Maastricht University

UNITED NATIONS UNIVERSITY

UNU-MERIT

Working Paper Series

#2020-021

Ph.D. research output in STEM: the role of gender and race in supervision

Giulia Rossello, Robin Cowan and Jacques Mairesse

Published 15 May 2020

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT to stimulate discussion on the issues raised.

Ph.D. research output in STEM: the role of gender and race in supervision*

Giulia Rossello[1,⋆], Robin Cowan[1,2,3,6] and Jacques Mairesse[1,4,5]

[1]UNU-MERIT United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology, Boschstraat 24, 6211 AX Maastricht, the Netherlands
[2]BETA Bureau d'Economie Théorique et Appliquée Université de Strasbourg, Avenue de la Forêt Noire 61, 67000 Strasbourg, France
[3]IUF Institut Universitaire de France, Descartes 1, 75231 Paris, France
[4]CREST-ENSAE Centre de Recherche en Économie et Statistique, 5 Ave. Henry Le Chatelier, 91120 Palaiseau, France
[5]NBER National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, Massachusetts, United States of America
[6]CREST Centre for Research on Evaluation of Science and Technology, Stellenbosch University, RW Wilcocks Building, 7600, Stellenbosch, South Africa
⋆**Corresponding author:** Rossello `rossello@merit.unu.edu`

May 13, 2020

*Financial support was provided through the Institut Universitaire de France. We gratefully acknowledge the comments and suggestions of participants of the workshop "APEX – Academic Pipeline Development for Excellence" at CREST-SciSTIP as well as those of Dr. Gloria Bernal, Dr. Tatenda Zimyemba, Prof. Kristof de Witte, Dr. Arjan Schakel, and Prof. Louis Volante.

Abstract

We study whether student-advisor gender and race couples matter for publication productivity of Ph.D. students in South Africa. We consider the sample of all Ph.D.s in STEM graduating between 2000 and 2014, after the recent systematic introduction of doctoral programs in this country. We investigate the joint effects of gender and race for the whole sample and looking separately at the sub-samples of (1) white-white; (2) black-black; and (3) black-white student-advisor couples. We find early career productivity differences: while female students publish on average 10% to 20% fewer articles than males, this is true mainly for female students working with a male advisor, not for those working with a female one. These disparities are similar, though more pronounced, when looking at the joint effects of gender and race for the white-white and black-black student-advisor pairs. We also explore whether publication productivity differences change significantly for students with a high, medium, or low "productivity-profile", and find that they are U-shaped. Female students with a high (or low) "productivity-profile" studying with female advisors are as productive than male students with a high (or low) "productivity-profile" studying with male advisors.

JEL codes: A14, I23, I24, J15, J16, J24, O32.

Keywords: Gender and race, Student Advisor, South Africa, Doctoral research productivity, Role models.

1 Introduction

The gender gap in publishing is well documented: depending on the context, discipline, geography or era; female scientists are found to produce fewer papers per year than their male colleagues (Allison and Stewart, 1974; Cole and Zuckerman, 1984; Fox and Faver, 1985; Mairesse and Pezzoni, 2015; Holman et al., 2018; Lerchenmueller and Sorenson, 2018; Mairesse et al., 2019; Pezzoni et al., 2016).

However, little previous research focuses on scientists' productivity during their doctoral studies and how it relates to advisor characteristics. In this paper, we study whether Ph.D. students' early career productivity is affected by the gender (and/or race) of student and supervisor. In the first analysis, we simply ask whether there is a correlation between students' publication output and the gender (race) of the student, and, independently, the gender (race) of the supervisor. In the second analysis, we ask whether there is an observable effect of the student-supervisor pair. Our data are drawn from an emerging economy, namely South Africa, where resource constraints in the science system generally, and universities, in particular, are much more severe than they are in developed countries. One might expect that in the presence of resource constraints, the "privileged group" will have better access and, therefore, a higher productivity relative to others. We observe that the academic science system in South Africa is relatively small — in 2012 there were only 2174 full professors,[1] and the production of Ph.D.s is concentrated in a relatively small number of institutions (Cowan and Rossello, 2018).[2] These features are typical of many developing countries (Nchinda, 2002; Gonzalez-Sauri and Rossello, 2019).

While concerns with gender in science are common to many countries today, given the history of apartheid and its on-going legacy, in South Africa, there is a second axis of concern, namely race.[3] In South Africa, people of color make up 90% of the population and apartheid essentially excluded them from academia. Until 1994 there were "black universities", but they were severely underfunded and not expected to do any meaningful research. One of the on-going efforts of governments since 1994 has been the transformation of the university system to include more of the black population in the "top" (formerly white) universities. Part of the challenge has been the academic "pipeline". Whether or not faculties are trying to hire previously excluded groups, if there is no supply of them, the system will not transform from a white male bastion to a more inclusive institution. Given that academic appointments are often heavily based on performance during graduate studies, understanding gender and race effects on Ph.D. student publishing becomes something of significant importance in this context.

Pezzoni et al. (2016) have done a similar analysis using data from the California Institute of Technology (Caltech), an elite institution in the US, and our work is modelled on theirs. They found that compared to the male-male student-advisor couple: female

[1] These data are available at `https://africacheck.org/reports/how-many-professors-are-there-in-sa/` last access November 2019.

[2] For a further discussion on the South African system see Rossello and Cowan (2019), and the report Mouton et al. (2015)

[3] "Race" is sometimes considered a contentious concept (and word) but in the context of South Africa it is well understood as central to the construct of the society, so we will employ the word and concept here in the way it is done in South Africa.

students working with male advisors publish 8.5% less; and male students working with female advisors publish 10% more. Their data were constrained to a single, rather specific (in terms of student and faculty quality, and finances, only to mention two dimensions) institution, namely Caltech. Ours involves the entire national academic science system, and so might be considered more representative of national trends and effects. In addition, our statistical analysis differs from theirs in an important respect. They study the relation between Ph.D. student productivity and student-advisor gender couple, controlling for several variables such as advisor past productivity and using for the estimation OLS panel regressions. We also thought interesting to implement a quantile regression analysis, in order to explicitly assess the student productivity distribution across genders (or races) throughout the population being studied. We can test in particular if differences in gender-specific productivity significantly vary depending on the "productivity-profile" (high, medium or low) of the student. Differential effects across different population groups are also more likely to be relevant where the output variable is skewed and has a fat tail, as is the case with publications (Petscher and Logan, 2014).

Note that we can explore both gender and race dimensions because the student body has a close to balance population in those characteristics.

We study productivity differences across gender-couples in the whole sample and separately for sub-samples of same- and cross-racial couples.[4]

As a preview, our main findings are the following: Female students on average publish 10%-20% fewer articles than males. This average gap is mostly driven by female students working with male advisors. Considering the joint effect of gender and race, it disappears for female students working with female advisors. Productivity differences with a high, medium, or low "productivity-profile" are U-shaped. While there is a productivity gap between female students with a medium "productivity-profile" studying with a female advisor and male students with the same "productivity-profile" studying with a male ad-

[4]In a specular way, we repeat the same for the comparison of racial-couples, doing the analysis for the whole sample and for the sub-samples of same- and cross-gender couples.

visor, there no such gaps in both cases of female and male students with a high or a low "productivity-profiles".

The remainder of the paper is organized as follows. In Section 2, we discuss how the paper relates to the existing literature. Section 3 presents the data, and Section 4 describes the methodology. In Section 5, we present the results and discuss their magnitudes. Finally, Section 6 concludes.

2 Early career productivity and student-advisor gender composition

Previous research examines the gender gap in publications between male and female scientists (Allison and Stewart, 1974; Cole and Zuckerman, 1984; Fox and Faver, 1985; Mairesse and Pezzoni, 2015; Holman et al., 2018; Lerchenmueller and Sorenson, 2018; Mairesse et al., 2019; Pezzoni et al., 2016).

However, the ultimate sources of this gap remain elusive, though Mairesse and Pezzoni (2015) have found that when biases in promotion decisions, and the frequency of "idle periods" are controlled for, women are more productive than men.[5] They admit, though, that their context is specific, and they do not claim to have presented the universal explanation.

It is common to observe in studies of the gender gap that age plays a role in publishing productivity and that the gap can change with age (Kelchtermans and Veugelers, 2011). This observation, combined with the well-known Matthew Effect (Merton, 1988) suggest that productivity gaps might originate very early in the career. An important open issue then is whether we observe publishing productivity gaps early in the career (David, 1993; Conti and Visentin, 2015), and if so, how to understand them. We can get at this issue by

[5]The context of their study is France 1982-2005 and they look at 2811 scholars in Physics in universities and Centre national de la recherche scientifique (CNRS). A similar study has been done in Mexico and South Africa, and finds that, after controlling for promotion biases, female are 8% more productive than male and that there are no differences in terms of publication quality (Rivera León et al., 2017)

examining publications of scientists during the course of their doctorates.

While the study of the gender gap focuses on single scientists, it must be acknowledged that much publishing involves more than the focal author (Wager et al., 2015; Chuang and Ho, 2014; Larivière, 2012). Not only co-authors, but research assistants, co-workers, technicians, conference participants, and many others contribute with work, ideas, and suggestions. Of course, when we are considering Ph.D. students as a (co-)author, the thesis supervisor is very likely to provide significant input.

Often the thesis advisor is the first person with whom a student co-authors, but additionally, supervisors play a key role in introducing students into the profession. It seems very likely that the properties of the supervisor matter for a student's early success (Li et al., 2019). A priori, there are some obvious traits of the supervisor that will matter: the extent of supervision, publishing record, status in the profession, quality, and so on. But other literature suggests that gender (race) might also matter. For example, subtle gender and racial biases can distort the meritocratic evaluation of the students. An experiment in a sample of 127 biology, chemistry, and physics professors in the US, asks academics to evaluate the CV of students for a laboratory manager position, where gender was randomly assigned to CVs. It finds that both male and female faculty judge female students as less competent, less likely to be hired than an identical male student, and also offered her a smaller salary and less mentoring (Moss-Racusin et al., 2016). Such biases can also reduce a student's access to relevant information. A similar randomization experiment finds that black students are less likely to receive warning information from academic advisors than are white students when the race is randomly assigned to student academic records (Crosby and Monin, 2007).

Gender (or racial) bias can play a role through both sides of the relationship (Rossello and Cowan, 2019). From the student side, in education and learning the gender of the advisor can affect performance and beliefs (Gaule and Piacentini, 2018; Breda et al., 2018; Rossello and Cowan, 2019). For example, female role models are often more effective in inspiring female students (Bettinger and Long, 2005; Lockwood, 2006; Aguinis et al., 2018). A recent French experiment, among senior high school students, finds a reduction of stereotypes associated with jobs in science after students took a class with female scientist (Breda et al., 2018). In the same study, enrolment in a selective science programme increased by 30% among the higher achieving students. Furthermore, the share of female (male) students in STEM programs were 38% (28%) than that in classes that did not receive the intervention.

Thus, we might expect to see female students performing better with female advisors. In South African academia, after controlling for preferential attachment and institutional constraints, Rossello and Cowan (2019) find preferences for same gender (race) in student-advisor tie formation in a sample of bachelor, master and Ph.D. students and advisors based on enrolment data. In particular, male (white) students have a high tendency to form same-gender (race) relations, while among professors it is female (black) faculty who display the higher frequency.

From the advisor side, the gender of the student can also be relevant. Each Ph.D. student shares with others a thesis advisor who guides and supervises the research, provides access to knowledge (tacit in particular), co-authors, resources, and job opportunities. Thus, gender biases in this phase can limit the access of the student to resources and information. Past research has found that supervisors provide more psychological support to protégés of the same gender (Koberg et al., 1998; Aguinis et al., 2018); male advisors were more likely to agree to a mentoring meeting with a male student than with a female student with same characteristics (Milkman et al., 2015); and less willing to supervise female students (Moss-Racusin et al., 2016).

Exploring the relationship between gender (race) and performance in the student-supervisor pair is a step towards understanding productivity differences among different groups within academia. Past research in Science, Technology, Engineering and Mathematics (STEM) is available only for the US in a first-tier institution (Pezzoni et al., 2016) or for a single field (Gaule and Piacentini, 2018). Looking 20,000 Ph.D. graduated between 1999 and 2008 in US chemistry departments, Gaule and Piacentini (2018) find that same-gender couples tend to be more productive during the Ph.D., and that female students working with female advisors are more likely to become faculty members compared with female students working with a male advisor. In contrast, Pezzoni et al. (2016) study all fields in STEM with data based on 933 Ph.D. graduates and 204 advisors at the Caltech between 2004 and 2009. In terms of student publication productivity, they find no difference between the female-female and the male-male couples. However, they find that male students working with a female advisor perform better than male-male peers, while female students working with male supervisors perform worse than male students working with male advisors.

A difference in performance of students depending on the gender composition of the student-advisor couple can be the expression of multiple mechanisms. Past contributions underline the importance of student-supervisor personal relations; access to resources; differences in the career paths; different nature of the research output in terms of content (for example between basic or applied research which can translate into differential 'publishability').

The personal relations hypothesis is compatible with results in Rossello and Cowan (2019), which finds a same-gender (same-race) bias in supervision-tie formation, driven mainly by male (white) students and female (black) advisors. Bias in tie formation relates to group behaviour and socialization in the working environment which may disadvantage female students working with males (Blackburn et al., 1981; Van den Brink and Benschop, 2014; Zinovyeva and Bagues, 2015). More in general, social relations are embedded in networks which vary with gender and enhance or restrict access to resources, information, and collaborations (Jadidi et al., 2018).

Differences in productivity are often explained by differences in career paths induced by motherhood (see Pezzoni et al. (2016)). Past research has found that female productivity has a negative shock during the first 3 years of a newborn (Mairesse et al., 2019). A similar shock may be accommodated differently depending on whether the female student works with a male or with a female supervisor.[6]

A further mechanism can relate to the two-world hypothesis. This hypothesis states that there exists a gender or racial specialization in specific (sub-)disciplines (Moore et al., 2018). Thus, cross-gender couples may combine different (sub-)fields and knowledge. More in general, the management literature has found that diversity is associated with novelty and innovation because it is more likely to recombine distant knowledge and expertise (Rzhetsky et al., 2015; Shi et al., 2015; Uzzi et al., 2013; Chen et al., 2009; Fleming, 2001). In science, novelty is often a risk, particularly for a young scientist, and may have slower returns (Wang et al., 2017; Boudreau et al., 2016; Verhoeven et al., 2016; Azoulay et al., 2011). Taking risks early in the career may slow down productivity in the short-run affecting 'publishability' of the research. Different gender composition pairs may differently mitigate the risk.

All these mechanisms may individually and jointly explain the importance of student-advisor gender composition for the early career productivity of both male and female scientists. The first step in this direction is to access whether student-advisor gender couples matter for publication productivity of doctoral students in a sample with a good representation across fields and universities.

Table 1: Students and Advisors, by Race and Gender. A professor can supervise more than one student.

	Advisor					
	White Male	White Female	White	Black Male	Black Female	Black
Stud. White Male	179	53	232	13	4	17
	55%	37%	49%	54%	36%	49%
Stud. White Female	149	92	241	11	7	18
	45%	63%	51%	46%	64%	51%
Stud. White	328	145	473	24	11	35
	100%	100%	100%	100%	100%	100%
Stud. Black Male	123	45	168	97	17	114
	71%	64%	69%	64%	81%	66%
Stud. Black Female	51	25	76	54	4	58
	29%	36%	31%	36%	19%	34%
Stud. Black	174	70	244	151	21	172
	100%	100%	100%	100%	100%	100%

3 Material

3.1 Data

Our data originate from the National Research Foundation (NRF) database of South African Academia.[7] The NRF has a system, in which academics at South African universities apply to be "rated". This rating has (until recently) financial and prestige incentives, so most academics in South Africa who pursue a research career do apply. Overall, rated scholars comprise about 30% of South African scholars who produce roughly 90% of the country peer-reviewed output. The STEM fields have been part of the system longer than have SSH fields, and in these fields coverage appears to be more complete. Consequently, we restrict attention to STEM, where the agency has a primary role in funding research. We create a unique dataset using data supplied in the application process. The raw data include student Ph.D. supervision from 2000 to 2014 and publications from 1961 to 2014.

We match students and supervisors with NRF publication data. To be confident that our publication data are complete, we include in the analysis only Ph.D. students in STEM who became active scholars in the NRF system. They constitute 25% of the total Ph.D.

[6] In South Africa, female fertility rates peak at age 25-29 which corresponds to doctoral years (Lehohla, 2015).

[7] NRF is a state agency that has as its mission the promotion of research and the development of national research capacity. `https://www.nrf.ac.za/`

graduates over the period. Our final sample represents Ph.D. students within the enrolment period of 2000-2012 and with a graduation period up to 2014. In our sample, the average completion time is after 3.8 years of enrolment (with a median of 4 years and a maximum of 12 years). Our panel represents each student from her/his enrolment year to two years after her/his graduation, obtaining a total of 6049 observations representing 924 Ph.D.s and 549 thesis supervisors.

In the period 2000-2014 the number of Ph.D.s graduated increased rapidly.[8] Our sample, in table 5 in the appendix, shows similar trends and has a good representation in terms of the distribution of Ph.D. graduation over time. However, the last two periods have a lower number of graduates relative to national statistics. The discrepancy is because it takes several years after graduation before a faculty member is ready to apply for rating. So by restricting to students who eventually do apply for rating, we will under-sample the later years. Distributions of Ph.D.s graduated over disciplines, in table 6 (appendix B), are also in line with national statistics.[9]

Students in our sample are 58% (249 white and 282 black) male and 42% female (259 white and 134 black). Professors in our final sample are 73% male (298 white and 104 black) and 27% female (130 white and 17 black).[10] Table 1 shows the population composition in terms of student and advisor pairs. The majority of students are supervised by white male advisors (54%) followed by white female (23%), black male (19%), and black female advisors (3%).

[8]As reported in the data of the Council of Higher Education (CHE) available for the period 2008-2012 at https://www.che.ac.za/focus_areas/higher_education_data/2008/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2009/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2010/graduates https://www.che.ac.za/focus_areas/higher_education_data/2011/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2012/graduates.

[9]As reported in the data of the Council of Higher Education (CHE) available for the period 2008-2012 at https://www.che.ac.za/focus_areas/higher_education_data/2008/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2009/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2010/graduates https://www.che.ac.za/focus_areas/higher_education_data/2011/graduates; https://www.che.ac.za/focus_areas/higher_education_data/2012/graduates.

[10]The sample demographic composition is close to that of the system in that period. See Rossello and Cowan (2019) for further discussion.

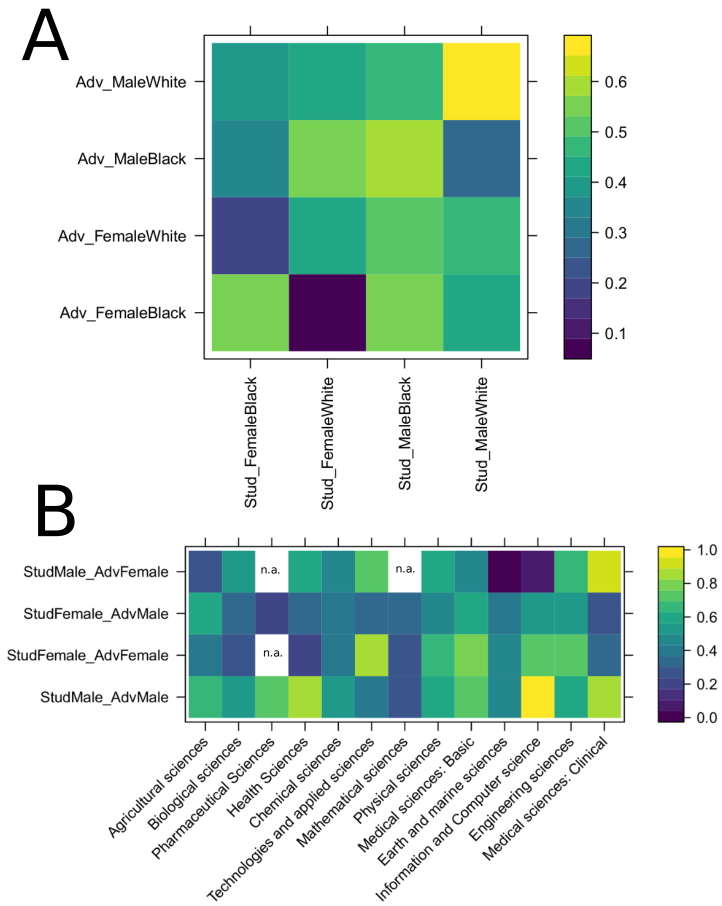


Figure 1: Heat-map of doctoral average annual productivity for student and advisor gender (racial) combinations. The color intensity of each entry represents the average annual productivity of each group. Darker (lighter) colors represent lower (higher) productivity values. Productivity is $log(1+pub_t)$, where pub_t is number of student publications between year t and $t+2$ inclusive, divided by 3. Rows in sub-figure A are advisors gender-race type while columns are student gender-race type. In sub-figure B rows are student advisor gender couples and columns are fields.

3.2 Variable description

For the average student, looking at her/his three-year moving windows between enrolment and graduation plus 2 years, the annual average number of publications is close to one; where white male students have the most (1.58) followed by black males (1.37), white females (1.21) and black females (0.75). The median values are close to zero for all groups (See Table 3 in the appendix). Publication data, also referring to a 3 year average, are skewed, and 44% (410 Ph.D.s) of the students do not publish at all between enrolment years and two years after graduation.[11]

We define student productivity as $log(1+pub_t)$ were pub_t is the number of student publications between year t and $t+2$ inclusive divided by 3. Raw differences in student productivity between different populations and student-supervisor pairs are presented in figure 1 and table 7 in the appendix. Figure 1(A) shows that same-type supervision (2nd diagonal) correlates with higher average productivity (lighter colors).[12] Further, looking at the type of students (by columns) and advisors (by rows), productivity displays a substantial heterogeneity across supervision couples, suggesting a complex joint effect of gender and race We explore this further in table 7(e) in the appendix: looking at black students working with white advisors, it is the couple (black) males with (white) female advisors who publish most. Similarly, in the population of female students working with male advisors (table 7(f) in the appendix), white (female) students working with black (male) advisors have the highest average productivity. Interestingly, the female students who stand out in terms of productivity, in the top decile of the productivity distribution, are those who have supervisors of the different race (Figure 6(d) in the appendix).

Figure 1(B) shows average productivity across student-advisor gender and disciplines. In 5 out of 13 fields, the couple female student with a female advisor has the highest average productivity.[13] In 2 fields, Mathematics and Medical: clinical, cross-gender ties are those with the highest averages. In the remaining 6 fields, the couple male student with male advisors has the highest average productivity. Overall, the couple female student with a male advisor has the lowest average productivity for 6 out of 13 fields.[14]

[11] Details on relative publication rates over time are shown in Figure 3 (a)(b) in the appendix.

[12] White females are an exception: they display higher averages when they work with black male supervisors, however the group has very few observations.

[13] The 5 fields are: Technologies and applied sciences, Physical sciences, Medical science: basic, Earth and marine sciences, Engineering

[14] In table 8 in the appendix we check whether there are any environmental effects at the level of university or field in terms of gender and racial likelihood of supervision association. There are not any identifiable environmental effects in terms of gender in our sample. But there are along racial lines, thus we run our analysis on separate racial sub-sample of the data.

4 Methods

The raw data indicate that female and black Ph.D. students publish less than male and white students. Nevertheless, these differences could be driven by many things. In the analysis that follows, we control for several factors that are likely to contribute to a scientists' publication productivity in order to isolate the effects of gender and race. Our variable of interest is the number of publications produced by a student in a year during their doctorate. Because this variable has a skewed distribution, we work with logs: $\log(1+pub_t)$. There is always a lag between the date of (completion of) the research and publication, so we include in our definition of pub_t publications of which the student was a (co-)author, between years t and $t+2$ inclusive. We normalize for annual output by dividing by 3.

In line with Pezzoni et al. (2016), we ask whether doctoral productivity differs for: (1) student gender; (2) advisor gender; (3) the genders of the student-advisor pair. We estimate panel OLS regressions with robust clustered standard errors.[15]

In all regressions, we control for discipline, enrolment year, time to graduation, whether the student had published previously, whether the student has more than one advisor, and advisor productivity as the log of average publications of the advisor lagged one year.[16]

In table 9 in appendix C, we control for the joint effects of gender and race exploring the interaction terms. This preliminary analysis shows that the main difference in productivity is between male and female students: race has no role. Since the end of apartheid, the progressive introduction of black was not uniform across gender. Black females are under-represented both among students and professors, particularly in STEM fields. For this reason, we also run the analysis on different sub-samples of the data to decompose the possible joint effects of gender and race.

[15] As robustness check we report Poisson panel regressions in appendix E with robust clustered standard errors. The results are not qualitatively different.

[16] In appendix J we show results of the OLS panel regressions with our main variables of gender and race and controlling only for field and enrolment years.

For the gender analysis, we look at the sub-samples: white students with white advisors, black students with black advisors, and black students with white advisors. Similarly, given the context of the country, we run a parallel analysis in section 5.1 to compare black and white. Here we look at the sub-samples: male students with male advisors, female students with female advisors, and female students with male advisors.[17]

As a further contribution to understanding where the gaps originate, we use quantile regression to examine the effects of the student-supervisor pair, where the quantiles are defined over productivity. The approach permits us to observe that much of the difference in average publication between male and female, white and black students is driven by differences in the right-hand tail of the output distribution.

The quantile regression formulation is:

$$Q_\tau(Y_{it}|Z_i, X_{it-1}) = \alpha_\tau + \gamma_\tau Z_i + \beta_\tau X_{it-1} + \varepsilon_{it} \tag{1}$$

where $Q_\tau(Y_{it}|Z_i, X_{it-1})$ is the τth quantile regression function, Z_i are time invariant covariates and X_{it-1} are time variant lagged controls and ε_{it} is the error term.[18]

5 Results

Table 2 shows the results of OLS estimations of three models. The models compare: student's gender, advisor's gender, and the student-advisor gender couples. The main independent variables are: the dummy StudFemale equal to 1 for female students and zero otherwise; the dummy AdvFemale, which is 1 for female advisors; and the dummies for the different student-advisor couples StudFemale_AdvFemale, StudFemale_AdvMale, StudMale_AdvFemale where the baseline category is the pair male students with male

[17]There are too few white students with black female advisors, and male students with black female advisors to give reliable results for those groups. Hence there were not included.

[18]For the estimation we use robust clustered standard errors to account for heteroskedasticity and intra-cluster correlation as described in Machado et al. (2011)

advisors. For each model, we show results for the whole population (ALL) and partitioning the data according to student-supervisor racial composition (WW for white students white advisors; BB for black students black advisors; BW for black students and white advisors).

Model 1 compares female with male students. Results for the whole sample (column 1) show that female students produce on average 11% fewer papers than male students. Looking at white students working with white advisors (column 1a), we find that female students produce on average 12% less than male students. Looking at black students working with black advisors (column 1b), we find that female students display a larger gap — 22% compared with males. Finally, among black students with white advisors (column 1c), there is no difference between male and female students.[19]

Table 2: Pooled OLS Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Models (1) for the Students comparison; Model (2) Advisor comparison; Model (3) Couples Comparison. Where Columns (a) On the sub-sample White Student White Professors; Columns (b) On the sub-sample Black Student Black Professors; Columns (c) On the sub-sample Black Student White Professors. Additional controls are moreAdv, logprofcumavgprod, DummyStudPrevPub, timegrad, field, and enrolment year.

	(1) ALL	(1)a WW	(1)b BB	(1)c BW	(2) ALL	(2)a WW	(2)b BB	(2)c BW	(3) ALL	(3)a WW	(3)b BB	(3)c BW
StudFemale	-0.113**	-0.116*	-0.217*	-0.0908								
	(0.0357)	(0.0497)	(0.0847)	(0.0686)								
AdvFemale					-0.0393	-0.0248	0.0989	-0.0678				
					(0.0416)	(0.0589)	(0.151)	(0.0720)				
StudFemale_AdvFemale									-0.120*	-0.113	-0.270	-0.123
									(0.0571)	(0.0745)	(0.239)	(0.129)
StudFemale_AdvMale									-0.136**	-0.142*	-0.198*	-0.112
									(0.0417)	(0.0580)	(0.0958)	(0.0810)
StudMale_AdvFemale									-0.0733	-0.0622	0.0768	-0.0931
									(0.0568)	(0.0899)	(0.172)	(0.0839)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6049	3083	1099	1641	6049	3083	1099	1641	6049	3083	1099	1641
R^2	0.284	0.354	0.260	0.263	0.280	0.349	0.244	0.260	0.285	0.355	0.260	0.264
Standard errors in parentheses												
* p<0.05	** p<0.01	*** p<0.001"										

Using the same structure, model 2 (table 2) compares male and female advisors. We find that female and male advisors have students that are not statistically different in productivity.[20]

[19] We should underline that in the BW sub-sample, outstanding students (top 10% more productive) are females and have a median productivity higher than males. However, they comprise less than 25% of their relative population (Appendix A figure 6).

[20] It is important not to draw hasty conclusions from this result. It is consistent with a situation in which male advisors favour male students, thus having productive male and unproductive female students, and

Model 3 in table 2 explores the gender pairs of students and advisors, where the baseline category is the student-supervisor pair male-male. Overall female students working with male advisors have the most significant gap compared with the male-male couple: they produce, on average, 14% fewer papers, while it is 12% fewer for female students working with female advisors. Male students working with female advisors do not differ in productivity with male-male. Decomposing the joint effect of gender and race, we find that the gap in productivity between female and male students mainly exists for female students working with male advisors in same-race supervisions. In particular, we find that when students and supervisors are both white (column 3a), female students working with male advisors produce 14% fewer papers than male students working with male supervisors. Similarly, among black-black supervision pairs (column 3b), female students working with male advisors produce, on average, 20% fewer papers than do male students working with male advisors. Interestingly, the group of black students working with white advisors (column 3c) display no significant difference in productivity between gender couples.

To go beyond average differences and to accommodate the skewness and fat tails of the dependent variable, we explore model (3) using quantile regressions with clustered standard errors. In this way, we look for the origin of this difference and ask whether the discrepancy between groups is stronger or weaker for different parts of the population, where the population is sorted into quantiles by publication productivity.

Figure 2 presents 40 quantile regression estimates (each 2.5% percentiles of the population) for the whole data and the sub-samples. It shows the coefficients of the dummies student-supervisor gender pairs with 95% confidence intervals, where zero represents the male-male supervision baseline. Results for the whole sample show that productivity differences of female-female and female-male with the male-male couple are u-shaped over student productivity (fig.2(a)(b)(c)). The u-shaped productivity gap is most pronounced

female advisors doing the reverse. This kind of homophilous preferential attention, were it to exist, would produce the results we see here. This observation should not be read as a conjectural explanation, though, but rather a caution against quickly interpreting this to mean that advisors are gender-blind.

for female students working with female advisors, who are not statistically different from male-male for low (<70th percentiles) and high (>90th percentiles) student "productivity-profile" (fig.2(a)). Female students working with male advisors overall display more significant gaps with the male-male couple in line with OLS results (fig.2(b)).

Results for data sub-samples look at the joint effect of gender and race. Overall we find that the productivity gap (with the baseline male-male couples) increases with student productivity. The figure shows that the productivity difference of female students working with female (fig.2(d)(g)) or male advisors (fig.2(e)(h)(k)) compared to males working with males occurs mostly after the 75th percentile of the productivity distribution and tends to grow with publication productivity. In line with OLS results, male students working with female advisors are not different in productivity compared to male-male pairs.[21]

We explore this evidence further in appendix F. For the sub-sample of white students working with white advisors, table 15 shows that differences in productivity of female students working with female advisors compared to male-male exists only among the most productive — top 10%, 5%, 1% (90th, 95th, 99th percentiles) of the students and ranges from a 27% to a 41% difference. We find a more heterogeneous and pronounced difference (from -20% to -47%) for the female students working with male advisors. The difference is significant also for the top 20% (80th percentile) productive students.

We find similar results in the sub-sample of black students working with black advisors (table 16) and with white advisors (table 17).

Results of the quantile regressions run on the entire population underline that the productivity difference between male and female students is most pronounced for females working with male advisors and u-shaped, in particular for females working with a female advisor. However, when student-supervisor gender is coupled with race, the gap with male-male is not u-shaped but rather downward sloping for same-race couples in particular. Where it exists only among the top productive (top 10-20%) students.

[21] However, this group displays significantly lower productivity than male-male at the 99th percentile for the sub-sample of white-white and black-black supervisions (see table 15 and 16 in appendix F).

The difference between figures for the whole sample and data sub-samples are of particular interest. These differences relate to the role of diversity or composition effects.

On the one hand, it may underline the potential role of cross-race supervision in attenuating gender differences in productivity. Our descriptive analysis in section 3.2 reveals that the most productive students are those with both cross-race and cross-gender supervision. Student-supervisor couples with those characteristics rarely form in the South African context.[22] When a similar relation is established, it is likely that both sides of the couple where involved in an active search before its formation, underlining that they might be particularly motivated in their research.

On the other hand, those differences may suggest some composition effects, reminiscent of the Simpson paradox. The Simpson paradox underlines that aggregate figures can show opposite trends to disaggregate ones. Indeed, one of the well-known instances of the paradox concerns gender or racial sorting into scientific disciplines and universities (Mullen and Baker, 2008). We test for such environmental effects at the level of university and field in table 8 in Appendix A, there were not any identifiable effects looking at gender, but we found some along racial lines.

5.1 Results looking at race

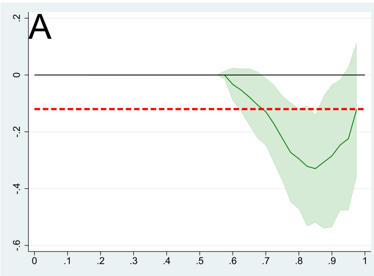
We perform the same analysis comparing black and white students and advisors for the whole sample and 3 sub-samples of gender couples: male-male (MM), female-female (FF), female students male advisors (FM). In appendix G and appendix I, we show respectively results for OLS and quantile regression estimation. The results show no difference in productivity between white and black students for all sub-samples. Our results are particularly relevant for policy in South Africa Academia. This year, in South Africa, many Ph.D. funding schemes (in social sciences in particular) are ending, and they will

[22] See Rossello and Cowan (2019) for further details on supervision tie-formation in the South African Universities.

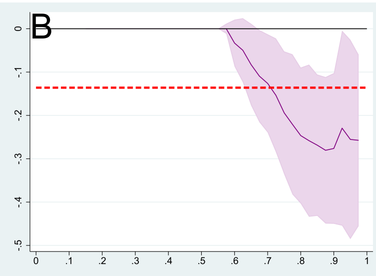
be re-discussed.[23] Funded Ph.D. programs are essential in a country with vast inequalities like South Africa. Surveys underline that black students identify financial constraints as the main reason preventing them from pursuing postgraduate education (Mouton et al., 2015). We should remark that in our context where there are no financial constraints because they are, to a great extent, removed. Indeed, doctoral programs in STEM are usually funded (Mouton et al., 2015). The fact that we do not find any difference between black and white students may underline the importance of such funding schemes, which guarantee access to postgraduate education for all.

[23]The NIHSS-SAHUDA funding program for example ends in 2020; available at `https://www.nihss.ac.za/content/nihss-sahuda-programme`. Last access December 2019.

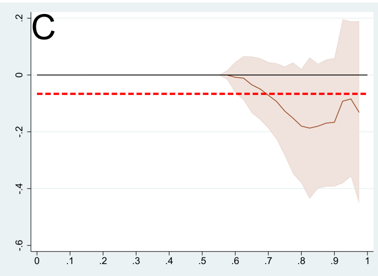
Figure 2: Quantile Regressions for student annual average doctoral productivity comparing student-advisor gender couples. Productivity is $log(1+pub_t)$, where pub_t is number of student publications between year t and $t+2$ inclusive, divided by 3. Each row shows results for a different data sample: All (A, B, C); only white student-advisor (D, E, F); only black student-advisor (G, H, I); and black student with white advisor (J, K, L). In each sub-figure, the horizontal axis represents percentiles and the vertical axis shows estimated productivity difference of student-advisor gender couple with the baseline Male-Male couple. The columns show respectively estimated coefficients for productivity difference for the dummy female-female (green), female-male (violet) and male-female (brown) student-advisor couple. Quantile regressions are done for each 2.5 percentile using robust clustered standard errors according to Machado et al. (2011) and estimates for the student-advisor gender are shown with 95% confidence intervals. The solid black line is zero, dashed red line is the (non-quantile) panel OLS estimation of Models 3 from table 2. Additional controls are: discipline, enrolment year, year, time to graduation, whether the student had published previously, whether the student have more than one advisor, the log of average publications of the advisor lagged one year. Corresponding regression tables are in Appendix section F



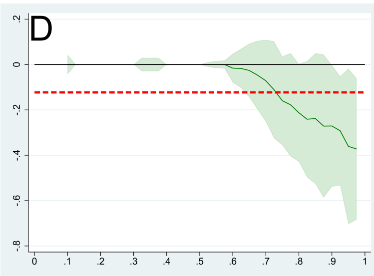
(a) All StudFemale_AdvFemale



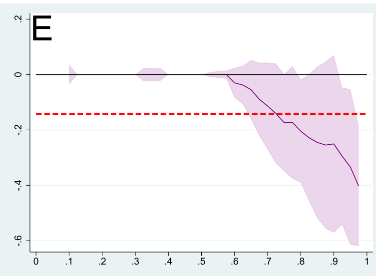
(b) All StudFemale_AdvMale



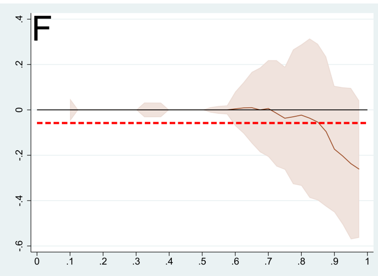
(c) All StudMale_AdvFemale



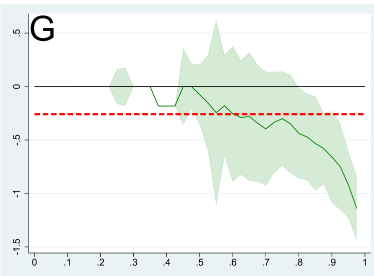
(d) WW StudFemale_AdvFemale



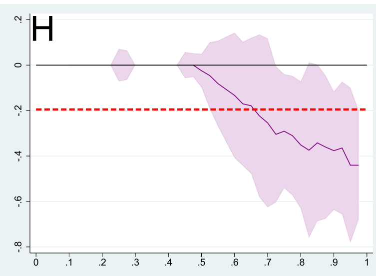
(e) WW StudFemale_AdvMale



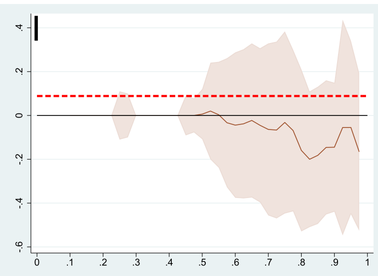
(f) WW StudMale_AdvFemale



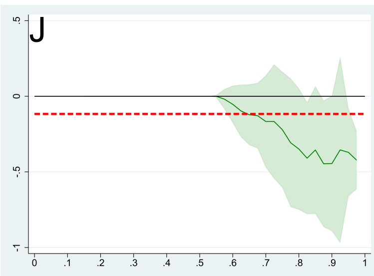
(g) BB StudFemale_AdvFemale



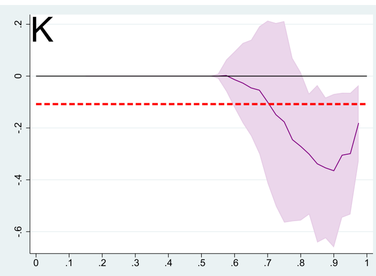
(h) BB StudFemale_AdvMale



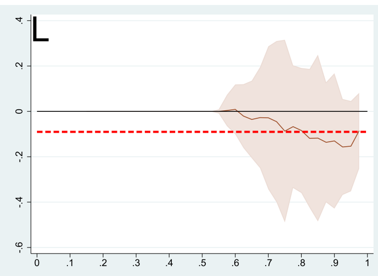
(i) BB StudMale_AdvFemale



(j) BW StudFemale_AdvFemale



(k) BW StudFemale_AdvMale



(l) BW StudMale_AdvFemale

6 Conclusion

We analyze gender productivity differences of 933 South African Ph.d.s students in STEM graduated between 2000 and 2014, we find, on average, evidence of lower publication productivity of female Ph.D. students compared to male fellows. Considering the gender composition of student-supervisor dyads or couple, we find that this difference is mostly attributable to female students working with male advisors. Female students with female advisors have publication records very similar to male students. Looking at the joint effects of gender and race, we find more significant gaps for female students when the student and the supervisor are of the same race, with again female students with male advisors having, on average, the lowest publication productivity.

Using quantile regressions to consider the productivity distribution underlying the mere average differences, we uncover two particularly striking observations. In the whole sample, female students with a high (or low) "productivity-profile" studying with female advisors are as productive as male students with a high (or low) "productivity-profile" studying with male advisors. Instead, in data sub-samples for same-race supervision (white-white and black-black couples), the gap productivity between female students compared to male students working with male advisors is mainly driven by a gap in the right-hand tails of the productivity distribution. That is, in the ”moderately productive” group of students, males and females have very similar numbers of publications per year. It is only when we look at the very highly productive (top 20%), that we find large, statistically significant, male-female disparities.

A simple restatement or reinterpretation of such a finding is that other things equal, female students are not treated as well as male students by male advisors, with less and lower quality supervision than male students. There are, of course, many other reasons that can individually and jointly account for it. We have touched on a number of them in our literature review. Going more in-depth in a real explanation is left for future work, and as usual, will need not only new data and indicators but also detailed case studies.

References

Aguinis, H., Y. H. Ji, and H. Joo (2018). Gender productivity gap among star performers in stem and other scientific fields. *Journal of Applied Psychology 103*(12), 1283.

Allison, P. D. and J. A. Stewart (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review 39*(4), 596–606.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics 42*(3), 527–554.

Bettinger, E. P. and B. T. Long (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review 95*(2), 152–157.

Blackburn, R. T., D. W. Chapman, and S. M. Cameron (1981). "cloning" in academe: Mentorship and academic careers. *Research in Higher Education 15*(4), 315–327.

Boudreau, K. J., E. C. Guinan, K. R. Lakhani, and C. Riedl (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science 62*(10), 2765–2783.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2018). Can female role models reduce the gender gap in science? evidence from classroom interventions in french high schools. *PSE Working Papers halshs-01713068*.

Chen, C., Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics 3*(3), 191–209.

Chuang, K.-Y. and Y.-S. Ho (2014). Bibliometric profile of top-cited single-author articles in the science citation index expanded. *Journal of Informetrics 8*(4), 951–962.

Cole, J. R. and H. Zuckerman (1984). The productivity puzzle. *Advances in Motivation and Achievement. Women in Science.*.

Conti, A. and F. Visentin (2015). Science and engineering ph. d. students' career outcomes, by gender. *PloS one 10*(8), e0133177.

Cowan, R. and G. Rossello (2018). Emergent structures in faculty hiring networks, and the effects of mobility on academic performance. *Scientometrics 117*(1), 527–562.

Crosby, J. R. and B. Monin (2007). Failure to warn: How student race affects warnings of potential academic difficulty. *Journal of Experimental Social Psychology 43*(4), 663–670.

David, P. A. (1993). *Positive feedbacks and research productivity in science: Reopening another black box*. MERIT Buxton.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science 47*(1), 117–132.

Fox, M. F. and C. A. Faver (1985). Men, women, and publication productivity: Patterns among social work academics. *The Sociological Quarterly 26*(4), 537–549.

Gaule, P. and M. Piacentini (2018). An advisor like me? advisor gender and post-graduate careers in science. *Research Policy 47*(4), 805–813.

Gonzalez-Sauri, M. and G. Rossello (2019). The role of early-career university prestige stratification on the future academic performance of scholars. *UNU-MERIT Working Paper Series 2019*(018).

Holman, L., D. Stuart-Fox, and C. E. Hauser (2018). The gender gap in science: How long until women are equally represented? *PLoS biology 16*(4), e2004956.

Jadidi, M., F. Karimi, H. Lietz, and C. Wagner (2018). Gender disparities in science? dropout, productivity, collaborations and success of male and female computer scientists. *Advances in Complex Systems 21*(03n04), 1750011.

Kelchtermans, S. and R. Veugelers (2011). The great divide in scientific productivity: Why the average scientist does not exist. *Industrial and Corporate Change 20*(1), 295–336.

Koberg, C. S., R. W. Boss, and E. Goodman (1998). Factors and outcomes associated with mentoring among health-care professionals. *Journal of Vocational Behavior 53*(1), 58–72.

Larivière, V. (2012). On the shoulders of students? the contribution of phd students to the advancement of knowledge. *Scientometrics 90*(2), 463–481.

Lehohla, P. (2015). Census 2011: Fertility in south africa. *Statistics South Africa. http://www. statssa. gov. za/publications/Report-03-01-63/Report-03-01-632011. pdf. Accessed September 27*, 2018.

Lerchenmueller, M. J. and O. Sorenson (2018). The gender gap in early career transitions in the life sciences. *Research Policy 47*(6), 1007–1017.

Li, W., T. Aste, F. Caccioli, and G. Livan (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications 10*(1), 1–9.

Lockwood, P. (2006). "someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly 30*(1), 36–46.

Machado, J., P. Parente, and J. S. Silva (2011, November). Qreg2: Stata module to perform quantile regression with robust and clustered standard errors. Statistical Software Components, Boston College Department of Economics.

Mairesse, J. and M. Pezzoni (2015). Does gender affect scientific productivity? *Revue économique 66*(1), 65–113.

Mairesse, J., M. Pezzoni, and F. Visentin (2019). Impact of family characteristics on the gender publication gap: evidence for physicists in france. *Interdisciplinary Science Reviews 44*(2), 204–220.

Merton, R. K. (1988). The matthew effect in science, ii: Cumulative advantage and the symbolism of intellectual property. *Isis 79*(4), 606–623.

Milkman, K. L., M. Akinola, and D. Chugh (2015). What happens before? a field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology 100*(6), 1678.

Moore, K. K., I. Cid-Martinez, J. Toney, J. A. Smith, A. C. Kalb, J. H. Shin, and R. M. Spalter-Roth (2018). Who climbs the academic ladder? race and gender stratification in a world of whiteness. *The Review of Black Political Economy 45*(3), 216–244.

Moss-Racusin, C. A., J. van der Toorn, J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman (2016). A “scientific diversity” intervention to reduce gender bias in a sample of life scientists. *CBE-LIFE SCIENCES EDUCATION 15*(3).

Mouton, B. Van Lill, V. Boshoff, Cloete, and Sheppard (2015). A study on the retention, completion and progression rates of south african postgraduate students. *Research Report*.

Mullen, A. L. and J. Baker (2008). Gender, race, and ethnic segregation of science fields in us universities. *Journal of Women and Minorities in Science and Engineering 14*(2).

Nchinda, T. C. (2002). Research capacity strengthening in the south. *Social science & medicine 54*(11), 1699–1711.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E 67*(2), 026126.

Petscher, Y. and J. A. Logan (2014). Quantile regression in the study of developmental sciences. *Child development 85*(3), 861–881.

Pezzoni, M., J. Mairesse, P. Stephan, and J. Lane (2016). Gender and the publication output of graduate students: A case study. *PLoS One 11*(1), e0145146.

Rivera León, L., J. Mairesse, and R. Cowan (2017). Gender gaps and scientific productivity in middle-income countries: Evidence from mexico. Technical report, Inter-American Development Bank.

Rossello, G. and R. Cowan (2019). Far from random? the role of homophily in student supervision. *UNU-MERIT Working Paper Series 2019*(024).

Rzhetsky, A., J. G. Foster, I. T. Foster, and J. A. Evans (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences 112*(47), 14569–14574.

Shi, F., J. G. Foster, and J. A. Evans (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks 43*, 73–85.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013). Atypical combinations and scientific impact. *Science 342*(6157), 468–472.

Van den Brink, M. and Y. Benschop (2014). Gender in academic networking: The role of gatekeepers in professorial recruitment. *Journal of Management Studies 51*(3), 460–492.

Verhoeven, D., J. Bakker, and R. Veugelers (2016). Measuring technological novelty with patent-based indicators. *Research Policy 45*(3), 707–723.

Wager, E., S. Singhvi, and S. Kleinert (2015). Too much of a good thing? an observational study of prolific authors. *PeerJ 3*, e1154.

Wang, J., R. Veugelers, and P. Stephan (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy 46*(8), 1416–1436.

Zinovyeva, N. and M. Bagues (2015). The role of connections in academic promotions. *American Economic Journal: Applied Economics 7*(2), 264–92.

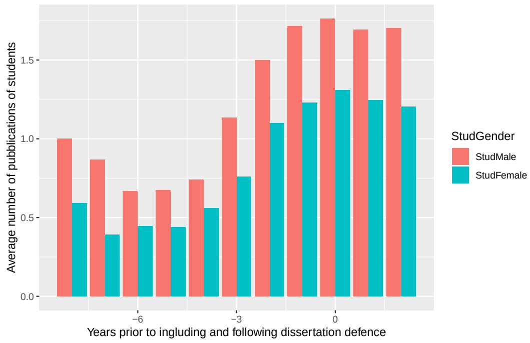
A Variables

Independent variable:

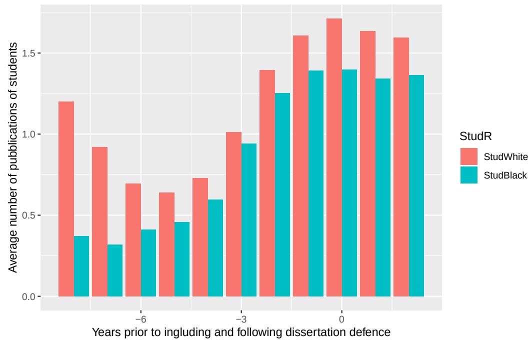
- **moreAdv** is a dummy variable equal to 1 if the students has more then one supervisor. One third of the students have more, the maximum number of advisor per student is 3
- **logprofcumavgprod** is the log of 1+ the lagged cumulative average productivity of advisor. The average cumulative number of paper is computed since the year of the first record in the publication data to t-1 and divided by the number of years.
- **DummyStudPrevPub** is a dummy equal to 1 if the student has published before. Overall the 28% of male students has already publish before starting the Ph.D.; while for female student this percentage is 25%. This suggest that the gap in publication could be originated before starting the Ph.D.
- **timegrad** time to graduation

B Additional Statistics on the data

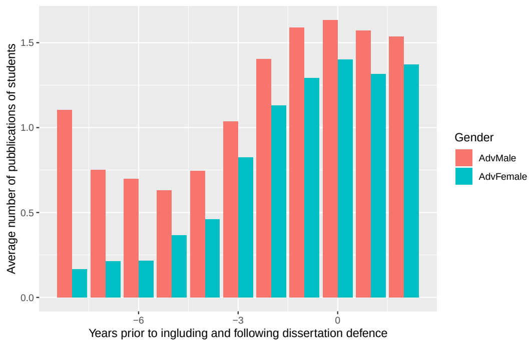
Figure 3: Average three years publications of students classified by gender(a) and race(b) of the student and gender(c) and race(d) of advisor. The average for the groups is calculated every year starting from 8 years before the thesis defence (d-8) until two years after the thesis defence (d+2). Where year of defence (d) is equal to zero.



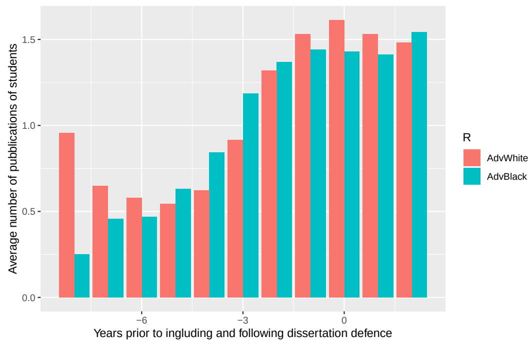
(a)



(b)



(c)



(d)

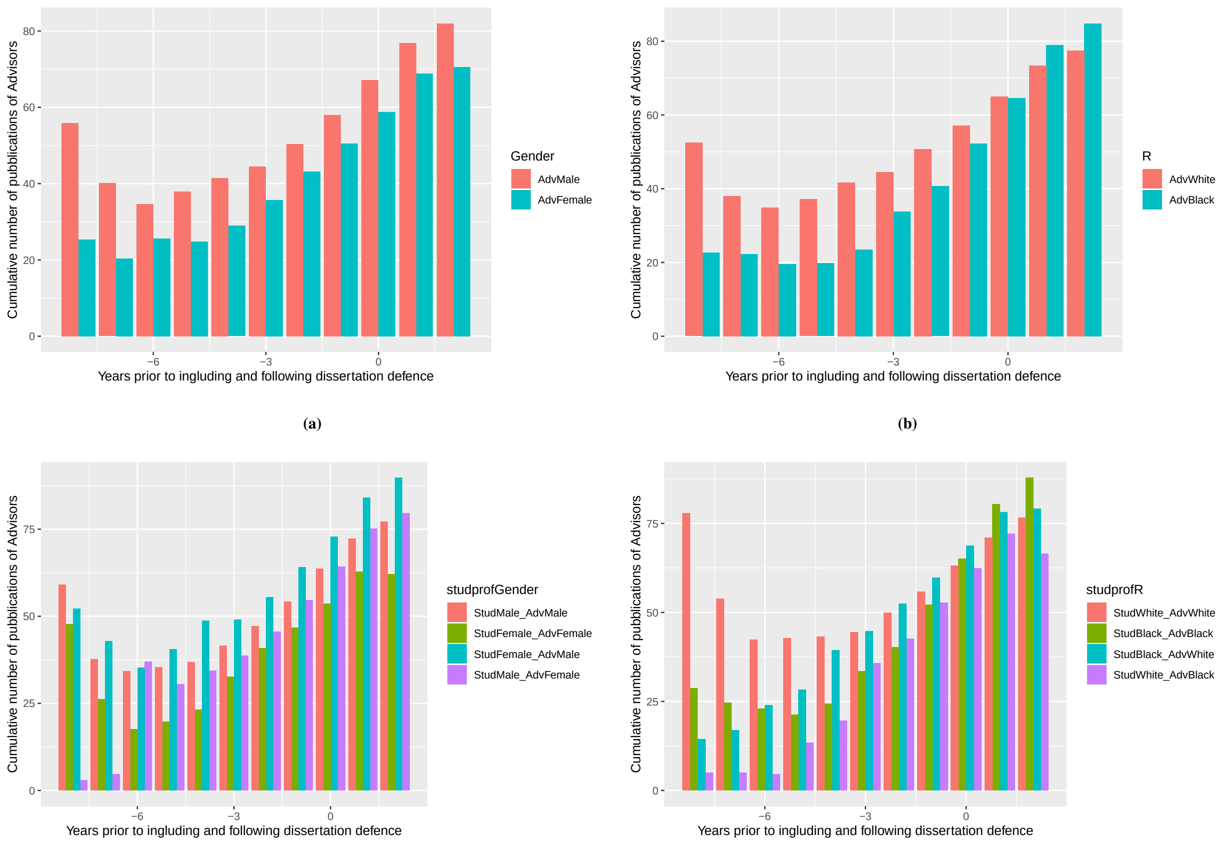
Table 3: Student 3 year average number of publications including and excluding zeros for white male, white female, black male, and black female. The logarithms are showed at the bottom of the table, excluding the zeros.

	White Male	White Female	F/M	Black Male	Black Female	F/M
			Including zeros			
Mean	1.58	1.21	0.77	1.37	0.75	0.55
Median	0.33	0		0	0	
Std dev.	2.87	3.30		2.81	2.12	
Obs.	1636	1673		1840	900	
			Excluding zeros			
Mean	3.16	2.86	0.91	2.89	1.82	0.63
Median	2.00	1.33	0.67	1.67	1.00	0.60
Std dev.	3.38	4.58		3.50	3.00	
Obs.	820	707		873	370	
			In logarithms Excluding zeros			
Mean	1.17	1.04	-0.13	1.10	0.81	-0.29
Median	1.10	0.85	-0.25	0.98	0.69	-0.29
Std dev.	0.69	0.69		0.67	0.57	
Obs.	820	707		873	370	

Table 4: Advisor Logarithm of 1+ cumulative average productivity from first record to t-1. It refers to the variable called Logprofcumavgprod

	White Male	White Female	F-M	Black Male	Black Female	F-M
			Including zeros			
Mean	1.29	1.14	-0.15	1.33	1.67	0.34
Median	1.36	1.17	-0.19	1.33	1.61	0.28
Std dev.	0.78	0.77		0.96	0.86	
Obs.	3311	1413		1120	205	
			Excluding zeros			
Mean	1.53	1.44	-0.09	1.67	1.77	0.1
Median	1.52	1.38	-0.14	1.58	1.65	0.07
Std dev.	0.59	0.57		0.76	0.77	
Obs.	2775	1108		881	193	

Figure 4: Cumulative number of publications of advisor classified by gender(a) and race(b) and by couple of student and advisor gender(c) and race(d). The average for the groups is calculated every year starting from 8 years before the thesis defence (d-8) until two years after the thesis defence (d+2). Where year of defence (d) is equal to zero.



(a) (b) (c) (d)

Table 5: Distribution of the study sample by year of thesis defence

	Student	Advisor
2001	1	1
2002	7	7
2003	23	23
2004	36	32
2005	58	54
2006	80	75
2007	76	69
2008	89	80
2009	90	77
2010	114	93
2011	117	106
2012	127	110
2013	91	79
2014	15	14

Table 6: Distribution of Ph.D. students and advisors by the discipline of the thesis.

	Student	Black Stud.	White Stud.	Female Stud.	Male Stud	Advisor	Black Adv.	White Adv.	Female Adv.	Male Adv.
Agricultural sciences	90	46	44	44	46	63	15	48	45	18
Biological sciences	278	96	182	154	124	142	19	123	92	50
Chemical sciences	117	66	51	73	44	49	16	33	38	11
Earth and marine sciences	67	28	39	35	32	44	6	38	36	8
Engineering sciences	69	26	43	51	18	51	10	41	46	5
Health Sciences	81	33	48	34	47	59	18	41	33	26
Information and Computer science	31	5	26	16	15	21	3	18	17	4
Mathematical sciences	24	14	10	18	6	20	6	14	19	1
Medical sciences: Basic	43	25	18	17	26	28	11	17	20	8
Medical sciences: Clinical	14	7	7	6	8	12	2	10	6	6
Pharmaceutical Sciences	15	15	0	10	5	3	2	1	3	0
Physical sciences	55	37	18	49	6	36	9	27	31	5
Technologies and applied sciences	40	18	22	24	16	24	4	20	18	6

Table 7: Average productivity for different sub-sample of the data and student advisor couple. Productivity is computed as $\log(1+pub_t)$ where pub_t is number of student publication between years t and $t+2$ inclusive.

		Advisor		
		Male	Female	Average
Student	Male	0.65	0.44	0.61
	Female	0.44	0.46	0.44
	Average	0.56	0.45	0.53

(a) Same-race pair, white student white advisor

		Advisor		
		White	Black	Average
Student	White	0.65	0.29	0.63
	Black	0.46	0.6	0.52
	Average	0.57	0.56	0.57

(b) Same-gender pair, male student male advisor

		Advisor		
		Male	Female	Average
Student	Male	0.6	0.54	0.59
	Female	0.35	0.55	0.37
	Average	0.51	0.54	0.52

(c) Same-race pair, black student black advisor

		Advisor		
		White	Black	Average
Student	White	0.44	0.09	0.42
	Black	0.19	0.55	0.23
	Average	0.38	0.26	0.37

(d) Same-gender pair, female student female advisor

		Advisor		
		Male	Female	Average
Student	Male	0.46	0.53	0.48
	Female	0.37	0.19	0.31
	Average	0.43	0.4	0.42

(e) Cross-race pair, black student white advisor

		Advisor		
		White	Black	Average
Student	White	0.44	0.57	0.45
	Black	0.37	0.35	0.36
	Average	0.43	0.39	0.42

(f) Cross-gender pair, female student male advisor

Table 8: Gender and Racial Assortativity Coefficient (Ass.) by universities and field. The assortativity coefficient is computed according to Newman (2003), while 95% confidence intervals are computed simulating 1000 times type-blind tie formation given supervision and population composition.

	Assortativity coefficient by Universities					
Univerisity	Ass. Gender	sign	95% CI Null Model	Ass. Race	sign	95% CI Null Model
CapePeninsulaUniversityOfTechnology;	-0.40		(-0.75 ; 0.65)	0.00		(-1.00 ; 1.00)
DurbanInstituteOfTechnology;	-0.39		(-0.57 ; 0.48)	0.76	*	(-0.65 ; 0.53)
NelsonMandelaMetropolitanUniversity;	0.30		(-0.41 ; 0.38)	0.33	*	(-0.33 ; 0.33)
NorthWestUniversity;	0.54	*	(-0.38 ; 0.36)	0.47	*	(-0.47 ; 0.47)
RhodesUniversity;	-0.11		(-0.28 ; 0.26)	0.26	*	(-0.16 ; 0.17)
TshwaneUniversityOfTechnology;	-0.45	*	(-0.45 ; 0.42)	0.60	*	(-0.47 ; 0.47)
UniversityOfCapeTown;	0.10		(-0.14 ; 0.13)	0.19	*	(-0.13 ; 0.12)
UniversityOfFortHare;	0.00		(0.00 ; 0.00)	0.66	*	(-0.89 ; 0.66)
UniversityOfJohannesburg;	0.17		(-0.39 ; 0.37)	0.47	*	(-0.33 ; 0.29)
UniversityOfKwaZuluNatal;	0.18		(-0.28 ; 0.28)	0.46	*	(-0.30 ; 0.33)
UniversityOfLimpopo;	0.00		(0.00 ; 0.00)	1.00		NA
UniversityOfPretoria;	0.11		(-0.14 ; 0.15)	0.20	*	(-0.15 ; 0.13)
UniversityOfSouthAfrica;	1.00		(-1.00 ; 1.00)	0.00		(0.00 ; 0.00)
UniversityOfStellenbosch;	0.05		(-0.14 ; 0.14)	0.20	*	(-0.13 ; 0.14)
UniversityOfTheFreeState;	0.22	*	(-0.20 ; 0.22)	0.05		(-0.11 ; 0.09)
UniversityOfTheWesternCape;	-0.07		(-0.23 ; 0.25)	0.02		(-0.24 ; 0.22)
UniversityOfVenda;	0.33		(-1.00 ; 1.00)	0.33		(-1.00 ; 1.00)
UniversityOfWitwatersrand;	0.16		(-0.56 ; 0.40)	0.00		(-0.31 ; 0.31)
UniversityOfZululand;	1.00		NA	1.00		NA
VaalUniversityOfTechnology;	1.00		NA	1.00		NA

	Assortativity coefficient by Field					
Field	Ass. Gender	sign	95% CI Null Model	Ass. Race	sign	95% CI Null Model
Agricultural sciences	0.14		(-0.17 ; 0.19)	0.52	*	(-0.19 ; 0.17)
Biological sciences	0.06		(-0.12 ; 0.12)	0.14	*	(-0.11 ; 0.11)
Chemical sciences	-0.07		(-0.19 ; 0.17)	0.46	*	(-0.16 ; 0.17)
Earth and marine sciences	0.05		(-0.20 ; 0.17)	0.19	*	(-0.19 ; 0.19)
Engineering sciences	0.11		(-0.22 ; 0.20)	0.52	*	(-0.23 ; 0.22)
Health Sciences	0.15		(-0.21 ; 0.20)	0.28	*	(-0.20 ; 0.20)
Information and Computer science	0.08		(-0.32 ; 0.28)	0.63	*	(-0.50 ; 0.50)
Mathematical sciences	0.24	*	(-0.36 ; 0.24)	0.14		(-0.33 ; 0.38)
Medical sciences: Basic	0.02		(-0.24 ; 0.23)	0.55	*	(-0.30 ; 0.28)
Medical sciences: Clinical	0.29		(-0.57 ; 0.57)	0.29		(-0.43 ; 0.43)
Pharmaceutical Sciences	0.00		(0.00 ; 0.00)	0.40		(-0.80 ; 0.60)
Physical sciences	0.16		(-0.42 ; 0.42)	0.26	*	(-0.18 ; 0.19)
Technologies and applied sciences	0.06		(-0.28 ; 0.28)	0.41	*	(-0.23 ; 0.25)

Figure 5: Three years number of publications of students classified by student gender/race (top) and couple student-advisor gender/race of advisor (bottom). The average for the groups is calculated every year starting from 8 years before the thesis defence (d-8) until two years after the thesis defence (d+2).

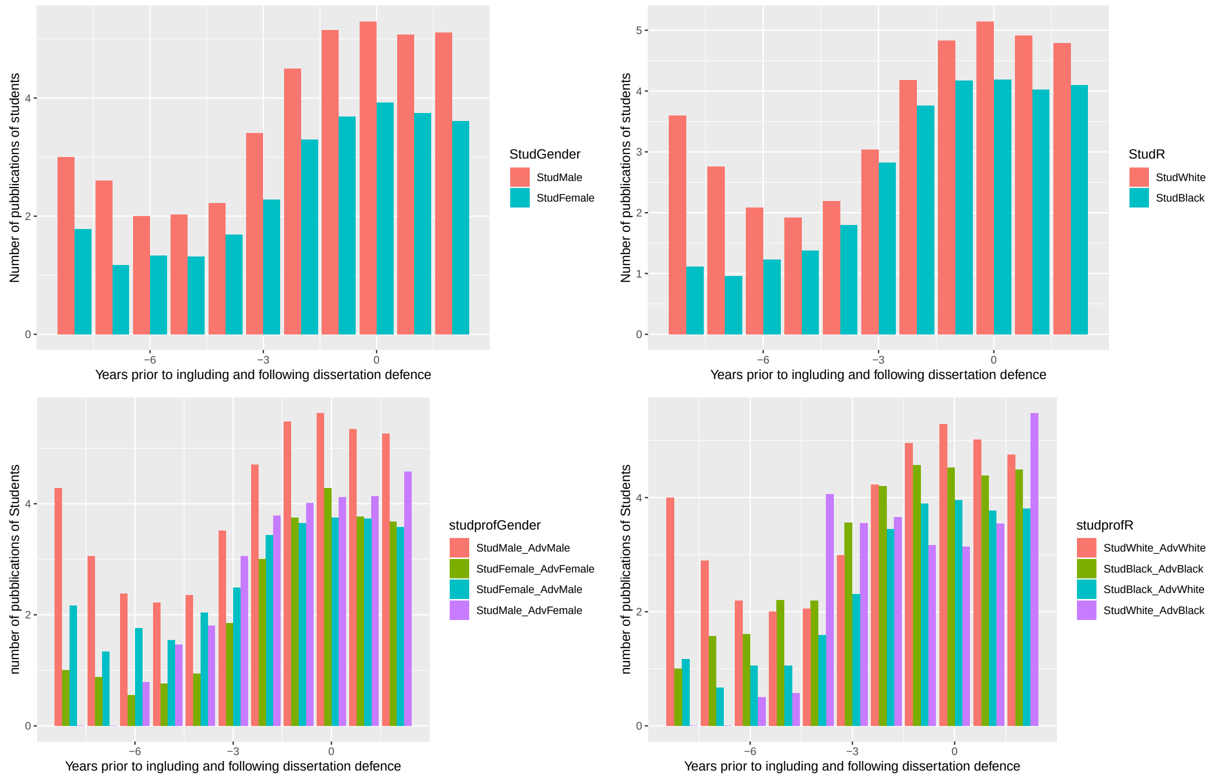
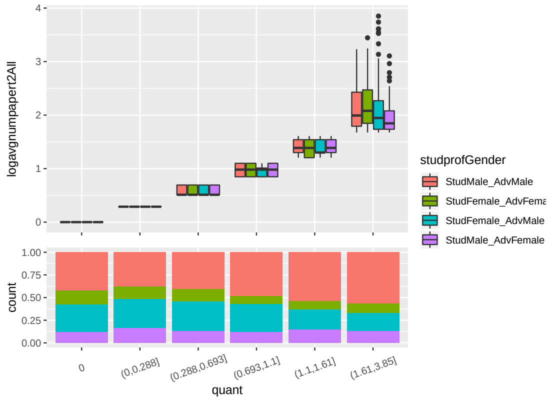
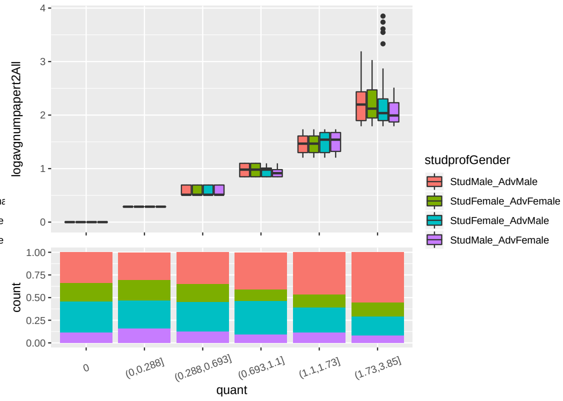


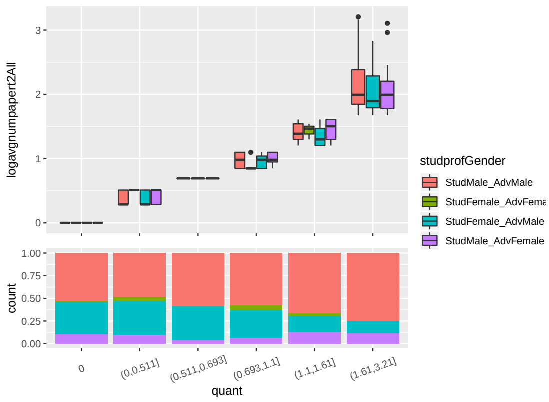
Figure 6: Distribution of the dependent variable over the deciles of its distribution by student-advisor gender couple. The bottom stack-plots represent the relative proportions of the gender couple in the population for each decile. (a) Whole sample, (b) sub-sample of white students with white advisors, (c) sub-sample of black students with black advisors, and (c) sub-sample of black students with white advisors.



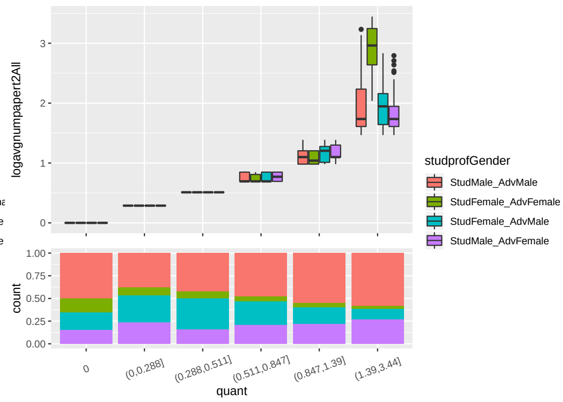
(a) All



(b) WW



(c) BB



(d) BW

C Regression with gender and race interaction on the whole sample

Table 9: Pooled OLS Panel Regression with robust clustered standard error on the whole sample with interaction terms of gender and race. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Models (1) for the Students comparison; Model (2) Advisor comparison; Model (3) Couples Comparison. Additional controls are moreAdv, logprofcumavgprod, DummyStudPrevPub, timegrad, field, and enrolment year.

	(1)	(2)	(3)
StudFemale	-0.105*		
	(0.0480)		
StudBlack	-0.00433		
	(0.0497)		
StudFemaleBlack	-0.0277		
	(0.0703)		
AdvFemale		-0.0291	
		(0.0452)	
AdvBlack		-0.000940	
		(0.0504)	
AdvFemaleBlack		-0.0787	
		(0.113)	
StudFemale_AdvFemale			-0.0855
			(0.0716)
StudFemale_AdvMale			-0.128*
			(0.0594)
StudMale_AdvFemale			-0.0508
			(0.0862)
StudBlack_AdvBlack			0.0154
			(0.0767)
StudBlack_AdvWhite			-0.0115
			(0.0665)
StudWhite_AdvBlack			-0.0000968
			(0.129)
StudFemale_AdvFemale # StudBlack_AdvBlack			-0.386
			(0.213)
StudFemale_AdvFemale # StudBlack_AdvWhite			-0.0425
			(0.139)
StudFemale_AdvFemale # StudWhite_AdvBlack			-0.229
			(0.151)
StudFemale_AdvMale # StudBlack_AdvBlack			-0.0572
			(0.106)
StudFemale_AdvMale # StudBlack_AdvWhite			0.0234
			(0.0968)
StudFemale_AdvMale # StudWhite_AdvBlack			-0.0577
			(0.263)
StudMale_AdvFemale # StudBlack_AdvBlack			0.0102
			(0.189)
StudMale_AdvFemale # StudBlack_AdvWhite			-0.0446
			(0.120)
StudMale_AdvFemale # StudWhite_AdvBlack			-0.185
			(0.219)
Constant	0.516***	0.393**	0.409***
	(0.150)	(0.140)	(0.112)
N	6049	6049	6049
R^2	0.285	0.281	0.287

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

D Main OLS regressions

Table 10: Pooled OLS Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Models (1) for the Students comparison; Model (2) Advisor comparison; Model (3) Couples Comparison. Where Columns (a) On the sub-sample White Student White Professors; Columns (b) On the sub-sample Black Student Black Professors; Columns (c) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	(1) ALL	(1)a WW	(1)b BB	(1)c BW	(2) ALL	(2)a WW	(2)b BB	(2)c BW	(3) ALL	(3)a WW	(3)b BB	(3)c BW
StudFemale	-0.113**	-0.116*	-0.217*	-0.0908								
	(0.0357)	(0.0497)	(0.0847)	(0.0686)								
AdvFemale					-0.0393	-0.0248	0.0989	-0.0678				
					(0.0416)	(0.0589)	(0.151)	(0.0720)				
StudFemale_AdvFemale									-0.120*	-0.113	-0.270	-0.123
									(0.0571)	(0.0745)	(0.239)	(0.129)
StudFemale_AdvMale									-0.136**	-0.142*	-0.198*	-0.112
									(0.0417)	(0.0580)	(0.0958)	(0.0810)
StudMale_AdvFemale									-0.0733	-0.0622	0.0768	-0.0931
									(0.0568)	(0.0899)	(0.172)	(0.0839)
moreAdv	0.0585	0.133*	0.143	-0.0929	0.0525	0.134*	0.0934	-0.0974	0.0583	0.131*	0.143	-0.0938
	(0.0431)	(0.0621)	(0.108)	(0.0675)	(0.0434)	(0.0629)	(0.113)	(0.0678)	(0.0430)	(0.0628)	(0.113)	(0.0684)
logprofcumavgprod	0.122***	0.151***	0.0544	0.143*	0.122***	0.153***	0.0492	0.140*	0.122***	0.151***	0.0528	0.143*
	(0.0286)	(0.0369)	(0.0651)	(0.0643)	(0.0287)	(0.0370)	(0.0666)	(0.0643)	(0.0287)	(0.0370)	(0.0664)	(0.0646)
DummyStudPrevPub	0.791***	0.825***	0.559***	0.772***	0.796***	0.834***	0.579***	0.790***	0.790***	0.822***	0.566***	0.788***
	(0.0443)	(0.0608)	(0.101)	(0.0921)	(0.0443)	(0.0611)	(0.101)	(0.0933)	(0.0442)	(0.0608)	(0.102)	(0.0952)
timegrad	-0.0401***	-0.0387*	-0.0511	-0.0236	-0.0401***	-0.0383*	-0.0594	-0.0236	-0.0395**	-0.0388*	-0.0525	-0.0201
	(0.0121)	(0.0168)	(0.0340)	(0.0241)	(0.0122)	(0.0167)	(0.0348)	(0.0244)	(0.0121)	(0.0168)	(0.0343)	(0.0241)
Constant	0.520***	0.294*	1.535**	0.489*	0.407***	0.148	1.190*	0.463*	0.421***	0.197	1.323**	0.416*
	(0.120)	(0.144)	(0.493)	(0.204)	(0.118)	(0.155)	(0.483)	(0.209)	(0.108)	(0.132)	(0.465)	(0.187)
N	6049	3083	1099	1641	6049	3083	1099	1641	6049	3083	1099	1641
R^2	0.284	0.354	0.260	0.263	0.280	0.349	0.244	0.260	0.285	0.355	0.260	0.264

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

E Poisson panel regressions Gender

E.1 Students

Table 11: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	ALL	(1) WW	(2) BB	(3) BW
StudFemale	-0.408***	-0.485**	-0.871**	-0.547*
	(0.124)	(0.166)	(0.324)	(0.243)
moreAdv	0.0734	0.264	0.292	-0.317
	(0.140)	(0.218)	(0.294)	(0.290)
logprofcumavgprod	0.427***	0.468***	0.186	0.805**
	(0.0965)	(0.135)	(0.201)	(0.264)
DummyStudPrevPub	1.655***	1.896***	1.448***	2.071***
	(0.117)	(0.165)	(0.278)	(0.291)
timegrad	-0.148***	-0.0920	-0.201	0.0476
	(0.0424)	(0.0559)	(0.124)	(0.0881)
Constant	1.415***	0.257	4.766***	0.447
	(0.398)	(0.530)	(0.798)	(0.704)
/				
lnalpha	1.471***	1.375*	1.199	1.442
	(0.400)	(0.547)	(0.902)	(0.838)
N	6049	3083	1099	1641
Standard errors in parentheses				
* p<0.05	** p<0.01	*** p<0.001"		

E.2 Advisors

Table 12: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	ALL	(1) WW	(2) BB	(3) BW
AdvFemale	-0.00650	-0.0205	-0.0947	-0.302
	(0.155)	(0.187)	(0.382)	(0.294)
moreAdv	0.0960	0.260	0.0887	-0.309
	(0.142)	(0.221)	(0.291)	(0.302)
logprofcumavgprod	0.428***	0.468***	0.187	0.802**
	(0.0966)	(0.135)	(0.201)	(0.266)
DummyStudPrevPub	1.664***	1.932***	1.410***	2.116***
	(0.117)	(0.168)	(0.277)	(0.309)
timegrad	-0.150***	-0.0870	-0.264*	0.0753
	(0.0419)	(0.0534)	(0.129)	(0.102)
Constant	0.919*	-0.468	4.069***	0.0270
	(0.404)	(0.519)	(0.901)	(0.717)
/				
lnalpha	1.481***	1.391*	1.230	1.451
	(0.398)	(0.548)	(0.897)	(0.833)
N	6049	3083	1099	1641
Standard errors in parentheses				
* p<0.05	** p<0.01	*** p<0.001"		

E.3 Student and Advisor couple

Table 13: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	ALL	(1) WW	(2) BB	(3) BW
StudFemale_AdvFemale	-0.335	-0.420	-1.303**	-0.992
	(0.209)	(0.216)	(0.505)	(0.683)
StudFemale_AdvMale	-0.437**	-0.504*	-0.828*	-0.581*
	(0.143)	(0.202)	(0.345)	(0.279)
StudMale_AdvFemale	0.0162	0.0635	-0.0460	-0.362
	(0.200)	(0.292)	(0.414)	(0.313)
moreAdv	0.0755	0.262	0.267	-0.309
	(0.140)	(0.218)	(0.298)	(0.291)
logprofcumavgprod	0.427***	0.468***	0.187	0.801**
	(0.0965)	(0.136)	(0.201)	(0.265)
DummyStudPrevPub	1.658***	1.900***	1.480***	2.135***
	(0.117)	(0.167)	(0.283)	(0.305)
timegrad	-0.146***	-0.0903	-0.207	0.0915
	(0.0421)	(0.0551)	(0.127)	(0.100)
Constant	0.997**	-0.245	3.884***	-0.0736
	(0.359)	(0.480)	(0.747)	(0.652)
/				
lnalpha	1.471***	1.375*	1.198	1.437
	(0.400)	(0.548)	(0.902)	(0.839)
N	6049	3083	1099	1641

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

F Quantile Regressions Gender

Table 14: Quantile regression with clustered standard errors on the complete sample. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(1)25th	(1)50th	(1)70th	(1)75th	(1)80th	(1)85th	(1)90th	(1)95th	(1)99th
StudFemale_AdvFemale	-2.44e-15	-3.18e-15	-0.129*	-0.223**	-0.295**	-0.330***	-0.285*	-0.224	-0.0191
	(3.76e-15)	(5.17e-15)	(0.0605)	(0.0766)	(0.0905)	(0.0969)	(0.128)	(0.129)	(0.108)
StudFemale_AdvMale	-2.79e-16	5.04e-15	-0.126*	-0.194**	-0.247**	-0.269**	-0.276**	-0.255*	-0.195**
	(3.22e-15)	(2.63e-15)	(0.0578)	(0.0723)	(0.0801)	(0.0832)	(0.0887)	(0.117)	(0.0754)
StudMale_AdvFemale	3.64e-16	-5.11e-16	-0.0713	-0.127	-0.180	-0.180	-0.166	-0.0843	-0.110
	(4.93e-15)	(3.95e-15)	(0.0596)	(0.0803)	(0.103)	(0.112)	(0.115)	(0.139)	(0.0950)
moreAdv	-2.09e-15	1.60e-15	0.0767	0.0574	0.0726	0.0830	0.0669	0.105	0.00277
	(3.54e-15)	(2.96e-15)	(0.0568)	(0.0626)	(0.0818)	(0.0848)	(0.100)	(0.116)	(0.0733)
logprofcumavgprod	6.71e-15	-2.11e-15	0.127***	0.138***	0.154***	0.181***	0.209***	0.219***	0.175**
	(3.84e-15)	(1.75e-15)	(0.0383)	(0.0352)	(0.0398)	(0.0476)	(0.0512)	(0.0528)	(0.0549)
DummyStudPrevPub	0.511***	0.981***	1.069***	1.031***	0.968***	0.946***	0.927***	0.913***	0.752***
	(1.91e-14)	(2.16e-14)	(0.0921)	(0.0789)	(0.0740)	(0.0880)	(0.0890)	(0.123)	(0.0669)
timegrad	3.38e-15	8.65e-16	-0.0117	-0.0311	-0.0527**	-0.0790***	-0.104***	-0.145***	-0.179***
	(2.13e-15)	(1.11e-15)	(0.0145)	(0.0176)	(0.0186)	(0.0209)	(0.0261)	(0.0272)	(0.0162)
N	6012	6012	6012	6012	6012	6012	6012	6012	6012
R^2	0.246	0.230	0.306	0.312	0.312	0.303	0.294	0.272	0.196
Standard errors in parentheses									
* p<0.05	** p<0.01	*** p<0.001"							

Table 15: Quantile regression with clustered standard errors on the sub-sample of White Student and White Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(1)25th	(1)50th	(1)70th	(1)75th	(1)80th	(1)85th	(1)90th	(1)95th	(1)99th
StudFemale_AdvFemale	-2.45e-15	6.67e-15	-0.0715	-0.160	-0.213	-0.238	-0.271*	-0.361*	-0.409***
	(2.01e-15)	(0.00646)	(0.0921)	(0.0998)	(0.110)	(0.147)	(0.137)	(0.175)	(0.100)
StudFemale_AdvMale	-1.61e-15	3.67e-15	-0.114	-0.174	-0.204*	-0.245	-0.250	-0.334*	-0.467***
	(1.56e-15)	(0.00471)	(0.0801)	(0.0895)	(0.0949)	(0.139)	(0.163)	(0.143)	(0.0749)
StudMale_AdvFemale	1.24e-16	2.99e-15	0.00621	-0.0367	-0.0230	-0.0541	-0.173	-0.237	-0.386***
	(2.41e-15)	(0.00698)	(0.109)	(0.116)	(0.159)	(0.176)	(0.142)	(0.170)	(0.0588)
moreAdv	4.20e-15	1.03e-14	0.0991	0.112	0.122	0.129	0.0814	0.00501	-0.0306
	(2.54e-15)	(0.00564)	(0.0679)	(0.0885)	(0.0927)	(0.117)	(0.169)	(0.133)	(0.0710)
logprofcumavgprod	-3.49e-16	-5.02e-16	0.108*	0.145**	0.162**	0.181*	0.236*	0.192**	0.184***
	(8.33e-16)	(0.00286)	(0.0425)	(0.0516)	(0.0583)	(0.0794)	(0.0951)	(0.0664)	(0.0216)
DummyStudPrevPub	0.511***	1.099***	1.192***	1.122***	1.067***	0.958***	0.924***	0.758***	0.495***
	(4.34e-15)	(0.0328)	(0.114)	(0.123)	(0.118)	(0.116)	(0.171)	(0.125)	(0.0711)
timegrad	-2.31e-16	3.72e-17	-0.00440	-0.0214	-0.0396*	-0.0625**	-0.0920**	-0.135***	-0.176***
	(4.39e-16)	(0.00169)	(0.0162)	(0.0162)	(0.0191)	(0.0211)	(0.0337)	(0.0361)	(0.0118)
N	3058	3058	3058	3058	3058	3058	3058	3058	3058
R^2	0.312	0.298	0.364	0.376	0.380	0.373	0.357	0.317	0.247
Standard errors in parentheses									
* p<0.05	** p<0.01	*** p<0.001"							

Table 16: Quantile regression with clustered standard errors on the subsample of Black Student and Black Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(2)25th	(2)50th	(2)70th	(2)75th	(2)80th	(2)85th	(2)90th	(2)95th	(2)99th
StudFemale_AdvFemale	1.09e-14	-0.0751	-0.396	-0.300	-0.438*	-0.532*	-0.659**	-0.919***	-0.865***
	(0.0840)	(0.144)	(0.272)	(0.228)	(0.218)	(0.224)	(0.218)	(0.160)	(0.0788)
StudFemale_AdvMale	1.99e-15	-0.0245	-0.254	-0.291*	-0.351*	-0.342	-0.376**	-0.440*	-0.310***
	(0.0364)	(0.0381)	(0.189)	(0.128)	(0.144)	(0.176)	(0.133)	(0.174)	(0.0426)
StudMale_AdvFemale	3.74e-15	0.00603	-0.0638	-0.0319	-0.159	-0.182	-0.145	-0.0549	-0.355***
	(0.0561)	(0.0586)	(0.200)	(0.210)	(0.187)	(0.160)	(0.150)	(0.202)	(0.0365)
moreAdv	6.48e-17	0.0172	0.104	0.134	0.0726	0.0446	0.0712	-0.0222	-0.0622**
	(0.0474)	(0.0443)	(0.146)	(0.134)	(0.148)	(0.170)	(0.111)	(0.108)	(0.0234)
logprofcumavgprod	-2.52e-15	0.0379	0.127	0.118	0.133	0.104	0.0796	0.103*	0.277***
	(0.0191)	(0.0370)	(0.0797)	(0.0742)	(0.0836)	(0.0696)	(0.0617)	(0.0450)	(0.0176)
DummyStudPrevPub	0.490***	0.801***	0.692**	0.571***	0.572**	0.654***	0.697***	0.655***	0.629***
	(0.0882)	(0.126)	(0.220)	(0.155)	(0.174)	(0.198)	(0.144)	(0.103)	(0.0353)
timegrad	-8.63e-16	0.0158	-0.0179	-0.0596	-0.0897	-0.108*	-0.121*	-0.178***	-0.185***
	(0.0116)	(0.0236)	(0.0858)	(0.0644)	(0.0522)	(0.0546)	(0.0503)	(0.0456)	(0.00787)
N	1091	1091	1091	1091	1091	1091	1091	1091	1091
R^2	0.166	0.177	0.273	0.268	0.240	0.244	0.234	0.215	0.187

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 17: Quantile regression with clustered standard errors on the subsample of Black Student and White Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(3)25th	(3)50th	(3)70th	(3)75th	(3)80th	(3)85th	(3)90th	(3)95th	(3)99th
StudFemale_AdvFemale	1.82e-15	3.80e-15	-0.167	-0.221	-0.349	-0.355	-0.445	-0.371*	-0.433***
	(1.36e-15)	(3.97e-15)	(0.155)	(0.195)	(0.205)	(0.215)	(0.228)	(0.149)	(0.0484)
StudFemale_AdvMale	3.03e-16	7.79e-16	-0.100	-0.176	-0.271	-0.338*	-0.365*	-0.299*	-0.157***
	(6.21e-16)	(3.07e-15)	(0.160)	(0.198)	(0.139)	(0.155)	(0.149)	(0.119)	(0.0329)
StudMale_AdvFemale	-2.95e-16	-2.00e-17	-0.0283	-0.0869	-0.0843	-0.118	-0.130	-0.153	-0.0521
	(1.09e-15)	(4.79e-15)	(0.161)	(0.205)	(0.141)	(0.187)	(0.151)	(0.101)	(0.0405)
moreAdv	-3.22e-16	-1.25e-15	-0.0839	-0.135	-0.165	-0.190	-0.271*	-0.179	-0.0589
	(7.26e-16)	(2.80e-15)	(0.114)	(0.121)	(0.103)	(0.0971)	(0.124)	(0.105)	(0.0393)
logprofcumavgprod	5.65e-16	2.69e-16	0.139	0.186	0.191*	0.163*	0.111	0.0983	0.123***
	(5.41e-16)	(1.91e-15)	(0.109)	(0.103)	(0.0813)	(0.0806)	(0.0936)	(0.0544)	(0.0147)
DummyStudPrevPub	0.288	0.847***	0.924***	0.931***	0.927***	0.984***	1.038***	1.030***	0.900***
	(.)	(1.90e-13)	(0.254)	(0.273)	(0.207)	(0.255)	(0.212)	(0.115)	(0.0446)
timegrad	2.01e-16	7.76e-16	0.0109	0.000221	-0.00841	-0.0159	-0.0600	-0.105**	-0.129***
	(1.81e-16)	(1.38e-15)	(0.0560)	(0.0520)	(0.0492)	(0.0546)	(0.0539)	(0.0323)	(0.00932)
N	1637	1637	1637	1637	1637	1637	1637	1637	1637
R^2	0.152	0.186	0.281	0.287	0.275	0.274	0.243	0.178	0.147

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

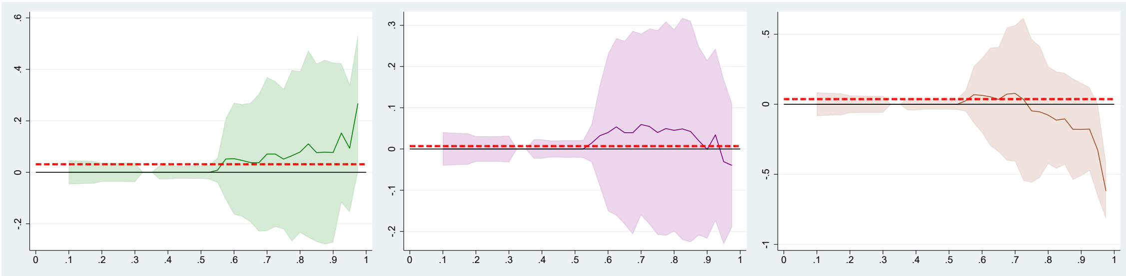
G Race Results

Table 18: Pooled OLS Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Models (1) for the Students comparison; Model (2) Advisor comparison; Model (3) Couples Comparison. Where Columns (a) On the sub-sample Male Student Male Professors; Columns (b) On the sub-sample Female Student Female Professors; Columns (c) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

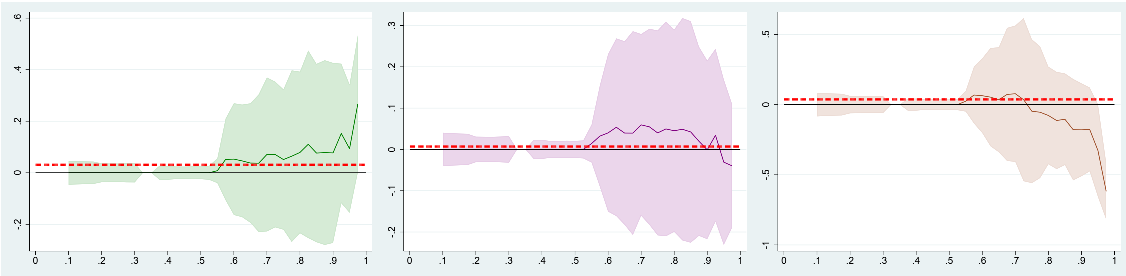
	(1) ALL	(1)a MM	(1)b FF	(1)c FM	(2) ALL	(2)a MM	(2)b FF	(2)c FM	(3) ALL	(3)a MM	(3)b FF	(3)c FM
StudBlack	0.00637	0.0175	-0.0567	-0.0128								
	(0.0366)	(0.0581)	(0.102)	(0.0623)								
AdvBlack					-0.00875	0.0439	-0.371*	-0.0533				
					(0.0455)	(0.0657)	(0.156)	(0.0766)				
StudBlack_AdvBlack									0.00567	0.0467	-0.482	-0.0571
									(0.0526)	(0.0784)	(0.259)	(0.0772)
StudBlack_AdvWhite									-0.00308	0.00510	-0.00275	0.0314
									(0.0416)	(0.0687)	(0.111)	(0.0731)
StudWhite_AdvBlack									-0.0787	0.0433	-0.320*	0.0288
									(0.0932)	(0.124)	(0.153)	(0.246)
moreAdv	0.0515	0.129	-0.00657	0.0402	0.0515	0.133	-0.0442	0.0442	0.0518	0.133	-0.0462	0.0419
	(0.0436)	(0.0744)	(0.0803)	(0.0798)	(0.0434)	(0.0747)	(0.0790)	(0.0806)	(0.0437)	(0.0749)	(0.0822)	(0.0821)
logprofcumavgprod	0.122***	0.182***	0.0714	0.0649	0.122***	0.183***	0.0749	0.0637	0.122***	0.183***	0.0735	0.0638
	(0.0287)	(0.0502)	(0.0455)	(0.0530)	(0.0287)	(0.0503)	(0.0453)	(0.0529)	(0.0288)	(0.0503)	(0.0456)	(0.0532)
DummyStudPrevPub	0.798***	0.853***	0.835***	0.743***	0.797***	0.853***	0.863***	0.741***	0.796***	0.854***	0.876***	0.741***
	(0.0445)	(0.0682)	(0.114)	(0.0890)	(0.0444)	(0.0671)	(0.110)	(0.0909)	(0.0449)	(0.0692)	(0.115)	(0.0889)
timegrad	-0.0407***	-0.0640***	-0.0470	-0.00598	-0.0405***	-0.0637***	-0.0392	-0.00599	-0.0400***	-0.0637***	-0.0388	-0.00641
	(0.0121)	(0.0189)	(0.0263)	(0.0192)	(0.0121)	(0.0188)	(0.0270)	(0.0192)	(0.0121)	(0.0191)	(0.0274)	(0.0192)
Constant	0.348**	0.414*	0.315	0.250	0.367***	0.379*	0.572**	0.298	0.351***	0.422*	0.213	0.244
	(0.111)	(0.185)	(0.220)	(0.195)	(0.109)	(0.182)	(0.207)	(0.206)	(0.104)	(0.167)	(0.197)	(0.180)
N	6049	2683	825	1748	6049	2683	825	1748	6049	2683	825	1748
R^2	0.279	0.317	0.417	0.298	0.279	0.317	0.433	0.297	0.279	0.317	0.435	0.299

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

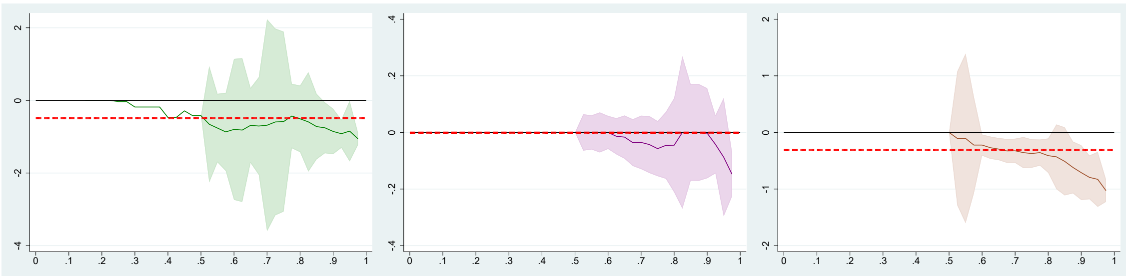
Figure 7: Quantile Regression with clustered standard errors. Results for group comparison where the baseline group is White Student with White Advisor. Quantile regressions are done for each 2.5 percentile. Full lines is zero, dotted lines are panel OLS estimation of Models (3) in table 18. Additional controls are: discipline, enrolment year, year, time to graduation, whether the student had published previously, whether the student have more than one advisor, the log of average publications of the advisor lagged one year.



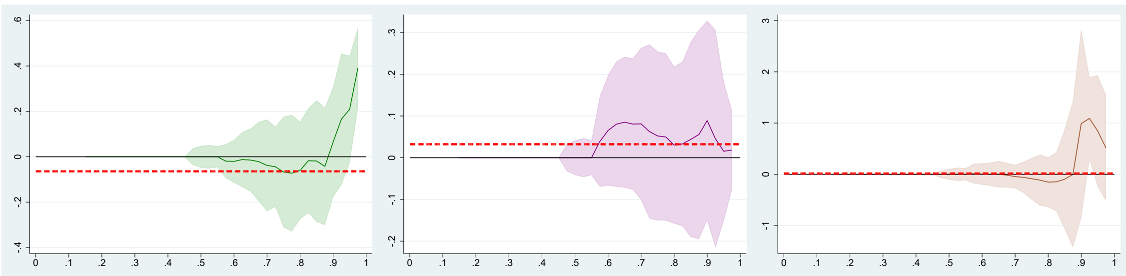
(a) All StudBlack_AdvBlack **(b)** All StudBlack_AdvWhite **(c)** All StudWhite_AdvBlack



(d) MM StudBlack_AdvBlack **(e)** MM StudBlack_AdvWhite **(f)** MM StudWhite_AdvBlack



(g) FF StudBlack_AdvBlack **(h)** FF StudBlack_AdvWhite **(i)** FF StudWhite_AdvBlack



(j) FM StudBlack_AdvBlack **(k)** FM StudBlack_AdvWhite **(l)** FM StudWhite_AdvBlack

Table 19: Quantile regression with robust clustered standard errors where a, b, and c are respectively 75th, 80th, and 99th. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. (1) is for the sub-sample White Student White Professors; (2) is the sub-sample Black Student Black Professors; and (3) is the sub-sample Black Student White Professors. Additional controls are field, enrolment year, and year.

	(1)a 75th MM	(1)b 80th MM	(1)c 99th MM	(2)a 75th FF	(2)b 80th FF	(2)c 99th FF	(3)a 75th FM	(3)b 80th FM	(3)c 99th FM
StudBlack_AdvBlack	0.0513	0.0789	0.267**	-0.584	-0.508	-1.262***	-0.0672	-0.0604	0.383***
	(0.142)	(0.161)	(0.0962)	(1.171)	(0.468)	(0.0638)	(0.125)	(0.109)	(0.0376)
StudBlack_AdvWhite	0.0399	0.0452	-0.0989	-0.0573	-0.0451	-0.136***	0.0522	0.0302	0.0649**
	(0.128)	(0.129)	(0.0531)	(0.0503)	(0.0848)	(0.0158)	(0.103)	(0.0949)	(0.0232)
StudWhite_AdvBlack	0.0471	0.0774	0.581***	0.373**	0.410**	1.050***	0.0868	0.152	-0.689***
	(0.260)	(0.179)	(0.126)	(0.125)	(0.155)	(0.0666)	(0.207)	(0.245)	(0.121)
moreAdv	0.0790	0.0955	0.293***	-0.0256	-0.0120	0.131***	0.107	0.203	0.0132
	(0.137)	(0.149)	(0.0704)	(0.0646)	(0.0887)	(0.0151)	(0.181)	(0.166)	(0.0280)
logprofcumavgprod	0.299***	0.327***	0.237***	0.0417	4.04e-16	0.0495*	0.0676	0.0779	0.126***
	(0.0616)	(0.0603)	(0.0307)	(0.0458)	(0.0624)	(0.0228)	(0.0708)	(0.0603)	(0.0182)
DummyStudPrevPub	0.943***	0.909***	0.920***	1.263***	1.299***	1.347***	1.034***	1.080***	1.019***
	(0.133)	(0.147)	(0.0871)	(0.321)	(0.232)	(0.0555)	(0.185)	(0.128)	(0.0300)
timegrad	-0.0852**	-0.115***	-0.204***	0.00337	-9.92e-16	0.00537	0.0150	0.00432	-0.0122
	(0.0319)	(0.0324)	(0.0382)	(0.0291)	(0.0307)	(0.00750)	(0.0265)	(0.0267)	(0.00984)
N	2668	2668	2668	811	811	811	1746	1746	1746
R^2	0.338	0.326	0.241	0.389	0.368	0.262	0.316	0.307	0.224

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

H Poisson panel regressions race

H.1 Students

Table 20: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	ALL	(1) MM	(2) FF	(3) FM
StudBlack	0.203	0.157	-0.689	0.522
	(0.129)	(0.183)	(0.622)	(0.268)
moreAdv	0.102	0.501*	0.268	0.112
	(0.143)	(0.221)	(0.448)	(0.287)
logprofcumavgprod	0.428***	0.399**	0.701*	0.425**
	(0.0965)	(0.146)	(0.301)	(0.138)
DummyStudPrevPub	1.700***	1.776***	3.188***	2.333***
	(0.122)	(0.175)	(0.537)	(0.272)
timegrad	-0.155***	-0.156**	0.0229	0.0250
	(0.0417)	(0.0575)	(0.135)	(0.0816)
Constant	0.557	0.753	-2.456	-1.247
	(0.409)	(0.683)	(1.796)	(0.753)
/				
lnalpha	1.479***	1.364*	1.291	1.356
	(0.397)	(0.580)	(1.137)	(0.753)
N	6049	2683	825	1748

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

H.2 Advisors

Table 21: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	ALL	(1) MM	(2) FF	(3) FM
AdvBlack	0.152	0.210	-3.972***	0.122
	(0.151)	(0.199)	(0.892)	(0.311)
moreAdv	0.0969	0.498*	0.0101	0.0643
	(0.141)	(0.215)	(0.444)	(0.277)
logprofcumavgprod	0.428***	0.399**	0.715*	0.426**
	(0.0965)	(0.146)	(0.299)	(0.137)
DummyStudPrevPub	1.675***	1.774***	3.746***	2.221***
	(0.118)	(0.168)	(0.675)	(0.256)
timegrad	-0.148***	-0.148*	0.0574	0.0354
	(0.0420)	(0.0605)	(0.139)	(0.0841)
Constant	0.673	0.700	0.765	-0.603
	(0.390)	(0.656)	(1.543)	(0.656)
/				
lnalpha	1.480***	1.364*	1.166	1.366
	(0.397)	(0.578)	(1.164)	(0.757)
N	6049	2683	825	1748

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

H.3 Student and Advisor couple

Table 22: Poisson Panel Regression with robust clustered standard error. The dependent variable is number of papers between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	ALL	(1) MM	(2) FF	(3) FM
StudBlack_AdvBlack	0.320	0.343	-3.422**	0.400
	(0.181)	(0.244)	(1.229)	(0.391)
StudBlack_AdvWhite	0.0997	0.0437	-0.211	0.666*
	(0.144)	(0.198)	(0.695)	(0.311)
StudWhite_AdvBlack	-0.400	-0.377	-4.190***	0.262
	(0.317)	(0.489)	(0.838)	(0.568)
moreAdv	0.112	0.544*	0.0651	0.0913
	(0.142)	(0.224)	(0.471)	(0.296)
logprofcumavgprod	0.428***	0.399**	0.712*	0.424**
	(0.0965)	(0.146)	(0.300)	(0.138)
DummyStudPrevPub	1.719***	1.777***	3.635***	2.316***
	(0.123)	(0.180)	(0.688)	(0.276)
timegrad	-0.148***	-0.150*	0.0603	0.0201
	(0.0416)	(0.0596)	(0.138)	(0.0866)
Constant	0.686	0.775	-3.266*	-0.713
	(0.351)	(0.590)	(1.497)	(0.649)
/				
lnalpha	1.476***	1.361*	1.163	1.353
	(0.397)	(0.579)	(1.174)	(0.754)
N	6049	2683	825	1748

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

I Quantile Regressions Race

Table 23: Quantile regression with clustered standard errors on the whole sample. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(1)25th	(1)50th	(1)70th	(1)75th	(1)80th	(1)85th	(1)90th	(1)95th	(1)99th
StudBlack_AdvBlack	-9.38e-16	-4.55e-18	-0.0105	0.00431	0.0443	0.0723	0.139	0.170	0.206
	(4.27e-15)	(2.78e-15)	(0.0636)	(0.106)	(0.105)	(0.145)	(0.115)	(0.127)	(0.120)
StudBlack_AdvWhite	6.27e-16	1.77e-15	0.00828	0.0277	0.0357	0.0382	0.0167	-0.0148	0.00657
	(3.19e-15)	(2.30e-15)	(0.0505)	(0.0679)	(0.0769)	(0.0922)	(0.0713)	(0.0914)	(0.170)
StudWhite_AdvBlack	9.22e-16	8.26e-15*	-0.114	-0.124	-0.112	-0.103	-0.112	-0.0827	-0.0101
	(8.10e-15)	(3.99e-15)	(0.0612)	(0.0951)	(0.137)	(0.142)	(0.159)	(0.315)	(0.187)
moreAdv	-2.13e-15	1.37e-15	0.0583	0.0529	0.0642	0.0754	0.0580	0.0384	0.0146
	(3.86e-15)	(2.59e-15)	(0.0598)	(0.0864)	(0.0993)	(0.115)	(0.0749)	(0.115)	(0.101)
logprofcumavgprod	6.99e-15	-7.50e-16	0.119***	0.152***	0.168***	0.197***	0.224***	0.234***	0.204*
	(3.72e-15)	(1.28e-15)	(0.0336)	(0.0422)	(0.0475)	(0.0533)	(0.0452)	(0.0582)	(0.0815)
DummyStudPrevPub	0.511***	0.981***	1.069***	1.037***	0.989***	0.944***	0.962***	0.906***	0.798***
	(2.25e-14)	(1.63e-14)	(0.0816)	(0.0959)	(0.0908)	(0.108)	(0.0978)	(0.118)	(0.0674)
timegrad	3.47e-15	1.13e-15	-0.00852	-0.0300	-0.0553*	-0.0861***	-0.115***	-0.149***	-0.190***
	(2.18e-15)	(8.56e-16)	(0.0134)	(0.0215)	(0.0269)	(0.0222)	(0.0180)	(0.0197)	(0.0179)
N	6012	6012	6012	6012	6012	6012	6012	6012	6012
R^2	0.246	0.230	0.299	0.306	0.308	0.299	0.287	0.260	0.202

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 24: Quantile regression with clustered standard errors on the sub-sample of Male Student and Male Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(1)25th	(1)50th	(1)70th	(1)75th	(1)80th	(1)85th	(1)90th	(1)95th	(1)99th
StudBlack_AdvBlack	-5.30e-15	2.65e-15	0.0711	0.0513	0.0789	0.0764	0.0771	0.0936	0.267**
	(0.0183)	(0.0126)	(0.152)	(0.139)	(0.160)	(0.176)	(0.178)	(0.127)	(0.0962)
StudBlack_AdvWhite	-4.30e-15	6.80e-15	0.0594	0.0399	0.0452	0.0423	-0.00101	-0.0307	-0.0989
	(0.0155)	(0.0105)	(0.112)	(0.127)	(0.125)	(0.137)	(0.110)	(0.102)	(0.0531)
StudWhite_AdvBlack	-8.47e-15	-6.90e-15	0.0776	-0.0471	-0.0774	-0.104	-0.179	-0.327	-0.581***
	(0.0308)	(0.0189)	(0.248)	(0.260)	(0.177)	(0.166)	(0.169)	(0.168)	(0.126)
moreAdv	-1.47e-15	3.29e-15	0.139	0.0790	0.0955	0.130	0.179	0.191	0.293***
	(0.0175)	(0.0124)	(0.138)	(0.137)	(0.141)	(0.196)	(0.174)	(0.122)	(0.0704)
logprofcumavgprod	6.32e-15	-1.70e-17	0.277***	0.299***	0.327***	0.316***	0.306***	0.265***	0.237***
	(0.00848)	(0.00595)	(0.0555)	(0.0610)	(0.0586)	(0.0694)	(0.0532)	(0.0586)	(0.0307)
DummyStudPrevPub	0.511***	1.099***	1.048***	0.943***	0.909***	0.926***	0.949***	0.924***	0.920***
	(0.0427)	(0.0600)	(0.127)	(0.131)	(0.147)	(0.183)	(0.156)	(0.137)	(0.0871)
timegrad	3.13e-15	4.75e-16	-0.0555	-0.0852**	-0.115***	-0.131***	-0.166***	-0.194***	-0.204***
	(0.00507)	(0.00401)	(0.0295)	(0.0314)	(0.0300)	(0.0279)	(0.0301)	(0.0231)	(0.0382)
N	2668	2668	2668	2668	2668	2668	2668	2668	2668
R^2	0.254	0.240	0.338	0.337	0.326	0.319	0.312	0.294	0.241

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 25: Quantile regression with clustered standard errors on the sub-sample of Female Student and Female Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(2)25th	(2)50th	(2)70th	(2)75th	(2)80th	(2)85th	(2)90th	(2)95th	(2)99th
StudBlack_AdvBlack	-0.0323	-0.421	-0.684	-0.584	-0.508	-0.721	-0.855**	-0.845*	-1.262***
	(.)	(.)	(1.487)	(1.309)	(0.468)	(0.459)	(0.312)	(0.426)	(0.0638)
StudBlack_AdvWhite	-3.07e-16	2.74e-15	-0.0352	-0.0573	-0.0451	2.74e-16	-0.00283	-0.0874	-0.136***
	(1.48e-15)	(3.69e-15)	(0.0481)	(0.0531)	(0.0848)	(0.0869)	(0.0815)	(0.107)	(0.0158)
StudWhite_AdvBlack	1.42e-15	2.03e-16	-0.324**	-0.373**	-0.410**	-0.511	-0.709**	-0.828***	-1.050***
	(2.17e-15)	(1.59e-15)	(0.109)	(0.131)	(0.155)	(0.306)	(0.249)	(0.248)	(0.0666)
moreAdv	1.00e-16	-3.68e-16	-0.0270	-0.0256	-0.0120	1.61e-15	0.00822	0.0747	0.131***
	(1.68e-15)	(2.04e-15)	(0.0699)	(0.0602)	(0.0887)	(0.0752)	(0.0905)	(0.0833)	(0.0151)
logprofcumavgprod	-1.02e-15	-2.02e-16	0.0349	0.0417	7.42e-16	8.25e-16	-0.0134	-0.00524	0.0495*
	(1.74e-15)	(1.23e-15)	(0.0450)	(0.0444)	(0.0624)	(0.0596)	(0.0670)	(0.129)	(0.0228)
DummyStudPrevPub	0.288***	0.981***	1.314***	1.263***	1.299***	1.345***	1.337***	1.264***	1.347***
	(1.97e-14)	(4.86e-14)	(0.246)	(0.369)	(0.232)	(0.218)	(0.143)	(0.209)	(0.0555)
timegrad	-6.85e-16	6.01e-18	0.0110	0.00337	-8.64e-16	-1.25e-15	-0.00539	0.00153	0.00537
	(7.53e-16)	(5.08e-16)	(0.0156)	(0.0291)	(0.0307)	(0.0293)	(0.0246)	(0.0377)	(0.00750)
N	811	811	811	811	811	811	811	811	811
R^2	0.280	0.303	0.384	0.386	0.368	0.368	0.286	0.275	0.262

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

Table 26: Quantile regression with clustered standard errors on the sub-sample of Female Student and Male Advisor. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Additional controls are field, enrolment year, and year.

	(3)25th	(3)50th	(3)70th	(3)75th	(3)80th	(3)85th	(3)90th	(3)95th	(3)99th
StudBlack_AdvBlack	-1.41e-16	2.35e-16	-0.0382	-0.0672	-0.0604	-0.0188	0.0648	0.209	0.383***
	(2.80e-15)	(0.0245)	(0.103)	(0.125)	(0.109)	(0.137)	(0.124)	(0.121)	(0.0376)
StudBlack_AdvWhite	7.19e-16	1.98e-16	0.0811	0.0522	0.0302	0.0438	0.0890	0.0155	0.0649**
	(1.97e-15)	(0.0212)	(0.0928)	(0.103)	(0.0949)	(0.119)	(0.122)	(0.0856)	(0.0232)
StudWhite_AdvBlack	1.12e-15	2.58e-15	-0.0449	-0.0868	-0.152	-0.0958	0.990	0.844	0.689***
	(4.16e-15)	(0.0543)	(0.117)	(0.206)	(0.245)	(0.497)	(0.941)	(0.555)	(0.121)
moreAdv	3.86e-16	-2.26e-16	0.0492	0.107	0.203	0.215	0.138	0.133	0.0132
	(2.02e-15)	(0.0230)	(0.103)	(0.180)	(0.166)	(0.237)	(0.170)	(0.143)	(0.0280)
logprofcumavgprod	-2.79e-15	-2.48e-16	0.0682	0.0676	0.0779	0.106	0.168	0.0919	0.126***
	(1.92e-15)	(0.0139)	(0.0634)	(0.0710)	(0.0603)	(0.0746)	(0.0859)	(0.0728)	(0.0182)
DummyStudPrevPub	0.288***	0.693***	0.959***	1.034***	1.080***	1.051***	1.021***	1.118***	1.019***
	(6.06e-15)	(0.0688)	(0.160)	(0.185)	(0.128)	(0.145)	(0.122)	(0.103)	(0.0300)
timegrad	-8.47e-16	-6.17e-16	0.0156	0.0150	0.00432	-0.0211	-0.0343	-0.0188	-0.0122
	(6.98e-16)	(0.00645)	(0.0209)	(0.0266)	(0.0267)	(0.0389)	(0.0390)	(0.0309)	(0.00984)
N	1746	1746	1746	1746	1746	1746	1746	1746	1746
R^2	0.225	0.242	0.311	0.316	0.307	0.299	0.271	0.249	0.224

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

J Without controls

Table 27: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
StudFemale	-0.195**	-0.236**	-0.169*
	(0.0613)	(0.0845)	(0.0813)
Constant	0.559***	1.659***	0.743***
	(0.143)	(0.393)	(0.189)
N	3083	1099	1641
R^2	0.0949	0.138	0.0893
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001"	

Table 28: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
AdvFemale	-0.119	0.137	-0.0473
	(0.0714)	(0.146)	(0.0834)
Constant	0.431**	1.198**	0.571**
	(0.159)	(0.402)	(0.206)
N	3083	1099	1641
R^2	0.0830	0.119	0.0764
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001"	

Table 29: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample White Student White Professors; Column (2) On the sub-sample Black Student Black Professors; Column (3) On the sub-sample Black Student White Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
StudFemale_AdvFemale	-0.252**	-0.0632	-0.291*
	(0.0941)	(0.230)	(0.137)
StudFemale_AdvMale	-0.238***	-0.240**	-0.101
	(0.0716)	(0.0923)	(0.0957)
StudMale_AdvFemale	-0.198	0.0319	0.0327
	(0.107)	(0.173)	(0.0989)
Constant	0.426***	1.411***	0.585***
	(0.125)	(0.368)	(0.167)
N	3083	1099	1641
R^2	0.0997	0.139	0.0941
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001"	

Table 30: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
StudBlack	-0.137	-0.146	-0.0791
	(0.0702)	(0.111)	(0.0770)
Constant	0.859***	0.337	0.599**
	(0.198)	(0.225)	(0.192)
N	2683	825	1748
R^2	0.0544	0.183	0.0884
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001"	

Table 31: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
AdvBlack	-0.0603	-0.252	-0.109
	(0.0792)	(0.163)	(0.0943)
Constant	0.732***	0.383	0.622**
	(0.188)	(0.207)	(0.203)
N	2683	825	1748
R^2	0.0491	0.184	0.0861
Standard errors in parentheses			
* p<0.05	** p<0.01	*** p<0.001"	

Table 32: Pooled Panel Regression with robust clustered standard error. The dependent variable is log of 1+ average productivity in terms of number of paper between the period t and t+2. Column (1) On the sub-sample Male Student Male Professors; Column (2) On the sub-sample Female Student Female Professors; Column (3) On the sub-sample Female Student Male Professors. Additional controls are field, and enrolment year.

	(1)	(2)	(3)
StudBlack_AdvBlack	-0.107	-0.0329	-0.141
	(0.0933)	(0.262)	(0.0949)
StudBlack_AdvWhite	-0.197*	-0.202	-0.0204
	(0.0812)	(0.120)	(0.0878)
StudWhite_AdvBlack	-0.361*	-0.393*	0.0174
	(0.155)	(0.161)	(0.320)
Constant	0.710***	0.175	0.526**
	(0.171)	(0.166)	(0.164)
N	2683	825	1748
R^2	0.0634	0.200	0.0890

Standard errors in parentheses
* p<0.05 ** p<0.01 *** p<0.001"

The UNU-MERIT WORKING Paper Series

2020-01 *Debating the assumptions of the Thirlwall Model: A VECM analysis of the Balance of Payments for Argentina, Brazil, Colombia, and Mexico* by Danilo Spinola
2020-02 *The La Marca Model revisited: Structuralist Goodwin cycles with evolutionary supply side and balance of payments constraints* by Danilo Spinola
2020-03 *Uneven development and the balance of payments constrained model: Terms of trade, economic cycles, and productivity catching-up* by Danilo Spinola
2020-04 Time-space dynamics of return and circular migration: Theories and evidence by Amelie F. Constant
2020-05 *Mapping industrial patterns and structural change in exports* by Charlotte Guillard
2020-06 *For real? Income and non-income effects of cash transfers on the demand for food* by Stephan Dietrich and Georg Schmerzeck
2020-07 *Robots and the origin of their labour-saving impact* by Fabio Montobbio, Jacopo Staccioli, Maria Enrica Virgillito and Marco Vivarelli
2020-08 *STI-DUI innovation modes and firm performance in the Indian capital goods industry: Do small firms differ from large ones?* By Nanditha Mathew and George Paily
2020-09 *The impact of automation on inequality across Europe* by Mary Kaltenberg and Neil Foster-McGregor
2020-10 *What matters in funding: The value of research coherence and alignment in evaluators' decisions* by Charles Ayoubi, Sandra Barbosu, Michele Pezzoni and Fabiana Visentin
2020-11 *The productivity impact of business visits across industries* by Mariacristina Piva, Massimiliano Tani and Marco Vivarelli
2020-12 *Technological revolutions, structural change & catching-up* by Jan Fagerberg and Bart Verspagen
2020-13 *Semi-endogenous growth models with domestic and foreign private and public R&D linked to VECMs with evidence for five countries* by Thomas Ziesemer
2020-14 *The economic impact of public R&D: an international perspective* by Luc Soete, Bart Verspagen and Thomas H.W. Ziesemer
2020-15 *Taking the challenge: A joint European policy response to the corona crisis to strengthen the public sector and restart a more sustainable and social Europe* by Jo Ritzen, Javi Lopez, André Knottnerus, Salvador Perez Moreno, George Papandreou and Klaus F. Zimmermann
2020-16 Migration of higher education students from the North Africa Region to the United Kingdom by Samia Satti Osman Mohamed Nour
2020-17 Overview of the Sudan Uprising by Samia Satti Osman Mohamed Nour
2020-18 *Inter-country Distancing, Globalisation and the Coronavirus Pandemic* by Klaus F. Zimmermann, Gokhan Karabulut, Mehmet Huseyin Bilgin and Asli Cansin Doker
2020-19 *How does innovation take place in the mining industry? Understanding the logic behind innovation in a changing context* by Beatriz Calzada Olvera & Michiko Iizuka
2020-20 Public universities, in search of enhanced funding by Jo Ritzen
2020-21 Ph.D. research output in STEM: the role of gender and race in supervision by Giulia Rossello, Robin Cowan and Jacques Mairesse