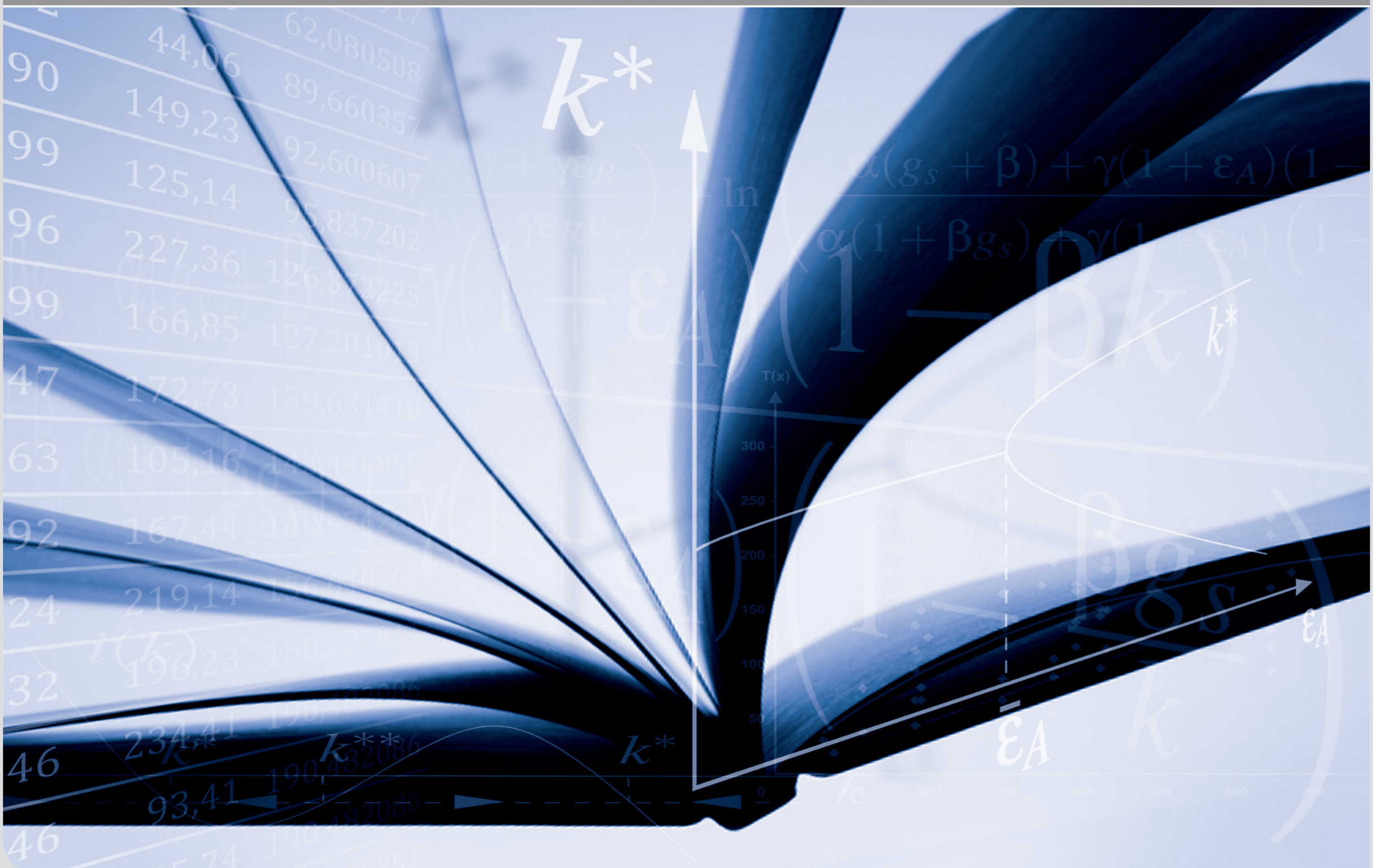
Karlsruher Institut für Technologie

On the value of foreign PhDs in the developing world: Training versus selection effects

by Helena Barnard, Robin Cowan, Moritz Müller

No. 82 | JANUARY 2016

WORKING PAPER SERIES IN ECONOMICS



Impressum

Karlsruher Institut für Technologie (KIT)
Fakultät für Wirtschaftswissenschaften
Institut für Volkswirtschaftslehre (ECON)

Schlossbezirk 12
76131 Karlsruhe

KIT – Die Forschungsuniversität in der Helmholtz-Gemeinschaft

Working Paper Series in Economics
No. 82, January 2016

ISSN 2190-9806

econpapers.wiwi.kit.edu

On the value of foreign PhDs in the developing world: Training versus selection effects

Helena Barnard[1], Robin Cowan[2,3,4] and Moritz Müller[2]

[1]GIBS, University of Pretoria, South Africa
[2]BETA, Université de Strasbourg, France
[3]Institut Universitaire de France
[4]UNU-MERIT, Maastricht University, Netherlands

January 25, 2016

Abstract

This paper compares the career effects of overseas and domestic PhD training for scientists working in an emerging economy, South Africa. Variations in scientific achievements of South African academics may arise because those who attend "better" PhD programmes receive better training, but it may also be because good students select into good universities. We examine selection and training effects for four tiers of South African and two tiers of foreign universities. Those who received PhDs from universities in industrialized countries tend to be more productive than those whose PhDs were locally granted, but universities from industrialized countries do not necessarily provide better training than local universities. Pure selection effects contribute to career outcomes nearly as much as training effects. When looking at training in isolation, PhDs from top South African universities produce a similar quantity and quality research output to those from leading universities in the developed world.[1]

JEL codes: F63; H52; I2; O15; O20; O30
keywords: Scientific mobility; Doctoral studies; University evaluation; Developing country; South Africa; Technological upgrading

[1]Financial support for the project was provided by the Institut Universitaire de France, and by the Chair in Economic Policy, ECON, KIT, Germany.
This paper is also circulating in the working paper series of BETA (http://www.beta-umr7522.fr/productions/workingpapers.php) and MERIT (http://www.merit.unu.edu/publications/working-papers/).

1 Introduction

The importance of universities in technological and economic development has long been recognised (Mowery and Rosenberg, 1999; Murmann, 2003; Rosenberg and Nelson, 1994) and in the context of increasingly knowledge-intensive economies, it is likely that universities will play an even greater role in the future upgrading of less developed countries. The importance of both foreign and indigenous universities in providing scientific training has been suggested on theoretical grounds (Nelson, 2005), but the field has thus far failed to ask whether there is a difference in the quality of PhD training provided locally versus abroad. This paper looks at the scientific achievements of South African academics and explores how the academic performance of scholars who have received PhD training in their home country compares to that of scholars who were trained in the developed world.

This question matters both practically — developing countries invest substantial resources to send their top students to presumed superior universities abroad — and theoretically. Probably because of the complex and multifaceted role played by universities, scholars have struggled to isolate the specific mechanisms by which universities contribute to society. To advance scholarship on the question of how universities support upgrading, a narrow focus may at this stage be useful. Much as linkages and supporting institutions are essential to allow the wider society to benefit from the work done at universities (Brundenius, Lundvall and Sutz, 2011), the quality of scientific training received at a university is foundational to what the university can contribute to its society. This paper helps to clarify at which universities scholars from developing countries are likely to receive the best training.

The scholarly achievement of academics following their PhDs may occur because they selected (and were selected by) good universities, or because of the training they had received at that university. Our econometric model allows for both selection and training effects, and we examine them for four tiers of South African and two tiers of foreign universities. We rely on the database of rated academic researchers of the National Research Foundation of South Africa. This database provides evidence of, inter alia, the quality level of scholars (as established through an intensive process of peer review), their research productivity and where they received their training. As can perhaps be expected given the perceptions of local versus foreign universities, the overall effect of a foreign PhD is more positive than that of a local PhD. But a decomposition into selection and training effects, shows that selection contributes to career outcomes nearly as much as training.

Interestingly, our results show that universities from the industrialized countries do not necessarily provide better training than local universities. When looking at training in isolation from selection, scholars with PhDs from the best South African universities produce a similar quantity and quality research output to those who had received their training from universities in the developed world.

Our findings suggest that it is perhaps too crude to differentiate along national lines between universities. The leading universities globally attract the best South African students, as they do for students across the world. But apart from that select group of universities, there are fewer differences between the research-intensive universities in developing countries and their counterparts in the developed world than have hitherto been acknowledged. A tiered differentiation between universities may be more useful in discussing the role of different universities.

2 Literature review

The important role of universities in technological and economic development has long been recognised (Mowery and Rosenberg, 1999; Murmann, 2003; Rosenberg and Nelson, 1994). Moreover, work on the growing importance of (scientific) knowledge to economic growth suggests that the role of universities will continue to increase in importance (Conceição and Heitor, 1999; Deiaco, Hughes and McKelvey, 2012). It follows that universities in middle-income countries are likely to play an especially important role in technology and knowledge upgrading (Altbach, 2013; Brundenius, Lundvall and Sutz, 2011).

The direct economic impact of universities is obviously of interest to scholars of economic development. Some form of direct economic benefit has long been an outcome (or desired outcome) of the university, with university spin-offs (Link and Scott, 2005), university-industry linkages (D'Este and Patel, 2007) and the 'entrepreneurial university' (Etzkowitz, 2003) all receiving scholarly attention. However, the economic impact of universities in industrialized versus developing countries cannot be directly compared. Much as there is a strong relationship between Silicon Valley and Stanford University, the economy of Bombay cannot be ascribed to the quality (or not) of the Indian Institute of Technology Bombay. The extent to which such benefits accrue is likely to be a reflection not only of the quality of the university, but also of a range of institutional and historical factors (Chen and Kenney, 2007).

The question of how universities in developing countries contribute to economic and technological upgrading is further complicated by the fact that universities play such a multifaceted role in society. In addition to their direct economic role, they contribute to a basic stock of knowledge, to a capacity for problem-solving and to skilled people (Salter and Martin, 2001). Universities are actors in nation-building and increasingly also in the building of cross-national connections (Scott, 2006). And while it is possible to learn from the rich historical evidence about the role of universities in the now-developed countries, such learning must be done with an awareness that globalization has changed both universities and how upgrading takes place. Thus Altbach (2004) cautions that globalization may lead to a form of academic 'neo-colonialism' where the institutions of developing countries model themselves in potentially inappropriate ways on the leading universities, while Mazzoleni and Nelson (2007) argue that globalization has resulted in an increased global connectedness of scientific communities for those with the 'appropriate training and connections' to get into the relevant networks. One effect of appropriate training and connections, presumably, is a more productive academic career. This raises the question of the definition of "appropriate" (which presumably contains reference to quality) and how training and connection-making are best carried out to further both individual careers and university contribution to development.

Moreover, apart from a few exceptions, e.g. Mathews and Hu (2007) and Mazzoleni and Nelson (2007), there is little empirical work on how universities in industrializing countries have been contributing to upgrading. The quite common practice of governments in underdeveloped countries (e.g. Delicado, 2010 re Portugal; Güngör and Tansel, 2008 re Turkey and Song, 1997 re South Korea) to support the study of students abroad suggests that indigenous universities may not be capable of training beyond a certain level. However, there are also concerns about the practice. For example, there are concerns about costs — should funds not rather be used to build local capacity? Also, the brain drain debate considers whether the potential higher human capital acquisition from a foreign PhD outweighs the potential loss of talent from those who choose to stay in the typically better resourced and better networked universities of the industrialised world.

One of the difficulties in advancing this debate has been that it is so hard to compare indigenous with foreign universities fairly. While recognizing the multifaceted nature of universities, we believe that there is benefit to a narrower focus. In this paper, we therefore focus on arguably the core function of the university, and certainly the one that differentiates the leading universities from others, namely the ability to create new knowledge. In comparing indigenous with foreign universities, two dimensions seem especially salient. The first is human capital development. Investment in human capital has long been recognized as a key enabler of development (Becker, 1962). Human capital development of course encompasses multiple activities, but given our desire to focus our study, we focus on the extent to which a university trains scholars who do cutting-edge research. A second important consideration relates to social capital. Not only is social capital instrumental in the development of human capital (Coleman, 1988), but in an era of globalization and the increased global connectedness of science networks (Mazzoleni and Nelson, 2007), universities are potentially important sites for the development of cross-national scholarly networks.

The point about networks and linkages is embedded in much of the historical work on upgrading. One of the central themes emerging from historical research on upgrading is the importance of cross-border contact and learning between people. Indeed, that point has been made far more commonly about other forms of connection than about universities. A concern with how cross-national connections take place has occupied scholars focusing on the nation as their unit of analysis (e.g. Lall, 2001; Ozawa, 1992), the firm (e.g. Athreye and Cantwell, 2007; Narula and Dunning, 2000; 2010) and the individual (e.g. Almeida, Phene and Li, 2014; Saxenian and Hsu, 2001). Learning and thus upgrading has long been recognized as resulting from the contact between more and less skilled people, or representatives from stronger and weaker institutions.

The issue of linkages and the connectedness of scholars not only to local 'users' like industry and policy makers, but importantly also to global scientific networks (Bernardes and Albuquerque, 2003; Mathews and Hu, 2007; Olds, 2007) is also a key theme in scholarship about the role of universities in the development of lower and middle income countries. Universities are seen as playing a significant role in facilitating cross-border learning. However, the universities of developing countries are often presumed to play a role primarily in the development of skilled personnel, while new knowledge is seen as being developed at foreign universities and transmitted to the less developed country through foreign-trained faculty members. The statement by Nelson (2005:27) exemplifies such a view:

"Indigenous universities will play a key role as the source of students who take advanced training abroad, and as the home of faculty who have been trained abroad."

Implicit in a statement like this is the belief that indigenous universities themselves are not capable of offering training beyond a certain level, or of developing new knowledge. Even faculty who have received training from leading universities globally are seen as teachers rather than creators of new knowledge once they return home. Given the scarcity of resources in developing countries, the need to pay faculty enough to at least live a middle class lifestyle (Altbach, 2013) and the high cost of developing and maintaining facilities like laboratories, this may well be the situation in developing countries.

This view of the respective roles of universities in different parts of the world is certainly is likely to have an effect on the process by which students from developing countries select the universities from which they wish to receive doctoral training. It is likely that many potential students will see training in a foreign university as desirable, and will actively explore options to study abroad. Given the limited resources in developing countries, government-sponsored programmes for PhD study abroad are typically very competitive, and only the top students are selected. The prestige of obtaining such funding may well reinforce the view that training at a local university is a less desirable choice. Even when students do not obtain government funding, it is almost the norm for foreign universities in the industrialized world to offer financial support to doctoral students, and such support is another avenue through which strong students from developing countries can receive doctoral training abroad.

But in addition to selection effects, the training that can be obtained at a given university also needs to be considered. One important dimension of such training is what Nelson and Sampat (2001) term 'social technologies'. They give the example of gaining mastery of the research methods of organic chemistry, and describe the social technology as 'the system of training young chemists in the relevant physical technology' (2001:50). They argue that 'social technologies', i.e. the customs and codes of practice in a field, are an essential complement to physical technologies and enable institutions to be effective. Nelson (2005) also argues that in less developed countries, universities are important mechanisms for the diffusion of such social technologies.

Under this view of the respective roles of universities in different parts of the world, universities in developing countries will vary in the effectiveness with which they use and disseminate such technologies. Indeed, a key dimension of our analysis is the conviction that different universities may operate at different tiers, (and we use the data to uncover these patterns). In the US, a highly stratified university system helps the higher education system to respond to rapid societal changes without compromising excellence (Conceicao and Heitor, 1999). In the case of fast-growing and under-resourced developing countries where there is a great need for skills development of the kind argued by Nelson (2005), a stratified system will allow the university system to play a similarly differentiated role. Research intensity may be the result of a deliberate government strategy, e.g. the role that the University of Sao Paolo is intended to play in Brazil, or of historical factors, e.g. the post-war role played by METU in Turkey that triggered a series of path-dependencies.

It is equally important to recognize that universities have historically developed instruments to address the challenges that inevitably emerge in the course of doing research (Rosenberg, 1992). Universities in developing countries may have to adapt not only physical but also social technologies to better address the specific challenges of their context. Certainly in the social sciences one of the challenges has to do with the multi-lingual nature of many developing societies. Supervisors in developing country universities are developing various strategies to supervise the work of students who are gathering data in a language the supervisor (and sometimes the student) may not know.

A similar point can be made not about the methods of doing research, but about its content. It has been repeatedly stated that universities in developing countries need to respond to local challenges to support upgrading (Brundenius, Lundvall and Sutz, 2011; Mazzoleni and Nelson, 2007; Nelson, 2004). The types of local challenges that are frequently mentioned (e.g. Lee, 2000; Mazzoleni and Nelson, 2007) are agriculture, medicine, and industry. Local conditions in aspects such as income, weather, demography, and epidemiology (Bernardes and Albuquerque, 2003:870) require the development of new technologies to improve on often inadequate solutions that have tended to originate from high income countries. There is general consensus that the best placed universities to develop such solutions are typically the leading universities of the developing world. However, research into the specific agricultural, medical or economic problems of a given underdeveloped region has long been seen as a purely local response to a local challenge by scholars and businesses from the industrialized world. Thus the problem, analysis of it, and solution have all been geographically localized in the developing world.

This situation is slowly changing. For example GE's India-based healthcare unit earned substantial returns on healthcare innovations that were developed to address some of the challenging conditions in India. These returns were earned not only in developing country markets, but also in industrialized economies. As evidence of the wider benefit of something like 'frugal innovation' (Zeschky, Widenmayer and Gassmann, 2011) increases, the wider scientific community is showing a greater interest in what was previously seen as the local challenges of developing countries. Application of solutions to local problems (of developing countries) are starting to be seen globally, implying integration at one end of the innovation process. Integration further upstream may also be valuable.

Scholarship advances in the tension between novelty and convention. Scholars need to demonstrate familiarity with the key works in the field, both their findings and their scholarly conventions, but must also advance beyond those contributions. The conventions guiding most fields are still shaped mainly in the developed world (Corbett, Cornelissen, Delios and Harley, 2014). But to the extent that scholars in developing countries are able to demonstrate an understanding and mastery of the current (global) customs and codes of practice of scholarship, i.e. social technologies, they are — arguably because rather than in spite of the challenges in their specific locations — well placed to contribute novel knowledge to their fields. This raises the question whether training locally or abroad will have greater impact on a scientists career. Certainly integrating international norms and standards will be crucial, but if that can be done within the local context, it may be that being embedded in the local knowledge context may provide advantages in terms of academic performance, at least in some fields. Local training may provide as many benefits as international training.

In the coming sections we develop an econometric model which dis-entangles selection and training effects at the PhD level. While a PhD from a top international university is presumed to do more for one's career, it is important to to understand the extent to which (if it is indeed true) this comes from superior training or merely selection and signalling. If training at good local universities is "as good as" training at foreign universities, the added expense of sending students to foreign PhD programmes may be a bad investment. With these issues in mind, we do not presume any ordering of the quality of training or the rigours of selection at universities of different status, and in particular we do not assume a priori that domestic universities of middle income countries are inferior to those of industrialized countries. Part of the issue is the role indigenous universities can play in upgrading. The effectiveness of a PhD from an indigenous university will be an indicator of the kind of role these universities can play in knowledge and technology upgrading.

3 The Model

3.1 Model set-up

The simple estimation of the effect of the quality of the PhD-granting university on a person's career is confounded by a selection effect. Something like "innate ability" will affect both which university selects a person, and that person's future career regardless of where the PhD was granted. To address this identification issue we use a structural equation model. The model includes an equation for individual ability in combination with a factor structure model, where ability is factored into both a selection and outcome equation.[2]

Our data are described in detail in Section 4, but to ease exposition we describe them briefly here. The data are generated by the rating system of the South African National Research Foundation (NRF). Roughly every four years, any researcher in South Africa wishing to apply for funds from the NRF, must apply to be "rated". This involves submitting a full CV, which is then refereed by half a dozen international experts. On the basis of their referee reports, the expert committee of the NRF assigns the researcher a rating (A, B, or C, *grosso modo*). In the analysis that follows, we use this rating as a measure of the quality of a researcher's scientific output, and take advantage of the detailed biographical information in the researcher's application to control for other confounding factors.

In the context of our data, we assume that the "quality" rating given by the NRF to a scientist depends on the individual's ability, ν_i, and a training effect

[2]See e.g. Song and Lee (2012) for measurement equations in structural equation models in combination with Aakvik et al. (2005) for factor structure models in treatment effect evaluation. Heckman et al. (2014) investigate a Bayesian approach very similar to ours for treatment evaluation.

which is specifc to the university at which the PhD was attained. Universities select doctoral students based on individual ability (ν_i), but universities are differently aggressive in their selection: perhaps the top universities are very tough, only taking students of the very highest ability; lesser-ranked universities cannot afford to be so selective as they need to fill their classrooms. This effect is captured by the parameter ζ_p. Generally, one would expect that institutions with a high training effect, (β_p), are able to implement higher selection pressure, (ζ_p), and attract more able students, (ν_i). The model allows for such positive correlation.[3]

Outline of the model The model is comprised of three equations which we estimate as a system. The basic structure contains equations for three variables: individual ability; PhD institution selection; and rating. Individual ability is latent through being unobserved. Selection and rating are modelled as outcomes of standard discrete-choice models, i.e. a multinomial and ordered logit respectively. The time ordering from ability (before PhD) to PhD institution selection and finally to subsequent ratings naturally results in a triangular system of three equations. We discuss each of the three equations below.

Initial ability — linear model Our concern is with the career effects of PhD study, or the added value of doing a PhD in particular types of universities. Thus for our purposes, "innate ability" is less interesting than an individual's "ability" level just before starting the PhD. Thus our ability equation (which is more a measurement equation than a modelling equation) contains the (type of) university granting the Master degree, and whether the student was awarded a distinction for that degree. We assume a linear model with normally distributed error term.

$$\nu_i = \boldsymbol{V}_i \boldsymbol{\gamma} + \epsilon_i, \tag{1}$$

where $\boldsymbol{V}_i$ includes our two indicators of ability (Master's institution and distinction in the Master degree), and age at Master degree.

Thus the model considers master studies separately from subsequent PhD studies. This is in accordance with the South African higher education system. In South Africa, differently from the US education system for example, holding a master degree or equivalent postgraduate qualification is a prerequisite for being admitted as a PhD student (Department of Education, 1998, Du Toit, 2012). Nevertheless a problem may arise in practice when the Master degree is seen as the

[3]We assume that selection, training, and individual effects are all stable over time. Since it might well be the case that the comparative advantage of South African universities has improved against foreign universities over time, allowing for a trend in the selection and training effects would be desirable. Unfortunately, the data do not provide sufficient information to estimate a reliable time trend.

first step in the PhD process, and so the Master and PhD are considered two parts of one educational programme. When this is the case selection takes place before rather than after the Master degree. We investigate how that may affect our estimations in the sensitivity analysis by restricting attention to individuals who have Master and PhD degrees from different institutions.

PhD university selection equation — multinomial logit We partition our PhD granting universities into P groups, described below. Thus an individual receives a PhD from one of the P university types. The selection process assigns a student to one of the types of universities. Selection is two-sided, in the sense that while it is based on individual ability, different types of universities will be differently selective on that variable. (We might assume that the top universities select only highest ability students, lower-ranked unversities are more catholic.) We treat selection using a multinomial logit model — any individual i has a probability of being selected into each of the P university types. Thus the dependent variable is a latent choice vector D_i^* with P elements. Each element $D_{i,p}^*$ can be interpreted as the fit between individual i and a university of type p, based on ability and other controls:

$$D_{i,p}^* = \boldsymbol{Z}_i \boldsymbol{\alpha}_p + \zeta_p \, \nu_i + \epsilon_{i,p}, \tag{2}$$

where $\boldsymbol{Z}_i$ is a (L)-vector of observed factors which vary over individuals, the corresponding coefficients $\boldsymbol{\alpha}_p$ vary over institutions. $\epsilon_{i,p}$ is an iid extreme value distributed error term. The usual relation between latent and observed choice vectors applies. The selection effect is modelled through the factor ζ_p. What $D_{i,p}^*$ captures is the jointly determined desirability of agent i attending university (of type) p. If we consider $\boldsymbol{Z}_i \boldsymbol{\alpha}_p + \zeta_p \, \nu_i$ (dropping ϵ for the moment) we can think of this as the term defining the probability that an agent with ability ν_i attends a university of type p. Thus, selection probability is jointly determined by the attractiveness of universities of type p to agent i, and the attractiveness of agents with ability ν_i to universities of type p. When ϵ is realized, agent i attends the university having the largest entry in the vector D_i^*.

The vectors $\boldsymbol{D}_i^*$ are not inherently interesting. Rather they are a means of extracting ζ_p and ν_i, which are otherwise unobserved.

Rating equation — ordered logit The rating is modelled as an ordered probit with a latent rating $R_{i,t}^*$ (which takes a real value). The rating, $R_{i,t}^*$, of individual i at time t is modelled as a function of the individual's ability ν_i, observed factors $\boldsymbol{X}_{i,t}$, and an unobserved error term $\epsilon_{i,t}$, assumed to be iid extreme value distributed.

$$R^*_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta}_p + \nu_i + \epsilon_{i,t}. \tag{3}$$

Differences in training across universities are modelled by letting coefficients $\boldsymbol{\beta}_p$ on observed factors (including the intercept) vary freely with the university p where the PhD was granted. $\epsilon_{i,t}$ is an iid extreme value distributed error term which results in an ordered logit model.The latent rating results in observed rating $R_{i,t}$ depending on which thresholds (c) the latent outcome has crossed. Simultaneously with the estimation of Equation 3 is the estimation of the vector of thresholds c which translate from the latent, real-valued, rating to the observed, discrete-valued actual rating granted by the National Research Foundation.

Table 1 summarises notation and indexing of model variables. Table 2 shows the interdependence of the variables across the three equations. The first column lists all endogenous and exogenous variables used in the model. The subsequent three columns correspond to the three equations which model the endogenous variables, i.e. Ability, PhD university category, and Rating respectively. Whenever a variable enters an equation the table entry is marked with a cross. Here we can see the system as a whole. We note that endogenous variables form a triangular system, and each equation includes at least one exogenous factor which is excluded from all other equations. Identification rests on both the triangular form and exclusion restrictions.

3.2 Identification and estimation

All three individual equations model latent outcomes, two of them with discrete observed outcomes. Therefore we impose some scale and level restriction in each equation; by fixing the standard deviation of the error terms and introducing a reference category in the choice settings respectively. The overall system of equations is then identified through the exclusion of factor(s) from one equation to subsequent equations within a triangular equation system.[4] For example we assume that the master degree university enters rating only through the (signal of) initial ability and not directly. Furthermore, the selection equation includes factors which do not enter the rating equation. We discuss these more in detail below in the variable description, and assess their empirical importance in the robustness section.

Measurement equation of initial ability All three equations are connected through the unobserved ability, ν.

[4]Only relying on structural form and error term distribution assumptions does not allow for reliable identification in practice.

Table 1: Model variables overview

Indexing:	
$i = 1, \ldots, N$	Individuals in the sample
$t = 1, \ldots, T_i$	Rating calls of individual i
$p = 1, \ldots, P$	PhD university categories
Observed variables:	
$R_{i,t}$	Rating of individual i at application t (ordered from C3 to A1)
D_i	University (p) where PhD obtained (unordered)
V_i	Factors related to innate ability
Z_i	Factors affecting PhD university choice
$X_{i,t}$	Factors affecting rating outcomes
Latent variables:	
$R^*_{i,t}$	Latent rating outcome
D^*_i	Latent PhD university indicator vector of length P
ν_i	Unobserved individual ability
Parameters:	
γ	Ability equation
α_1, α_p	Selection equation
ζ_p	PhD selection effect of university category p
β_p	Training effect of university category p
c	Hurdles in rating equation

The expected unobserved ability is anchored at zero for students obtaining their master from a 'typical' SA university (our university reference category), being of average age, and obtaining no master distinction. Ability is then freely estimated for master students deviating from this reference case. Thus, we do not model unobserved ability in absolute terms but relative to the ability of a reference master student. The error term is standard normal distributed. This way the variance of the combination of error and ability term is fixed in subsequent equations.

Selection equation Constraints implied by the multinomial logit model are standard (see Train, 2009). Since by definition the alternative with the highest 'utility' is chosen, only the ranking of alternatives can be modelled. Most factors which enter the PhD university decision describe the individual, and hence do not vary over alternatives within a decision. Differences in 'utility' are thus created through alternative-varying coefficients. Ranking of alternatives is then with respect to a base category whose 'utility' is set to zero. Our base category are South African technicons and universities. Furthermore, we fix the scale by using the

Table 2: Variables by model equation

Variables	Equation (1) Ability (ν) measurement eq.	Equation (2). PhD univ. cat. (D) selection eq.	Equation (3) Rating (R) outcome eq.
Endogenous			
Ability (ν)	-	x	x
PhD univ. cat (D)	-	-	x
Rating (R)	-	-	-
Exogenous	$\boldsymbol{V}$	$\boldsymbol{Z}$	$\boldsymbol{X}$
Age at master	x	-	-
Master distinction	x	-	-
Master univ. cat.	x	x	-
PhD period	-	x	-
Race	-	x	x
Gender	-	x	x
Scientific domain	-	x	x
Experience	-	-	x

standard logit formulation with an iid standard extreme value distributed error.

Rating equation Because we let the PhD training effect, β_p, vary freely for each university category, the level of latent rating outcomes is free. Translating from the latent, real-valued, rating to the observed, categorical, rating involves setting thresholds for the latent variable. Thresholds are set endogenously, and for identification we set the first threshold to zero. Thus a latent rating below zero becomes a 'Rating Unsuccessful'. Latent ratings that are positive but close to zero, become a 'C3'. All the remaining 7 thresholds are then estimated freely.

Estimation The structural model constituted by equations 3, 2 and 1 is formulated in classical terms, and given above assumptions, is identified under classical conditions. We estimate the model however within a Bayesian framework, using the Markov Chain Monte Carlo process and the No U-turn sampler.[5] This approach is due to computational efficiency and has been already followed in prior econometric work such as Hansen et al. (2004) for example. The Bayesian framework however necessitates to fix some prior distributions of our parameters and unobserved factors. We use flat, uninformative priors.[6]

[5]The results are not sensitive to the sampler used.

[6]These are the default prior distributions implemented in RStan, and in fact worked best of all those that we examined.

Proper convergence of the MCMC estimation is ensured by following the suggestions of Gelman and Rubin (1992). In particular we run multiple Markov Chains starting from different initial values and verify that the chains do not diverge but instead describe the same distribution (are 'well mixing'). Furthermore we make sure that the potential scale reduction is estimated to be close to one for all scalar estimates. Finally, estimation results of the complete model are found to be reasonably close to single equation estimations.

4 Data

4.1 Sample

We compare foreign-trained with the locally-trained PhDs active in the South African academic community, and use the rated researchers of its National Research Foundation (NRF) as our sample. The NRF (www.nrf.ac.za) is a state agency that has as its mission the promotion of research and the development of national research capacity. One of its key roles is to facilitate the 'rating' of researchers at universities and other public research institutions such as museums. The rating process is similar to the tenure process at North American universities, but it is focused only on research (not teaching or service to the institution), centrally administered (not by the institution) and valid only for a set period (the exact period has changed over the years, but is around five years).

NRF ratings are useful to understand the quality of research in the academic community in South Africa since essentially all active researchers in the South African academic research community are NRF-rated, partly because this renders researchers eligible for NRF grant funding, and partly because ratings are used by institutions, for example to benchmark relative to other institutions or as the basis for promotion decisions.

As part of the rating process, researchers provide the NRF with evidence of research outputs. A specialist review panel selects six (local and foreign) reviewers who are asked to read at least the self-identified most significant papers of the candidate. Each reviewer assesses the research outputs, and both an independent assessor and the specialist review committee consider the referee reports to assign a rating.

NRF ratings are a useful indicator of the quality of researchers in the academic community in South Africa because of the rigour of the review process, the fact that the NRF filings are detailed (including not only extensive personal detail but also information about outputs as diverse as peer-reviewed journal articles, books, conference presentations, patents, policy or technical reports, and publications in the public press) and suffer very little missing information. As part of the application process, researchers provide the NRF with demographic information, including the institution where they have obtained their degrees.

Our sample is based on the complete, digital files of the NRF spanning the years 2002 to 2012. The NRF maintains two rating systems.[7] We include only scientists who obtained PhDs after 1970 because the data suggests that we have a regime shift concerning the decision to go abroad around that year. The third restriction is that only students whose first academic degree observed in the data, i.e. Bachelor or Master, and last master prior to PhD has been obtained in South Africa.This way we ensure that we only consider students with South African education up to the master making the decision to go or not to go abroad. Finally we remove all individuals with missing information from the sample.

The main sample used to obtain the main results is restricted to 1. senior researchers, 2. PhD after 1970, 3. researchers who did their studies until master at a South African institution. In total these are 1189 scientists experiencing 2432 rating events.[8]

4.2 Variables

Our main focus is how being a PhD graduate of a certain university relates to a scientist's subsequent scientific performance.

Ideally, we would like to investigate this relationship for each university individually, but in practice we are forced to create groups of universities in order to combine a sufficient number of observations for the econometric analysis. In total we construct six PhD university categories: three SA university categories and two foreign university categories. We constructed categories before estimation based on the Shanghai ranking (AWRU). In particular for SA universities, this ranking is consistent with other prominent university rankings such as the Times Higher Education ranking. Resulting categories are summarised in Table 3.

SA research universities form the three university categories which are denoted first-tier, second-tier, and third-tier SA university. A fourth category includes all remaining SA universities and technikons. Foreign universities provide a benchmark for the SA university categories. In order to position SA universities within

[7] There are three common ratings: A, B and C; and three "special" ratings: L, Y and P. The latter group is dedicated to junior researchers, the former to senior researchers. Although there is no strict rule, most researchers shift from junior to senior status within five to ten years of obtaining their PhD. Our analysis focusses on senior researchers for simplicity and because we are mostly interested in the long-term effects of PhD training.

[8] The sensitivity section provides further estimation results on variations of the sample.

[9] 2011

[10] Dropped out of the AWRU in 2008, though it appears in earlier years.

[11] 4 British, 1 Canadian, 2 Japanese, 1 Swiss, 22 US universities

Table 3: University categories

University category	University	Shanghai ranking[9]
First tier SA university	University of Cape Town	201-300
	Witwatersrand University	201-300
Second tier SA university	University of Stellenbosch	401-500
	University of KwaZulu-Natal	401-500
Third tier SA university	University of Pretoria	–[10]
	Rhodes University	–
	University of Johannesburg	–
SA university/technikon	Remainder SA category	
First tier foreign university	30 foreign universities[11]	1-30
Foreign university	Remainder category	

the distribution of foreign universities, we need at least two foreign university categories. We limit ourselves to two categories, in order to keep the number of 'destinations' for PhD students in our model as small as possible. The group of 1st-tier foreign universities consists of all universities ranked among the top 30 universities in the 2011 Shanghai ranking. The rest enters the remainder group, 'other' foreign universities.[12]

Scientific performance is measured using the **NRF ratings**. The rating system categorises researchers according to three broad categories, i.e. from C (established researchers), over B (internationally acclaimed researcher), to A (leading international researcher). Obtaining even a C-rating is considered a success, since ratings may also be unsuccessful. We therefore consider an unsuccessful rating as a separate, lowest category.

Figure 1 shows the percentage of ratings falling in each broader category, starting from 'Rating unsuccessful' over 'C', 'B', up to 'A', for each university category where PhD has been obtained. Across university categories, the majority of scientists obtain a C-rating (60 to 70 percent), B-ratings are relatively common (10 to 30 percent), and A-ratings are rarely awarded (1 to 10 percent).

In general the chance of obtaining a higher rating increases with having a PhD from a university of higher reputation. The ordering is strict for SA university

[12]Out of 1189 scientists in the sample 155 scientists obtained a foreign PhD. Nearly all students with foreign PhD obtained their degree at a Shanghai ranked university (86%). Roughly one third (45 students) at a university ranked among the best 30 institutions, one third ranked between positions 30 and 100, and the remaining between 100 up to 500 (the last Shanghai rank). Main results and statistics are based on including the top 30 universities of the Shanghai ranking into the 1st-tier foreign university category; all of which having a very high international reputation for both research and education. The Sensitivity section presents results from variations of the definition of 1st-tier foreign university.

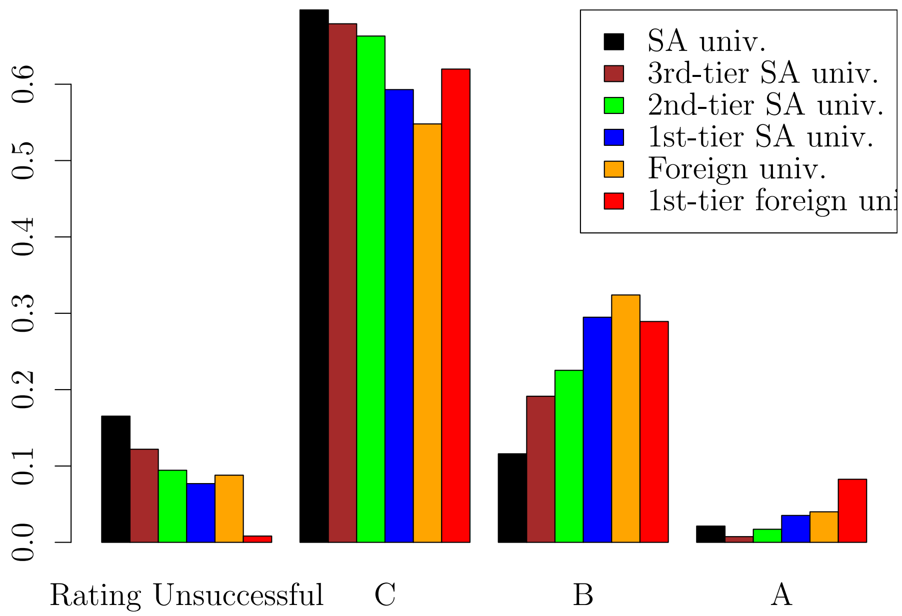


Figure 1: Rating shares by PhD university category. Based on all rating observations. The four bars for each university type should sum to one.

categories. For example in Figure 1, a B-rating is obtained by about 10 percent of the graduates from 'other' SA universities, 20 percent of third-tier SA university graduates, 25 percent of second-tier university graduates and 30 percent of first-tier university graduates. As expected, graduates from top-tier foreign institutions outperform all others.

The econometric model is estimated on fine rating categories as outcomes of the performance equation, Eq. 3. The fine ratings further distinguish within C, B, and A categories by excellence and internationality to arrive at the following categories: Rating Unsuccessful, C3, C2, C1, B3, B2, B1, A2, and A1. In the Appendix, Table 11 provides the relation between PhD university category and rating on this more detailed level. Finer ratings follow the same pattern as the broader ratings.

If the rating of established scientists would coincide with the quality of their PhD training, the analysis could stop here. However, the ordering of scientific achievements by PhD university category may result not only from training effects, i.e. differences in the quality of PhD training, but also from selection effects, i.e. heterogeneity of students' scientific ability before their PhD training.

There is some support for the selection effect explanation. For example, first-tier foreign universities attract in particular young students who obtained their master degree with distinction.[13] It is also the case that researchers with distinction at the master level tend to have higher NRF ratings, suggesting that ability matters both for PhD location and for future ratings.

Whereas age and distinction at master might well serve as signals of scientific ability on the basis of which a (foreign) university can select it's PhD students, it seems unlikely that these characteristics exert a strong influence on an international expert committee which judges scientific excellence of an established researcher twenty years later. Consequently, we allow age at master and master distinction to be indicators of ability in the measurement equation. **Age at master** is the age when the first master degree was obtained. For easier interpretation this variable is transformed in the estimation such that we count years in decades from age twenty. The youngest master graduate being twenty one hence obtains a value of 0.1. We might expect the coefficient to be negative since, all else equal, we expect more able people to obtain their degrees earlier.

As a further signal of ability after master we indicate **master distinction**, i.e. whether or not the master degree has been obtained with distinction; (roughly half masters degrees in our sample were awarded distinction). Neither age at master nor master distinction is assumed to influence directly subsequent selection or performance equations.[14]

We also assume that students who obtain masters degrees from the same university will have similar post-Master ability. As for the PhD it seems reasonable that students' human capital is clustered by universities due to both selection and training effects. Master university categories are constructed exactly as the PhD university categories. **Master university category** enters the PhD university selection equation, capturing some of the differences in the benefit and costs of

[13]Overall, fifty percent of students obtained their master degree between 25 and 30 years (interquartile range) with a median age of 27 years. PhD students at first-tier foreign universities obtained their master between 24 and 26 years with a median age of 26 years. Furthermore, two third of PhD students at first-tier foreign universities received a master degree with distinction, compared to an average of about fifty percent in the sample. Differences in distinction do not seem too large for other categories.

[14]Clearly, distinction in a master's degree does enter the deliberations of a PhD admissions committee, but we assume that it enters as a signal of innate ability and so should enter our ability equation. Age might influence the student's decision of PhD university category. Older students are more likely to have a family for example and therefore less willing to leave the country than younger students. The year when the PhD decision has been made is however not available, and this decision is not necessarily made directly after master. The best available proxy is the age when PhD degree has been obtained. Age at master is significantly correlated with, age at PhD (with a correlation coefficient of about 0.5). Introducing that variable into the PhD selection equation however is doubtful because the estimated coefficient of age at (end of) PhD is particularly strong and significant for top-tier foreign universities but not for other foreign universities. This signals that age at PhD is either a selection criterion or an outcome of the PhD program. Therefore we do not introduce this variable.

PhD university choices across master students.

Table 4 shows the transition from master university to PhD university category for focal South African students. Each row in the table captures students from one type of Master institution, a cell in a row indicates where students from that type of Master institution did their PhDs. The transition pattern is relatively ordered: master students from SA universities with higher reputation are more likely to go abroad to foreign universities of high reputation. For example seven percent of master students from an SA top-tier university did their PhD at a top-tier foreign university (fourth row, sixth column), whereas only four percent (five percent) of students obtaining their master from 2nd- (3rd-) tier SA universities dis so (sixth column). Such a pattern provides a loading of the master university category factor in the ability equation.

Table 4: Transition from master to PhD university category, frequency (row percentage).

	PhD SA university				PhD foreign university		Total
	Others	3rd-tier	2nd-tier	1st-tier	Others	1st-tier	
Others	314 (0.73)	49 (0.11)	23 (0.05)	11 (0.03)	30 (0.07)	6 (0.01)	433 (1)
3rd-tier	39 (0.15)	179 (0.68)	11 (0.04)	11 (0.04)	14 (0.05)	11 (0.04)	265 (1)
2nd-tier	26 (0.09)	17 (0.06)	178 (0.62)	18 (0.06)	36 (0.12)	14 (0.05)	289 (1)
1st-tier	5 (0.02)	15 (0.07)	19 (0.09)	119 (0.59)	30 (0.15)	14 (0.07)	202 (0.99)
Total	384 (0.32)	260 (0.22)	231 (0.19)	159 (0.13)	110 (0.09)	45 (0.04)	1189 (0.99)

A further pattern in Table 4 is that most master students stay at 'their' university (category) for the PhD. There are several explanations. One relates again to ability, in that there may be a match between the ability of a student obtaining a master from a certain university with the requirements of that university regarding the ability of their future PhDs. Further explanations have nothing to do with scientific ability as such. Firstly, master studies orient the student to subjects treated within the faculty and thereby create a thematic fit. Secondly, uncertainty regarding both the doctoral program and the student may be reduced due to the master experience. Finally, students may have established favourable social and economic conditions during their master studies which increases their opportunity costs of choosing another university. Therefore, 'Master university category' enters also the PhD university selection equation, capturing some of the differences in the benefit and costs of PhD university choices across master students.

South Africa suffered an "academic boycott" between the 1960s and 1990. Precisely what its effects were is debated,[15] but we do observe that the distribution of where scientists received their PhDs does vary over time. There may be many causal processes beneath this variation, among which the academic boycott, but

[15]See for example Sha, 1986.

in general we might expect a temporal effect on the decision to go abroad (or the propensity of foreign universities to select South African students). Unfortunately, data limitations prevent us from estimating time trends of selection and training effects. As a result, our estimations provide an intuition on the selection and training effects that took place during the last forty years in average. Accordingly the variable **Period of PhD**, dummies indicating the five year interval in which the PhD has been obtained, enters the selection equation but is excluded from the rating equation, satisfying the exclusion restriction to aid identification.

Experience captures the post PhD experience of the scientist at the time of rating. We measure experience with a set of dummies indicating whether the rating takes place five to nine, ten to fourteen, etc. years after PhD. We find Experience to have a strong positive relationship with fine rating outcomes. Clearly, increasing seniority allows for higher potential scientific achievements. Since time passes for everybody, it is an exogenous variable.

Finally, **Race**, **Gender**, and **Scientific Domain** are three control variables that enter the PhD university selection equation as well as the rating performance equation.[16]

5 Main Result

Deatiled discussion of the model and coefficient estimates is presented in sections 5.2 and 5.3. Before turing to those details, though, we discuss how well the model fits the observations. We compare predicted values and observed values for two variables: where students did their PhD training; and, given where the PhD was granted, their rating outcomes.

5.1 Model fit

Before we turn to estimates of specific coefficients, we assess how the model fits the data more generally. Table 5 compares the fitted probability of selecting a PhD university category with the observed choice. For example, of the students who received their PhDs from "other SA" (type 1) universities, 63 percent were predicted by the model to attend other SA; 14 percent were predicted to attend third-tier SA, and so on. What is striking here is that the looking at any column, which captures where the model predicts students will go, the heaviest entry is on the diagonal — when the model predicts that a student goes to a university of type x, this is in fact where we are most likely to see him.

In fact, the probability assigned by the model, of attending a particular PhD university (type) is throughout highest for those which did in fact graduate from

[16] Out of the 1189 scientists, most are white (1026), male (800), in natural sciences (703).

Table 5: Fitted probabilities of PhD university category, averaged over observed PhD university. To be read as: of those observed to graduate from type 1, (other SA), 63 % were predicted to graduate from type 1, 14 % from type 2, 9% from type three etcetera.

	Fitted Probability					
Observed choice	1	2	3	4	5	6
Other SA (1)	**0.63**	0.14	0.09	0.04	0.07	0.03
3rd-tier SA (2)	0.21	**0.50**	0.09	0.07	0.08	0.04
2nd-tier SA (3)	0.15	0.12	**0.50**	0.08	0.10	0.04
1st-tier SA (4)	0.13	0.12	0.13	**0.48**	0.09	0.05
Other foreign (5)	0.24	0.16	0.24	0.17	**0.14**	0.05
1st-tier foreign (6)	0.15	0.22	0.23	0.20	0.12	**0.07**

that type. In particular for the SA PhD university categories, we find a high probability mass of around 50 percent on the categories which in fact have been observed (see values on the diagonal of Table 5). Fitted probabilities to attend a foreign university category are highest for those that actually went abroad (seen by reading down the last two columns.)

Table 6 compares for each group of PhD university category graduates the actual and fitted average rating outcomes. Fitted shares follow the observed ones relatively closely.[17] χ^2 tests are not able to reject the hypotheses that observed and fitted ratings stem from the same distributions, except for rating outcomes of 'other SA university' PhDs.[18]

5.2 Regression results

Table 7 shows the coefficient estimates of the model. In addition, the first two lines of Table 7 provide two further statistics derived from the estimates: average ability and average training for each PhD university category — capturing the selection and training effects in latent ratings outcomes.

[17] The (slight) over- and under-fitting in some cases can be traced back to the restrictions which we impose on the distribution of the latent rating outcome (R^*). In particular, we model the mean of latent rating outcomes and not the variance and skewness; be it directly through the error term in the rating function or indirectly through the ability equation. We experimented with expanding the model in that direction, but relaxing the distributional assumption comes at the cost of weaker identification, lower estimation precision, and potential over-fitting which we found to be too high.

[18] The χ^2-statistics are calculated as follows: $\chi^2 = \sum_i \frac{N(p_i - \hat{p}_i)^2}{\hat{p}_i}$, where the sum is over ratings i, p_i and $\hat{p}_i$ denote observed and fitted probabilities respectively, and N denotes the number of ratings observed for the respective PhD group. The degrees of freedom of the χ^2 distribution is three in all cases.

Table 6: Average rating outcome probability for PhD university categories with χ^2 statistics

	R.U.	C	B	A	χ^2	p-value
Other SA university (N= 898)						
Observed	0.1653	0.6973	0.1160	0.0213		
Fitted	0.1632	0.6910	0.1353	0.0105	10.6251	0.0139
3rd tier SA university (N= 652)						
Observed	0.1220	0.6792	0.1914	0.0075		
Fitted	0.1177	0.6775	0.1903	0.0145	1.8965	0.5942
2nd tier SA university (N= 580)						
Observed	0.0944	0.6631	0.2253	0.0172		
Fitted	0.0885	0.6660	0.2265	0.0189	0.2693	0.9657
1st tier SA university (N= 433)						
Observed	0.0769	0.5929	0.2949	0.0353		
Fitted	0.0652	0.6065	0.2935	0.0348	0.7583	0.8594
Other foreign university (N= 234)						
Observed	0.0880	0.5480	0.3240	0.0400		
Fitted	0.0792	0.5962	0.2881	0.0364	2.4237	0.4892
1st tier foreign university (N= 97)						
Observed	0.0083	0.6198	0.2893	0.0826		
Fitted	0.0288	0.5428	0.3613	0.0671	5.2727	0.1529

Average ability is the expected ability of PhD students in each university category. Universities of higher reputation in general are found to attract more able students. In South Africa, average ability is estimated to be lowest at ‘other SA universities’ and highest at ‘1st-tier SA universities’. ‘Other foreign universities’ attract students of same quality as the best SA universities (average ability of 0.4 and 0.42, respectively). ‘1st-tier foreign universities’ host the most able students, as they have an average ability which is significantly higher than that of any other university. Overall, the estimates of ability over universities is relatively precise, though not always significant.

Average training is calculated as the average over experience, our time- and university-dependent intercept in the rating equation. We can observe that the quality of the PhD granting university does have a direct effect on the rating a scientist receives. What is striking, however is that there is not a complete dominance by foreign over South African universities. While we observe that highest ratings are indeed associated with the best (foreign) universities, we also observe that the best-considered South African universities are better than a generic foreign university (estimates of 3.99 and 3.79, respectively). Estimates of average training are relatively close given their standard errors. Thus, while estimated average training provides a reasonable order of universities, there may be considerable variation at the student level within each university type.

Table 7: Estimation results

	SA university				Foreign university	
	other	3rd-tier	2nd-tier	1st-tier	other	1st-tier
	Selection and training effects in latent rating (R^*) for white, male, nat. sci. researchers					
Avg. ability	-0.34 (0.122)	-0.01 (0.192)	0.16 (0.248)	0.42 (0.257)	0.4 (0.214)	0.88 (0.293)
Avg. training	3.26 (0.192)	3.57 (0.265)	3.69 (0.33)	3.99 (0.341)	3.79 (0.332)	4.09 (0.492)
	Ability equation (γ)					
Master university	0	0.23 (0.184)	0.31 (0.236)	1.07 (0.238)	–	–
Master distinction	0.68 (0.103)	0.68	0.68	0.68	–	–
Age at master	-0.73 (0.113)	-0.73	-0.73	-0.73	–	–
	Selection equation (α_p, ζ_p)					
Inertia value	2.32 (0.181)	1.99 (0.2)	2.38 (0.207)	2.94 (0.285)	–	–
1970-75	0	0.06 (0.568)	-0.89 (0.596)	-0.91 (0.714)	1.02 (0.491)	-0.42 (0.708)
1975-79	0	1.1 (0.396)	-0.29 (0.491)	-1.11 (0.679)	0.72 (0.448)	-0.5 (0.679)
1980-84	0	0.26 (0.399)	-0.71 (0.432)	-0.37 (0.453)	-1.29 (0.725)	-1.42 (0.788)
1985-89	0	0.16 (0.324)	-0.38 (0.373)	-0.19 (0.366)	-1.08 (0.498)	-1.12 (0.513)
1990-94	0	-0.19 (0.341)	-0.46 (0.342)	-0.36 (0.355)	-0.89 (0.402)	-0.67 (0.449)
1995-99	0	0.09 (0.267)	-0.82 (0.323)	-0.81 (0.341)	-0.15 (0.313)	-1.48 (0.507)
2000	0	0.2 (0.325)	-0.25 (0.35)	-0.27 (0.36)	0.18 (0.314)	-0.84 (0.457)
Non-white	0	0.05 (0.369)	0.94 (0.329)	0.3 (0.364)	1.14 (0.322)	1.03 (0.47)
Female	0	0.31 (0.217)	0.01 (0.274)	0.01 (0.293)	-0.03 (0.269)	-1.16 (0.521)
Social sciences	0	-0.57 (0.224)	-0.11 (0.228)	-1.08 (0.285)	-0.04 (0.249)	-0.19 (0.38)
ζ	0	0.08 (0.137)	0.31 (0.154)	0.02 (0.183)	0.38 (0.175)	0.67 (0.226)
	Rating equation (β)					
05-09 years experience	1.82 (0.22)	1.78 (0.261)	2.29 (0.355)	2.32 (0.35)	2.1 (0.45)	2.4 (0.564)
10-14 years experience	2.46 (0.204)	2.56 (0.278)	3.28 (0.344)	3.33 (0.366)	2.82 (0.419)	2.8 (0.575)
15-19 years experience	2.94 (0.223)	3.38 (0.297)	3.39 (0.36)	4.06 (0.39)	3.84 (0.43)	3.99 (0.585)
20-24 years experience	3.32 (0.24)	3.8 (0.302)	3.85 (0.35)	4.21 (0.406)	3.72 (0.402)	4.26 (0.734)
25-29 years experience	3.72 (0.314)	3.89 (0.348)	4.28 (0.412)	4.24 (0.527)	4.49 (0.542)	5 (0.73)
30-34 years experience	4.2 (0.402)	4.77 (0.462)	4.41 (0.527)	4.53 (0.633)	4.51 (0.53)	5.28 (0.834)
35+ years experience	4.37 (0.478)	4.81 (0.685)	4.34 (0.691)	5.22 (0.813)	5.07 (0.536)	4.91 (0.906)
Non-white	-0.65 (0.303)	-0.62 (0.6)	-0.78 (0.328)	-1.41 (0.396)	-0.39 (0.407)	0.08 (0.547)
Female	-0.32 (0.187)	-0.31 (0.216)	-0.25 (0.252)	-0.38 (0.304)	-1.16 (0.391)	-0.58 (0.82)
Social sciences	-0.3 (0.173)	-0.72 (0.221)	-0.48 (0.221)	-0.09 (0.275)	0.26 (0.386)	-0.09 (0.543)
	Rating equation hurdles (c)					
C3	C2	C1	B3	B2	B1	A
0	1.74 (0.072)	3.78 (0.093)	4.85 (0.107)	5.65 (0.113)	6.85 (0.134)	8.4 (0.202)
	No. individuals: 1189		No. observations: 2432		log-Lik: -5531.78	

Notes: i) Std.dev in (). Coeff. without std.dev. are fixed across universities or fixed to zero
ii) – if coefficient not applicable.

We turn now to the estimates of coefficients in the three equations of the model. Section 'ability equation' in Table 7 shows that ability, at least as judged by PhD admissions committees, is particularly high for young students who received a distinction on their masters degree. These effects are fixed over university categories. Although coefficients are positive for 3rd and 2nd tier master universities, as might be expected, they are not significantly different from zero due to the relatively high standard errors. First tier SA universities issue (signals of) ability significantly higher than all other SA university categories.

Section 'Selection equation' in Table 7 contains the regression results pertaining to the PhD selection equation. The first column provides the variables, and each subsequent column corresponds to one type of PhD university. Coefficients of factors that vary only across individuals and not over alternatives are set to zero for the baseline category 'other SA university'.

In detail, inertia value contains four dummy variables which indicate whether the PhD university category in question was the same as the student's master university category. The tendency of master students to stay at 'their' university (type) for their PhD is reflected in the positive and significant coefficient estimates for all four SA university categories. These coefficients can be seen as capturing the value of inertia in PhD university choice.[19]

PhD period has some influence on the probability that a student is selected into a particular class of PhD university, but changes from one period to the next are not overly strong, when considering the standard deviation of coefficients. Female students are less likely to enter 1st-tier foreign universities, and black students obtained their degree relatively often from second tier SA universities or abroad. Social science PhDs are less sought at third and first tier SA universities.

As expected, in general the aggression with which universities examine ability at the end of the Master degree increases as we read the estimates for ζ from left to right on the bottom row. Roughly speaking this accords with the intuition that higher ranked universities will apply stricter standards for PhD admission. The curiosity in this row is that 1st-tier SA universities obtain a coefficient that is much smaller than those below them in status, insignificant and close to zero. This however does not mean that 1st-tier SA universities fall short in talented master students, as they simply recruit (perhaps "excessively") their own master students. Average ability, the first statistic presented in the table, in combination with Inertia value, suggests that the strong tendency of first-tier SA master students to stay for PhD offsets the selection factor ζ.[20]

The two subsequent parts of Table 7 pertain to the rating equation. The

[19]Recall that the sample includes only students who completed their Masters degrees in South Africa, so there can be no inertia for foreign universities.

[20]In the sensitivity analysis below, restricting the sample to those who changed universities between Master and Doctorate increases this coefficient to its "natural" place in the order.

coefficient estimates of the rating equation, β, represents the estimation of the real-valued, latent R^*. The bottom panel gives the threshold values that convert from R^* to the letter rating given in the raw data.

To interpret the β coefficients it is necessary to make reference to the thresholds c. For example: the coefficient of five to nine years of experience after PhD for students from an 'other SA university' is 1.82. Ignoring individual ability, we would expect a 'C2' rating (since $C2 = 1.74 < 1.82 < 3.78 = C1$). It takes an expected 30 years of experience for that scientist to obtain a C1 rating. In contrast, a scientist holding a PhD from a first tier SA university can be expected to obtain a C1 within 15 years — half the time.

Taking into account individual ability, which adds to training effects at each rating event, further changes expectations: PhD graduates from 'other SA university' would never be expected to obtain a B3 rating (4.37 (35+ years of experience) − 0.34 (avg. ability) $= 4.03 < 4.85 = B3$), while PhDs from '1st-tier university' would be expected to obtain a B3 rating or higher after 15 years of experience (3.99 ($15 - 19$ years of experience) + 0.88 (avg. ability) $= 4.87 > 4.85 = B3$).

Overall the pattern of experience effects seems coherent. Reading coefficients down the columns, we find for any PhD university type that experience increases the likelihood of getting a better rating. Reading the coefficients along the line, we find that doing a PhD from universities of higher reputation is advantageous for subsequent career achievements due to training (independent of ability). Differences however are relatively small and uncertain overall. The overall pattern is well characterised by the average training statistic in the second row of Table 7.

Finally, being female reduces one's latent rating, as does being non-white. The racial effect has in general a higher significance level than the gender effect, neither being statistically significant at standard levels.

5.3 Training and selection effects

Speaking loosely, a scientist's rating is affected by his or her innate ability and his or her training. Where ability stops and training starts is in general an open question. We have defined ability as measured at the end of a Master degree. Training then refers to training received during the course of completing a PhD. In this section we identify the magnitudes of those two effects separately.

The econometric model identifies treatment and selection effects through the joint estimation of the parameters β and ζ respectively. Given the model and data, we obtain the conditional probability of obtaining a certain rating for each individual:

$$P(R_{i,t} = j | X_{i,t}, \nu_i, \beta_p),$$

where X is the vector of controls, ν_i individual ability, and β_p training effects of the university of type p.

The probability is conditioned on both ability and training (among other things), so taking one marginal distribution (conditional on either ability or training) will give us the distribution of one or the other effects. The expected value of that marginal distribution will be the average size of the effect.[21]

Put another way, what we refer to as the training effect captures the following thought experiment. Send a student randomly chosen from the population to different universities to do a PhD. Different universities give different training. Depending on which university he or she goes to, we would make different predictions regarding his or her rating. Table 8 shows these probabilities (estimations, with confidence intervals beneath each estimate). Each column represents the probability distribution of rating outcomes given that a randomly chosen student attends a particular type of PhD institution. Thus each column in principle sums to one.

Students most likely to receive a "rating unsuccessful" (R.U.) are those who receive their PhDs from a generic South African university/technikon or a third tier SA institution. Least likely are those who go to a top university, be it in South African or abroad.

Reading down any column reveals a non-monotonic pattern. In every column the peak is at rating level C2. This is simply because in the data this is the most common rating. Reading across reveals that the higher ratings (A and B) are more likely to be achieved by those who went to Foreign top universities. However again we see evidence that "foreign" does not in any simple sense dominate "South African". High ratings are more likely to be achieved by those whose PhDs are from top South African universities than those receiving them from generic foreign universities. Again university quality rather than location is what matters.

The sorting parameter ζ creates differences in expected ratings conditional on PhD university ex ante, or net of, the treatment. As for the rating effect, the selection effect corresponds to the expected value of the marginal distribution, but

[21] More formally, the training effect is defined here as

$$P(R = j|D = p) = 1/N \sum_{i \in N} 1/T_i \sum_{t \in T_i} P(R_{i,t} = j|\boldsymbol{X}_{i,t}, \nu_i, \boldsymbol{\beta}_p).$$

We define a selection effect similar to the training effect, but this time we integrate out the training effect. The selection effect enters through the fact that the distribution of unobserved ability ν, $\Phi(\nu_p)$, can be expected to vary across PhD universities.

$$P(R = j|\Phi(\nu_p)) = 1/N \sum_{i \in N} 1/T_i \sum_{t \in T_i} P(R_{i,t} = j|\boldsymbol{X}_{i,t}, \nu_p, \boldsymbol{\beta}_s)\ ,$$

where ν_p is normally distributed with mean average ability at university p, and empirical standard deviation of ability in the population. Training $\boldsymbol{\beta}_s$ is from a randomly drawn university. Averages are obtained in each simulation/estimation iteration within MCMC.

Table 8: Training effect estimates, 90 percent confidence interval in ()

	PhD SA university				PhD foreign university	
	Others	3rd-tier	2nd-tier	1st-tier	Others	1st-tier
RU	0.128	0.131	0.091	0.086	0.105	0.086
	(0.109, 0.148)	(0.102, 0.159)	(0.069, 0.116)	(0.062, 0.107)	(0.072, 0.136)	(0.045, 0.134)
C3	0.216	0.206	0.176	0.159	0.179	0.16
	(0.196, 0.237)	(0.187, 0.223)	(0.148, 0.204)	(0.134, 0.183)	(0.152, 0.207)	(0.116, 0.195)
C2	0.331	0.318	0.326	0.309	0.309	0.301
	(0.318, 0.346)	(0.304, 0.333)	(0.31, 0.342)	(0.291, 0.328)	(0.286, 0.328)	(0.26, 0.329)
C1	0.135	0.135	0.155	0.158	0.147	0.153
	(0.121, 0.148)	(0.118, 0.151)	(0.135, 0.174)	(0.142, 0.179)	(0.128, 0.164)	(0.127, 0.173)
B3	0.072	0.075	0.089	0.096	0.087	0.094
	(0.06, 0.084)	(0.061, 0.086)	(0.075, 0.105)	(0.083, 0.112)	(0.073, 0.105)	(0.077, 0.118)
B2	0.067	0.072	0.088	0.099	0.09	0.101
	(0.055, 0.079)	(0.062, 0.086)	(0.07, 0.108)	(0.084, 0.116)	(0.073, 0.111)	(0.077, 0.131)
B1	0.037	0.043	0.052	0.063	0.057	0.069
	(0.028, 0.047)	(0.034, 0.052)	(0.039, 0.068)	(0.046, 0.084)	(0.044, 0.071)	(0.046, 0.095)
A	0.015	0.019	0.022	0.029	0.027	0.036
	(0.01, 0.02)	(0.012, 0.026)	(0.016, 0.03)	(0.019, 0.038)	(0.018, 0.037)	(0.021, 0.055)

this time we integrate out the training effect of PhD universities.

The question we answer here is: What would I achieve with having just any kind of training but the brain of a Harvard/other foreign/first tier SA/... student?

Selection effects displayed in Table 9 follow the same pattern as training effects. For example, "unsuccessful ratings" are, again, most likely for PhD students that obtained their PhD from the base category 'other' SA universities. This time however, as probabilities are based on 'average training effects', so the reason for the pattern lies in the lower estimated ability of the students selected into those universities. Overall, selection is estimated to introduce considerable heterogeneity of scientific performance across PhD university categories.

First tier foreign universities attract the brightest students. For the lowest and higher ratings (R.U. C3, B, A) this result is also statistically significant: the 90 percent confidence intervals of first-tier foreign university students so not overlap with the intervals of the generic or third tier SA universities.

A further noticeable result is how the ranking of 'other' foreign university changes for the selection effect compared to the training effects. While the training effects of other foreign university PhD students have been approximately on the same level as for *2nd tier* SA universities, selection effects of other foreign university PhD students compare rather to the *1st tier* SA universities. Thus, although 'other' foreign universities offer a training which compares to that of a 2nd tier SA university, average foreign universities tend to attract students which are of the same innate ability level as 1st tier SA university students.

Table 9: Selection effect estimates, 90 percent confidence interval in ()

	PhD SA university				PhD foreign university	
	Others	3rd-tier	2nd-tier	1st-tier	Others	1st-tier
RU	0.145	0.116	0.105	0.086	0.09	0.061
	(0.113, 0.181)	(0.092, 0.141)	(0.084, 0.129)	(0.063, 0.107)	(0.072, 0.108)	(0.045, 0.079)
C3	0.217	0.195	0.185	0.164	0.169	0.134
	(0.194, 0.239)	(0.172, 0.219)	(0.162, 0.204)	(0.14, 0.186)	(0.148, 0.189)	(0.113, 0.154)
C2	0.326	0.326	0.324	0.316	0.319	0.296
	(0.311, 0.34)	(0.311, 0.339)	(0.309, 0.34)	(0.299, 0.335)	(0.3, 0.336)	(0.275, 0.314)
C1	0.134	0.146	0.152	0.16	0.158	0.17
	(0.117, 0.152)	(0.127, 0.163)	(0.137, 0.167)	(0.142, 0.175)	(0.142, 0.173)	(0.153, 0.186)
B3	0.071	0.082	0.087	0.096	0.094	0.109
	(0.059, 0.087)	(0.068, 0.096)	(0.073, 0.101)	(0.082, 0.113)	(0.082, 0.109)	(0.095, 0.124)
B2	0.063	0.077	0.083	0.097	0.093	0.118
	(0.051, 0.079)	(0.062, 0.097)	(0.068, 0.099)	(0.078, 0.113)	(0.079, 0.109)	(0.097, 0.141)
B1	0.033	0.042	0.046	0.057	0.054	0.076
	(0.024, 0.042)	(0.031, 0.055)	(0.035, 0.057)	(0.046, 0.07)	(0.041, 0.067)	(0.059, 0.098)
A	0.012	0.016	0.018	0.024	0.022	0.035
	(0.008, 0.016)	(0.012, 0.021)	(0.012, 0.027)	(0.017, 0.031)	(0.016, 0.031)	(0.023, 0.049)

Finally, comparing tables 9 and 8 we observe very similar patterns. Examining any row (B3 for example) in the two tables we observe that the magnitudes of the coefficients are very similar. In addition, the pattern across the row, which indicates the relative value of training or selection at a the different university types, is also very similar in the two tables. The same is true if we compare two columns in the two tables. What this suggests is that the effect of a foreign PhD on rating arises from both selection and from training, and that the magnitudes of the two effects are roughly equal: selection is as important as training.

6 Sensitivity Analysis

This section provides a short discussion on the validity of our results in light of additional analyses. A more detailed discussion along with result tables of analyses is provided in the Appendix of this working paper.

First, we conducted a **Monte Carlo analysis** to verify that the model, as implemented, produces unbiased estimates given the true data generation process (dgp) of simulated data corresponds to that assumed by the model. One particular restriction introduced into the model was to fix the scale of latent, individual ability relative to the error term in the latent rating equation. While identification in theory is possible without this restriction, it was needed in practice for identification and convergence. The Monte Carlo analysis shows that when the relative influence in the true dgp differs from that assumed in estimation, a bias is created.

When the relative importance of ability in fact is higher than assumed, treatment effects will be over- and selection effects under-estimated.

Therefore, we estimated several **alternative models**. Estimating the same model as in the main text with various scalings of ability in the rating equation, we observe that the likelihood improves with increasing scales of ability. This indicates considerable individual-level heterogeneity in research outcomes beyond that captured by the restricted model. However, results get extreme in that there is the tendency to attribute all achievements to individual (signals of) ability and nothing to training. One reason the model is not able to separate ability from training may be that the training effect of a university actually depends on the students' ability.

We therefore test an alternative model which allows for such an ability dependent training effect: the idea is that the PhD training actually transforms the students' ability as signalled before the PhD, into scientific competence after PhD. For each university, we model training as a projection of three levels of ex-ante ability — low, medium, and high ability — onto scientific competence which influences subsequent ratings. Selection of ability remains as in the main model. Since levels of competence are freely estimated, the relative scale of ability and the error in the rating equation is no longer an issue. This advantage comes at the cost of assuming a common shape of the career trajectory subsequent to PhD. The training effect is then in the level of the trajectory given by the competence acquired through training of ones (ex-ante) ability. Estimation results suggest that ability and training are indeed interdependent. Yet, although this alternative model differs considerably in spirit from our main model, they produce the same pattern and magnitudes of training and selection effects. Therefore, we consider the main model as estimated in the text as a useful simplification which is robust to alternative frameworks.

Finally, we tried variations on the **sample and coding of variables** which we estimated on the model in the main text. The more noticeable checks are discussed in the following; always focussing on how the alternative estimates support or speak against our interpretation of the main model estimates.

Our estimates provide training and selection averaged over a longer time period; from 1970 to roughly the 2000s. Yet, training and selection evolves over time, and given South Africa's turbulent recent history probably in a non-continuous fashion. The number of observations is not sufficient for estimation of reliable time trends or period effects. Therefore we varied the sample of scientists by the period of PhD obtained to investigate potential cohort effects within the limits naturally imposed by our focus on long-term career achievements.

Changing the sample by including all scientists which obtained a PhD either after 1960 or after 1975 has very little impact on estimation results. One reason is of course that the sample changes only slightly, for example the 1975 restriction excludes about five percent of scientists and ten percent of rating events from the sample. Further restrictions of the sample, such as PhD obtained after 1980, however keeps very few rating events with scientists of an experience of 25 years or more — the period when higher ratings are typically achieved. Using the sample of scientists with PhD obtained after 1980 we obtain coefficients which remain close to those of the main results (within one standard deviation). The order and magnitude of average ability remains the same as in the main estimation. Estimations for experience, our period-varying intercepts driving training effects, are less favourable though for foreign universities and improve for South African universities with respect to the first 20 years of experience. Training estimates should however not be emphasised too much since there are signs of over-fitting in particular for foreign universities.

Given that (relative) training quality changes over time, one might wonder whether Period of PhD is rightly excluded from the rating equation. In order to test for the relevance of the exclusion of Period of Phd for the results, we removed this variable from the selection equation. This had a very small effect on our estimates because identification rests mainly on the ability equation, rather than the exogenous shifter 'Period of Phd' in the selection equation. Without the ability equation, i.e. estimating a more traditional factor structure model, training and selection are indeed not separated; all ex-post heterogeneity is attributed to training and nothing to selection. Thus, the validity of our results regarding the sorting of ability into university types rests on the assumption that age at master, distinction at master, and master university influences PhD university selection and training only through (signals of) ability (except for a general tendency to stay at a master university for PhD for reasons different from ability).

The assumed separation of master from PhD studies is institutionalised in South Africa but may not hold always in practice. Yet, this assumption does seem correct for master students who switched university from master to PhD. To check, we re-do the analysis on a restricted sample: including only those who hswitched universities between Master and Doctorate. Thisleaves 482 scientists experiencing 979 rating events (which is relatively few observations given the complexity of our model). Observations remain the same for those going abroad for a PhD, and we also estimate ability and training to be the same as in the main estimation for these students. In South African universities however average ability decreases by more than one standard error. This drop is consistent with the idea that students of higher (lower) ability than their peers tend to switch to higher (lower) quality institutions for their PhD. Training at generic and third tier South African universities is estimated to be considerably higher (more than one standard deviation), but we would not generalise this result as there are signs of over-fitting of period-specific intercepts.

What exactly should be considered a top tier institution is somehow an arbitrary decision. We therefore varied the categorisation of top foreign universities to include only the top 20 Shanghai ranked universities as well as the top 50 Shanghai ranked universities. In either case changing this definition has no effect on the estimates of South African universities. Considering the top 20 Shanghai ranked universities as top foreign universities, we note that average ability increases slightly for generic foreign universities and by about one standard deviation for top foreign universities. On the other hand, we observe an (insignificant) increase of training for generic foreign universities and a decrease for top foreign universities. One explanation might be that competition among high quality institutions is more on selection than on training. Considering the top 50 Shanghai ranked universities tends to decrease both selection and training for top foreign universities. Generic foreign and top foreign universities become more similar in both aspects. This is reasonable given that nearly all foreign universities in our data are Shanghai listed and about one half below rank 50.

Which insights remain from the main analysis given the sensitivity checks? Two statements found broad support throughout: Firstly, more able students tend to do a PhD at universities of higher reputation. This sorting by ability explains heterogeneity of career achievements as much or more than differences in PhD training across universities. That universities of a strong, world-wide reputation attract the most able students found ample support. Whether generic foreign universities attract more able students than the best SA universities might be subject for debate. Secondly, SA universities do not necessarily provide worse training than foreign universities in Europe or US. In some estimation results, estimated training effects of foreign universities appear unreasonably low. This might be caused by our sample which includes only scientists returning from a foreign PhD, creating a negative bias. Therefore, we need to interpret our results from the perspective of the developing country: sending students abroad rather than supporting local PhD programs does not necessarily imply a human capital gain.

7 Discussion

Scholarship about the role of universities in development has been stalled as scholars struggled to disentangle the functioning of the manifold local and global institutions and influences shaping indigenous universities. In this paper, we deliberately choose a narrow focus, namely the academic success (as measured in their publication record) of scholars working in developing countries who have completed their doctorate at least five year prior. In particular, we consider the human capital that is developed in foreign versus local universities in a developing country context.

Using a unique dataset, we have been able to differentiate between the training and selection effects of receiving a PhD from different institutions. Our econometric model is an extension of a model that has often been used in the evaluation literature (notably in the schooling effects literature where it was introduced by Heckman) in order to account for a confounding selection effect: the factor structure model with a selection and outcome equation, where unobserved individual characteristics (typically interpreted as individual ability) are factored into a selection and an outcome equation. We expand this model by an additional measurement equation which allows for including proxies of individual ability which is not possible in the more traditional approach. Table 10 displays our results in stylized form.

		Local universities		**Foreign universities**
Effect of PhD **selection** on subsequent career success	1			Top foreign universities
	2	1st tier local universities	=	All other foreign universities
	3	2nd tier local universities		
	4	3rd tier local universities		
	5	All other local universities		
Effect of PhD **training** on subsequent career success	1	1st tier local universities	=	Top foreign universities
	2	2nd tier local universities	=	All other foreign universities
	3	3rd tier local universities		
	4	All other local universities		

Table 10: Summary of results: ranking effects of selection and training.

We find that the quality of the PhD granting institution is indeed correlated with future career success as an academic in South Africa. However, our evidence suggests that the fairly common view that foreignness equates to quality does not capture the nuances of university training in a developing country, at least not in terms of the quantity and quality of research produced by academics. As can be expected, the perception that foreign universities offer superior training has a strong effect in terms of the (self)selection of students. Students are particularly attracted to the universities that would be found at the top of any quality ranking — Oxford, Harvard, Stanford and so on — but this can perhaps be expected. Students from advanced economies also strongly prefer to study at those institutions. However, even the foreign universities that are not as highly ranked exert a very strong selection effect. Given that study abroad, even when financial support is offered, is expensive and disruptive, especially for students with partners and families, this finding is in alignment with the view that indigenous universities offer a less desirable PhD training.

However, once selection effects are controlled for, we also find very clear evidence that the leading local universities are "world-class" in the training that they offer. The evidence of the high quality of scholarly training at a number of the local universities may appear counter-intuitive at first; the understanding has long been that universities in developing countries are likely to develop "technicians" of research, rather than scholars who can meaningfully advance their field. However, to the extent that a university offers training in the fundamental skills needed to do meaningful work in the field, it is plausible that scholars from those universities could produce research that is deemed interesting by their peers globally. In an increasingly globally connected world the challenges faced by developing country universities are likely to attract more attention than before. Scholars from those universities may be able not only to join the informal global networks where knowledge in their fields is advanced, but also find that their particular contextual perspective is deemed to be of value to the field.

Indeed, from the perspective of career prospects, a PhD from the top South African university is more desirable than one from a non-top foreign university, where "top" foreign refers to the roughly thirty top-ranked international universities. However, the selection effect is very strong, and the training effect, while significant, determines less of the future career trajectory than might reasonably have been expected.

We believe that our work makes important contributions to both theory and practice. In terms of theory, debates in the past have often built on a vertical division between developing and developed countries. A strong reason is of course that universities tend to be associated at the national level, in terms of aspects like funding, the regulatory context and eventually general reputation. National boundaries do exist. However, a better understanding may be obtained from doing a horizontal division, separating first and second-tier universities. Our evidence suggests that this is a more appropriate division and therefore serves better to understand the development process.

Practically speaking, "foreign" is not necessarily "better", which implies that the PhD programmes eligible for financial assistance for doing the PhD abroad should be carefully considered. Not all foreign programmes are worthy of support, at least no more worthy of support than their domestic equivalents. Naturally this observation applies not only at the ministry or university level, but also at the student level.

Our evidence also suggests that rating systems do tap into some dimension of academic quality. The NRF data that we used maps individual rather than institutional academic performance; indeed, in the uneven post-Apartheid academic landscape, one of the main purposes of the NRF is to provide a national (rather than institutional) benchmark of academic quality. However, there was nonetheless correlation between individual and institutional performance. The University of Cape Town is consistently the top-rated South African university in the various global rankings, with the University of Witwatersrand and Stellenbosch University appearing in those rankings, but much lower down. The Tier 3 South African universities all appear in at least one of the three commonly-used ranking systems (Shanghai, QS and Times Higher Education), but the methodological differences between the rankings result in substantial variability in the performance of those universities. The other South African universities do not appear on any rankings. This hierarchy was also evident in our results. As academic administrators and policymakers battle with the issue of how much credence to give to various academic rankings, our findings suggest that they do offer useful guidance about academic quality, especially once a threshold level of quality has been obtained.

The strength of the selection effect in our results permits some optimism regarding possibilities for universities in emerging economies to catch up with their advanced economy competitors. Some local universities are good, and improved support (e.g. visa regulations to support researcher mobility and increased financing) might allow the best universities in emerging economies to become recognized worldwide for the quality of science and scientific training conducted there. The top global universities will always attract the best students, but as the country and its education system gains confidence in its (best) universities, the best local universities should be able to attract better students than the competent but non-leading foreign ones.

The study is not without limitations. In terms of methodology, three are worth mentioning. First, we only observe PhD students that stayed in South Africa or came back to South Africa. One could imagine that the most promising recipients of PhDs from South African universities could compete on the international job market, and leave the country to pursue their careers abroad. While perceptions about the quality of PhD training in developing countries make such scenarios less likely, in those cases the training and selection effect of South African universities would be underestimated. By contrast, if only scientists who did not "make it" in the US return to South Africa, then the training and selection effects of foreign universities are underestimated. These dimensions cannot be controlled for with the current data.

Second, the disentanglement of selection from training effects is only as good as the instrument and observed ability measures. Our main instruments, "period of PhD" and the ability measures "Master institution" and "distinction in Master degree", are good predictors of university selection but of course less than perfect. Thus it is possible that treatment effects in particular of foreign top tier universities are somewhat overestimated. This is an issue shared with any study of our kind and little can be done to formally address it. Overall, however, our results are very robust, and there is no evidence that the highest ranked universities of the world do not have a positive effect on their bright students.

Thirdly, we have assumed some stability in the system. That is, the effects we are measuring are assumed not to change over the period of our sample. Particularly given the history of South Africa, which has included some fairly serious structural reforms of universities post-Apartheid, this is not strictly true. Some of the historical evolution will be picked up in the "period of PhD" variable, but not all. Unfortunately the data are not strong enough to include this possible instability in the analysis.

In terms of theoretical limitations, although we believe that the paper gains from our narrower theoretical focus, there are costs to such an approach. The first relates to the fact that the paper does not actually engage with the role of indigenous universities in economic or even technological development. The evidence is very clear that universities contribute to upgrading not simply through the quality of academic scholarship, but through linkages with other institutions and in particular industry (Bernardes and Albuquerque, 2003; Brundenius, Lundvall and Sutz, 2011; Mazzoleni and Nelson, 2007; Mowery and Rosenberg, 1999; Murmann, 2003; Rosenberg and Nelson, 1994). We provide evidence of one building block in the process of upgrading — the quality of local science — but recognize the continued importance of an integrative approach to the role of universities in upgrading.

For example, in addition to the selection effects of foreign versus local universities, a local selection effect also seems likely. The top local universities are likely to benefit from a virtuous cycle where the top local students who do not wish (or are not able) to go abroad will prefer study at the leading local institutions. Policymakers may respond to local differentiation by nurturing a small group of select institutions so that they indeed become globally acknowledged for the quality of scholarship conducted there, or may try to increase the number of high quality local institutions by spreading financial and other support more widely. The likely developmental implications of these choices are very different, and may well centre around the opportunities for linkages offered by different institutions rather than on pure academic success. Further research is needed to understand how development is best served.

A related limitation is that this paper considers only the development of human capital. But social capital is an arguably equally important outcome of academic training. In addition to training in the social technologies of a field, PhD training allows scholars to develop relationships that can provide them with collaborators or at least access to the informal scientific networks in a field. It seems likely that PhD training abroad provides scholars with a richer social network than local PhD training, which would increase the value of foreign training. We believe that further research is needed into how local versus foreign PhD training shapes the social capital of researchers.

8 Conclusion

This paper contributes to scholarship about the role of universities in development by comparing PhD training in the developed world to that in an emerging economy, South Africa. We focus on a narrow outcome measure, the scientific achievements of South African academics, which requires of us to sacrifice a general understanding of how universities from developing countries engage with their underdeveloped context in order to isolate whether the quality of academic scholarship is shaped by the type of university where the PhD has been obtained. Looking at four tiers of South African and two tiers of foreign universities, we find that better scholars indeed emerge from better universities.

This may be because good students select (and are selected by) good universities, or because of the better training they receive there. Our model allows us to examine both selection and training effects, and we find that universities from industrialized countries are preferred over local universities. Indeed, our estimates show that pure selection effects contribute to career outcomes nearly as much as training effects. But universities from industrialized countries do not necessarily provide better training than do local universities. When looking at training rather than selection, PhDs from top South African universities produce a similar quantity and quality research output to those from leading universities in the developed world.

This finding allows some confidence about the role of local universities in the upgrading of developing countries. It suggests that an investment in the local science system is likely to result not only in the development of fairly basic local skills, as was previously believed, but can also result in cutting-edge science. While there is value to foreign training, countries need to consider carefully to which foreign universities they send PhD students, as not all foreign PhD training is equally useful. Indeed, from a theoretical perspective, our study suggests that the discussion about the role of universities in upgrading may be better served by differentiating universities in terms of the roles they serve in their economy rather than along the better established “local” and “foreign” dimension. The boundary experienced for students and scientists between Oxford and the University of Cape Town may well be more permeable than the boundaries between the University of Buxtehude and Oxford.

References

[1] Aakvik, A., Heckman, J. J., and Vytlacil, E. J. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, 125(1), 15-51.

[2] Almeida, P., Phene, A., and Li, S. (2014). The Influence of Ethnic Community Knowledge on Indian Inventor Innovativeness. Organization Science, 26(1), 198-217.

[3] Altbach, P. G. (2013). Advancing the national and global knowledge economy: the role of research universities in developing countries. Studies in higher education, 38(3), 316-330.

[4] Athreye, S., and Cantwell, J. (2007). Creating competition?: Globalisation and the emergence of new technology producers. Research Policy, 36(2), 209-226.

[5] Becker, G. S. (1962). Investment in human capital: A theoretical analysis. The journal of political economy, 9-49.

[6] Bernardes, A. T., and Albuquerque, E. D. M. (2003). Cross-over, thresholds, and interactions between science and technology: lessons for less-developed countries. Research Policy, 32(5), 865-885.

[7] Brundenius, C., Lundvall, B. ., and Sutz, J. (2011). The role of universities in innovation systems in developing countries: developmental university systems?empirical, analytical and normative perspectives. Handbook of Innovation Systems and Developing Countries: Building Domestic Capabilities in a Global Setting, 311-336.

[8] Chen, K., and Kenney, M. (2007). Universities/research institutes and regional innovation systems: the cases of Beijing and Shenzhen. World development, 35(6), 1056-1074.

[9] Coleman, J. S. (1988). Social capital in the creation of human capital. American journal of sociology, S95-S120.

[10] Conceio, P., and Heitor, M. V. (1999). On the role of the university in the knowledge economy. Science and Public Policy, 26(1), 37-51.

[11] Corbett, A., Cornelissen, J., Delios, A., and Harley, B. (2014). Variety, novelty, and perceptions of scholarship in research on management and organizations: An appeal for ambidextrous scholarship. Journal of Management Studies, 51(1), 3-18.

[12] D?Este, P., and Patel, P. (2007). University?industry linkages in the UK: What are the factors underlying the variety of interactions with industry? Research policy, 36(9), 1295-1313.

[13] Deiaco, E., Hughes, A., and McKelvey, M. (2012). Universities as strategic actors in the knowledge economy. Cambridge Journal of Economics, 36(3), 525-541.

[14] Department of Education (1998) "Formal Degree/Diploma/Certificate Programme Classification Structure Manual", Manual SAPSE-004 (95/10), 2nd edition, Department of Education, Pretoria, Republic of South Africa.

[15] Dos Santos, T. 1970, "The Structure of Dependence." *American Economic Review, Papers and Proceedings*, pp, 231-2360.

[16] Du Toit, Andre (2012) "The PhD and the degree structure of South African higher education: A brief and rough guide", Memorandum presented at the 'Center for Higher Education Transformation'-seminar 'Knowledge Production in South African Higher Education' on 23 February 2012, accessed 29 April at http://www.chet.org.za/files/PhD%20Andre%20du%20Toit.pdf.

[17] Etzkowitz, H. (2003). Research groups as ?quasi-firms?: the invention of the entrepreneurial university. Research policy, 32(1), 109-121.

[18] Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457-472.

[19] Gngr, N. D., and Tansel, A. (2008). Brain drain from Turkey: an investigation of students? return intentions. Applied Economics, 40(23), 3069-3087.

[20] Hansen, K. T., Heckman, J. J., and Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1), 39-98.

[21] James J. Heckman, J. J., Lopes, H. F. and Piatek, R. (2014). Treatment Effects: A Bayesian Perspective. *Econometric Reviews*, 33:1-4, 36-67.

[22] Hyslop, J., Vally, S., and Hassim S. (2006) "The South African Boycott Experience", *Academe* 92(5):59-70.

[23] Jaffe, A., Trajtenberg, M., Henderson, R. (1993) “Geographic localization of knowledge spillovers as evidenced by patent citations”, *Quarterly Journal of Economics*, 63, 577-598.

[24] Kumar, Nagesh, 1985, “Cost of Technology Imports: The Indian Experience,” *Economic and Political Weekly*, 20 (35), M103-M114,

[25] Lall, S. (2001). Competitiveness, Technology and Skills. Edward Elgar Publishing, Williston, VT.

[26] Lee, S. Y., and Song, X. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences.* John Wiley & Sons.

[27] Lee, W-Y. (2000). The role of science and technology policy in Korea?s industrial development. In: Kim, L., and Nelson, R. R. (eds). Technology, learning, and innovation: Experiences of newly industrializing economies. Cambridge University Press.

[28] Link, A. N., and Scott, J. T. (2005). Opening the ivory tower's door: An analysis of the determinants of the formation of US university spin-off companies. Research Policy, 34(7), 1106-1112.

[29] Luiz, J. M. (2015). The impact of ethno-linguistic fractionalization on cultural measures: Dynamics, endogeneity and modernization. Journal of International Business Studies.

[30] Mathews, J. A., and Hu, M. C. (2007). Enhancing the role of universities in building national innovative capacity in Asia: the case of Taiwan. World Development, 35(6), 1005-1020.

[31] Mazzoleni, R., and Nelson, R. R. (2007). Public research institutions and economic catch-up. Research policy, 36(10), 1512-1528.

[32] Merton, Robert K. (1968). “The Matthew Effect in Science”. *Science* 159 (3810), 56?63.

[33] Mowery, D. C., and Rosenberg, N. (1999). Paths of innovation: Technological change in 20th-century America. Cambridge University Press.

[34] Murmann, J. P. (2003). Knowledge and competitive advantage: The coevolution of firms, technology, and national institutions. Cambridge University Press.

[35] Narula, R., and Dunning, J. H. (2000). Industrial development, globalization and multinational enterprises: new realities for developing countries. Oxford development studies, 28(2), 141-167.

[36] Narula, R., and Dunning, J. H. (2010). Multinational enterprises, development and globalization: some clarifications and a research agenda. Oxford Development Studies, 38(3), 263-287.

[37] Nelson, R. (2004). The challenge of building an effective innovation system for catch-up. Oxford Development Studies, 32(3), 365-374.

[38] Nelson, R. (2005). The roles of research in universities and public labs in economic catch-up. In: Santangelo, G.D. (Ed.), Technological Change and Economic Catch-Up. Edward Elgar, Cheltenham, UK, 19-32.

[39] Nelson, R., (2005) The roles of research in universities and public labs in economic catch-up. In: Santangelo, G.D. (Ed.), Technological Change and Economic Catch-Up. Edward Elgar, Cheltenham, UK, pp. 1932.

[40] Olds, K. (2007). Global assemblage: Singapore, foreign universities, and the construction of a ?global education hub?. World Development, 35(6), 959-975.

[41] Ozawa, T. (1992). Foreign direct investment and economic development. Transnational Corporations, 1(1), 27-54.

[42] Rosenberg, N. (1992). Scientific instrumentation and university research. Research Policy, 21(4), 381-390.

[43] Rosenberg, N., and Nelson, R. R. (1994). American universities and technical advance in industry. Research policy, 23(3), 323-348.

[44] Salter, A. J., and Martin, B. R. (2001). The economic benefits of publicly funded basic research: a critical review. Research policy, 30(3), 509-532.

[45] Saxenian, A. (2007). The new argonauts: Regional advantage in a global economy. Harvard University Press.

[46] Saxenian, A., and Hsu, J. Y. (2001). The Silicon Valley?Hsinchu connection: technical communities and industrial upgrading. Industrial and corporate change, 10(4), 893-920.

[47] Scott, J. C. (2006). The mission of the university: Medieval to postmodern transformations. The journal of higher education, 77(1), 1-39.

[48] Song, H. (1997). From brain drain to reverse brain drain: Three decades of Korean experience. Science Technology and Society, 2(2), 317-345.

[49] Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

[50] Zeschky, M., Widenmayer, B., and Gassmann, O. (2011). Frugal innovation in emerging markets. Research-Technology Management, 54(4), 38-45.

A Descriptive statistics

The cross-tabulation of rating outcomes by PhD university category is shown in Table 11. This is the equivalent of Figure 1, but using the fine-grained ratings. Taken as such, the top foreign universities are clearly outperforming all others since their PhD graduates achieved relatively many A and B ratings. Similarly, the typical SA university produced proportionally less Bs and As than other university categories. With respect to the remainder categories, the pecking order becomes less clear. Whereas the top-tier SA university and two foreign university have A PhD graduates, the second-tier SA university gathers the highest percentage of B PhD graduates. For estimations we combine the rarely achieved 'A1' and 'A2' ratings into a single rating category 'A'.

Table 11: Rating outcomes by PhD university category

		SA University			Foreign University		
	Other	third-tier	2nd-tier	first-tier	foreign	first-tier	Total
R.U.	124 (0.17)	65 (0.12)	44 (0.09)	24 (0.08)	22 (0.09)	1 (0.01)	280 (0.12)
C3	192 (0.26)	114 (0.21)	85 (0.18)	38 (0.12)	44 (0.18)	12 (0.1)	485 (0.2)
C2	255 (0.34)	173 (0.32)	152 (0.33)	96 (0.31)	63 (0.25)	37 (0.31)	776 (0.32)
C1	76 (0.1)	75 (0.14)	72 (0.15)	51 (0.16)	30 (0.12)	26 (0.21)	330 (0.14)
B3	31 (0.04)	43 (0.08)	35 (0.08)	33 (0.11)	28 (0.11)	15 (0.12)	185 (0.08)
B2	33 (0.04)	31 (0.06)	44 (0.09)	41 (0.13)	34 (0.14)	10 (0.08)	193 (0.08)
B1	23 (0.03)	28 (0.05)	26 (0.06)	18 (0.06)	19 (0.08)	10 (0.08)	124 (0.05)
A2	14 (0.02)	2 (0)	8 (0.02)	9 (0.03)	10 (0.04)	7 (0.06)	50 (0.02)
A1	2 (0)	2 (0)	0 (0)	2 (0.01)	0 (0)	3 (0.02)	9 (0)
Total	750 (1)	533 (0.98)	466 (1)	312 (1.01)	250 (1.01)	121 (0.99)	2432 (1.01)

Table 12 provides the number of students which obtained their PhD in a given period for any of the six university categories. The share of foreign PhDs has been relatively low during the 80s, when the academic boycott was most severe. The share of foreign PhDs recovered during the 90s for the generic foreign university category but not for top tier foreign universities. The share of PhD graduates from

SA universities has been relatively stable over time, except for a steady relative decay at third tier SA universities.

Table 12: Period of PhD by PhD University Category

	PhD SA university				PhD foreign university		Total
Period of PhD	Others	3rd-tier	2nd-tier	1st-tier	Others	1st-tier	
1970-75	15 (0.26)	8 (0.14)	9 (0.16)	6 (0.1)	15 (0.26)	5 (0.09)	58 (1.01)
1975-79	24 (0.3)	21 (0.26)	15 (0.19)	5 (0.06)	11 (0.14)	4 (0.05)	80 (1)
1980-84	31 (0.33)	27 (0.29)	17 (0.18)	12 (0.13)	3 (0.03)	3 (0.03)	93 (0.99)
1985-89	51 (0.3)	43 (0.25)	41 (0.24)	24 (0.14)	6 (0.04)	6 (0.04)	171 (1.01)
1990-94	79 (0.36)	46 (0.21)	38 (0.18)	33 (0.15)	10 (0.05)	11 (0.05)	217 (1)
1995-99	97 (0.35)	57 (0.2)	55 (0.2)	35 (0.12)	30 (0.11)	7 (0.02)	281 (1)
2000+	87 (0.3)	58 (0.2)	56 (0.19)	44 (0.15)	35 (0.12)	9 (0.03)	289 (0.99)
Total	384 (0.32)	260 (0.22)	231 (0.19)	159 (0.13)	110 (0.09)	45 (0.04)	1189 (0.99)

B Monte Carlo analysis

In theory the set-up of the econometric model presented in the main text allows for varying freely the relative influence of 'innate ability' and the (fixed-scale) error term on rating. In practice, for identification on the observed data, we fix the scale of innate ability. The following Monte Carlo analysis shows that this approach corresponds potentially to a wrong assumption about the true data generation process which may result in biased coefficient and treatment effect estimates. This result prompts us to robustify our main results in the subsequent section C, 'Alternative Models'.

Table 13 summarises the settings for data simulation in both studies. The equations to generate the data correspond exactly to those of the econometric model used for estimation, except for the additional coefficient τ used to vary the scale of ability ν in the rating equation. In the first MC study (MC 1), we generate data with $\tau = 1$. The second MC study (MC 2) generates data with $\tau = 2$. Estimation results are obtained in both cases on the same restricted econometric model with $\tau = 1$. Except for the τ coefficient, all parameters are identical for both MC studies. Generated data is similar in size and complexity to our observed data.

Our discussion of the Monte Carlo results focusses on the rating equation (Table 14) and ATE (Table 15).

We first turn to the rating equation results in Table 14. In both Monte Carlo studies, MC 1 and MC 2, true β_1 is a university specific intercept in the rating equation increasing from 1 for university 1 to 5 for university 5. Since $\beta_2 = 1$ for all universities, β_1 is the sole source of the training effect in the simulated dgp.

Table 13: Settings for Monte Carlo simulations

Equations	
$\nu_i = V_i + \epsilon_{i,\nu}$ (ability eq.)	
$\boldsymbol{D}_i^* = \boldsymbol{W}_i + \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{\zeta}\nu_i + \boldsymbol{\epsilon}_{i,D}$ (selection eq.)	
$R_{i,t}^* = \boldsymbol{X}_{i,t}\boldsymbol{\beta}\, I_{D_i=p} + \tau\nu_i + \epsilon_{i,t,R}$ (rating eq.)	
Coefficients	
$\tau = 1$ in MC 1, $\tau = 2$ in MC 2	
$\boldsymbol{\alpha} = \begin{pmatrix} 0\,1\,2\,3\,4 \\ 0\,1\,1\,1\,1 \end{pmatrix}$, $\boldsymbol{\zeta}^t = \begin{pmatrix} 0\;0\;0.5\;1.0\;1.5 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} 1\,2\,3\,4\,5 \\ 1\,1\,1\,1\,1 \end{pmatrix}$, $\boldsymbol{c} = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7 \end{pmatrix}$	
Indices	
S	1000 MC simulation runs
N	1000 individuals
NT	2000 rating observations (2 ratings per individual)
Q	2 factors in ability equation
P	5 universities
L	2 factors in selection equation (indiv.-varying)
K	2 factors in rating equation
Regressors	
V_i	scalar $N(0,1)$
$\boldsymbol{W}_i$	Vector with $P = 5$ iid $N(0,1)$ distr. entries
$\boldsymbol{Z}_i$	Vector with $L = 2$ iid $N(0,1)$ distr. entries
$\boldsymbol{X}_{i,t}$	Vector with intercept and $N(0,1)$ scalar (K=2)
Error terms	
$\epsilon_{i,\nu}$	$N(0,1)$ in ability equation
$\boldsymbol{\epsilon}_{i,D}$	P-vector, iid extr. value distr. entries (multinomial-logit)
$\epsilon_{i,R}$	iid extr. value distr. (ordered logit)

Table 14: Rating coefficients in Monte Carlo studies

	MC 1 ($\tau = 1$)					MC 2 ($\tau = 2$)				
	Uni. 1	Uni. 2	Uni. 3	Uni. 4	Uni. 5	Uni. 1	Uni. 2	Uni. 3	Uni. 4	Uni. 5
True										
β_1	1.000	2.000	3.000	4.000	5.000	1.000	2.000	3.000	4.000	5.000
β_2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\tau E[\nu]$	-0.398	-0.634	-0.170	0.279	0.696	-0.799	-1.277	-0.350	0.561	1.396
Mean										
β_1	0.994	1.991	3.029	3.962	4.996	0.815	1.646	2.926	4.163	5.417
β_1	1.008	1.006	1.001	1.003	1.008	0.997	1.005	0.997	1.004	0.997
$E[\nu]$	-0.391	-0.620	-0.190	0.338	0.726	-0.556	-0.880	-0.239	0.435	1.020
SD										
β_1	0.116	0.145	0.177	0.181	0.160	0.148	0.183	0.212	0.212	0.190
β_2	0.098	0.117	0.129	0.117	0.092	0.126	0.154	0.167	0.152	0.121
$E[\nu]$	0.785	0.763	0.748	0.744	0.760	1.034	1.015	0.987	0.986	1.014
Bias										
β_1	0.006	0.009	-0.029	0.038	0.004	0.185	0.354	0.074	-0.163	-0.417
β_2	-0.008	-0.006	-0.001	-0.003	-0.008	0.003	-0.005	0.003	-0.004	0.003
$E[\nu]$	-0.007	-0.014	0.019	-0.059	-0.030	-0.243	-0.397	-0.111	0.126	0.376
RMSE										
β_1	0.119	0.145	0.183	0.185	0.162	0.246	0.406	0.245	0.284	0.461
β_2	0.099	0.116	0.128	0.121	0.093	0.125	0.149	0.163	0.148	0.122
$E[\nu]$	0.067	0.086	0.112	0.120	0.073	0.272	0.425	0.212	0.209	0.393

The selection effect in the simulation manifests itself in the scaled expected ability, $\tau E[\nu]$, conditional on the university. The econometric model restricts $\tau = 1$ and, hence, only allows for unscaled ability, $E[\nu]$. In MC 1 $\tau = 1$, and $\tau E[\nu]$ ranges between -0.634 for University 2 and 0.696 for University 5.[22] Since ν is scaled by $\tau = 2$ in MC 2, $\tau E[\nu]$ is twice as much.

The estimates averaged over MC simulations (Mean), MCMC standard deviations averaged over MC simulations (SD), the difference between true value and mean estimates (Bias), and the root mean squared error of estimates to true values over MC simulations (RMSE) are used to judge the performance of the econometric model.

In the first Monte Carlo study, MC 1 (left side of Table 14) estimated coefficients β and estimated expected ability ν are close to their true values. Also the SD of β within single estimations reflects well the variation of the estimates around the true value (RMSE). SD of $E[\nu]$ differs from RMSE of $E[\nu]$ because the former gives the average variation of individual abilities within a single estimation while the latter gives the variation of average ability across estimations. Overall, the model performs very well in MC 1.

[22] True $E[\nu]$ is calculated as the average of simulated ν_i within university categories and across simulation runs.

Table 15: ATE in Monte Carlo studies

	MC 1 ($\tau = 1$)					MC 2 ($\tau = 2$)				
Ratings	Uni. 1	Uni. 2	Uni. 3	Uni. 4	Uni. 5	Uni. 1	Uni. 2	Uni. 3	Uni. 4	Uni. 5
True										
1	0.341	0.207	0.112	0.054	0.024	0.341	0.207	0.112	0.054	0.024
2	0.160	0.134	0.095	0.057	0.030	0.160	0.134	0.095	0.057	0.030
3	0.160	0.160	0.134	0.095	0.057	0.160	0.160	0.134	0.095	0.057
4	0.134	0.160	0.160	0.134	0.095	0.134	0.160	0.160	0.134	0.095
5	0.095	0.134	0.160	0.160	0.134	0.095	0.134	0.160	0.160	0.134
6	0.057	0.095	0.134	0.160	0.160	0.057	0.095	0.134	0.160	0.160
7	0.030	0.057	0.095	0.134	0.160	0.030	0.057	0.095	0.134	0.160
8	0.014	0.030	0.057	0.095	0.134	0.014	0.030	0.057	0.095	0.134
9	0.010	0.024	0.054	0.112	0.206	0.010	0.024	0.054	0.111	0.206
Bias										
10	0.001	-0.000	0.002	-0.002	-0.000	-0.061	-0.104	-0.081	-0.055	-0.032
11	0.000	0.000	0.002	-0.001	-0.000	0.041	0.022	0.004	-0.007	-0.010
12	-0.000	0.000	0.002	-0.001	-0.000	0.043	0.041	0.026	0.009	-0.002
13	0.000	0.001	0.001	-0.001	0.000	0.029	0.044	0.041	0.029	0.013
14	-0.001	-0.000	-0.000	-0.000	-0.000	0.009	0.032	0.042	0.043	0.032
15	-0.000	-0.000	-0.001	0.000	-0.000	-0.007	0.013	0.027	0.041	0.044
16	-0.000	-0.000	-0.001	0.002	0.000	-0.014	-0.003	0.007	0.024	0.040
17	-0.000	-0.000	-0.001	0.001	-0.000	-0.014	-0.011	-0.009	0.002	0.021
18	-0.000	-0.000	-0.002	0.002	0.001	-0.025	-0.033	-0.058	-0.084	-0.106
RMSE										
19	0.018	0.017	0.014	0.008	0.003	0.064	0.106	0.083	0.057	0.033
20	0.011	0.010	0.009	0.007	0.004	0.042	0.024	0.009	0.010	0.011
21	0.010	0.010	0.009	0.008	0.006	0.044	0.042	0.027	0.011	0.006
22	0.009	0.010	0.010	0.009	0.007	0.030	0.045	0.042	0.030	0.015
23	0.008	0.009	0.010	0.010	0.009	0.011	0.033	0.043	0.043	0.033
24	0.006	0.008	0.010	0.011	0.010	0.009	0.014	0.029	0.042	0.045
25	0.003	0.006	0.009	0.010	0.011	0.015	0.007	0.011	0.026	0.041
26	0.002	0.004	0.007	0.009	0.011	0.014	0.012	0.012	0.009	0.023
27	0.002	0.004	0.009	0.013	0.014	0.025	0.034	0.060	0.086	0.107

The second Monte Carlo study, MC 2 (right side of Table 14), shows that restricting the scale τ to one in the estimation when in the true dgp it is not creates a bias.[23] In this example where $\tau > 1$ the role of the treatment effects parameter β_1 is overestimated and average ability $E[\nu]$ is less relevant than in the true dgp. Hence, the bias of the treatment parameter β_2 is negative for universities one to three and positive for universities four and five. The opposite holds for $E[\nu]$.

[23] In MC 2, Mean and SD are normalised by the average β_2 across universities, 0.784. Bias and RMSE are calculated based on these normalised values. The normalisation allows for a comparison of the estimates to the true values and the relative influence of the intercept, β_1, and average ability $E[\nu]$ in the rating equation. The approach follows the same logic as the normalisation of coefficients to compare results of a logit and a probit model, for example.

Compared to MC 1, the larger bias creates a larger RMSE in MC 2.

Table 15 shows that the bias in the coefficient estimates carries over to the ATE estimates. The true ATE is the same in both simulation studies since the data is simulated with the same true β coefficients (upper part of Table 15). Probabilities of obtaining a certain rating conditional on university are correctly estimated in MC 1 since the bias is small overall. In MC 2 however the shape of the conditional probability mass function tends to be smoother than the true probability mass function. Despite this bias, the relative tendencies across universities are still captured, for example low probability of obtaining a high rating conditional on university 1 and high probability of obtaining a high rating conditional on university 5. The relatively low RMSE, mostly below 5%, supports the idea that the overall pattern estimated is consistent with the true conditional probabilities.

The results of the Monte Carlo analyses prompt us to estimate alternative models; presented in the next section.

C Alternative models

C.1 Traditional Factor Structure Model

An alternative to the model presented in the main text is the more traditional Factor Structure Model (FSM) which excludes the ability equation and instead assumes an unconditional distribution of ability over individuals. We estimate the model using the same Bayesian approach, simply removing the ability equation. We find that FSM identifies the same order of universities by their training effect but fails to separate treatment from selection effects.

Table 16 provides the estimation results. We judge university training effects on the average training effect, i.e. the average (period-specific) intercept in the rating equation for each university. The average training effect suggests the same order between universities as found in the main results, i.e. 1st-tier foreign university (4.86) $\succsim$ 1st-tier SA university (4.48) $\succsim$ 'other' foreign university (4.35) $\succsim$ 2nd-tier SA univ. (3.87) $\succsim$ 3rd-tier SA univ. (3.55) $\succsim$ 'other' SA univ. (3.09). The difference between FSM and main results is in differentiating selection from treatment effects. Estimation results of FSM ascribe most ex-post heterogeneity to treatment effects, while estimated selection effects contribute very little to explanation. Average ability is centred around zero for all universities.

Table 16: Alternative model - FSM

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
	Selection and training effects in latent rating (R^*) for white, male, nat. sci. researchers					
Avg. ability	0.01 (0.099)	0.02 (0.125)	0 (0.136)	-0.02 (0.179)	-0.03 (0.198)	-0.05 (0.305)
Avg. training 3.08 (0.186)	3.56 (0.216)	3.86 (0.238)	4.48 (0.289)	4.34 (0.306)	4.88 (0.429)	
	Ability equation (γ)					
Master university	–	–	–	–	–	–
Master distinction	–	–	–	–	–	–
Age at master	–	–	–	–	–	–
	Selection equation (α_1,α_2,ζ)					
inertia value	–	–	–	–	–	–
1970-75	0	-0.53 (0.457)	-0.48 (0.435)	-0.81 (0.508)	0 (0.383)	-1.14 (0.568)
1975-79	0	0.03 (0.308)	-0.4 (0.34)	-1.45 (0.521)	-0.81 (0.376)	-1.81 (0.589)
1980-84	0	0.11 (0.282)	-0.47 (0.323)	-0.63 (0.364)	-2.48 (0.678)	-2.3 (0.692)
1985-89	0	0.07 (0.229)	-0.05 (0.236)	-0.44 (0.274)	-2.21 (0.47)	-1.93 (0.494)
1990-94	0	-0.29 (0.211)	-0.56 (0.222)	-0.55 (0.236)	-2.12 (0.372)	-1.7 (0.377)
1995-99	0	-0.23 (0.201)	-0.41 (0.208)	-0.72 (0.232)	-1.26 (0.268)	-2.41 (0.455)
2000	0	-0.01 (0.222)	-0.27 (0.232)	-0.33 (0.245)	-1.05 (0.295)	-1.96 (0.466)
Non-white	0	-0.94 (0.323)	0.27 (0.25)	-0.01 (0.291)	0.6 (0.301)	0.39 (0.473)
Female	0	0.11 (0.179)	-0.37 (0.195)	-0.04 (0.211)	-0.14 (0.259)	-1.2 (0.489)
Social sciences	0	-0.6 (0.169)	-0.2 (0.173)	-0.84 (0.204)	-0.07 (0.229)	-0.31 (0.339)
ζ	0	0.02 (0.17)	-0.01 (0.176)	-0.02 (0.212)	-0.04 (0.222)	-0.06 (0.302)
	Rating equation (β)					
05-09 years after phd	1.54 (0.21)	1.77 (0.238)	2.42 (0.275)	2.57 (0.316)	2.65 (0.4)	3.2 (0.546)
10-14 years after phd	2.22 (0.208)	2.55 (0.241)	3.47 (0.268)	3.73 (0.306)	3.39 (0.373)	3.69 (0.475)
15-19 years after phd	2.73 (0.223)	3.34 (0.251)	3.62 (0.28)	4.49 (0.339)	4.32 (0.424)	4.84 (0.518)
20-24 years after phd	3.15 (0.247)	3.81 (0.282)	4 (0.304)	4.71 (0.391)	4.27 (0.412)	5.03 (0.626)
25-29 years after phd	3.53 (0.301)	3.92 (0.322)	4.44 (0.364)	4.72 (0.489)	5.01 (0.499)	5.67 (0.636)
30-34 years after phd	4.04 (0.37)	4.78 (0.429)	4.65 (0.482)	5.18 (0.632)	5.05 (0.532)	6.05 (0.82)
35+ years after phd	4.36 (0.487)	4.72 (0.642)	4.45 (0.605)	5.93 (0.758)	5.66 (0.499)	5.69 (0.9)
Non-white	-1.08 (0.32)	-1.07 (0.519)	-1.18 (0.35)	-1.47 (0.43)	-0.86 (0.404)	-0.39 (0.627)
Female	-0.37 (0.197)	-0.34 (0.231)	-0.16 (0.269)	-0.19 (0.299)	-1.26 (0.413)	-0.54 (0.749)
Social sciences	-0.46 (0.185)	-0.83 (0.234)	-0.62 (0.231)	-0.21 (0.306)	0.17 (0.375)	0.02 (0.535)
	Rating equation hurdles (c)					
C3	C2	C1	B3	B2	B1	A
0	1.69 (0.07)	3.67 (0.09)	4.72 (0.1)	5.51 (0.11)	6.7 (0.131)	8.22 (0.179)
	No. individuals: 1189		No. observations: 2432		log-Lik: -6150.78	

Notes: i) Figures in () are std.dev. of coefficients, avg. ability std.dev. are averaged over individuals,
ii) coefficients without std.dev. are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

This result is caused by a relatively weak exogenous shifter in the selection equation. Our exogenous shifter is 'period of phd obtained'. Indeed, estimates of the FSM model without the exogenous shifter 'period of phd obtained' in the selection equation yields very similar results (no table provided). Therefore we conclude for the FSM estimations that ATE estimates are likely to be biased upwards while ASE estimates are biased downwards.

C.2 Models with ability scaled in the rating equation

The Monte Carlo analysis yields the result that assuming a scale of one for ability in the rating equation could potentially introduce a bias in the coefficient estimates and, consequently, ATE and ASE estimates. Unfortunately, estimations on our data do not converge for freely estimated scaling factors τ. We therefore estimate the model again with different scaling factors (see Appendix 'Monte Carlo simulation' for the corresponding equation).

We estimated the model with a scaling of $\tau = 0.5$ which performs worse considering the log-likelihood but does not change considerably the results. We found that increasing the scaling factor τ tends to increase the log-likelihood of the estimated model. For example increasing the scaling from $\tau = 1$ to $\tau = 2$ improves the log-likelihood from -5532.88 to -4969.1 (see Table 17). The reason for this improvement is that unexplained individual-level heterogeneity in research achievements is actually rather large, and therefore increasing the scale of individual ability, which is akin to introducing more noise in the model, yields a higher log-likelihood.

Increasing the scaling factor τ also tends to increase the role of the selection effect compared to the training effect. Consider for example the results for a scaling of $\tau = 2$ displayed in Table 17. The overall level of (unscaled) average ability and average training is rather similar to those presented in the main estimation results. Yet, (scaled) ability is scaled by two in the rating equation and, hence, twice as influential as before.

The ordering of universities based on the 'average training' statistic changes. In particular '1st-tier foreign universities' are estimated to provide in average a similar training effect as 'other foreign universities' and '2nd-tier SA universities' (both with avg. training of 4.3). Investigating the period varying intercepts of the rating equation, the ordering of universities appears less consistent. For example for the first 5 to 9 years after phd 'other SA universities' are estimated to provide a larger training effect than '1st-tier foreign universities', while the opposite is estimated to hold for the '25-29 years after phd'. This result seems unreasonable.

One explanation for this behaviour may be that training varies not only across universities but also with the students' ability. Thus ex-post individual-level heterogeneity due to training might be loaded on ex-ante heterogeneity. The next section takes up that idea.

Table 17: Estimation results of 'ability scaled 2' model

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
	Selection and training effects in latent rating (R^*) for white, male, nat. sci. researchers					
Avg. ability (scaled)	-0.45 (0.224)	0.16 (0.374)	0.65 (0.484)	0.75 (0.564)	0.96 (0.45)	1.86 (0.625)
Avg. training	4.14 (0.284)	4.32 (0.432)	4.33 (0.568)	4.95 (0.632)	4.46 (0.568)	4.45 (0.77)
	Ability equation (γ)					
Master university	0	0.24 (0.154)	0.38 (0.19)	0.83 (0.223)	–	–
Master distinction	0.45 (0.076)	0.45	0.45	0.45	–	–
Age at master	-0.45 (0.085)	-0.45	-0.45	-0.45	–	–
	Selection equation (α_1,α_2,ζ)					
Inertia value	2.32 (0.194)	2 (0.2)	2.35 (0.208)	2.92 (0.291)	–	–
1970-75	0	0 (0.557)	-0.94 (0.59)	-0.92 (0.649)	1.04 (0.478)	-0.44 (0.71)
1975-79	0	1.02 (0.402)	-0.33 (0.492)	-1.22 (0.666)	0.66 (0.448)	-0.6 (0.696)
1980-84	0	0.27 (0.407)	-0.74 (0.477)	-0.4 (0.508)	-1.34 (0.711)	-1.49 (0.791)
1985-89	0	0.15 (0.341)	-0.39 (0.377)	-0.19 (0.387)	-1.15 (0.523)	-1.19 (0.611)
1990-94	0	-0.21 (0.328)	-0.48 (0.35)	-0.34 (0.356)	-0.93 (0.424)	-0.8 (0.5)
1995-99	0	0.09 (0.296)	-0.82 (0.352)	-0.81 (0.357)	-0.18 (0.33)	-1.59 (0.56)
2000	0	0.17 (0.332)	-0.31 (0.371)	-0.31 (0.38)	0.12 (0.359)	-1.01 (0.555)
Non-white	0	0.05 (0.388)	0.92 (0.344)	0.3 (0.398)	1.15 (0.344)	1.04 (0.513)
Female	0	0.31 (0.238)	0.01 (0.259)	0.01 (0.288)	-0.02 (0.282)	-1.05 (0.506)
Social sciences	0	-0.58 (0.222)	-0.13 (0.235)	-1.1 (0.278)	-0.08 (0.248)	-0.23 (0.364)
ζ	0	0.12 (0.198)	0.39 (0.22)	0.05 (0.263)	0.46 (0.23)	0.77 (0.296)
	Rating equation (β)					
05-09 years after phd	2.19 (0.308)	1.81 (0.447)	2.22 (0.581)	2.7 (0.64)	2.2 (0.633)	1.99 (0.841)
10-14 years after phd	3.05 (0.301)	2.89 (0.449)	3.57 (0.578)	4.02 (0.637)	3 (0.605)	2.83 (0.791)
15-19 years after phd	3.79 (0.311)	4.06 (0.454)	3.9 (0.587)	4.83 (0.651)	4.39 (0.647)	4.32 (0.826)
20-24 years after phd	4.36 (0.335)	4.75 (0.476)	4.52 (0.602)	5.14 (0.686)	4.47 (0.651)	4.75 (0.919)
25-29 years after phd	4.81 (0.382)	4.94 (0.509)	5.17 (0.639)	5.35 (0.764)	5.36 (0.719)	5.6 (0.942)
30-34 years after phd	5.27 (0.459)	5.82 (0.591)	5.55 (0.732)	6.06 (0.899)	5.42 (0.742)	6.04 (1.093)
35+ years after phd	5.53 (0.573)	5.97 (0.808)	5.35 (0.858)	6.54 (1.023)	6.42 (0.719)	5.57 (1.15)
Non-white	-0.68 (0.444)	-0.54 (0.736)	-0.82 (0.485)	-2.06 (0.607)	-0.43 (0.575)	0.43 (0.962)
Female	-0.41 (0.286)	-0.24 (0.335)	-0.25 (0.384)	-0.54 (0.431)	-1.44 (0.586)	-0.68 (1.118)
Social sciences	-0.44 (0.27)	-0.94 (0.34)	-0.61 (0.338)	0.06 (0.448)	0.39 (0.533)	0 (0.782)
	Rating equation hurdles (c)					
C3	C2	C1	B3	B2	B1	A
0	2.22 (0.091)	4.9 (0.12)	6.35 (0.137)	7.47 (0.153)	9.08 (0.182)	10.98 (0.236)
	No. individuals: 1189		No. observations: 2432		log-Lik: -4970.12	

Notes: i) Figures in () are std.dev. of coefficients, avg. ability std.dev. are averaged over individuals,
ii) coefficients without std.dev. are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

C.3 Model with ability dependent treatment effects

The above econometric models build on a questionable assumption. The long-term effect of an individual PhD university on subsequent research achievements is assumed to be invariant to the ability of the student. In the alternative model we allow for a university specific treatment effect that varies with the ability of the students, i.e. $\tau_p(\nu_i)$.

This way, treatment effects are allowed to vary not only for students with same 'initial ability' studying at different universities, but also for students with different 'initial ability' studying at the same university. For example a more able student might benefit more from a demanding phd program than a less able student.

To cap model complexity, we simplify the rating equation by restricting the coefficient vector to be the same across all universities ($\boldsymbol{\beta}_p = \boldsymbol{\beta}$), and by dropping the additive 'innate ability' which is subsumed in the new function $\tau_p(\nu_i)$.

$$R^*_{i,t,p} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \tau_p(\nu_i) + \epsilon_{i,t,p},$$

In the spirit of a non-parametric regression, we do not assume an analytical function for τ but instead estimate a fixed τ value for each university category for different levels of ν. The continuous latent variable ν may fall into $1 \dots q \dots Q$ levels defined by $(Q-1)$ hurdles, and we define:

$$\tau_p(\nu_i) = \tau_{p,q} \text{ if } (q-1) < \nu_i < q,$$

where for $\tau_{p,1}$ the lower limit is $-\infty$ and for $\tau_{p,Q}$ the upper limit is $+\infty$.

We fix the hurdles ex ante to the estimation. Hurdles need to be chosen such that each level of ν at each university has sufficient data support. We explored extensively models with three levels of ν: low ability ($-\infty < \nu \leq 0$), medium ability ($0 < \nu \leq 2$), and high ability ($2 < \nu \leq \infty$).

Given $\tau_{p,q}$, ν_i, and the hurdles q one can easily calculate the rating probabilities $P(R_i = r|X_i, \nu_i)$. In theory, one could directly estimate the model with the transformed rating equation through MCMC. In practice however the estimation becomes unstable because variation in abilities cause a rugged optimisation landscape as individuals cross the hurdles from one MCMC step to the next. We establish a smoother landscape by integrating out the ability ν in the rating equation, taking the current state of ability in the MCMC step as expectation and a variance of one which corresponds to the variance in the original ability equation (Equation 1). The rating probabilities with ability integrated out then reads:

$$P(R_i = r|X_i) = \sum_{q \in Q} P(\nu_i \in q|V_i)\; P(R_i = r|X_i, \nu_i \in q)$$

We turn now to the estimation results displayed in Table 18. We first note that indeed the treatment effect varies with student ability as the treatment factor τ differs significantly over ability levels within the same university category. The order of universities by treatment effect remains largely the same as observed in the prior models. Furthermore, average ability in university categories follows largely the pattern previously observed.

Table 18: Estimation results with alternative model 'ability dependent treatment effects'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
	Selection and training effects in latent rating (R^*) for white, male, nat. sci. researchers.					
Avg. ability (ν)	-0.01 (0.181)	0.48 (0.191)	0.93 (0.278)	1.09 (0.246)	0.94 (0.255)	1.64 (0.31)
$\tau(\nu < 0)$	- 3.18 (0.625)	- 3.08 (0.628)	- 2.94 (0.722)	- 2.33 (0.712)	- 2.65 (0.691)	- 1.5 (0.684)
$\tau(0 < \nu < 2)$	1.05 (0.612)	1.27 (0.647)	1.33 (0.729)	2.24 (0.705)	2.37 (0.8)	2.43 (0.799)
$\tau(\nu > 2)$	4.59 (0.374)	4.53 (0.381)	5.01 (0.387)	5.28 (0.389)	5.28 (0.398)	5.43 (0.419)
	Ability equation (γ)					
Master university	0.36 (0.174)	0.67 (0.22)	1.44 (0.231)	0.91 (0.101)	–	–
Master distinction	0.06 (0.366)	0.06	0.06	0.06	–	–
Age at master	-0.45 (0.177)	-0.45	-0.45	-0.45	–	–
	Selection equation (α_1,α_2,ζ					
inertia value	2.23 (0.208)	1.96 (0.181)	2.35 (0.224)	2.95 (0.268)	0	0
1970-75	0	-0.12 (0.537)	-1.39 (0.629)	-1.03 (0.636)	0.68 (0.493)	-1.17 (0.832)
1975-79	0	0.9 (0.436)	-0.71 (0.544)	-1.27 (0.687)	0.35 (0.51)	-1.29 (0.812)
1980-84	0	0.19 (0.406)	-1.1 (0.518)	-0.5 (0.525)	-1.65 (0.8)	-2.17 (0.886)
1985-89	0	0.08 (0.337)	-0.75 (0.395)	-0.32 (0.381)	-1.42 (0.543)	-1.78 (0.655)
1990-94	0	-0.29 (0.324)	-0.81 (0.428)	-0.45 (0.373)	-1.22 (0.47)	-1.38 (0.589)
1995-99	0	0.05 (0.305)	-1.13 (0.409)	-0.94 (0.373)	-0.43 (0.368)	-2.16 (0.655)
2000	0	0.12 (0.381)	-0.58 (0.449)	-0.35 (0.381)	-0.09 (0.426)	-1.52 (0.714)
Non-white	0	0.1 (0.373)	1.09 (0.376)	0.41 (0.39)	1.31 (0.302)	1.27 (0.531)
Female	0	0.28 (0.24)	-0.02 (0.264)	-0.01 (0.284)	-0.04 (0.283)	-1.11 (0.534)
Social sciences	0	-0.56 (0.225)	-0.07 (0.216)	-1.03 (0.253)	-0.02 (0.243)	-0.17 (0.361)
ζ	0	0.17 (0.107)	0.54 (0.163)	0.18 (0.136)	0.5 (0.164)	0.92 (0.224)
	Rating equation β					
	Years after phd 05-09: 3.11 (0.654)		Years after phd 10-14: 4.17 (0.63)		Years after phd 15-19: 5.14 (0.658)	
	Years after phd 20-24: 5.38 (0.66)		Years after phd 25-29: 5.83 (0.653)		Years after phd 30-34: 6.47 (0.69)	
	Years after phd 35+: 7.04 (0.696)		Non-white: -0.93 (0.277)		Female: -0.78 (0.229)	
	Social sciences -0.46 (0.158)					
	Rating equation hurdles (c)					
C3	C2	C1	B3	B2	B1	A
0	2.37 (0.171)	6.41 (0.432)	8.25 (0.495)	9.43 (0.539)	10.91 (0.58)	12.53 (0.608)
	No. individuals	1189	No. observations	2432	log-Likelihood	-5733.24

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

Table 19: ATE from alternative model 'ability dependent treatment effects'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
RU	0.128	0.123	0.12	0.082	0.098	0.05
	(0.11, 0.145)	(0.103, 0.146)	(0.092, 0.146)	(0.058, 0.113)	(0.068, 0.134)	(0.027, 0.08)
C3	0.202	0.199	0.199	0.177	0.179	0.16
	(0.183, 0.219)	(0.182, 0.215)	(0.182, 0.217)	(0.152, 0.196)	(0.162, 0.2)	(0.131, 0.187)
C2	0.332	0.332	0.32	0.313	0.294	0.341
	(0.306, 0.353)	(0.309, 0.352)	(0.297, 0.343)	(0.284, 0.342)	(0.257, 0.336)	(0.297, 0.382)
C1	0.14	0.146	0.135	0.157	0.157	0.163
	(0.124, 0.157)	(0.13, 0.162)	(0.118, 0.157)	(0.141, 0.173)	(0.13, 0.179)	(0.136, 0.192)
B3	0.074	0.077	0.076	0.09	0.092	0.094
	(0.065, 0.084)	(0.067, 0.088)	(0.066, 0.088)	(0.078, 0.107)	(0.072, 0.112)	(0.076, 0.112)
B2	0.068	0.069	0.076	0.09	0.091	0.094
	(0.058, 0.079)	(0.06, 0.08)	(0.066, 0.086)	(0.076, 0.104)	(0.071, 0.108)	(0.078, 0.108)
B1	0.038	0.038	0.048	0.06	0.059	0.064
	(0.031, 0.047)	(0.029, 0.047)	(0.039, 0.057)	(0.049, 0.072)	(0.044, 0.073)	(0.044, 0.085)
A	0.017	0.016	0.024	0.031	0.03	0.035
	(0.011, 0.022)	(0.012, 0.021)	(0.017, 0.032)	(0.022, 0.042)	(0.02, 0.042)	(0.023, 0.051)

Considering ATE and ASE estimates, displayed in Tables 19 and 20, we note that the advantage of 1st-tier SA universities over 'other' foreign universities in terms of training effect is somewhat reduced. ATE is estimated to be on the same level for 1st-tier SA and 'other' foreign universities, while ASE remains larger for 'other' foreign universities only in the very high rating categories B1 and A.

In sum, the results of the alternative model with ability dependent treatment effects are in line with those presented in the main section. Therefore we consider the main model as a useful simplification. The next section investigates to what extent insights of the main model pertain under sample and variable variations.

D Variations on the sample and variables

D.1 Scientists with phd obtained after 1960, 1975, or 1980

The main results have been obtained on a sample including scientists which obtained their PhD after 1970. In order to see how results change when the sample covers differing time periods, we estimated the main model for samples including all scientists obtaining their PhD after i) 1960, ii) 1975, iii) and 1980. We provide a short discussion on the results of each but for brevity include a result table only for the sample restriction 'phd obtained after 1975' (Table 21).

Expanding the sample to include all researchers with PhD obtained in or after 1960 includes 1210 researchers with 2537 rating events. Estimation results for individual equations change only slightly. Remarkable is that 1st-tier SA universities

Table 20: ASE from alternative model 'ability dependent treatment effects'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
RU	0.131	0.108	0.086	0.08	0.086	0.059
	(0.114, 0.151)	(0.092, 0.126)	(0.071, 0.1)	(0.069, 0.092)	(0.074, 0.101)	(0.044, 0.071)
C3	0.247	0.21	0.173	0.163	0.173	0.126
	(0.228, 0.265)	(0.192, 0.231)	(0.154, 0.192)	(0.145, 0.178)	(0.152, 0.192)	(0.1, 0.147)
C2	0.337	0.335	0.327	0.324	0.327	0.305
	(0.32, 0.356)	(0.318, 0.353)	(0.312, 0.341)	(0.305, 0.339)	(0.309, 0.343)	(0.281, 0.322)
C1	0.122	0.141	0.16	0.164	0.16	0.183
	(0.109, 0.135)	(0.128, 0.158)	(0.145, 0.172)	(0.15, 0.179)	(0.145, 0.175)	(0.165, 0.204)
B3	0.061	0.075	0.09	0.094	0.09	0.111
	(0.053, 0.069)	(0.062, 0.086)	(0.079, 0.102)	(0.081, 0.106)	(0.077, 0.101)	(0.096, 0.129)
B2	0.055	0.069	0.086	0.091	0.086	0.112
	(0.048, 0.062)	(0.061, 0.08)	(0.076, 0.098)	(0.081, 0.103)	(0.076, 0.097)	(0.095, 0.129)
B1	0.032	0.042	0.053	0.056	0.053	0.07
	(0.027, 0.037)	(0.035, 0.05)	(0.045, 0.064)	(0.047, 0.065)	(0.043, 0.064)	(0.057, 0.083)
A	0.015	0.02	0.026	0.027	0.026	0.034
	(0.012, 0.018)	(0.015, 0.024)	(0.02, 0.031)	(0.022, 0.033)	(0.019, 0.031)	(0.027, 0.042)

obtain a higher estimated treatment effect than 1st-tier Foreign universities in this sample; but only due to a change of about one half of their respective standard errors.

Reducing the sample to include all researchers with PhD obtained in or after 1975 yields very similar results as those of the main estimation. In particular the score ranking of universities based on comparison of the intercept as well as average ability across universities remains the same. This can be expected as the sample changed only slightly, a drop of about 5% in the number of individuals and a reduction by about 10% in the number ratings.

Table 21: Estimation results with sample restriction ‘phd obtained after 1975’

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
Selection and training effects in latent rating (R^*) for white, male, natural science researchers.						
Avg. ability	-0.31 (0.142)	0.11 (0.201)	0.25 (0.262)	0.43 (0.296)	0.41 (0.251)	1.11 (0.312)
Avg. training	3.16 (0.222)	3.46 (0.372)	3.63 (0.349)	3.87 (0.459)	3.82 (0.419)	4.1 (0.559)
Ability equation (γ)						
Master university	0	0.29 (0.181)	0.31 (0.228)	1.06 (0.267)	–	–
Master distinction	0.76 (0.106)	0.76	0.76	0.76	–	–
Age at master	-0.7 (0.117)	-0.7	-0.7	-0.7	–	–
Selection equation (α_1,α_2,ζ						
inertia value	2.28 (0.201)	1.96 (0.2)	2.32 (0.215)	2.95 (0.307)	–	–
1975-79	0	1.02 (0.405)	-0.33 (0.496)	-1.16 (0.664)	0.64 (0.458)	-0.87 (0.744)
1980-84	0	0.28 (0.406)	-0.73 (0.479)	-0.37 (0.499)	-1.34 (0.713)	-1.76 (0.823)
1985-89	0	0.18 (0.339)	-0.36 (0.376)	-0.14 (0.384)	-1.13 (0.528)	-1.39 (0.645)
1990-94	0	-0.18 (0.328)	-0.45 (0.35)	-0.31 (0.356)	-0.9 (0.43)	-0.94 (0.518)
1995-99	0	0.12 (0.294)	-0.76 (0.351)	-0.77 (0.358)	-0.13 (0.332)	-1.68 (0.584)
2000	0	0.22 (0.334)	-0.24 (0.367)	-0.25 (0.376)	0.18 (0.36)	-1.04 (0.574)
Non-white	0	0.07 (0.386)	0.95 (0.35)	0.31 (0.397)	1.2 (0.346)	0.94 (0.535)
Female	0	0.28 (0.238)	-0.05 (0.256)	-0.02 (0.288)	-0.08 (0.282)	-1.14 (0.511)
Social sciences	0	-0.6 (0.225)	-0.14 (0.235)	-1.18 (0.281)	-0.14 (0.258)	-0.08 (0.378)
ζ	0	0.14 (0.16)	0.38 (0.177)	0.02 (0.213)	0.43 (0.202)	0.84 (0.256)
Rating equation β						
05-09 years after phd	1.77 (0.226)	1.68 (0.283)	2.18 (0.365)	2.2 (0.388)	1.95 (0.443)	2.25 (0.571)
10-14 years after phd	2.45 (0.223)	2.51 (0.29)	3.2 (0.36)	3.28 (0.39)	2.89 (0.45)	2.7 (0.539)
15-19 years after phd	2.82 (0.237)	3.28 (0.299)	3.36 (0.367)	3.98 (0.417)	3.83 (0.528)	4.08 (0.596)
20-24 years after phd	3.18 (0.268)	3.72 (0.327)	3.77 (0.393)	3.99 (0.466)	3.39 (0.553)	3.85 (0.779)
25-29 years after phd	3.6 (0.325)	3.69 (0.369)	4.09 (0.441)	3.76 (0.562)	4.38 (0.726)	5.03 (0.868)
30-34 years after phd	4.22 (0.454)	4.74 (0.505)	4.39 (0.6)	3.76 (0.843)	4.88 (0.786)	4.86 (0.978)
35+ years after phd	4.09 (0.791)	4.63 (1.936)	4.42 (0.891)	6.15 (1.993)	5.46 (0.959)	5.91 (2.088)
Non-white	-0.65 (0.33)	-0.63 (0.525)	-0.78 (0.358)	-1.39 (0.437)	-0.44 (0.422)	0.13 (0.733)
Female	-0.33 (0.199)	-0.35 (0.233)	-0.29 (0.273)	-0.3 (0.296)	-1.16 (0.439)	-0.66 (0.777)
Social sciences	-0.31 (0.193)	-0.7 (0.241)	-0.42 (0.24)	0.14 (0.316)	0.61 (0.41)	-0.15 (0.578)
Rating equation hurdles (c)						
C3	C2	C1	B3	B2	B1	A
0	1.73 (0.073)	3.82 (0.094)	4.95 (0.109)	5.79 (0.122)	7.01 (0.149)	8.61 (0.219)
	No. individuals	1131	No. observations	2173	log-Likelihood	-4989.55

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

Estimating the model for an even later period, such as PhD obtained after 1980, however leaves very few rating observations with researchers more than 25 years of experience - which are the ones achieving the higher ratings. Therefore we reduce the model and include one dummy for experience of 25 years or later. Looking at Table 22 shows that ability is estimated very close to the main results. Although coefficient changes compared to the main model remain within one standard deviation, there is a tendency of a reduced training effect (lower period after phd coefficients) for foreign universities. Taking the reduced estimates at face value suggests that the training effect of foreign universities relative to South African universities has been higher during the 70s than during the 80s and 90s. We hesitate however to attach great importance to that result due to the relatively small number of observations in some time periods. Indeed the non-monotonic pattern of experience for foreign universities suggest some over-fitting.

Table 22: Estimation results with sample restriction ‘phd obtained after 1980’

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
Selection and training effects in latent rating (R^*) for white, male, natural science researchers.						
Avg. ability	-0.38 (0.148)	0.1 (0.228)	0.23 (0.28)	0.47 (0.299)	0.4 (0.262)	1.22 (0.324)
Avg. training	2.88 (0.207)	3.03 (0.293)	3.46 (0.347)	3.47 (0.364)	2.65 (0.499)	2.93 (0.508)
Ability equation (γ)						
Master university	0	0.36 (0.202)	0.36 (0.241)	1.2 (0.269)	–	–
Master distinction	0.8 (0.116)	0.8	0.8	0.8	–	–
Age at master	-0.8 (0.12)	-0.8	-0.8	-0.8	–	–
Selection equation (α_1,α_2,ζ						
inertia value	2.28 (0.215)	2.06 (0.215)	2.41 (0.227)	2.93 (0.319)	–	–
1980-84	0	0.22 (0.411)	-0.75 (0.496)	-0.33 (0.501)	-1.32 (0.721)	-1.89 (0.891)
1985-89	0	0.11 (0.355)	-0.38 (0.397)	-0.1 (0.39)	-1.08 (0.535)	-1.49 (0.7)
1990-94	0	-0.24 (0.341)	-0.45 (0.365)	-0.27 (0.361)	-0.84 (0.443)	-0.99 (0.567)
1995-99	0	0.08 (0.307)	-0.77 (0.364)	-0.72 (0.361)	-0.05 (0.341)	-1.72 (0.623)
2000	0	0.18 (0.344)	-0.22 (0.385)	-0.19 (0.386)	0.3 (0.371)	-1.04 (0.607)
Non-white	0	0.11 (0.398)	1.01 (0.356)	0.34 (0.405)	1.17 (0.36)	1.02 (0.552)
Female	0	0.25 (0.247)	-0.1 (0.267)	-0.03 (0.29)	-0.18 (0.294)	-1.13 (0.526)
Social sciences	0	-0.54 (0.238)	-0.19 (0.246)	-1.23 (0.291)	-0.18 (0.273)	-0.19 (0.402)
ζ	0	0.16 (0.169)	0.41 (0.184)	0.06 (0.214)	0.47 (0.215)	0.95 (0.279)
Rating equation β						
05-09 years after phd	1.92 (0.24)	1.91 (0.325)	2.29 (0.384)	2.23 (0.397)	1.82 (0.484)	1.8 (0.601)
10-14 years after phd	2.57 (0.238)	2.6 (0.328)	3.34 (0.379)	3.33 (0.398)	2.99 (0.502)	2.39 (0.57)
15-19 years after phd	2.9 (0.256)	3.25 (0.342)	3.49 (0.394)	4.15 (0.424)	3.47 (0.635)	3.27 (0.655)
20-24 years after phd	3.18 (0.297)	3.68 (0.378)	3.68 (0.422)	4.13 (0.478)	2.65 (0.692)	3.12 (0.824)
25+ years after phd	3.83 (0.384)	3.71 (0.419)	4.49 (0.488)	3.52 (0.579)	2.32 (1.28)	4.08 (0.977)
Non-white	-0.66 (0.329)	-0.71 (0.525)	-0.86 (0.362)	-1.44 (0.434)	-0.44 (0.461)	0.67 (0.754)
Female	-0.37 (0.203)	-0.48 (0.242)	-0.36 (0.286)	-0.35 (0.307)	-1.13 (0.483)	-0.28 (0.784)
Social sciences	-0.34 (0.202)	-0.65 (0.255)	-0.32 (0.249)	0.06 (0.322)	0.97 (0.462)	-0.04 (0.623)
Rating equation hurdles (c)						
C3	C2	C1	B3	B2	B1	A
0	1.69 (0.079)	3.91 (0.104)	5.02 (0.119)	5.91 (0.136)	7.05 (0.166)	8.78 (0.265)
	No. individuals	1051	No. observations	1847	log-Likelihood	-4304.54

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

D.2 Scientists switching university from master to PhD

Throughout the analyses we assumed that the decision and training effect of doing a master at a certain university can be separated from the decision and effect of following a PhD program subsequently. While it is true that in South Africa the Master is in general separated from the PhD, rather than integrated as is often the case in the US, the decision where to do a master might be intermingled with the PhD decision. For PhD students switching university from master to PhD this is certainly not an issue. This argument suggests the alternative sample restriction of including only those PhD students which switched university from master to PhD. This criterion reduces the number of cases considerably to 482 scientists involved in 979 rating events.

Table 23 provides estimation results on this reduced sample. First consider average ability which follows an order that seems natural. It is lowest, -0.74 at 'other SA universities' and subsequently increasing up to 1.02 at '1st-tier foreign universities'. Note that this spread of about 1.7 is larger compared to the main estimation results having a spread of about 1.3. While estimated average ability increases for both foreign university categories, it decreases for SA universities in general and in particular for '1st-tier SA universities'. The reason is that the movement from master university to phd university typically is from lower tier master universities to higher tier PhD universities, and that lower tier SA master universities are estimated to issue master students of lower expected average ability.

The change of average ability caused by restricting the sample to switching students is actually consistent with the idea that a sorting of students after master takes place. More able master students of a university tend to switch to institutions of higher reputation while less able master students might be prompted to do a PhD at a lower ranked institution. The fact that students in general are likely to stay at their master universities for PhD, be it due to switching costs, better fit, or information, seem to help all SA universities to maintain a pool of PhD students with relatively high ability.

Turning to average training (second row in Table 23) we estimate a relatively low average training for foreign universities, even lower than the training of '3rd-tier SA' and '2nd-tier SA universities'. This result suggests that the model might not be able to separate correctly selection from training effects. In fact, foreign university PhD students appear to be advantaged compared to SA students when we combine average ability and training. While ability factors into the rating equation constantly over time, we estimate a training effect which varies over periods after PhD. Thus, ability in our model adds on the training throughout the career. Yet, there is the possibility that selection does not take place on realised scientific ability but on signals of potential ability and that PhD training actually helps to realise that potential. With this interpretation average ability could be interpreted as a training effect which is constant over time, while the period varying intercept after PhD rather describes differences in career trajectories.

The issue is probably amplified by the relatively low number of observations in this sample. The fact that university-period varying intercepts become less ordered from one period to the next indeed suggests some over-fitting.

Table 23: Estimation results with sample restriction 'switching university from master to phd studies'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
	Selection and training effects in latent rating (R^*) for white, male, natural science researchers.					
Avg. ability	-0.7 (0.273)	-0.28 (0.216)	0.13 (0.298)	0.19 (0.322)	0.54 (0.238)	1.05 (0.321)
Avg. training	3.36 (0.377)	3.81 (0.335)	3.71 (0.462)	2.77 (0.523)	3.42 (0.35)	3.56 (0.455)
	Ability equation (γ)					
Master university	0	0.43 (0.234)	0.67 (0.219)	1.26 (0.223)	–	–
Master distinction	0.5 (0.154)	0.5	0.5	0.5	–	–
Age at master	-0.83 (0.171)	-0.83	-0.83	-0.83	–	–
	Selection equation (α_1,α_2,ζ					
inertia value	–	–	–	–	–	–
1970-75	0.82 (0.457)	-0.33 (0.605)	-2.45 (1.404)	1.03 (0.494)	-0.48 (0.732)	
1980-84	-0.04 (0.56)	-1.04 (0.786)	-0.79 (0.799)	-1.28 (0.822)	-1.66 (1.004)	
1985-89	0.25 (0.482)	-0.22 (0.561)	0.68 (0.515)	-0.72 (0.631)	-0.78 (0.753)	
1990-94	-0.3 (0.419)	-0.74 (0.49)	-0.16 (0.467)	-0.78 (0.509)	-0.72 (0.605)	
1995-99	0.12 (0.388)	-0.45 (0.469)	-0.02 (0.468)	0.16 (0.446)	-1.38 (0.701)	
2000	0.05 (0.426)	-0.04 (0.474)	0.04 (0.503)	0.3 (0.471)	-0.85 (0.683)	
Non-white	-0.28 (0.472)	0.79 (0.472)	-0.52 (0.614)	0.92 (0.445)	0.76 (0.609)	
Female	0.52 (0.326)	-0.17 (0.4)	-0.31 (0.428)	0.08 (0.36)	-0.99 (0.574)	
Social sciences	-0.7 (0.298)	-0.52 (0.349)	-1.18 (0.387)	-0.5 (0.319)	-0.57 (0.432)	
ζ	0	0.29 (0.253)	0.69 (0.278)	0.68 (0.297)	0.95 (0.259)	1.32 (0.342)
	Rating equation β					
05-09 years after phd	1.75 (0.448)	2.24 (0.412)	2.39 (0.562)	1.87 (0.51)	1.97 (0.433)	2.09 (0.57)
10-14 years after phd	2.63 (0.462)	2.87 (0.409)	3.66 (0.579)	2.44 (0.546)	2.63 (0.412)	2.46 (0.509)
15-19 years after phd	3.16 (0.474)	3.74 (0.419)	4.11 (0.594)	3.41 (0.599)	3.59 (0.463)	3.69 (0.552)
20-24 years after phd	3.62 (0.518)	4.21 (0.468)	3.39 (0.684)	3.75 (0.734)	3.46 (0.447)	3.85 (0.658)
25-29 years after phd	3.99 (0.589)	4.25 (0.529)	4.45 (0.694)	2.06 (0.976)	4.26 (0.532)	4.56 (0.675)
30+ years after phd	5.01 (0.594)	5.54 (0.631)	4.27 (0.87)	3.12 (1.449)	4.59 (0.457)	4.7 (0.698)
Non-white	-1 (0.61)	-0.59 (0.665)	-0.46 (0.581)	-1.44 (0.903)	-0.36 (0.407)	0.2 (0.627)
Female	0.07 (0.398)	-0.58 (0.378)	0.29 (0.55)	0.24 (0.555)	-1.2 (0.41)	-0.7 (0.743)
Social sciences	-0.32 (0.349)	-0.58 (0.367)	-0.26 (0.462)	0.71 (0.566)	0.24 (0.37)	0.03 (0.534)
	Rating equation hurdles (c)					
C3	C2	C1	B3	B2	B1	A
0	1.68 (0.12)	3.66 (0.15)	4.76 (0.168)	5.56 (0.183)	6.75 (0.214)	8.19 (0.271)
	No. individuals	482	No. observations	979	log-Likelihood	-2512.97

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

D.3 Shanghai ranking of 1st-tier foreign universities

The main results use a categorisation where all universities having a Shanghai ranking of 30 or lower are considered as '1st-tier foreign university'. We varied the coding in both directions. In one estimation only universities with a Shanghai ranking of at most 20 are labelled '1st-tier foreign university', in another estimation all universities among the top 50 in the Shanghai ranking are put in that category.

We discuss first the results of the 'top 20 foreign universities', Table 24. Different from what one might expect, average training of the top 20 universities is estimated to be lower than that of the top 30 universities. The average training of 'other foreign universities' increases as much as '1st-tier foreign universities' lose. The reason however is not in a worse performance of PhDs from 1st-tier foreign universities. The different ordering is due to a higher estimated selection effect of 1st-tier foreign universities. A relatively high estimate of the selection factor ζ for 1st-tier foreign universities results in a higher expected ability for respective PhDs and consequently a higher ASE. All other estimates remain unchanged. This result suggests that differences in graduates between universities of very high reputation increasingly are driven by selection rather than training effects.

Table 24: Estimation results with alternative coding '1st-tier foreign among top 20 Shanghai ranked universities'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
Selection and training effects in latent rating (R^{*}) for white, male, nat. sci. researchers.						
Avg. ability	-0.31 (0.137)	0.07 (0.196)	0.3 (0.253)	0.54 (0.285)	0.47 (0.216)	1.22 (0.335)
Avg. training	3.24 (0.193)	3.48 (0.259)	3.54 (0.322)	3.86 (0.36)	3.85 (0.316)	3.58 (0.461)
Ability equation (γ)						
Master university	0	0.29 (0.177)	0.43 (0.218)	1.16 (0.252)	–	–
Master distinction	0.69 (0.102)	0.69	0.69	0.69	–	–
Age at master	-0.71 (0.115)	-0.71	-0.71	-0.71	–	–
Selection equation (α_1,α_2,ζ						
inertia value	2.29 (0.196)	2 (0.202)	2.35 (0.21)	2.95 (0.3)	0	0
1970-75	0	-0.04 (0.557)	-0.99 (0.591)	-0.94 (0.649)	0.99 (0.473)	-0.69 (0.769)
1975-79	0	1.01 (0.406)	-0.34 (0.502)	-1.18 (0.661)	0.64 (0.456)	-0.77 (0.756)
1980-84	0	0.26 (0.403)	-0.77 (0.476)	-0.4 (0.494)	-1.03 (0.629)	-2.15 (0.968)
1985-89	0	0.15 (0.34)	-0.4 (0.374)	-0.2 (0.389)	-0.85 (0.465)	-1.7 (0.733)
1990-94	0	-0.21 (0.325)	-0.48 (0.348)	-0.35 (0.358)	-0.75 (0.399)	-0.99 (0.551)
1995-99	0	0.09 (0.296)	-0.81 (0.344)	-0.83 (0.353)	-0.14 (0.322)	-1.86 (0.655)
2000	0	0.18 (0.331)	-0.27 (0.366)	-0.29 (0.378)	0.11 (0.351)	-0.83 (0.587)
Non-white	0	0.07 (0.386)	0.93 (0.343)	0.31 (0.397)	1.36 (0.332)	-0.05 (0.735)
Female	0	0.3 (0.241)	0 (0.258)	-0.01 (0.284)	-0.07 (0.273)	-1.14 (0.561)
Social sciences	0	-0.58 (0.223)	-0.12 (0.236)	-1.09 (0.277)	-0.05 (0.244)	-0.28 (0.406)
ζ	0	0.11 (0.155)	0.36 (0.171)	0.06 (0.205)	0.41 (0.175)	0.82 (0.267)
Rating equation β						
05-09 years after phd	1.78 (0.219)	1.7 (0.281)	2.16 (0.351)	2.18 (0.38)	2.18 (0.405)	1.85 (0.594)
10-14 years after phd	2.42 (0.215)	2.49 (0.282)	3.13 (0.347)	3.17 (0.38)	2.88 (0.379)	2.28 (0.522)
15-19 years after phd	2.91 (0.229)	3.28 (0.291)	3.27 (0.353)	3.94 (0.401)	3.97 (0.418)	3.53 (0.557)
20-24 years after phd	3.3 (0.252)	3.72 (0.317)	3.71 (0.371)	4.06 (0.443)	3.77 (0.418)	3.72 (0.697)
25-29 years after phd	3.7 (0.303)	3.82 (0.35)	4.09 (0.423)	4.1 (0.533)	4.62 (0.49)	4.39 (0.699)
30-34 years after phd	4.2 (0.374)	4.69 (0.451)	4.29 (0.532)	4.47 (0.661)	4.45 (0.527)	4.9 (0.836)
35+ years after phd	4.36 (0.492)	4.68 (0.668)	4.17 (0.641)	5.09 (0.785)	5.11 (0.503)	4.41 (0.925)
Non-white	-0.64 (0.325)	-0.62 (0.522)	-0.8 (0.354)	-1.41 (0.438)	-0.38 (0.377)	0.16 (0.955)
Female	-0.34 (0.195)	-0.33 (0.232)	-0.27 (0.264)	-0.35 (0.296)	-1.13 (0.41)	-0.96 (0.827)
Social sciences	-0.32 (0.188)	-0.72 (0.238)	-0.5 (0.23)	-0.1 (0.307)	0.18 (0.349)	0.08 (0.604)
Rating equation hurdles (c)						
C3	C2	C1	B3	B2	B1	A
0	1.73 (0.073)	3.77 (0.092)	4.84 (0.103)	5.64 (0.114)	6.84 (0.134)	8.37 (0.182)
	No. individuals	1189	No. observations	2432	log-Likelihood	-5524.55

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

Results for the alternative coding 'top 50 foreign universities' are as expected and shown in Table 25. Compared to the main estimation results, selection as well as treatment effects of 1st-tier foreign universities decrease. Average training of top 50 foreign universities is estimated to be at the same level as 'other foreign universities' (mostly among the top 200, nearly all Shanghai ranked). This training level is comparable to the top SA universities. Average ability however remains highest for 'top 50 foreign universities'.

Table 25: Estimation results with alternative coding '1st-tier foreign among top 50 Shanghai ranked universities'

	SA university				Foreign university	
	other	3rd-tier at	2nd-tier	1st-tier	other	1st-tier
Selection and training effects in latent rating (R^*) for white, male, nat. sci. researchers.						
Avg. ability	-0.32 (0.139)	0.05 (0.199)	0.27 (0.266)	0.51 (0.294)	0.54 (0.242)	0.75 (0.272)
Avg. training	3.25 (0.191)	3.5 (0.261)	3.57 (0.336)	3.89 (0.369)	3.83 (0.35)	3.8 (0.406)
Ability equation (γ)						
Master university	0	0.28 (0.177)	0.41 (0.226)	1.14 (0.258)	–	–
Master distinction	0.67 (0.1)	0.67	0.67	0.67	–	–
Age at master	-0.71 (0.117)	-0.71	-0.71	-0.71	–	–
Selection equation (α_1,α_2,ζ						
inertia value	2.3 (0.199)	2 (0.201)	2.35 (0.21)	2.95 (0.298)	0	0
1970-75	0	-0.04 (0.561)	-0.97 (0.594)	-0.95 (0.655)	0.86 (0.495)	-0.09 (0.621)
1975-79	0	1.02 (0.407)	-0.34 (0.498)	-1.2 (0.662)	0.38 (0.486)	-0.03 (0.571)
1980-84	0	0.26 (0.403)	-0.75 (0.476)	-0.4 (0.5)	-1.41 (0.718)	-1.33 (0.749)
1985-89	0	0.15 (0.342)	-0.38 (0.381)	-0.19 (0.392)	-1.65 (0.622)	-0.73 (0.505)
1990-94	0	-0.21 (0.326)	-0.48 (0.35)	-0.36 (0.358)	-1.36 (0.484)	-0.43 (0.417)
1995-99	0	0.09 (0.295)	-0.79 (0.347)	-0.83 (0.359)	-0.4 (0.353)	-0.9 (0.435)
2000	0	0.18 (0.332)	-0.27 (0.367)	-0.3 (0.377)	0.01 (0.375)	-0.62 (0.471)
Non-white	0	0.07 (0.384)	0.93 (0.341)	0.33 (0.397)	1.22 (0.36)	1.04 (0.442)
Female	0	0.31 (0.239)	0 (0.256)	0 (0.288)	-0.05 (0.3)	-0.67 (0.391)
Social sciences	0	-0.58 (0.222)	-0.12 (0.233)	-1.09 (0.278)	0 (0.264)	-0.26 (0.314)
ζ	0	0.11 (0.155)	0.35 (0.175)	0.05 (0.207)	0.46 (0.19)	0.56 (0.205)
Rating equation β						
05-09 years after phd	1.79 (0.219)	1.72 (0.281)	2.18 (0.36)	2.2 (0.388)	2.07 (0.47)	2.19 (0.481)
10-14 years after phd	2.43 (0.216)	2.5 (0.284)	3.16 (0.355)	3.19 (0.383)	2.81 (0.423)	2.65 (0.44)
15-19 years after phd	2.91 (0.227)	3.3 (0.293)	3.3 (0.366)	3.96 (0.408)	3.82 (0.474)	3.96 (0.482)
20-24 years after phd	3.31 (0.25)	3.74 (0.317)	3.73 (0.382)	4.09 (0.451)	3.69 (0.451)	4.03 (0.589)
25-29 years after phd	3.71 (0.298)	3.83 (0.356)	4.12 (0.431)	4.13 (0.536)	4.58 (0.539)	4.61 (0.611)
30-34 years after phd	4.21 (0.372)	4.71 (0.453)	4.32 (0.533)	4.5 (0.665)	4.61 (0.563)	4.69 (0.743)
35+ years after phd	4.37 (0.494)	4.7 (0.672)	4.19 (0.654)	5.12 (0.8)	5.21 (0.536)	4.44 (0.8)
Non-white	-0.64 (0.325)	-0.61 (0.518)	-0.79 (0.355)	-1.39 (0.441)	-0.49 (0.432)	0.03 (0.548)
Female	-0.33 (0.197)	-0.32 (0.233)	-0.26 (0.266)	-0.34 (0.296)	-1.02 (0.471)	-1.34 (0.583)
Social sciences	-0.31 (0.187)	-0.72 (0.237)	-0.5 (0.232)	-0.1 (0.309)	0.33 (0.402)	-0.1 (0.47)
Rating equation hurdles (c)						
C3	C2	C1	B3	B2	B1	A
0	1.73 (0.072)	3.77 (0.092)	4.84 (0.103)	5.65 (0.114)	6.85 (0.135)	8.38 (0.182)
	No. individuals	1189	No. observations	2432	log-Likelihood	-5544.08

Notes: i) Figures in () are standard errors, ii) coefficients without standard errors are fixed across universities or fixed to zero, iii) – if coefficient not applicable.

Working Paper Series in Economics

recent issues

No. 82 *Helena Barnard, Robin Cowan, Moritz Müller:* On the value of foreign PhDs in the developing world: Training versus selection effects, January 2016

No. 81 *Enno Mammen, Christoph Rothe, Melanie Schienle:* Semiparametric estimation with generated covariates, January 2016

No. 80 *Carsten Bormann, Julia Schaumburg, Melanie Schienle:* Beyond dimension two: A test for higher-order tail risk, January 2016

No. 79 *Frank Betz, Nikolaus Hautsch, Tuomas A. Peltonen, Melanie Schienle:* Systemic risk spillovers in the European banking and sovereign network, January 2016

No. 78 *Armin Falk and Nora Szech:* Pleasures of skill and moral conduct, January 2016

No. 77 *Kirill Borissov, Mikhail Pakhnin, Clemens Puppe:* On discounting and voting in a simple growth model, December 2015

No. 76 *Nikolaus Schweizer and Nora Szech:* A quantitative version of Myerson regularity, December 2015

No. 75 *Tim Deeken:* Schumpeterian growth with technological interdependence: An application to US states, November 2015

No. 74 *Tim Deeken:* Knowledge spillovers: On the impact of genetic distance and data revisions, November 2015

No. 73 *Christian Feige and Karl-Martin Ehrhart:* Voting and transfer payments in a threshold public goods game, November 2015

No. 72 *Steffen Huck, Nora Szech, Lukas M. Wenner:* More effort with less pay: On information avoidance, belief design and performance, September 2015