university. Thus, all workers in this sub-group attend public university.

1.2.5.3 Public University's Decision

Recall that public universities set h^* in order exactly to fill the available number of seats, $\bar{U}$. In Appendix A.5, I derive the value of h^* for each of the three possible equilibria. Note that the value of h^* will be highest in Case 3, when demand for public university is highest, and lowest in Case 1, when demand for public university is lowest.

Case 1:

$$h^* = \frac{1-\theta-\bar{U}}{1-\theta} \tag{1.23}$$

Case 2:

$$h^* = 1-\theta-\bar{U} + \frac{\theta x_3^*}{\beta(\gamma_3-\gamma_2)} \tag{1.24}$$

Case 3:

$$h^* = 1 - \bar{U} \tag{1.25}$$

1.2.5.4 Private Universities' Decision

Demand for private university comes from two sub-groups of worker: rich, high-ability workers and rich, low-ability workers. Within each group, willingness to pay is increasing in ability, due to the complementarity between tertiary education and ability in the wage function. Nevertheless, *for a given level of ability*, rich, low-ability workers have a more inelastic demand for private university than do rich, high-ability workers.[4] This is unsurprising, given that they face a more favourable outside option than their low-ability counterparts.

Conditional on exogenous parameter values, the private university can then choose among values of x_3 that either capture both markets (i.e. pooling equilibria) or values that isolate the demand of rich, low-ability workers (i.e. separating equilibria). Equilibria which isolate the demand of rich, high-ability workers, however, do not exist. In Appendix A.6, I derive the private university's equilibrium tuition fee, x_3, for each of the three possible cases.

Case 1: This is a pooling equilibrium in which all rich, high-ability workers and some rich, low-ability workers attend private university. The equilibrium tuition fee is:

$$x_3^* = min\left[\frac{\beta\gamma_3 - (1+\beta)w_1 + c}{2}, \frac{\beta(\gamma_3 - \gamma_2)(1-\theta-\bar{U})}{(1-\theta)}, \frac{\beta\gamma_3(1-\theta-\bar{U})}{(1-\theta)} - (1+\beta)\right] \tag{1.26}$$

Case 2: This is a pooling equilibrium in which some rich, high-ability workers and some rich, low-ability workers demand private university. Appendix A.6 derives the equilibrium tuition fee under different parameter conditions.

Case 3: This is a separating equilibrium, in which only rich, low-ability workers demand private university. The equilibrium tuition fee is:

$$x_3^* = \frac{\beta\gamma_3(1-\bar{U}) - (1+\beta)w_1 + c}{2} \quad \text{or} \quad x_3^* = (1-\bar{U})\beta\gamma_3 - (1+\beta) \qquad \text{or} \quad x_3^* = \beta(\gamma_3 - \gamma_2) \tag{1.27}$$

[4] To see this, compare the slopes of each group's demand function. Rich, high-ability workers' demand is given by $x_3 \leq \beta(\gamma_3 - \gamma_2)h_i$, while rich, low-ability workers' demand is given by $x_3 \leq \beta\gamma_3 h_i - (1+\beta)$. Clearly, the former group has a less steeply-sloping demand curve.

1.2.6 Summary

Thus, conditional on the particular realisation of parameter values, an economy may be in any one of these three cases, where the unique equilibrium is the one that maximises private university profit. Given exogenous parameter values, there is then always a unique equilibrium for the model economy.

Two main predictions emerge from this model framework. The first is that the composition of workers in private universities and, consequently, their average productivity, may vary widely across countries in different cases. The second concerns the relative efficiency of these equilibria. While all three cases involve all agents making optimal decisions given their constraints, efficiency in terms of human capital allocation is decreasing in the distance from Case 1. This is because Cases 2 and 3 describe outcomes in which non-zero fractions of high-ability workers attend public university. Under the assumption that private universities supplement worker productivity by a larger factor than do public universities, graduate-sector output is thus lower than it would be if some of these workers attended private university instead. Thus, there may be a role for governments to subsidise private tertiary education or to seek to adopt private university technology into the public tertiary education sector. Discussing the results from the structural estimation conducted in Section 4 affords an opportunity for considering these predictions, and their implications, in more detail. First, however, I consider an extension of the model framework by allowing for asymmetric information regarding workers' ability.

1.3 Model Extension: Asymmetric Information on Worker Ability

I now relax the assumption of full information by introducing the following information asymmetry into the model framework: employers can no longer observe individual's innate ability, although the distribution from which ability draws are made remains common knowledge. Graduate wages are now set as follows:

$$w_k = \gamma_k E[h_i | i \in n_k, s'] \quad \forall k = \{1, 2\} \tag{1.28}$$

$$w_k = 1 \quad \forall k = \{1\} \tag{1.29}$$

where n_k denotes the graduating cohort of education type k, and s' is the information set of the employer. As individual ability is now unobservable to employers, their perception of the ability of worker i is exactly equal to their perception of the average ability of cohort k

(private university, public university, or non-graduate) from which she is drawn. This in turn is determined by the choice behaviour of different groups of workers. (Common) knowledge of such behaviour is contained in the state variable, s'. Thus, the new state variable for employers is:

$$s' = \{k_a, k_b, k_c, k_d; w_k, h^*, x_3^*, y_i\} \tag{1.30}$$

where, as before, k denotes the optimal education choice $k = \{1, 2, 3\}$ and the subscripts refer to groups of workers undertaking these choices: a describes workers with $y_i = y^L$ and $h_i < h^*$, b describes workers with $y_i = y^L$ and $h_i \geq h^*$, c describes workers with $y_i = y^H$ and $h_i < h^*$, and d describes workers with $y_i = y^H$ and $h_i \geq h^*$.

As before, h_i is a random variable drawn from a uniform distribution: $h_i \sim U[0, 1]$.

The state variable for workers and universities is denoted s'' and contains, in addition to the information in s', knowledge of individual workers' ability endowment:

$$s'' = \{k_a, k_b, k_c, k_d; w_k, h^*, x_3^*, y_i, h_i\} \tag{1.31}$$

As before, it is reasonable to assume that the discounted wage of a public university graduate exceeds the discounted wage of a non-graduate, such that remaining a non-graduate is a strictly dominated strategy for this group of workers. This allows me to solve for the set of "reasonable" equilibria as in the baseline model, specifically excluding any in which high-ability, rich workers prefer being non-graduates to attending free public university.. Formally, this assumption amounts to a public university capacity constraint that satisfies the following condition, which replaces Assumption 6 from the baseline model (see Appendix A.7 for derivation):

Assumption 6$'$:

$$\bar{U} \leq \frac{2(1-\theta)[\beta\gamma_2 - (1+\beta)]}{\beta\gamma_2} \tag{1.32}$$

1.3.1 Characterising the Equilibrium

As before, it is necessary to consider the optimal behaviour of four sub-groups of worker, each of whom has a different feasible choice set, given her endowment of income and ability. These are described in full in Appendix A.8, along with optimal worker sorting behaviour in this model extension. This gives rise, as in the baseline model, to three distinct cases. All low-wealth workers maintain the same optimal behaviour across all three cases, namely that workers with $y_i = y^L$ and $h_i < h^*$ always remain non-graduates, while the sub-group with $y_i = y^L$ and $h_i \geq h^*$ always attends public university. Thus, I now characterise the three candidate equilibria in terms of the sorting behaviour of the two remaining sub-groups: namely, rich workers, of both high and low ability:

Case $1'$**:** In this case, all rich workers (regardless of ability) attend private university.

Case $2'$**:** In this case, all rich, high-ability workers attend public university, while rich, low-ability workers remain non-graduates. This defines the shut-down condition for private universities, as its demand goes to zero.

Case $3'$**:** In this case, all rich, high-ability workers attend public university, while rich, low-ability workers attend private university.

Derivation of the public university's optimal entry requirement and of the private university's optimal tuition fee may be found in Appendix A.9. The question of which choice of tuition fee yields the highest profit and, consequently, which equilibrium results, depends on the relative value of exogenous model parameters. Formally, the unique equilibrium of the economy depends on θ, the fraction of rich workers in the population. If θ is very low (but non-zero), the equilibrium described by Case $3'$ results. If θ is high, the equilibrium described by Case $1'$ results. If $\theta = 0$, we have the equilibrium described by Case $2'$. Formally:

$$x_3^* = \begin{cases} x_3^{*2} & \text{if } \theta = 0 \\ x_3^{*3} & \text{if } 0 < \theta \leq X \\ x_3^{*1} & \text{if } 1 > \theta > X \end{cases} \tag{1.33}$$

where $X \equiv 1 - \frac{\beta\gamma_2\bar{U}}{\beta\gamma_3\bar{U}^2 - 2\beta\gamma_3\bar{U} - 2(1+\beta)(1-\bar{U}) + 2\bar{U}c + 2\beta\gamma_2}$

1.3.2 Summary

As is the case with symmetric information, there are three candidate equilibria, each of which is the unique equilibrium under different relative parameter values. Case $3'$ describes an equilibrium that, like Case 3 in the baseline model, is inefficient in its allocation of human capital, as all high-ability workers attend public university, thereby gaining a relatively smaller productivity boost than they would at private university. Thus, graduate-sector output would increase if some high-ability workers moved to private university. In this model extension, the possibility of the inefficient equilibrium arises from workers being forced to trade-off the benefit of a higher productivity boost from private university against being perceived by employers to be of high-ability by entering public university. In particular, the inefficient equilibrium results when the fraction of rich workers in the population is small, which is likely to be true in many developing countries. As before, this suggests a role for government to deal with the implicit market failure, either by subsidising private tertiary education or by adopting its technology, which would require increased investment into the public tertiary education system.

1.4 Data

1.4.1 Data Description

The data I use is from the World Bank's "Skills Towards Employment and Productivity" (STEP) Skills Measurement household survey, administered between 2012 and 2014 in ten developing countries: Armenia, Bolivia, Colombia, Georgia, Ghana, Kenya, Laos, Sri Lanka, Vietnam and Yunnan Province, China. I drop the last four countries due to their very small sample size of private university graduates, and use only the first six in the analysis that follows.

My sample consists of all individuals who possess an undergraduate degree or who may be classified as a non-graduate. For the purposes of this study, the latter is defined as a person who completed secondary schooling but has obtained no further education. I exclude individuals with missing reports on education, wages, household wealth at age 15, or my chosen measures of ability.

Wealth at the time of tertiary education choice-making is captured by a recall measure of the respondent's family's socio-economic rank at age 15, which ranges from 0 to 10. I refer to this hereafter as the *household rank*. I use two separate measures of ability. The first is a composite quintile score based on the direct reading assessment conducted by the World Bank in conjunction with this survey. I refer to this hereafter as the *ETS score*. The reading assessment consists of a core literacy assessment unit intended to sort the least literate from those with higher levels of reading skill; and a reading components section, which consists of the following three sub-units: vocabulary, sentence processing and passage comprehension. The score I construct here is the aggregation of the individual's quintile score for each of these four measurements. Thus, it ranges between 4 and 20 points. The second, used as a robustness check only in Appendix A.10, is a recall measure which records how respondents feel they compared academically to their peers in the highest primary/secondary grade attended. Response options are "excellent/among the best in class", "above average", "average" and "below average". I use this data to construct a scale of ability which ranges from 1 to 4 points. I refer to this hereafter as the *ability score*. Finally, wages are taken from individuals' reported period wage earnings, and are recalculated in terms of daily, US dollar units.[5]

1.4.2 Summary Statistics

In this section, I present some preliminary summary statistics. This initial presentation is not intended to establish causal relationships; rather, it aims to test whether the data supports the underlying assumptions of the model.

Firstly, the model assumes that private university provides a larger boost to worker productivity than does public university. One explanation for this difference is that private university graduates possess more or better soft skills than their public counterparts - skills that are highly sought-after by private-sector employers in developing countries. Table 1.1 shows that 68 per cent of private university graduates in the STEP dataset work in the private sector, where average graduate wages are highest for this sample of countries, compared to only 37 per cent of public university graduates. Further, Table 1.2 summarises the skills used by graduates in their current jobs. The cognitive skills score measures reading, writing and numeracy on the job. The computing skills score is an aggregation of fifteen binary measures of skills ranging from the use of email, word processing and spreadsheets to programming, graphic design and network management. Finally, the general skills score measures inter-personal relations, autonomy, computer use, tasks that involve learning, and tasks that involve at least 30 minutes of thinking. We see that private university graduates have, on average, higher cognitive, computing and general skills scores than do public university graduates.

Lastly, Table 1.3 shows that private university graduates' hourly average wages (in USD) are nearly twice that of their public university counterparts. In fact, private university graduates in this sample have both a lower unemployment rate and a shorter school-to-work transition than public university graduates, so these differences exist along extensive margins as well as intensive ones. Collectively, therefore, the summary statistics presented here are broadly in line with the model assumption that private university boosts worker productivity more than public university does.

Table 1.4 presents some additional summary statistics for each country, disaggregated by level of education. Here, ability is measured by ETS score and wealth by household rank.

[5] In the case of Armenia alone, I use hourly wage data, due to the unreliability of the daily wage data in this instance. As the analysis that follows makes use only of relative wages, this does not affect the comparability of results across countries.

Table 1.1: Employment Sector by University Type

	frequency	percentage	avg. daily wage (US$)
public graduates			
private sector	466	37.40	14.50
public sector	780	62.60	11.53
Total	1246	100	
private graduates			
private sector	124	67.76	22.84
public sector	59	32.24	19.55
Total	183	100	

Table 1.2: Soft Skills by University Type

	mean	sd	min	max	difference
Cognitive Skills					-0.622***
public	5.65	2.14	0	9	
private	6.28	1.76	0	9	
Computing Skills					-1.801***
public	2.56	3.65	0	15	
private	4.39	4.14	0	15	
General Skills					-1.044***
public	10.03	2.62	0	15	
private	11.09	2.51	0	15	

Table 1.3: Daily Wages by Education Type

	mean	sd	min	max
public university graduate	12.48	10.40	.14	95.45
private university graduate	21.18	16.43	.26	90

Table 1.4: Summary Statistics by Country and Level of Education

	private graduate	public graduate	non-graduate
Armenia			
proportion (full population)	0.01	0.37	0.62
proportion (grad. population)	0.03	0.97	-
hourly wage (USD)	1.36	1.64	1.28
ability	7.46	7.62	7.49
wealth	5.86	6.72	5.75
N	24	777	1306
Bolivia			
proportion (full population)	0.05	0.09	0.85
proportion (grad. population)	0.36	0.64	-
daily wage (USD)	12.93	7.90	2.87
ability	7.88	7.81	6.25
wealth	5.33	4.97	4.22
N	76	134	1230
Colombia			
proportion (full population)	0.06	0.03	0.91
proportion (grad. population)	0.62	0.38	-
daily wage (USD)	14.48	14.07	3.32
ability	7.98	8.08	6.79
wealth	5.42	4.75	4.18
N	121	73	1968
Georgia			
proportion (full population)	0.02	0.20	0.78
proportion (grad. population)	0.11	0.89	-
daily wage (USD)	3.35	3.46	1.06
ability	6.88	7.01	6.59
wealth	6	6.01	5.55
N	34	274	1092
Ghana			
proportion (full population)	0.01	0.07	0.92
proportion (grad. population)	0.14	0.86	-
daily wage (USD)	4.83	5.61	0.73
ability	11.32	11.39	7.52
wealth	6.28	5.41	4.87
N	25	156	2095
Kenya			
proportion (full population)	0.01	0.08	0.92
proportion (grad. population)	0.08	0.92	-
daily wage (USD)	13	12.45	1.85
ability	7.9	7.93	6.44
wealth	6.3	5.21	4.46
N	20	230	2722

1.4.3 Education Choice Regression and Case Predictions

The summary statistics allow us to make some (limited) predictions about the outcome of each country. We expect countries with a positive private-public wealth differential, a negligible private-public ability differential and a positive private-public wage differential to be closer to Case 1. This is because average wages for private graduates rise when rich, high-ability workers attend private university, and wealth rather than ability drives sorting across university types. Equally, we expect countries with a negligible private-public wealth differential, a negative private-public ability differential and a negative private-public wage differential to be closer to Case 3. This is because average wages for private graduates fall when rich, high-ability workers attend public university, and ability rather than wealth drives sorting across university types. We cannot, however, easily distinguish a Case 2 county from these other cases.

Table 1.5: Wage Ratios and Predicted Cases

	Predicted Case	Private Grad/Public Grad Wage Ratio
Bolivia	1	1.64
Colombia	1	1.03
Kenya	1	1.04
Armenia	3	0.82
Georgia	3	0.97
Ghana	3	0.86

Bolivia, Colombia and Kenya have a positive private-public graduate wage differential, while Armenia, Georgia and Ghana have a negative one, as Table 1.5 shows. Thus, we expect the former group to be closer to Case 1 and the latter to be closer to Case 3. I use a multinomial logit specification to estimate tertiary education choice for the two groups, and to test whether the coefficients on ability and wealth are in line with the predictions above. The results, using public university as the base outcome, are provided in Table 1.6. The key independent variables in terms of model assumptions are ability and income, and I include a set of control variables, including a full set of country dummies.

First, note that the coefficients on ability and wealth for non-graduates are negative and significant, as expected. Second, for the choice between private and public university, the coefficient on ability is positive and significant for both Case 1 and Case 3 countries, but the latter coefficient is larger, in line with the predictions above. The coefficient on wealth is positive and significant for Case 1 countries, but zero for Case 3 countries, as predicted. Thus, the reduced-form results support the model's prediction of different equilibria across countries.

Table 1.6: Education Choice Regression

Dependent Variable:	highest education level	
	(1)	(2)
	case 1 countries	case 3 countries
non-graduate		
ability	-0.936***	-1.087***
	(-15.15)	(-22.20)
wealth	-0.247***	-0.114***
	(-8.09)	(-5.95)
private university graduate		
ability	-0.283**	-0.400**
	(-2.89)	(-2.80)
wealth	0.144**	-0.004
	(2.96)	(-0.06)
country dummies	Yes	Yes
Observations	6333	5481

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

controls: household size at age 12, gender, age

1.5 Estimation and Results

1.5.1 Static Model Estimation

The model framework consists of workers who are heterogeneous in terms of ability and initial wealth. Thus, I begin by defining corresponding cut-offs for wealth and ability in the data (see Appendix A.11 for details).

Estimation of the unknown parameters occurs in three steps. Firstly, I set some parameters equal to standard values in the literature or to those taken directly from the data. Table 1.7 summarises the values used. Real interest rates for each country are obtained from World Bank Development Indicators data.[6] The capacity of public universities is set to equal the proportion of public university graduates in each country sample. The fraction of rich workers is set to equal the fraction of the country sample whose value of the household socio-economic rank variable is greater than or equal to 7, as explained in Appendix A.11.

Secondly, I estimate the main parameters of the model independently for each country using simulated method of moments (SMM). The chosen estimator thus minimises the distance

[6]Real interest rate data for Ghana is missing, so I replace it with the average of the other five countries' values.

Table 1.7: Fixed Parameters

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
discount factor, β	0.4	0.4	0.4	0.4	0.4	0.4
real interest rate, r	0.13	0.08	0.09	0.08	0.09	0.08
public university capacity, $\bar{U}$	0.37	0.09	0.03	0.20	0.07	0.08
fraction of initial rich, θ	0.40	0.11	0.12	0.32	0.20	0.11

between a set of selected moments from the data and the moments generated by simulation of the model (McFadden (1989)). The SMM estimator takes the form:

$$\hat{\theta}_{S,T}(W) = argmin_{\theta}[\hat{\mu}^d_T - \hat{\mu}^s_{S,T}(\theta)]'W_T[\hat{\mu}^d_T - \hat{\mu}^s_{S,T}(\theta)] \tag{1.34}$$

where $\hat{\mu}^d_T$ is a vector of moments from the data with T observations, $\hat{\mu}^s_{S,T}$ is the corresponding vector of moments from S simulations of T observations, and W_T is a weighting matrix. Here, I set $S = 1$ and W_T is the identity matrix.

The parameters to be estimated structurally are the university's marginal cost (c), the productivity contribution of private university (γ_3) and the productivity contribution of public university (γ_2). To identify these three parameters, I use the following five moments: the proportion of non-graduates in the sample, the proportion of high-ability workers in private university, the proportion of rich workers in public university, the ratio of private university graduate wages to public university graduate wages, and the ratio of private university graduate wages to non-graduate wages.

Table 1.8 shows the values of these moments for each country, alongside the corresponding simulated values from the estimation. The "distance" row shows the value of the minimisation criterion in each case.

Table 1.9 displays the resulting parameter estimates. Importantly, these results are in line with my initial hypothesis that private universities provide a larger productivity boost than do their public counterparts (i.e. $\gamma_3 > \gamma_2$). The only exception is Ghana, where the estimated value of γ_2 is slightly higher than the corresponding estimate of γ_3.

Finally, I compute the remaining parameters as follows. The initial wealth of rich workers is set at a value just above the highest optimal tuition fee charged by private universities in any equilibrium; this is to ensure that the choice of wealth level does not determine the equilibrium. The initial wealth of poor workers is set to be equal to the value of the private university's marginal cost, estimated structurally. Finally, the lump-sum tax is equal to the equilibrium fraction of public university graduates, multiplied by the marginal cost of

Table 1.8: Matched Moments

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
prop. non-grads (data)	0.62	0.85	0.91	0.78	0.92	0.92
prop. non-grads (sim.)	0.62	0.86	0.91	0.78	0.92	0.86
prop. rich public grads (data)	0.52	0.08	0.12	0.43	0.22	0.2
prop. rich public grads (sim.)	0.41	0.08	0	0.35	0.15	0
prop. high-ability private grads (data)	0.17	0.28	0.44	0.01	0.16	0.25
prop. high-ability private grads (sim.)	0	0.28	0.06	0	0	0.15
private/public grad wage ratio (data)	0.82	1.64	1.03	0.97	0.86	1.04
private/public grad wage ratio (sim.)	0.80	1.64	1.08	0.97	0.86	1.07
private grad/non-grad wage ratio (data)	1.06	4.51	4.34	3.17	6.62	7.03
private grad/non-grad wage ratio (sim.)	1.07	4.51	4.33	3.17	6.62	7.03
distance	0.04	0.00	0.16	0.01	0.03	0.05

Table 1.9: Estimated Structural Parameters

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
university marginal cost, c	0.10	2.4	0	2.1	6.3	0.07
private university productivity, γ_3	5.94	19.53	21.00	14.66	26.25	34.86
public university productivity, γ_2	5.69	10.20	14.37	12.69	27.89	24.00

educating each graduate and divided by the size of the country sample. The results are shown in Table 1.10.

Table 1.10: Fixed Parameters

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
initial wealth of rich, y_H	10	10	10	10	10	10
initial wealth of poor, y_L	0.10	2.4	0	2.1	6.3	0.07
lump-sum tax, τ	0.04	0.22	0	0.39	0.43	0.01

1.5.2 Multiple Cases and Efficiency

The model predicts that any one of three possible cases may manifest in a particular country, conditional on its exogenous parameter values. In the basic model with observable ability, these are as follows: in case 1, all high-ability, rich workers and some low-ability, rich workers above a threshold level of ability attend private university, while poor, high-ability workers attend public university and poor, low-ability workers remain non-graduates. In case 2, behaviour is unchanged except that a fraction of the high-ability, rich workers now attend public rather than private university. In case 3, behaviour is again unchanged from case 1 except that all high-ability, rich workers attend public rather than private university. Recall that these different cases result directly from the optimal tuition fee set by the private university, which in turn depends on the relative value of the productivity boosts offered by private and public university (i.e. the values of γ_3 and γ_2) and on the value of its marginal cost parameter.

The results from the structural estimation highlight this wide variability of outcomes across countries as a consequence of variation in exogenous parameter values: Colombia and Kenya are in case 1, Bolivia is in case 2, and Armenia, Georgia and Ghana are all in case 3, as Table 1.11 shows.

Table 1.11: Cases

	Case
Armenia	3
Bolivia	2
Colombia	1
Georgia	3
Ghana	3
Kenya	1

The existence of multiple equilibria across countries is important primarily because, as discussed earlier, the resulting tertiary education outcomes deviate to varying degrees from the social planner's optimum. Specifically, in all countries for which $\gamma_3 > \gamma_2$, case 1 is the most, and case 3 the least, desirable from the social planner's perspective.

To see this more clearly, consider the output in Table 1.12. Here, I consider the outcomes in each country under case 1 and case 3, respectively. To do so, I set the tuition fee for case 1 equal to its highest possible level, and the tuition fee for case 3 equal to its lowest possible level, thus obtaining a lower bound on the size of the efficiency loss manifest by moving from case 1 to 3. I calculate aggregate output under each of these equilibria, and then compute the percentage deviation from aggregate output under the social planner's optimal allocation of education (see Appendix A.12). It is clear that the movement from case 1 to 3 within each country (excluding Ghana) involves an increase in the deviation from output under the social planner's optimum. This occurs through two channels: a fall in the total number of graduates and (potentially) a fall in the productivity of existing graduates.

Thus, differences in parameter values across countries affect the efficiency of resulting equilibria, with the size of the efficiency loss increasing as the economy moves further away from case 1 if $\gamma_3 > \gamma_2$.

The grouping of countries into separate equilibria is interesting for a second reason, namely that it shows very clearly the different roles private universities may adopt in different

Table 1.12: Measuring Efficiency Loss

country	equilibrium	tuition fee	% Δ from benchmark output	total grads.	output per grad.
Armenia	1	-0.48	-5.68	1022	4.37
Armenia	3	0.10	-10.32	786	4.66
Bolivia	1	3.18	-65.10	199	11.65
Bolivia	3	3.55	-68.01	173	10.79
Colombia	1	2.54	-70.22	199	14.68
Colombia	3	2.64	-70.76	190	14.63
Georgia	1	0.64	-27.69	568	10.48
Georgia	3	0.90	-37.10	455	10.41
Ghana*	1	-	-	-	-
Ghana	3	0	-57.55	545	17.44
Kenya	1	3.97	-71.64	407	23.63
Kenya	3	4.34	-73.16	377	23.25

*equilibrium 1 impossible in Ghana, which has $\gamma_3 < \gamma_2$

economies. In case 1 countries, for instance, private universities are clearly the preferred option for tertiary education across all levels of ability, and only wealth constraints prevent public university applicants from switching to private university. At the other end of the spectrum, private universities in case 3 countries are populated solely by wealthy workers of relatively low ability, while all the most able workers in the economy attend public university. Finally, in case 2 countries private universities manage to attract both high- and low-ability workers, yet there remains a sub-group of the wealthy that finds it optimal to attend public university instead, such that public universities receive a mix of students from across the income distribution. The ability to predict such a diverse range of outcomes is one of the most important features of the model framework, and is central to the analysis presented in the remainder of this section.

1.5.3 Interpreting Results

The differences in parameter values across countries is also important for another reason: namely, that they underpin wide variation in the wages of workers with different types of education, as well as in the cost of private tertiary education. Table 1.13 shows that the estimated private university tuition fee as a percentage of rich workers' initial wealth ranges from a mere 1 per cent in Armenia to 73 per cent in Ghana. Equally, the average private graduate wage premium over non-graduate pay ranges from 7 per cent in Armenia to 603 per cent in Kenya. There is also a large amount of variation in the difference between average private and public graduate wage rates in different countries: in Armenia, the average private graduate earns approximately 20 per cent less than her public university counterpart while, in Bolivia, she earns 63 per cent more than the average public university graduate.

Table 1.13: Comparing Private University Costs and Wage Premia

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
tuition fee as % of y_H	1.00	32.23	25.51	27.06	73.41	39.67
private grad./non-grad. wage premium (%)	7.14	346.69	332.64	216.79	561.64	602.63
private grad./public grad. wage premium (%)	-19.63	62.59	8.63	-3.22	-13.90	7.31

1.5.4 Model Dynamics

In order better to understand the impact of the expansion over time of private universities in developing countries, it is necessary to consider the evolution of these model economies in a dynamic setting. Unfortunately, the STEP survey provides only a single cross-section of data, so it is not possible to obtain structural parameter estimates based on actually observing the evolution of the moments in the data. Nevertheless, it is still instructive to compare outcomes over time across the STEP countries using the parameter estimates obtained from the cross-section, while recognising the limitations of such an approach. This is because such an exercise allows me to highlight the fact that the expansion of private universities in developing countries does not have a clear dynamic effect on aggregate productivity and income inequality. Rather, such effects depend crucially on relative parameter values, as will be demonstrated in this section. Thus, the results in this section should not be read as predictive of the future outcomes of these particular countries but, rather, as a demonstration of the dynamic outcomes one might expect in economies characterised by the state variables we observe in the cross-sectional data.

I proceed by using the structural parameter estimates from Table 1.9 to run the OLG model over six generations, independently for each country as in the static case.

1.5.4.1 Generational Equilibria and Social Efficiency

An interesting and important result that emerges from this exercise is that countries move away from case 1 over time. Specifically, of the three countries that begin in case 3, two remain there in every generation and, in the third (Armenia), private universities cease entirely to operate. By contrast, the two countries that begin in case 1 move to case 2 very quickly, while Bolivia, the only country to begin in case 2, moves briefly to case 3 but then returns to case 2 in the next generation, as shown in Table 1.14. Thus, case 3 is persistent in a manner that case 1 is not. This is important because, as just discussed, case 1 is the closest to, and case 3 the furthest from, the social planner's optimal allocation when $\gamma_3 > \gamma_2$. Ghana, as earlier noted, is an exception to this rule, as it is the only one of the six countries for which public universities provide a greater contribution to productivity than do their private counterparts; thus, case 3 is both persistent and desirable here. The first main result, therefore, is that for countries in which private universities are more productivity-enhancing than public ones, efficiency weakly declines over time.

Table 1.14: Generational Equilibria

	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 5	Gen. 6
Armenia	3	-	-	-	-	-
Bolivia	2	3	2	2	2	2
Colombia	1	2	2	2	2	2
Georgia	3	3	3	3	3	3
Ghana	3	3	3	3	3	3
Kenya	1	2	3	2	2	2

A second important result is that there is wide variation in the success of the private university sector across countries: the sector may remain small, expand significantly or collapse entirely. This is, in fact, mirrored in the observed trajectory of private universities across developed countries. Consider, for instance, Figure 1-1, which shows enrolments in private universities as a percentage of total tertiary education enrolments across the OECD at two points in time: 2003 and 2011. First, there is large variation in the 2003 cross-section, and second, between 2003 and 2011, there is movement in different directions - the sector shrinks in Portugal and Switzerland, for instance, but grows in New Zealand and the Slovak Republic.

Similarly, in this simulation using data from the STEP countries, private university enrolment grows strongly in Kenya and Colombia, remains at a low, constant level in Ghana, Bolivia and Georgia, and collapses to zero in Armenia, with the private university shutting down in the second generation in the latter country, as Figure 1-2 shows.

1.5.4.2 Output Per Worker

Now, having established that dynamic outcomes for private universities vary widely across the STEP countries, I turn my attention to the question of what this means for productivity and income inequality.

First, Figure 1-3 shows the path of output per worker for each of the six countries. The two countries which experienced strong growth in the private university sector - Kenya and Colombia - also display much higher output per worker than other countries. Bolivia, despite being in the same case (i.e. 2) as Kenya and Colombia fails to take off in either dimension. This is because, despite having similar values of private and public university productivity to

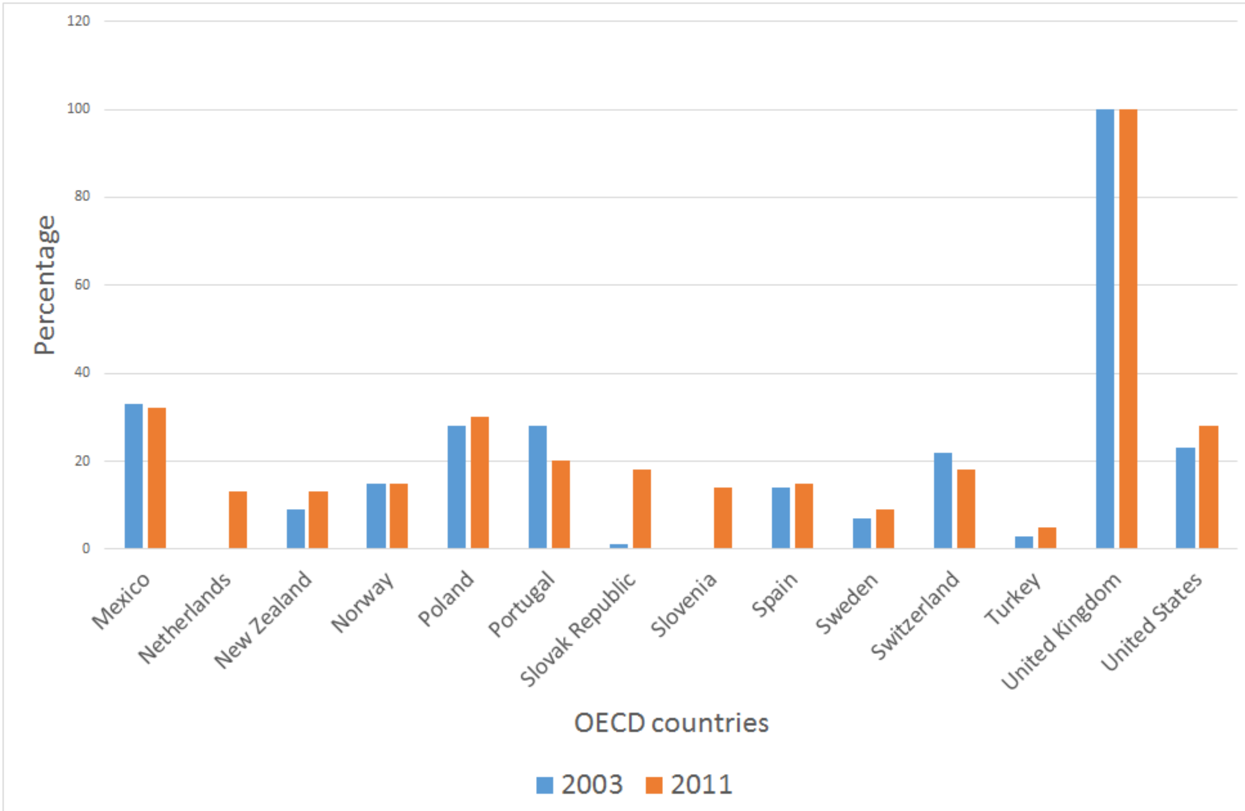


Figure 1-1: Percentage of Students in Independent or Government-Assisted Private Tertiary Education

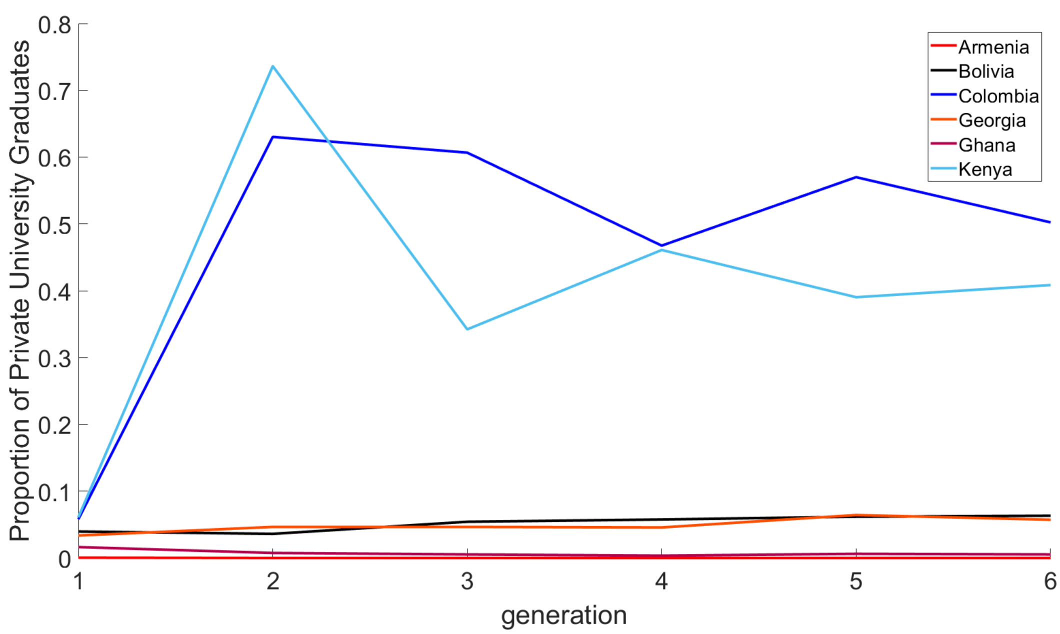


Figure 1-2: Proportion of Private University Graduates

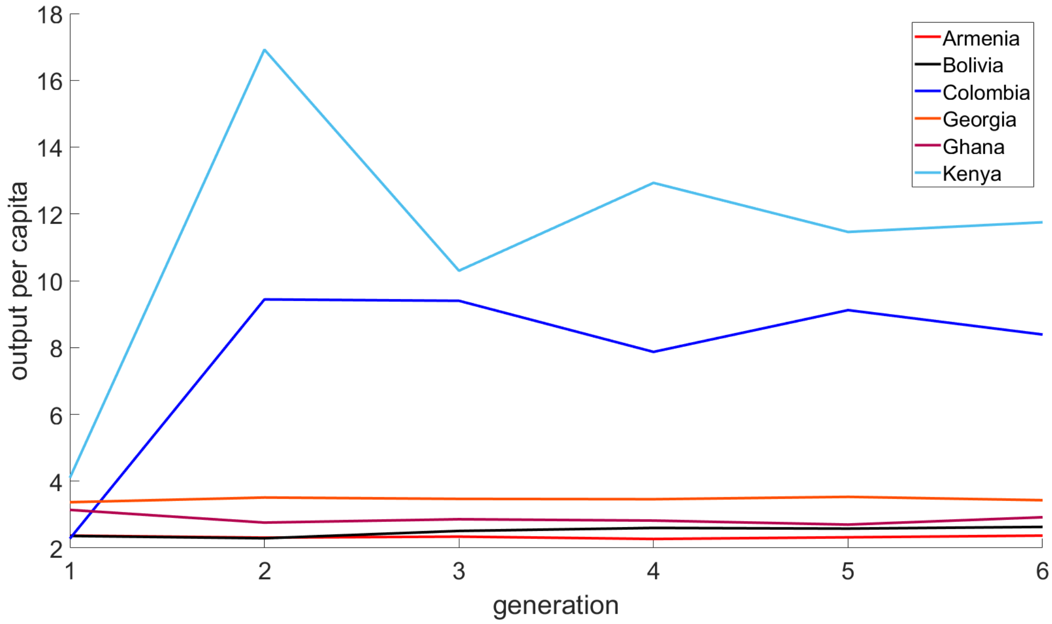


Figure 1-3: Output per Capita

Colombia, it has a higher marginal cost (see Table 1.9) and so, must set a higher equilibrium tuition fee. This constrains the growth of private university enrolment and, consequently, output per worker. The three case 3 countries remain at a constant, low level of output per worker (and private university enrolment) over time.

Correspondingly, Figure 1-4 shows the percentage deviation of output per worker in each country from a counterfactual scenario in which private universities do not exist. Output per worker is higher than in the counterfactual case for all countries except Armenia where, since private universities shut down after the first generation, the counterfactual is equal to the observed result. This is unsurprising, given that the emergence of private universities increases the total number, and (except in Ghana and Armenia) the productivity, of graduates. Thus, the benefits of private universities derive from two separate (and complementary) channels. More interesting, however, is to observe that the gap between observed and counterfactual output per worker is, as before, largest in Colombia and Kenya.

A third result emerging from the dynamic simulation, therefore, is that the effect of private universities on output per worker varies widely, and a large positive effect depends crucially on whether γ_3 and γ_2 (and, indeed, their difference) is sufficiently large, and the marginal cost parameter sufficiently small.

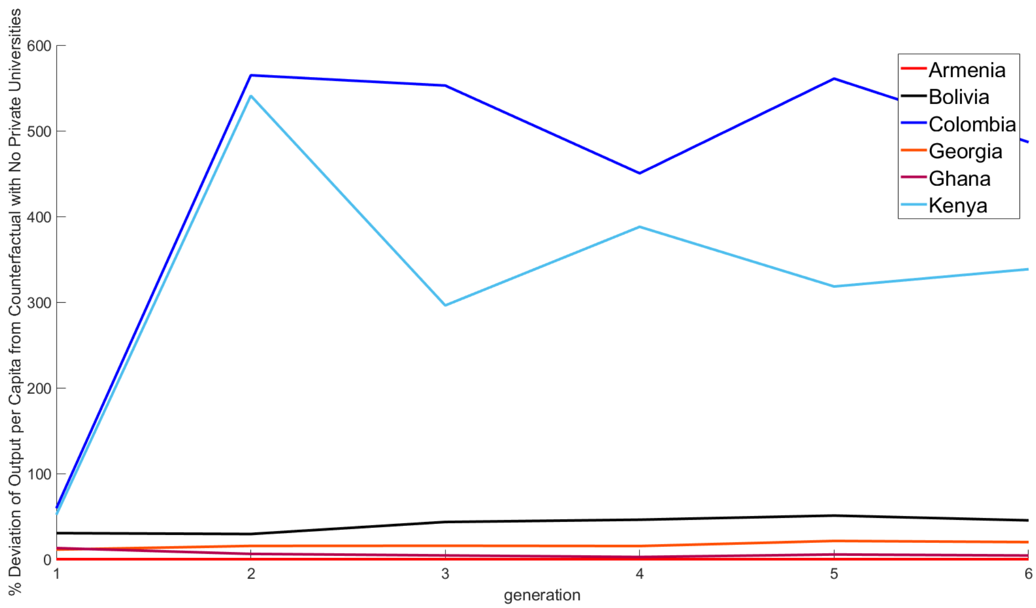


Figure 1-4: Output per Capita: Counterfactual Comparison

1.5.4.3 Workers' Income Inequality

Private education is often associated with perpetuating income inequality. While it is true that, in all STEP countries in which private universities survive, income inequality is higher compared to the counterfactual of no private universities, the reality is more nuanced, as Figure 1-5 shows. This figure plots the percentage deviation of the Gini coefficient from its counterfactual value in the case of no private universities for each of the STEP countries. In Ghana, for instance, inequality is high and rising over time in both cases, and there is very little difference between the case with and without private universities. As with output per capita, the gap between the counterfactual and observed Gini coefficient is highest in Kenya and Colombia, where private universities hold greater sway over the economy.

A caveat regarding these results is that this simple framework cannot model the manner in which the graduate wage premium may respond to the graduate labour supply expansion caused by the arrival of private universities. Studies that consider such effects for the U.S. and for OECD countries have shown, however, that graduate wages, rather than falling in response to such an increase in graduate labour supply, have remained remarkably constant. For instance, Barth and Lucifora (2006) find a negligible change in the graduate premium over the period 1985-2000, when the tertiary education supply grew on average by 20 per cent each year. Similarly, Machin and McNally (2007) show that the graduate wage premia for the UK and the US actually increase slightly over 1980-2004, despite similar increases in tertiary education supply. Both sets of authors argue that this is due to a large and

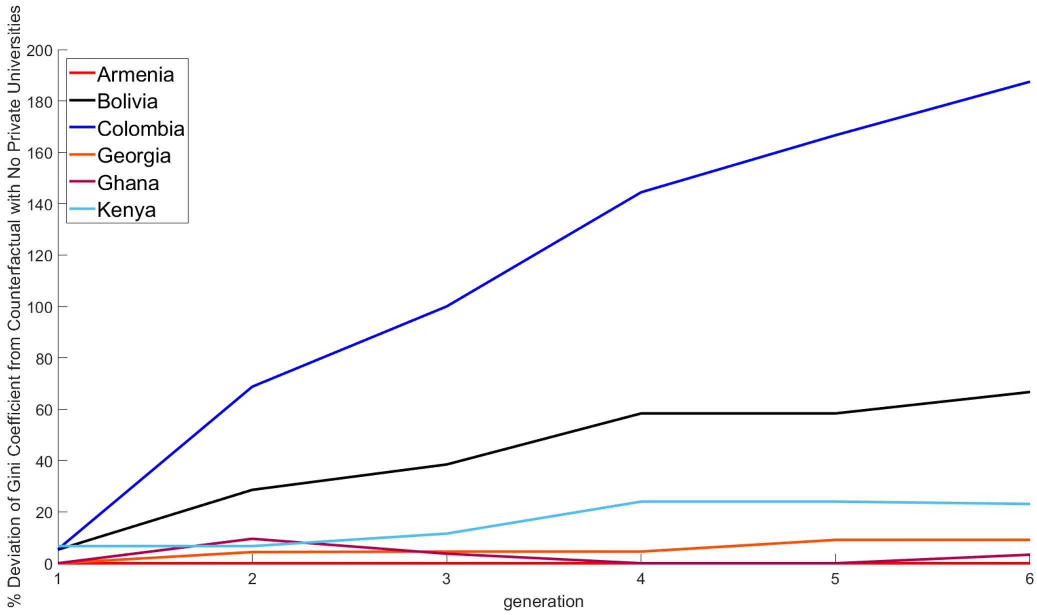


Figure 1-5: Gini Coefficient for Workers: Counterfactual Comparison

offsetting increase in graduate labour demand, driven by skill-biased technological change. It may similarly be argued that structural transformation and technological advancement in developing countries over the period in which private universities have appeared there implies a similar labour demand shift to keep returns high. Indeed, Machin (2002) argues that "the shifts in skill demand in the developing world appear to be correlated with the shifts seen in the developed world", while Ra et al. (2015) provides further evidence of this for Asia in particular. Otherwise, we would expect the effects shown in this section to be somewhat muted by a fall in graduate wages: evidence from the OECD suggests an extensive-margin elasticity of about 2 in the absence of any demand-side effects (Barth and Lucifora (2006)).

Looking now across the STEP countries, we would expect that workers' income inequality is higher in those with a larger private graduate wage premium and a higher marginal cost for private universities (the latter because this drives up the optimal tuition fee, thus reducing the number of graduates relative to non-graduates). The capacity of the public university is also likely to play a role in this respect. Figure 1-6 broadly supports this interpretation. The Gini coefficient is highest in Kenya, the country with the highest private graduate to non-graduate wage ratio in the data. Armenia begins with the second-highest value (somewhat mechanically, as it has the highest fraction of wealthy workers in the static model), but quickly becomes the country with the lowest level of inequality, in keeping with its very low graduate wage premium. Conditional on having similar graduate wage premia, the university's marginal cost parameter becomes relatively more important, broadly determining

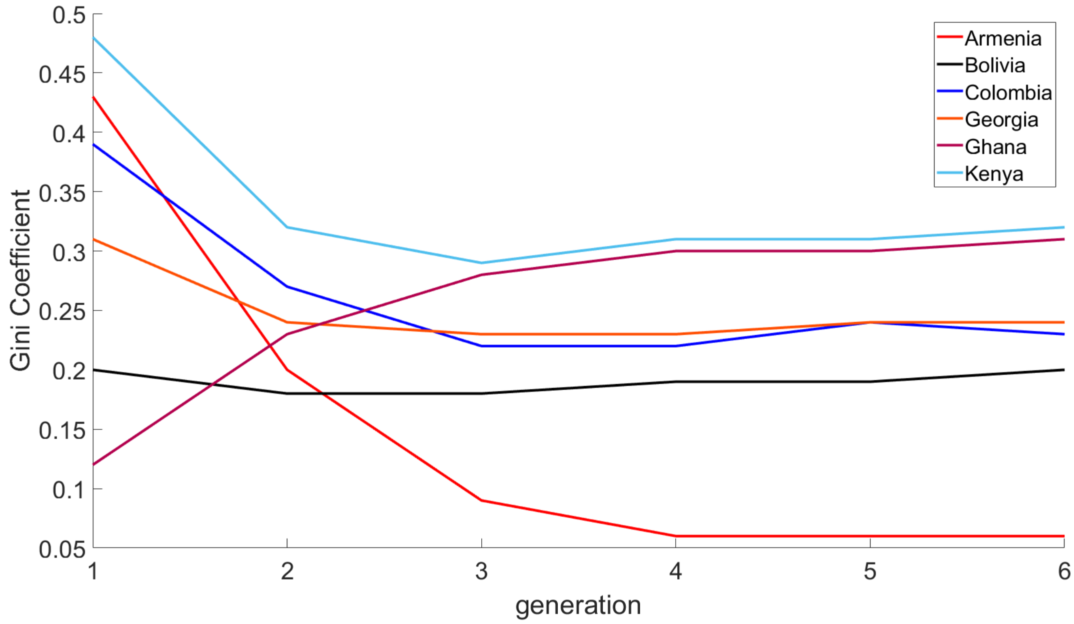


Figure 1-6: Gini Coefficient for Workers

the ordering of the remaining countries here.

Workers' inequality decreases over time in all countries but Ghana, where it is increasing at a decreasing rate. Importantly, this difference is not driven by the fact that $\gamma_2 > \gamma_3$ in Ghana (the effect persists if I switch the values of the two parameters). Rather, this is simply a feature of a case 3 country with a sufficiently high value of public university productivity. The reason for this is that, in case 3 countries, many wealthy graduates now attend public university, at zero cost. If the public university wage premium is sufficiently high, the first generation rich will leave a bequest larger than its own initial wealth, causing income inequality to increase over time. Thus, in cases where the private university tuition fee is high, the profit-maximising behaviour of private universities hinders the growth of income inequality, by capturing a large portion of rents which would otherwise accrue to graduates and their descendants.

Thus, while having private universities weakly raises workers' income inequality relative to the counterfactual with no such institutions, the relationship between private university enrolment and income inequality is not unambiguous. Specifically, with low enrolment, as is the case in case 3 countries, the gap between observed and counterfactual income inequality will be smaller but, if the graduate wage premium is sufficiently high, the limited power of private universities to capture excess rents may lead to an increase in income inequality over time.

A final result emerging from this section, therefore, is that, as with output per capita, the effect of private universities on workers' income inequality is always positive compared to the counterfactual with no private universities, but the magnitude of the effect varies widely across countries, and is driven primarily by the size of the private graduate wage premium, but also by the size of the university's marginal cost parameter. Finally and importantly, it is *not* simply the case that countries with higher private university enrolments (i.e. case 1 countries) are more unequal than countries in case 3, as the latter may experience growth in inequality over time.

1.6 Conclusion

This paper has explored the effect of private universities operating in developing countries on human capital allocation, aggregate productivity and workers' income equality. I have shown, firstly, that the growth experience of private universities may vary dramatically across countries: even in this small sample, we observe these institutions growing strongly, remaining at a low, relatively constant level, or shutting down entirely over time in different countries.

Secondly, I have demonstrated that the worker composition of these universities also varies across countries. There are three possible cases: in case 1, private universities attract wealthy workers of all abilities, in case 2, some wealthy workers of relatively high ability attend public university but others remain at the private one and, in case 3, the private university consists solely of low-ability, wealthy workers, while all high-ability workers attend public university.

Thirdly, private universities always increase output per worker, by raising the number and (if the difference between private and public university productivity parameters is sufficiently positive) the productivity of graduates. Nevertheless, the size of the effect is heterogeneous across countries, and is largest when the productivity parameter of private and public universities (and their difference) is large, while the marginal cost of tertiary education is small.

Fourthly, workers' income inequality is weakly higher than the counterfactual with no such institutions, but the relationship between private university enrolment and inequality is nuanced - at low levels of enrolment, the gap with the counterfactual is smaller, but the limited ability of private universities to capture excess rents may lead to income inequality increasing over time if the graduate wage premium is sufficiently large.

While the specific results of the model are considered here only in relation to a small group of developing countries, the central premise that small differences in underlying parameter values underpin large differences across countries in the growth of private universities, their cohort composition, and their effects on output and inequality may be applied more generally to explain observed patterns in other developing countries across the world.

Further research may use longitudinal data to consider how deep structural parameters may vary over time, and extend the model to deal with more complex interactions, such as how a non-competitive labour market - in which workers' wages no longer are constrained by marginal product - or cohort effects, may affect workers' sorting behaviour.

Chapter 2

A Job Worth Waiting for: Parental Wealth and Youth Unemployment in Ghana

Abstract

Youth unemployment in Ghana increases in parental wealth. This occurs because, in the absence of unemployment insurance, only workers with a sufficiently high stock of parental wealth can afford to remain unemployed, and do so in order to search for scarce, high-productivity jobs. I build and estimate a structural model of endogenous education, employment and occupational choice to quantify this effect, and to demonstrate that it leads to a number of undesirable labour market outcomes: namely, low educational attainment, high income inequality and low match efficiency among workers of heterogeneous ability. I use the estimation results to decompose the effect of wealth on average lifetime earnings into an education channel and an unemployment channel, and show that the latter accounts for 37% of the total effect. Further, I use the estimated model to compare the effectiveness of two alternative policy interventions: an education subsidy and unemployment insurance. I find that the education subsidy is most effective at increasing aggregate productivity, but comes at the cost of increasing income inequality, while unemployment insurance has a smaller effect on aggregate productivity but also decreases income inequality.

Keywords: Youth Unemployment, Occupational Choice, Human Capital Investment, Credit Constraints, Unemployment Insurance

JEL codes: J24, I24, J64

2.1 Introduction

The first-time transition from school to the labour force often involves a period of searching for a "good" job offer, regardless of where in the world it takes place. In many developing countries, however, institutional failures such as search and information frictions, and a shortage of desirable employment options can severely protract the required waiting time, while a lack of credit or unemployment insurance renders long periods of unemployment more costly. The outcome of these tensions is high dispersion and low average productivity among accepted jobs. In Sub-Saharan Africa, such a state of affairs bears a particularly high economic cost, as the region has the youngest population of any in the world: approximately 70% of its citizens are under 30 years of age, and the ILO estimates that 11 million young people will enter the labour market each year over the next decade. Consequently, the question of how best to channel youth into stable and productive employment is an important consideration for researchers and policy-makers alike.

The primary focus of existing research and policy dialogue concerning youth in this region has been unemployment (see, for instance, Baah-Boateng (2013) and Baah-Boateng (2016)). This is epitomised by the president of Ghana who, in his 2018 State of the Nation address, declared that "the number of young people who cannot find work is staggering and a threat to our national security".[1] Nevertheless, there is an informal consensus among academics and policy-makers that, in Sub-Saharan Africa, unemployment is often a luxury that few can afford (Hart (1973), Udall and Sinclair (1982) and Fox et al. (2016)).[2] In this paper, I use data from Ghana to present a set of stylised facts that formalise this view. I show that youth unemployment in this region is relatively low; consequently, the low productivity of youth employment represents a much graver and more salient concern than unemployment.[3] Further, I demonstrate that the probability of youth unemployment is increasing in parental wealth: a 1% rise in parental assets is associated with a 0.08 percentage point increase in the probability of unemployment among 15-29 year-old Ghanaian males. This relationship arises precisely because the scarcity of high-productivity employment requires workers to wait (often several years) for a "good" job. Youth whose families do not have a large stock of assets to draw down will then accept worse jobs in order to avoid a prolonged period of unemployment. In the absence of unemployment insurance these effects are likely to be

[1] https://www.dw.com/en/ghana-president-vows-to-step-up-fight-against-youth-unemployment/a-42519951, last accessed 26 August 2018

[2] See also Serneels (2007) for a treatment of this subject in the context of workers queueing for public-sector employment in urban Ethiopia. He argues, however, that unemployment is concentrated among the middle-classes, rather than the wealthiest strata of society.

[3] This is in line with Falco and Teal (2012), who argue that, for young job-seekers in urban Ghana, the binding constraint is not a shortage of jobs, but the low wages offered by many of them.

particularly pronounced. Finally, I show that educational attainment is low, a fact that may also be explained by an inability to reap the returns of education when they must be preceded by a costly stint in unemployment.

I build and structurally estimate a model that quantifies these effects and demonstrates that the positive link between parental assets and unemployment has negative implications for education choices, income inequality and aggregate productivity. Individuals in the model are heterogeneous in initial family assets and ability. They make decisions over whether or not to invest in education, which sector of employment to accept work in, and how long to remain unemployed after the end of education. A tension between the desire for higher consumption while unemployed and the desire to save in order to fund a continued search for a better job underlies, and motivates, these decisions. The consequence of this is a positive relationship between parental wealth and unemployment, which results in youth from wealthier families obtaining higher-productivity jobs, on average, than their poorer counterparts. For the latter, this often implies a career in low-productivity agriculture or small-scale, low-capital entrepreneurship. In particular, when the expected return to education is low and earnings in high-education jobs are relatively dispersed, there is a high degree of income inequality, low educational participation and an inefficient allocation of workers to jobs, as low-wealth, high-ability workers end up in low-productivity employment.

I use the estimated model to decompose the effect of parental wealth on a worker's average lifetime earnings into two channels: higher education and greater willingness to wait in unemployment for a good job. I show that, for Ghana, the former accounts for 63% of the total effect, while the latter contributes the remaining 37%. The model also provides a framework within which to consider counter-factual policy experiments. Accordingly, I use it to compare the effectiveness of two alternative policy interventions: an education subsidy and unemployment insurance. I find that the education subsidy is most effective at improving aggregate productivity – a subsidy of 20% increases productivity by 25%, but also increases income inequality by 6%. Unemployment insurance of the same total value increases aggregate productivity by only 1%, but also reduces income inequality by 1%. The relatively small effects generated by unemployment insurance are due to the fact that the take-up rate of this policy is higher than that of the education subsidy. With a fixed budget, this results in a relatively small per-person payment compared to that of the latter policy. Differences in the size of the productivity effect across the two policies are also due to gains from raising the aggregate level of education outweighing the gains from extending unemployment (while holding education constant) in order to access higher-productivity employment. The opposing effects on equity are due to workers in the bottom 20% of the initial wealth distribution being unresponsive to the education subsidy, such that the former policy disproportionately benefits wealthier individuals, while the latter disproportionately benefits the bottom of the wealth distribution. Thus, the optimal policy prescription depends on a government's relative preferences over efficiency and equity.

This paper draws on, and brings together, three related strands of existing literature: research on the role of assets in determining labour market outcomes, research on education choice and its implications for labour market search, and research on unemployment in developing-country contexts.

In the first of these strands, Danforth (1979) is one of the earliest papers to analyse job search with risk-averse workers, and shows that reservation wages are increasing in wealth.[4] Eeckhout and Sepahsalari (2015) come closest to the spirit of this paper: they set up a search-and-matching framework with wage-posting, in which workers choose whether to direct their search towards low- or high-productivity jobs. There is a trade-off between job productivity and the probability of receiving an offer in each sub-market, such that initial assets determine workers' sorting behaviour across job types. They demonstrate the conditions under which workers with low levels of initial assets sort into low-risk, low-productivity jobs, while high-asset workers prefer jobs with high levels of both risk and productivity, and unemployment duration increases in initial asset-holdings. Relatedly, Herkenhoff et al. (2017) use an empirical approach and US data to show that, when credit access tightens during an economic downturn, employment levels recover relatively quickly, while output and productivity remain low; this is due to low-asset workers exiting unemployment relatively quickly, and taking jobs at low-productivity firms. A separate strand of the literature, based on Baily (1978), shows that the receipt of unemployment insurance extends job search duration: for instance, Chetty (2008) shows that UI disproportionately extends the search of low-asset households, while Crossley and Low (2011) demonstrate the effects of UI on inter-temporal consumption smoothing and job search duration in the presence of credit constraints.

The other two papers in this grouping focus on aggregate-level structural change: Banerjee and Newman (1993) use a dynamic model of occupational choice to explain how the initial distribution of wealth among individuals determines the long-run development path of the economy, which, under certain conditions, leads to prosperity, and, under others, results in stagnation. The mechanism driving these outcomes is the occupational choices made by individuals in the presence of imperfect credit markets. Similarly, Ghatak and Jiang

[4]See Mortensen (1986) for a review of related research, and Bloemen and Stancanelli (2001), Algan et al. (2002) and Rendon (2006) for similar findings. Lentz and Tranaes (2005) show that search effort decreases in wealth for risk-averse workers when utility is additively separable. Wolpin (1987) estimates an asset-dependent model of the transition from high-school to employment for US workers.

(2002) set up an inter-generational model of poverty traps based on Banerjee and Newman (1993), in which credit-constrained workers must choose among three sectors of employment: agricultural subsistence, entrepreneurship and wage work. The entrepreneurial sector utilises a more efficient, but also more costly, technology than the subsistence sector. Consequently, the question of whether or not the economy converges to a prosperous equilibrium depends on the initial wealth distribution: a threshold amount of investment in entrepreneurship is required in order to push the economy onto the high-growth path. Finally, Galor and Zeira (1993) demonstrate the way in which the initial wealth distribution drives long-run aggregate output in the presence of imperfect credit markets and indivisibility in human capital investment.

The second strand of literature I build on deals with endogenous education choice. Keane and Wolpin (1997) set up and estimate a dynamic, discrete-choice model of schooling, employment and occupational choice.[5] They use this to explain observed patterns in the labour-market outcomes of young men in the US, and to forecast patterns of employment and wages. The primary focus of the paper is its augmentation of the standard human capital investment model, however, rather than a consideration of inequality. As such, there is no treatment of wealth. Flinn and Mullins (2015) augment a standard search-and-matching framework with an endogenous education choice that enables access to different labour sub-markets. As in the model of this paper, workers are heterogeneous in ability, and earnings are complementary in ability and match productivity with the firm. Workers sort across sub-markets, taking into account the relative costs, wages, and unemployment rates. The authors then estimate the model using US data to consider the impact of minimum wage policies and education subsidies on welfare. Again, however, differences in wealth are not considered. A closely related paper is Lee (2005), who estimates a dynamic model of schooling, employment and occupational choice for US data, in order to consider the effect of changes in cohort size on these decisions.[6]. Lastly, a number of papers such as Carneiro and Heckman (2002) and Yang (2017) also analyse the effect of parental wealth on post-secondary schooling choice and, consequently, lifetime earnings, but without consideration of unemployment behaviour.

The third, and final, body of research focuses on unemployment in developing countries. Feng et al. (2018) use cross-country household survey data to investigate the relationship between unemployment and GDP per capita. They find a positive relationship, and show further that this is driven primarily by unemployment among low-educated workers. They

[5]Keane and Wolpin (2001), Lee and Wolpin (2006) and Sullivan (2010) represent related models of school, employment and occupational choice.

[6]Other relevant papers include Charlot and Decreuse (2010), who build a model with heterogeneous agents that demonstrates how the presence of search frictions in the labour market leads to over-investment in education by the rich, and under-investment by the poor

build a two-sector model (traditional and modern), in which workers differ in terms of ability (proxying education), while countries vary in terms of their modern-sector productivity. As productivity rises, production shifts to the modern sector and, under the assumption of stronger labour market frictions here than in the traditional sector, unemployment rates rise, both overall, and disproportionately for low-ability workers. Girsberger and Meango (2017) seek to explain the "puzzle of educated unemployment in West Africa".[7] Accordingly, they present a search-and-matching framework with three labour market sectors and heterogeneity in workers' education levels. The model is estimated using data on two West African countries: Senegal and Burkina Faso. They conclude that the relatively high unemployment rates among highly-educated workers are driven by a combination of low job arrival rates, poor search efficiency in self-employment, and differential job destruction. Lastly, Falco and Teal (2012) investigate the nature of youth unemployment in urban Ghana, showing that, while measured unemployment for this group is low, so are earnings and the quality of employment.[8]

The contribution of this paper is, firstly, to provide a framework within which to consider the issue of youth unemployment in sub-Saharan Africa. This is achieved by presenting a set of stylised facts that motivate an understanding of youth unemployment as a luxury that increases in parental wealth, and by formalising these facts into a model that allows for endogenous education, savings and occupational decisions based on parental wealth. This framework draws together the three disparate strands of the labour-market literature I have just described in order to facilitate a richer understanding of the school-to-work transition in a developing-country setting. Secondly, the paper provides estimates of the model for Ghana, which allows quantification of both the "waiting effect" on earnings and the aggregate efficiency and equity consequences of the link between parental wealth and unemployment. Finally, it offers a comparison of two policy interventions, and quantifies their relative effects. The remainder of this paper is organised as follows: Section 2.2 presents a trio of stylised facts about labour markets in Ghana (and sub-Saharan Africa more generally) that motivates this research, Section 2.3 outlines the theoretical framework, Section 2.4 introduces the data and presents the results from a structural estimation of the model, Section 2.5 evaluates a number of labour market policy alternatives, and Section 2.6 concludes.

[7]Fan and Stark (2012) offer a discussion of this phenomenon for developing countries more generally.

[8]Falco and Teal (2012) also argue that waiting in unemployment increases neither one's chance of getting a job, nor one's earnings later in life, which appears to contradict the findings of this paper. Yet Falco and Teal (2012) include workers out of the labour force in their definition of unemployment, while this paper is concerned solely with active job-seekers.

2.2 Data and Facts

The paper is motivated by a trio of stylised facts concerning young labour-force participants in Ghana.[9] The data I use comes from the 2012 round of the Ghana Living Standards Survey (hereafter, GLSS 2012). This nationally representative, cross-sectional dataset covers approximately 17,000 Ghanaian households, and includes a rich labour market module. The selected sample is working-age males (aged 15-60). I drop those with missing information on education, current employment and the variables used to compute parental wealth. I also drop those currently in education, as well as casual workers, apprentices and individuals not in the labour force. This yields a sample size of approximately 10,500 individuals.

Workers' education is measured as the last completed level of schooling. For the structural model, this is converted into a binary variable that defines individuals as "highly-educated" if they have completed (as a minimum) senior high school, and as "low-educated" otherwise. Employment status is defined as follows: an individual is classified as "unemployed" if he has not worked during the last week, does not have a job on hold, and is available to work.[10] Parental wealth for all individuals is obtained as follows: first, I regress a measure of household assets on education, occupation and geographic region for males aged 35 to 60. I then predict parental wealth using the same explanatory variables for sampled individuals' fathers, collected in the household module of the GLSS 2012. Details are given in Appendix B.1.

2.2.1 Fact 1: Low Youth Unemployment

The first fact of interest is that youth unemployment, in Sub-Saharan Africa on average, and in Ghana in particular, is moderately low. In 2012, unemployment among men aged 15-24 years was 12.1% in the SSA region and only 9% in Ghana (Kühn et al. (2016)). By comparison, the relevant figures for the European Union and the United States were 25.2% and 17.5%, respectively. Nevertheless, relatively low levels of unemployment in the context of low-income, developing countries are not surprising but, rather, to be expected: in the absence of unemployment insurance, remaining unemployed is simply an unaffordable luxury for the majority of the labour-force. Instead, the more important and relevant statistic is the productivity of employment, and it is here that the region truly lags behind others: it has the

[9] While Ghana constitutes the principal focus of this research, however, these findings are not unique to it but rather, are also applicable more broadly to other countries in the Sub-Saharan African region. To illustrate this, I replicate the stylised facts detailed in this section for Uganda in Appendix B.2.

[10] This is a slightly more relaxed version of the standard ILO definition of unemployment, which additionally requires individuals to have actively searched for work in the last 7 days. I adopt this version because it includes discouraged workers and those who may not have searched in the last 7 days due to slow job arrival rates, but who are still "unemployed" in the sense that they want a job but do not have one. This is, further, the definition used by the Ghana Statistical Service in their 2014 Labour Force Report based on this data.

highest working poverty rate among youth of any region in the world, with approximately 70 per cent of youth living on a daily income of less than US$ 3.10 (Kühn et al. (2016)).

2.2.2 Fact 2: Low Educational Attainment

The second fact is that educational attainment in Ghana is very low, despite numerous policy attempts to raise it, including making education up to the age of 15 compulsory, and providing education up to the end of senior high school available free of charge. As Figure 2-1 shows, approximately 76% of Ghanaian men between the ages of 20 and 60 have not completed senior high school, and 36% have no formal education at all. Regionally, it is estimated that, in 2014, 58% of secondary school-aged children were not in school; this is the highest proportion of any region in the world (UNESCO Institute for Statistics (2016)).

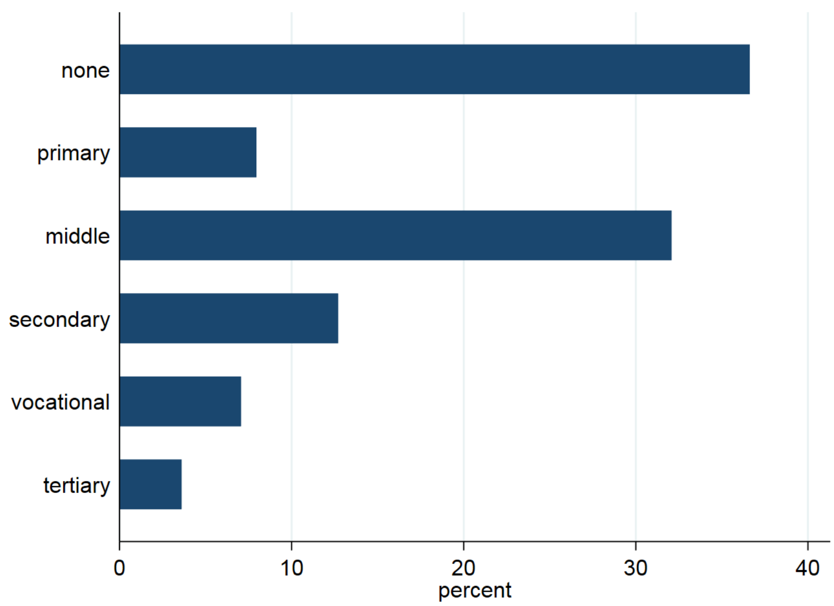


Figure 2-1: Male Educational Attainment in Ghana: Age 20-60
Data Source: GLSS 2012

2.2.3 Fact 3: Youth Unemployment Increases with Wealth

The third stylised fact is that Ghanaian youth from wealthy families are more likely to be unemployed early in life than are their counterparts from poorer backgrounds. Table 2.1 shows the results from a linear probability regression of parental wealth on the probability of men across different age groups being unemployed.[11] In both cases, the dependant variable is equal to 1 if the individual is currently unemployed and reports being available to work,

[11] The construction of the parental wealth measure is described in detail in Appendix B.1

and zero if currently employed. I control for potentially confounding variables, including age, region, and whether the individual lives in an urban or rural environment. Columns 3 and 4 add controls for education. While education decisions are endogenous, adding these controls is still useful in terms of showing that differences in unemployment are not accounted for by education alone; rather, as the model presented in the next section will show, it is one of the channels through which parental wealth affects unemployment.

In Column 1, there is a positive and highly significant correlation between parental wealth and unemployment among men aged 15-29: a 1% increase in parental wealth raises the probability of unemployment by 0.08 percentage points. Comparing this to the results in Column 2, however, reveals that this relationship completely disappears after the age of 30. Columns (3) and (4) show that the same pattern persists even after controlling for education. The results are also robust to using a probit specification, as shown in Appendix B.2, Table B.3.

Table 2.1: Parental Wealth and Male Unemployment

	(1) Age 15-29 Unemployment = 1	(2) Age 30-60 Unemployment = 1	(3) Age 15-29 Unemployment = 1	(4) Age 30-60 Unemployment = 1
log parental wealth	0.077***	0.009	0.059**	0.009
	(0.019)	(0.006)	(0.020)	(0.006)
age	-0.007***	0.000	-0.008***	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
urban	0.070***	0.004	0.062***	0.003
	(0.012)	(0.003)	(0.012)	(0.003)
region dummies	Yes	Yes	Yes	Yes
education dummies	No	No	Yes	Yes
mean unemployment	0.072	0.020	0.072	0.020
N	3328	7172	3328	7172

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Data Source: GLSS 2012

In the next sections, therefore, I build and estimate a model that formalises the mechanisms underlying this trio of stylised facts. As previously stated, the primary focus of this model is the way in which parental wealth affects youth unemployment in a developing-country setting, and the consequences of such a relationship for the equity and efficiency of labour market outcomes.

2.3 Model

In this section, I present a discrete-choice, life-cycle model that considers the effect of parental wealth on youth unemployment. In this model, workers rely on family wealth in order to fund consumption while transitioning from education into employment. The primary mechanism underlying the unemployment motive is a desire to trade-off higher present consumption (by accepting a job) against a better wage offer (by waiting in unemployment). Consequently, wealthy workers remain unemployed longer, on average, in order to access high-paying jobs, while workers from poor families are willing to accept less desirable jobs in order to transition out of unemployment, which has consequences for education choices. As previously discussed, the combination of credit market failure and an absence of unemployment benefits amplifies the role of family wealth in this developing-country setting.

2.3.1 Model Set-Up

The model economy consists of a finitely-lived, working-age population. There are no credit markets.

Endowments:

Workers are heterogeneous in their initial endowments of family wealth and ability, each of which is an independent, random draw. Ability, denoted h_i, may be low or high, such that $h_i \in \{H^L, H^H\}$, and is observable. The proportion of low-ability types in the population is q_L. The initial family wealth of worker i, denoted $a_{i,0}$, is drawn from a log-normal distribution with cumulative distribution function F_{a0}, mean μ_{a0} and variance σ^2_{a0}.

Preferences:

Workers are risk-averse, and have logarithmic preferences over consumption in every period. The period discount factor is denoted β.

$$\sum_{t=1}^{T} \beta^{t-1} \ln c_t \tag{2.1}$$

Education:

Workers make a binary education decision at time $t = 0$. Let $k = \{0, 1\}$ indicate a choice of low and high education, respectively. For the purposes of this paper, I define a worker of "high education" to be one who invests in schooling beyond the compulsory level in Ghana of junior high school. A worker of "low education" is equivalently defined as one who has the compulsory level of schooling or lower. Workers incur a fixed cost of education, denoted E^E_k, where $E^E_0 = 0$ and $E^E_1 > 0$. Those who choose to invest in high education spend T_E periods at the start of life in school, with one period in this model set equal to a single year of working life; simultaneous employment is not permitted. A worker's choice of education determines the set of occupations in which she later searches for employment (this is described in more detail in the following subsection on jobs). Finally, I assume no depreciation of skills during the transition from school into employment.

Jobs:

The labour market in this model consists of three sectors: wage employment, agricultural production and entrepreneurship, where the latter is defined as self-employment in a non-agricultural business.[12] Entrepreneurship incurs a fixed start-up cost, denoted E_S. Let $m = \{u, w, a, e\}$ denote the set of possible employment statuses, corresponding to unemployment, wage employment, agriculture and entrepreneurship, respectively. Note that, while workers may remain unemployed indefinitely, employment is considered an absorbing state, such that there is no mobility across occupational alternatives once the worker has moved from unemployment to some form of employment. This is a reasonable assumption because job-to-job mobility in this context is low: for instance, among individuals sampled in the 2004 round of the Ghana Urban Household Panel Survey, which includes a detailed job history module, 30-40 year-olds had had an average of 1.9 jobs in their lifetime (see Table B.7 in Appendix B.3). This figure would be even lower in rural regions (which is where 66% of the estimation sample lives), given the relative scarcity of job opportunities there.

Workers in this model economy may be employed in one of six distinct occupations. An occupation is defined as a sector-education level pairing, each with its own distribution of firm productivities. The productivity distribution for each alternative – from which each unemployed worker receives a random draw each period – is stationary, and is assumed to follow a log-normal distribution with mean $\mu_{m,k}$, variance $\sigma^2_{m,k}$ and cdf $F_{m,k}$ where, as before, m denotes an employment sector and k denotes workers' education level. Thus, as previously described, while workers may choose to be employed in any sector, their education choice restricts access to one or other subset of three (of the possible six) occupations.

The return to workers in each of the wage and entrepreneurial occupations is complementary in firm productivity and worker ability, while the equivalent return for the agricultural occupations depends solely on the firm productivity draw.[13] Furthermore, I allow for growth

[12]Gindling and Newhouse (2014) and Nguimkeu (2014) highlight the important role played by the heterogeneous informal sector in a broad range of developing countries, and in West Africa, respectively. For an analysis of the role of self-employment in the Ghanaian labour market, with particular treatment of the question of whether it represents a sector of "opportunity" or of "last resort", see Falco and Haywood (2016). Finally, Kerr (2012) argues the importance of distinguishing between self-employment and wage-employment in small firms, which are traditionally treated collectively as the "informal sector".

[13]Such an assumption is common in the literature for developing countries. For instance, Feng et al. (2018)

in earnings over time (and within a given job) by including an occupation-specific growth term, $G_{m,k}$. Thus, an employed worker i's total period earnings in a given job, $W_{i,m,k,t}$, may be written as:

$$W_{i,m,k,t} = \begin{cases} G_{m,k}^{t-d} w_{i,m,k,d} h_i & \text{if } m \in \{w, e\} \\ G_{m,k}^{t-d} w_{i,m,k,d} & \text{if } m \in \{a\} \end{cases} \tag{2.2}$$

where t indicates the current period, and d the period in which worker i transitioned into employment.

Choices and Timing:

During their lifetime, workers face three decisions. The first is a choice of education level. The second is a two-part decision while in unemployment: whether or not to accept a given job offer, and which sector to accept an offer from. Search is random, and every unemployed worker receives an offer simultaneously from each of the three sectors in each period prior to accepting one. They receive no unemployment benefits, and so, must rely solely on their initial wealth draw to finance consumption until entering employment. The third, and final, decision is a trade-off between consumption and savings in each period. When unemployed, consumption will be an increasing function of $a_{i,0}$, h_i, and expected future earnings in employment. Once employed, the optimal consumption path is deterministic, and is increasing in $a_{i,0}$, h_i and accepted earnings.

The timing of the model is as follows: at $t = 0$, workers receive their endowments of wealth and ability ($a_{i,0}$ and h_i). They must choose whether or not to invest in high education at a cost of E_1^E. At $t = 1$, workers enter the labour market. They receive an independent, random firm productivity draw from each sector: wage employment, agriculture and entrepreneurship. Workers then decide whether to accept one of these offers, or to remain unemployed for the period. In all subsequent periods up to and including $t = T$, employed workers remain in their chosen job. Unemployed workers continue to receive employment offers each period until they accept one. At the end of $t = T$, all workers die.

---

equally assume that agricultural returns do not depend on ability while returns to other sectors do, and Banerjee and Newman (1993) make an equivalent assumption, albeit using effort rather than ability.

2.3.2 Solving the Model: Optimal Education Choice

Workers seek the optimal strategy k to maximise the discounted sum of future pay-offs. I define the value of strategy k at time $t = 0$ as:

$$VS_{i,k} = E\left[\sum_{t=1}^{T} \beta^{t-1} \ln(c_{i,k,t})\right] - E_k^E \tag{2.3}$$

The optimal strategy is then:

$$VS_i^* = max\{VS_{i,0}, VS_{i,1}\} \tag{2.4}$$

2.3.3 Solving the Model: Optimal Savings Choice

Unemployed Workers

Let $c_{i,m,k,t}$ denote consumption at time t of an individual i with education k and employment status m. While unemployed, workers then solve the following maximisation problem, based on their expectations of future earnings:

$$\begin{aligned} \max_{c_{i,0,k,t}} \quad & E\left[\sum_{t=1}^{T} \beta^{t-1} \ln(c_{i,0,k,t})\right] \\ \text{subject to} \quad & a_{i,t} = (1+r)a_{i,t-1} + W_{i,m,k,t} - c_{i,0,k,t} \\ & a_T = 0 \quad \forall \quad i = 1, ..., N \\ & a_t \geq 0 \quad \forall \ t = 1, ..., T \ and \ \forall i = 1, ..., N \end{aligned} \tag{2.5}$$

This gives rise to an Euler equation:

$$E_t[c_{t+1}] = \beta R c_t \tag{2.6}$$

This problem has no explicit solution for $c^*_{i,0,k,t}$, and is solved computationally, by backward induction.

Employed Workers

Once employed, the savings choice is deterministic, and may be solved explicitly. Workers solve the following maximisation problem, in which z denotes the current period of employed life and $Z \leq T$ denotes the total number of periods spent in employment:

$$
\begin{aligned}
\max_{c_{i,m,k,z}} \quad & \sum_{z=1}^{Z} \beta^{z-1} \ln(c_{i,m,k,z}) \\
\text{subject to} \quad & a_{i,z} = (1+r)a_{i,z-1} + W_{i,m,k,z} - c_{i,m,k,z} \\
& a_{i,Z} = 0 \quad \forall \quad i = 1, ..., N \\
& a_{i,z} \geq 0 \quad \forall \ z = 1, ..., Z \ \ and \ \forall i = 1, ..., N
\end{aligned} \tag{2.7}
$$

This yields the following interior and corner solutions for the optimal consumption path, where $R = (1+r)$ and r is the real interest rate on savings:

$$
c^{I}_{i,m,k,z} = \frac{\beta^{z-1} R^{z-1}}{R^{Z-1} \sum_{z=0}^{Z-1} \beta} [R^{Z} a_{i,0} + \sum_{z=1}^{Z} R^{Z-z} W_{i,m,k,z}] \tag{2.8}
$$

$$
c^{C}_{i,m,k,z} = \begin{cases} Ra_{i,0} + W_{i,m,k,z} & \text{if } z = 1 \\ W_{i,m,k,z} & \text{if } z = 2, ..., Z \end{cases} \tag{2.9}
$$

Thus, the employed worker's optimal consumption in any period z is given by:

$$
c^{*}_{i,m,k,z} = min\{c^{I}_{i,m,k,z}, c^{C}_{i,m,k,z}\} \tag{2.10}
$$

2.3.4 Solving the Model: Optimal Employment and Sectoral Choice

Finally, I consider jointly the decision by workers between employment and unemployment, as well as between sectors. Workers seek the optimal strategy m to maximise the discounted sum of future pay-offs. I define the value of strategy m at time z as:

$$
VM_{i,m,k,z} = E\left[\sum_{t=z}^{T} \beta^{t-1} \ln(c_{i,m,k,z})\right] \tag{2.11}
$$

The optimal strategy is then:

$$
VM^{*}_{i,k,t} = max\{VM_{i,u,k,t}, VM_{i,w,k,t}, VM_{i,a,k,t}, VM_{i,3,e,t}\} \tag{2.12}
$$

The model is solved computationally, by backward induction, and the results are discussed in the next section.

2.3.5 Model Predictions

In this section, I consider the ways in which the two dimensions of worker heterogeneity (wealth and ability) affect choices over education and unemployment duration. This is important for understanding the mechanisms that underlie observed individual choices, for determining the sign and magnitude of the effects on equity and efficiency, and, later, for making sense of the differential effects of alternative policy interventions.

2.3.5.1 Wealth and Education Choice

The sign of the relationship between wealth and education choice is ambiguous; it depends on the cost of education and the relative parameters of the offered earnings distributions in high- and low-education occupations. There are thus two channels of effect; I term these the "fixed cost" channel and the "cost of waiting" channel, respectively. The first of these depends on E_1^E and T_E, the monetary and time costs of education. The existence of these fixed costs implies a threshold value of wealth above which workers can afford to enter high education, and below which they cannot. The "cost of waiting" channel is, however, governed by the relative means and variances of offered earnings in high- and low-education occupations. In particular, when returns to education are highly uncertain (i.e. they have a low mean and a high variance), less-wealthy workers will prefer not to obtain high education, due to the high waiting cost of a good draw.

To show the effect on education choice of each of these channels in turn, I begin by considering a case in which the fixed cost of education is zero and the distribution of offered earnings in high education has the same mean but a high variance compared to the equivalent distribution in low education. This shuts down the "fixed cost" channel, isolating the "cost of waiting" effect. As Figure 2-2 illustrates, this generates the standard result of a threshold value of wealth, a_0^*, above which individuals always choose high education, and below which, equivalently, they always choose low education. Increasing the fixed cost of education then shifts the education-wealth threshold to the right: the minimum level of wealth which induces workers to invest in high education is now higher.

It should be noted, however, that, while Figure 2-2 shows a case in which the "cost of waiting" channel induces a positive relationship between wealth and education choice, it is also possible for this effect to be negative. In particular, if low-education occupations have a sufficiently higher "cost of waiting" than high-education occupations (that is, the low-education earnings distribution is relatively more dispersed than the equivalent high-education one) and the "fixed cost" effect is small or zero, there is a negative relationship between education and wealth, in which poorer workers choose high education, while the

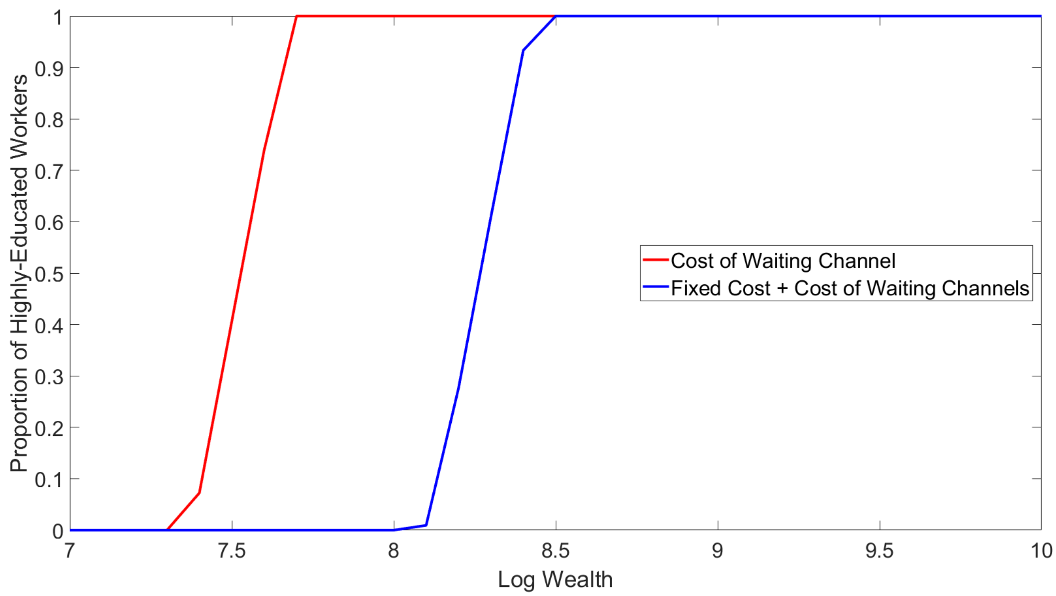


Figure 2-2: Wealth and Education Choice

wealthiest workers, who can afford to wait a longer time in unemployment for a high-paying job, find it optimal to remain at a low level of education, where there is now a chance of getting a very high draw.

2.3.5.2 Wealth and Unemployment Choice

In considering the effect of wealth on workers' choice of employment status, it is important to distinguish between the effects *within* and *across* education categories. Conditional on education status, average unemployment duration is always increasing in initial wealth. Across education groups, however, the sign of the relationship depends on relative parameter values of the earnings distributions, as these determine how long an individual must wait in unemployment at each education level for a job he is willing to accept.

In particular, when offered earnings in high- and low-education have the same mean, but the former has a higher variance, the cross-education wealth effect is positive: wealthy workers are now more likely to get educated, given the fixed costs of high-education, and also remain unemployed longer than their poorer counterparts, incentivised by the relatively high earnings variance in their sub-market. This case is illustrated in Panel (i) of Figure 2-3.

Conversely, consider a case in which high-education offered earnings have a high mean and a low variance relative to low-education earnings.[14] Wealthy workers remain more likely

[14] A more natural comparison to the former case would be to assume constant means across the two education levels, but a higher variance for low-education earnings. In such a case, however, the fraction of educated workers is zero, so it is not possible to consider cross-group variation.

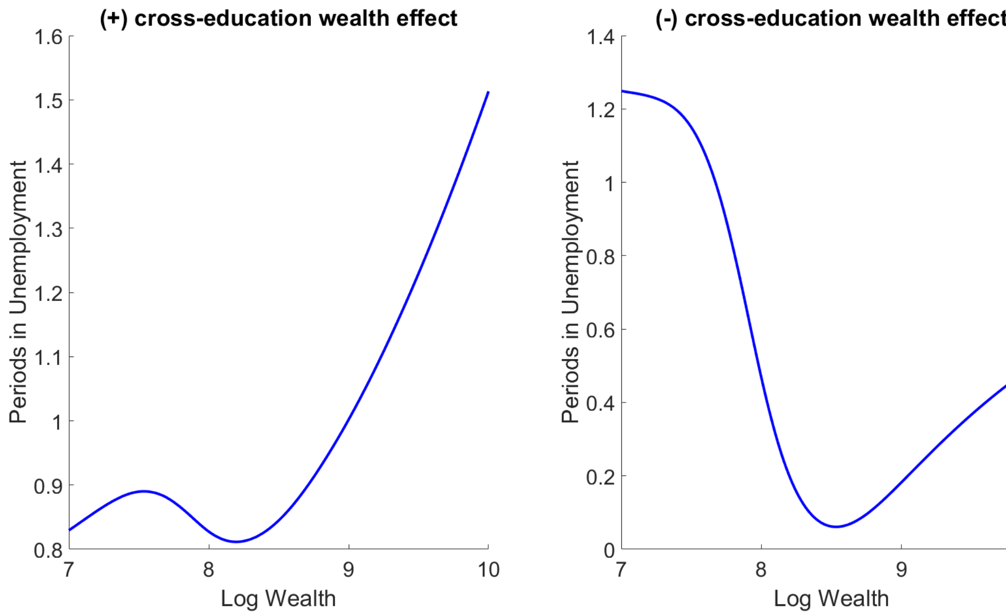


Figure 2-3: Wealth and Unemployment Duration

to get educated, but now receive good job offers relatively quickly, so do not remain long in unemployment. Poorer workers, by contrast, do not invest in high education, and then find it optimal to wait relatively longer in order to secure a job. Here, the cross-education wealth effect is negative. Panel (ii) of Figure 2-3 illustrates this case; with the negative cross-education wealth effect dominating the positive within-education wealth effect, except at the top of the wealth distribution (where everyone is highly-educated, such that the cross-group effect is zero and the positive within-group effect dominates).

2.3.5.3 Ability and Education Choice

In terms of ability, the model predicts two opposing effects on education choice, rendering the net effect ambiguous. I term these the "fixed cost" channel and the "asset de-accumulation" channel, respectively. The "fixed cost" channel summarises the way in which, conditional on wealth, the existence of an education fixed cost, E_1^E makes high-ability workers more likely than their low-ability counterparts to undertake high education. This is shown in Figure 2-4: increasing the monetary or time cost of high education reduces the proportion of highly-educated workers, but high-ability workers are less sensitive to these higher costs than are their low-ability counterparts.

The "asset de-accumulation" channel, on the other hand, captures the fact that, all else equal, a high-ability worker draws down his assets faster in unemployment than does a low-ability worker, in anticipation of relatively higher future wages. This means that a high-ability

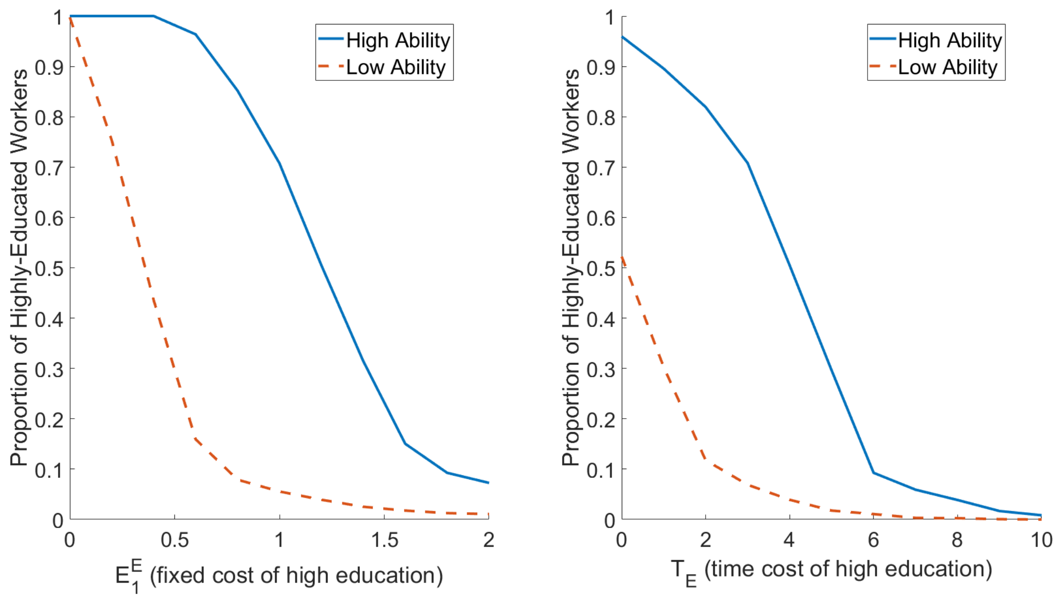


Figure 2-4: Cost of Education and Education Choice by Ability

worker behaves, for the purposes of education choice, as though he were effectively poorer than a low-ability worker with the same level of assets. Therefore, as in the relationship between wealth and education choice described above, when offered high-education earnings are sufficiently uncertain (low $\mu_{m,1}$ and high $\sigma^2_{m,1}$) relative to low-education earnings, a high-ability worker prefers to remain uneducated, as it allows him to transition more quickly into employment, while a low-ability worker with the same wealth chooses high education. Figure 2-5 presents this channel; beginning with a case in which high-education returns have a lower mean but a higher variance than low-education returns, an increase in the variance of high-education returns induces workers of both high- and low-ability to get educated, but the latter group is more sensitive to this change than is the former.[15]

Thus, the question of which effect dominates depends on the relative parameters of education costs and the earnings distributions; consequently, the probability of obtaining high education may be increasing or decreasing in workers' ability.

2.3.5.4 Ability and Unemployment Choice

As with wealth, it is important to distinguish between the effects of ability on unemployment *within* and *across* education categories. First, the relationship with unemployment conditional on education is ambiguous. There are two channels that drive this relationship, which I term the "asset de-accumulation" channel and the "sectoral shift" channel, respectively.

[15]For simplicity, I set offered returns equal to zero in all sectors except wage-employment.

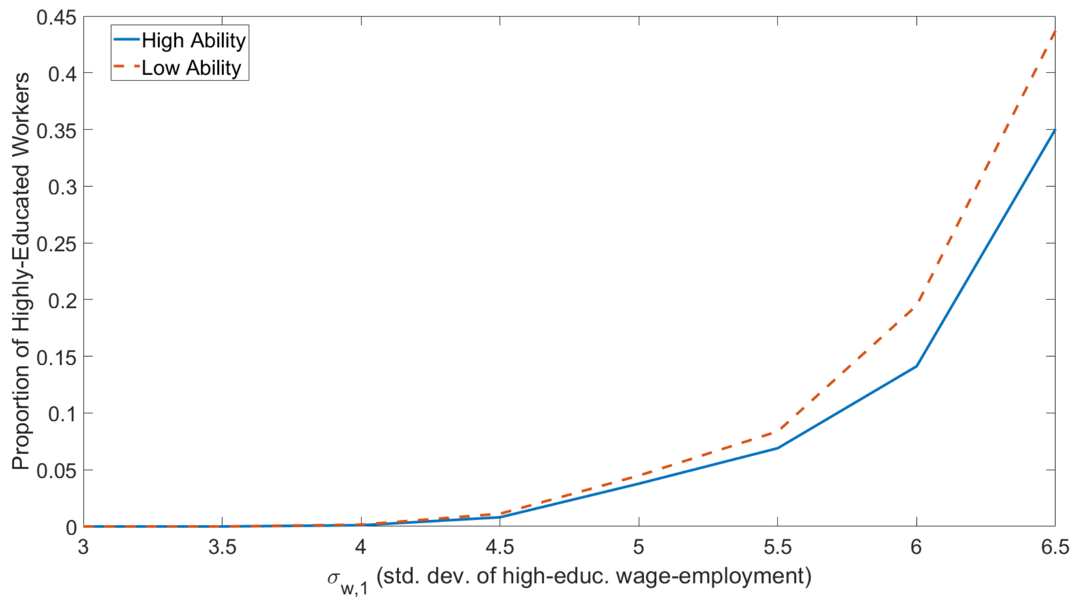


Figure 2-5: Uncertainty in Returns to Education and Education Choice by Ability

The first of these has already been explained: a high-ability worker consumes more in unemployment than does a low-ability worker with the same wealth, due to having a higher expected future income stream. Thus, he draws down his assets more quickly, which lowers his reservation wage and pushes him into employment. This means that a high-ability worker behaves, for the purposes of employment choice, as though he were effectively poorer than a low-ability worker with the same level of assets.

The "sectoral shift" channel arises due to the assumption that ability does not affect earnings in the agricultural production sector. Consequently, a low-ability worker facing an employment offer from this sector is now *more* likely to accept it than an equivalent high-ability worker, because he receives exactly the same value of employment, but the former has a lower outside value of unemployment, due to lower expected earnings in the other two sectors. This shifts low-ability workers towards employment in the agricultural sector, particularly when offered returns in this sector have a relatively high mean.

Figure 2-6 shows the relationship between ability and unemployment duration, conditional on workers' education choice. In the first panel, I retain the assumption that ability does not affect earnings in agricultural production, such that the results in this panel derive from a combination of the "asset de-accumulation" and "sectoral shift" channels. In the second panel, however, I relax this assumption (thereby removing the "sectoral shift" effect). The removal of this channel leads to higher unemployment durations among low-ability workers in the second panel compared to the first, but there is no change in the outcomes of high-ability workers.

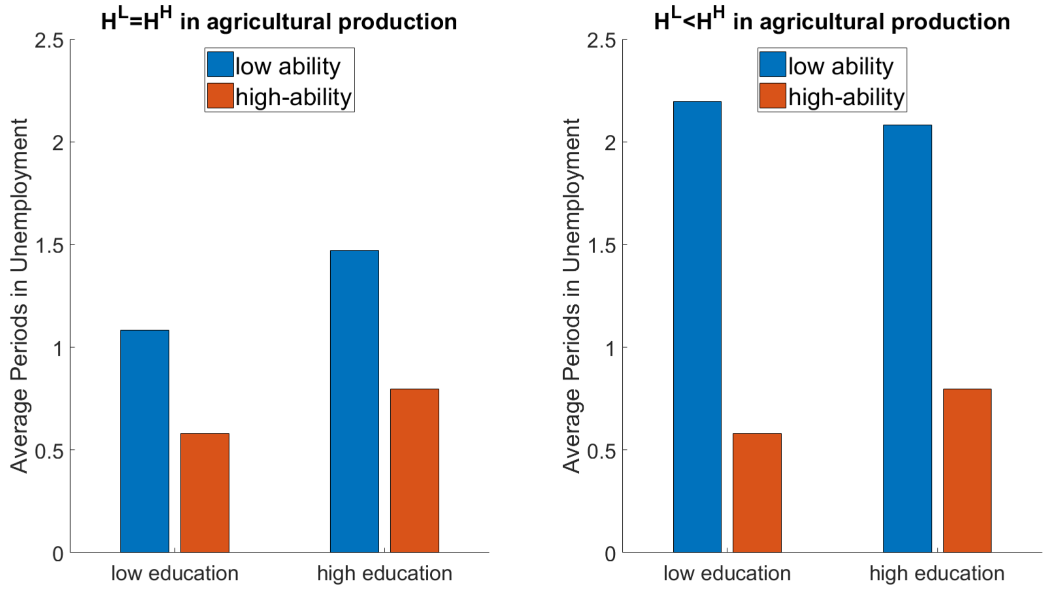


Figure 2-6: Ability and Unemployment Duration

Second, if we now consider the relationship *across* education groups, an additional, compositional effect arises. The average unemployment duration across education sub-markets is driven by the average asset-holdings of individuals in each sub-market, as well as by the offered earnings parameters of its occupations. Thus, if high-ability workers are disproportionately likely to invest in high education (as detailed in the previous section), and offered earnings in this sub-market have relatively low means and high variances compared to offered earnings in the low-education sub-market, the cross-education effect of ability on unemployment duration is positive.

2.4 Data Description and Estimation

As previously described, the data comes from the sub-sample of working-age males (aged 15-60) in the GLSS 2012. For employed workers, information about annual earnings and sector of employment comes from survey questions relating to work during the last week and the total number of hours worked in the last year. Workers who report having multiple occupations (representing about 10% of the sample) are classified as belonging to their main sector of employment, but total earnings across both sectors are used. For workers in agriculture, earnings information is supplemented by data from a survey module on household agricultural income. Lastly, for individuals recorded as “unpaid workers” in household agriculture or non-agricultural businesses, earnings are computed as the total household earnings for that activity, divided by the number of household members engaged in it.

2.4.1 Summary Statistics

In this section, I present some summary statistics for the GLSS 2012 data. First, Ghana has - as does the Sub-Saharan African region more broadly - a particularly youthful population. As Figure 2-7 shows, 68% of individuals in the full sample are 30 years old or younger, a feature that lends additional significance to the consideration of youth labour-market outcomes.

The remainder of these summary statistics focuses on the sub-sample of male labour-force participants aged 16-30, which is the estimation sub-sample (equivalent statistics for the full sample are presented in Appendix B.4). As previously discussed, average education is low, with 70.73% of sampled workers having attained a maximum of middle-school education (see Table 2.2). 34.33% are urban-dwellers, while the rest live in rural areas.

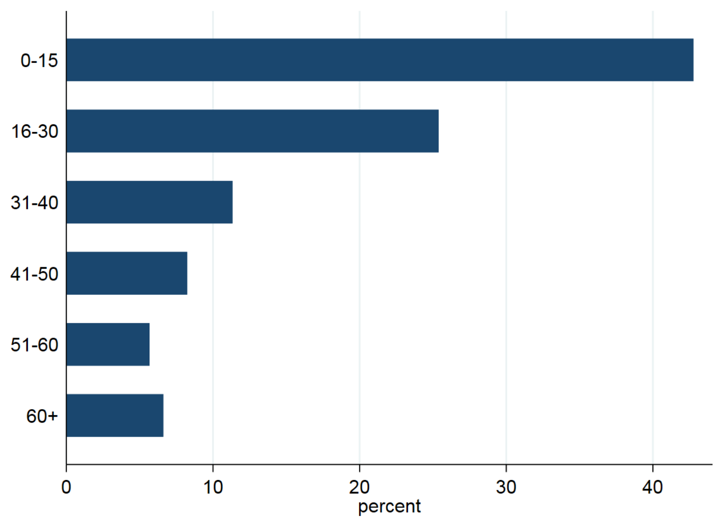


Figure 2-7: GLSS 2012 by Age Group

Table 2.2: Summary Statistics: Male Labour-Force Participants Aged 16-30

observations	3728	
urban (%)	34.33	
highly-educated (%)	29.27	
	mean	s.d.
age (years)	24.26	4.07
education (years)	8.47	5.90
median parental assets (US$)	620.59	553.76

Figure 2-8 presents the sectoral composition of this labour market; the most striking feature is that 45% of all individuals work in agricultural production (which accords with the high proportion of rural-dwellers). The proportion of agricultural workers falls significantly with education, however: 64% of low-educated workers are in this sector compared to 28% of their highly-educated counterparts. Education shifts the sectoral composition away from agricultural production and towards wage-employment, with 47% of highly-educated workers employed in the wage sector, compared to only 18% of the low educated. The share of entrepreneurs, by contrast, is approximately constant across education groups.[16]

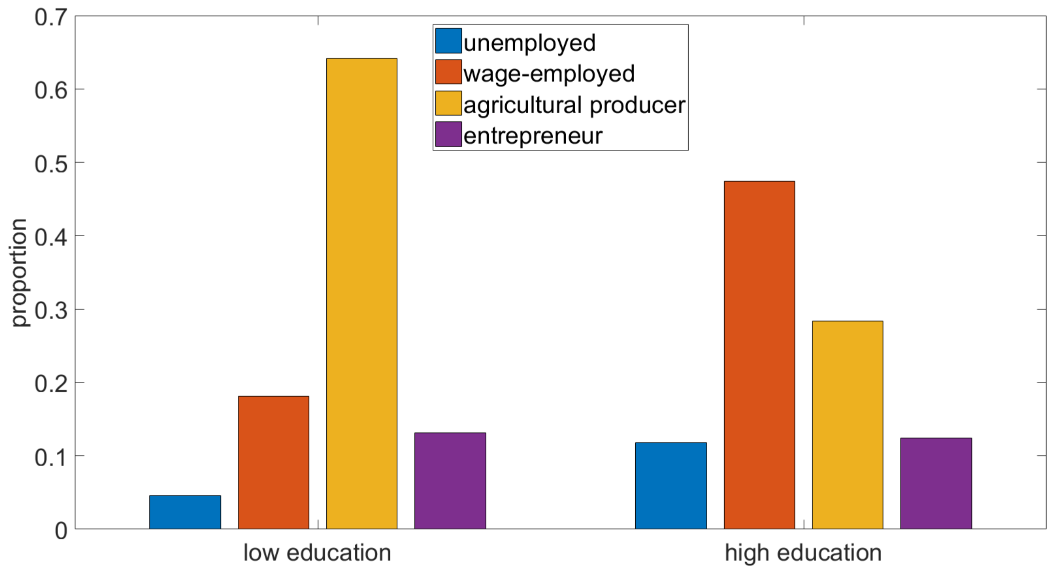


Figure 2-8: Employment Status by Education

2.4.2 Model Predictions and Empirical Patterns of Parental Wealth

The model predicts that parental wealth drives workers' choices over unemployment, education and sector of employment and, consequently, leads to higher lifetime earnings for the wealthy. In this section, I show that the empirical patterns of workers' outcomes are consistent with these predictions.

First, parental wealth drives higher unemployment rates through two channels: first, by making wealthy workers more likely to invest in education, when earnings in high-education occupations have relatively high variances, and second, conditional on educational status, by making wealthy workers wait longer in unemployment for a better job. I have already demonstrated the positive correlation between parental wealth and unemployment in Section 2.2. In fact, unemployment rates in the data also differ significantly across education groups,

[16]Poschke (2013), for instance, argues, using US data, that entrepreneurship rates are U-shaped with respect to education, due to the co-existence of necessity-based and talent-based entrepreneurs.

with 12% of highly-educated workers unemployed compared to only 5% of low-educated workers, as Figure 2-8 has shown.

Second, Table 2.3 presents a linear probability regression of education choice on the log of parental wealth and geographic controls. There is, as expected, a positive and significant correlation between parental wealth and the probability of choosing to be highly-educated.

Third, the model predicts that differences in wealth drive sorting behaviour across sectors of employment, with wealthier workers more likely to enter sectors in which returns are relatively dispersed and, therefore, require a longer wait in unemployment in order to secure a high-productivity draw. Table 2.4 presents the results from a multinomial logit regression of employment sector on the log of parental wealth and geographic controls. Among low-educated workers (the first column), higher parental wealth is associated with a shift out of agriculture and into wage employment and (to a lesser extent) entrepreneurship. Among highly-educated workers, higher parental wealth remains correlated with a shift out of agriculture and into wage employment, but there is no significant change in the probability of being an entrepreneur. This is consistent with the model prediction, as the estimation results in the next section show that offered earnings are, in fact, more dispersed in the wage and entrepreneurial sectors than in agriculture.

Finally, I consider the relationship between parental wealth and lifetime earnings. Table 2.5 displays the results from a regression of the log of annual earnings on the log of parental wealth and geographic controls. As parental wealth has a positive relationship with both education choice and unemployment, the model predicts that workers from wealthier families earn more than their counterparts from poorer backgrounds on average, and this wage gap is increasing over the life-cycle. The latter result is due primarily to higher unemployment among the wealthy (such that they are more likely to record zero earnings early in life), but may also be affected by wealthy workers being relatively more willing to accept employment in sectors with low initial earnings but high earnings growth. Table 2.5 presents just such a pattern: in the first column (workers aged 16-30), parental wealth is negatively correlated with earnings; in the second column (workers aged 31-50), however, wealthy workers not only catch up with, but significantly outstrip, their poorer counterparts.

2.4.3 Identification and Structural Estimation Procedure

The model is estimated by the Method of Simulated Moments (MSM). It has a total of 28 parameters; 14 of these are identified structurally, by targeting a set of 25 data moments, while the remaining 14 are external parameters fixed ex-ante in accordance with the data or

Table 2.3: Education Choice and Parental Wealth

	(1) Age 16-30 Educated=1
log parental wealth	0.331***
	(0.019)
urban	0.179***
	(0.018)
region dummies	Yes
mean(educated)	0.293
N	3728

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

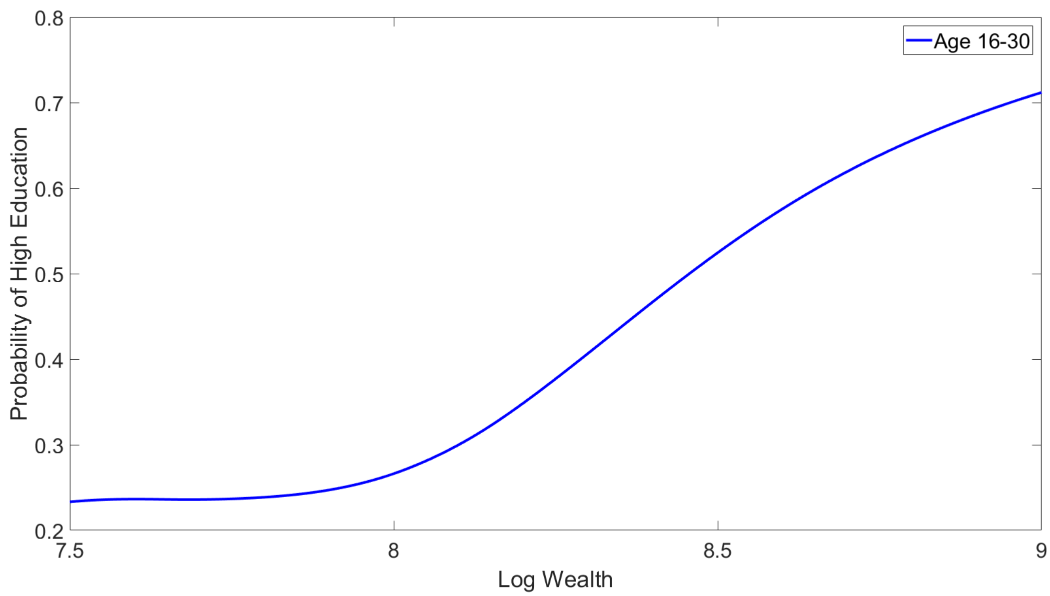


Figure 2-9: Education Choice and Parental Wealth

Table 2.4: Employment Sector Choice and Parental Wealth

	(1) education=0 Employment Sector	(2) education=1 Employment Sector
log parental wealth		
1.wage employment	0.093***	0.170***
	(0.023)	(0.036)
2. agriculture	-0.136***	-0.140***
	(0.028)	(0.038)
3. entrepreneurship	0.044*	-0.030
	(0.022)	(0.021)
urban		
1.wage employment	0.130***	0.169***
	(0.013)	(0.030)
2. agriculture	-0.299***	-0.269***
	(0.014)	(0.024)
3. entrepreneurship	0.169***	0.101***
	(0.013)	(0.024)
region dummies	Yes	Yes
N	2516	962

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
data: employed males aged 16-30 (GLSS 2012)

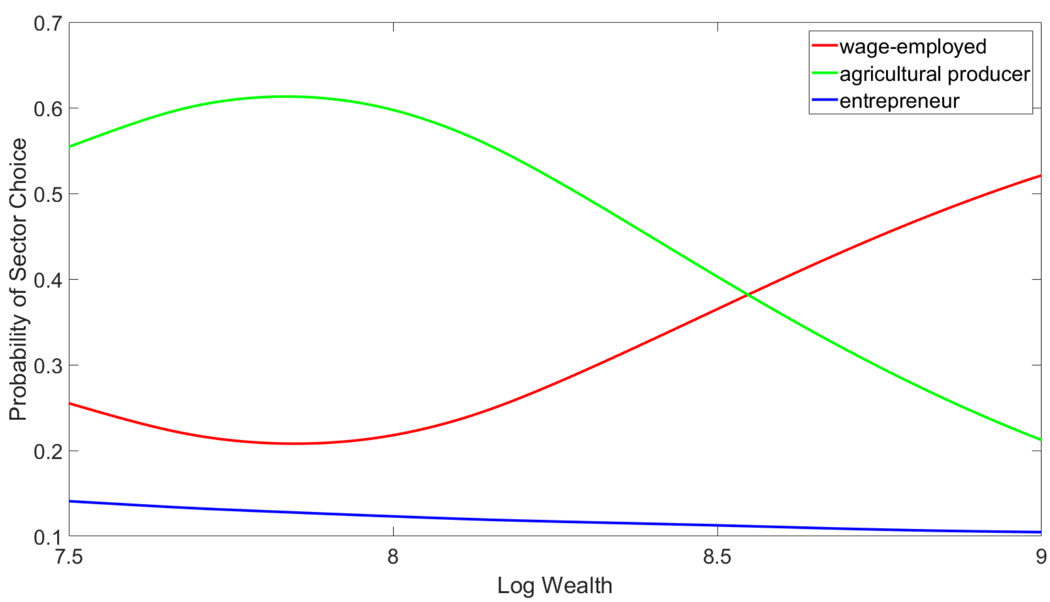


Figure 2-10: Sectoral Choice and Parental Wealth

Table 2.5: Log Earnings and Parental Wealth

	(1) Age 16-30 Log Annual Earnings	(2) Age 31-50 Log Annual Earnings
log parental wealth	-0.479*	0.260**
	(0.196)	(0.088)
age	0.324***	-0.002
	(0.022)	(0.006)
urban	-0.194	0.671***
	(0.187)	(0.076)
region dummies	Yes	Yes
mean(log earnings)	6.050	7.388
N	2237	4329

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

data: all males aged 16-50 (GLSS2012)

Figure 2-11: Log Annual Earnings and Parental Wealth

the literature.

Table 2.6 summarises the external parameters of the model. The real interest rate is obtained from national statistics. The wealth distribution is described by my measure of parental wealth. The cost of education is the median amount spent on tuition, textbooks, transport, boarding and other schooling costs for male family members currently in public schooling at the level of senior high school or above (the variable captures expenditure within the last 12 months, and this figure is multiplied by the time cost of education to obtain the parameter estimate for E_1^E). The GLSS 2012 does not collect information about the fixed cost of entrepreneurship. Thus, the estimate of E_S is the average start-up cost reported by entrepreneurs in the Ghana Urban Household Panel Survey (2004-2006). The model is simulated over 35 periods, covering the age range 16-50. The time cost of education is the average number of additional years spent in education by high-educated individuals (those with a minimum of a senior high school certificate).

Finally, the growth rate of earnings in each occupation is computed by regressing the log of earnings on age for men between the ages of 31 and 50 (see Appendix B.5 for further details). The challenge here is to separately identify the wage growth due to selection - that is, remaining unemployed longer to get a good earnings draw - and that governed by the underlying growth parameter. As transition into the labour market has been completed by age 30, I use the age range 31-50 to compute the growth parameters. Estimation of the structural earnings parameters thus relies on the assumption that the growth rate of earnings is constant over the life-cycle.

Table 2.6: External Parameters 1

Parameter	Description	Value	
β	discount factor	0.95	
r	real interest rate	5%	
F_{a0}	initial wealth endowment	from GLSS 2012	
E_1^E	cost of high education	1200 GHc ($\approx$ 250 USD)	
E_S	cost of entrepreneurship	210 GHc ($\approx$ 44 USD)	
T	number of periods	35	
T_E	time (period) cost of high education	4	
H^H	level of high ability	1	
		educated=0	educated=1
$G_{w,k}$	earnings growth	1%	3%
$G_{a,k}$	earnings growth	0%	5%
$G_{e,k}$	earnings growth	1%	2%

Estimation of the model's structural parameters is conducted as follows: first, I select the sub-sample of all male labour-force participants aged 16-30. As the model focuses on the initial transition from education to the labour-force, this is the age category whose characteristics I aim to match in the estimation. This yields a sample size of $N = 3728$. I proceed to solve the model by backward induction for an initial parameter set. This yields a full set of simulated outcomes for each sampled individual over his life-cycle. As the GLSS 2012 is a cross-sectional dataset, however, I then convert the simulated data into a comparable cross-sectional form based on the age at which each individual is observed in the empirical data. The next step is to construct the moments required for estimating the structural parameters, for both the simulated and empirical data. Simulated moments are computed for labour-force participants aged 16-30, excluding those currently in education, in order to match the data. Table 2.7 lists these parameters, while Tables 2.8 and 2.9 display the moments used to estimate their values.

The estimated parameters minimise a quadratic loss function that measures the weighted distance between the set of moments from the simulated dataset and their empirical counterparts. I use the Nelder-Mead algorithm to simulate the model (and obtain the set of simulated moments) repeatedly, varying the set of structural parameters until the loss function is minimised, such that the simulated moments are as close as possible to the empirical ones. Formally, the simulated moments estimator $\hat{\theta}_{S,N}(W)$ solves:

$$\hat{\theta}_{S,N}(W) = argmin_{\theta}[\hat{\psi}_N^d - \hat{\psi}_N^s(\theta)]' W_N [\hat{\psi}_N^d - \hat{\psi}_N^s(\theta)] \tag{2.13}$$

in which S is the number of simulations for a given parameter set (I choose $S = 10$), N is the sample size, $\hat{\psi}_N^d$ is the set of empirical moments, $\hat{\psi}_N^s(\theta)$ is the set of simulated moments, and W is the weighting matrix. For the last of these, I use the inverse of the variance-covariance matrix of the empirical moments. This puts more weight on moments that are precisely estimated, while also adjusting the weighting for correlation between moments.

The moments employed in estimation are as follows: the means and standard deviations of log earnings in each of the six occupations, the proportion of workers in each occupation, the proportion of highly-educated workers in the sample, the proportion of unemployed workers aged 16-24 by education level, the proportion of low-education workers in the top wealth decile, and the mean log earnings ratio of workers in the top wealth decile by sector. The last two sets of moments are particularly useful in estimating the parameters of the ability distribution; identification of these parameters thus relies on the assumption that ability is uncorrelated with the initial distribution of parental wealth. While the model simulates the schooling process of those who choose high education, simulated moments are computed using only the sub-sample of out-of-school workers, in order to ensure comparability with the data moments, which are computed for exactly this group of workers.

2.4.4 Estimation Results and Model Fit

The estimation results are presented below. Table 2.7 shows the estimated parameter values, and Tables 2.8 and 2.9 list the values of the matched model moments alongside their empirical counterparts and a 95% confidence interval around them. Standard errors (in parentheses) are bootstrapped using 200 replications.

Table 2.7: Structural Parameters 1

Parameter	Description	Estimate	
q_L	proportion of low types	0.42 (0.01)	
H^L	low ability level	0.54 (0.03)	
		educated=0	educated=1
$\mu_{w,k}$	mean log earnings (wage-emp.)	0.86 (0.06)	4.87 (0.08)
$\mu_{a,k}$	mean log earnings (agric.)	4.56 (0.03)	3.03 (0.27)
$\mu_{e,k}$	mean log earnings (entrep.)	0.11 (0.01)	0.73 (0.05)
$\sigma_{w,k}$	sd log earnings (wage-emp.)	3.79 (0.11)	2.97 (0.12)
$\sigma_{a,k}$	sd log earnings (agric.)	2.02 (0.03)	2.55 (0.14)
$\sigma_{e,k}$	sd log earnings (entrep.)	4.10 (0.06)	3.69 (0.11)

One of the most striking features of this set of results is the degree of labour-market selection that underlies it. The wage offer distributions for all occupations have high estimated standard deviations; in fact, with the exception of high-education wage-employment and the two

Table 2.8: Matched Moments 1

	educated=0			educated=1		
Moment	Fitted	Data	{LB,UB}	Fitted	Data	{LB,UB}
prop. (wage-emp.)	0.13	0.13	$\{0.12, 0.14\}$	0.17	0.14	$\{0.13, 0.15\}$
prop. (agric.)	0.46	0.45	$\{0.44, 0.47\}$	0.06	0.08	$\{0.07, 0.09\}$
prop. (entrep.)	0.11	0.09	$\{0.08, 0.10\}$	0.03	0.04	$\{0.03, 0.04\}$
mean log earn. (wage-emp.)	7.26	7.76	$\{7.66, 7.86\}$	8.07	8.03	$\{7.92, 8.13\}$
mean log earn. (agric.)	6.67	6.54	$\{6.45, 6.63\}$	7.25	6.64	$\{6.63, 6.90\}$
mean log earn. (entrep.)	7.41	7.85	$\{7.66, 8.03\}$	7.76	7.99	$\{7.75, 8.23\}$
sd log earn. (wage-emp.)	1.68	1.04	$\{0.97, 1.14\}$	1.69	1.12	$\{1.02, 1.24\}$
sd log earn. (agric.)	1.14	1.17	$\{1.13, 1.22\}$	1.27	1.25	$\{1.14, 1.38\}$
sd log earn. (entrep.)	1.77	1.52	$\{1.36, 1.79\}$	1.46	1.29	$\{1.12, 1.30\}$

note: UB and LB are the upper and lower bounds of a 95% confidence interval around the data moment

Table 2.9: Matched Moments 2

Moment	Fitted	Data	{LB, UB}
proportion of high-educ. workers	0.28	0.29	$\{0.28, 0.31\}$
proportion of low-educ. workers in top wealth decile	0.29	0.33	$\{0.28, 0.38\}$
high/low-educ. wage earnings ratio in top wealth decile	1.15	1.08	$\{1.02, 1.15\}$
high/low-educ. agric. earnings ratio in top wealth decile	1.08	0.99	$\{0.83, 1.19\}$
high/low-educ. entrep. earnings ratio in top wealth decile	1.04	0.99	$\{0.86, 1.07\}$
proportion of low-educ. unemployed workers aged 16-24	0.15	0.16	$\{0.13, 0.19\}$
proportion of high-educ. unemployed workers aged 16-24	0.06	0.06	$\{0.05, 0.09\}$

note: UB and LB are the upper and lower bounds of a 95% confidence interval around the data moment agricultural occupations, the standard deviation of each occupation exceeds its corresponding mean. By contrast, the accepted wage distributions (see Table 2.8) have low standard deviations and high means relative to their offered counterparts. This demonstrates the degree to which individuals, conditional on wealth, are willing to wait in unemployment in order to secure a higher-paying job.

The model fits the data quite well; almost all the moment estimates lie within a 95% confidence interval around their empirical counterparts, as Tables 2.8 and 2.9 show. Exceptions are the mean of log earnings in high-education agriculture, and the standard deviation of log earnings in the wage sector, which the model struggles to fit. The fit of the remaining earnings moments is quite good, and the unemployment moments, in particular, are very closely matched.

Figure 2-12: Log Wealth and Labout Market Outcomes

The manner in which the level of parental wealth influences youth labour market outcomes is the central concern of this paper. Figure 2-12 thus shows the positive correlation between parental wealth and each of the following: education choice, life-cycle earnings and probability of unemployment. As before, results from the simulated model provide a good match to their empirical counterparts.

The probability of education and of unemployment are both increasing in wealth. Average lifetime earnings are also increasing in wealth, and this effect is strengthened by conditioning on employment - this is as expected, given that wealthy workers are more likely to experience unemployment in their youth.

2.4.5 Wealth Effects on Education and Employment Choice: Consequences for Equity and Efficiency

These empirical results draw attention to the fact that differences in initial parental wealth underlie a strong selection effect in terms of education choice, sectoral choice and lifetime earnings, driven by a combination of labour market search frictions and credit constraints on workers. As discussed in the model set-up, wealth affects these outcomes through two main channels: education choice and unemployment duration, which jointly offer access to higher-productivity jobs.

Figure 2-13 illustrates the importance of the latter channel in this context, by comparing the distributions of offered and accepted earnings in each occupation. In each case, the distribution of accepted earnings lies significantly to the right of the offer distribution, indicating a preference among workers for rejecting low-productivity offers and waiting in unemployment for better ones.

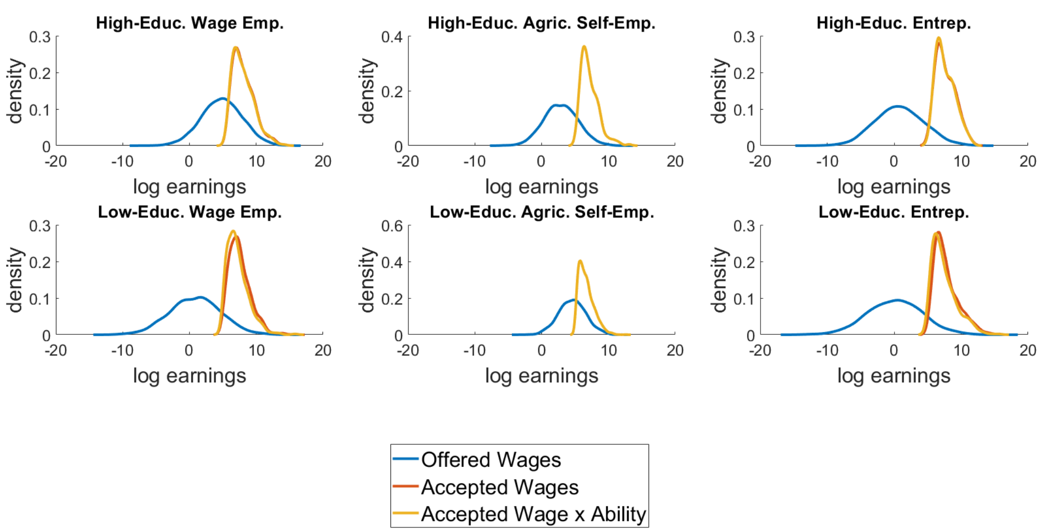


Figure 2-13: Job Selection Effects by Occupation

Table 2.10 decomposes the wealth effect into its two channels: education choice and unemployment duration. This is achieved by regressing the log of annual earnings (averaged over the life-cycle, from age 16 to age 50) in the simulated model on the log of parental wealth, shutting down each of the education and unemployment channels in turn. Thus, the first column of Table 2.10 shows the total effect of wealth on earnings; the second shows the effect of wealth on earnings when all workers accept employment offers immediately, with no time spent in unemployment, and leaving education choices unchanged from (1);

finally, the third column shows the effect of wealth on earnings when all workers are low-educated. This decomposition reveals that 63% of the effect of wealth on earnings comes through an increased probability of being highly-educated, while the remaining 37% is attributed to higher unemployment among the wealthy. Thus, the ability to wait for a good job is an important component of the earnings premium gained by workers from wealthier family backgrounds.

Table 2.10: Wealth Effect Decomposition

	(1) *total wealth effect* log annual earnings	(2) *education effect* log annual earnings	(3) *unemployment effect* log annual earnings
log parental wealth	0.38^{***}	0.24^{***}	0.14^{***}
	(0.02)	(0.02)	(0.02)
% of total wealth effect	100	63	37
N	3728	3728	3728

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Differences in initial (parental) assets therefore drive a significant inequality in earnings. Furthermore, the combination of wealthy individuals delaying entry into the labour-market and accessing jobs not only with higher earnings levels but also higher earnings growth than their low-asset counterparts means that this inequality is increasing over the life-cycle, as Figure 2-14 shows. Individuals in the lowest wealth decile actually begin with higher earnings than those in the top wealth decile (due to high early-life unemployment among the latter group), but are overtaken by their mid-20s. Thus, the earnings gap between the richest and poorest 10% of the wealth distribution increases over the life-cycle, thereby exacerbating the problem in the next generation, whose starting wealth is precisely a function of these, more unequal, earnings.

Recall that the theoretical framework predicts an ambiguous relationship between ability and unemployment, whose overall sign is governed by the relative magnitudes of an "asset de-accumulation" effect and a "sectoral shift" effect. When high-ability workers are relatively more impatient than low-ability workers to become employed, the issue of earnings inequality is somewhat ameliorated: the earnings gap between low- and high-ability workers is then lower compared to the case with positive assortative matching between workers and jobs and, consequently, the aggregate earnings distribution is relatively compressed. This is due to the complementarity between worker and firm productivities in the returns to

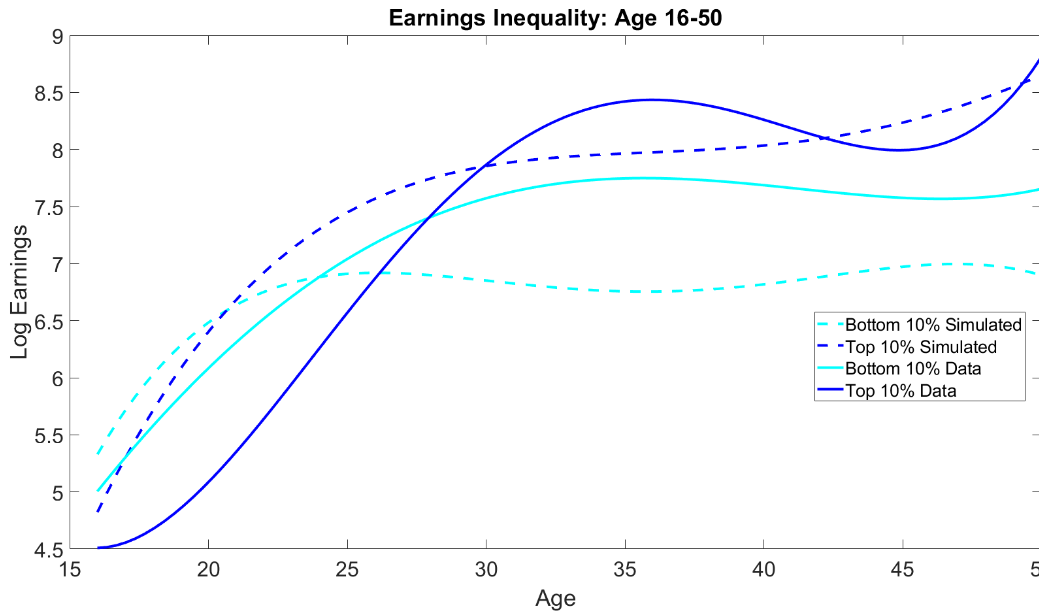


Figure 2-14: Earnings Inequality Over the Life-Cycle

wage-employment and entrepreneurship. By contrast, when low-ability workers exit unemployment more quickly than their high-ability counterparts, the earnings gap widens. As Figure 2-15 shows, under the estimated parameters for Ghana, the "asset de-accumulation" channel is dominant over the "sectoral shift channel", such that low-ability workers have a longer school-to-work transition than their high-ability counterparts, except at the very top of the wealth distribution.[17] Thus, the earnings gap is smaller than it would have been in the opposite case.

The effect of wealth on individual decision-making has undesirable consequences, not just for equity, however, but also for efficiency. The first, and most obvious, manifestation of this is the fact that the search frictions reduce total productivity compared to a case in which they did not exist, partly because workers must now spend in unemployment time that could have been spent on productive labour, and partly because workers now accept lower-productivity jobs because they cannot afford to wait for better ones. Under these parameter estimates, workers spend an average of 0.98 years in unemployment.

The complementarities in the production functions for wage and entrepreneurial employment mean that this issue of lower total productivity is exacerbated whenever high-ability workers are relatively more reluctant to remain unemployed (which, as previously discussed, is true in this case). This is because, even though high-ability workers earn more on average (they have a higher total match productivity as earnings in the wage and entrepreneurial sectors

[17]See Appendix B.6 for a discussion of why the relationship between ability and unemployment varies along the distribution of initial wealth

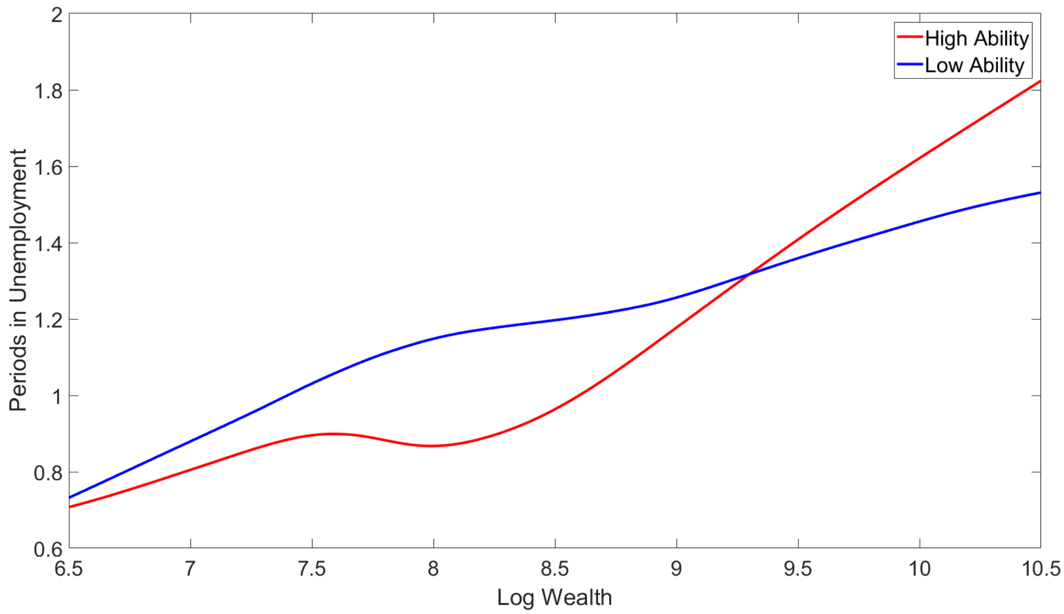


Figure 2-15: Log Wealth and Unemployment Duration by Ability

are complementary in firm productivity and worker ability), low-ability workers match with weakly higher-productivity firms. This is illustrated in Figure 2-16; only at higher levels of wealth do high-ability workers catch up with their low-ability counterparts in terms of the productivity of matched firms. Thus, the size of the welfare loss is exacerbated by the fact that it is low-wealth, *high-ability* workers who are unable to access high-productivity jobs.

Finally, wealth affects efficiency through determination of education choice: to the extent that high-productivity jobs are located among the high-education occupations, low-wealth workers are inefficiently matched to low-productivity jobs in low-educated occupations.[18] The fact that 45% of sampled workers are employed in low-education agricultural production, the occupation with the lowest average earnings, clearly exemplifies this problem.

[18]The model's prediction that low-wealth workers fail to invest in education when returns are highly dispersed is in line with Banerjee et al. (2016), who use a cross-country approach to make a similar argument: namely, that relatively high unemployment risk in high-education jobs located in poor countries is an important determinant of low educational take-up in the presence of high returns to education. Relatedly, Meghir et al. (2015) show, using a search-and-matching framework estimated on Brazilian data, that the existence of an informal sector contributes significantly to the number of workers in low-productivity employment, thereby reducing output and welfare.

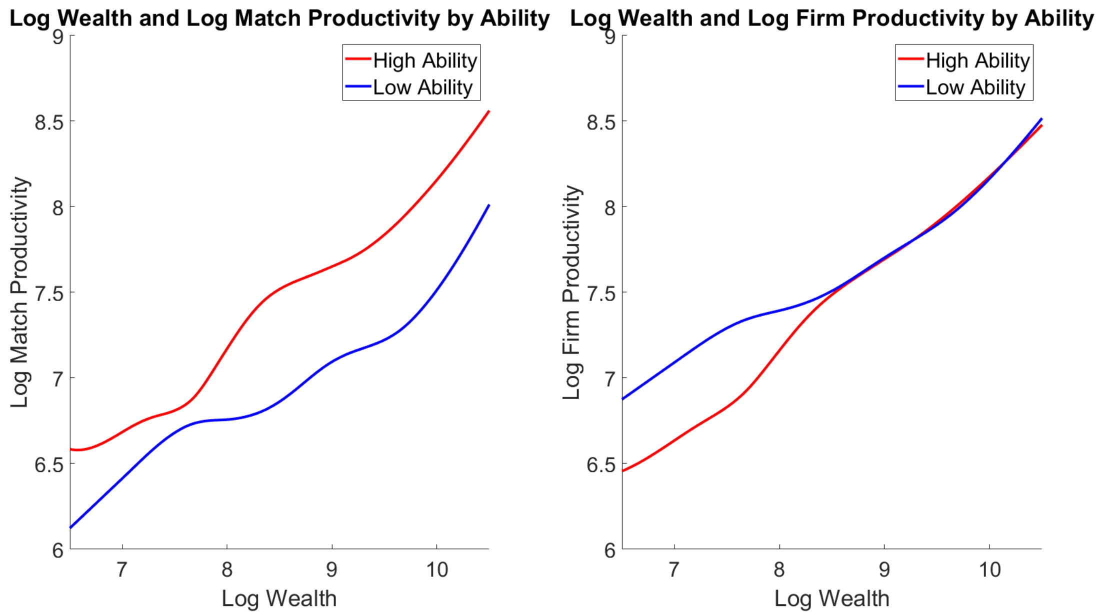


Figure 2-16: Log Wealth and Log Productivity by Ability

2.5 Counterfactual Policy Simulations

In this section, I examine the effects of a pair of policy interventions – namely, an education subsidy and unemployment insurance – that target earnings inequality and aggregate productivity. In particular, I consider which of these policies has the greatest impact when implemented by a government with a resource constraint. I denote the government's total policy budget by G.

With a fixed resource constraint, the government may face a trade-off between equity and efficiency in selecting the optimal policy target group. If solely concerned with improving the equity of labour market outcomes, it may target resources towards the poorest individuals. By contrast, a government motivated by improving aggregate productivity could focus its resources on improving the job prospects of high-ability workers, given the complementarity between worker and firm productivity in the wage and entrepreneurial sectors. Thus, in the discussion that follows, I consider three targeting options for each policy: first, a baseline policy in which the policy is offered to everyone, second, an alternative in which the policy is offered only to individuals of high-ability, and finally, an alternative in which the policy is offered to low-wealth individuals.

Tables 2.11 and 2.12 show the effect of the policy interventions considered in this section on a number of key outcomes: educational attainment, youth unemployment, income inequality and aggregate productivity. The first column of results in each table shows the baseline outcome, and the following rows show the outcome for the relevant policy counterfactual under each of the three targeting options.

My measure of inequality is the "20/20 ratio" (the ratio of average earnings in the top 20% of the income distribution to average earnings in the bottom 20%). This measure is used for the United Nations Development Programme's Human Development Indicators. The baseline value of inequality is 25.74: that is, the top 20% of earners earn 25.74 times more on average than the bottom 20%. Aggregate productivity is measured by the sum of individuals' average income across all years in the labour-force.

2.5.1 Counterfactual Simulation 1: Subsidise Education Cost

The first policy I examine is a subsidy on the cost of education. This mimics a policy actually implemented by the Ghanaian government: in 2017, it introduced a 100% subsidy on senior high-school education, in an attempt to raise educational attainment. The parsimonious model framework of this paper does not, however, allow for a consideration of the labour market effects that such a policy would generate as a result of increasing the supply of educated workers. Thus, I consider instead a 20% education subsidy, in order to avoid this additional complexity.

First, I set the government's policy budget, G, at 396,960 Ghanaian cedis (approximately US$ 82,324). This is the total cost of a 20% education subsidy when offered to the entire sample of 3728 individuals, of whom 1654 take it up; I fix G at this value for the remainder of this section in order to simplify comparison across policies.[19]

As previously discussed, I consider three separate targeting options: first, the government offers a 20% subsidy to all individuals, regardless of their characteristics. Second, it aims to improve worker-firm matching efficiency by targeting the subsidy at high-ability workers. This change in targeting raises the value of the subsidy to 21%, when G is held constant. Finally, the government attempts to address equity concerns by targeting the subsidy at individuals drawn from the bottom of the initial wealth distribution. There are now two options for determining the eligibility cut-off: first, retain the subsidy value used when targeting the group of high-ability workers and allow differential compliance to the policy, or second, adjust the subsidy value such that the same number of targeted individuals (who chose low education in the baseline) take up education as under efficiency targeting. I employ the former option, in order to compare the targeting alternatives in terms of the number, as well as the composition, of individuals who respond to the subsidy by altering their education choices.

[19] As shown in Table 2.11, a total of 1654 individuals choose high education when the subsidy is offered to everyone - the total cost outlay is thus $0.2 * 1200 * 1654 = 396,960$ GHc.

Table 2.11: Counterfactual 1 (Education Subsidy): Policy Outcomes

	baseline	target all	target efficiency	target equity
Proportion of Educated Workers	0.2751	0.4013	0.3967	0.4053
(% change)		(46%)	(44%)	(47%)
Proportion of Unemployed Workers (Age 16-25)	0.0845	0.0909	0.0903	0.0911
(% change)		(8%)	(7%)	(8%)
20/20 ratio (Income Inequality)	25.74	27.38	27.51	27.75
(% change)		(6%)	(7%)	(8%)
Aggregate Productivity	65946.78	82198.89	82465.08	82813.97
(% change)		(25%)	(25%)	(26%)

The first, and perhaps most obvious, outcome of this policy intervention is that the proportion of highly-educated workers increases by 12 percentage points, from 28% to 40%. The newly-educated are almost all high-ability individuals; less than 2% of low-ability workers respond to the subsidy when offered it (see Figure 2-17). Further, youth unemployment increases slightly, from 8% to 9%. Two mechanisms underlie this: firstly, more workers now enter a labour sub-market in which there are greater benefits to waiting in unemployment, and secondly, a substitution effect – some recipients of the subsidy chose to be educated even in the baseline case; thus, they now channel the savings made on education costs into funding a lengthier school-to-work transition. While higher youth unemployment may seem undesirable from a policy-maker's perspective, it is important to observe that any demand-side interventions in this labour-market framework must raise unemployment in order to achieve either (or both) higher output and lower inequality.

The expected effect on income inequality of this policy is ambiguous: first, we might expect the education subsidy to reduce inequality by allowing lower-wealth individuals to access better-paying (high-education) jobs. Conversely, however, targeting the subsidy at everyone results in the substitution effect just described, which allows wealthier individuals to wait longer in unemployment for even better jobs. As Table 2.11 shows, the aggregate result is an increase in inequality. This is unsurprising given that individuals in the bottom 20% of the wealth distribution do not respond to the subsidy (see Figure 2-17); thus, the inequality measure reflects only the income gains at the top of the distribution.

Finally, I consider the effects of the policy on aggregate productivity. Unsurprisingly, there is an increase (of 25% above the baseline) in productivity, as workers responding to the policy change access better jobs, through a combination of accessing occupations with higher returns and waiting longer for them.

Strikingly, for this policy, changing the targeting group does not appear to have much effect on these key outcomes. This is because the composition of workers induced to alter their education choice in response to the subsidy is virtually the same under all three targeting options, as take-up is almost exclusively among high-ability workers.

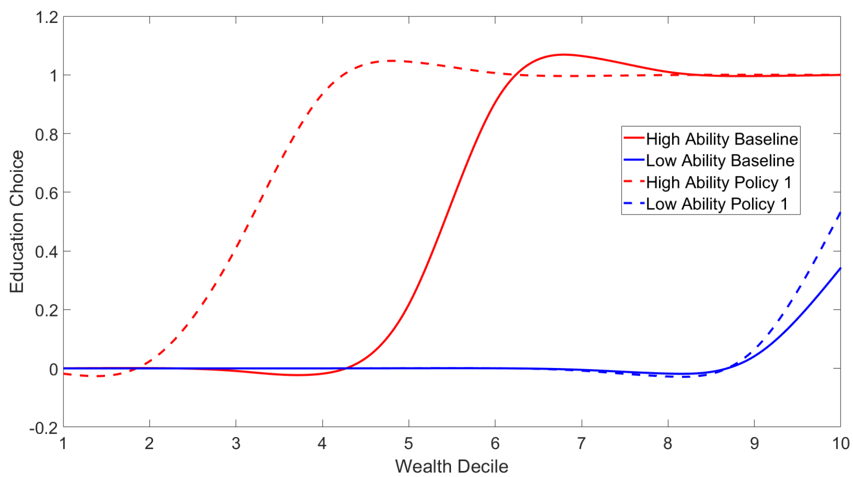


Figure 2-17: Education Choice by Wealth and Ability

In summary, therefore, we see that, while some high-ability workers respond to the education subsidy, take-up among low-ability workers is very low, such that the aggregate effect on productivity is muted and, further, that income inequality rises. These results are very much in keeping with the predictions of the model, which are that an education subsidy alone is unlikely to be a sufficient policy tool in this context. This is because, when returns to educated employment require a lengthy wait in unemployment, take-up of education may remain relatively low even when subsidised. Further, even if the subsidy induces additional take-up of education, the fact that initial wealth continues to drive job quality conditional on education means that it has limited effects on aggregate productivity and income inequality.

2.5.2 Counterfactual Simulation 2: Introduce Unemployment Insurance

The second policy I consider is the introduction of unemployment insurance.[20] Table 2.12 shows the results of this intervention. As before, I allow for three different types of targeting.

[20] see Shimer and Werning (2007) and Shimer and Werning (2008) for a discussion of optimal unemployment insurance when workers are risk-averse.

In the first case, with everyone eligible for UI, the per period payment is 105 GHc. In the remaining two cases, where the target group is either high-ability or low-wealth individuals, the relevant amount is 188 GHc.

Table 2.12: Counterfactual 2 (Unemployment Insurance): Policy Outcomes

	baseline	target all	target efficiency	target equity
Proportion of Educated Workers	0.2751	0.2790	0.2790	0.2790
(% change)		(1%)	(1%)	(1%)
Proportion of Unemployed Workers (Age 16-25)	0.0845	0.0889	0.0893	0.0888
(% change)		(5%)	(6%)	(5%)
20/20 ratio (Income Inequality)	25.74	25.51	25.66	25.33
(% change)		(1%)	(0%)	(2%)
Aggregate Productivity	65946.78	66679.07	67038.09	66790.92
(% change)		(1%)	(2%)	(1%)

Firstly, the effect of this policy on educational attainment is negligible: the model predicts that raising consumption in unemployment (here, via unemployment insurance) can alter education choices by making individuals more likely to enter a labour sub-market in which returns are highly dispersed, but this result does not materialise. This is due to the small size of the policy change; increasing the per-period payment to 1000 GHc, for instance, causes the proportion of highly-educated workers to fall to zero, and the unemployment rate to rise sharply, to 28%, as Table 2.13 shows.[21]

Second, the policy change has a comparable effect on youth unemployment to that of the education subsidy – the proportion of unemployed 16-25 year-olds increases to approximately 9% - a 5% change. The direction of this effect is as expected – unemployment insurance allows individuals to remain unemployed longer in order to access higher-productivity jobs.

It follows then that the policy should also reduce income inequality and increase aggregate productivity; Table 2.12 shows that this is, in fact, what occurs. The increase in productivity is much smaller than under the education subsidy (a 1% change compared to 25%) but, while the larger increase in the latter case came at the expense of higher income inequality, there is

[21] The proportion of highly-educated workers drops to zero because workers face a trade-off in terms of expending resources on education and on drawing out the time spent in unemployment. Given the fixed cost of education and the relative means and variances of high-education and low-education earnings, the introduction of relatively high-valued unemployment insurance makes it optimal for workers to remain at a low level of education and allocate more resources towards staying unemployed, rather than to expend them on higher education.

no equity-efficiency trade-off under this policy option. This is because the resulting increase in unemployment is proportionally larger at the bottom of the wealth distribution than at the top, as Figure 2-18 shows.[22]

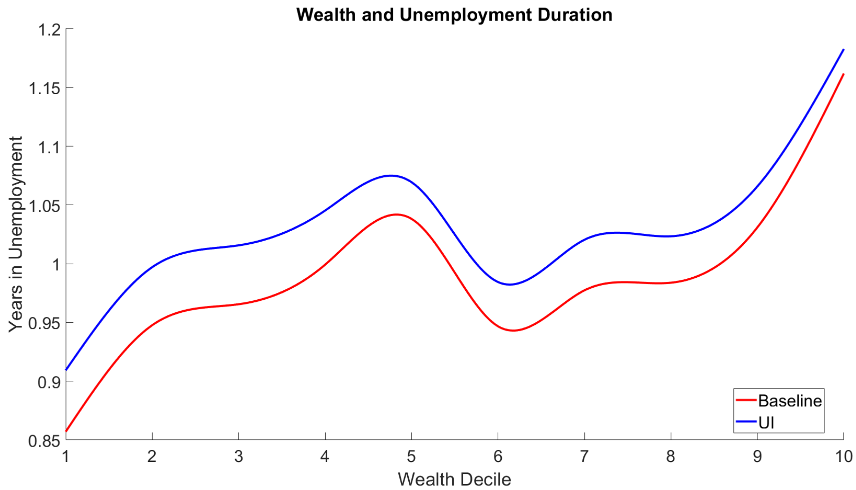


Figure 2-18: Unemployment Duration and Wealth

As expected, targeting the policy to high-ability individuals is most effective in raising aggregate productivity, while targeting low-wealth individuals is better at reducing income inequality. Finally, targeting everyone is a strictly dominated option.

Table 2.13: Key Outcomes with $UI = 1000$ GHc

	high-educated (%)	unemployed (16-25) (%)	aggregate prod. 1000s of GHc	20/20 ratio
	0	27.81	101454.39	11.80
% change from baseline	(100%)	(229%)	(54%)	(54%)

In conclusion, therefore, with a fixed resource constraint of 396,960 GHc, an education subsidy is the best option in terms of increasing aggregate productivity, while unemployment insurance is the optimal choice for reducing income inequality. Thus, the optimal policy prescription for a resource-constrained government will depend on its preferences over equity and efficiency; in the absence of such constraints, a combination of the education subsidy and unemployment insurance would be effective in tackling both issues.

[22]Relatedly, Browning and Crossley (2001) show that unemployment insurance improves consumption smoothing among the poor but not the wealthy.

2.6 Conclusion

In this paper, I set up and estimate a discrete-choice model of education and occupation choice, in order to quantify the relationship between youth unemployment and parental wealth in Ghana. I show that this relationship is positive, in line with descriptive evidence from studies of developing-country labour markets. This is because, as in McCall models of search, higher levels of liquid assets allow workers to remain unemployed for longer, in an attempt to secure a high-productivity job. Using the results from a structural estimation of the model's parameters, I show that this waiting behaviour leads to a number of undesirable labour market outcomes, compared to a case without credit or search frictions in the labour market. These are, firstly, a high degree of income inequality that grows over the life-cycle; secondly, low average educational attainment, as high dispersion in the returns to education discourages investment in education by workers who cannot afford to wait in unemployment for long; and finally, low aggregate productivity. This last outcome is due to a combination of two factors: first, output loss from time spent waiting in unemployment, and second, output loss from inefficient worker-firm matching, as high-ability workers are less willing to wait in unemployment than their low-ability counterparts and, as such, match with lower-productivity firms.

Further, I use the estimated model to decompose the effect of parental wealth on average lifetime earnings into an education and a waiting channel, and show that the latter is an important channel, accounting for 37% of the total effect. Finally, I evaluate the effects of two alternative labour-market policies: an education subsidy and unemployment insurance. I find that the first of these is most effective at raising aggregate productivity, but does so at the expense of also raising income inequality. Unemployment insurance has a smaller impact on productivity, but has the additional benefit of lowering income inequality. The optimal policy tool thus depends on the government's preferences over efficiency and equity.

The results of this paper suggest a number of important considerations for policy. First, policy-makers seeking to improve educational attainment in such a setting would need to take account of the additional, indirect costs of investment in education caused by the need to wait for "good" jobs, which are disproportionately high for the poor. Second, in terms of optimal labour market policy, it it clear that demand-side policies are a second-best solution to the problems generated by wealthy workers waiting for better jobs, and each has its own shortcomings: education subsidies increase inequality, while unemployment insurance may improve worker-firm match efficiency and inequality, but results in an output loss due to more time spent in unemployment. By contrast, policy efforts focused on improving the average quality of high-education jobs are more likely – by targeting labour market frictions directly

– to be effective.

Possible directions for further research on this subject are to consider the ways in which parental wealth affects the likelihood of unemployment later in life, rather than simply during the transition from school to work; to expand the model framework to allow for job-to-job transitions, in order to understand how the possibility of on-the-job search affects the relationship between parental wealth and youth unemployment; and to expand the employment decision space to include the possibility of labour market inactivity, in order to consider how parental wealth may affect the decision to delay labour market entry (i.e. searching for employment) among youth, rather than simply the choice between unemployment and employment for those currently active in the labour market.

Chapter 3

Learning from Rejection: Information Cascades in Organ Transplantation

Abstract

This paper examines the consequences for efficiency of information cascades among UK hospitals in organ transplant queues. We build a structural model in which organs are sequentially assessed by centres of heterogeneous ability, and show that herding behaviour plays an important role in observed organ wastage: once an organ is rejected by one or more centres, subsequent centres emulate their behaviour, ignoring their own assessment of the organ's quality. We employ administrative data from the NHS Blood and Transplant (NHSBT) that covers the universe of abdominal organs donated in the UK during 2006-2016 to provide reduced-form evidence of herding behaviour among transplant centres. Further, we undertake a set of counter-factual analyses to demonstrate that, while herding behaviour is common among UK transplant centres, the resulting increase in discard rates is not substantially higher than that of the full-information benchmark. Equally, it performs the important function of preventing centres from accepting organs of poor quality such that, overall, the benefits derived from observing predecessors' decisions outweigh the costs of herding traditionally emphasised in the theoretical literature. In contrast with this literature, therefore, we find that, in this context, herding is efficiency-enhancing overall.

Keywords: Social learning, Herd behavior, Information Cascades, Organ transplant decisions

JEL codes: J12, J16, D31, I3

3.1 Introduction

Allocating donated organs in an efficient and equitable manner to recipients on a waiting list is the primary objective of any organ transplant program. In the United Kingdom, the organ transplant program is organized around a nationwide network of centers (hospitals). When a deceased-donor organ becomes available, all patients on the National Transplant Registry are assigned a priority rank based on a predetermined allocation algorithm. Transplant centers are offered the organ in order of their patients' priority, until the organ either is accepted or, having deteriorated with time, is no longer viable and, as such, is discarded. Currently, demand outstrips the availability of both livers and kidneys, the two organs that dominate transplantation activity in the United Kingdom and that constitute the focus of our analysis. National Health Service Blood and Transplantation (NHSBT) statistics indicate that, five years after being listed, approximately 20% of patients on the National Registry have either died or been removed from the waiting list, as their condition has deteriorated below the minimum eligibility criterion for transplantation. Simultaneously, however, approximately 50% of livers and 30% of kidneys are discarded. Organs may be discarded because they are of poor quality and, thus, are unsuitable for transplantation. In this paper, we examine an alternative possibility, namely that some viable organs fail to be utilized because, once an organ has been rejected by one or more centers, subsequent centers ignore their own assessment of organ quality and herd in line with their predecessors.

There is a growing economics literature on organ transplantation and two-sided matching, including Roth et al. (2004), Roth et al. (2005) and Roth et al. (2007). This literature, which is largely focused on organ transplants that involve live donors, studies the manner in which donors and recipients are optimally matched in the absence of market mechanisms (which, in the case of organs, do not exist for obvious ethical reasons). Our research shifts the organ transplant literature in a new direction, by focusing on the information-based inefficiencies that may arise when deceased-donor organs are offered sequentially to centers on a waiting list. Such information-based inefficiencies may arise, more generally, whenever an object is offered sequentially to decision-makers and past decisions are public knowledge. Bikhchandani et al. (1992) motivate their seminal paper on herding with an example in which a research paper, once rejected by one or more journals, is more likely to be rejected by other journals as well. The labor market is another obvious setting for this phenomenon. In this paper, we provide a strategy to detect information-based herding in settings in which decisions are characterized by a sequence of rejections, followed either by an acceptance or by termination of the process, as well as a methodology for quantifying the efficiency consequences of such herding.

In order to understand why transplant centers might rationally condition their decisions on those of the centers that preceded them, and why such herding behavior could generate inefficiencies, suppose that there are two types of organ: good (G) and bad (B). Each transplant center makes an assessment of the quality of the organ that is made available to it. This assessment – which, to be consistent with the existing literature on herding, may equivalently be treated as an information signal – is not directly observed by other transplant centers. As in Bikhchandani et al. (1992) and Anderson and Holt (1997), we assume that signals are binary: good (g) and bad (b). Centers are not systematically misinformed; they are thus more likely to receive a g (b) signals when an organ is good (bad). For expositional convenience, we assume that transplant centers have a common negative prior about the quality of offered organs; thus, in the absence of any other information, each center's decision is to reject the organ. Additionally, we assume that a g signal dominates a center's negative prior, such that it always accepts the offered organ upon receipt of such a signal.

It follows that the first center to be offered an organ rejects it on receipt of a b signal (which reinforces its prior), but accepts it on receipt of a g signal. The second center in line is offered the organ only if the first center rejects it, and knows that the first center only rejects after a b signal. Thus, if the second center also receives a b signal, this reinforces the information contained in both the first center's b signal and the negative prior, and it will certainly reject the organ. If, however, the second center receives a g signal, it knows that its signal is not aligned with that of the first center. Under the assumption that all transplant centers receive signals of equal precision (i.e. that they are all equally competent in assessing the quality of an organ), the first and second centers' signals cancel each other out, and the second center also rejects (based on its negative prior). Next, consider the third center's decision: it knows that the second center rejects the organ regardless of its signal, such that the second center's decision gives the third center no additional information. Accordingly, the third center now behaves as if it were second in line. Following the preceding argument, it also rejects the organ, regardless of the signal it receives. This process is replicated along the entire waiting list, regardless of the sequence of signals received by centers.

While it is individually rational for centers to ignore their signals in this manner, herding can result in the under-utilization of viable organs. To see why this is the case, suppose that the first center in line for an organ receives a b signal, but all the centers that follow receive g signals. This organ is rejected by all centers, despite the fact that it is evidently a G organ. The phenomenon just described is referred to in the literature as an "information cascade". The pioneering papers in this literature, Banerjee (1992) and Bikhchandani et al. (1992) also provide similar illustrations of information cascades. These cascades are a special case of a general phenomenon known as "herding", in which agents condition their decisions on their predecessors' decisions (Çelen and Kariv (2004)). The distinguishing feature of an information cascade is that agents completely ignore their own signals when they herd behind their predecessors; consequently, agents that follow them in the decision-making process can learn nothing from their decisions, resulting in an information inefficiency.

The identification of information-based herding is a challenging statistical problem because, for many reasons, agents' decisions may be correlated. For example, a strand of the finance literature, which includes Lakonishok et al. (1992), Grinblatt et al. (1995) and Wermers (1999), has interpreted clustering in asset investment decisions as being indicative of herding. Such clustering does not, however, imply that agents learn from each other; rather, agents could make the same decisions simply because they have correlated characteristics or receive correlated information signals. More recently, Cipriani and Guarino (2014) estimate a structural model of asset investment decisions to examine herding in financial markets. Their identification strategy relies on a particular (i.i.d.) specification of unobserved information shocks within a day. If, given this specification, investors are found to trade against their own information and systematically to follow their predecessors' decisions, the authors infer that herding is present. As the timing of decisions is endogenous (not random, as assumed in the model), investors who choose to buy or sell at particular times of the day likely receive correlated signals. Clustering in investment decisions within a day would then spuriously be attributed to information-based herding.[1]

There is a long tradition in development economics of using theoretical models to derive robust tests of herding in the adoption of new agricultural technology, with particular attention paid to potential omitted variable biases of the sort described above; examples include Foster and Rosenzweig (1995), Munshi (2004), Conley and Udry (2010). Although this literature provides credible evidence of herding, the information frictions incorporated into the models of technology adoption are distinct from those in the canonical theoretical models of information cascades. Specifically, farmers draw signals from a continuous distribution and make decisions along a continuum. Thus, neighbors' signals can perfectly be recovered from their decisions. Informational inefficiencies arise because neighbors' signals are less informative than the farmer's own signals. Identifying information cascades, which arise when signals and decisions are binary, requires novel tests. An important contribution of our research is to develop such tests, which are based explicitly on the fact that agents ignore their signals entirely when information cascades occur (such that their decisions are useless to the agents

[1] In general, structural models (conditional on having been previously validated) are most useful for quantification. In this paper, we independently validate our model with reduced-form tests before proceeding to the structural estimation.

who follow them).

The most direct implication of herding is that agents' decisions should be correlated with their predecessors' decisions. Quite apart from the alternative non-learning interpretations of these correlations, the structure of our data does not even allow this test to be implemented. This is because there is so little variation in transplant decisions – an acceptance is observed once, at most, for each organ (and only ever at the end of a sequence). To identify herding, which, in our setting, is equivalent to information cascades, we thus incorporate heterogeneity in center ability. In particular, suppose that centers are distinguished by their ability to identify good and bad organs. In our model, this implies that the precision of the information signal varies across centers. We would then expect a center in second position to be more likely to reject the organ when it follows a higher-ability center in first position, compared to the case in which it follows a lower-ability center in the same position. Letting p_2 be the probability that center 2 rejects (conditional on being offered the organ) and q_1 be center 1's ability, this implies $\alpha_1 > 0$ in equation (3.1):

$$p_2 = \alpha_0 + \alpha_1 q_1. \tag{3.1}$$

While an estimated $\alpha_1 > 0$ is consistent with herding, a special feature of our application is that this result could also be obtained without learning – that is, even if centers simply follow their own signals with no regard for the decisions of their peers. This is because a higher-ability center is more likely to accept a good organ and less likely to accept a bad organ. Thus, compared to a center of lower ability, it passes on a worse pool of organs when in first position. Our first test of herding consequently is based on an augmented specification of the preceding equation:

$$p_2 = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_1 \cdot q_2, \tag{3.2}$$

where q_2 is center 2's ability. Our focus is on the interaction term, $q_1 \cdot q_2$. When centers follow their own signals, center 2 is more responsive to the deterioration in the organ pool that results from center 1 being of relatively high ability when center 2 is itself of higher ability. That is, we expect α_3 to be positive. When centers ignore their own signals and herd in line with their predecessors, however, this effect may be reversed, resulting in a negative value for α_3. This is because lower-ability centers in second position are more likely to abandon their signals and reject with certainty, especially when following higher-ability centers.

Our second test of herding is based on the behavior of centers in third position. To implement this test, we estimate the following equation, with the probability of rejection in third position, p_3, as the dependent variable:

$$p_3 = \tilde{\alpha}_0 + \tilde{\alpha}_1 q_1 + \tilde{\alpha}_2 q_2. \tag{3.3}$$

Consider a case in which centers follow their own signals. As previously noted, a higher-ability center passes on a relatively worse pool of organs to subsequent centres. Thus, $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ are both positive. Furthermore, as long as a center's position does not vary systematically with its ability (i.e. average q_1 equals average q_2), as verified below, the coefficients $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ are also equal to each other. Once we introduce the possibility of herding (this can happen in second, but not first, position), $\tilde{\alpha}_2$ is strictly smaller than $\tilde{\alpha}_1$. This is because centers that herd (and reject with certainty, regardless of their signal) do not alter the quality of the organ pool or, consequently, the rejection probability of center 3.

We implement the tests of information cascades using administrative data obtained from NHSBT. This data covers the universe of deceased-donor livers and kidneys offered between 2006 and 2015. The data includes the sequence of centers that was offered each organ, as well as their decisions, which – with the possible exception of the final center in every sequence – must necessarily be rejections. The first step in estimating equations (3.2) and (3.3) is to construct a measure of center ability. Higher-ability centers are better than their low-ability counterparts at detecting both G and B organs. Thus, when the pool of organs is poor, higher-ability centers reject more often than do lower-ability centers. Conversely, when the pool of organs has high average quality, higher-ability centers accept more often. To determine which scenario is relevant for a given organ type – livers or kidneys – we take advantage of equation (3.1), which indicates that centers in second position reject more often when following a higher-ability center. We find that, for livers, a center's ability increases in its first-position probability of rejection, while, for kidneys, the reverse is true. Our measure of ability is thus, for livers, precisely this probability and, for kidneys, one minus the corresponding probability. Using this measure, we successfully test both predictions of the model. Based on the estimates of equation (3.1), the model additionally predicts that the interaction coefficient in equation (3.2), α_3, should be negative for livers and positive for kidneys. We also provide empirical support for this prediction.

Having established that information cascades exist, the next step is to estimate their prevalence and to quantify their efficiency consequences. Accordingly, we estimate the structural parameters of the model and then conduct counter-factual simulations. Heterogeneity in center quality is a distinguishing feature of our model; while this helps us to identify information cascades, however, it also renders the structural estimation more challenging. Our model has two ability parameters – the probability of receiving a g signal with a G organ and the probability of receiving a b signal with a B organ – that must be estimated separately for each center. Due to the large number of centers, we estimate these ability parameters outside the model. Indices of organ quality based on optimally-determined donor and organ characteristics have recently been proposed in the organ transplant literature for both livers and kidneys. These indices are constructed from retrospective NHSBT data that covers organ transplant decisions and subsequent patient outcomes over many years. Although individual centers base their decisions on many of the characteristics that are incorporated in the risk indices, their decisions are nonetheless limited by their own experiences. In general, we expect decisions taken by higher-ability centers to track more closely with the risk index. We use this intuition to construct center-specific measures of ability, to characterize the organ-center specific distribution of signals, and to compute the fraction of G organs in the population of organs.

The risk indices allow us to obtain internally-consistent estimates of all the model's parameters, with one exception: the threshold belief that an organ is good, above which centers accept it. An increase in this threshold has no bearing on the decisions of centers that follow their own signal; they will accept if they receive a g signal and reject if they receive a b signal. It does, however, increase the fraction of centers that herd and ignore their own signal, thus rejecting with certainty. An increase in the threshold belief is therefore accompanied by an increase in the rejection rate and, hence, there exists a unique threshold value at which the rejection rate predicted by the model matches the data. This is our best estimate of the threshold belief, which is estimated using the simulated method of moments (based on repeated draws of the information signals). We verify that the estimated threshold satisfies a key assumption of the model, which is that all centers follow their signals in first position. We also verify that the model matches the data well; indeed, our model's goodness of fit is substantially better than that of the alternative "no-learning" model, in which centers ignore their predecessors' decisions and always follow their signals.

To measure the prevalence of information cascades, we compute the fraction of decisions in our data for which centers are predicted to have ignored their signals (given our estimate of the threshold belief). Based on this estimate, it is common for centers to ignore their signals: this occurs 66% of the time for livers, and 54% of the time for kidneys, with an increase in these statistics at higher positions. Arguably, however, a more important question is whether such herding has substantial efficiency consequences. To answer this question, it is necessary to specify the full-information benchmark. In our model, centers know which of their predecessors followed their own signals and, consequently, which must have received a b signal when they rejected. The missing information is associated with centers who herd, as the centers that follow them learn nothing from their decisions. With full information, by contrast, signals received by all preceding centers are utilized in the decision-making process. We construct the full-information benchmark by drawing signals for those centers who are predicted by the model to herd (this is possible because the risk index provides us with more information than was historically available to individual centers). The signals we draw are also used to predict decisions in the alternative no-learning model, in which centers always follow their own signals.

With herding, there are too many rejections relative to the full-information benchmark. This is because herding centres ignore their own g signals, and this has spillover effects on the centers that follow. Conversely, with no learning, there are too many acceptances, because useful information contained in previous rejection decisions is ignored. Our estimates quantify these opposing effects: we find that over-rejection with the herding model is modest, while over-acceptance with the no-learning model is substantial. These findings motivate the final step of the analysis, in which we compare discard rates under the herding and no-learning models against the full-information benchmark. The decision to discard an organ is taken by NHSBT and thus exists outside our model. Our interest is in whether an organ that is discarded in the data (with herding) would have been accepted at an earlier position with the alternative models. As decisions are similar under herding and the full-information benchmark, we do not expect discard rates to diverge substantially. As expected, discard rates with herding are roughly 10% higher than the full-information benchmark. By contrast, discard rates under the assumption of no learning are 35% lower than the same benchmark, on account of the many false acceptances that arise when centers ignore the information that is contained in their predecessors' decisions. These results collectively indicate that herding contributes substantially to the high discard rate observed in the data, but that this increase in the discard rate is not necessarily inefficient. Centers often ignore their own signals, but their reliance on their predecessors actually protects them from accepting bad organs. The concluding section of the paper discusses the generalizability of these findings.

3.2 Institutional Setting

The shortage of suitable donor organs has always been the primary challenge faced by organ transplant programs. In response to this challenge, many countries, including the United Kingdom, have established national allocation schemes for the distribution of organs supplied by deceased donors. Organs obtained from deceased donors are classified, according to the manner of death, as either DBD (donation after brain death) or DCD (donation after cardiac death).[2] In the absence of circulation, cells switch from aerobic to anaerobic metabolism, with a resulting accumulation of toxic metabolites and lactic acid. To reduce the accompanying loss in cellular integrity, organs from deceased donors must be transplanted as quickly as possible. Organs are removed from DBD donors with the heart still beating (with the support of a ventilator) and are then immediately cooled to 4^{o}C. Although these organs do deteriorate, through a process known as cold ischaemia, the metabolic rate at 4^{o}C is less than 10% of what it is at normal temperature and, hence, deterioration is relatively slow. DCD organs, by contrast, are subject to an initial period of warm ischaemia once the heart stops beating – which, in practice, could exceed an hour – before the organ is retrieved and cooled. Deterioration is relatively rapid during this period and, moreover, organs that suffer warm ischaemia subsequently tolerate cold ischaemia very poorly. Thus, while DBD and DCD organs may not vary systematically with respect to *ex ante* quality, and while the same broad allocation protocols are utilized by the National Health Service Blood and Transplant (NHSBT) for both types of organ, DCD organs are useable for a shorter period of time before they must be discarded from the donor pool and set aside for research.

Our analysis focuses on livers and kidneys, for which donors have been matched to recipients in the United Kingdom through a national allocation scheme since the late 1990s. These two types of organ continue to dominate transplantation activity: NHSBT statistics indicate that over 80% of livers and kidneys obtained from DBD donors in 2014-2015 were transplanted, while the corresponding statistics for pancreases, hearts, and lungs were less than 35%. For DBD livers and kidneys, a Transplant Benefit Score (TBS), which puts weight on both the patient's need for a transplant and the patient's organ-specific quality of life after the transplant, is used to rank all patients listed on the National Registry when a given organ becomes available. The TBS is calculated using a fixed set of donor and recipient characteristics. When an organ becomes available, it is offered to patients in order of their priority. Each patient's hospital (transplant center) has 45 minutes to accept or decline the offer, based on the information that is made available; this process continues until the organ has been accepted or until too much time has elapsed for it to remain useable. Transplantation delays are substantially more costly for DCD organs and, hence, the proximity between donor and recipient is also a factor in drawing up the priority list. For livers, DBD (DCD) organs should be transplanted within 12 (6) hours, while the corresponding cutoffs for kidneys are 18 (12) hours. Due to greater urgency, DCD organs are offered regionally before being added to national lists. Nevertheless, a common feature of the allocation procedure for all organ types is that the order of recipients and, by extension, the order of centers, is exogenously determined. This order varies from one organ to the next; we exploit this variation in our

[2]The description of DBD and DCD organs that follows is based on Watson and Dark (2012).

empirical analysis.

Our analysis is based on the universe of deceased-donor organs that was offered to patients on the National Registry between 2006 and 2015. A new National Kidney Allocation Scheme was initiated in 2006 and a new National Liver Allocation Scheme was initiated after 2015. The analysis thus covers a period during which both livers and kidneys are allocated in a uniform manner.[3] We see in Figure 3-1 that there was a substantial increase in the number of organs offered per annum over this period – from 900 to 1800, and from 1200 to 2400, for livers and kidneys, respectively.[4] Nevertheless, the fraction of patients who either die while on the wait list or are removed from the wait list remains high. Patients are generally listed on the Registry, and can remain on the Registry, if they have more than a 50% chance of surviving for five years post-transplant. NHS Blood and Transplant (2010) indicates that 23% of liver patients listed on the National Registry in 2007-2008 had either died or had been removed from the list three years later. Among the kidney patients listed in 2004-2005, 12% had the same outcome. Nearly a decade later, NHS Blood and Transplant (2018a) and NHS Blood and Transplant (2018b) indicate that the shortage of available organs remains largely unchanged. 17% of liver patients who were listed on the National Registry in 2015-2016 had either died or were removed from the list three years later, while 12% of kidney patients listed in 2014-2015 had the same outcome two years later.

Figure 3-2 provides an explanation for the continued organ shortage, which exists despite the fact that the number of donations has increased over time. This figure reports the fraction of organs that was discarded (i.e. set aside for research), and we see that there has been a substantial increase in this fraction over the period of our analysis (2006-2015). While 22% of donated livers were discarded in 2006, 48% were discarded by the end of our analysis period. Discard rates are lower for kidneys than for livers.[5] Simultaneously, however, there has been a steep increase in the discard rate for kidneys, from 8% in 2006 to 29% by 2015. Some of this increase in the discard rate may be attributed to the worsening pool of organs over time, as DCD organs and organs that would previously have be considered unsuitable are now being retrieved (Watson and Dark (2012)). We posit a different mechanism, however,

[3]Starting in 2012, organs that have been declined by a fixed number of centers or have been subject to cold ischaemia for a sufficient amount of time are offered to all the centers that remain on the list, through a process termed the "Fast Track" allocation scheme. As discussed below, this scheme, as currently implemented, can nevertheless be modelled as a sequential process. Providing support for this argument, our model, based on sequential decision-making, matches the data at least as well in the Fast Track period as it does before that period (results available from the authors).

[4]There has been an increase in the number of live kidney donors in recent years. However, most kidneys continue to be transplanted from deceased donors in the United Kingdom.

[5]Recipients of liver transplants are usually much sicker than recipients of kidney transplants, who can remain on dialysis while they are waiting. Liver transplants are also more complex operations than kidney transplants. Not surprisingly, transplant surgeons tend to be more conservative with livers than kidneys.

Figure 3-1: Total Number of Organs Offered Over Time

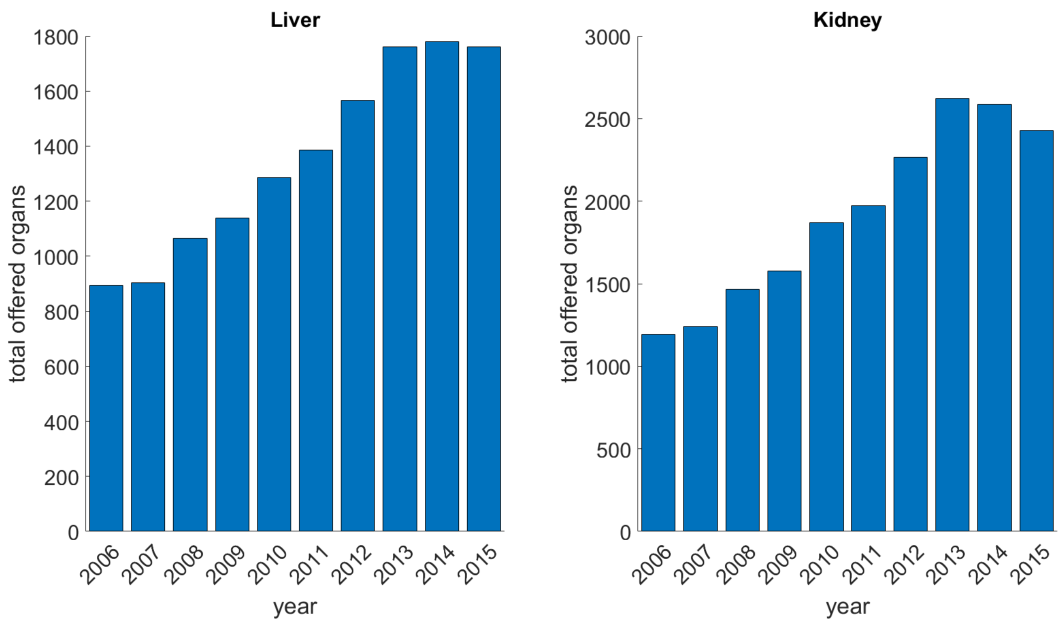
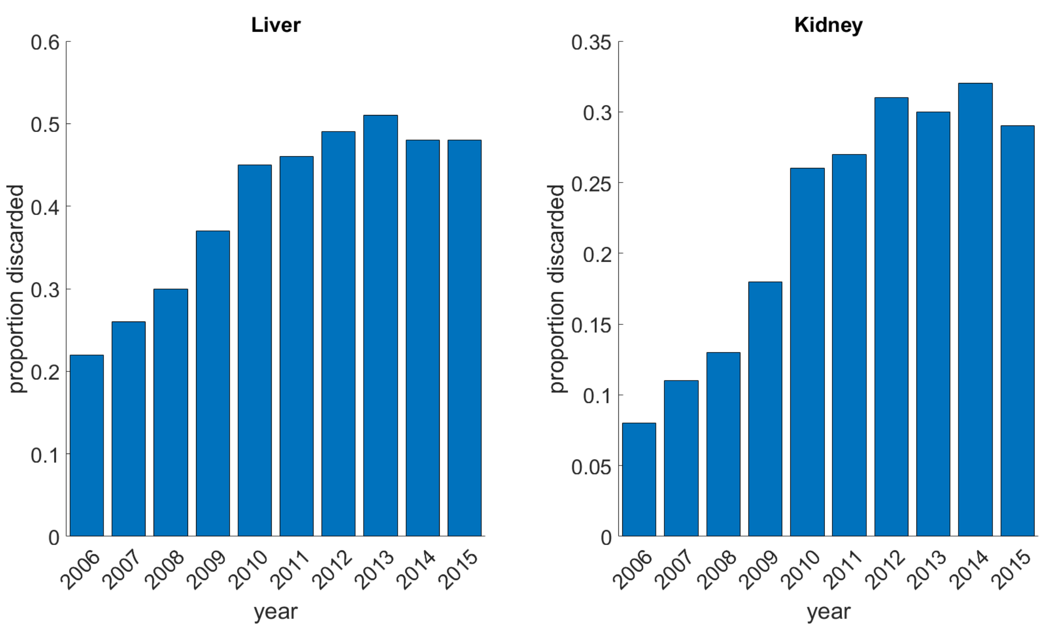


Figure 3-2: Proportion of Discarded Organs Over Time



which is based on the notion of herding in transplant decisions. The sequential nature of the decision-making process means that, once one or more centers have declined an organ, following centers may (rationally) emulate their decision to decline. The analysis that follows tests for the presence of such herding, and quantifies its contribution to the increase in the discard rate over time, as well as its consequences for welfare.

3.3 A Model of Organ Transplantation

3.3.1 Organs, Centers and Signals

There is a pool of organs of unknown quality. Organs can be either of good (G) or bad (B) quality.[6] The outcome of an organ transplant is denoted by H if the organ is good, and by L if it is bad, with $H > 0 > L$. Payoffs H and L are realized independently of the hospital (center) undertaking the transplant and the identity of the patient who receives the organ. We normalize the outcome of not transplanting an organ to 0. The prior on organ quality is denoted by π, i.e. this is the probability that a random organ from the pool is a G organ. We define the *cut-off* belief $\tilde{\pi}$ as the belief at which every hospital is indifferent between accepting or rejecting an organ – that is, $\tilde{\pi}H + (1-\tilde{\pi})L = 0$, or $\tilde{\pi} = \frac{-L}{H-L}$.

Centers individually assess organ quality before making a decision. Specifically, each center independently receives a signal $s \in \{g, b\}$, for which a g signal indicates that the organ is good, while b indicates that it is bad. Centers differ in their ability to distinguish G organs from B organs, and this heterogeneity is captured by an underlying ability (or type) parameter $q_j \in [\underline{q}, \overline{q}] \subset \mathbb{R}$ for each center j. A center's type q_j determines the probability $\gamma_j \in (0,1)$, with which it correctly identifies a G organ, and the probability $\beta_j \in (0,1)$, with which it correctly identifies a B organ. Formally, for center j, $prob(g \mid G) = \gamma_j = \gamma(q_j)$ and $prob(b \mid B) = \beta_j = \beta(q_j)$, with strictly increasing functions $\gamma, \beta : [\underline{q}, \overline{q}] \to (0,1)$.

A center with type q_j updates its beliefs about organ quality upwards upon receiving a g signal and downwards upon receiving a b signal, so that with

$$\begin{aligned}
\pi_j(G|g) &= \frac{\pi\gamma_j}{\pi\gamma_j + (1-\pi)(1-\beta_j)}, \\
\pi_j(G|b) &= \frac{\pi(1-\gamma_j)}{\pi(1-\gamma_j) + (1-\pi)\beta_j},
\end{aligned}$$

[6] Given that decisions are binary (accept or reject) and that these are the only outcomes that we consider in the analysis, there is no gain from enriching the organ type-space. For example, if we assumed that organ quality was measured on a continuum, then all values above (below) a threshold would be observationally equivalent to G (B) organs. Centers would accept (reject) if they knew that an organ was G (B).

we have

$$\pi_j(G|g) > \pi > \pi_j(G|b). \tag{3.4}$$

The inequalities in (3.4) are satisfied if

Assumption 1: *For all* j, $\beta_j \geq 1 - \gamma_j$.

We further assume that, in the absence of any other information, centers always follow their own signals, such that each center accepts the organ if it receives a g signal and declines the organ upon receipt of a b signal.[7] This requires that:

Assumption 2: *For all* j, $\pi_j(G|g) \geq \tilde{\pi} > \pi_j(G|b)$

3.3.2 Transplant Decisions

Organs are offered sequentially to centers on the basis of a predetermined (algorithmic) order. The priority list, both for a given organ and across organs, is independent of center ability. Center 1 receives a signal and, given Assumption 2, accepts after a g signal and declines after a b signal. If the organ is accepted, it is transplanted by center 1 and results in payoff H or L, depending on its quality. If it is declined, an administrator from NHSBT decides either to offer the organ to the next center or to set it aside for research. The decision to discard an organ is based on its condition or useability, which depends on the duration of ischaemia, as discussed above. NHSBT administrators monitor the condition of the organ during the offering process (this information is not available to centers), discarding it as soon as it is considered to be unsuitable for transplantation. Because the decision to discard an organ is orthogonal to its quality, this will not affect the next center's prior on organ quality and, as such, has no impact on the analysis.

Centers positioned further along in the sequence learn from the decisions of their predecessors; the particular structure of our data means that these decisions can only be rejections. Each center knows the identity of its predecessors and the order in which they made their decisions.[8] As centers are ordered in a sequence, we identify a center by its position in this sequence,

[7] As decisions are binary and there are two organ types, there is no gain to enriching the signal space. For example, suppose that signals are bounded and distributed on a continuum, with centers receiving a higher signal on average when an organ is good. Then there exists a threshold signal value above (below) which the organ is accepted (rejected) by a given center, absent any other information. All signal values above (below) the threshold are then observationally equivalent to g (b) signals.

[8] Results (available from the authors) are the same, but the analysis is more involved, when this assumption is relaxed.

such that the center at position j has ability q_j. We use an iterative process to describe centers' equilibrium beliefs and strategies. The equilibrium concept that characterizes the learning process is Perfect Bayesian Nash Equilibrium.

Center 2 knows that center 1 must have received a b signal, given Assumption 2. Its prior belief (before it receives its own signal) is therefore

$$\pi_2 = \pi_1(b) = \frac{\pi(1-\gamma_1)}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{3.5}$$

Its posterior belief when it receives a g signal is

$$\pi_2(g) = \frac{\pi_2\gamma_2}{\pi_2\gamma_2 + (1-\pi_2)(1-\beta_2)},$$

Finally, its posterior belief when it receives a b signal is

$$\pi_2(b) = \frac{\pi_2(1-\gamma_2)}{\pi_2(1-\gamma_2) + \pi_2\beta_2}.$$

Center 2 always rejects the organ if it receives a b signal, because its prior belief, π_2 (which is lower than $\tilde{\pi}$), is downgraded even further following a b signal. Center 2 could reject the organ even if it receives a g signal – which implies that it is herding – if this updating does not raise its posterior above $\tilde{\pi}$. In summary, therefore, center 2's optimal decision is only to accept the organ if it received a g signal and $\pi_2(g) \geq \tilde{\pi}$, and to decline otherwise.[9]

Furthermore, center 3 knows center 2's decision-making process and its prior belief, π_2, but does not necessarily know center 2's signal. If center 2 herds, its decision provides no information about its signal to center 3, and the latter's prior belief is therefore equal to π_2. If, on the other hand, center 2 uses its signal to make its decision ($\pi_2(g) \geq \tilde{\pi}$), center 3 infers from center 2's rejection that it must have received a b signal, and therefore has a prior belief equal to $\pi_2(b)$:

$$\pi_3 = \begin{cases} \pi_2, & \text{if } \pi_2(g) < \tilde{\pi} \\ \pi_2(b) & \text{otherwise.} \end{cases}$$

We can easily extrapolate from the discussion above to understand the manner in which center $n > 3$, given its prior belief (π_n), forms its posterior belief (either $\pi_n(g)$ or $\pi_n(b)$),

[9]With a binary signal structure, centers either follow their own signal or ignore their signal entirely, which implies an information cascade. With a continuum of signals, centers could place weight on both their own signals and those of their predecessors. Herding and cascades could co-exist in such a case but, in order to identify both phenomena simultaneously, as in Çelen and Kariv (2004), we would require information that is only available in a laboratory setting.

and then chooses optimally either to accept or to decline the organ. Center $n+1$'s prior thus depends (as just discussed for center 3) on whether its predecessor, center n, herds or not.

3.3.3 Center Heterogeneity

When agents herd, they place weight on the decisions of their predecessors in making their own decisions. For the case in which the decision is binary (accept or reject), an agent is more likely to accept (reject) when his predecessors also have accepted (rejected). Due to the particular structure of our data, in which an organ can be accepted only once, this correlation in decisions cannot be tested directly. Instead, in order to identify herding, we exploit heterogeneity in center ability. In particular, we expect a center to put more weight on a predecessor with higher ability because a rejection decision by such a center is more informative about the state of the world (i.e. that the organ is bad).[10] The first step, therefore, is to establish that such heterogeneity among centers does indeed exist.

The first measure of a center's ability that we construct is based on its probability of rejection when in first position. Given Assumption 2, the probability that center 1 (with ability q_1) rejects is equal to the probability that it receives a b signal, which is:

$$p_1(q_1) = \pi(1-\gamma_1) + (1-\pi)\beta_1. \tag{3.6}$$

In order to be able to use a center's rejection probability in position 1 as a measure of its ability we need to assume that this probability is monotonic – either increasing or decreasing – for all centers. The probability of center 1 rejecting the organ is increasing in its own ability if, for any $q_1 \in [\underline{q}, \overline{q}]$,

$$\frac{dp_1(q_1)}{dq_1} = -\pi\gamma'(q_1) + (1-\pi)\beta'(q_1) \geq 0. \tag{3.7}$$

If the preceding condition is reversed for all q_1, a center's probability of rejection in position 1 will be decreasing in its ability.

To determine whether center ability is increasing or decreasing in the probability of rejection in first position, we examine the decisions of centers in second position. To begin with, assume

[10] The implicit assumption here is that center ability is common knowledge. Past research on herding; e.g. Munshi (2004) has generated informational inefficiencies by assuming that neighbors' types are imperfectly observed. In such a situation, similar neighbors (in type-space) receive more weight, as do higher-ability predecessors in our framework.

that center 2 follows its own signal. This would be the case if it ignores its predecessors' decisions or if it does not herd; that is, its posterior belief upon receiving a g signal exceeds $\tilde{\pi}$. In this case, the probability that center 2 rejects the organ is the probability that it receives a b signal, conditional on center 1 also having received a b signal. Applying Bayes' Rule yields:

$$p_2(q_1, q_2) = \frac{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{3.8}$$

We can now compute, using Assumption 2, the manner in which center 2's rejection probability varies with center 1's ability:

$$\frac{\partial p_2(q_1, q_2)}{\partial q_1} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2 - (1-\gamma_2))}{(\pi(1-\gamma_1) + (1-\pi)\beta_1)^2} \geq 0, \tag{3.9}$$

Furthermore, an increase in center 1's ability makes it more likely that center 2 herds and rejects with certainty. This is because a higher-ability predecessor's rejection has a bigger impact on center 2's prior belief, thereby increasing the likelihood that its posterior belief will remain below $\tilde{\pi}$ even when it receives a g signal. In general, center 2 is more likely to reject when center 1 has high ability, regardless of whether centers learn from their predecessors or not.

If we observe that centers in second position are more (less) likely to reject when they follow centers with a higher probability of rejection when in first position, condition (3.7) implies that this probability is positively (negatively) associated with center ability. We will see later that the sign of this relationship is positive for livers and negative for kidneys. We therefore use the probability of rejection in first position to measure center ability for liver transplants and one minus this probability to measure center ability for kidney transplants in our tests of information cascades.

3.3.4 A Test of Cascades (based on decisions in second position)

The preceding discussion demonstrates that a center positioned at 2 is more likely to decline an organ when it follows a higher-ability center, both with and without herding. To test for information cascades, and for herding more generally, we must thus put more structure on the relationship between center 2's rejection decision and center 1's quality, q_1. In particular, we examine the way in which this relationship varies with center 2's quality, q_2. In order to implement this test, we estimate equation (3.2), which includes q_1, q_2, and $q_1 \cdot q_2$ as covariates, and focus on the interaction term.

First, assume that center 2 does not herd, such that its rejection probability is described by

(3.8). The cross-partial with respect to q_1 and q_2, which is essentially the coefficient on the interaction term, is then

$$\frac{\partial^2 p_2(q_1, q_2)}{\partial q_1 \partial q_2} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2' + \gamma_2')}{(\pi(1-\gamma_1) + (1-\pi)\beta_1)^2} > 0, \tag{3.10}$$

That is, the effect of an increase in center 1's ability on center 2's rejection probability is larger when center 2 has higher ability. In order to obtain some intuition for this result, we decompose the effect of an increase in center 1's ability on center 2's rejection probability as follows: an increase in q_1 decreases π_2 (the quality of the organ pool passed on to centre 2), which, in turn, raises centre 2's rejection probability. Therefore, $\partial p_2/\partial q_1$ is increasing in q_2 if and only if $-\frac{\partial p_2}{\partial \pi_2}$ also is, where

$$p_2 = \pi_2(1-\gamma_2) + (1-\pi_2)\beta_2,$$

and π_2 (see (3.5)) depends on q_1 but not on q_2. Taking derivatives,

$$-\frac{\partial p_2}{\partial \pi_2} = \beta_2 - (1-\gamma_2),$$

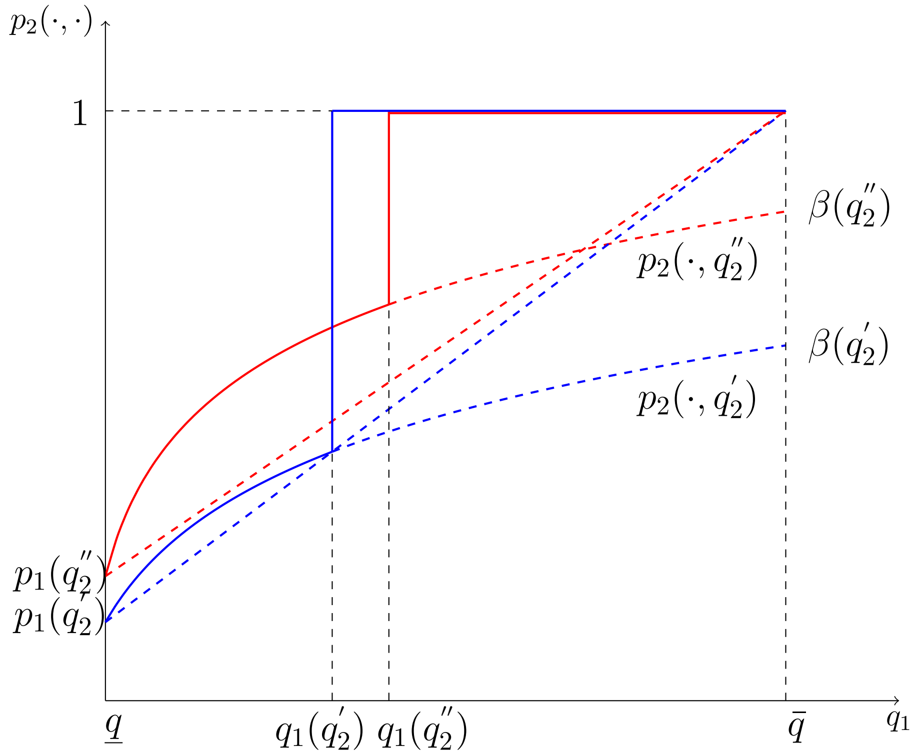
which is increasing in q_2 because both β_2 and γ_2 are. Intuitively, the rejection decision of a high-ability center is more sensitive to the quality of its organ pool than is the rejection decision of a low-ability center.

Although (3.10) shows that the cross-partial is positive in the absence of herding, this result does not necessarily hold when center 2 herds. In particular, there are now two effects: the *pure cross-partial effect* and the *herding effect*. The former, which we described above, implies that a low-ability center reacts less to an increase in its predecessor's ability than does a high-ability center, because the former is less sensitive to the quality of its organ pool. The latter effect works in the opposite direction: as the rejection by a high-ability center 1 represents worse news about underlying organ quality than does the rejection of a low-ability center, it is more likely that a weak center 2 herds and also rejects. The question of which effect dominates depends on the underlying organ pool and on the ability of the relevant centers.

This is best seen in the following two figures. In drawing these figures, we have assumed that the lowest-ability center is completely uninformed and takes a random decision, and that the highest-ability center is perfectly informed, although the results do not rely on those assumptions.[11] The figures show the rejection probabilities of two centers at position 2 as a function of center 1's ability q_1. The centers at position 2 differ in their own ability, and we assume that $q_2' < q_2''$. The curves labelled $p_2(\cdot, q_2')$ and $p_2(\cdot, q_2'')$ are the centers' respective rejection probabilities *without herding*, given by the expression in (3.8). If center 2 follows center 1 with $q_1 = \underline{q}$ it faces an organ pool with quality π and, consequently, draws signals as if it were in first position. This implies that $p_2(\underline{q}, q_2) = p_1(q_2)$, which pins down the intercept of each curve. Note that Figure 3 assumes that $p_1(q_2') < p_1(q_2'')$, which we later see applies to livers, while Figure 4 assumes that the inequality is reversed, which is relevant for kidneys. This represents the only difference between the two figures.

If center 2 follows center 1 with $q_1 = \bar{q}$, the organ is a B organ for certain, so center 2 draws signals from a B organ. This explains why the curves reach a height of $\beta(q_2)$, where $\beta(q_2') < \beta(q_2'')$. *With herding* the rejection probability of each center 2 jumps to 1 at a threshold q_1. This happens at $q_1(q_2)$, where $q_1(q_2') < q_1(q_2'')$, because a lower-ability center positioned at 2 starts herding sooner than does a high-ability one. Thus, for $q_1 < q_1(q_2)$, center 2 uses its signal and rejects according to the expression in (3.8) while, for $q_1 \geq q_1(q_2)$, center 2 herds and always rejects.

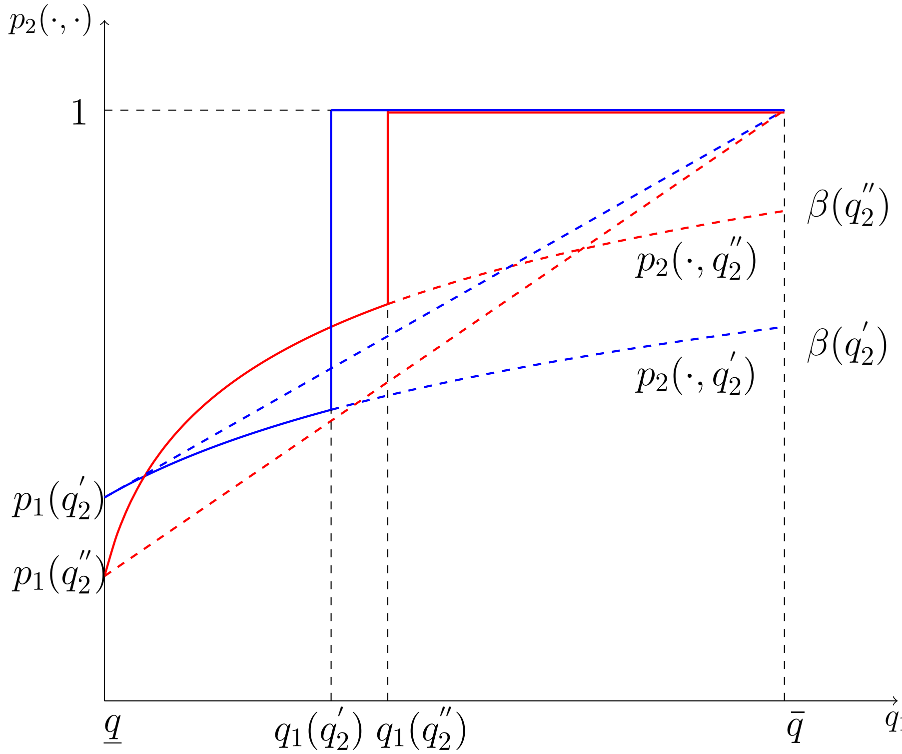
Figure 3-3: $\frac{dp_1(q)}{dq} > 0$



We now use the figures to derive the effect of the interaction term $q_1 \cdot q_2$, on center 2's rejection

[11]Formally, $1 - \gamma(\underline{q}) = \beta(\underline{q})$ and $\gamma(\overline{q}) = \beta(\overline{q}) = 1$.

Figure 3-4: $\frac{dp_1(q)}{dq} < 0$



probability with herding. Expression (3.10) allows us to obtain a measure of this effect for each pair (q_1, q_2) in the absence of herding. With herding, we cannot use the same method, because center 2's rejection probability contains a jump when center 2 starts to herd and, as such, the derivative is not well-defined at each point. Instead, we compute the"average effect" of an increase in q_1; this is the slope of the line starting at $(\underline{q}, p_1(q_2))$ and going to $(\bar{q}, p_2(\bar{q}, q_2))$. We then examine how this slope varies with q_2 (q_2' versus q_2''). Consistent with the cross-partial expression in (3.10), the slope with respect to q_1 is increasing in q_2 in the absence of herding in both figures:

$$\frac{p_2(\bar{q}, q_2') - p_1(q_2')}{\bar{q} - \underline{q}} < \frac{p_2(\bar{q}, q_2'') - p_1(q_2'')}{\bar{q} - \underline{q}}$$

When we incorporate the effect of herding, however, we see that the slope with respect to q_1 in Figure 3-3 is decreasing in q_2:

$$\frac{1 - p_1(q_2')}{\bar{q} - \underline{q}} > \frac{1 - p_1(q_2'')}{\bar{q} - \underline{q}}.$$

Thus, when center ability is *increasing* in the rejection probability in first position,($\frac{dp_1(q)}{dq} > 0$, as observed for livers), we predict that the interaction effect is reversed with herding: lower-ability centers at position 2 react more to an increase in center 1's ability because the herding effect dominates. By contrast, when center ability is *decreasing* in the rejection probability in first position, ($\frac{dp_1(q)}{dq} < 0$, as observed for kidneys), we predict that the interaction effect goes in the same direction as it does in the absence of herding; the pure cross-partial effect then dominates, such that higher-ability centers at position 2 react more to an increase in center 1's ability. In summary, therefore:

Proposition 1 *Without herding, the average cross-partial effect of an increase in both center* 1*'s and center* 2*'s ability on center 2's rejection probability is strictly positive. With herding, the average cross-partial effect is strictly positive if center ability is decreasing in rejection probability at first position and negative if center ability is increasing in rejection probability at first position.*

3.3.5 A Test of Cascades (based on decisions in third position)

Our second test of cascades is based on decisions at position 3, in particular, on the relationship between these decisions and center abilities at position 1 (q_1) and position 2 (q_2), as expressed in equation (3.3). In deriving this test we assume that center 3 does not herd. Centers that herd at third position always reject, and their decision is thus unaffected by marginal changes in q_1 and q_2.

To develop our second test of herding we investigate the effect of a marginal increase in q_1 and q_2 on center 3's rejection probability. We first consider the case without herding, in which center 3's rejection probability, denoted by p_3, is given by the probability that it receives a b signal, conditional on both center 1 and center 2 also having received b signals. Again, applying Bayes' Rule:

$$p_3(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_2)(1-\gamma_3) + (1-\pi)\beta_1\beta_2\beta_3}{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}. \tag{3.11}$$

It is easy to see that an increase in either center 1's or center 2's ability decreases the quality of the organ pool passed on to center 3, which increases the latter's rejection probability. Formally,

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_1} = \Theta\beta_2(1-\gamma_2)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)), \tag{3.12}$$

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_2} = \Theta\beta_1(1-\gamma_1)(\gamma_2'\beta_2 + \beta_2'(1-\gamma_2)), \tag{3.13}$$

with $\Theta = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)(1-\gamma_2)+(1-\pi)\beta_1\beta_2]^2}$. Both expressions clearly are strictly positive, but their exact magnitudes depend on the abilities of both centers. For $q_1 \neq q_2$, either effect could be larger than the other; for $q_1 = q_2$, however, expressions (3.12) and (3.13) are identical, such that the two effects are the same.

Now consider a situation in which center 2 herds. Our intuition suggests that, with herding – contrary to the environment in the absence of herding – a marginal increase in center 1's ability should have a bigger impact on center 3's rejection probability than would the same increase in center 2's ability. Intuitively, if center 2 herds, an increase in its ability has no impact on the organ pool passed on to center 3. By contrast, center 1 always uses its signal, such that an increase in its ability always worsens center 3's organ pool. Formally, center 3's rejection probability when center 2 herds, denoted by p_3^h, is equal to the probability that center 3 receives a b signal, conditional on only center 1 having received a b signal:

$$p_3^h(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_3) + (1-\pi)\beta_1\beta_3}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{3.14}$$

In this case, the effect of an increase in the ability of center 1, though different to the case without herding (see (3.12)), is still positive, while the effect of an increase in the ability of center 2 is zero:

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_1} = \Pi(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)), \tag{3.15}$$

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_2} = 0, \tag{3.16}$$

where $\Pi = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)+(1-\pi)\beta_1]^2}$. Our prediction, therefore, is as follows:

Proposition 2 *Assume that centers at position* 1 *and* 2 *have identical abilities. Then, without herding, the effect of an increase in center* 1*'s ability on center 3's rejection probability equals the effect of an increase in center 2's ability. With herding, the effect of an increase in center* 1*'s ability is larger than the effect of an increase in center* 2*'s ability.*

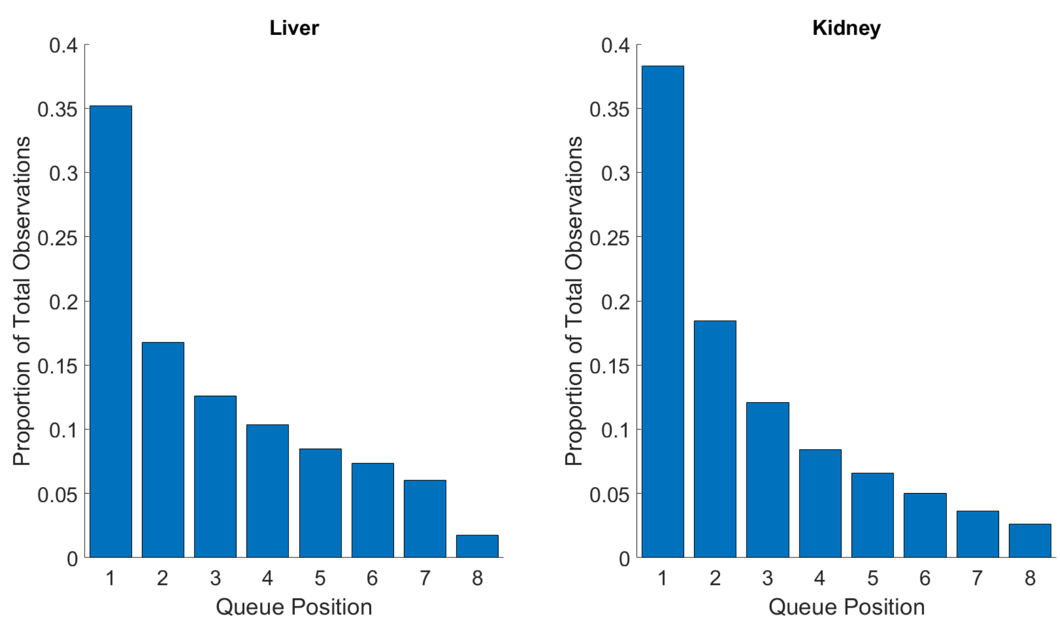
3.4 Testing the Model

3.4.1 The Data

The data that we use to test the model consists of the sequence of decisions taken by centers for each deceased-donor organ (liver or kidney) that was offered for transplantation in the 2006-2015 period. Each center that is offered an organ can either accept or reject it. If an organ is rejected, it is offered to the next center in line, unless NHSBT assesses that the condition of the organ has deteriorated to the point that it is no longer useable, in which case it is discarded (i.e. set aside for research). There are thus two possible end-points for an organ: it is accepted or it is discarded. Prior to either end-point, all decisions must necessarily be rejections.

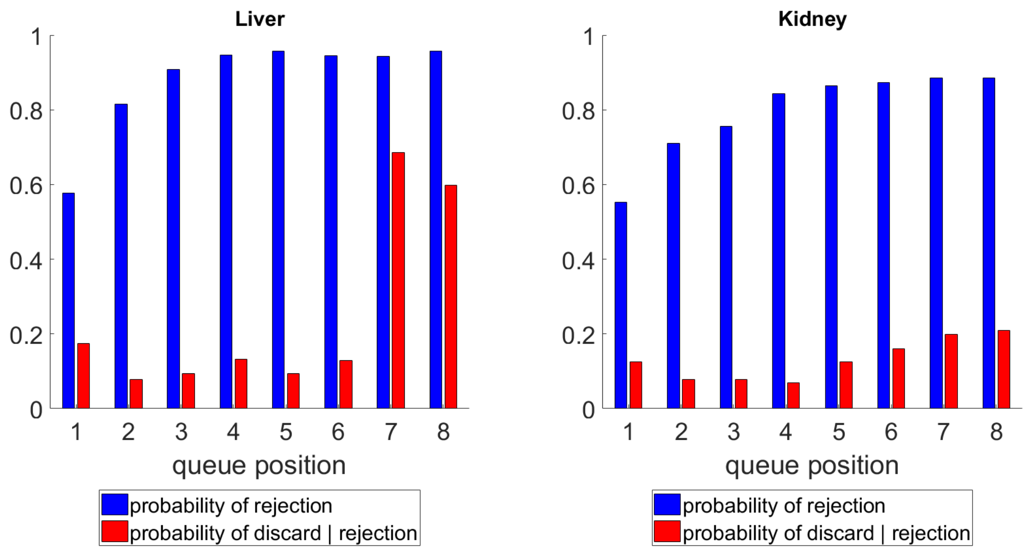
The deterioration that results in an organ being discarded can be caused by delays in retrieving the organ (warm ischaemia) or by subsequent delays in transplantation (cold ischaemia). In either case, the amount of time available for transplantation is limited; hence, sequence lengths tend to be short. With such short sequence lengths, most decisions must necessarily be concentrated at early positions: this is, in fact, precisely what we observe in Figure 3-5. For livers, 35% of observations are in first position, with a steep decline in the fraction of decisions at higher positions. For kidneys, 40% of observed decisions are in first position, followed by an even steeper decline in the fraction of decisions at higher positions. There are relatively few decisions past the eighth position for either type of organ.

Figure 3-5: Proportion of Observations by Queue Position



Another useful way to describe the data is to plot the fraction of rejections and the fraction of discarded organs (conditional on rejection) at each position. We see in Figure 3-6 that organs are discarded as early as the first position, presumably because such organs enter the offering sequence in relatively poor condition. Discard rates remain fairly stable at higher positions, except for a spike at positions 7 and 8 for livers. Rejection rates, by contrast, which start at around 60% at first position for both livers and kidneys, increase steadily by position. Notice that rejection rates are systematically lower for kidneys; this results in shorter sequence lengths for this type of organ, which explains the relatively steep decline in the fraction of organs by position that we document for kidneys in Figure 3-5.

Figure 3-6: Probability of Rejection and Discarding by Queue Position

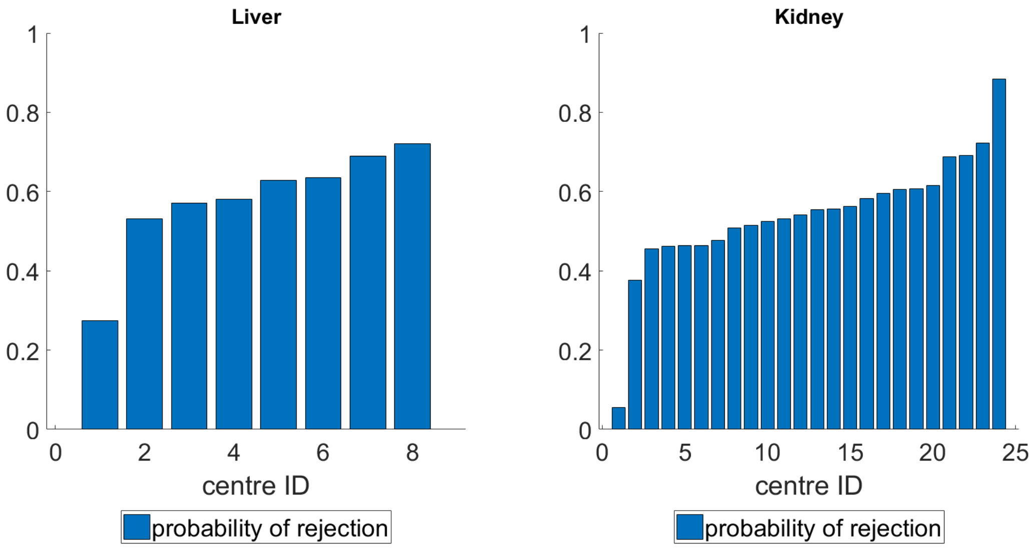


3.4.2 Center Ability

Our measure of center ability is based on the center's probability of rejection when in first position. Figure 3-7 plots this statistic for all transplantation centers in the United Kingdom, separately for livers and kidneys. There are eight liver transplant centers in total, and we see that their probability of rejection in first position ranges from 0.3 to nearly 0.8. This range is quite wide, even if we ignore one outlying center with a relatively low rejection probability. There are many more kidney transplantation centers (twenty-four), and these also display wide variation in the probability of rejection. Ignoring two centers with relatively low or high probabilities, the rejection probabilities range from around 0.4 to 0.8. Thus, there appears to be substantial variation in our measure of ability across transplantation centers.

As noted, the rejection probability in first position may be either positively or negatively

Figure 3-7: Rejections by Centers in First Position



associated with center ability, depending on whether or not inequality (3.7) is satisfied. We test this by estimating the relationship between the rejection decision in second position and the preceding center's overall probability of rejection in first position. The unconditional relationship is reported in columns 1 and 3 of Table 3.1. We see that the coefficient on the first-position center's probability of rejection is positive and significant for livers, and negative and significant for kidneys. Columns 2 and 4 report estimates for an augmented specification that includes center fixed effects (to account for the direct effect of center 2's ability), as well as a measure of organ quality that is based on an independently-constructed Risk Index, which is described in greater detail below. The implicit assumption made in using the probability of rejection in first position to measure a center's ability is that centers receive organs of equal quality on average. Suppose, instead, that centers are homogeneous with respect to ability, but that some centers receive more bad organs than others when at first position. These centers will reject more often and, because they pass on worse organs on average, their successors will also be more likely to reject. The augmented specification accounts for this possibility and, while the estimated coefficient on the first-position probability of rejection does increase in absolute magnitude for both livers and kidneys, the unconditional and conditional estimates are statistically indistinguishable.

Centers in second position are unambiguously more likely to reject when they follow a higher-ability center. The results in Table 3.1 thus imply that center ability is increasing (decreasing) in the first-position probability of rejection for livers (kidneys). We measure center ability by this probability for livers and by one minus the probability for kidneys.

Table 3.1: Measuring Center Ability

Dependent variable:	probability of rejection in second position			
Organ:	liver		kidney	
	(1)	(2)	(3)	(4)
Probability of rejection in first position	0.300***	0.440***	-0.281***	-0.374**
	(0.068)	(0.089)	(0.036)	(0.038)
Constant	0.652***	0.537***	0.889***	1.316***
	(0.041)	(0.093)	(0.022)	(0.052)
Center 2 fixed effects	No	Yes	No	Yes
Organ risk index	No	Yes	No	Yes
N	6383	5684	9257	8764

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.4.3 Testing for Cascades

Our first test of information cascades is based on the manner in which the relationship between center 2's rejection decision and center 1's ability varies with the ability of center 2. As described in Proposition 1, we expect the cross-partial effect (i.e. the effect of the interaction term $q_1 \cdot q_2$ in equation (3.2)) to be negative when each center's ability is increasing in its first-position rejection probability. Based on the preceding results, this will be the case with livers. By contrast, we expect the cross-partial effect to be positive for kidneys, as center ability is negatively associated with the first-position rejection probability.

Table 3.2 reports the estimated relationship between the rejection decision in second position and center ability in first and second position, together with the interaction term. The coefficient on q_1 applies to the case in which q_2 equals zero. In the absence of herding, a center with $q_2 = 0$ effectively chooses to accept or reject randomly, independently of q_1. With herding, however, the probability that these centers reject with certainty is increasing in q_1. We find that β_1, the coefficient on q_1, is small and imprecisely estimated for kidneys, and is much larger and significant at the one percent level for livers. β_2, the coefficient on q_2, applies to the case in which q_1 is at its lowest possible level. In this case, the first center's decision has no effect on the quality of the organ pool that is passed on and, moreover, it does not increase the likelihood that the center that follows will herd. When $q_1 = 0$, center 2 effectively behaves as if it were at first position and β_2 corresponds to the intercept in

Figures 3 and 4, which is increasing (decreasing) in q_2 for livers (kidneys). As predicted, the coefficient on q_2 is positive and significant for livers, and negative and significant for kidneys. Our test of cascades is, however, based on the interaction coefficient. As predicted by the model, the interaction coefficient is negative and significant for livers, and positive and significant for kidneys. By contrast, the interaction coefficient would be positive and significant for both livers and kidneys in the absence of herding.

Table 3.2: First Test of Information Cascades (based on decisions in second position)

Dependent variable:	probability of rejection in second-position	
Organ:	liver	kidney
	(1)	(2)
Ability of center in first position (q_1)	2.135**	-0.0245
	(0.713)	(0.097)
Ability of center in second position (q_2)	2.588***	-0.999***
	(0.688)	(0.102)
($q_1 \times q_2$)	-2.668**	1.003***
	(1.103)	(0.246)
Constant	-1.081**	0.977***
	(0.447)	(0.040)
N	6383	9257

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.05$

Our second test of information cascades is based on the decisions of centers in third position, as specified in equation (3.3). As described in Proposition 2, we expect these decisions to place more weight on first-position ability, q_1, than on second-position ability, q_2, when there is some amount of herding at position 2. The implicit assumption here, which we verify below, is that centers always follow their signals at position 1. By contrast, decisions in third position would place equal weight on q_1 and q_2 in the absence of cascades.

Table 3.3 reports the estimated relationship between the rejection decision in third position and each of q_1 and q_2. Center ability, for livers and kidneys, is measured as in Table 3.2. As predicted by the model, when cascades are present, the coefficient on q_1 is substantially larger than the coefficient on q_2; it is twice as large for livers and 50% larger for kidneys. The coefficients on q_1 and q_2 (β_1 and β_2 respectively) are imprecisely estimated for livers, and we cannot reject the hypothesis that $\beta_1 \leq \beta_2$. The corresponding coefficients for kidneys, however, are statistically significant: we can reject the hypothesis that $\beta_1 \leq \beta_2$ at the 5 per cent level.

Table 3.3: Second Test of Information Cascades (based on decisions in third position)

Dependent variable:	probability of rejection in third position	
Organ:	liver (1)	kidney (2)
First position center ability (q_1)	0.104 (0.066)	0.352*** (0.045)
Second position center ability (q_2)	0.0529 (0.064)	0.220*** (0.046)
Constant	0.820*** (0.067)	0.541*** (0.024)
F-statistic ($\beta_1 \leq \beta_2$) p-value	0.47 [0.247]	3.43 [0.032]
$\bar{q_1}$	0.60	0.40
$\bar{q_2}$	0.65	0.40
N	4819	6084

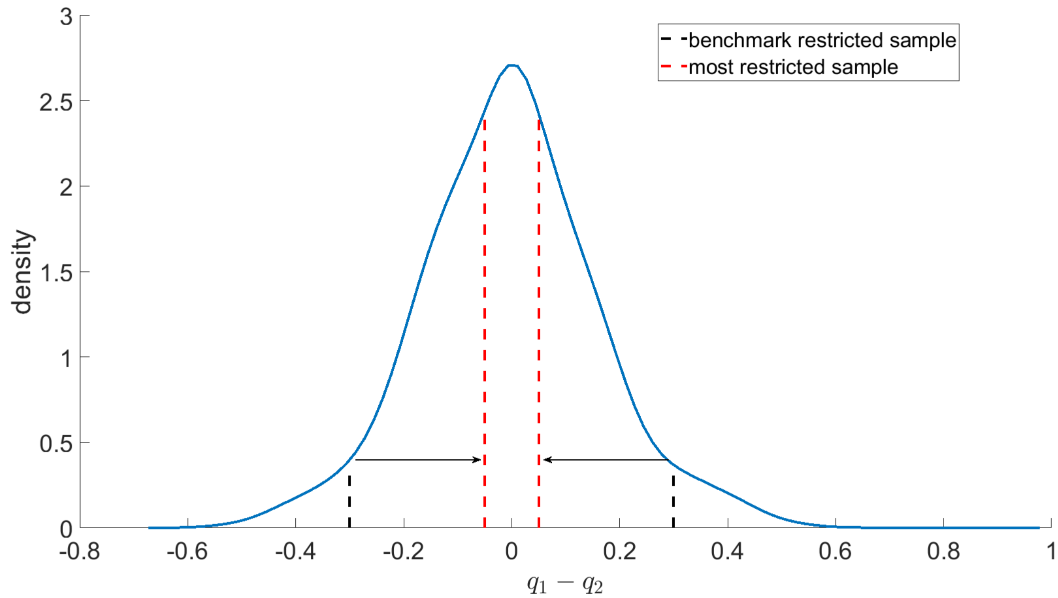
Note: β_1, β_2 are the coefficients on q_1, q_2, respectively
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The data requirements for implementing the second test of information cascades are quite stringent: (i) a large fraction of centers should herd in second position, (ii) a large fraction of centers should *not* herd in third position (as, if they did, the variation in q_1 and q_2 would have no consequence for their decisions), and (iii) substantial variation in decisions (accept versus reject) in third position is required in order for the test to have statistical power.

Based on the estimated structural model, we will see that conditions (i) and (ii) are satisfied for both livers and kidneys. An important difference between the two organ types is that, by the third position, over 90 per cent of decisions for livers are rejections, while the same is not true of kidneys. This lack of variation in liver decisions may explain the fact that the coefficients on q_1 and q_2 in column 1 are imprecisely estimated. Due to this data limitation, we focus on kidneys alone for this test of information cascades.

Proposition 2 is derived for the case in which centers at positions 1 and 2 have equal ability

Figure 3-8: Distribution of Ability Differential ($q_1 - q_2$)



Note: sample includes all kidneys that reached third position.

Table 3.4: Second Test (restricted samples)

Dependent variable:	probability of rejection in third position					
($q_1 - q_2$) range:	[-0.30,0.30] (1)	[-0.25,0.25] (2)	[-0.20,0.20] (3)	[-0.15,0.15] (4)	[-0.10,0.10] (5)	[-0.05,0.05] (6)
Center 1 ability (q_1)	0.396***	0.418***	0.427***	0.575***	0.638***	1.020**
	(0.053)	(0.057)	(0.064)	(0.090)	(0.136)	(0.468)
Center 2 ability (q_2)	0.153**	0.139**	0.132**	0.0817	0.00181	-0.437
	(0.055)	(0.059)	(0.065)	(0.091)	(0.140)	(0.470)
Constant	0.554***	0.550***	0.550***	0.506***	0.513***	0.515***
	(0.025)	(0.027)	(0.027)	(0.035)	(0.039)	(0.045)
F-statistic ($\beta_1 \leq \beta_2$)	7.27	7.97	6.81	9.25	5.94	2.44
p-value	[0.004]	[0.002]	[0.004]	[0.001]	[0.007]	[0.059]
N	5603	5399	5071	4063	3069	1665

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Alternative samples restricted to kidneys within a pre-specified ability differential ($q_1 - q_2$) range (that is, $q_1 = q_2$). Although average ability in first and second position ($\bar{q_1}$ and $\bar{q_2}$, respectively) are equal for kidneys in column 2,the more stringent requirement for testing the model is that these abilities should be equal for each organ.

Figure 3-8 describes the distribution of the ability differential ($q_1 - q_2$) for all kidneys that reached at least third position in our data (and, thus, are used in the second test of information cascades). Although the distribution is centered at zero, there is substantial variation in the ability differential statistic

Table 3.4 takes account of the preceding variation by implementing the test of cascades with an increasingly restricted sample of organs, by sequentially narrowing the ability differential range). Reassuringly, the key result in Table 3.3, which is that the coefficient on q_1 is significantly larger than that on q_2 for kidneys, remains stable as we reduce the sample in this manner. Indeed, this result is obtained even with the extremely narrow ability differential presented in column 6, for which the sample is just one-quarter of the full sample of kidneys.

3.5 Structural Estimation and Quantification

3.5.1 Estimation

Our model features two types of organ: good (G) and bad (B). Centers are heterogeneous in their ability to identify organ type; β_j is the probability that center j receives a b signal when the organ is bad and γ_j is the probability that it receives a g signal when the organ is good. The fraction of good organs in the population of organs is π and the cutoff belief (that the organ is good) above which centers accept an organ is $\tilde{\pi}$.

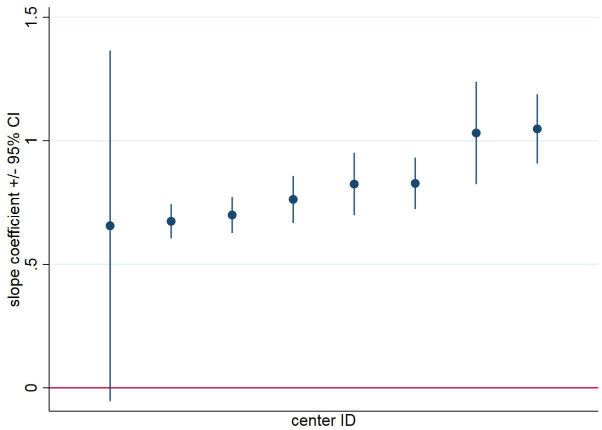
We begin by describing the construction of β_j and γ_j. The reduced-form tests of the model are based on a single ability measure, which is derived from each center's probability of rejection in first position. The structural estimation, by contrast, is based on a complete specification of the model, which utilises two ability measures. In order to construct separate measures of ability, β_j and γ_j, we take advantage of two indices of organ quality that have previously been constructed specifically for the United Kingdom: the UK KDRI (Kidney Donor Risk Index) and the UK DLI (Donor Liver Index).

Indices of liver and kidney quality were first constructed in the United States, but have recently been adapted to the U.K. population. The UK KDRI is based on U.K. National Transplant Registry data covering over 7000 recipients who received deceased-donor kidneys between January 1, 2000 and December 31, 2007 (Watson et al. (2012)). Various recipient and transplant factors were included in a model of transplant success, measured by patient survival, and the UK KDRI consists of those donor and organ characteristics that were found to be significant determinants of success (with appropriate, estimated weights on each of those characteristics). More recently, data from all liver transplants from deceased donors between January 1, 2000 and December 31, 2014 have been used to construct the UK DLI (Collett et al. (2017)). As with the UK KDRI, donor, recipient, and transplant data were used to identify factors associated with graft survival. Those donor and organ characteristics that were found to be significant determinants of transplant success, appropriately weighted, are included in the UK DLI.

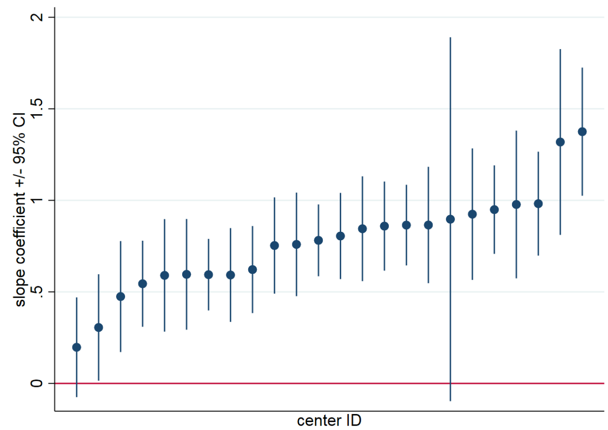
We begin by using the probit model to estimate the relationship between the probability that an organ is rejected and its risk index, by center, restricting the sample to decisions that were made when centers were at first position (and therefore, by assumption, following their signals, as verified below). The risk indices were originally developed to aid centers in their decision-making, and a proposal to incorporate the UK KDRI into the National Kidney Offering Scheme was presented at the 2018 Blood and Transplantation Congress. At the time of writing, however, neither the UK KDRI nor the UK DLI – the latter of which was developed in 2017 – are made available to transplant surgeons at the time of their decision-making. While centers may thus not have had explicit knowledge of the risk indices during the period of our analysis (2006-2015), we expect them to have put weight on many of the factors incorporated into the UK KDRI and the UK DLI. Figure 3-9 reports probit estimates of the Risk Index coefficient, with the corresponding 95% confidence interval, by center. As expected, the coefficient is positive and significant, almost without exception, for both livers and kidneys.

The risk indices are optimally constructed on the basis of outcomes generated by thousands of transplants over many years. This information is not available to transplant centers, who must base their decisions on past experiences with a limited set of outcomes. Higher-ability centers will, nonetheless, make decisions that track more closely with the risk indices. We see in Figure 3-9 that there is substantial heterogeneity in the estimated slope coefficients; we interpret larger coefficients as corresponding to higher center ability. Although we do not explicitly map the risk index onto underlying organ types (G versus B), in the limit, extremely high levels of the risk index would correspond to a B organ, while extremely low levels would correspond to a G organ. For a given center j, the predicted probability of rejection at the top of the risk index distribution thus provides an estimate of β_j; i.e. the probability of receiving a b signal with a B organ. Similarly, at the bottom of the risk index distribution, the predicted probability of rejection provides an estimate of $1-\gamma_j$; (the probability of receiving

Figure 3-9: Center-Specific Probit Slope Coefficient Estimates



(a) liver



(b) kidney

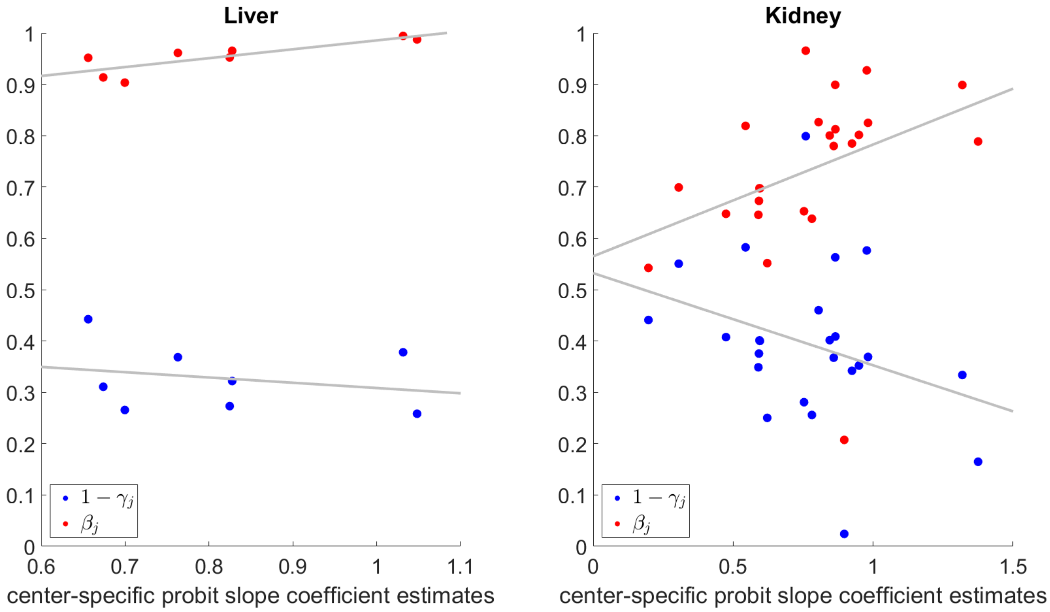
Note: estimates based on the relationship between the probability of rejection in first position and the risk index.

a b signal with a G organ). Figure 3-10 reports the estimated β_j and $1-\gamma_j$ for all centers, separately for livers and kidneys. These predictions are computed at the 95^{th} percentile and the 5^{th} percentile of the risk index distribution, respectively. Assumption 1 in the model states that $\beta_j > 1-\gamma_j$ (that is, that centers are not systematically misinformed, such that the probability of receiving a b signal when an organ is bad must exceed the corresponding probability when an organ is good). We see in Figure 3-10 that this assumption is satisfied for each center, both for livers and for kidneys.[12]

Having constructed center-specific measures of ability, β_j and γ_j, the next step is to characterize the signal-generating process. If the type of organ (G versus B) were observed, then the model would provide us with the probability that a given center j would receive a g signal or, conversely, a b signal. In practice, however, the type of organ is not observed by the econometrician. As with the construction of the center ability measures, therefore, we turn to the risk indices to characterize the distribution of signals. Recall that we have already estimated the relationship between the probability of rejection in first position and the risk index for each center. Given this relationship, the predicted organ-center specific probability of rejection for a given organ (with an associated risk index) will determine the probability that the center receives a b signal. This probability does not rely on the center's position in the decision-making sequence for that organ.

[12] Notice also that, while β_j is increasing and $(1-\gamma_j)$ is decreasing in the estimated risk index coefficient, the relationships are not monotone. This is because the β_j, γ_j measures that we construct are based on both the risk index coefficients – i.e. the slope and constant (intercept) terms of the estimated probit model. The slope alone cannot be used to construct our measures of ability.

Figure 3-10: Estimates of Center Ability



The risk indices can also be used to compute π (the fraction of G organs in the population of organs). In constructing β_j and γ_j we assumed that an organ at the top of the risk index distribution is a B organ, while an organ at the bottom of the distribution is a G organ. At intermediate levels, the probability that an organ is a G organ will be decreasing in the risk index. Based on our model, this allows us to characterize decision-making in first position as follows:

$$p_j(R_i) = \pi_i(1-\gamma_j) + (1-\pi_i)\beta_j, \tag{3.17}$$

where $p_j(R_i)$ is the probability that center j rejects organs with risk index R_i when in first position, while π_i is the probability that the underlying organ is of type G. Although, in principle, the π_i corresponding to a given R_i should be the same for all centers, noise in the estimated β_j and γ_j could generate some variation in practice. Our best estimate of π_i is thus the average across all centers; the average of this statistic across all risk indices R_i then yields an estimate of π.

The characterization of decision-making in equation (3.17) is not inconsistent with our description of decision-making in the model. In the model, each organ is of a particular type – G or B – and, while centers may not know this type, the signals that they receive (which are all that matter for decision-making) correctly reflect the underlying organ type. The econometrician does not observe the underlying organ type either, but the risk index allows us to characterize the (organ-center-specific) distribution of signals and to estimate the probability that an organ is a G organ (given R_i). This probability, averaged over all risk indices, provides an estimate of π, the fraction of G organs in the population of organs.

Having constructed measures of center ability (β_j and γ_j), characterized the signal-generating process, and estimated π, all that now remains is to estimate $\tilde{\pi}$, the cutoff belief that an organ is a G organ above which centers accept the offer. All centers follow their signal in first position in the model, which implies that their belief following a g (b) signal lies above (below) $\tilde{\pi}$. While some centers continue to follow their signals in later positions, others will begin to herd (i.e. to reject offers regardless of whether they receive a g or a b signal). This is because their beliefs always lie below $\tilde{\pi}$. As $\tilde{\pi}$ increases, the fraction of centers that herd thus increases, with an accompanying increase in the rejection rate (the decisions of centers that follow their signals remain unchanged). To estimate $\tilde{\pi}$ we thus match the overall rejection rate in the data to the rejection rate predicted by the model; there is a unique value of $\tilde{\pi}$ at which the actual and predicted rejection rates match and this will be our best estimate of the $\tilde{\pi}$ parameter.

We employ the simulated method of moments to estimate $\tilde{\pi}$. We draw signals that are organ-center-specific, as described above, and then predict decisions at each position (given the previously-estimated values of β_j, γ_j, and π). The average over multiple draws of the signals predicts the overall rejection rate for a given $\tilde{\pi}$; we then search over all $\tilde{\pi}$ to find the value at which the actual and predicted rejection rates match. The data are effectively generated by a single draw from the signal distribution. Even if the model were correctly specified and the correct value of $\tilde{\pi}$ selected by the econometrician, actual and predicted decisions evidently would not match at each organ-position. As long as a large number of centers follow their signals, however, this sampling error will wash out, such that actual and predicted rejection rates will match overall when the correct $\tilde{\pi}$ is selected.[13] Table 3.5 reports $\tilde{\pi}$ estimates, with bootstrapped standard errors, separately for livers and kidneys. As we estimate only a single parameter, we require a single moment for matching; we may therefore employ the rejection rate at any position for this purpose. The benchmark specification matches on the rejection rate at second position, as the largest fraction of centers follow their signals at this position (smoothing out the sampling error). We observe, however, that the parameter estimates remain stable when we match on additional positions (moments) up to position 5.[14]

[13] A special feature of our data is that the order of centers who would have been approached after the position at which an organ is accepted or discarded, is unavailable. Thus, if the model predicts a rejection at the final position in a sequence for which an organ was accepted in the data, then we can go no further. To preserve symmetry, therefore, whenever the model predicts an acceptance at a position at which an organ was rejected in the data, we proceed no further (and subsequent positions are not utilized for estimation).

[14] When matching on multiple positions, we compute the error in the rejection rate at each position and then take the unweighted average across all positions in order to compute the overall error. Our best estimate of $\tilde{\pi}$ is the value that minimizes the overall error.

Table 3.5: Structural Parameter Estimates

Organ:	liver				kidney			
Moment count:	one (1)	two (2)	three (3)	four (4)	one (1)	two (2)	three (3)	four (4)
$\tilde{\pi}$	0.87 (0.0038)	0.86 (0.0058)	0.86 (0.0055)	0.86 (0.005)	0.58 (0.0044)	0.52 (0.0018)	0.50 (0.0066)	0.50 (0.0015)
N	5029	3780	3109	2548	7691	5031	3508	2747

Bootstrapped standard errors in parentheses

Note: to match moments, we begin with the probability of rejection in second position, and sequentially add the corresponding probabilities in third, fourth and fifth positions

3.5.2 Validation and Goodness of Fit

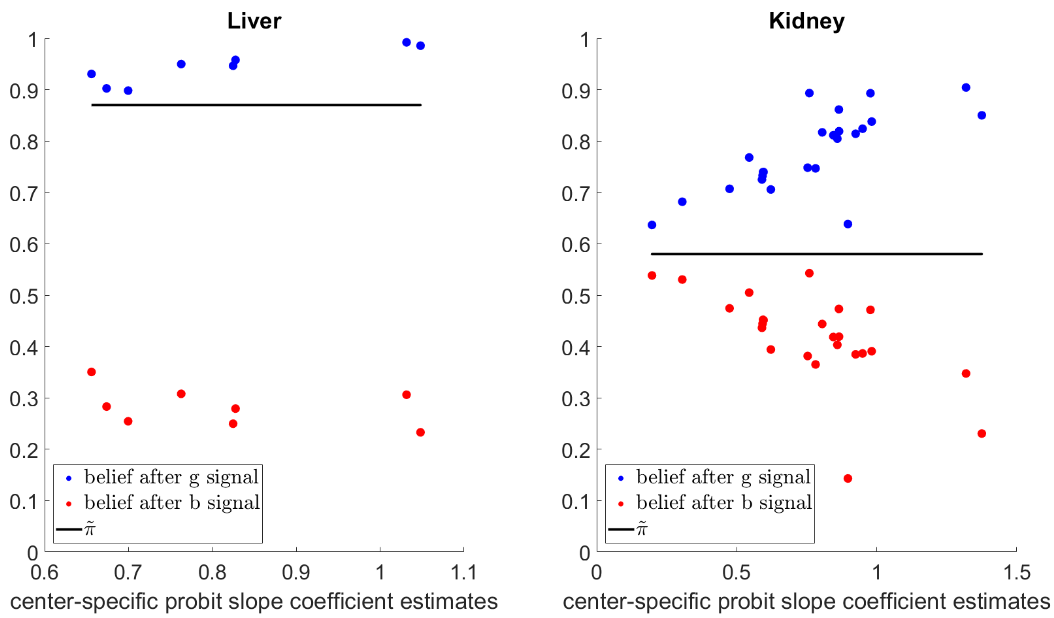
As multiple positions (moments) are available for estimation, we might think it possible simultaneously to estimate both $\tilde{\pi}$ and π. To see why this is infeasible, however, suppose that we choose a value of π, but leave the estimated β_j and γ_j unchanged. The generated signals would no longer be organ-specific; nevertheless, as π increases, the rejection rate among centers that follow their signals will decrease. To bring the overall rejection rate back in line with the data, therefore, $\tilde{\pi}$ must increase (and, with it, the fraction of centers that herd). It follows that, for every value of π, there exists a $\tilde{\pi}$ such that the actual and predicted rejection rates match. Therefore, $\tilde{\pi}$ and π cannot be estimated simultaneously, highlighting the important role played by the risk index in our analysis.[15]

As we need only estimate a single parameter, $\tilde{\pi}$, it is straightforward to establish that the parameter is identified – that is, that a unique value exists at which the model best fits the data. This value must, however, also satisfy additional restrictions implied by the model. In particular, Assumption 2 requires that each center's belief that the organ is good lie above (below) $\tilde{\pi}$ when it receives a g (b) signal in first position. Figure 3-11 verifies that this important assumption is satisfied for each center, both for livers and for kidneys.

Having estimated the model and validated its key assumptions, the next step is to assess the

[15] Cipriani and Guarino (2014) estimate a structural model of herding in financial markets in which investors draw signals from a continuous distribution. However, this additional flexibility in the signal structure does not allow them to estimate additional parameters. Once we reduce their model to match the key elements of our model, ignoring parameters associated with the presence of noise from traders who never receive signals as well as from changes in the fundamental value of the financial asset (which corresponds to whether the organ is good or bad) we see that they estimate only one parameter, which characterizes the distribution of signals. The parameter that corresponds to $\tilde{\pi}$ in our model is the price of the asset, which is observed. Thus, Cipriani and Guarino measure $\tilde{\pi}$ and estimate π, while we do the opposite.

Figure 3-11: Updated Beliefs in First Position, by Center



model's goodness of fit with the data. As we use the overall rejection rate to estimate $\tilde{\pi}$, we begin by comparing the actual and predicted rejection rates, by position, separately for livers and kidneys. The benchmark $\tilde{\pi}$ estimate is obtained by matching rejection rates at second position; all the results that follow are based on this estimate. Thus, we would expect to find a close match at second position, though not necessarily at higher positions. As a basis for comparison, we also report rejection rates from an alternative model with no learning. In this model, center quality and the signal-generating process are captured just as in the baseline specification, although decisions are now based exclusively on the signals received by each center (without regard to the decisions of preceding centers). We see in Figure 3-12 that rejection rates predicted by the herding model exceed the corresponding rates in the data, particularly at higher positions, while the alternative no-learning model systematically under-predicts rejection. The error associated with our model is not, however, substantial: overall, predicted rejection rates exceed actual rejection rates by just 3% for kidneys and 7% for livers

An alternative metric for comparing the performance of these models is the fraction of correctly-predicted decisions – that is, acceptances (rejections) in the data that are predicted to be acceptances (rejections). Based on this metric, the herding model clearly out-performs the no-learning model, both for livers and for kidneys, as shown in Figure 3-13. This result complements the reduced-form tests of the model in providing independent evidence that herding is an important component of centers' decision-making.

Figure 3-12: Goodness of Fit (Probability of Rejection): Alternative Models

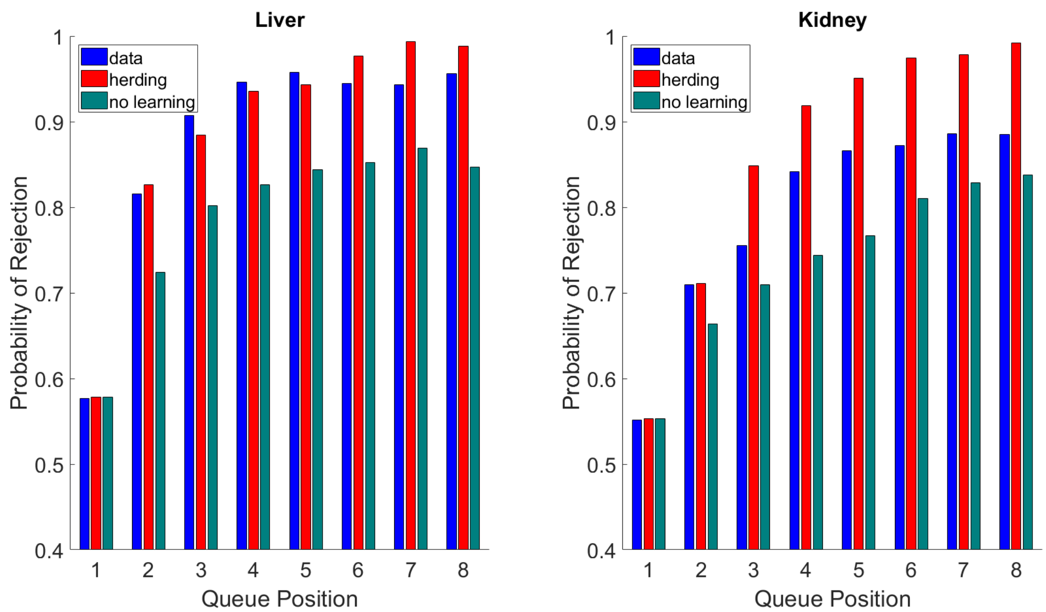
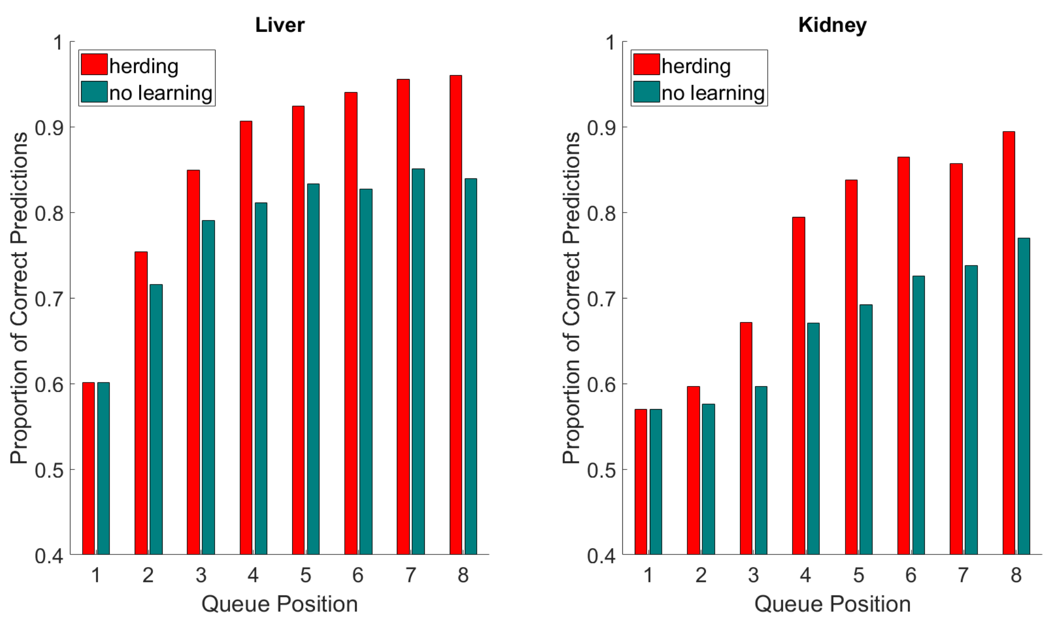


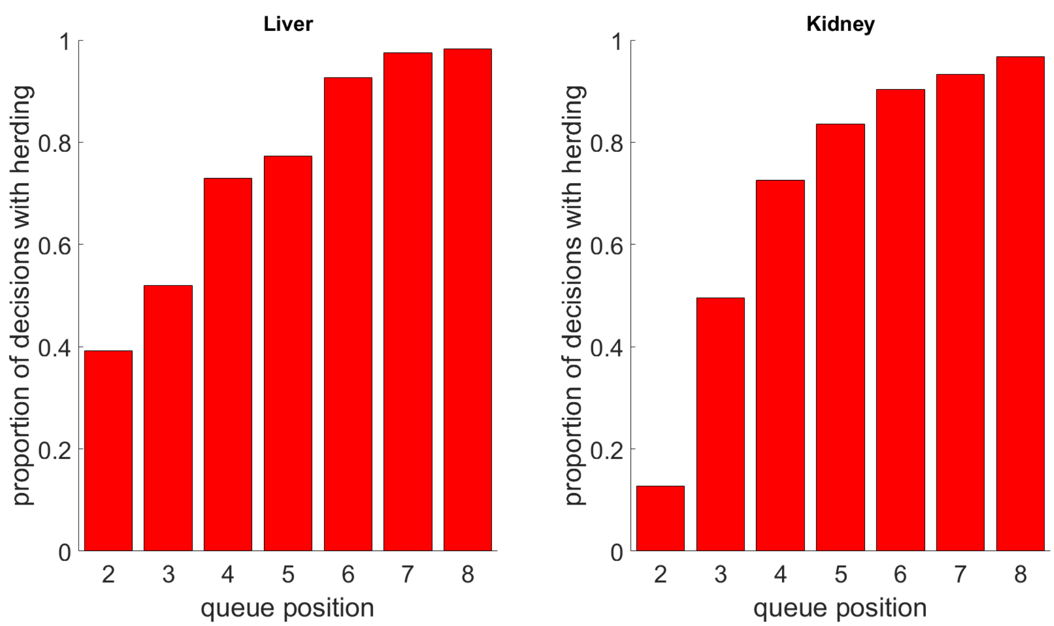
Figure 3-13: Goodness of Fit (Proportion of Correct Predictions): Alternative Models



3.5.3 Quantification and Counter-Factual Simulations

While both the reduced-form tests and the tests of the structural model against the no-learning alternative indicate that centers are herding, we would also like to quantify the prevalence of this behavior. Given the estimates of β_j, γ_j, π, $\tilde{\pi}$ and the sequence of centers associated with each organ, we can determine whether a given center at a given position is herding – that is, that its prior belief based on preceding decisions is so far below $\tilde{\pi}$ that it will reject regardless of the signal it receives. Figure 3-14 reports the prevalence of such herding, by position, for livers and kidneys. Herding is very common. For livers, about 40% of centers at second position herd. There is a steep increase in herding at higher positions and, by the sixth position, almost all centers herd. Herding is less prevalent, on average, for kidneys than for livers. Nevertheless, over 10% of kidney centers still herd at second position, with a sharp increase to 50% at third position, and over 90% herding by the sixth position.

Figure 3-14: Prevalence of Herding



Having measured the prevalence of herding in organ transplant decisions, we now turn to the efficiency consequences of such behavior. We do so by comparing organ discard rates under the status quo (with herding) with discard rates under full information (in which case centers observe, and use, the signals of all their predecessors) and under no-learning (for which centers rely exclusively on their own signals).

The standard practice for counter-factual analysis is to compare outcomes predicted by the model with simulated outcomes under alternative scenarios. A complication arises in predicting discards with our model, however, namely that the order of centers beyond the point at which an organ is accepted or discarded in the data, is unavailable. As predicted discards are thus based on observed data, and as observed sequences consist exclusively of rejections (with the possible exception of an acceptance at the end of a sequence), the model will under-predict discards and sequence lengths, even when correctly specified (by predicting acceptances in place of rejections simply by chance).[16]

Given the special structure of our data, we therefore use the actual discard rate as the benchmark for the counter-factual analysis. With full information, the signals received by centers that herd will be available to those that follow; as some of these will be g signals, subsequent rejections observed in the data may be reversed, threby lowering the discard rate. With no learning, centers follow their own signals. Here, some centers who currently herd, but draw a g signal, will reverse their decision. In either case, we are interested in the reduction in the discard rate that occurs under both full information and no-learning.

We begin by examining the decisions that underlie discard rates under alternative scenarios. In predicting the decisions associated with the full-information model, we draw signals only when required, in order to be consistent with our herding (limited-information) model. Thus, when a center accepts an organ, we assume that it must have followed its (g) signal. Similarly, when our model predicts that a center was following its signal, and it was observed to reject the organ, we assume that it received a b signal. We draw signals only when our model predicts that a center was herding; here, we use the organ's risk index and the center's quality to determine the probability that it received a b signal, in the manner described above. Decisions generated by the full-information model for each draw of the signals are compared with actual decisions at each organ-position, and the resulting discrepancy is then averaged over multiple draws to compute the fraction of false rejections and false acceptances with herding.[17] The same procedure is used to compare the no-learning outcome with that of full information, with the exception that, in the former case, centers always follow their signals.

Figure 3-15 reports the fraction of false rejections among all rejections, by position, in the

[16]Some acceptances in the data will be predicted by the model to be rejections, but this will be less frequent than reversals in the opposite direction, as rejections are more common. Moreover, we cannot infer that an organ for which an acceptance in the data is reversed by the model will necessarily be subsequently discarded.

[17]Observed decisions are determined by a single draw from an underlying signal-generating process. Thus, there will always be a discrepancy, due to sampling error, between the observed decision and the decision predicted by the model, based on repeated draws at a given organ-position (even when the model is correctly specified). This discrepancy washes out, however, when averaged over many organ-positions. By the same argument, we obtain a consistent estimate of the discrepancy between the alternative models under consideration as long as the number of organ-positions for which we draw signals is large. This will indeed be the case, due to the high prevalence of herding documented at each position in Figure 3-14.

herding and no-learning models, relative to the full-information benchmark. The herding model will generate a false rejection whenever a preceding center ignored its signal *and* when this signal would have changed the subsequent decision (from a rejection to an acceptance) if it had been observed to be g. The wide prevalence of herding that we have documented implies that such false rejections are likely to be common. With no learning, by contrast, there will be a false rejection whenever a center that rejects an organ on the basis of the b signal it received would have reversed that decision if the signals received by preceding centers had been observed and utilized for decision-making. This is because, in the absence of herding, all signals associated with prior decisions (which would necessarily be rejections) would be b signals, thus reinforcing the decision to reject. A false rejection could thus only be obtained if a sufficiently large fraction of preceding centers were herding *and* received positive signals, thereby off-setting the negative information received both from prior centers that did not herd and from the center's own b signal. False rejections are therefore likely to be relatively uncommon under the assumption of no learning. This is indeed what we observe in Figure 3-15: the proportion of false rejections is much greater for the herding model than for its no-learning counterpart. Although the fraction of false rejections in the herding model is quite large at higher positions, especially for livers, most organ sequences are short. Thus, the efficiency consequences of the false rejections may not be substantial (as verified below).

Figure 3-15: False Rejections as a Proportion of All Rejections (relative to full-information benchmark)

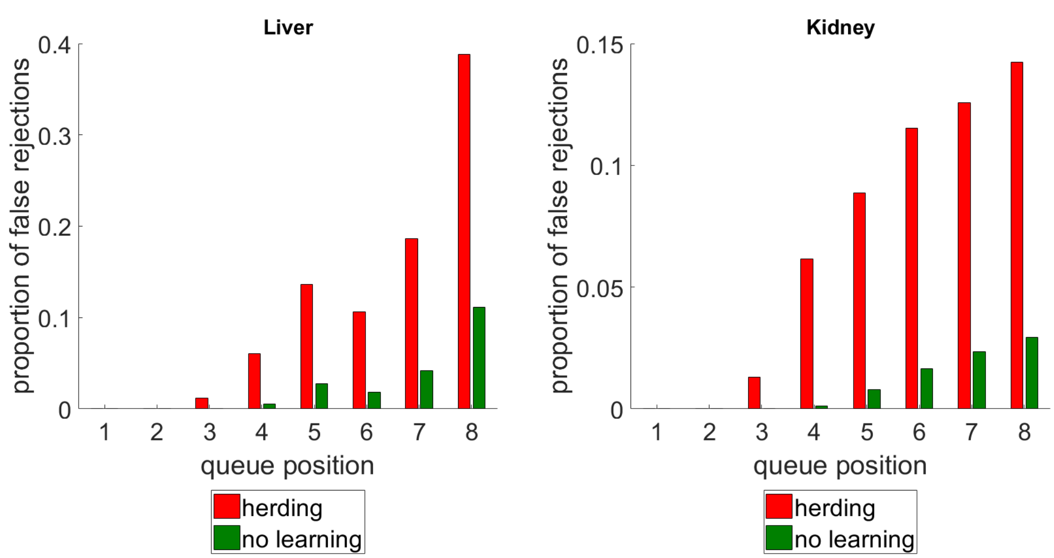
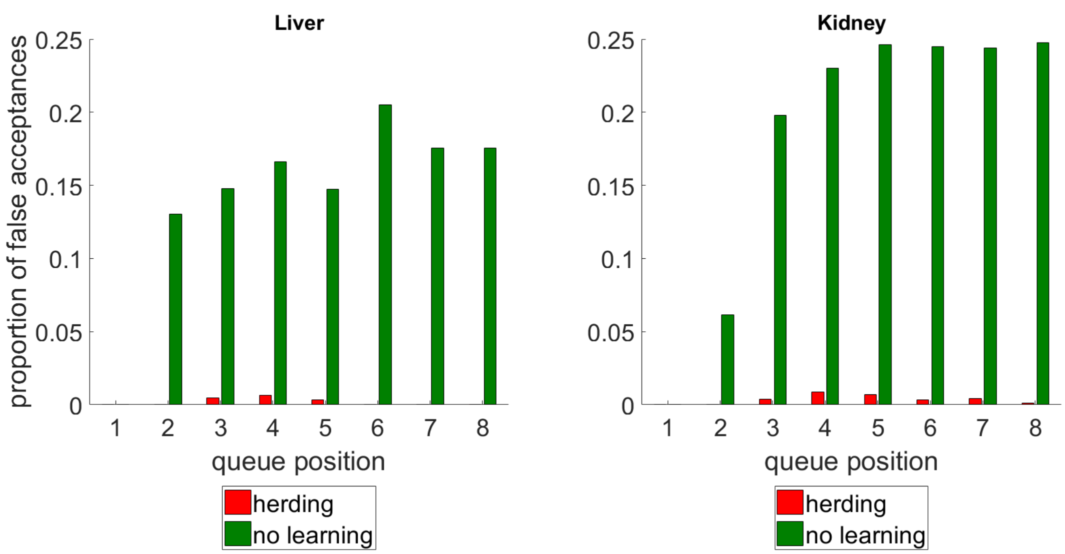


Figure 3-16 reports the fraction of false acceptances among all acceptances, comparing –

as before – the herding and no-learning models to the full-information benchmark. As an acceptance must, beginning at position 2, follow one or more rejections, accepting centers must have received a g signal in all the models we consider. The herding model would then generate a false acceptance only if a sufficiently large fraction of centers were herding *and* received b signals, with this new information shifting the accepting center's belief that the organ was a G organ from above $\tilde{\pi}$ to below $\tilde{\pi}$. This would appear to be an unlikely event. By contrast, we expect false acceptances for the no-learning model to be quite common. This is because, under this model, centers ignore the information contained in preceding decisions, which are always rejections. We see in Figure 3-16 that false acceptances are indeed rare with the herding model but are quite common with the no-learning model, for which they begin as early as position 2.

Figure 3-16: False Acceptances as a Proportion of All Acceptances (relative to full-information benchmark)



While the position-specific analysis described above paints a comprehensive picture of decision-making under different models relative to the full-information benchmark, the important consideration from a social welfare perspective is that good organs should be accepted and bad organs discarded. For instance, the case in which a center herds in line with its predecessors and rejects a good organ is costly for its patient. There is no welfare loss associated with this behavior, however, as long as patients are treated interchangeably and the organ is accepted further down the line. We therefore complete the analysis by comparing discard rates under full information and no-learning against the data.

In our data, an organ is either accepted for transplantation or discarded (set aside for research). The decision to discard an organ is taken by NHSBT and is based on its usability, which depends, in turn, on its condition (which is distinct from its quality). Thus, the discard decision is treated as exogenous in our analysis. If we observe that an organ has been discarded in the data, but one of the alternative models (no-learning or full-information) predicts that it would have been accepted at one or more earlier positions, we assume that the alternative model would have generated an acceptance for that organ. If the alternative model does not predict an acceptance at any position, the organ is assumed to have been discarded. By the same logic, if an organ was accepted in the data and an alternative model predicts that it would have been accepted at one or more positions, up to the point at which it was accepted, the organ is assumed to have been accepted with the alternative model. Finally, if an organ was accepted in the data, but an alternative model does not predict an acceptance at any position up to that point, we assume that it would have been accepted further down the line.[18]

Figure 3-17 reports the discard rate in the data (with herding), for both the full-information and no-learning models, across the range of risk indices and separately for livers and kidneys. The herding model and the full-information model differ only when there are false rejections and false acceptances in the data. As observed in Figures 3-15 and 3-16, the former is much more likely. Nevertheless, discard rates for the herding model are only slightly higher than those for the full-information model: 11% for livers and 10% for kidneys. This contrasts with discard rates for the no-learning model, which are substantially lower than those for the full-information model: 31% for livers and 35% for kidneys. This is due to the false acceptances in the no-learning model, as documented in Figure 3-16, which appear to be concentrated at higher risk indices, precisely where they are most dangerous. With organ transplantation, herding appears to protect centers from costly false acceptance decisions, without substantially raising the discard rate through false rejections.

Figure 3-18 closes the analysis by quantifying the contribution of herding to the observed increase in the discard rate over time. The model is now estimated separately for each year, providing us with distinct β_j, γ_j, π and $\tilde{\pi}$ at each point in time. Changes in the discard rate over time under the no-learning model then reflect changes in center quality (β_j, γ_j), changes in the quality of the organ pool (reflected in changes in the signal-generating process), and changes in decision-making (due to changes in $\tilde{\pi}$). We see that these changes, which are outside our model, resulted in an increase in the discard rate over the 2006-2012 period, followed by a mild decline in recent years, which tracks the overall discard

[18]We make this assumption in order to generate a conservative estimate of the discard rate for the full-information model. Nevertheless, we will see that the discard rate predicted by this model is not too far below what we observe in the data.

Figure 3-17: Organ Discard Rate

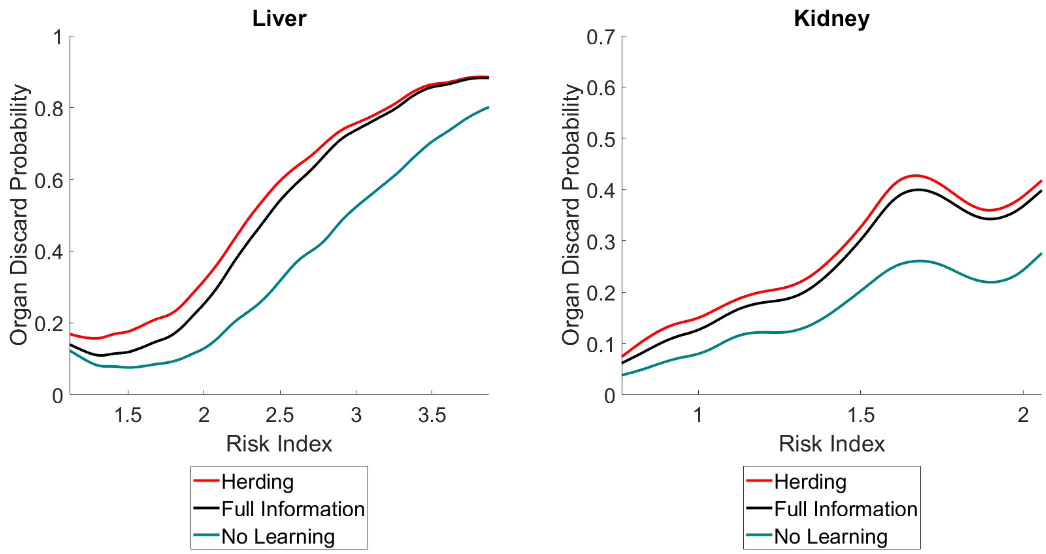
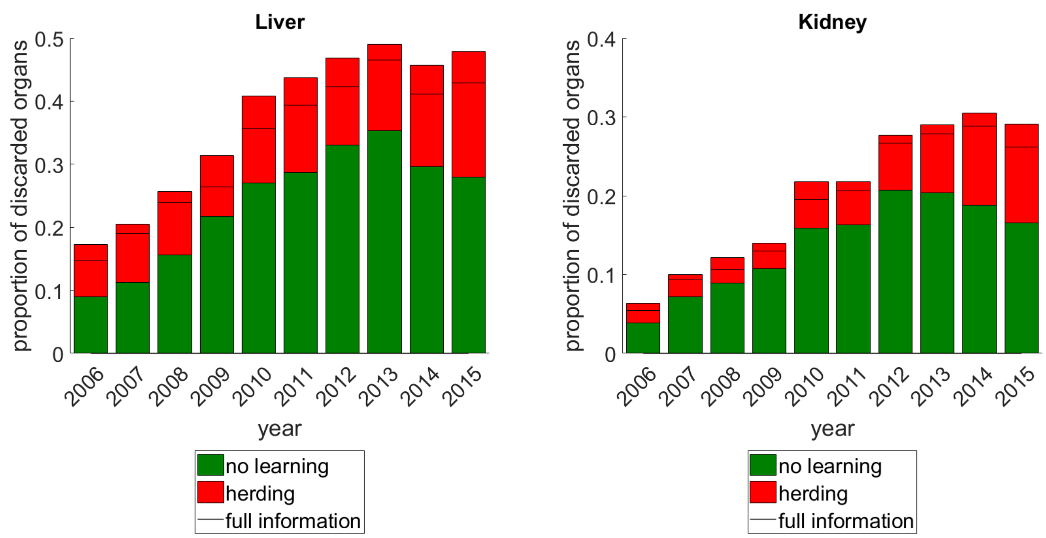


Figure 3-18: Organ Discard Rate, by Year



rate. The gap between the observed discard rate and the discard rate predicted by the no-learning model quantifies the contribution of herding to the discard rate. This contribution is substantial in each year and increases in relative importance in the most recent years. Notice, however, as in Figure 3-17, that herding does not push the discard rate very far above the full information benchmark in any year. Centers do herd behind their peers but, in this context, the herding is beneficial (relative to no learning) and is not particularly costly (relative to full information).

3.6 Conclusion

Organs obtained from deceased donors continue to be discarded at high rates in the United Kingdom, despite the long waiting list for transplantation. We examine one explanation for these high discard rates, which is information-based herding; once one or more transplant centers have rejected an organ, those that follow rationally ignore their own assessment of the organ's quality and reject as well. We find evidence that such herding is relatively common. It does not, however, push discard rates much above the efficient full-information benchmark; instead, it is protective, preventing centers from accepting organs of poor quality.

The literature on information-based herding has been active for nearly three decades and now has well-established theoretical and empirical components. Nevertheless, while the canonical models (and much of the theoretical literature that followed them) have focussed on information cascades, in which agents completely ignore their own signals and instead follow their predecessors, resulting in inefficiencies, the empirical literature has tested a different class of learning models. In particular, the empirical development literature, which has provided the most compelling evidence in support of herding to date, has studied environments in which agents always pay some attention to their own signals and in which neighbors' signals may be perfectly recovered from their decisions. The informational inefficiencies that arise in these models are distinct from the informational inefficiencies associated with information cascades (which arise because the signals received by those who herd fail to be passed on to those that follow).

Our analysis explicitly incorporates information cascades in a setting in which organs are assessed sequentially by centers with varying ability. As noted, we find that cascades are common, but that the benefits derived from predecessors' decisions outweigh the costs of herding that are emphasized in the theoretical literature. A special feature of our environment is that the order of centers is determined by the priority of their patients and, as such, is independent of their ability. In other settings for which our model is relevant, such as the selection of job candidates in the labor market, higher-ability decision-makers (firms with better reputations) are able to choose earlier on average. This increases the incidence of herding, but does not necessarily reduce efficiency (because decision-makers that move first are better informed). There is, therefore, no obvious reason why our broad empirical findings should not be replicated elsewhere.

Bibliography

Alderman, Harold, Peter F Orazem, and Elizabeth M Paterno, "School quality, school cost, and the public/private school choices of low-income households in Pakistan," *Journal of Human Resources*, 2001, pp. 304–326.

Algan, Y., A. Chéron, Hairault J.-0., and Langot F., "Wealth Effect on Labour Market Transitions," *Review of Economic Dynamics*, 2002, *6*, 156–178.

Anderson, Lisa R and Charles A Holt, "Information cascades in the laboratory," *The American Economic Review*, 1997, pp. 847–862.

Baah-Boateng, William, "Determinants of Unemployment in Ghana," *African Development Review*, 2013, *25* (4), 385–399.

_ , "The Youth Unemployment Challenge in Africa," *The Economic and Labour Relations Review*, 2016, *27* (4), 413–431.

Baily, Martin Neil, "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 1978, *10* (3), 379–402.

Banerjee, Abhijit V, "A simple model of herd behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

_ **and Andrew F Newman**, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 1993, *101* (2), 274–298.

Banerjee, Anurag, Parantap Baus, and Elisa Keller, "Cross-Country Disparities in Skill Premium and Skill Acquisition," 2016. Unpublished Working Paper.

Banerji, Arup, *Stepping up skills: For more jobs and higher productivity*, World Bank, 2010.

Barth, Erling and Claudio Lucifora, "Wage dispersion, markets and institutions: the effects of the boom in education on the wage structure," *IZA Discussion Paper*, 2006.

Bedi, Arjun S and Ashish Garg, "The effectiveness of private versus public schools: The case of Indonesia," *Journal of Development Economics*, 2000, *61* (2), 463–494.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bloemen, H. and E. Stancanelli, "Individual Wealth, Reservation Wages, and Transitions into Employment," *Journal of Labour Economics*, 2001, *19*, 400–439.

Brewer, Dominic J, Eric R Eide, and Ronald G Ehrenberg, "Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings," *Journal of Human Resources*, 1999, pp. 104–123.

Browning, M. and T.F. Crossley, "Unemployment Insurance Benefit Levels and Consumption Changes," *Journal of Public Economics*, 2001, *80* (1), 1–23.

Calónico, Sebastián and Hugo Ñopo, "Where did you go to school? Private-Public differences in schooling trajectories and their role on earnings," *Well-Being and Social Policy*, 2007, *3* (1), 25–46.

Carneiro, Pedro and James J Heckman, "The evidence on credit constraints in post-secondary schooling," *The Economic Journal*, 2002, *112* (482), 705–734.

Çelen, Boğaçhan and Shachar Kariv, "Distinguishing informational cascades from herd behavior in the laboratory," *American Economic Review*, 2004, *94* (3), 484–498.

Charlot, Olivier and Bruno Decreuse, "Over-Education for the Rich, Under-Education for the Poor - A Search-Theoretic Microfoundation," *Journal of Labour Economics*, 2010, *17*, 886–896.

Chetty, Raj, "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 2008, *116* (2), 173–234.

Chevalier, Arnaud, Gavan Conlon et al., "Does it pay to attend a prestigious university?," Technical Report, London School of Economics and Political Science, LSE Library 2003.

Cipriani, Marco and Antonio Guarino, "Estimating a structural model of herd behavior in financial markets," *American Economic Review*, 2014, *104* (1), 224–51.

Collett, David, Peter J Friend, and Christopher JE Watson, "Factors associated with short-and long-term liver graft survival in the United Kingdom: development of a UK donor liver index," *Transplantation*, 2017, *101* (4), 786–792.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Crossley, Thomas F and Hamish Low, "Borrowing Constraints, the Cost of Precautionary Saving and Unemployment Insurance," *International Tax and Public Finance*, 2011, *18* (6), 658–687.

Dale, Stacy Berg and Alan B Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics*, 2002, pp. 1491–1527.

Danforth, J., "On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search," *Studies in the Economics of Search*, 1979, *123*, 109–131.

Durlauf, Steven N, "A theory of persistent income inequality," *Journal of Economic Growth*, 1996, *1* (1), 75–93.

Eeckhout, Jan and Alireza Sepahsalari, "Unemployment Risk and the Distribution of Assets," 2015. Unpublished Working Paper.

Epple, Dennis and Richard E Romano, "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, 1998, pp. 33–62.

Falco, Paolo and Francis Teal, "Unemployment, Not Working and Jobs for the Young in Urban Africa: Evidence from Ghana," 2012. Unpublished Working Paper.

_ **and Luke Haywood**, "Entrepreneurship versus Joblessness: Explaining the Rise in Self-Employment," *Journal of Development Economics*, 2016, *118*, 245–265.

Fan, C. Simon and Oded Stark, "The Prospect of Migration, Sticky Wages and 'Educated Unemployment'," 2012. University of Tübingen Working Papers in Economics and Finance, No. 9.

Feng, Ying, David Lagakos, and James E. Rauch, "Unemployment and Development," 2018. Unpublished Working Paper.

Flinn, Christopher and Joseph Mullins, "Labor Market Search and Schooling Investment," *International Economic Review*, 2015, *56* (2), 359–398.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Fox, Louise, Lemma W Senbet, and Witness Simbanegavi, "Youth Employment in Sub-Saharan Africa: Challenges, Constraints and Opportunities," *Journal of African Economies*, 2016, *25* (AERC Supplement 1), i3–i15.

Galor, Oded and Joseph Zeira, "Income Distribution and Macroeconomics," *The Review of Economic Studies*, 1993, *60* (1), 35–52.

Ghatak, Maitreesh and Neville Nien-Huei Jiang, "A Simple Model of Inequality, Occupational Choice, and Development," *Journal of Development Economics*, 2002, *69* (1), 205–226.

Gindling, T.H. and David Newhouse, "Self-Employment in the Developing World," *World Development*, 2014, *56*, 313–331.

Girsberger, Mirjam and Romuald Meango, "The Puzzle of Educated Unemployment in West Africa," 2017. Unpublished Working Paper.

Glomm, Gerhard and Balasubrahmanian Ravikumar, "Public versus private investment in human capital: endogenous growth and income inequality," *Journal of Political Economy*, 1992, pp. 818–834.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, "Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior," *The American Economic Review*, 1995, pp. 1088–1105.

Hart, Keith, "Informal Income Opportunities and Urban Employment in Ghana," *The Journal of Modern African Studies*, 1973, *11* (1), 61–89.

Herkenhoff, Kyle, Gordon Phillips, and Ethan Cohen-Cole, "How Credit Constraints Impact Job Finding Rates, Sorting & Aggregate Output," 2017. Unpublished Working Paper.

Keane, Michael P and Kenneth I Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy*, 1997, *105* (3), 473–522.

_ and _ , "The effect of parental transfers and borrowing constraints on educational attainment," *International Economic Review*, 2001, *42* (4), 1051–1103.

Kerr, Andrew, "A Model of Comparative Advantage with Matching in the Urban Tanzanian Labour Market," 2012. CSAE Working Paper WPS/2012-21.

Kühn, Stefan, Santo Milasi, Richard Horne, and Sheena Yoon, "World Employment and Social Outlook 2016: Trends for Youth," Technical Report 2016.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, "The impact of institutional trading on stock prices," *Journal of Financial Economics*, 1992, *32* (1), 23–43.

Lee, Donghoon, "An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice," *International Economic Review*, 2005, *46* (1), 1–34.

_ **and Kenneth I. Wolpin**, "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 2006, *74* (1), 1–46.

Lentz, Rasmus and Torben Tranaes, "Job Search and Savings: Wealth Effects and Duration Dependence," *Journal of Labour Economics*, 2005, *23* (3), 467–489.

Loury, Glenn C, "Intergenerational transfers and the distribution of earnings," *Econometrica: Journal of the Econometric Society*, 1981, pp. 843–867.

Machin, Stephen, "Labour Market Inequality and Changes in the Relative Demand for Skills," in "Paper for the Royal Statistical Society Conference on Explanations of Rising Economic Inequality" 2002.

_ **and Sandra McNally**, "Tertiary education systems and labour markets," *Education and Training Policy Division, OECD*, 2007.

McFadden, Daniel, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 1989, pp. 995–1026.

Meghir, Costas, Renata Narita, and Jean-Marc Robin, "Wages and Informality in Developing Countries," *American Economic Review*, 2015, *105* (4), 1509–1546.

Mortensen, Dale, "On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search," *Handbook of Labour Economics*, 1986, *2*, 849–919.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Nguimkeu, Pierre, "A Structural Econometric Analysis of the Informal Sector Heterogeneity," *Journal of Development Economics*, 2014, *107*, 175–191.

NHS Blood and Transplant, "Transplant Activity in the UK Activity Report 2009/2010," Technical Report 2010.

_ , "Annual Report on Kidney Transplantation 2017/18," Technical Report 2018.

_ , "Annual Report on Liver Transplantation 2017/18," Technical Report 2018.

Poschke, Markus, "Who Becomes an Entrepreneur? Labour Market Prospects and Occupational Choice," *Journal of Economic Dynamics and Control*, 2013, *37*, 693–710.

Praphamontripong, Prachayani, "Private Higher Education across Asia: Expanding Access, Searching for Quality," Technical Report, Asian Development Bank 2012.

Ra, Sungsup, Brian Chin, and Amy Liu, "Challenges and Opportunities for Skills Development in Asia: Changing Supply, Demand, and Mismatches," Technical Report 2015.

Rendon, Silvio, "Job Search and Asset Accumulation under Borrowing Constraints," *International Economic Review*, 2006, *47* (1), 233–263.

Roth, Alvin E, Tayfun Sönmez, and M Utku Ünver, "Kidney exchange," *The Quarterly Journal of Economics*, 2004, *119* (2), 457–488.

_ , _ , and _ , "Pairwise kidney exchange," *Journal of Economic Theory*, 2005, *125* (2), 151–188.

_ , _ , and _ , "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *American Economic Review*, 2007, *97* (3), 828–851.

Serneels, Pieter, "The Nature of Unemployment among Young Men in Urban Ethiopia," *Review of Development Economics*, 2007, *11* (1), 170–186.

Shimer, R. and I. Werning, "Reservation Wages and Unemployment Insurance," *The Quarterly Journal of Economics*, 2007, *122* (3), 1145–1185.

_ and _ , "Liquidity and Insurance for the Unemployed," *American Economic Review*, 2008, *98* (5), 1922–1942.

Spence, Michael, "Job market signaling," *The Quarterly Journal of Economics*, 1973, pp. 355–374.

Sullivan, Paul, "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search," *International Economic Review*, 2010, *51* (1), 289–317.

Udall, Alan T. and Stuart Sinclair, "The 'Luxury Unemployment' Hypothesis: A Review of Recent Evidence," *World Development*, 1982, *10* (1), 49–62.

UNESCO Institute for Statistics, "Global Education Monitoring Report," Technical Report 2016.

Varghese, NV, *Growth and expansion of private higher education in Africa*, UNESCO, International Institute for Educational Planning, 2006.

Watson, Christopher JE, Rachel J Johnson, Rhiannon Birch, Dave Collett, and J Andrew Bradley, "A simplified donor risk index for predicting outcome after deceased donor kidney transplantation," *Transplantation*, 2012, *93* (3), 314–318.

Watson, CJE and JH Dark, "Organ transplantation: historical perspective and current practice," *British Journal of Anaesthesia*, 2012, *108* (suppl_1), i29–i42.

Wermers, Russ, "Mutual fund herding and the impact on stock prices," *Journal of Finance*, 1999, *54* (2), 581–622.

Wolpin, Kenneth I., "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 1987, *55* (4), 801–817.

World Bank, "STEP Skills Measurement Surveys," 2012/2014. data retrieved from World Bank Central Microdata Catalog, http://microdata.worldbank.org/index.php/catalog/step; last accessed: 29 March 2016.

Yang, Guanyi, "Endogenous Skills and Labor Income Inequality," 2017. Unpublished Working Paper.

Yusuf, Shahid, William Saint, and Kaoru Nabeshima, *Accelerating catch-up: Tertiary education for growth in Sub-Saharan Africa*, World Bank, 2009.

Appendix A

Chapter 1 Appendix

A.1 Deriving Optimal Bequests

The budget constraint for period 1 is given as follows, where d_k represents a dummy variable for action k and we require $s_1 \geq 0$ due to the absence of credit markets.

$$s_1 = y_i + d_1 w_1 - d_3 x_3 \tag{A.1}$$

The budget constraint for period 2 is given by:

$$c_2 + y_2 = (1 + r)s_1 + d_1 w_1 + d_2 w_2 + d_3 w_3 - \tau \tag{A.2}$$

where y_2 denotes the bequest and τ is a lump-sum tax levied on all workers by the government in order to cover the cost of public university education. It is calculated as follows:

$$\tau = \frac{cn_2}{N} \tag{A.3}$$

where c is the university's marginal cost and n_2 is the sum of public graduates in the current generation.

The individual's problem is thus to choose the size of consumption and bequest that maximise lifetime utility subject to the lifetime budget constraint:

$$\begin{aligned} \max_{y_2, c_2} \quad & u(c_2) + \beta u(y_2) \\ \text{subject to} \quad & c_2 + y_2 = (1 + r)(y_1 + d_1 w_1 - d_3 x_3) + d_1 w_1 + d_2 w_2 + d_3 w_3 - \tau \end{aligned} \tag{A.4}$$

Note that the RHS of the lifetime budget constraint is simply lifetime income, which I denote I. Separability allows me to consider the agent's choice of education first. Then, taking this as given, I can solve for the optimal levels of consumption and bequest. Specifying logarithmic utility, I write the Lagrangian as follows:

$$\mathcal{L} = lnc_2 + \beta lny_2 - \lambda[c_2 + y_2 - (1+r)(y_1 + d_1w_1 - d_3x_3) + d_1w_1 + d_2w_2 + d_3w_3 - \tau] \quad \text{(A.5)}$$

First-order conditions are given by:

$$\frac{\partial \mathcal{L}}{\partial c_2} = \frac{1}{c_2} - \lambda = 0 \quad \text{(A.6)}$$

$$\frac{\partial \mathcal{L}}{\partial y_2} = \frac{\beta}{y_2} - \lambda = 0 \quad \text{(A.7)}$$

Solving the FOCs simultaneously yields the following expressions for optimal consumption and bequest:

$$y_2^* = \frac{\beta}{1+\beta} I \quad \text{(A.8)}$$

$$c_2^* = \frac{1}{1+\beta} I \quad \text{(A.9)}$$

Thus, I can write optimal bequests by choice of education as follows: For non-graduates:

$$y_2^* = \frac{\beta}{1+\beta}(1+r)(y_1 + w_1) + w_1 - \tau \quad \text{(A.10)}$$

For public university graduates:

$$y_2^* = \frac{\beta}{1+\beta}(1+r)y_1 + w_2 - \tau \quad \text{(A.11)}$$

For private university graduates:

$$y_2^* = \frac{\beta}{1+\beta}(1+r)(y_1 - x_3) + w_3 - \tau \quad \text{(A.12)}$$

A.2 Deriving Workers' Education Choice Rules

A worker with ability h_i and wealth y_i chooses d_k^* to maximise discounted lifetime earnings. Thus, the optimal choice is given by:

$$d_k^* \begin{cases} = d_3 & \text{if } w_3 > w_2 + x_3/\beta \text{ and } w_3 > ((1+\beta)w_1 + x_3)/\beta \text{ and } d_3 \in D^i \\ = d_2 & \text{if } w_2 > w_3 - x_3/\beta \text{ and } w_2 > (1+\beta)w_1/\beta \text{ and } d_2 \in D^i \\ = d_1 & \text{if } w_1 > (\beta/(1+\beta))w_2 \text{ and } w_1 > (\beta/(1+\beta))w_3 - x_3/(1+\beta) \end{cases} \tag{A.13}$$

where β is the discount factor between periods. Substituting in for w_2 and w_3, gives the final decision rules for each of the four sub-groups of workers defined by variation in feasible choice sets.

A.3 Identifying Reasonable Equilibria

As described in section 2.5.2 of the main paper, equilibria in this framework are determined by the relative ordering of workers' education choice cut-offs. There are 120 permutations; however, I restrict the set of equilibria to a subset of "reasonable" permutations by imposing the public university capacity constraint in the main text (derived in Appendix A.4). This reduces the number of cut-off permutations to the following three cases:

- **Case 1:** $h^* > max\left\{\frac{x_3}{\beta(\gamma_3-\gamma_2)}, \frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$
- **Case 2:** $1 \geq \frac{x_3}{\beta(\gamma_3-\gamma_2)} > h^* > max\left\{\frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$
- **Case 3:** $\frac{x_3}{\beta(\gamma_3-\gamma_2)} > 1 > h^* > max\left\{\frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$

A.4 Deriving the Public University's Capacity Constraint for Baseline Model

Cases 1 and 3 represent the lowest and highest equilibrium demand for public university, and thus the lowest and highest equilibrium values of h^*, respectively. Thus, we can use these two cases to determine appropriate bounds for $\bar{U}$. In order to limit the set of equilibria to the sub-set in which the public university's entry requirement is "binding", we require that the following expression holds for every equilibrium value of h^* and x_3^*:

- $h^* \geq \frac{(1+\beta)w_1+x_3}{\beta\gamma_3}$
- $h^* \geq \frac{(1+\beta)w_1}{\beta\gamma_2}$

In Case 1, this amounts to the following set of conditions:

$$\frac{(1-\theta-\bar{U})}{1-\theta} \geq \frac{(1+\beta)}{\beta\gamma_3} + \frac{\beta\gamma_3-(1+\beta)+c}{2\beta\gamma_3} \tag{A.14}$$

$$\frac{(1-\theta-\bar{U})}{1-\theta} \geq \frac{(1+\beta)}{\beta\gamma_3} + \frac{\beta(\gamma_3-\gamma_2)(1-\theta-\bar{U})}{\beta\gamma_3(1-\theta)} \tag{A.15}$$

$$\frac{(1-\theta-\bar{U})}{1-\theta} \geq \frac{(1+\beta)}{\beta\gamma_2} \tag{A.16}$$

$$\frac{(1-\theta-\bar{U})}{1-\theta} \geq \frac{(1+\beta+x_3)}{\beta\gamma_3} + \frac{(1-\theta-\bar{U})}{1-\theta} - \frac{(1+\beta)}{\beta\gamma_2} \tag{A.17}$$

In Case 3, this amounts to the following set of conditions:

$$(1-\bar{U}) \geq \frac{(1+\beta)}{\beta\gamma_3} + \frac{\beta\gamma_3(1-\bar{U})-(1+\beta)+c}{2\beta\gamma_3} \tag{A.18}$$

$$(1-\bar{U}) \geq \frac{(1+\beta)}{\beta\gamma_3} + \frac{\beta(\gamma_3-\gamma_2)}{\beta\gamma_3} \tag{A.19}$$

$$(1-\bar{U}) \geq \frac{(1+\beta)}{\beta\gamma_2} \tag{A.20}$$

$$(1-\bar{U}) \geq \frac{(1+\beta)}{\beta\gamma_3} + \beta\gamma_3(1-\bar{U})-(1+\beta) \tag{A.21}$$

It is now possible to eliminate some of these conditions. First, it is easy to see that the constraint on $\bar{U}$ implied by condition A.16 is stricter than that implied by condition A.20, so the latter may be ignored. Secondly, the value of x_3 in condition A.14 is only an equilibrium if it is smaller than the values of x_3 in both condition A.15 and condition A.17. Thus, these latter conditions imply stricter constraints on $\bar{U}$, so condition A.14 may be ignored. Thirdly, the value of x_3 in condition A.18 is an equilibrium only if it is smaller than the value of x_3 in condition A.21. Thus, condition A.21 is a stricter constraint on $\bar{U}$, so condition A.18 may also be ignored. Fourthly, the value of x_3 in condition A.18 is an equilibrium only if it is larger than the value of x_3 in condition A.19, so condition A.18 is a stricter constraint than condition A.19. Thus, condition A.19 may also be ignored. Fifthly, conditions A.15 and A.16 may be re-arranged to give exactly the same constraint on $\bar{U}$, so condition A.16 may also be dropped. Sixthly, condition A.17 may be rearranged to give $0 \geq 0$, which is always true, so this condition may also be ignored. We are now left with conditions A.15 and A.21, which imply the following capacity constraint:

$$\bar{U} \leq min\left[\frac{[\beta\gamma_3-(1+\beta)]}{\beta\gamma_3}, \frac{(1-\theta)[\beta\gamma_2-(1+\beta)]}{\beta\gamma_2}\right] \tag{A.22}$$

Now observe that, if $\beta\gamma_3 > (1+\beta)$, condition A.15 implies a stricter capacity constraint than does condition A.21. This simply states that a worker with the highest endowment of ability ($h_i = 1$) must prefer being a private university graduate to being a non-graduates when $x_3 = 0$. This must be true, else no-one would ever attend private university. Thus, condition A.21 may be ignored, and we obtain the final capacity constraint:

$$\bar{U} \leq \frac{(1-\theta)[\beta\gamma_2 - (1+\beta)]}{\beta\gamma_2} \tag{A.23}$$

It is intuitive that the strictest constraint on public university capacity is one implied by Case 1 rather than Case 3 as, when h^* is lower, it is more difficult for the two preference cut-offs to lie below h^*, as specified above.

A.5 Deriving Public University Admission Cut-Offs

Case 1: $h^* > max\left\{\frac{x_3}{\beta(\gamma_3-\gamma_2)}, \frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$
Now all rich, high-ability workers attend private university, so demand for public university comes exclusively from workers with income y^L. Thus, h^* is set to satisfy:

$$\bar{U} = (1-\theta)(1-h^*) \tag{A.24}$$

Rearranging gives the equilibrium value of h^*:

$$h^* = \frac{1-\theta-\bar{U}}{1-\theta} \tag{A.25}$$

Case 2: $1 \geq \frac{x_3}{\beta(\gamma_3-\gamma_2)} > h^* > max\left\{\frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$
Now some fraction of rich workers both above and below h^* attend private university, so h^* is set to satisfy:

$$\bar{U} = (1-\theta)(1-h^*) + \theta\left(\frac{x_3}{\beta(\gamma_3-\gamma_2)} - h^*\right) \tag{A.26}$$

Applying the uniform and solving simultaneously with x_3^* gives:

$$h^* = 1-\theta-\bar{U} + \frac{\theta x_3^*}{\beta(\gamma_3-\gamma_2)} \tag{A.27}$$

where we can substitute the appropriate value of x_3^* as derived for Case 2 in Appendix A.6, below.

Case 3: $\frac{x_3}{\beta(\gamma_3-\gamma_2)} > 1 > h^* > max\left\{\frac{(1+\beta)w_1+x_3}{\beta\gamma_3}, \frac{(1+\beta)w_1}{\beta\gamma_2}\right\}$

If the private university chooses an equilibrium tuition fee that is sufficiently high to shut rich workers with $h_i \geq h^*$ out of the market, h^* is set to satisfy:

$$\bar{U} = (1 - h^*) \tag{A.28}$$

Rearranging gives the equilibrium value of h^*:

$$h^* = 1 - \bar{U} \tag{A.29}$$

A.6 Deriving the Private University's Optimal Tuition Fee for Baseline Model

As the relative ordering of the cut-offs determines the demand function for private universities, I solve for the equilibrium in the following way. First, I postulate an equilibrium under each of the three possible orderings. I then derive the optimal tuition fee and profit for the profit-maximising private university in that case. Finally, I check whether it meets the conditions for being an equilibrium (i.e. that no-one has an incentive to deviate, and that the relative ordering of the cut-offs is satisfied).

- Case 1:

  The total group of workers demanding private university education at a given price x_3 is:

$$\eta = \theta \left(1 - \frac{(1+\beta)w_1 + x_3}{\beta\gamma_3}\right) \tag{A.30}$$

  The firm's profit-maximisation problem is given by:

$$\begin{aligned}
&\max_{x_3} \qquad \pi = \theta \left(1 - \frac{(1+\beta)w_1 + x_3}{\beta\gamma_3}\right)(x_3 - c) \\
&\text{subject to} \quad \frac{(1-\theta-\bar{U})}{(1-\theta)}\beta(\gamma_3 - \gamma_2) \geq x_3 \qquad \text{and} \qquad \frac{\beta\gamma_3(1-\theta-\bar{U})}{(1-\theta)} - (1+\beta) \geq x_3
\end{aligned} \tag{A.31}$$

  This yields an interior solution and two corner solutions, and the final expression for x_3^* is given as:

$$x_3^* = min\left[\frac{\beta\gamma_3 - (1+\beta)w_1 + c}{2}, \frac{\beta(\gamma_3-\gamma_2)(1-\theta-\bar{U})}{(1-\theta)}, \frac{\beta\gamma_3(1-\theta-\bar{U})}{(1-\theta)} - (1+\beta)\right] \tag{A.32}$$

The equilibrium exists iff the following condition is satisfied:

$$\frac{1-\theta-\bar{U}}{1-\theta} \geq min\left[\frac{\beta\gamma_3-(1+\beta)w_1+c}{2\beta(\gamma_3-\gamma_2)}, \frac{\beta\gamma_3(1-\theta-\bar{U})-(1-\theta)(1+\beta)}{(1-\theta)\beta(\gamma_3-\gamma_2)}\right] \quad \text{(A.33)}$$

- Case 2: The demand function is now:

$$\eta = \theta\left(h^* + 1 - \frac{(1+\beta)w_1+x_3}{\beta\gamma_3} - \frac{x_3}{\beta(\gamma_3-\gamma_2)}\right) \quad \text{(A.34)}$$

Substituting for h^* gives:

$$\eta = \theta\left(1 + 1 - \theta - \bar{U} + \frac{\theta x_3}{\beta(\gamma_3-\gamma_2)} - \frac{(1+\beta)w_1+x_3}{\beta\gamma_3} - \frac{x_3}{\beta(\gamma_3-\gamma_2)}\right) \quad \text{(A.35)}$$

The profit-maximisation problem is then given by:

$$\begin{aligned}
&\max_{x_3} && \pi = \theta\left(2-\theta-\bar{U}+\frac{\theta x_3}{\beta(\gamma_3-\gamma_2)} - \frac{(1+\beta)w_1+x_3}{\beta\gamma_3} - \frac{x_3}{\beta(\gamma_3-\gamma_2)}\right)(x_3-c)\\
&\text{subject to} && \frac{x_3}{\beta(\gamma_3-\gamma_2)} \leq 1 \quad \text{and} \quad x_3 \geq \frac{(1-\theta-\bar{U})\beta(\gamma_3-\gamma_2)}{(1-\theta)}\\
&\text{and} && x_3 \geq \frac{(\gamma_3-\gamma_2)[(1+\beta)-\beta\gamma_2(1-\theta-\bar{U})]}{\gamma_2\theta} \text{and} \quad \text{constraint 4}
\end{aligned} \quad \text{(A.36)}$$

where

$$\text{constraint 4} = \begin{cases} x_3 \leq \frac{(\gamma_3-\gamma_2)[\beta\gamma_3(1-\theta-\bar{U})-(1+\beta)]}{[(1-\theta)\gamma_3-\gamma_2]} & \text{if } [(1-\theta)\gamma_3-\gamma_2] > 0 \\ x_3 \geq \frac{(\gamma_3-\gamma_2)[\beta\gamma_3(1-\theta-\bar{U})-(1+\beta)]}{[(1-\theta)\gamma_3-\gamma_2]} & [(1-\theta)\gamma_3-\gamma_2] < 0 \end{cases} \quad \text{(A.37)}$$

This yields an interior solution and two corner solutions. If $[(1-\theta)\gamma_3-\gamma_2] > 0$, we have:

$$x_3^* = \frac{(\gamma_3-\gamma_2)[(2-\theta-\bar{U})\gamma_3-(1+\beta)+c]+(1-\theta)\gamma_3 c}{2[(1-\theta)\gamma_3+\gamma_3-\gamma_2]} \quad \text{(A.38)}$$

$$x_3^* = min\left[\beta(\gamma_3-\gamma_2), \frac{(\gamma_3-\gamma_2)[\beta\gamma_3(1-\theta-\bar{U})-(1+\beta)]}{[(1-\theta)\gamma_3-\gamma_2]}\right] \quad \text{(A.39)}$$

$$x_3^* = max\left[\frac{\beta(\gamma_3-\gamma_2)(1-\theta-\bar{U})}{(1-\theta)}, \frac{(\gamma_3-\gamma_2)[(1+\beta)-\beta\gamma_2(1-\theta-\bar{U})]}{\gamma_2\theta}\right] \quad \text{(A.40)}$$

Conversely, if $[(1-\theta)\gamma_3 - \gamma_2] < 0$, we have:

$$x_3^* = \frac{(\gamma_3-\gamma_2)[(2-\theta-\bar{U})\gamma_3 - (1+\beta) + c] + (1-\theta)\gamma_3 c}{2[(1-\theta)\gamma_3 + \gamma_3 - \gamma_2]} \tag{A.41}$$

$$x_3^* = \beta(\gamma_3 - \gamma_2) \tag{A.42}$$

$$\begin{aligned} x_3^* = max\Bigg[\frac{\beta(\gamma_3-\gamma_2)(1-\theta-\bar{U})}{(1-\theta)}, &\frac{(\gamma_3-\gamma_2)[\beta\gamma_3(1-\theta-\bar{U})-(1+\beta)]}{[(1-\theta)\gamma_3-\gamma_2]}, \\ &\frac{(\gamma_3-\gamma_2)[(1+\beta)-\beta\gamma_2(1-\theta-\bar{U})]}{\gamma_2\theta}\Bigg] \end{aligned} \tag{A.43}$$

The equilibrium exists iff the following condition is satisfied:

$$\beta(\gamma_3-\gamma_2)h^* \leq x_3^* < \beta(\gamma_3-\gamma_2) \tag{A.44}$$

where $h^* = 1-\theta-\bar{U} + \frac{\theta x_3^*}{\beta(\gamma_3-\gamma_2)}$

- Case 3: The demand function is given by:

$$\eta = \theta \int_{\frac{(1+\beta)w_1 + x_3}{\beta\gamma_3}}^{h^*} f(h)dh \tag{A.45}$$

The profit maximisation problem is then:

$$\begin{aligned} &\max_{x_3} \quad \pi = \theta\left(1-\bar{U}-\frac{(1+\beta+x_3)}{\beta\gamma_3}\right)(x_3-c) \\ &\text{subject to} \quad \frac{x_3}{\beta(\gamma_3-\gamma_2)} > 1 \quad \text{and} \quad \beta\gamma_3[1-\bar{U}-(1+\beta)] > x_3 \end{aligned} \tag{A.46}$$

This yields an interior solution and two corner solutions, as follows:

$$x_3^* = \frac{\beta\gamma_3(1-\bar{U}) - (1+\beta)w_1 + c}{2} \tag{A.47}$$

$$x_3^* = \beta(\gamma_3-\gamma_2) \tag{A.48}$$

$$x_3^* = \beta\gamma_3(1-\bar{U}-(1+\beta)) \tag{A.49}$$

The equilibrium exists iff the following condition is satisfied:

$$1-\bar{U} < 1 \leq \frac{x_3^*}{\beta(\gamma_3-\gamma_2)} \tag{A.50}$$

A.7 Deriving Assumptions $6'$ for Model Extension

To begin, I use the uniform distribution to calculate values for the conditional expected ability of each cohort. First, I consider the expected ability of workers in the private university cohort. There are two possible realisations of k_c and k_d respectively, as rich, high-ability workers may attend private or public university, while rich, low-ability workers may attend private university or remain non-graduates. Thus, I can write the expected ability of the private university cohort in each of these scenarios as:

- $E[h_i|i \in n_3, s' = \{k_a, k_b, 1_c, 2_d; w_k, h^*, x_3^*, y_i\}]$ is undefined, as the private university cohort$= 0$.
- $E[h_i|i \in n_3, s' = \{k_a, k_b, 1_c, 3_d; w_k, h^*, x_3^*, y_i\}] = E[h_i|h^* \leq h_i \leq 1] = \frac{1+h^*}{2}$
- $E[h_i|i \in n_3, s' = \{k_a, k_b, 3_c, 2_d; w_k, h^*, x_3^*, y_i\}] = E[h_i|0 \leq h_i \leq 1] = \frac{h^*}{2}$
- $E[h_i|i \in n_3, s' = \{k_a, k_b, 3_c, 3_d; w_k, h^*, x_3^*, y_i\}] = E[h_i|0 < h_i < h^*] = \frac{1}{2}$

Next, I consider the expected ability of workers in the public university cohort. This is unaffected by the value of the state variable, s' (although the equilibrium value of h^* will still be determined by other workers' choices). The conditional expectation is given by:

$$E[h_i|i \in n_2, s'] = E[h_i|h_i \geq h^*] = \int_{h^*}^{1} hf(h)dh = \frac{1+h^*}{2} \tag{A.51}$$

Now I calculate the capacity constraint required to appropriately restrict the set of equilibria.

Assumption $6'$ - Public University's Capacity Constraint
The purpose of the capacity constraint is to restrict the model equilibria to be reasonable, in the sense that they are ones in which high-ability, rich workers prefer free public university to remaining non-graduates. This requires that the following inequality hold in every equilibrium (that is, for the model parameters and every possible equilibrium value of h^*):

$$\beta\gamma_2 E[h_i|i \in n_2, s'] > (1+\beta) \tag{A.52}$$

Substituting in for the expected ability of a public university graduate yields:

$$\beta\gamma_2 \frac{1+h^*}{2} > (1+\beta) \tag{A.53}$$

h^* is lowest in the equilibrium described by Case 1, so if the inequality binds for this value of h^*, it will do so for all other equilibrium values of h^*. Thus, we require:

$$\frac{\beta\gamma_2}{2}\left\{1+\frac{(1-\theta-\bar{U})}{(1-\theta)}\right\} > (1+\beta) \tag{A.54}$$

Rearranging then yields the capacity constraint for public universities:

$$\frac{2(1-\theta)[\beta\gamma_2-(1+\beta)]}{\beta\gamma_2} > \bar{U} \tag{A.55}$$

A.8 Characterising the Equilibrium of the Model Extension

A.8.1 Workers' Decision Rules

A fraction $(1-\theta)h^*$ of the population consists of workers with $y_i = y^L$ and $h_i < h^*$. As in the previous model, this group always remains non-graduates in equilibrium, as this is the only choice available to them.

A fraction $(1-\theta)(1-h^*)$ of the population consists of workers with $y_i = y^L$ and $h_i \geq h^*$. The optimal choice for this group is:

$$d_k^* = \begin{cases} d_2 & \text{if } \beta\gamma_2 E[h_i|i \in n_2, s'] > (1+\beta)w_1 \\ d_1 & \text{otherwise} \end{cases} \tag{A.56}$$

A fraction θh^* of the population consists of workers with $y_i = y^H$ and $h_i < h^*$. The optimal choice for this group is:

$$d_k^* = \begin{cases} d_3 & \text{if } \beta\gamma_3 E[h_i|i \in n_3, s'] - x_3 > (1+\beta)w_1 \\ d_1 & \text{otherwise} \end{cases} \tag{A.57}$$

Finally, a fraction $\theta(1-h^*)$ of the population consists of workers with $y_i = y^H$ and $h_i \geq h^*$. This group can choose freely amongst all three education alternatives. Nevertheless, Assumption 6′ makes being a non-graduate a strictly dominated strategy for workers in this group. Thus, this optimisation problem is now reduced to a simple binary choice between private and public university. The optimal choice for this sub-group is:

$$d_k^* = \begin{cases} d_3 & \text{if } \gamma_3 E[h_i | i \in n_3, s'] - \frac{x_3}{\beta} > \gamma_2 E[h_i | i \in n_2, s'] \\ d_2 & \text{otherwise} \end{cases} \tag{A.58}$$

A.8.2 Worker Sorting in Equilibrium

As in the baseline model, the feasible choice sets and decision rules described above result in a set of three distinct cut-off values for relative pay-offs:

1. $\beta\gamma_2 E[h_i|i \in n_2, s'] > (1+\beta)w_1$
2. $x_3 < \beta\gamma_3 E[h_i|i \in n_3, s'] - (1+\beta)w_1$
3. $x_3 < \beta\gamma_3 E[h_i|i \in n_3, s'] - \beta\gamma_2 E[h_i|i \in n_2, s]$

This allows me to define eight potential Bayesian Nash equilibria for the model economy, based on the subset of these three inequalities that is satisfied by the model parameters. Nevertheless, Assumption $6'$ allows the elimination of some of these equilibria.

The first candidate equilibrium is one in which none of the three inequalities is satisfied. The second is one in which only the second inequality is satisfied, and the third is one in which only the third inequality holds. The fourth candidate equilibrium is one in which only the first two inequalities holds, while the fifth is one in which only the second and third inequalities are satisfied. These are all impossible by Assumption $6'$, which requires the first inequality always to be satisfied. The sixth candidate equilibrium is one in which only the first inequality holds. This equilibrium is possible, and defines the shut-down condition for private universities. All high-ability workers attend public university, while all low-ability workers remain non-graduates. The seventh candidate equilibrium is one in which only inequalities 1 and 3 hold, and the eighth is one in which all three inequalities hold. These are both possible.

Thus, I have reduced the number of candidate equilibria to three. I now consider workers' sorting behaviour under each of these. All workers with $y_i = y^L$ maintain the same optimal behaviour behaviour across all three equilibria, namely that all workers with $y_i = y^L$ and $h_i < h^*$ always remains non-graduates, while the sub-group with $y_i = y^L$ and $h_i \geq h^*$ always attends public university. Thus, I now characterise the three candidate equilibria in terms of the sorting behaviour of the two remaining sub-groups: namely, rich workers, of both high and low ability:

- Case $1'$: Inequalities 1, 2 and 3 are satisfied

In this equilibrium, all rich workers (regardless of ability) attend private university.

- Case $2'$: Inequality 1 is satisfied
  In this equilibrium, all rich, high-ability workers attend public university, while rich, low-ability workers remain non-graduates. This defines the shut-down condition for private universities.
- Case $3'$: Inequalities 1 and 2 are satisfied
  In this equilibrium, all rich, high-ability workers attend public university, while rich, low-ability workers attend private university.

A.9 Public and Private University Equilibrium Behaviour in Model Extension

A.9.1 Public University's Decision

Public universities set h^* in order exactly to fill the available number of seats, $\bar{U}$. There are 2 possible equilibrium values of h^*.

- Case $1'$
  h^* is chosen to satisfy the following equation:

$$\bar{U} = (1-\theta)(1-h^*) \tag{A.59}$$

  This yields the equilibrium value of h^*:

$$h^* = \frac{1-\theta-\bar{U}}{1-\theta} \tag{A.60}$$

- Case $2'$ or Case $3'$:
  h^* is chosen to satisfy the following equation:

$$\bar{U} = (1-h^*) \tag{A.61}$$

  This yields the equilibrium value of h^*:

$$h^* = 1-\bar{U} \tag{A.62}$$

A.9.2 Private University's Decision

As in the baseline model, the demand function for private university consists of two groups. Recall that rich, high-ability workers choose private university if the following condition is satisfied:

$$x_3 \leq \beta\gamma_3 E[h_i | i \in n_3, s'] - \beta\gamma_2 E[h_i | i \in n_2, s] \tag{A.63}$$

Conversely, rich, low-ability workers with ability choose private university if the following condition is satisfied:

$$x_3 \leq \beta\gamma_3 E[h_i | i \in n_3, s'] - (1+\beta)w_1 \tag{A.64}$$

The private university now has three options. It can set $x_3 = \beta\gamma_3 E[h_i|i \in n_3, s'] - \beta\gamma_2 E[h_i|i \in n_2, s]$ and serve both groups (this yields the pooling equilibrium described by Case 1′); it can set $x_3 = \beta\gamma_3 E[h_i|i \in n_3, s'] - (1+\beta)w_1$ and serve only rich, low-ability workers (this yields the separating equilibrium of Case 3′), or it can shut down (this yields the equilibrium of Case 2′). Thus, to see which equilibrium is optimal for the private university, I now calculate and compare equilibrium profit for each case.

Firstly, however, it is important to note that, in any equilibrium in which the private university stays open, one of the two inequalities above must bind and the other must be slack. (Suppose not. Then the private university can always profitably increase x_3, which means that the original value of x_3 is not an optimum.)

Case 1′:
The equilibrium tuition fee and profit are given by:

$$x_3^* = \frac{\beta\gamma_3}{2} - \frac{\beta\gamma_2(1-\theta-\bar{U})}{2(1-\theta)} - \frac{\beta\gamma_2}{2} \tag{A.65}$$

$$\pi^* = \theta\left(\frac{\beta\gamma_3}{2} - \frac{\beta\gamma_2(1-\theta-\bar{U})}{2(1-\theta)} - \frac{\beta\gamma_2}{2} - c\right) \tag{A.66}$$

The set of necessary and sufficient conditions for the existence of this equilibrium is:

$$x_3 \leq \frac{\beta\gamma_3}{2} - (1+\beta) \tag{A.67}$$

$$x_3 \leq \frac{\beta\gamma_3}{2} - \frac{\beta\gamma_2(1+h^*)}{2} \tag{A.68}$$

$$1 - \frac{\beta\gamma_2\bar{U}}{\beta\gamma_3\bar{U}^2 - 2\beta\gamma_3\bar{U} - 2(1+\beta)(1-\bar{U}) + 2\bar{U}c + 2\beta\gamma_2} < \theta < 1 \tag{A.69}$$

The third condition describes the requirement that profit under Case 1′ exceeds profit under

Case 3′.

To see that it is indeed an equilibrium, recall that, by Assumption 6′, high-ability rich workers always prefer some form of university education to remaining a non-graduate. When h^* is set sufficiently low, such that the condition above binds, and expected ability of the private university cohort is unaffected by the value of h^*, the relative gain in terms of expected ability from moving to a public university is low. Thus, their best response is to choose private university over public. This is because the larger productivity boost provided by private university outweighs the benefit from being perceived to be of high average ability through attending public university.

Low-ability, rich workers prefer private university to being a non-graduate by Assumption 4, and are constrained from deviating to public university. High-ability, poor workers prefer public university to being a non-graduate by Assumption 4, and prefer public to private by the condition above. Low-ability, poor workers never have any choice but to remain non-graduates. Thus, no-one deviates.

Case 2′:
The equilibrium tuition fee and profit are given by:

$$x_3^* > \frac{\beta\gamma_3(1-\bar{U})}{2} - (1+\beta) \tag{A.70}$$

$$\pi^* = 0 \tag{A.71}$$

The set of necessary and sufficient conditions for this equilibrium is:

$$x_3 > \beta\gamma_3 E[h_i | i \in n_3, s = k_a, k_b, 3_c, 3_d] - (1+\beta)w_1 \tag{A.72}$$

$$x_3 > \beta\gamma_3 E[h_i | i \in n_3, s = k_a, k_b, 3_c, 3_d] - \beta\gamma_2 E[h_i | i \in n_2, s] \tag{A.73}$$

In this equilibrium, private universities shut down. Both rich and poor workers with $h_i \geq h^*$ attend public university, while all other workers remain non-graduates. Note that, in this equilibrium, the high demand for public university pushes up the cutoff in order to keep intake equal to the university's capacity constraint, $\bar{U}$.

Case 3′:

The equilibrium tuition fee and profit are given by:

$$x_3^* = \frac{\beta\gamma_3(1-\bar{U})}{2} - (1+\beta) \tag{A.74}$$
$$\pi^* = \theta(1-\bar{U})\left(\frac{\beta\gamma_3(1-\bar{U})}{2} - (1+\beta) - c\right) \tag{A.75}$$

Recall that, in this equilibrium, all high-ability workers attend public university, all rich, low-ability workers attend private university, and all poor, low-ability workers remain non-graduates. The set of necessary and sufficient conditions for the existence of this equilibrium is:

$$x_3 \leq \frac{\beta\gamma_3 h^*}{2} - (1+\beta) \tag{A.76}$$
$$x_3 > \frac{\beta\gamma_3 h^*}{2} - \frac{\beta\gamma_2(1+h^*)}{2} \tag{A.77}$$
$$0 < \theta \leq 1 - \frac{\beta\gamma_2\bar{U}}{\beta\gamma_3\bar{U}^2 - 2\beta\gamma_3\bar{U} - 2(1+\beta)(1-\bar{U}) + 2\bar{U}c + 2\beta\gamma_2} or \theta = 1 \tag{A.78}$$

The third condition describes the requirement that profit under Case $3'$ exceeds profit under Case $1'$. This equilibrium is inefficient because now all of the highest-ability workers in the economy attend public university, which gives a lower productivity boost than does private. To see that it is indeed an equilibrium, recall that, by Assumption $6'$, rich, high-ability workers always prefer some form of university education to remaining a non-graduate. When h^* is set sufficiently low, such that the condition above is satisfied, and the expected ability of the private university cohort is decreasing in h^*, attending private university implies a signal of such low ability that their best response is to choose public university over private. This is because the benefit from being perceived to be of higher average ability outweighs the larger productivity boost provided by private university. Low-ability, rich workers prefer private university to being a non-graduate by Assumption 4, and are constrained from deviating to public university. High-ability, poor workers prefer public university to being a non-graduate by Assumption 4, and prefer public to private by the condition above. Low-ability, poor workers never have any choice but to remain non-graduates. Thus, no-one deviates.

A.10 Robustness Check

As a robustness check, I re-estimate the structural parameters using the ability score, rather than the ETS score, as the chosen measure of ability. As Table A.1 shows, the results are largely invariant to this change.

Table A.1: Estimated Structural Parameters using Ability Score

	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya
university marginal cost, c	0.10	2.4	0	2	6.7	0-0.4
private university productivity, γ_3	5.94	19.01	20.82	14.92	25.97	34.60
public university productivity, γ_2	5.69	10.12	14.22	12.69	27.89	23.61

A.11 Defining "Rich" and "High-Ability" in the Data

To determine who is "rich" in the data, I proceed in the following manner. First, I obtain auxiliary data on private university tuition fees in the STEP countries. I do so by searching the websites of 2-3 randomly-selected private universities in each country, and then taking an average of the posted tuition fees. I then compute the fraction of workers whose annual wage income is greater than or equal to this average. Finally, I select as the cut-off for "rich" the value of the household rank score which matches this fraction. This results in a cut-off score of 7.

To determine who is of "high ability" in the data, I consider my two measures of ability (ETS score and ability score) separately. The ability score ranges between 1 and 4, so I define any worker with a score of 4 to be a "high-ability" worker. This includes approximately 24 per cent of the workforce across all countries.

The ETS score ranges between 4 and 20, but the maximum value observed is much lower, and varies across countries. Thus, for each country, I define a "high-ability" worker to be one with the highest observed ETS score for that particular country. This results in a cut-off score of 8 for Georgia, 9 for Armenia, Bolivia, Colombia and Kenya, and 13 for Kenya.

A.12 Calculating the Social Planner's Optimal Allocation

The social planner's optimal allocation of human capital is one in which each worker's choice of education is informed solely by his marginal productivity in the graduate and non-graduate sectors and the marginal cost of education. I assume, as with public university education, that this cost is financed via a lump-sum tax on all workers.

The difference in marginal productivity between working in the graduate and non-graduate sector is given by $\gamma_3 h_i - 2$, as non-graduates contribute a single unit of output in both periods. The marginal cost is the change in the tax burden. This is simply c, assuming the government takes over the running of these universities. Thus, a worker should attend private university if and only if $\gamma_3 h_i > 2 + c$, and should remain a non-graduate otherwise.

Appendix B

Chapter 2 Appendix

B.1 Computing Parental Wealth

First, I compute household wealth for families in the GLSS 2012 using a rich household asset module. Household wealth is defined as the sum of the listed value of all household durable goods, savings, livestock and agricultural equipment.

Second, I regress the log of this wealth measure on the education level, occupation and geographic region of males aged 35 to 60 (35 is the youngest reasonable age for a father of workers aged at least 15). Education is measured on a 12-point scale ranging from no education to a post-graduate degree. Occupations are classified by 8 major groups based on the International Standard Classification of Occupations (ISCO-88). Results are shown in Table B.1.

Table B.1: Wealth Regression for Men Aged 35 to 60

	(1) log wealth
Constant	8.692^{***}
	(0.148)
education dummies	Yes
occupation dummies	Yes
region dummies	Yes
N	5701

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Third, I use the coefficients from this regression to predict the wealth of sampled individuals' fathers based on information regarding the latter's education and occupation, and the former's birth region. Figures B-1, B-2 and B-3 show the variation in household wealth over fathers' education, occupation and region, respectively.

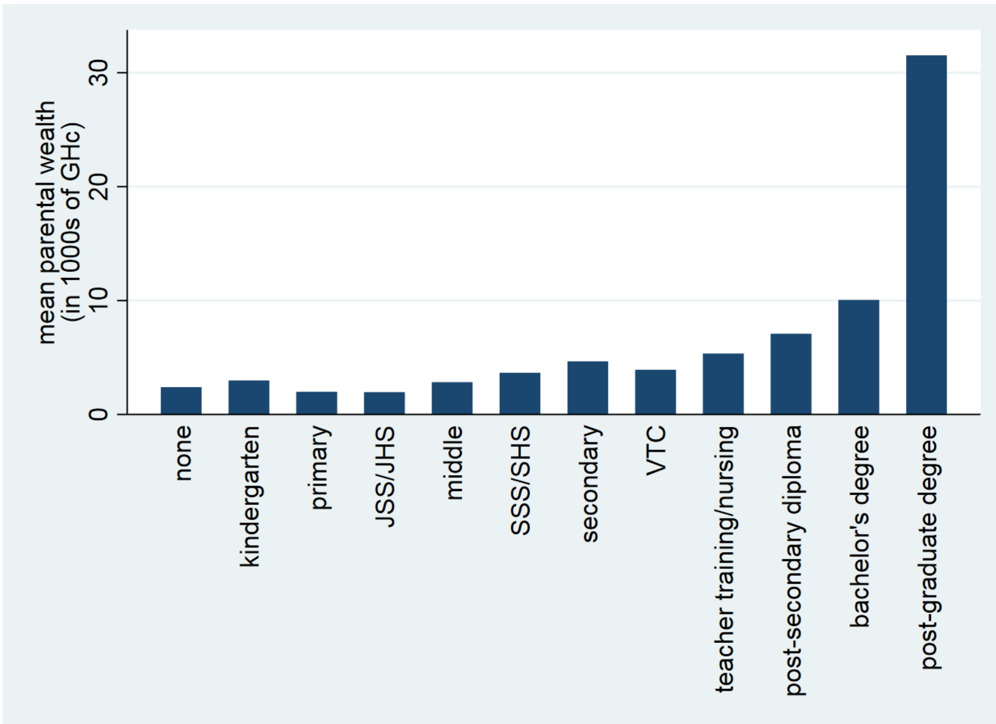


Figure B-1: Parental Wealth and Father's Educational Attainment for Males Aged 16-50

To test the performance of this measure, I consider its correlation with the child's educational attainment. Table B.2 shows that this correlation is strongly positive, as expected.

Table B.2: Parental Wealth and Educational Attainment for Males Aged 16-50

	(1)
	years of education
log parental wealth	1.657***
	(0.137)
_cons	-4.714***
	(1.073)
N	9059

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

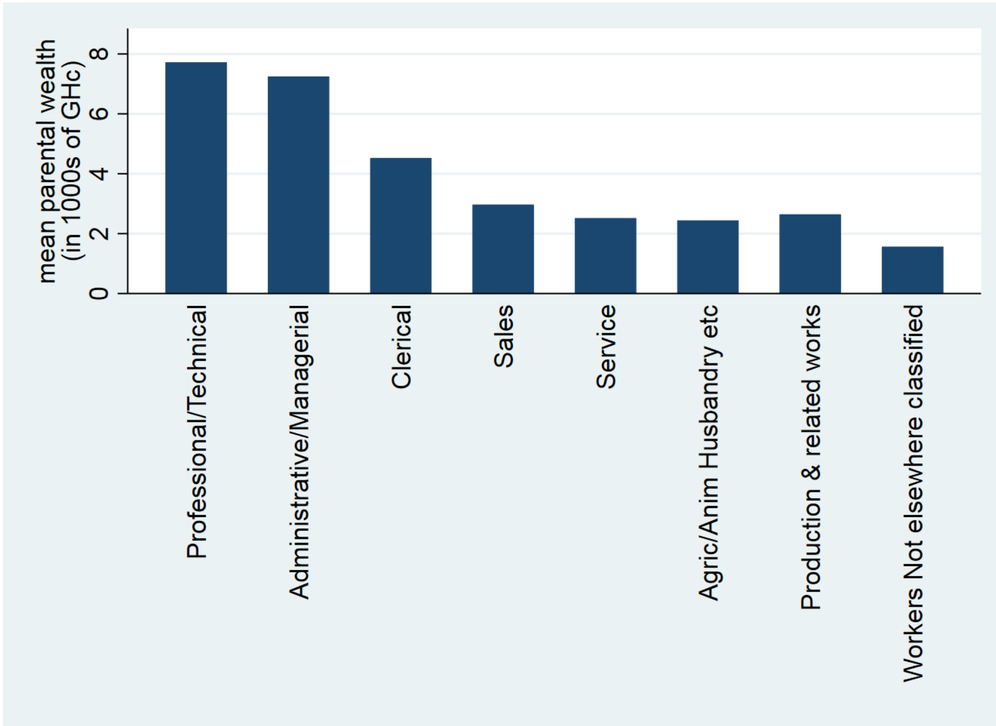


Figure B-2: Parental Wealth and Father's Occupation for Males Aged 16-50

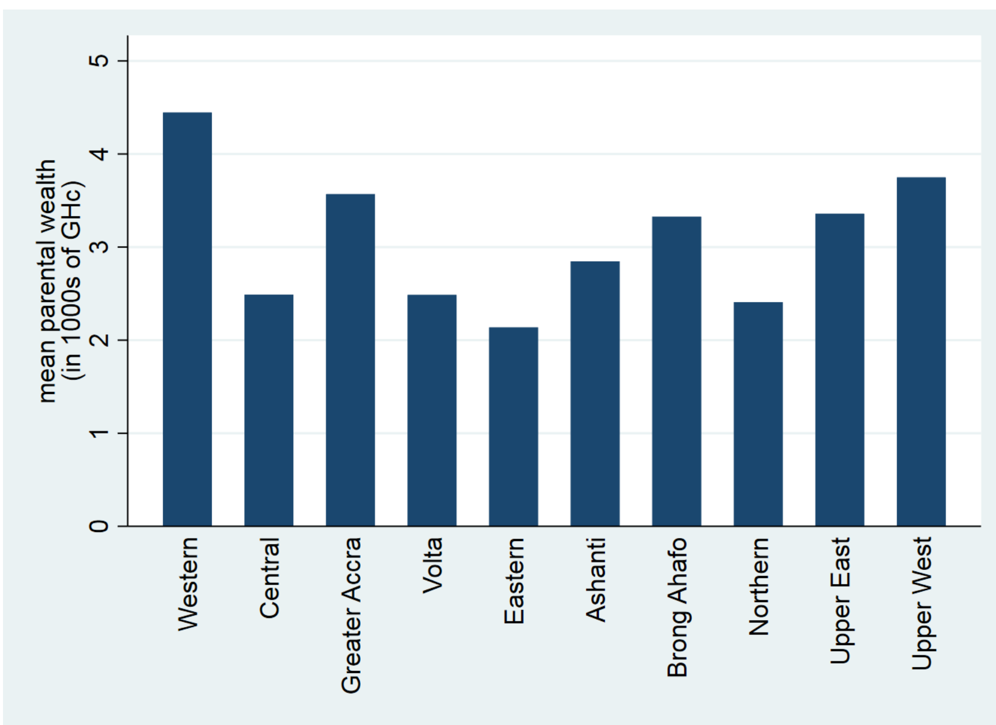


Figure B-3: Parental Wealth and Birth Region for Males Aged 16-50

B.2 Stylised Facts: Robustness Checks

Table B.3 replicates the results for Ghana from Table 2.1 using a probit specification in place of a linear probability model.

Table B.3: GLSS 2012 Parental Wealth and Male Unemployment: Probit (Average Marginal Effects)

	(1) Age 15-29 Unemployed = 1	(2) Age 30-60 Unemployed = 1	(3) Age 15-29 Unemployed = 1	(4) Age 30-60 Unemployed = 1
log parental wealth	0.055***	0.009	0.039**	0.009
	(0.011)	(0.005)	(0.012)	(0.005)
age	-0.007***	0.000	-0.009***	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
urban	0.070***	0.006	0.060***	0.006
	(0.011)	(0.004)	(0.010)	(0.004)
region dummies	Yes	Yes	Yes	Yes
education dummies	No	No	Yes	Yes
N	3328	7172	3328	7172

Marginal effects; Standard errors in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4 uses a different dependent variable, namely the number of years between the end of schooling and the start of the first job, and demonstrates a positive and significant relationship between parental wealth and this new measure of youth unemployment.

Table B.5 replicates the results from Table 2.1 using a secondary data source, namely the first three rounds of the Ghana Urban Household Panel Survey (2004-2006), which include information on respondents' labour market histories. Standard errors are clustered at the individual level. The results are very similar to those in Table 2.1.

To show that these patterns are not unique to the Ghanaian context but, rather, may also be present in other sub-Saharan African countries, I replicate the trio of stylised facts outlined in Section 2.2 for Uganda, using data from the fourth round of the Uganda National Panel Survey (2013/2014). Note that this data source has no substantive data available for workers older than 50 years.

First, the 2012 unemployment rate among Ugandan men aged 15-24 is 4.7%. Second, I replicate the educational attainment results from Figure 2-1 in Figure B-4. Finally, I replicate

Table B.4: GLSS 2012 Parental Wealth and Male School-to-Work Transition Duration

	(1) Age 30-60 transition years	(2) Age 30-60 transition years
log parental wealth	7.748***	7.078**
	(2.301)	(2.212)
log parental wealth^2	-0.472***	-0.420**
	(0.141)	(0.135)
age	0.021**	0.022**
	(0.007)	(0.007)
urban	0.680***	0.665***
	(0.127)	(0.132)
region dummies	Yes	Yes
education dummies	No	Yes
mean (transition years)	3.800	3.800
N	3598	3598

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Data Source: GLSS 2012

Table B.5: Ghana UHPS: Parental Wealth and Male Unemployment

	(1) Age 15-29 Unemployed = 1	(2) Age 30-60 Unemployed = 1	(3) Age 15-29 Unemployed = 1	(4) Age 30-60 Unemployed = 1
log parental wealth	0.048**	0.028	0.044**	0.034
	(0.016)	(0.025)	(0.017)	(0.025)
age	-0.007***	-0.000	-0.006**	0.000
	(0.002)	(0.001)	(0.002)	(0.001)
region dummies	Yes	Yes	Yes	Yes
education dummies	No	No	Yes	Yes
N	6846	7297	6797	7255

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the results from Table 2.1 in Table B.6. The patterns are very similar across both countries: youth unemployment is low; educational attainment is very low, with only a small proportion of workers completing high school; and parental wealth is a positive and significant determinant of youth unemployment, but does not affect workers' unemployment status later in life. Note that these patterns hold despite the stark sub-regional differences between Ghana and Uganda, the former a part of West Africa, and the latter located in East Africa.

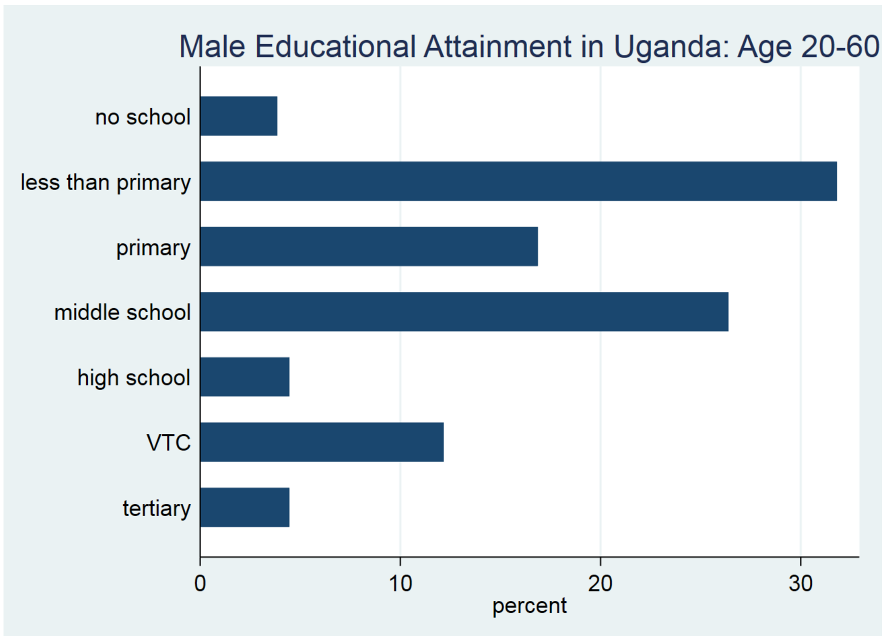


Figure B-4: Educational Attainment of Ugandan Males Aged 20-60

Table B.6: Uganda National Panel Survey: Parental Wealth and Unemployment

	(1) Age 15-29 Unemployed = 1	(2) Age 30-60 Unemployed = 1	(3) Age 15-29 Unemployed = 1	(4) Age 30-60 Unemployed = 1
log parental wealth	0.183***	0.054	0.129**	0.055
	(0.044)	(0.045)	(0.045)	(0.049)
age	-0.007	-0.003	-0.012*	-0.003
	(0.005)	(0.003)	(0.005)	(0.003)
urban	0.031	0.007	-0.010	0.009
	(0.041)	(0.034)	(0.042)	(0.037)
region dummies	Yes	Yes	Yes	Yes
education dummies	No	No	Yes	Yes
N	422	251	422	251

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.3 Model Assumptions: Employment is an Absorbing State

To justify the assumption that workers stay in a single job for their lifetime in the model framework of this paper, I use data from the Ghana Urban Household Panel survey (2004) to show that job-to-job transitions in this context are very low, as Table B.7 shows. Note that I use this auxiliary dataset because the GLSS 2012 (the main dataset used in this paper) does not contain information about individuals' labour market histories.

I restrict the sample to workers aged between 30 and 40 years, because the labour market history data goes back to a maximum of 20 years, such that including older workers would lead to a potential underestimation of the number of jobs per worker. Similarly, including younger workers, who have spent only a short while in the labour force, would also lead to an underestimation.

Table B.7: Lifetime Jobs per Worker: Age 30-40

N	mean	s.d.
575	1.95	0.95

Data Source: Ghana Urban Household Panel Survey 2005

B.4 Summary Statistics: Male Labour-Force Participants Aged 16-50

Table B.8: Summary Statistics: Male Labour-Force Participants Aged 16-50

observations	9059	
urban (%)	38.03	
highly-educated (%)	24.33	
	mean	s.d.
age (years)	33.47	9.22
education (years)	8.24	6.19
median parental assets (US$)	513.97	508.78

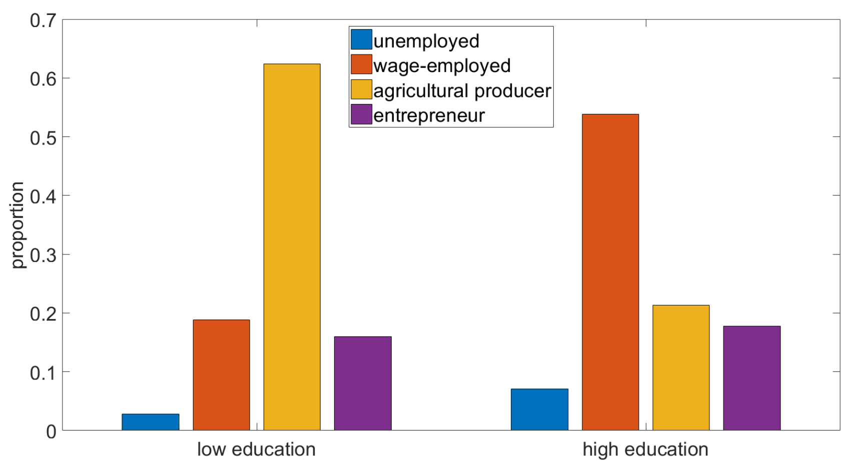


Figure B-5: Employment Status by Education

B.5 Estimating Earnings Growth Parameters

Table B.9: Log Earnings and Age for Low-Education Males Aged 31-50

	(1) wage-employed	(2) agric. self-employed	(3) entrepreneurs
age	0.00638	0.00442	0.0101
	(0.007)	(0.006)	(0.009)
_cons	7.903***	6.701***	7.994***
	(0.271)	(0.234)	(0.363)
N	680	2018	627

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.10: Log Earnings and Age for High-Education Males Aged 31-50

	(1) wage-employed	(2) agric. self-employed	(3) entrepreneurs
age	0.0276***	0.0431	0.0236
	(0.007)	(0.028)	(0.017)
_cons	7.791***	5.414***	7.693***
	(0.282)	(1.100)	(0.670)
N	574	124	206

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.6 Ability and Unemployment Duration by Wealth

As discussed in Section 2.3.5.4, the relationship between ability and unemployment duration has an ambiguous sign, as the "asset de-accumulation" channel is negative, and the "sectoral shift" channel is positive. Further, the relationship is not constant along the support of initial wealth; specifically, the "sectoral shift" channel becomes relatively more, and the "asset de-accumulation" channel relatively less, important at higher levels of wealth. This is because the difference between high- and low-ability workers in the value of wage employment and entrepreneurship is decreasing in wealth. Formally:

$$\frac{\partial^2 (VM_{i,m,k,t}|h_i = H^H)}{\partial a_{i0}^2} < \frac{\partial^2 (VM_{i,m,k,t}|h_i = H^L)}{\partial a_{i0}^2} \quad \forall \quad m \in \{w, e\} \tag{B.1}$$

Figure B-6 shows this quite clearly for the high-education wage-employment sector. The intuition underlying this is that, as the value of employment for high- and low-ability workers draws closer together at high wealth levels, the differences across ability types in consumption behaviour while unemployed become negligible, such that the "asset de-accumulation" channel diminishes in influence, such that high-ability workers become relatively more likely to remain unemployed.

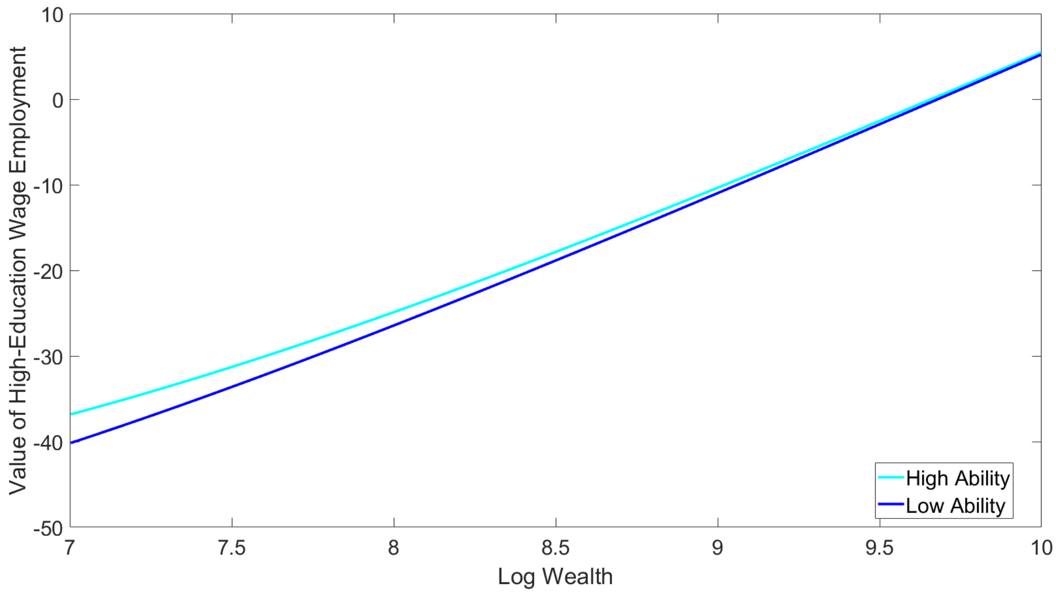


Figure B-6: Value of High-Education Wage Employment by Log Wealth