Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Evaluating Professor Value-added: Evidence from Professor and Student Matching in Physics

Yuta Kikuchi, Ryo Nakajima

5 February, 2016
DP2016-002
http://ies.keio.ac.jp/en/publications/5323

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
5 February, 2016

Evaluating Professor Value-added: Evidence from Professor and Student Matching in Physics

Yuta Kikuchi, Ryo Nakajima

Keio-IES DP2016-002
5 February, 2016
JEL classification: D83; I23; J24
Keywords: knowledge creation; postgraduate education; faculty quality; research apprenticeship

Abstract

This paper estimates a professor's value added to a postgraduate student's research achievement growth using unique panel data on matched advisor-advisee pairs in a world-leading physics graduate program. To address an identification problem related to the endogenous selection of advisors and advisees, we use professor turnover and estimate a semi-parametric lower bound of the variance in advisor quality affecting advisee research performance. We find that a one-standard-deviation increase in professor quality results in a 0.54 standard deviation increase in a doctoral student's research achievement growth, increasing the number of first-authored papers that are published in top journals by 0.64 at the doctoral level.

Yuta Kikuchi
GISS, Yokohama National University
79-1 Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan
kikuchi-yuta-vy@ynu.jp

Ryo Nakajima
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
nakajima@econ.keio.ac.jp

Acknowledgement: This research was supported by Grants-in-Aid for Scientific Research Nos. 14J08184 and 26380313 from the Japan Society for the Promotion of Science.

Evaluating Professor Value-added: Evidence from Professor and Student Matching in Physics *

Yuta Kikuchi† Ryo Nakajima‡

February 17, 2016

Abstract

This paper estimates a professor's value added to a postgraduate student's research achievement growth using unique panel data on matched advisor-advisee pairs in a world-leading physics graduate program. To address an identification problem related to the endogenous selection of advisors and advisees, we use professor turnover and estimate a semi-parametric lower bound of the variance in advisor quality affecting advisee research performance. We find that a one-standard-deviation increase in professor quality results in a 0.54 standard deviation increase in a doctoral student's research achievement growth, increasing the number of first-authored papers that are published in top journals by 0.64 at the doctoral level.

Keywords: knowledge creation; postgraduate education; faculty quality; research apprenticeship

JEL codes: D83; I23; J24

---

*The authors thank Donna Ginther and Daiji Kawaguchi for supportive comments and helpful suggestions for improving the manuscript. We are grateful to Kenichiro Aoki and Kensuke Kobayashi for providing helpful comments on postgraduate education and academic culture in Japanese physics communities, especially regarding the the University of Tokyo physics department.

†GISS, Yokohama National University, E-mail address: kikuchi-yuta-vy@ynu.jp

‡Department of Economics, Keio University, E-mail address: nakajima@econ.keio.ac.jp

1 Introduction

1.1 Overview

How is knowledge created? Economists have had a particular and long-standing interest in knowledge production. Indeed, the new economic growth theory literature regards the way in which knowledge is created and accumulated as crucial for a nation to grow (Romer, 1990; Lucas, 1988; Grossman and Helpman, 1991).

Recently, there has been an increase in the number of empirical studies on knowledge creation in the field of science and technology. They have investigated how an individual's knowledge creation is affected by knowledge created by others, with a particular emphasis on knowledge spillovers between individuals within and across institutions. The evidence obtained thus far, however, has been mixed. Some studies (e.g., Azoulay, Zivin and Wang, 2010; Moser, Voena and Waldinger, 2014; Borjas and Doran, 2014) provide evidence in favor of positive knowledge spillovers, while others (Waldinger, 2012; Borjas and Doran, 2012) do not.

Surprisingly little attention has been devoted to knowledge *reproduction* processes across generations. As is often argued, scientific and technological knowledge is tacit (Polanyi, 1958, 1966). It is not easily translated and thus needs to be intentionally articulated, codified and diffused. Therefore, knowledge has long been reproduced through a deliberate process of education and learning whereby those with knowledge take voluntary action to pass it on to those who do not. Thus, to investigate the creation and diffusion of scientific and technological knowledge, it seems natural to distinguish *vertical* knowledge flow (i.e., the knowledge flow from an individual with high expertise to one with low expertise) from *horizontal* knowledge flow (i.e., the knowledge flow among individuals with the same level of expertise). If the two lines of knowledge flow differ in the efficiency of the transmission of know-how, the mixed results obtained by prior studies on the extent of spillovers might be explained with regard to such differences.

This paper focuses on the knowledge reproduction process whereby knowledge is conveyed through vertical relationships, including master-apprentice, teacher-student and senior-junior-collaborator relationships. We specifically focus on the *advisor-advisee* relationship in post-graduate education to examine its effectiveness in expanding scientific frontiers. Under the hypothesis that a professor's "quality" has a consequential impact on the growth of a student's research achievement, we estimate the professor's (advisor's) *value added* as the contribution to student's (advisee's) progress on research outcomes.

Empirically estimating a graduate school professor's value added is complicated by the endogenous selection process involving students, professors and schools. We anticipate that students of promise will apply and be admitted to highly ranked schools. Moreover, professors with good academic standing are likely to have faculty positions at prestigious schools. These types of selective recruitment will lead to nonrandom sorting of students and professors across graduate programs. Furthermore, the existence and extent of sorting can be reinforced by the advisor-advisee matching process *within* a school, whereby students will choose and be chosen by faculty members.[1] Therefore, the mere association of research productivity, measured by, say, publication records, between a professor and a student does not necessarily imply a causal relationship whereby the professor's advising and mentoring could enhance the research capabilities of his or her students.

To disentangle the influence of professors on students from the sorting and matching effects, we use an identification strategy that exploits professor *turnover* from events such as retirement, relocation, or death. We borrow this idea from Rivkin, Hanushek and Kain (2005, henceforth, RHK), who estimate a lower bound of the teacher quality effect, which can be either observed or unobserved, on student achievement gains by exploiting teacher turnover.

This paper estimates within-school professor value added at a world-leading postgraduate program in physics in Japan using unique panel data on matched advisor-advisee pairs. Japanese graduate schools provide an ideal setting for applying RHK's strategy of turnover-based value-added estimation. When an advisor exits a Japanese graduate school due to turnover, the advisees usually remain in the same program and continue their research projects under the supervision of new advisor. Therefore, the advisees who experience advisor turnover are influenced by two advisers of different quality. Thus, the student's research achievement growth, under the varied influences of different advisors, would be more volatile than that of advisees who did not suffer advisor turnover. To measure the magnitude of advisor impact on advisee research performance, we exploit the degree to which the student's research achievement growth differs across cohorts.

Certainly, factors other than advisor quality might affect an advisee's research performance. This paper employs a semi-parametric education production function, which is widely used in the economics of education literature, that attributes the student's achievement gains to various fixed effects. Repeated observations of an individual student's research outcomes, which are measured by publication records, in master's and doctoral degree programs enable us to eliminate student fixed effects by taking the difference of the research outcomes of a given student from one degree program to another.

This paper employs a quasi-experimental design. We base our analysis on a *lab*, defined by a cohort of students who were assigned to the same advisor. We assign labs to a treatment group, in which the advisor was replaced due to turnover, and a control group, in which

[1] These choices give rise to "assortative matching" between students and professors within a school with respect to research ability.

the advisor was not replaced. We use cohort-to-cohort variation in the average gain in student research achievement to measure the effect of the advisor quality on advisee research achievement gains. We demonstrate that the squared double-difference in student research achievement gains between the degree programs (i.e., master's and doctoral programs) and between cohorts is larger for the treatment group than for the control group and is driven primarily by the change in advisor quality following turnover.

For a treatment and control study to be valid, some degree of randomness in the treatment assignment is necessary, which is equivalent to random advisor replacement in our empirical context. To address the lack of sufficient randomness in the actual school environment, we employ propensity score matching, which selects a subset of the control group that has a similar likelihood of the treatment being offered to that for individuals in the treatment group. We compute the propensity score as the estimated likelihood of advisor turnover occurring and match the sample of control group labs with the treatment group labs based on the estimated propensity score. We implement regression analysis based on the matched sample to obtain an unbiased estimate of the lower bound of the variance in advisor effectiveness on advisee research achievement gains.

We estimate a professor's value added to student research achievement growth in the department of physics, University of Tokyo (henceforth, UTokyo), which is a prestigious research and educational institution in physics. The estimation results provide strong evidence for the existence of professor value added, which is consistent with the expectation that knowledge and ideas are transmitted vertically from advisor to advisee. Indeed, the results consistently demonstrate that the difference in the quality of a student's advisor makes a notable difference in the student's research outcome growth at the doctoral level. Specifically, our estimates indicate that a one-standard-deviation increase in advisor quality will increase an doctoral advisee's research achievement by 0.54 standard deviations. We also find that a one-standard-deviation increase in advisor quality entails an increase in the number of articles published by a doctorate student in top journals as a first author by 0.64. The findings of this paper are robust to different definitions of student research outcomes and are also insensitive to many different model specifications. The results are also robust to a falsification exercise that examines whether the timing of the increased variability in the double-differenced student research achievement gain agrees with that of advisor turnover, as predicted by the empirical model.

We also investigate alternative mechanisms for knowledge transmission other than that based on learning through the advisor-advisee relationship. The data indicate that advisor turnover does not have a significant unidirectional, positive or negative, impact on an advisee's research achievement gain, *per se*, as is consistent with the mechanism that our value-added model postulates, and is thus not fully explained by the other mechanisms such as that emphasizing the recombination role of various extant pieces of knowledge in new knowledge creation (e.g., Weitzman, 1998). Further analysis reveals that the effect of knowledge transmission from advisor to student *within* a lab tends to outweigh that from non-advisor to student *across* labs.

The remainder of the paper proceeds as follows. A brief literature review is provided in the remainder of this section. Section 2 describes the institutional background of postgraduate physics education in Japan. Section 3 presents the empirical model and describes a regression-based approach to estimate the lower bound of professor value added. Section 4 explains the data set used for the analysis. Section 5 discusses some empirical issues concerning value-added estimation. Section 6 presents the estimation results and provides robustness checks. Section 7 concludes.

1.2 Related Literature

This paper contributes to the literature by measuring the effectiveness of professors in promoting students' research productivity growth at a postgraduate institution. The most closely related work to this paper is Waldinger (2010), who estimates the causal effect of prominent professors on the research outcomes of Ph.D. students in mathematics at German universities during the Nazi era. Although we share his view that "university quality is believed to be one of the key drivers for a successful professional career of university graduates "(Waldinger, 2010, p.787), we highlight the importance of direct interactions between advisor and advisee as a medium whereby knowledge is memorized, transferred and accumulated. Indeed, anecdotal evidence (e.g., Zuckerman, 1977) suggests the importance of vertical social ties in scientific enterprises at academic institutions. However, to the best of our knowledge, no systematic quantitative study, especially one that carefully controls for endogenous matching between master (teacher, advisor or senior collaborator) and apprentice (student, advisee or junior collaborator), has been conducted to date.

Our findings validate the view of earlier studies (e.g., Azoulay et al., 2010; Moser et al., 2014; Borjas and Doran, 2014) that vertical social interactions among scientists are enduring and consequential for scientific and technological knowledge to be created and diffused. For example, a recent study by Moser et al. (2014), who estimate the effect of German Jewish émigrés on U.S. innovation, suggests that knowledge externalities occurred and were amplified through educational and collaborative ties in scientist networks such that U.S. junior scientists were trained by and collaborated with prominent Jewish senior scientists who emigrated. Borjas and Doran (2014) study the impact of the influx of Soviet mathematicians into the United States after the collapse of the Soviet Union and conclude that positive knowledge spillovers are generated through the relationships among collaborating mathematicians who regularly interact when at least one of them is an outstanding knowledge producer.

This study is also related to a voluminous education economics literature that evaluates teacher value added (e.g., Hanushek and Rivkin, 2006, 2010). We base our empirical analysis on the value-added model approach that is widely employed in the literature. Specifically, as mentioned above, we adopt a semi-parametric value-added model and employ the turnover estimator proposed by RHK. However, we depart from the previous literature on teacher value added in that we focus on value added at a level higher than secondary education. Although numerous studies estimate value added at the primary and secondary education levels (e.g., Hanushek and Rivkin, 2012, for a recent review), few studies (e.g., Hoffmann and Oreopoulos, 2009; Carrell and West, 2010) estimate a professor's value added in the context of post-secondary institutions. While these studies on professor value added attempt to estimate the effectiveness of professors in improving students' grade gains at the *undergraduate* level, we turn to professors' value added to students' research achievement gains at the *postgraduate* level and thus evaluate the effectiveness of professors in terms of their "quality" in advising or mentoring graduate students' research projects.

To the best of our knowledge, no studies assess the impact of professor quality on graduate student research productivity growth by shedding light on the value-added contribution. A partial exception is the study by Hilmer and Hilmer (2009), who find a positive effect of an advisor's research prominence on advisees' early career publication success in U.S. economics Ph.D. programs. While they are successful in disentangling the effect of advisor quality from that of program quality on Ph.D. students' publication outcomes, they do not address endogenous advisor-advisee matching between professors and students within and across institutions. Thus, it seems questionable to interpret their finding of a positive correlation between the research productivity of advisors and advisees as causal.

2 Institutional Background

2.1 Postgraduate Physics Education in Japan

Postgraduate education in Japan, including in physics, has a two-tiered structure, that is, a two-year master's degree program followed by a doctoral program that typically lasts three or four years.[2] Leading Japanese research universities typically offer both master's and doctoral courses. In most cases, students enrolled in a doctoral degree program graduate with a master's degree from the same school. However, they are institutionally separated. Thus, a master's student seeking to pursue a doctorate must take an entrance examination, which is

[2]The basic structure has remained unchanged since World War II, although the organizational structure of universities has been reformed (see Ushiogi, 1993; Ogawa, 2002)

largely based on a master's thesis, to be admitted to a doctoral course even if it is offered by the same institution. In a sense, the master's degree program implicitly serves as a screening device for doctoral programs in Japan.

Three features are notable for graduate education in physics for master's programs in Japan. First, Japanese physics master's students are closely linked to their faculty advisors immediately after enrollment in a program. Indeed, applicants to a master's degree program must declare their desired field of specialization and submit a short list of faculty advisors from whom mentorship is sought upon admission. Only those who are approved for support by designated advisors are admitted to a graduate school.[3]

Second, physics education in Japan at the master's level is best characterized by research-based apprentice training, which is often contrasted by coursework-based training in the U.S. (Abe and Watanabe, 2012). Although Japanese master's students in physics are required to take some "coursework" credits toward their degrees, they can earn most of their credits through learning-by-doing style research "seminars" taught by a faculty advisor.[4]

Finally, for Japanese physics graduate students, a thesis is required to complete the master's program. It is expected to be original, as a doctoral thesis should be, although they are evaluated according to different criteria of scholarly maturity. Students are encouraged to begin original research in their chosen fields at an early stage of the master's degree program under the instruction and guidance of a faculty advisor. Because the master's thesis is a critical factor for admittance to doctoral programs, Japanese students and professors attach great importance to a master's thesis as a pathway to doctoral study.

In contrast, the doctoral programs in physics at Japanese universities are more similar to their counterparts in Western countries than are the master's programs. Specifically, Japanese doctoral students and American Ph.D candidates are considered comparable in that there is no coursework requirement. Japanese students at the doctoral level, similar to Ph.D. candidates in the U.S., begin the research for their doctoral dissertations under the supervision of their research advisors and continue the research topic they pursued in their master's thesis in their doctoral dissertation. In general terms, Japanese physics students are required to write several articles published in refereed journals as a prerequisite for a doctoral degree. These publications are usually included in a doctoral thesis.

[3]This contrasts with U.S. graduate students, who are matched with their supervisors through the rotation of faculty labs after they complete their coursework and become Ph.D. candidates (see Gumport, 1993).

[4]For example, for the master's degree program in physics at UTokyo, students must take at least thirty credits of coursework at the master's degree level. However, lab-based research "seminars" offered by thesis advisors constitute two thirds of their total credits.

2.2 Physics Labs in Japanese Universities

Interaction between a graduate student and a faculty advisor is lab-oriented in Japanese physics graduate programs. Upon enrollment in the master's program, Japanese physics students are assigned individually to a lab, and the lab's leader (or sometimes sub-leader) becomes their thesis research advisor. Students acquire the knowledge necessary to conduct their research through frequent interaction with their advisors in a lab setting. The content of this lab-based teaching and learning includes basic research skills, such as how to read scientific articles, how to select research topics, how to present results at seminars and conferences, and how to write publishable papers, as well as the culture of physics such as the style of work, mode of thought, and a taste for "good" physics (Abe and Watanabe, 2012).

While apprenticeship-style education is also employed in Western countries,[5] it is particularly personalized in Japan. It is typical to refer to a lab using the lab leader's family name.[6] Indeed, a research lab is often referred to as an "*ie*", which means a household in Japanese: the leader (a faculty member who is a full or associate professor) is the father, the sub-leader (associate professor or research associate) is the mother, the doctoral course students (and postdocs if any) are the older brother or sisters, and the master's students are the younger siblings. In Japanese universities, the everyday activities of graduate students are organized around a lab (Kawashima and Maruyama, 1993).

Although Japanese physics labs are often likened to a household, they are generally democratic, not feudal, in tone. The "laboratory democracy" in Japanese physics communities can be traced back to the end of World War II, the period when there were immediate and insistent calls for the creation of a new "scientific Japan" under the control of the allied occupation (Low, 2005).[7] To place this in perspective, it is broadly understood that Japanese physics labs are less prescriptive and less hierarchical than their U.S. counterparts.[8] For example, Sharon Traweek, an anthropologist who studied various research groups of elementary

[5]See Gumport (1993) for the U.S.; Becher (1993) for the UK; and Gellert (1993) for Germany.

[6]For instance, if the last name of a lab leader is *Nakajima*, the lab is usually called the *Nakajima Lab* in Japanese universities.

[7]Low (2005) also notes that professor Shouichi Sakata at the physics department of Nagoya Imperial University, an influential physicist at that time, played an important role in developing the new democratic lab system in the Japanese physics community. Sakata, who was under the philosophical influence of Marxism, introduced a charter for the physics department at Nagoya in 1946. The charter holds that democracy should serve as the guiding principle in department affairs; all faculty members and students should be treated equally concerning physics research (Department of Physics, Nagoya University, 2015). The idealism of Sakata's "laboratory democracy" then spread. Soon after the Nagoya Charter was announced, several physics departments at other universities introduced similar systems. See Tanabashi (2012) for details on Sakata's laboratory democracy.

[8]Regarding the difference in lab cultures between Japan and America, it can be insightful to contrast the description of Kawashima and Maruyama (1993) with that of Gumport (1993).

particle physicists in Japan and the U.S., reports that decision-making in Japanese physics labs was based on the consensus of the members. There is no strict division of labor among lab members, even between faculty members and graduate students, in Japanese physics labs. Traweek (1988) offers a first-hand account of the democratic nature of labs in Japan by asking group leaders of a lab for the source of new ideas for experimental design or data analysis. Traweek (1988) writes, (p.147) " [lab leaders] generally credited the graduate students ... they said the group then responds to their ideas, perhaps modifying or amplifying them".

Hence, although it is not uncommon for the research topics of master's and doctoral theses to be suggested by advisors as a part of a large, ongoing project in a given lab, Japanese physics graduate students are, generally, given some autonomy to pursue their own research based on their original ideas.

3 Empirical Model

In this section, we introduce a simple value-added model that associates growth in student research achievement with the "quality" of the professor supervising the student. Then, we present a regression-based approach to estimate a lower bound of the variance in professor quality, which can be interpreted as the extent to which any professor differences matter in determining student research outcome growth.

3.1 Value-added Specification

Following the standard value-added modeling approach (e.g., Hanushek and Rivkin, 2010), we employ a semi-parametric specification of a professor's contribution to a student's achievement growth.

Consider graduate student i who entered the master's program of a graduate school in year c. Below, we treat year c as the student's cohort. We denote the research outcome growth of a graduate student in the master's degree program by $g = m$ and in the doctoral degree program by $g = d$. The growth is measured by the *gains* in research output from the previous degree program to the current degree program.[9] Let $\Delta outcome^c_{iag}$ be the research outcome growth of student i under the supervision of professor $a \in \mathscr{A}$ in degree program $g \in \{m, d\}$ in cohort $c \in \mathscr{C}$. We assume that it is given by the following function:

$$\Delta outcome^c_{iag} = \gamma_i + \theta_{ag} + \nu^c_{iag}, \tag{1}$$

[9] We assume that the research output of students at the bachelor level is zero. We compute a publication-based research proficiency score, which is explained in detail in Section 4, for students in the sample when they are undergraduate students and find that it is negligible.

where γ_i is student i's individual fixed effect, θ_{ag} is professor a's quality that influences the student research outcome growth in degree program g, and ν_{iag}^c is an idiosyncratic random shock.

The specification highlights the components that affect a student's research outcome growth. The model is very simple given its additive structure. First, note that other effects, such as school fixed effects and research field fixed effects, are not included in the value-added model. We opt not to include these fixed effects because they are subtracted out of the estimation model in the process of "differencing", as presented below. Second, professor quality, θ_{ag}, and student quality, γ_i, will be correlated. Specifically, because of endogenous matching between professor (advisor) and student (advisee), we expect that θ_{ag} and γ_i are positively correlated. Finally, as in the standard specification of the value-added model, the idiosyncratic shock, ν_{iag}^c, is assumed to be uncorrelated with the student fixed effect, γ_i, or the advisor fixed effect, θ_{ag}.

We assume that matching between student and professor is many-to-one, that is, multiple students are assigned to one advisor. Let us define a *lab* as a group of students (advisees) in the same cohort who were assigned to the same professor (advisor). Specifically, we use $\ell(a,c)$ to denote a lab in which students are in cohort c and assigned to professor a as an advisor. Let L be the number of all labs in a school, and let students in lab $\ell(a,c)$ be indexed by $i = 1, \cdots, I^{\ell(a,c)}$, where $I^{\ell(a,c)}$ is the number of students in lab $\ell(a,c)$. We use $\mathscr{I}^{\ell(a,c)} \equiv \{1, \cdots, I^{\ell(a,c)}\}$ to denote the set of students in the lab.

We take the average of Equation (1) over all students in the same lab $\ell(a,c)$. Because the students in the same lab have the same advisor quality, we have the following equation for the lab-level average of the student research outcome growth:

$$\overline{\Delta outcome}_{ag}^{\ell(a,c)} = \bar{\gamma}^{\ell(a,c)} + \theta_{ag} + \bar{\nu}_{ag}^{\ell(a,c)}, \tag{2}$$

where the overbar notation indicates the group average.

Note that the *superscript a* denotes the *initial* advisor to whom the students in lab $\ell(a,c)$ were assigned, while the *subscript a* denotes the advisor who supervised the students in degree program g. Thus, the advisors represented by the superscript and subscript could be different. For example, suppose that a turnover incident causes the students in lab $\ell(a,c)$ to switch their research advisor from professor a in the master's degree program to professor b in the doctoral degree program. Here, the average student research outcome gain at the doctoral level, which is the left-hand side of Equation (2), is given by $\overline{\Delta outcome}_{bd}^{\ell(a,c)}$, where the index a in the superscript differs from the index b in the subscript.

We use the event of professor turnover (e.g., retirement, relocation and decease) to identify the variance in professor quality. We implicitly assume that, when a professor exits a graduate program due to turnover, the students in the lab whom he or she initially supervised are re-assigned to a new advisor and continue their research projects in the same program.[10] In what follows, we therefore assume that an event of professor turnover on the faculty side leads to an event of advisor *switch* on the student side. In other words, we treat these two events, advisor turnover and advisor switch, identically. When advisor turnover occurs in a lab, two faculty members, whose quality levels are generally different, advised students in the lab.[11]

It should be noted that the professor, say b, who was assigned to the students in the lab of a professor, say a, after the latter exited due to turnover was not necessary drawn at random from a pool of professors available at the school at that time. Indeed, the newly assigned professor might select the students that he or she is willing to take over. We thus allow the student fixed effect, γ_i, to be correlated with the quality of the re-assigned professor, θ_{bd}, in the same way as we assume it to be correlated with the quality of the original advisor, θ_{ad}.

3.2 A Lower-bound Estimation of the Variance in Advisor Quality

We are interested in decomposing the total variation in student outcome gains into the variation that can be attributed to professor quality, θ_{ag}. First, take the difference of Equation (2) between the master's degree and doctoral degree programs. Doing so eliminates the student fixed effect, γ_i, because it is constant across degree programs for a given student. If advisor turnover did not occur in lab $\ell(a, c)$, it is given by the following between-degree difference equation:

$$\overline{\Delta outcome}_{ad}^{\ell(a,c)} - \overline{\Delta outcome}_{am}^{\ell(a,c)} = (\theta_{ad} - \theta_{am}) + (\bar{\nu}_{ad}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}). \tag{3}$$

In contrast, assume that there was advisor turnover in lab $\ell(a, c)$. As the students switched their advisors from advisor a in the master's program to advisor b in the doctoral program, the between-degree difference equation, corresponding to Equation (3), is given by:

$$\overline{\Delta outcome}_{bd}^{\ell(a,c)} - \overline{\Delta outcome}_{am}^{\ell(a,c)} = (\theta_{bd} - \theta_{am}) + (\bar{\nu}_{bd}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}). \tag{4}$$

Comparing Equations (3) and (4) shows that advisor turnover influences the development of student research achievement in different ways. There is a clear difference in student research outcome growth, which appears on the left-hand side of each equation, that responds differently to a change in advisors due to the difference in degree-level advisor effects, $(\theta_{ad} -$

[10] A joint transfer of faculty and students is quite rare in Japanese universities, and hence, even if a faculty member changes affiliation, the students usually remain in the same program.

[11] Based on the observed pattern of advisor replacement in UTokyo's physics graduate program, when advisor turnover occurred, the students were usually either assigned to a junior faculty member or the sub-leader of the same lab or they were moved to a different lab in closely related research fields within the same institution and were supervised by the faculty member who managed that lab.

θ_{am}) and ($\theta_{bd} - \theta_{am}$), which are generally not equal. This plays a key role in the identification of the effect of advisor quality on student research outcome growth at each degree level.

The point is illustrated by Figure1, which depicts three labs with different cohorts, c_0, c_1 and c_2, whose initial advisor is professor a. In the figure, each lab is portrayed by a connected line segment, which represents the two-year master's degree program (the first half of the segment) and the three-year doctoral degree program (the last half of the segment).[12] Here, advisor turnover did not occur in labs $l(a, c_0)$ or $l(a, c_1)$ before cohort c_2, and hence, the students in these labs were supervised by the same professor, a, throughout both the master's and doctoral programs. However, in lab $l(a, c_2)$, professor a exited the school due to turnover, and professor b took charge of the doctoral students.

Note that, on average, the research outcome gains of lab $l(a, c_0)$ and $l(a, c_1)$ students are the same, which is given by $(\theta_{ad} - \theta_{am})$, whereas, following advisor turnover, the average student research outcome gain of lab $l(a, c_2)$, which is given by $(\theta_{bd} - \theta_{am})$, could be better or worse than those of the previous cohorts, depending on whether the supervising quality of the newly assigned professor, b, is higher than that of the departing professor, a. In either case, irrespective of whether the achievement growth is positive or negative, an instance of turnover triggers a change in professor quality at the doctoral level and could thus result in a disparity in the between-degree research achievement gains between cohorts. We will use the induced divergence in research outcome growth as evidence of an advisor's impact on an advisee.

Insert Figure 1

To improve the identification, we use the *double-differencing* approach as proposed by RHK to estimate a lower bound of the variance in unknown teacher quality. We take the difference of Equations (3) and (4) with respect to cohort year. Let c' denote the cohort before c, and let τ be the years between c and c'. For professor a, consider two labs, $\ell(a, c)$ and $\ell(a, c')$. Let $W^{\ell(a,c,c')}$ denote a dummy variable indicating a change in advisor due to turnover: it takes value one if professor a is replaced in lab $\ell(a, c)$ due to turnover and zero otherwise. Without loss of generality, we assume that supervisor replacement is from professor a to professor b such that, if there were advisor turnover, the students would have been supervised by two different professors, a and b, in the master's and doctoral degree programs, respectively. Then,

[12]For the ease of exposition, the labs' cohorts are not overlapped in the figure, although this is not necessarily the case in the actual sample.

we have the following double-differenced (DD) average student research outcome growth:

$$
\begin{aligned}
& DD\overline{\Delta outcome}^{\ell(a,c,c')} \\
= \quad & \begin{cases} [\overline{\Delta outcome}_{bd}^{\ell(a,c)} - \overline{\Delta outcome}_{am}^{\ell(a,c)}] - [\overline{\Delta outcome}_{ad}^{\ell(a,c')} - \overline{\Delta outcome}_{pm}^{\ell(a,c')}] & \text{if } W^{\ell(a,c,c')} = 1 \\ [\overline{\Delta outcome}_{ad}^{\ell(a,c)} - \overline{\Delta outcome}_{am}^{\ell(a,c)}] - [\overline{\Delta outcome}_{pd}^{\ell(a,c')} - \overline{\Delta outcome}_{pm}^{\ell(a,c')}] & \text{if } W^{\ell(a,c,c')} = 0 \end{cases} \\
= \quad & \begin{cases} (\theta_{bd} - \theta_{ad}) + error\ term & \text{if } W^{\ell(a,c,c')} = 1 \\ error\ term & \text{if } W^{\ell(a,c,c')} = 0, \end{cases}
\end{aligned}
\tag{5}
$$

where the *error term* is a catchall random noise term that combines the average idiosyncratic errors.

Equation (5) shows that all of the fixed effects, except for doctoral-level advisor quality, are eliminated after the double difference is taken with respect to degree programs and cohorts. The double-differenced measure is more variable,on average, for the pair of labs with and without a change in advisor ($W^{\ell(a,c,c')} = 1$) than that for the pair of labs without such a change ($W^{\ell(a,c,c')} = 0$). The gap is attributable to a discrete change in doctoral-level advisor quality from θ_{ad} to θ_{bd} due to advisor turnover. Note that advisors' quality levels can be correlated with the lab averages of student fixed effects, $\bar{\gamma}^{\ell(a,c)}$ and $\bar{\gamma}^{\ell(a,c')}$, and they can also be correlated with one another, that is, $\text{Corr}(\theta_{ad}, \theta_{bd}) \neq 0$. In what follows, we ascribe the sample variation in the double-differenced measure as a series of variance and covariance components of advisor quality and idiosyncratic shocks.

The Assumption on Advisor Quality

We make the following assumptions concerning the distribution of advisor quality.

> assumption 1.1: The expectation and variance of advisor quality are given by $\text{E}(\theta_{ag}) = \mu_g$ and $\text{Var}(\theta_{ag}) = \sigma_g^2$, for any $a \in \mathscr{A}$, $g \in \{m, d\}$, and $c, c' \in \mathscr{C}$.
>
> assumption 1.2: The correlation of advisor quality across professors, $a \neq b \in \mathcal{A}$, is given by $\text{Corr}(\theta_{ag}, \theta_{bg}) = \rho_g$, for any $a, b \in \mathscr{A}$, $a \neq b$, $g \in \{m, d\}$ and $c, c' \in \mathscr{C}$.

These assumptions state the stationarity of the advisor quality distribution, which characterizes the notion that the professors' advising quality levels are drawn from a common distribution for each degree type. It requires that the grade-program-specific mean and variance do not vary across cohorts and that the correlation with any given advisor is constant. Specifically, we interpret μ_g and σ_g^2 as the long-run mean and variance of the stationary distribution of advisor quality in degree program g within a school. The stationarity assumption simplifies the estimation of professor value added because it reduces the number of parameters to be considered. In the empirical section of the paper that follows, we estimate a lower bound of the variance in the advisor effect, σ_d^2, which is a measure of a professor's effectiveness in improving a student's research achievement growth at the doctoral level.

The Assumption on the Random Shock

The following assumptions impose restrictions on the moments of the idiosyncratic shock after *demeaning* by each cohort. Let $\bar{\nu}_g$ be the average of the random shock ν_{iag}^c, the average of which is taken over all cohorts in each degree program, g, such that the *demeaned* random shock is given by $\tilde{\nu}_{iag}^c = \nu_{iag}^c - \bar{\nu}_g$.

assumption 2.1: The conditional expectation and variance of the demeaned random shock, $\tilde{\nu}_c$, for student $i \in \mathscr{I}^{\ell(a,c)}$ are $\mathrm{E}(\tilde{\nu}_{iag}^c | W^{\ell(a,c,c')}) = 0$ and $\mathrm{Var}(\tilde{\nu}_{iag}^c | W^{\ell(a,c,c')}) = \phi_g^2$, respectively, for any $a \in \mathscr{A}$, $g \in \{m, d\}$, and $c, c' \in \mathscr{C}$.

assumption 2.2: The covariance of the demeaned random shocks *between* degree programs *within* the same student, $i \in \mathscr{I}^{\ell(a,c)}$, is given by $\mathrm{Cov}(\tilde{\nu}_{iam}^c, \tilde{\nu}_{iad}^c | W^{\ell(a,c,c')}) = \phi_{md}$ for any $a \in \mathscr{A}$, and $c, c' \in \mathscr{C}$.

assumption 2.3: The covariance of the demeaned random shocks between different students $i \in \mathscr{I}^{\ell(a,c)}$ and $j \in \mathscr{I}^{\ell(a,c)}$ who are advised by the *same* professor in degree program g is given by $\mathrm{Cov}(\tilde{\nu}_{iag}^c, \tilde{\nu}_{jag}^c | W^{\ell(a,c,c')}) = \mathrm{Cov}(\tilde{\nu}_{iag}^c, \tilde{\nu}_{jag}^{c'} | W^{\ell(a,c,c')}) = \psi_g$, for any $a \in \mathscr{A}$, $g \in \{m, d\}$, and $c, c' \in \mathscr{C}$.

assumption 2.4: The covariance of the demeaned random shocks between different students $i \in \mathscr{I}^{\ell(a,c)}$ and $j \in \mathscr{I}^{\ell(a',c')}$ who are advised by *different* professors in degree program g is zero, that is,

$$\mathrm{Cov}(\tilde{\nu}_{iag}^c, \tilde{\nu}_{ja'g}^c | W^{\ell(a,c,c')}) = \mathrm{Cov}(\tilde{\nu}_{iag}^c, \tilde{\nu}_{ja'g}^{c'} | W^{\ell(a,c,c')}) = 0,$$

for any $a, a' \in \mathscr{A}$, $a \neq a'$, $g \in \{m, d\}$, and $c, c' \in \mathscr{C}$.

assumption 2.5: The covariance of the demeaned random shocks *between* different students $i \in \mathscr{I}^{\ell(a,c)}$ and $j \in \mathscr{I}^{\ell(a',c')}$ *between* degree programs is zero, that is,

$$\mathrm{Cov}(\tilde{\nu}_{iam}^c, \tilde{\nu}_{ja'd}^{c'} | W^{\ell(a,c,c')}) = \mathrm{Cov}(\tilde{\nu}_{iad}^c, \tilde{\nu}_{ja'm}^{c'} | W^{\ell(a,c,c')}) = 0$$

for any $a, a' \in \mathscr{A}$, and $c, c' \in \mathscr{C}$.

The random shocks demeaned by cohort are assumed to be independent of turnover incidents (assumption 2.1). They can be serially correlated between degree programs within a student (assumption 2.2) and between students in each degree program if they are supervised by the same advisor (assumption 2.3). However, they are neither cross- nor serially correlated (assumptions 2.4 and 2.5). Note that, even if the demeaned random shock, $\tilde{\nu}^c_{iag}$, is uncorrelated with others under assumptions 2.4 and 2.5, the original random shock, ν^c_{iag}, is allowed to be correlated through the common mean factor, $\bar{\nu}_g$.

The Regression Model

Finally, given the assumptions presented above, we square both sides of Equation (5) and take the expectation conditional on the occurrence of turnover. We have the following result:[13]

$$\mathrm{E}\left[\left(DD\overline{\Delta outcome}^{\ell(a,c,c')}\right)^2 |W^{\ell(a,c,c')}\right] = \alpha\left(\frac{1}{I^{\ell(a,c)}} + \frac{1}{I^{l(a,c')}}\right) + \left\{2\sigma_d^2(1-\rho_d)\right\} W^{\ell(a,c,c')}, \quad (6)$$

where we define $\alpha = \{\phi_d^2 + \phi_m^2 + 4(\psi_d + \psi_m) - 2\phi_{dm}\}$.

Equation (6) provides a basis for estimating the variance in advisor quality at the doctoral level. Using the cohort examples, c_0, c_1, and c_2, that are depicted by Figure1 for illustration, the squared difference measure of student research outcome growth, which is the right-hand side of Equation (6), is greater for $\ell(a, c_1, c_2)$ than that for $\ell(a, c_0, c_1)$ by $2\sigma_d^2(1-\rho_d)$. We can therefore ascribe the large sample variation of the right-hand side of Equation (6), if any, to the variance in doctoral-level advisor quality, σ_d^2, unless the correlation coefficient, ρ_d, is equal to one.

We now present a regression model to obtain a lower-bound estimate of the variance of σ_d^2. Consider the following:

$$\left(DD\overline{\Delta outcome}_n\right)^2 = \alpha X_n + \beta W_n + \varepsilon_n, \quad (7)$$

where $n = 1, \cdots, N$ is the index of observations. Here, the unit of observation is each element of (a, c, c') for any advisor $a \in \mathscr{A}$ and cohort c, c' such that $0 < c - c' \leq \tau$, where τ is the period over which the difference is taken.[14]. Note that, analogous to Equation (6), the covariate $X_n = 2(1/I^{\ell(a,c)} + 1/I^{l(a,c')})$ is introduced into the regression. The random term ε_n is interpreted as the prediction error between the expected and observed values of the divergence measures, that is:

$$\varepsilon_n \equiv \mathrm{E}\left[\left(DD\overline{\Delta outcome}_n\right)^2 \middle| W_n\right] - \left(DD\overline{\Delta outcome}_n\right)^2.$$

Assume for a moment that the advisor switch indicator, W_n, is independent of the prediction error, ε_n. If the value of ρ_d were known perfectly, the OLS estimate $\hat{\beta}$ in Equation (7)

[13] See Appendix A for the derivation.

[14] To obtain the double-differenced average of the research outcome gain, which is the left-hand side of Equation (7), we take the difference between all cohorts within a period of τ years. As $\binom{\tau+1}{2} = \frac{\tau!}{(\tau-2)!2!}$ samples are created for each lab, the total sample size of the regression is given by $N = \frac{\tau! L}{(\tau-2)!2!}$, where L is the total number of labs.

would provide a consistent estimate of σ_d^2 through the following equation:

$$\hat{\beta} = \left\{2\hat{\sigma}_d^2(1-\rho_d)\right\}. \tag{8}$$

As the correlation is imperfect ($\rho_d < 1$), a lower-bound estimate of σ_d^2 is given by the last term of the following equation:

$$\hat{\sigma}_d^2 = \frac{\hat{\beta}}{2(1-\rho_d)} \geq \frac{\hat{\beta}}{4}. \tag{9}$$

In other words, a lower-bound estimate of the within-school variance of faculty quality at the doctoral level is equal to the estimated coefficient, $\hat{\beta}$, of the regression model (7) divided by four.

4 Data

We assemble data sets of professors and students in a graduate program in physics in Japan. Among the numerous Japanese research universities that offer both master's and doctoral programs in the field of physics, we focus on the graduate program at UTokyo, which is the oldest institution of its kind in the country and has enjoyed high prestige in the global academic community.[15]

The graduate program in physics at UTokyo consists of the department of physics as its core and other physics-related research institutes on campus.[16] The average number of graduates in recent years is 105.6 for the master's program and 58.4 for the doctoral program[17]. At present, there are more than 130 full-time faculty members. Many subfields of physics are covered by laboratories in UTokyo's physics graduate programs, such as nuclear physics, particle physics, condensed matter physics, and biophysics.

4.1 Data on Advisor and Advisee Pairs

To extract the information on matched advisor-advisee pairs, we use the master's and doctoral thesis catalogs for graduate students in UTokyo's physics program.[18] For each thesis entry in the catalog, the available information includes the degree date, the title of the thesis, the name of the student, and the name of the faculty advisor who supervised the student.

[15] According to several world university rankings, UTokyo has been in the top 10 in the discipline of physics. The alumni include five Nobel laureates in physics as of 2015.

[16] The institutes are the Institute of Cosmic Ray Research, (ICRR), the Institute of Solid State Physics (ISSP), and the International Center for Elementary Particle Physics (ICEPP).

[17] These are the average figures over the period from 2010 to 2014.

[18] The catalogs are available on the department's website at `http://www.phys.s.u-tokyo.ac.jp/TOSHO/ronbun.html`.

We compile the thesis data for the students who obtained their doctoral degrees in the cohorts between 1970 and 2004 (35 years). Among all of the graduate students who were listed in both the master's and doctoral thesis catalogs, we restrict our attention to those who earned doctorates within six years of enrollment. In addition, we restrict the analysis to those who were supervised by faculty members with the ranks of full and associate professors in the physics department or on-campus physics-related research institutions.

4.2 Data on Advisor Turnover and Switch

We obtain information on faculty turnover from the University Personnel Directory Book ("*Zenkoku Daigaku Shokuin Roku*") published by *Koujyun Sha*, which includes information on the full name, rank, department, school, specialized fields and year of birth of all staff members at every Japanese university, public or private, in a given year. By compiling the roster of faculty members at UTokyo, we can obtain their turnover information.

We identify turnover as a case in which a faculty member left UTokyo. We classify the reasons for turnover into the following three categories: (1) retirement if the instance of turnover occurred at the mandatory retirement age predetermined by UTokyo;[19] (2) move if turnover occurred before the retirement age and the faculty name began to reappear on other universities' roasters beginning in the year after the turnover instance; and (3) decease/quit otherwise.[20]

Figure 2 presents the graphs that plot the number of turnover incidents in each year of the sample period, broken down by the reasons.[21]

Insert Figure 2

The matched advisor-advisee data reveal that approximately 14.4 percent of graduate students switched advisors between the master's program and the doctoral program. Instances of professor turnover are responsible for some, although not all, of the students' observed

[19]Before fiscal year 2000, the mandatory retirement age at UTokyo was 60. After the 2001 fiscal year, it was increased by one year every three years until it reached 65. As of 2004, which is the end of the sample period, the retirement age was 61.

[20]However, note that the reasons for faculty turnover are not perfectly distinguishable. Indeed, the majority of faculty members categorized as "retire" did not actually retire from academic life and were reemployed at other universities or research institutions. This is possible because of the gap in retirement ages between universities: UTokyo set its faculty retirement age at 60 during the most of the sample period, while other Japanese universities, public and private, adopted retirement ages that were several years older.

[21]There is a considerable number of incidents in 1997, when the Institute for Nuclear Study (INS) at UTokyo, which was one of the on-campus research institutes affiliated with the physics department, was closed and merged with the High Energy Accelerator Research Organization (also known as the KEK (Kō Enerugī Kasokuki Kenkyū Kikō.), and some of the faculty members at the INS chose to leave UTokyo for the KEK.

changes in advisors. As mentioned previously, in Japanese universities, a joint transfer of faculty member and student is quite rare. If a faculty member exits a graduate program, another other faculty member – usually a sub-leader of the same lab or, sometimes, a faculty member from a different lab in the same institution whose research area is closely related to the professor who exited – becomes the new advisor of the students who are left behind. In either case, the student remains in the same program.[22]

We identify an advisor switch due to turnover if a student's master's thesis advisor exited UTokyo before the student earned a doctoral degree. Such cases account for 53.2 percent of all advisor switches in the sample. We exclude students who switched advisors on their own initiative from the sample observations, as such student-side advisor switches are likely to be caused by a mismatch between advisor and advisee and could be correlated with student research outcomes. Ultimately, the resulting sample contains 801 students and 158 advisors, and this sample is used to estimate professor value added in what follows.

4.3 Data on Student Research Achievement

To measure a graduate student's research achievement, we use the number of journal articles that he or she published. To obtain this information, we employ the Thomson Reuters Web of Science (WoS) archive. We collect physics articles with author names that match the name of the graduate student under consideration. We further restrict our attention to those articles published around the period when the target student was enrolled.

The articles selected by author name matching may contain false positives: these articles could have an author who coincidentally has the same name as the graduate student in the sample but is in fact a different person. To minimize such identification errors, we add a further restriction; that is, for an article to be identified as written by the student in question, we impose a restriction that the words in the article title should overlap to some extent with those in the title of the master's or doctoral thesis. [23]

Based on a student's publication records, we define the *research proficiency score* as the number of publication counts during *a given year*. Here, we employ two quality adjustment methods. First, we limit the publications to those published in twelve high-quality peer-reviewed journals, including three high-reputation general-interest science journals and nine highly ranked physics journals.[24] Second, we consider a student's share of credit for an article

[22]It is often noted that Japanese graduate students are loyal not to their advisors but to their labs. Cultural norms dictate that each member of a lab is expected to keep its resources intact and pass them on to the next generation (see, e.g., Traweek p.148), which is congruent with the analogy of *ie* (the household) to labs in Japanese universities, as described in the previous section.

[23]See Appendix B for details on the score of word overlap in titles.

[24]*Nature*, *Science* and *Proceedings of the National Academy of Sciences of the United States of America*

if there are multiple authors. In physics, as in other scientific disciplines, papers are usually written by a group of authors whose contributions are not necessarily equal. We follow a standard bibliometric method (e.g., Liu and Fang, 2012; Waltman, 2012) based on the byline hierarchy rule to quantify an coauthor's share of credit for an article with multiple authors.[25]

Figure 3 plots the average research proficiency score for our sample graduate students in each year. Note that, in the figure, we begin the graduate school year index at one in the year when a student entered the master's program and increase it throughout the duration of the graduate program. For the sake of expedience, the graduate school year is also defined for the postdoctoral period after the student obtained a doctorate degree. In the figure, it corresponds to the period after the 6th year.

Insert Figure 3

The figure illustrates the time pattern of how physics graduate students at UTokyo develop their research outcomes: the achievement curve rises and reaches its peak in the years near the completion of the doctoral degree (D1 and P1). Then, the research outcomes begin to decline during the postdoctoral periods (P1-P5). We suspect that this reflects two types of lag structure: the first relates to a publication lag, that is, the time lag from the submission to publication of articles in journals. The second concerns a gestation lag, that is, the time lag between project inception and completion.

5 Empirical Issues

In this section, as a starting point for our empirical analysis, we describe the empirical issues involved in estimating a lower bound of professor quality based on the regression model in Equation (7). We first address how to construct the squared difference measure of the student outcome growth variable, which is used as the dependent variable in the regression model.

(PNAS) are included as the general-interest science journals, and *Physical Reviews A, B, C, D, and E*; *Physical Reviews Letters*; and *Physics Letters A and B* are included as the top physics journals. We received advice from physicists regarding the selection of the top journals.

[25]What follows illustrates how the coauthor's credit share is constructed. Suppose that the names of the authors are ordered alphabetically. Then, the contribution weight is fractional: each author receives equal credit. Suppose this alphabetical approach is not used. Then, each author receives a share of credit that decreases in the authorship ranking. Following Liu and Fang (2012), the credit formula is given by $n^{-1/k}r^{-(1-1/k)}$ for the r-th author of a paper with n authors. The integral constant, k, controls the declining rate of credit allocated in proportion to that of the first author. According to the suggestion of Liu and Fang (2012), we set $k = 3$ for our analysis. Waltman (2012) notes that authorship could unintentionally be alphabetical, especially when the number of authors is small, despite the authors' intention to list their names based on a non-alphabetical criterion. Therefore, we account for the probabilities of both such incidental and intentional alphabetical authorship and use the expected value as the final research outcome measure.

We next discuss the non-randomness of professor turnover, which could cause an endogeneity problem and thus threaten the validity of the estimates. We then propose a method to address this endogeneity concern.

5.1 Student Research Outcome Variable

As presented in Section 3.2, the regression model is based on the double-differenced student research outcome measure, which requires systematic difference — the first-stage difference is taken with respect to the degree program g, and then, the second stage is taken regarding cohort c.

Two issues arise: (i) the choice of years over which the student research outcomes are aggregated at the program level [26] and (ii) the choice of interval years between the pair of cohorts that are differenced.

Regarding the first issue, which publications should we count as research outcomes of the master's program and which as those of the doctoral program?

Figure 4 presents the student average research proficiency scores that are decomposed into those related to the master's thesis and those related to the doctoral thesis.[27] The findings indicate that the proficiency score associated with the master's thesis peaks in the second year of the doctoral program (D2) and decreases thereafter, while the score related to the doctoral thesis continues to increase. We thus opt to aggregate the research proficiency scores over the period from the first year of the master's program (M1) to the second year of the doctoral program (D2) to compute the research outcome at the master's level. However, for the research outcome at the doctoral level, we assemble the research proficiency scores from the first year of the doctoral program (D1) up to the fourth year of the postdoctoral period (P4). We choose a rather long aggregation period at the doctoral level in light of the lag between the time of article publication and the time the degree is awarded, as seen in Figure 3.

Insert Figure 4

In sum, our benchmark student research outcomes are aggregated over the period from M1 to D2 and the period from D1 to P4 for the master's degree and doctoral degree programs, respectively. Table 1 presents the descriptive statistics. Figure 5 presents the box plots of the

[26] Because the value-added model focuses on the student research achievement gain while in school, the magnitude might be minute and unnoticeable if it is measured by the annual gain. We thus select the unit of measure as each *degree program* period.

[27] As explained in Section 4.3, to implement the decomposition, we classify student articles as those related to the master's thesis and those related to the doctoral thesis by considering the overlap of the title of the article and that of the thesis.

research outcome distributions at the master's and doctoral levels.

Insert Table 1

We turn to the second issue concerning the interval in years between cohorts. In Section 3.2, τ denotes the number of years between two cohorts, c and c', such that $c - c' \leq \tau$ when determining the double-difference student research outcome growth. Note that there is no theoretical rule for which year should be used as τ. On the one hand, the longer the interval is, the more efficient the estimator because it yields more samples for the regression analysis.[28] On the other hand, the shorter interval is, the better because it requires a weaker assumption on the covariance stationarity of the distribution of the demeaned random shocks (assumption 2.3).[29] In light of balance, we adopt the adjacent cohort period of $\tau = 3, 4$ and 5 years as the benchmark when implementing the regression.

Insert Figure 5

5.2 Non-Random Turnover

Thus far, we have assumed that professor turnover is independent of various factors in the value-added model and thus does not affect student research performance except through the change in advisor quality. However, the assumption might be untenable. Arguably, a professor's decision of whether to retire, move, or remain at a graduate program might be endogenous to the student's performance.

Consequently, the regression model in Equation (7) might suffer from the standard endogenous variable problem, as the catch-all error term, ε_n, which influences student research outcome growth, will be confounded by the advisor switch dummy variable, W_n, through the heterogeneity of advisors, who systematically differ between those with and without turnover. In this case, we might not be able to obtain an unbiased estimate of β from the regression and thus be unable to obtain a reliable estimate of the lower bound of advisor quality.

Table 2 reports the descriptive statistics for some characteristics of advisors and compares those of advisors when turnover occurred and the corresponding advisor characteristics when it did not.[30] We find that, for some characteristics, the differences in means between the two groups, professors with turnover in column (1) and those without in column (2), are

[28] As presented in footnote 14, the total sample size is given by $N = \frac{\tau! L}{(\tau-2)!2!}$, which is an increasing function of the adjacent period, τ, ceteris paribus.

[29] To be more precise, assumption 2.3 states that the covariance of the demeaned error terms is constant between any two students, i and j, in different cohorts, c and c'. This assumption might be reasonable only for adjacent cohorts.

[30] The research proficiency scores of professors are computed in the same way as those of students. The score is, in essence, the number of publications in top general and physics journals, with the coauthor's credit share being adjusted. The data source is the WoS.

statistically significant at the 5 percent level. We also find that the absolute values of the standardized differences, reported in column (3), are large for some characteristics.[31] Therefore, this suggests that the sample is not balanced, that is, there are systematic differences between the groups with and without professor turnover on some characteristics.

Insert Table 2

To make the sample balanced and comparable, we employ a propensity score matching method. The basic idea is to match a turnover case with a case of no turnover that has approximately the same conditional likelihood, typically called the propensity score, that an incident of advisor turnover would have occurred. After constructing a new balanced sample based on the propensity score matching procedure, we estimate the regression model in Equation (7) using the balanced sample, as if advisor changes due to turnover occurred at random.

Note that, to account for the endogeneity of the advisor switch dummy variable in the regression model, we only control for advisor characteristics. It is potentially justifiable not to balance the sample on student characteristics because we exclude all cases in which a change in advisor occurs for a student's own reasons, as described in Section 4.2. The sample restriction can eliminate the possibility that student factors are confounded with the occurrence of an advisor switch, and therefore, it is deemed to occur exclusively for reasons on the faculty side. Hence, we control for the professor's characteristics in the propensity score analysis.

Following standard practice in the literature, we estimate the propensity scores using a logit model. We include all of the characteristics presented in Table 2 when estimating the propensity scores. We determine a baseline specification of the model by a stepwise likelihood-test-based procedure, suggested by Imbens (2014) and Imbens and Rubin (2015).[32] The results of the logit estimation of the propensity score can be found in Appendix C.3. Given the estimated propensity scores, we match a case with $W_n = 1$ (a lab with an advisor switch) to one with $W_n = 0$ (a lab without an advisor switch) that share approximately identical estimated propensity scores. We employ a one-to-one nearest-neighbor matching method.

[31] The standardized difference considers the size of the difference in means of a conditioning variable, scaled by the square root of the variances of the treatment and control groups in the original sample. According to the suggestion of Rosenbaum and Rubin (1985), an absolute value of the standardized difference greater than 0.2 should be considered "large".

[32] Specifically, in the first step, we begin with a set of basic covariates and add an additional linear term based on a likelihood ratio test for the null hypothesis that the coefficient of the added variable is equal to zero. In the second step, we proceed to the choice of the quadratic and cross-product terms and apply the same type of likelihood test as that used in the first step. We follow the suggestion of Imbens and Rubin (2015) that the threshold values for the likelihood ratio test should be $C_L = 1.0$ and $C_Q = 2.71$ for the linear and quadratic terms, respectively.

To assess the quality of the propensity score matching, we present Figure 6 that depicts the absolute values in the standardized differences of the variables for the original and matched samples. The imbalance between the treatment and control cases is attenuated on many professor characteristics. For example, professor's age differs between the treatment and control labs by more than the average standard deviation (the absolute standard deviation is 1.129) before matching, whereas the difference is considerably reduced (the absolute standard deviation is 0.006) after matching.

Insert Figure 6

Figure 7 presents the distributions of the estimated propensity scores for the treatment labs (left) and control labs (right) in each case of the adjacent period, $\tau = 3, 4$, and 5. The top and bottom groups in the graphs correspond to those before and after matching, respectively. Before matching, the shapes of distributions differ considerably between the treatment and control groups. Nevertheless, the propensity score distributions have some degree of overlap. Moreover, after matching, the dissimilarity of the distributions between the treatment and control groups is considerably reduced.

Insert Figure 7

One might worry that the spread of the common support of the propensity score distributions should not be across the full range [0 1] and hence that the observations of the treatment group, especially those with high propensity scores, are matched forcibly with those of the control group, the propensity scores of which are not sufficiently close. To address the problem caused by the limited common support of the propensity score distribution, we employ a systematic approach proposed by Crump et al. (2009) and discard all observations with estimated propensity scores outside the range of [0.1 0.9].

6 Estimation Results

6.1 Benchmark Results

This section presents the estimation results for professors' value added to the students' research achievement gains. We estimate the econometric model (7) using the propensity score matching method that we described in the previous section. The main estimate of interest is the lower bound of the variance in advisor quality at the doctoral level, which is given by one-fourth of the coefficient of the advisor switch indicator variable in the regression model.

Table 3 presents the baseline results. We report the regression estimates in rows (1) and (2). Columns (1), (2) and (3) are used to report the estimation results for the three cases of adjacent periods between cohorts, τ = 3, 4 and 5 years, respectively. As the estimated propensity scores are used for the true values, we compute resampling-based standard errors to correct for the additional sampling variability arising from estimation.[33] All estimates of βs are positive and statistically significant from zero at the 10 percent level except for one case.

Insert Table 3

Row (3) of Table 3 presents the estimated lower bound of advisor quality variance at the doctoral level. As the variance must be non-negative, we perform one-sided tests on the lower-bound estimates such that $\sigma_d^2 = 0$ against the alternative $\sigma_d^2 > 0$. The results indicate that the null hypothesis is rejected at least at the 5 percent level for all cases, indicating that a professor's quality has a measurable effect on the research performance growth of the student to whom he or she is assigned.

For the results that we have presented thus far, we base the student research outcome on the research proficiency scores that are adjusted for the share of credit of each author. Alternatively, we can quantify the research outcome of a student *without* credit share adjustment. To this end, we count the number of *first-authored* articles that the student published as a lead author in the selected top general and field journals in physics. While the alternative research outcome measure might be crude and subject to a certain amount of noise — it might underrate the research achievement of a student because it ignores the articles for which he or she is not a lead author, or it might overrate the student's attainment because it accords him or her all of the credit, even for multi-authored articles, irrespective of how many coauthors are involved — it nonetheless serves as a simple and easily interpreted yardstick.

The estimation results using the alternative research outcome measure are presented in columns (4) to (6) of Table 3. The regression estimates are larger than previous results that adjusted the author's credit share. This is unsurprising because the first-author-based measure is greater than the original measure to the extent that the credit share is not weighted.[34] The estimated values of the lower bound of σ_d^2, reported in row (3), are correspondingly larger than those previously reported. Reassuringly, the null hypothesis that the variance in advisor quality is zero cannot be rejected at least at the 5 percent level. We therefore obtain qualitatively similar evidence on the professor's value added as previously.

The results presented above indicate the effectiveness of professors in improving doctoral students' research productivity growth. Indeed, better advisor quality causally affects

[33] Abadie and Imbens (2008) demonstrate that the bootstrap method generates biased estimates of the standard errors for a nearest-neighbor matching estimator and suggest the subsampling method developed by Politis and Romano (1994). We therefore use the subsampling method whereby we draw fewer observations than the same size at each iteration without replacement.

[34] The mean and standard deviation of the first-author-based research outcome at the doctoral level are 0.39 and 0.96.

advisees' research achievement gains in graduate school. If we use 0.0489 as the most conservative estimate of the lower bound of the advisor quality variance among those reported in columns (1) to (3) of Table 3, we find that a one-standard-deviation increase in professor quality raises the average student research achievement gain at the doctoral level by at least 0.221, which corresponds to approximately 0.54 standard deviations of the total doctoral program research outcome distribution.

If we base the estimation results on the first-author-based research outcome measure reported in columns (4) to (6) of Table 3, we find that, if professor quality increases by one standard deviation, the average student publishes 0.64 more first-authored articles in top journals at the doctoral level.[35] We are thus able to conclude that professor's value added to graduate student research outcomes is substantial.

Our estimates of value added provide an interesting comparison with the professor value-added estimates at the undergraduate level reported by previous studies. For example, Hoffmann and Oreopoulos (2009) estimate professor value added to student's achievement gains, measured by undergraduate course grades in a large Canadian university. They report that a one-standard-deviation increase in professor quality yields an approximately 0.05 standard deviation increase in a student's grade. Carrell and West (2010) obtain a similar value-added estimate for professors at the U.S. Air Force Academy who teach introductory courses at the undergraduate level. They report that the standard deviation of value added is approximately 0.05. Therefore, our estimates of professor-value added at the postgraduate level are substantially larger than those standard-deviation estimates at the undergraduate level.

The observed difference in the estimates might not be too surprising considering several factors that make our study distinct from other studies. First, the professor quality that we measure is different. We evaluate the dimension of professor quality that promotes a student's *research* capability, whereas those previous studies assess the aspect of quality that enhances a student's *academic* capability. Second, closely related to the first point, the student outcome is different. We focus on the research achievement gains of postgraduate students, while previous studies investigate the academic achievement gains of undergraduate students. Finally, the estimation method is different. Our estimation method, following that of RHK, is based on professor turnover and provides a lower bound of professor quality. By contrast, the approach employed by Hoffmann and Oreopoulos (2009) is based on the covariance estimation procedure proposed by Page and Solon (2003) and is interpreted as an upper bound of professor quality. The estimation method used in Carrell and West (2010) is a random effect estimation of unobserved professor quality, relying on the fact that the courses are randomly assigned to students and, therefore, that no issue of self-selection arises.

[35] When computing the standard deviation increase, we use 0.410 as the estimated value of the lower bound of advisor quality variance.

We hope further studies will add evidence on the difference in professor value added between undergraduate and postgraduate education.

6.2 Robustness Tests

This section provides various robustness checks for the benchmark results. First, we implement a falsification test that investigates whether a false instance of an advisor switch predicts an increase in the volatility of student research outcomes between programs and cohorts. Second, we perform specification checks to examine whether the benchmark results are robust to alternative definitions of the student research outcome. Third, we discuss the possibility that the lower bound of the estimate of the advisor quality variance might be overestimated.

Falsification Test

In our estimation framework, the variance in advisor quality is identified by an increase in the squared difference of the student research outcome gain at the time of advisor turnover. We thus implement a falsification exercise that examines whether the timing agrees with what is predicted by the empirical model.

To do so, we construct a *false* advisor switch dummy variable, $\tilde{W}_n$, that takes value one for the lab in one cohort before the actual incident and zero otherwise. Specifically, given lab $\ell(a,c)$, where advisor turnover occurred, the variable $\tilde{W}_n$ is one in the *latest* cohort, c', in which advisor a supervised at least one student before cohort c. We estimate a regression similar to regression model (7) using the dummy variable $\tilde{W}_n$ as the regressor instead of using the true advisor switch dummy variable, W_n, with $\tilde{\beta}$ being the coefficient of the variable $\tilde{W}$.

We present the results in Table 4, where we adopt the same definition of the student research outcome measures as the baseline case, and replicate the regression results except that we use the false advisor switch dummy variable. Columns (1) to (3) show the results for the credit-share-based research outcome measure, and columns (4) to (6) show those for the first-authored-paper-based research outcome measure. The false advisor switch dummy variable is sometimes negative and has no systematic impact on the the squared difference of the student research outcome gain. Indeed, in all cases except one, the false advisor switch dummy variable is not statistically significant, suggesting that the results survive the falsification test. As there is no clear sign that the previous results can be explained by a spurious trend, we might be able to conclude that the timing of increased volatility in the student research outcome gains is consistent with that of advisor turnover.

Insert Table 4

Specification Checks

One might wonder whether our estimates are sensitive to specific assumptions on the definition of the student research outcome. To verify the robustness of the estimates to these assumptions, we consider alternative configurations in terms of the period over which the research proficiency scores are aggregated for each degree program. Specifically, in addition to the benchmark case (M1-D2 for the master's program and D1-P4 for the doctoral program), we examine alternative cases that change the aggregation period at the master's and doctoral levels.

Table 5 summarizes the set of lower-bound estimates of advisor quality under various definitions of student research outcomes. We employ the same specification and the same propensity-score-based estimation method as in the baseline case. For the purpose of comparison, the first row reports the corresponding estimate from the baseline case. We examine several different aggregation periods for both the baseline and alternative research outcome measures. All of the results are qualitatively similar to the previously reported findings. The null hypothesis that the variance in doctoral-level advisor quality is zero is rejected at the 10 percent level in all cases.

Insert Table 5

We also provide additional robustness tests regarding whether the results are driven by a specific value of the threshold that is used to compute students' research proficiency scores. As explained above, we consider research articles that are actually published by a target student if the author's name matches the student's name and, in addition, the degree of word overlap in the titles between the article and the student's thesis exceeds some predetermined threshold value. While the default value is set to minimize both type 1 and type 2 errors, we employ both over-matching and under-matching criteria in the robustness exercise. The results presented in Table C.2 in Appendix C.4 show that, while some estimates are not statistically significant in the cases in which the adjacent cohort period is five and the over-matching criterion is used, they tend to be positive and statistically significant. Despite the insignificant estimates, our conclusion regarding an advisor's effectiveness in improving an advisee's research productivity growth appears to be supported on the grounds that the value-added estimates are lower-bound estimates.

We turn to issues concerning the quality of research publications when computing student research achievement. In the benchmark case, we select twelve top journals (three general-interest science and nine physics journals). To examine the sensitivity of our estimates to the particular choice of top journals, we replicate the baseline analysis by narrowing the coverage to nine journals (two general-interest science and seven field journals) instead of twelve journals.[36]

Table 6 presents the estimation results. Although the estimates as a whole become smaller than those for the case of broader journal coverage, they are qualitatively unchanged, indicating that the findings from the regression model are not merely artifacts of the specific choice of top journals.[37]

Insert Table 6

In summary, considering all of the estimation results presented above, we can conclude that the specification of the student research outcome measure has little or no systematic effect on the estimation of professor value added.

Factors That Might Lead to Upward Bias

As we are interested in estimating a lower bound of the variance in advisor quality, downward bias would not be problematic, as is the case for the original turnover estimator that RHK propose. There is, however, a set of potential sources of upward bias.

The first possibility is that the assumption on the time-invariance of advisor quality, given by assumption 1.1, might be violated. Suppose, contrary to the assumption, that it varies across cohorts *within* a professor. In particular, if it fluctuates as the end of a professor's research career approaches, the squared difference measure, the dependent variable in regression model (7), becomes more volatile in the last cohorts before a professor's turnover. In this case, the regression coefficient of the advisor switch dummy variable might overstate the lower bound of advisor quality variance, as the increase in the dependent variable, which is indeed caused by within-advisor quality change, is mistakenly attributed to a systematic and discrete change in advisor quality due to turnover, despite that it should not be.

To shed some light on this concern, we augment the regression model in Equation (7) by including a set of dummy variables that capture the possible change in advisor quality variance in the period near turnover. Specifically, the dummy variable $D_k^{(a,c,c')}$ takes value one if cohort c is within k years before professor a exited and zero otherwise, for $k = 1, 2$, and 3.

The estimation results from the augmented specification are presented in Table 7, with δ_k

[36]The three of the original twelve journals excluded here are *PNAS* in the general-interest science journal category and *Physics Letters A and B* in the field journal category. This is based on suggestions that we received from several physics researchers.

[37]Table C.3 in Appendix C.4 summarizes the estimation results for the lower-bound estimates of advisor quality at the doctoral level for the case in which the student research outcome is based on the top nine journals and aggregation years are allowed to vary. The results demonstrate that the null hypothesis that the variance in doctoral-level advisor quality is zero is rejected at the 10 percent level for the majority of the cases, although we hasten to add that it cannot be rejected for some cases. Nevertheless, all of the estimates of the variance are positive.

being the coefficient of the dummy variable $D_k^{(a,c,c')}$. As reported in rows (3) to (5), for both the baseline and alternative research outcome measures, none of the coefficients concerning the added dummy variables are statistically significant. Moreover, row (6) indicates that the estimated coefficient of the advisor switch dummy variable is not substantially affected by the inclusion of the cohort-specific dummy variables. Furthermore, encouragingly, the null hypothesis that advisor quality in the doctoral program has no effect on student research outcome growth is rejected at least at the 10 percent level in all cases.[38] On the basis of this evidence, we obtain the same conclusion regarding professor value added even if we allow for the possibility of time-varying advisor quality.

Insert Table 7

Another possibility that might introduce upward bias into the lower bound of the variance in advisor quality concerns the allocation of the research *credit share* between advisor and advisee. Note that our empirical study relies on the assumption that the student made an original and substantial contribution to his or her thesis research projects and that the articles with titles that are closely associated with the master's thesis or doctoral dissertation can be used as an unbiased yardstick to gauge the student's in-school research achievement. The assumption appears somewhat reasonable for physics departments in Japanese universities, where, as described in Section 2.2, graduate students are typically accorded a fair amount of autonomy when choosing a research topic and approach.

Nevertheless, the assumption might not be tenable. One could imagine that students are merely given a part of a larger research project, or subtopic, that the advisor has pursued, and thus, their contribution to the project in collaboration with their advisors is marginal.[39] If this is true, our turnover estimator for the lower bound of the variance in advisor quality might suffer from systematic upward bias, as we would then mistakenly ascribe the advisor's research contribution to the student's research achievement.

Because the actual collaboration process is not observed for joint research activities, it is impossible for us to allocate the true share of credit to each member of an advisor-advisee pair that engaged in a joint research project. We therefore consider an extreme case in which the student's contribution is *zero* whenever he or she collaborated with a research advisor to highlight the sensitivity of the previous estimation results to the assumption on the allocation of research credit.

[38] Table C.4 in appendix C.4 presents the estimation results for the lower bound of advisor quality variance when we change the aggregation period for the student research outcomes. The estimates are qualitatively similar to those reported in Table 7.

[39] The view that attributes substantially greater credit for knowledge contribution to an accomplished senior researcher than to a less-known junior researcher is referred to as the "Matthew effect," a term coined by sociologist Robert K. Merton.

Table 8 presents the estimation results for the cases of the baseline and alternative student research outcome measures, assuming that the research proficiency score of student publication is equal to zero if it is coauthored with the advisor.[40] Looking across the columns of the table, the size of the estimated coefficients and the lower bound of advisor quality variance tend to be lower. Nonetheless, the one-sided test of the null hypothesis that doctoral-level advisor quality has no effect on an advisee's research achievement growth is rejected at the 10 percent level. Because we consider a severe restriction on the allocation of the credit share to the side of advisees, which is overly severe for the advisees in terms of their research contributions, the reported evidence of positive professor value added reassuringly supports the conclusion that professors enhance their students' research achievement gains by advising and mentoring their research projects at the postgraduate level.

Insert Table 8

6.3 Additional Evidence for Professors' Influence on Students

Other Mechanisms

The estimation results have shown that advisor turnover generates significant variations in an advisee's research achievement gains at UTokyo's department of physics. According to a standard value-added model, we ascribe the increased diversity of student research achievement gains to the discrete change in advisor quality at the time of turnover. Admittedly, however, there may remain other mechanisms that create such a pattern.

One possibility is that professor turnover always has a *positive* effect on students' research capacity and thus increases the variability in student achievement gains between cohorts with and without turnover. The positive advisor turnover effect could be caused by a mechanism that reflects a well-known understanding that innovation (and thus economic growth) is due to the recombination of existing ideas (e.g., Weitzman, 1998). It follows from this view that, as new innovation is likely to arise from recombining old knowledge elements, students who are supervised by different professors would have access to a wider variety of knowledge and ideas and can thus enhance their research capabilities.

Another mechanism is the one that yields a *negative* effect of professor turnover on students' research achievement gains. As is often noted in the education literature (e.g., Wisker

[40] In Table C.5 in appendix C.4, we report the estimation results of the lower bound of the advisor's quality variance when the aggregation period for the student research outcome is allowed to differ. These estimates are reassuringly statistically significant at the 10 percent level in more than half of the cases. Note that, in particular, if the research outcomes are measured by the number of the first-authored papers, then all estimates are statistically significant.

and Robinson, 2013), if an advisor is lost due to turnover, an advisee who becomes an "orphan" occasionally perceives this as a traumatic event and suffers from psychological problems that might occasionally result in under-development of academic achievement. If this understanding is correct, advisor turnover would retard the advisee's research progress, irrespective of how high the quality of the newly assigned advisor is, and thus generate a noticeable gap in student research outcome gains between cohorts with and without turnover.

Recall that, according to the mechanism captured by the value-added model, the advisee's research outcome growth can be positive or negative after turnover – indeed, as explained in Section 3.1, the direction of growth depends decisively on the relative levels of advisor quality that were switched when turnover occurred and will thus not be predicted *a priori* unless the information on the exact quality levels is available. In sum, because under all of the mechanisms presented above, advisor turnover can generate divergence in advisee research achievement gains, the baseline regression specification given in Equation (7) would misattribute the effects of turnover that are essentially attributable to different mechanisms.

In the analysis that follows, we investigate which mechanism is more likely by estimating a regression similar to the regression model in Equation (7), except with the dependent variable being *in levels*, $\left(DD\overline{\Delta outcome_n}\right)$, not *in squares* $\left(DD\overline{\Delta outcome_n}\right)^2$. To identify the mechanism in place, we focus on the sign of the turnover effect on the advisor's research achievement gain. As explained, if the first alternative mechanism (i.e., a student's research derives from the recombination of advisors' ideas) dominates the others, the coefficient of the advisor switch indicator dummy variable will be positive in the regression model with the double-differenced student achievement measure in levels as the dependent variable. However, if the second mechanism (i.e., a student's research progress is hampered by a traumatic experience triggered by advisor turnover) dominates, that coefficient will be negative.

Table 9 presents the regression results for which all estimated coefficients of the advisor switch indicator are shown to be positive but are not statistically significant in all cases. We can interpret the results as indicating that, contrary to the predictions of the alternative mechanisms, advisor turnover can have a positive or negative impact on an advisee's research productivity growth. As the individual impacts cancel one another out, the aggregate effect, as reflected by the integration, is not significantly different from the null in levels. It thus appears to confirm that the mechanism that the value-added model postulates should be a main driver of the empirical findings obtained thus far and to endorse the conclusion that professor quality plays a distinct role in enhancing a student's research capacity in the doctoral program.

Insert Table 9

Indirect Influence

Our analysis thus far has concentrated on the advisor-advisee relationship within a lab and intended to measure the effectiveness of knowledge transmission through a direct interaction channel within a lab. However, knowledge might be transmitted beyond lab-oriented master-apprenticeship-style contact. There might also exist an indirect transmission route across labs. For instance, students will learn, formally or informally, research skills and expertise in their discipline from faculty members who are not their supervisors through coursework, lectures or collaboration opportunities.

To measure such an indirect effect from non-advisor faculty members on students across labs within the same institution, consider the following augmented model of student research outcome gains:

$$\overline{\Delta outcome}_{eg}^{\ell(e,c)} = \bar{\gamma}^{\ell(e,c)} + \theta_{eg} + \sum_{f \in \mathscr{A}} \pi_{ef}\theta_{fg} + \bar{\nu}_{eg}^{\ell(e,c)}, \tag{10}$$

where we consider lab $\ell(e,c)$ of professor $e \in \mathscr{A}$ in cohort $c \in \mathscr{C}$. We modify the baseline specification in Equation (2) by incorporating an "indirect" effect of professor f on the average research outcome gain in program g for students in lab $\ell(e,c)$ who are supervised directly by professor e, where $e \neq f \in \mathscr{A}$. The magnitude of the indirect influence from non-advisor faculty member f is captured by the parameter π_{ef}, which can vary across professors, depending on the type of relationship the students in lab $\ell(e,c)$ have with professor f.

In what follows, for the purpose of simplicity, we restrict the scope of indirect influence to that between professors and students within the *same* research field (or subfields). We, particularly, assume that $\pi_{ef} = \pi$ if the research field or subfield of professor f is the same as or closely related to that of the direct advisor, e, and $\pi_{ef} = 0$ otherwise.

Analogous to Equation (6), we compute the conditional expectation of the squared double-differenced average student research outcome growth and construct a regression model based on the comparison of the conditional expectations between labs with and without a "treatment" assignment. To achieve this aim, let us use $V^{\ell(e,c,c')}$ to denote an assignment indicator of an "indirect" turnover incident. Specifically, define $V^{\ell(e,c,c')} = 1$ if a professor whose research subfield is the same as that of professor e is replaced due to turnover in cohort c and $V^{\ell(e,c,c')} = 0$ otherwise. In other words, the binary indicator variable $V^{\ell(e,c,c')}$ represents an instance of turnover in which a professor from the same research field has an indirect impact on the average student research outcome growth from lab $\ell(e,c')$ to lab $\ell(e,c)$.

We obtain the following result under the same assumptions as above on the distributions of advisor quality and the idiosyncratic error terms:

$$\begin{aligned}&\mathrm{E}\left[\left(DD\overline{\Delta outcome}^{\ell(e,c,c')}\right)^2 | W^{\ell(e,c,c')}=0, V^{\ell(e,c,c')}\right]\\&=\alpha\left(\frac{1}{I^{\ell(e,c)}}+\frac{1}{I^{l(e,c')}}\right)+\pi^2\left\{2\sigma_d^2(1-\rho_d)\right\}V^{\ell(e,c,c')},\end{aligned} \tag{11}$$

where α is the same as that given in Equation (6).

This, in turn, leads to the following regression model using the subsamples that consist of labs in which advisor turnover did *not* occur ($W^{\ell(e,c,c')}=0$):

$$\left(DD\overline{\Delta outcome}_m\right)^2 = \alpha_{ind}X_m + \beta_{ind}V_m + \varepsilon_m, \tag{12}$$

where $m = 1, \cdots, M$ is the index of observations[41].

Comparing Equations (11) and (12) leads to the parameter relationship that $\beta_{ind} = \pi^2\{2\sigma_d^2(1-\rho_d)\}$. Let us use $\hat{\beta}_{dir}$ to denote an estimate of the coefficient of W_n from the baseline regression model given by Equation (7), and let $\hat{\beta}_{ind}$ be an estimate of the coefficient of V_m from Equation (12) presented above. We therefore obtain $\hat{\pi} = \sqrt{\hat{\beta}_{ind}/\hat{\beta}_{dir}}$, which can be used as a measure of indirect knowledge transfer from a non-advisor professor in the same research field.

An empirical challenge is to identify a group of professors whose research subjects were close enough to that of the professor experiencing turnover. As the type of data necessary to judge the similarity between research subjects is absent or rarely present, we adopt a simple and heuristic method for identifying the same research subject groups, which exploits the information revealed by the actual turnover events. It is conceivable that, when an instance of professor turnover occurred, the students in the lab of the professor who exited were highly likely to be re-assigned to a professor whose research area was closely related to that of the original professor. In the empirical analysis that follows, we therefore assume that the original professor who exited and the re-assigned professor were working in the same research area.

The situation is illustrated in Figure 8, which parallels that illustrated in Figure 1. In the former figure, there are three cohorts, c_0, c_1 and c_2. As we have assumed previously, an instance of turnover involving professor a occurred in cohort c_2, and the students in lab $\ell(a, c_2)$ switched their research advisor from professor a to professor b in the doctoral program. Then, professor b, whose research area is the same as that of professor a, took over the students in lab $\ell(a, c_2)$, whereas he had supervised two labs, $\ell(b, c_0)$ and $\ell(b, c_1)$, before the incident occurred, and oversaw another lab, $\ell(b, c_2)$, at the time of the incident. We assume that professor a's turnover affects the doctoral research productivity of the students in lab $\ell(b, c_2)$ because

[41] Here, the unit of observation is each element of (e, c, c') such that $W^{\ell(e,c,c')} = 0$ for any advisor $e \in \mathscr{A}$ and cohorts c, c' such that $0 < c - c' \leq \tau$.

the indirect influence from the professor, θ_{ad}, ceases to exist after turnover.[42] In this case, to identify the magnitude of the indirect impact, we essentially compare the gap in student research outcome growth between labs $\ell(b, c_2)$ and $\ell(b, c_1)$ (treatment group with $V^{\ell(b,c_2,c_1)} = 1$) with the same gap between labs $\ell(b, c_1)$ and $\ell(b, c_0)$ (control group with $V^{\ell(b,c_1,c_0)} = 0$).

Table 10 presents the regression estimates. We adopt the default setting for the student research outcome and use the same estimation method as before.[43] As shown, the estimates are ambiguous for β_{ind}. One of the estimates is negative, and in the case in which the estimates are positive, they are not statistically significant at the 10 percent level in any specification except one. The estimates of the squared indirect influence parameter, π^2, are reported in row (3).[44] Again, the signs of the estimates differ. The maximum estimate is 0.269, while the value where $\hat{\beta}_{ind}$ is statistically significant is as low as approximately 0.1, as reported in column (6). This result implies that the indirect knowledge transfer effect from non-advisor faculty is $\hat{\pi} = 0.33$, suggesting that it is, at most, less than one-third of the direct effect from the advisor.

On balance, therefore, there appears to be little or no indirect influence from non-advisor faculty members across labs on doctoral student research productivity growth.

Insert Table 10

7 Conclusion

In this paper, we investigated the extent to which professors can affect the development of the research performance of the graduate students whom they supervise. By using detailed data on professors and students at UTokyo's department of physics, we estimated a lower bound of the professor value added to student research achievement growth while in school. The estimation results consistently show that postgraduate research education based on an advisor-advisee relationship is quite effective — professors have a substantial impact on the students' achievement gains in terms of the number of publications in top journals in physics. This corroborates the view of earlier studies (e.g., Azoulay et al., 2010; Moser et al., 2014;

[42]In the estimation that follows, we choose the lab of professor b that was influenced "indirectly" by professor a's turnover if the turnover occurred while the students in the lab were in the doctoral program (i.e., from the first doctoral year to the final year of the doctoral program). We require this because the indirect influence from professor a at the doctoral degree level, not the master's degree level, needs to be changed.

[43]The research outcomes in the master's degree and doctoral degree programs are aggregated over the period from M1 to D2 and the period from D1 to P4, respectively. Furthermore, the set of "top journals" here consists of twelve journals.

[44]We compute $\hat{\pi}^2$ as the quotient of the estimate $\hat{\beta}_{ind}$ over the estimate $\hat{\beta}_{dir}$ that appear in the corresponding number of columns in Table 3 and Table 10, respectively.

Borjas and Doran, 2014) that research interactions among scientists in vertically aligned relationships, including senior-and-junior-collaborator, teacher-student, and adviser-advisee relationships, matter for the creation and diffusion of scientific ideas and knowledge.

Our findings also suggest that the accumulation of prominent scientists in a comparatively small number of universities is explained, at least partially, by the results of successful education at the postgraduate level. For example, in Japan, five out of ten Nobel Prize winners in physics completed their doctoral degrees at UTokyo, and four earned their doctorate degrees at Nagoya University. Given our results on the effectiveness of professors in enhancing students' research capability growth, we can speculate that the relatively high concentration of physics Nobel laureates in these two universities in Japan might be caused not only by the processes of students' self-selection or schools' selective recruitment but also by the beneficial reproduction of elite physicists, which was enabled by a deliberate process of teaching and learning in a lab. While previous studies (e.g., Waldinger, 2010) suggest that high-quality universities can facilitate human capital accumulation among graduate students, our paper specifically adds that this outcome is based on advisor-advisee-based education.

We need to highlight some limitations of this paper. First, our analysis of the professor's value added is essentially short run. Although the estimation results reveal that research advisors can influence the research development of their students, the impact might be limited to the short span of time while the student is in graduate school or several years after the completion of graduate school. It is left to future research to examine whether a professor's supervision during a graduate program has a long-term impact on student research performance during their postgraduation careers.

Second, the analysis in the paper is limited to a small, albeit prominent, group of physicists. Thus, our conclusion regarding a professor's value added might not be generalizable to groups of other scientists from different disciplines or other graduate schools. We hope that the findings of this paper regarding the efficacy of professors in promoting student progress in research performance will be helpful to stimulate further research in related areas including the economics of higher education and the economics of science and technology.

Tables

Table 1: Descriptive Statistics for the Student Research Outcomes in Levels and in Differences

	Research Outcome at the Master's Level	Research Outcome at the Doctoral Level	Research Outcome Gain at the Doctoral Level
	$outcome^c_{iam}$	$outcome^c_{iad}$	$\Delta outcome^c_{iad}$
Mean	0.0677	0.2202	0.1481
S.D.	0.2184	0.5075	0.4068
Min	0.0000	0.0000	-0.4738
Max	2.3175	4.7303	4.7303
Sample Size	1019	1019	1019

Note:
1) The research outcome at each degree level is computed based on the research proficiency scores. The aggregation years are M1-D2 for the master's level and D1-P4 for the doctoral level, respectively.
2) The research outcome gain at the doctoral level is given by the difference of the research outcome from the doctoral level to the master's level. Since the research outcome at the bachelor's level is normalized as zero, the research outcome gain at the master's level is equal to the research oucome at the master's level.

Table 2: Descriptive Statistics of Advisors: Comparison between Advisors When Turnover Occurred and When It Did Not

Variable	Description	With Turnover	Without Turnover	t-stat		Absolute Standardized Difference
Age	Professor Age	53.72 (6.37)	47.06 (5.77)	-9.00	***	1.10
Num_Stud	Number of Students	1.16 (0.41)	1.27 (0.52)	1.63		0.22
*Outcome*5	Profeessor's Research Outcome (5 years average)	0.18 (0.29)	0.21 (0.40)	0.50		0.07
Rank_Assoc	Associate Professor Dummy	0.21 (0.41)	0.44 (0.50)	3.73	***	0.51
Rank_Prof	Full Professor Dummy	0.79 (0.41)	0.53 (0.50)	-4.27	***	0.59
Dept_Phys	Department of Physics Dummy	0.72 (0.45)	0.75 (0.43)	0.62		0.08
Inst_Solid	Institute of Solid State Dummy	0.72 (0.41)	0.22 (0.41)	0.27		0.03
Inst_Other	Other Institutes Dummy	0.34 (0.48)	0.29 (0.45)	-0.85		0.11
Period_70s	70's Dummy	0.15 (0.36)	0.21 (0.41)	1.19		0.16
Period_80s	80's Dummy	0.16 (0.37)	0.26 (0.44)	1.71	*	0.23
Period_90s	90's Dummy	0.57 (0.50)	0.35 (0.48)	-3.72	***	0.46
Period_00s	00's Dummy	0.12	0.19	1.46		0.20

Note:
1) $***\ p < .01$, $**\ p < .05$, $*\ p < .10$
2) The standardized difference is given by the size of the difference in means of a conditioning variable, scaled by the square root of the variances in the original samples (Ronsenbaum and Rubin 1985)

Table 3: Baseline Estimation Results: The Effect of Advisor Turnover on Student Research Outcome Growth at the Doctoral Level

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
(1) α	0.0667 ***	0.0742 ***	0.0960 ***	0.0570	0.1267 **	0.4025 ***
	(0.0237)	(0.0225)	(0.0134)	(0.0682)	(0.0545)	(0.0720)
(2) β	0.3371 *	0.2663 **	0.1956 **	2.3091 *	2.1322 **	1.6401 **
	(0.1746)	(0.1204)	(0.0985)	(1.3249)	(0.9220)	(0.8251)
(3) Lower bound of σ_d^2	0.0843 **	0.0666 **	0.0489 **	0.5773 **	0.5331 **	0.4100 **
	[0.0268]	[0.0135]	[0.0236]	[0.0407]	[0.0104]	[0.0234]
Sample Size	104	186	271	104	186	271

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Table 4: Falsification Test Results

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
(1) α	0.1827 ***	0.1902 ***	0.1736 ***	0.2819 ***	0.6911 ***	0.8996 ***
	(0.0418)	(0.0409)	(0.0443)	(0.0458)	(0.1828)	(0.2287)
(2) $\tilde{\beta}$	0.2979	0.2039	0.0814	0.6315 **	-0.0969	-0.7784
	(0.2123)	(0.1725)	(0.1270)	(0.2616)	(0.3965)	(0.4350)
Sample Size	422	763	603	422	763	603

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Table 5: The Estimates of the Lower Bound of Advisor Quality Variance for Various Aggregation Periods

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
M1-D2/D1-P4†‡	0.0843 **	0.0666 **	0.0489 **	0.5773 **	0.5331 **	0.4100 **
	[0.0268]	[0.0135]	[0.0236]	[0.0407]	[0.0104]	[0.0234]
M1-D2/D1-P3†	0.0689 **	0.0498 **	0.0257 *	0.4578 **	0.4099 ***	0.2793 **
	[0.0232]	[0.0181]	[0.0973]	[0.0310]	[0.0080]	[0.0338]
M1-D1/D1-P4†	0.1460 **	0.1243 **	0.0910 **	0.9274 **	0.8522 **	0.7525 **
	[0.0488]	[0.0210]	[0.0389]	[0.0450]	[0.0133]	[0.0145]
M1-D1/D1-P3†	0.1215 **	0.0966 **	0.0572 *	0.7711 **	0.6911 **	0.5745 **
	[0.0485]	[0.0271]	[0.0913]	[0.0384]	[0.0119]	[0.0175]

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The p-values are in square brackets
†)(master's level aggregation period) / (doctoral level aggregation period)
‡) The baseline case.

Table 6: Estimation Results: When Only 9 Top Journals Are Included in Student Research Outcomes

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
(1) α	0.0151 ***	0.0154 ***	0.0648 ***	0.0491 *	0.0533 ***	0.3727 ***
	(0.0061)	(0.0040)	(0.0096)	(0.0348)	(0.0185)	(0.0628)
(2) β	0.1797 ***	0.1451 ***	0.0843 **	0.7449 ***	0.6364 ***	0.0963
	(0.0551)	(0.0320)	(0.0349)	(0.2798)	(0.1851)	(0.1942)
(3) Lower bound of σ_d^2	0.0449 ***	0.0363 ***	0.0211 ***	0.1862 ***	0.1591 ***	0.0241
	[0.0005]	[0.0000]	[0.0079]	[0.0039]	[0.0003]	[0.3100]
Sample Size	104	186	271	104	186	271

Note:

1) $*** \ p < .01, \ ** \ p < .05, \ * \ p < .10$

2) The standard errors are in parentheses, and p-values are in square brackets

3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Table 7: Estimation Results: When a Change in Advisor Quality Variance Is Allowed during the Period Near Turnover

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
(1) α	0.0717 ***	0.0604 ***	0.0982 ***	0.0622	0.1395 **	0.4297 ***
	(0.0234)	(0.0156)	(0.0144)	(0.0674)	(0.0614)	(0.0752)
(2) β	0.3391 *	0.2431 *	0.2145 **	2.3320	2.1209 **	1.7108 *
	(0.1839)	(0.1294)	(0.1065)	(1.4204)	(0.9704)	(0.8976)
(3) δ_1	—	-0.1207	-0.0085	—	-0.2789	-0.5772
		(0.032)	(0.022)		(0.123)	(0.181)
(4) δ_2	-0.1512	0.4291	-0.2504	-0.4563	-0.1015	-1.4842
	(0.2259)	(0.296)	(0.152)	(1.5777)	(0.938)	(1.093)
(5) δ_3	-0.1195	—	—	-0.1036	—	—
	(0.0431)			(0.0843)		
(6) Lower bound of σ_d^2	0.0848 **	0.0608 **	0.0536 **	0.5830 *	0.5302 **	0.4277 **
	[0.0326]	[0.0301]	[0.0220]	[0.0503]	[0.0144]	[0.0283]
Sample Size	104	186	271	104	186	271

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)
4) The coefficient of the dummy variable $D_k^{(a,c,c')}$ is given by δ_k for $k = 1, 2$ and 3. The dummy variable $D_k^{(a,c,c')}$ is omitted from the regression if there is no cohort with the dummy variable being one in the matched sample.

Table 8: Estimation Results: The Student Proficiency Score is Set to Zero If the Student Coauthored with the Advisor

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$ (1)	$\tau = 4$ (2)	$\tau = 5$ (3)	$\tau = 3$ (4)	$\tau = 4$ (5)	$\tau = 5$ (6)
(1) α	0.0544 ***	0.0670 ***	0.0439 ***	0.0215	0.0713 *	0.0528 **
	(0.0230)	(0.0225)	(0.0091)	(0.0448)	(0.0419)	(0.0250)
(2) β	0.2057	0.1665	0.2374 **	1.3776	1.3374 **	1.3690 **
	(0.1727)	(0.1187)	(0.1022)	(0.9758)	(0.6807)	(0.5814)
(3) Lower bound of σ_d^2	0.0514	0.0416 *	0.0593 **	0.3444 *	0.3343 **	0.3422 ***
	[0.1168]	[0.0803]	[0.0101]	[0.0790]	[0.0247]	[0.0093]
Sample Size	104	186	271	104	186	271

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Table 9: Estimation Results: the Double-Difference Measure in Levels Is Used as the Dependent Variable

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]$			First-authored-paper Based $[DD\overline{\Delta outcome}]$		
Adjacent Period	$\tau = 3$ (1)	$\tau = 4$ (2)	$\tau = 5$ (3)	$\tau = 3$ (4)	$\tau = 4$ (5)	$\tau = 5$ (6)
(1) α	-0.0462	-0.0186	0.0362 ***	-0.0655	-0.0693	0.0963
	(0.0237)	(0.0225)	(0.0134)	(0.0961)	(0.0856)	(0.0893)
(2) β	0.1439	0.1480	0.0310	0.2655	0.3968	0.2276
	(0.1746)	(0.1204)	(0.0985)	(2.1883)	(1.5374)	(1.3783)
Sample Size	104	186	271	104	186	271

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Table 10: Estimation Results: Effect of Non-Advisor Turnover on Student Research Outcome Growth at the Doctoral Level

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
(1) α_{ind}	0.0434 ***	0.0356 ***	0.0335 ***	0.0988 ***	0.0863 ***	0.0755 ***
	(0.0175)	(0.0056)	(0.0060)	(0.0358)	(0.0094)	(0.0136)
(2) β_{ind}	-0.0230	0.0274	0.0527	0.0764	0.1007	0.1768 *
	(0.0374)	(0.0320)	(0.0336)	(0.1192)	(0.0924)	(0.0924)
(3) $\pi^2 = \beta_{ind}/\beta_{dir}$	-0.0682	0.1030	0.2694	0.0331	0.0472	0.1078
Sample Size	145	282	288	145	282	288

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The standard errors are in parentheses, and p-values are in square brackets
3) The standard errors are computed by the subsampling method of Politis and Romano (1994)

Figures

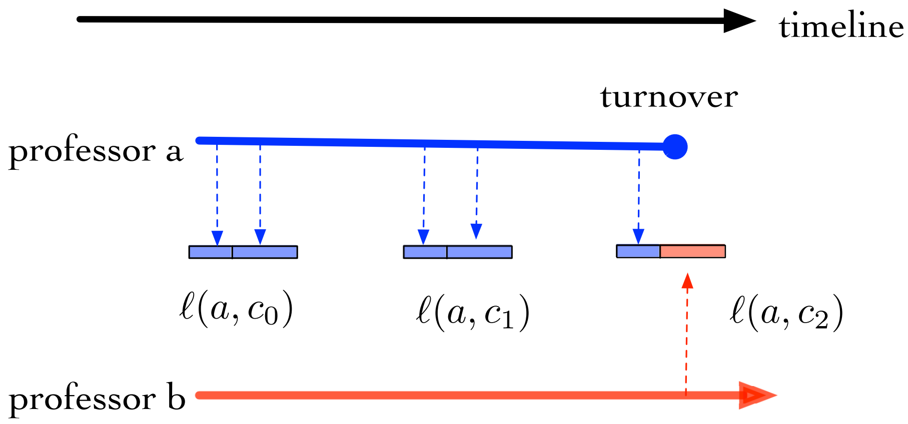

Figure 1: Example of Labs with and without Turnover

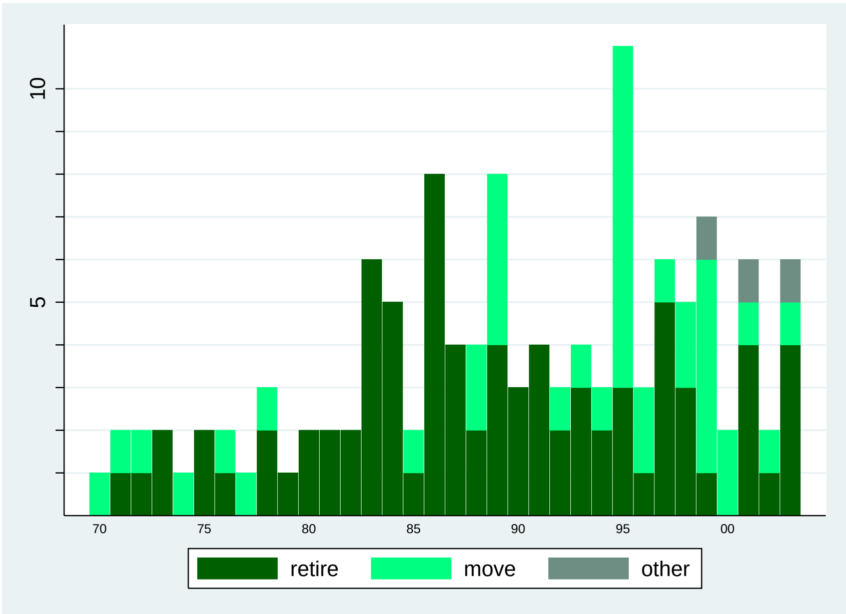

Figure 2: Number of Turnover Incidents in Each Year (1970-2004)

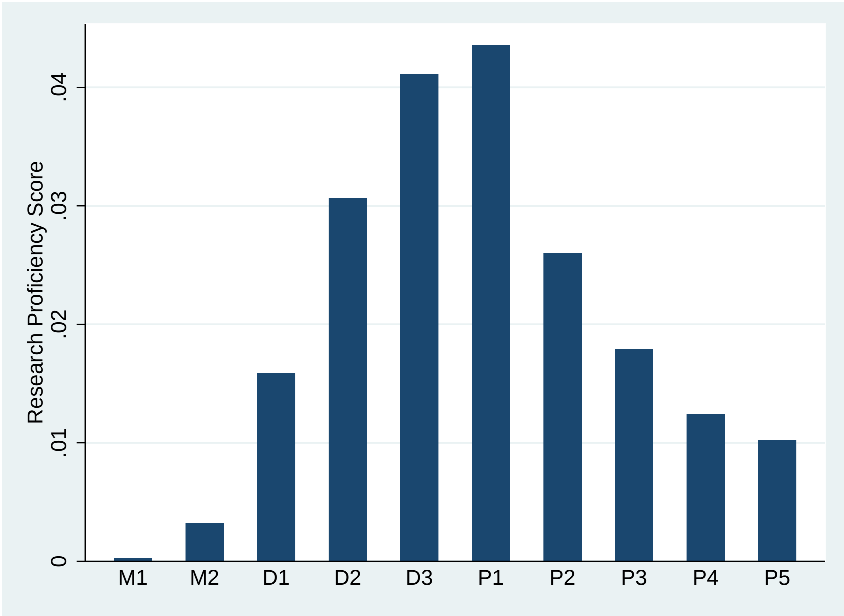

Figure 3: Average Student Research Proficiency Scores

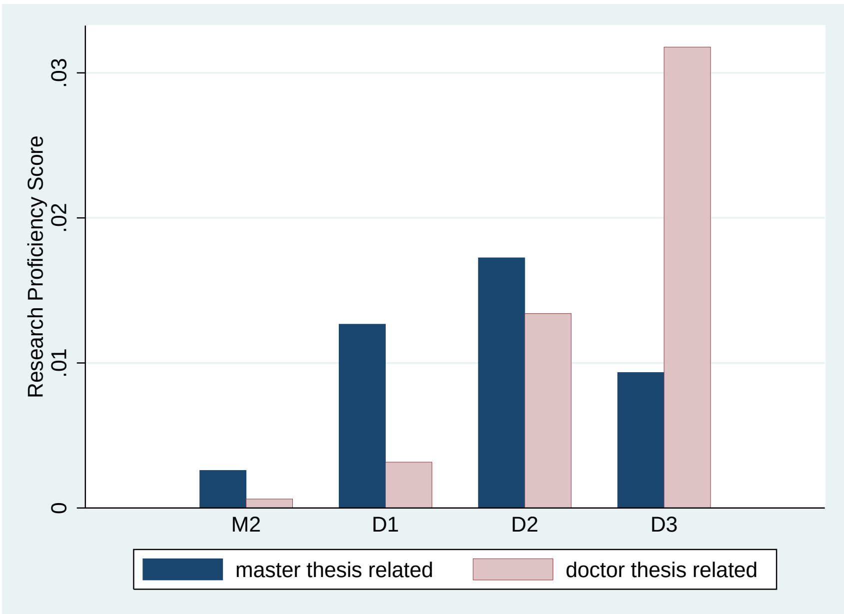

Figure 4: Decomposition of the Student Research Proficiency Scores: Those Related to Master's and Doctoral Theses

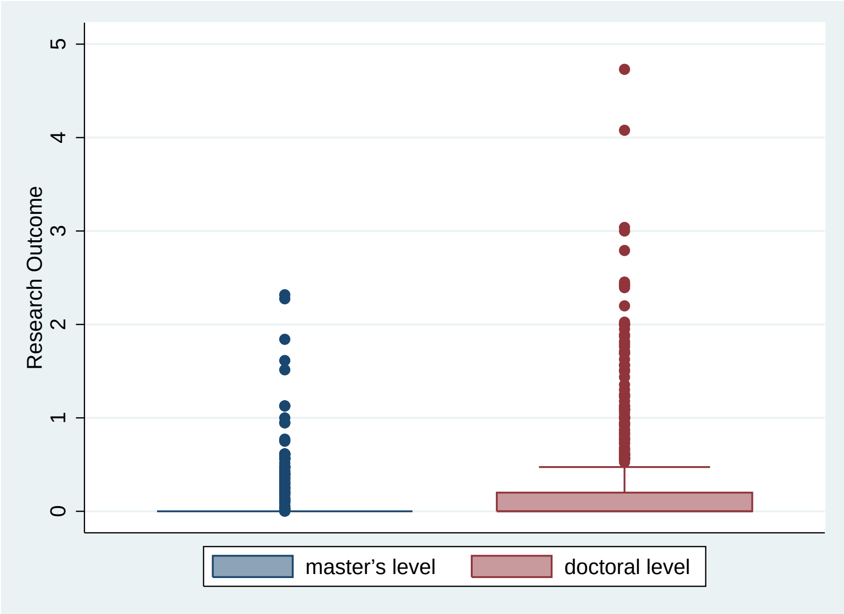

Figure 5: Box Plots of the Student Research Outcome Distributions in the Master's and Doctoral Programs

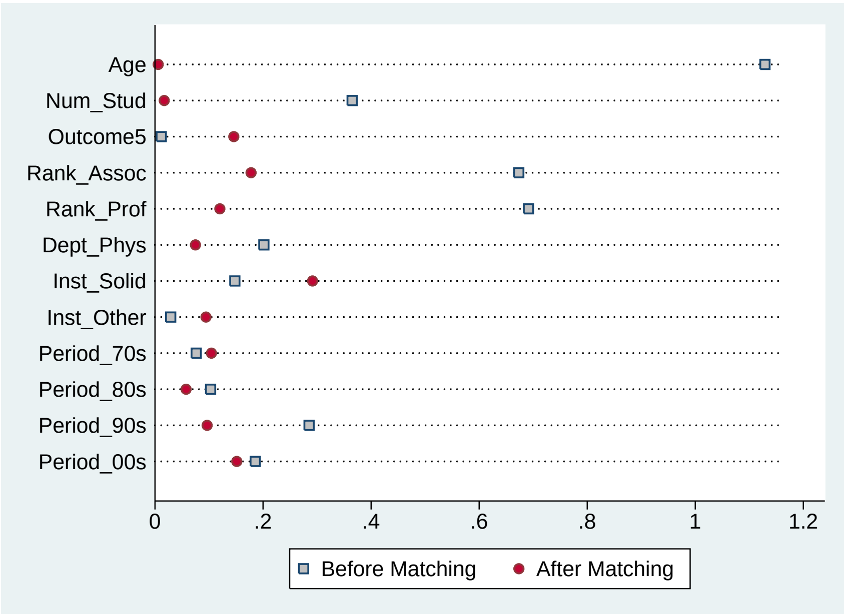

Figure 6: Comparison of the Absolute Values of the Standardized Differences between Treatment and Control Groups

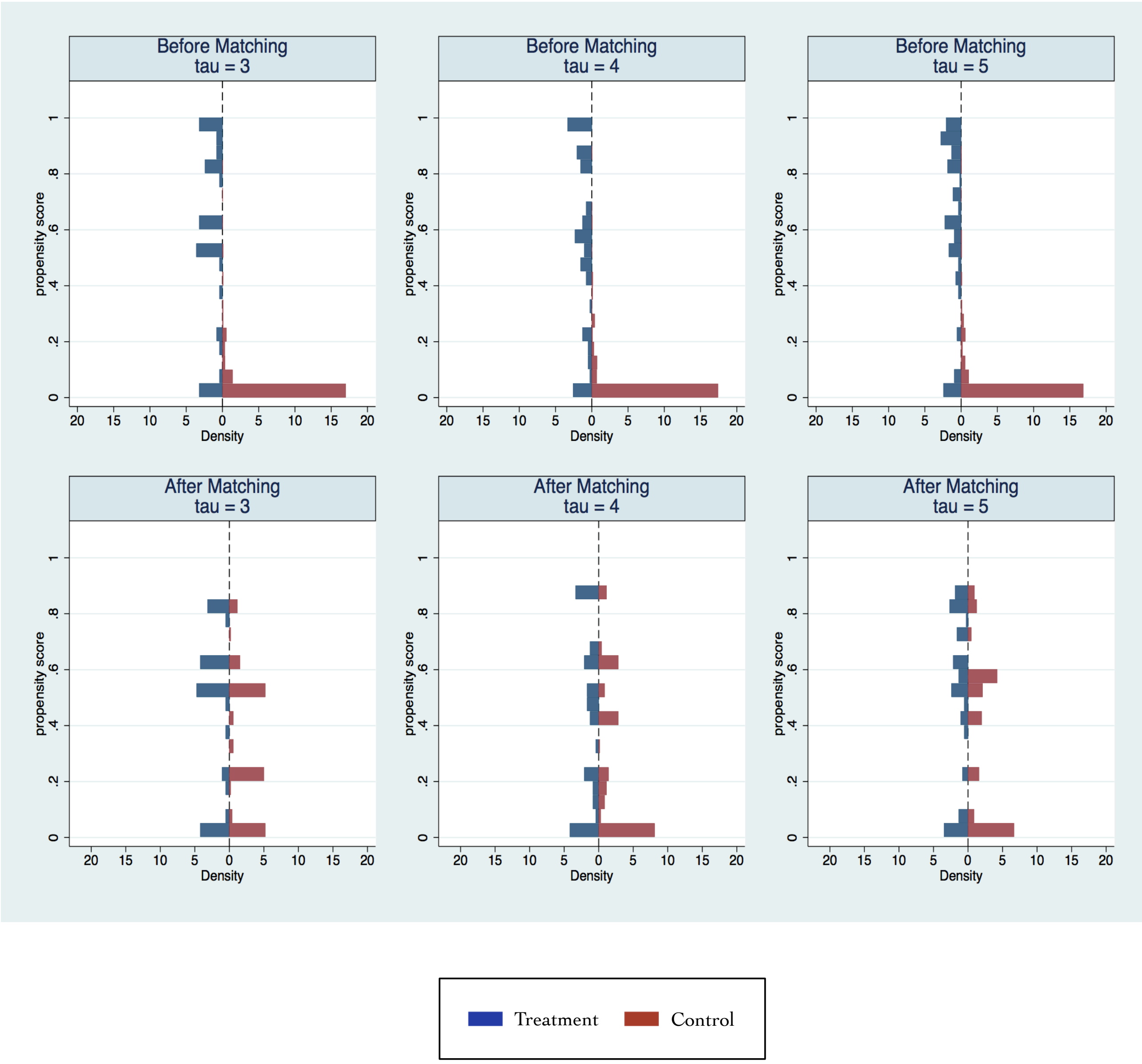

Figure 7: Distribution of the Propensity Score for the Treatment and Control Groups: Before and After Matching

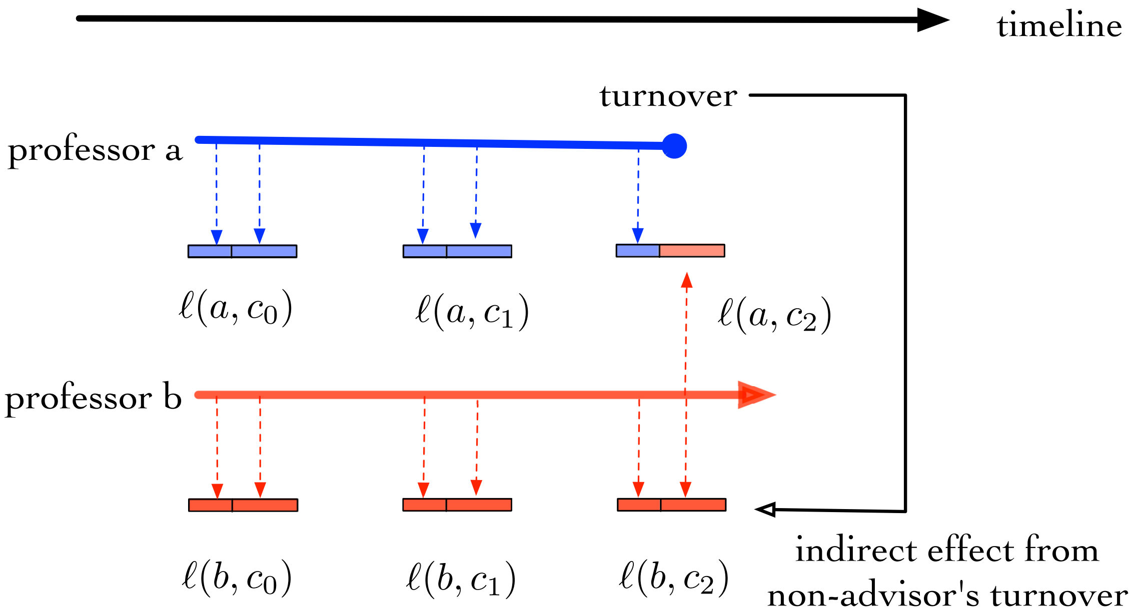

Figure 8: Example of Labs with and without Turnover for Advisors and Non-Advisors

Appendix

A Derivation of Equation (6)

We compute the conditional expectation of the squared left-hand side of Equation (5). Under the assumption that the random shock, ν_{iag}^{c}, is orthogonal to advisor quality, θ_g, for any student $i \in \mathscr{I}^{\ell(a,c)}$, professor $a \in \mathcal{A}$, cohort $c \in \mathcal{C}$ and program $g \in \{m, d\}$, the conditional expectation is given as follows:

$$
\begin{aligned}
&\mathrm{E}\left[\left(DD\overline{\Delta Outcome}^{\ell(a,c,c')}\right)^2 \middle| W^{\ell(a,c,c')}\right] \\
&= \mathrm{E}\left[(\theta_{bd} - \theta_{ad})^2 \middle| W^{\ell(a,c,c')} = 1\right] \cdot W^{\ell(a,c,c')} \\
&+ \mathrm{E}\left[\left\{\left(\bar{\nu}_{bd}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}\right) - \left(\bar{\nu}_{ad}^{\ell(a,c')} - \bar{\nu}_{am}^{\ell(a,c')}\right)\right\}^2 \middle| W^{\ell(a,c,c')} = 1\right] \cdot W^{\ell(a,c,c')} \\
&+ \mathrm{E}\left[\left\{\left(\bar{\nu}_{ad}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}\right) - \left(\bar{\nu}_{ad}^{\ell(a,c')} - \bar{\nu}_{am}^{\ell(a,c')}\right)\right\}^2 \middle| W^{\ell(a,c,c')} = 0\right] \cdot (1 - W^{\ell(a,c,c')}). \quad \text{(A.1)}
\end{aligned}
$$

Under assumption 1.1-1.2, we can compute the first part of Equation (A.1) as follows:

$$
\mathrm{E}\left[(\theta_{bd} - \theta_{ad})^2 \middle| W^{\ell(a,c,c')} = 1\right] = 2\sigma_d^2(1 - \rho_d). \tag{A.2}
$$

We turn to the second part of Equation(A.1), which is related to the conditional expectation of when turnover occurred, $W^{\ell(a,c,c')} = 1$. We have the following equality concerning the value within the expectation operator:

$$
\begin{aligned}
&\left\{\left(\bar{\nu}_{bd}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}\right) - \left(\bar{\nu}_{ad}^{\ell(a,c')} - \bar{\nu}_{am}^{\ell(a,c')}\right)\right\}^2 \\
&= \left\{\left(\bar{\tilde{\nu}}_{bd}^{\ell(a,c)} - \bar{\tilde{\nu}}_{am}^{\ell(a,c)}\right) - \left(\bar{\tilde{\nu}}_{ad}^{\ell(a,c')} - \bar{\tilde{\nu}}_{am}^{\ell(a,c')}\right)\right\}^2 \\
&= \left\{\left(\bar{\tilde{\nu}}_{bd}^{\ell(a,c)}\right)^2 + \left(\bar{\tilde{\nu}}_{am}^{\ell(a,c)}\right)^2 - 2\left(\bar{\tilde{\nu}}_{bd}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{am}^{\ell(a,c)}\right)\right\} + \left\{\left(\bar{\tilde{\nu}}_{ad}^{\ell(a,c')}\right)^2 + \left(\bar{\tilde{\nu}}_{am}^{\ell(a,c')}\right)^2 - 2\left(\bar{\tilde{\nu}}_{ad}^{\ell(a,c')}\right)\left(\bar{\tilde{\nu}}_{am}^{\ell(a,c')}\right)\right\} \\
&-2\left\{\left(\bar{\tilde{\nu}}_{bd}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{ad}^{\ell(a,c')}\right) - \left(\bar{\tilde{\nu}}_{bd}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{am}^{\ell(a,c')}\right) - \left(\bar{\tilde{\nu}}_{am}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{ad}^{\ell(a,c')}\right) + \left(\bar{\tilde{\nu}}_{am}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{am}^{\ell(a,c')}\right)\right\}, \quad \text{(A.3)}
\end{aligned}
$$

where $\bar{\tilde{\nu}}_{ag}^{\ell(a,c)}$ is the lab $\ell(a,c)$ average of $\tilde{\nu}_{iag}^{c}$, and hence, we use $\bar{\nu}_{ag}^{\ell(a,c)} = \bar{\tilde{\nu}}_{ag}^{\ell(a,c)} + \bar{\nu}_g$ in the computation above. We take the conditional expectation of each piece of the last term of Equation (A.3) under assumptions 2.1 - 2.5.

1. Consider the squared term of the average demeaned error, $\bar{\tilde{\nu}}_{pg}^{\ell(a,t)}$, where professor $p \in$

$\{a,b\}$, program $g \in \{d,m\}$ and cohort $t \in \{c,c'\}$. We have the following equation:

$$\begin{aligned}
\left(\bar{\tilde{\nu}}_{pg}^{\ell(a,t)}\right)^2 &= \left[\frac{1}{N^{\ell(a,t)}}\sum_{i\in I^{\ell(a,t)}}\left(\tilde{\nu}_{pig}^{t}\right)\right]^2 \\
&= \left(\frac{1}{N^{\ell(a,t)}}\right)^2\left\{\sum_{i\in I^{\ell(a,t)}}\left(\tilde{\nu}_{pig}^{t}\right)^2 + 2\sum_{j\in I^{\ell(a,t)}}\sum_{k\neq j\in I^{\ell(a,t)}}\left(\tilde{\nu}_{jpg}^{t}\right)\left(\tilde{\nu}_{jpg}^{t}\right)\right\}.
\end{aligned}$$

Assumptions 2.1 and 2.3 lead to the following conditional expectation:

$$\mathrm{E}\left[\left(\bar{\tilde{\nu}}_{pg}^{\ell(a,t)}\right)^2 \middle| W^{\ell(p,c)}=1\right] = \frac{\phi_g^2+2\psi_g}{N^{\ell(a,t)}}. \tag{A.4}$$

2. Consider the cross-term of the average demeaned errors, $\bar{\tilde{\nu}}_{pd}^{\ell(a,t)}$ and $\bar{\tilde{\nu}}_{am}^{\ell(a,t)}$, between master's and doctoral programs *within* lab $\ell(a,t)$ for cohort $t \in \{c,c'\}$, where professor $p=b$ if the professor switched from a to b due to turnover and $p=a$ if not. Then, we have:

$$\begin{aligned}
\left(\bar{\tilde{\nu}}_{pd}^{\ell(a,t)}\right)\left(\bar{\tilde{\nu}}_{am}^{\ell(a,t)}\right) &= \left[\frac{1}{N^{\ell(a,t)}}\sum_{i\in I^{\ell(a,t)}}\left(\tilde{\nu}_{ipd}^{t}\right)\right]\cdot\left[\frac{1}{N^{\ell(a,t)}}\sum_{i\in I^{\ell(a,t)}}\left(\tilde{\nu}_{iam}^{t}\right)\right] \\
&= \left(\frac{1}{N^{\ell(a,t)}}\right)^2\left\{\sum_{i\in I^{\ell(a,t)}}\left(\tilde{\nu}_{ipd}^{t}\right)\left(\tilde{\nu}_{iam}^{t}\right) + \sum_{j\in I^{\ell(a,t)}}\sum_{k\neq j\in I^{\ell(a,t)}}\left(\tilde{\nu}_{jpg}^{t}\right)\left(\tilde{\nu}_{kam}^{t}\right)\right\}.
\end{aligned}$$

Given Assumption 2.2 , the conditional expectation is given by:

$$\mathrm{E}\left[\left(\bar{\tilde{\nu}}_{pd}^{\ell(a,t)}\right)\left(\bar{\tilde{\nu}}_{pm}^{\ell(a,t)}\right) \middle| W^{\ell(p,c)}=1\right] = \frac{\phi_{md}}{N^{\ell(a,t)}}. \tag{A.5}$$

3. Consider the cross-term of the average demeaned errors between $\bar{\tilde{\nu}}_{pg}^{\ell(a,c)}$ and $\bar{\tilde{\nu}}_{p'g'}^{\ell(a,c')}$ *across* cohorts c and c', where professors $p \in \{a,b\}$ and $p' \in \{a,b\}$ and grad programs $g \in \{d,m\}$ and $g' \in \{d,m\}$. It is equal to:

$$\begin{aligned}
\left(\bar{\tilde{\nu}}_{pg}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{p'g'}^{\ell(a,c')}\right) &= \left[\frac{1}{N^{\ell(a,c)}}\sum_{i\in I^{\ell(a,c)}}\left(\tilde{\nu}_{ipg}^{c}\right)\right]\cdot\left[\frac{1}{N^{\ell(a,c')}}\sum_{j\in I^{\ell(a,c')}}\left(\tilde{\nu}_{jp'g'}^{c'}\right)\right] \\
&= \left(\frac{1}{N^{\ell(a,c)}}\right)\left(\frac{1}{N^{\ell(a,c')}}\right)\left\{\sum_{i\in I^{\ell(a,c)}}\sum_{j\neq i\in I^{\ell(a,c')}}\left(\tilde{\nu}_{ipg}^{c}\right)\left(\tilde{\nu}_{jp'g'}^{c'}\right)\right\}.
\end{aligned}$$

The conditional expectation is zero under Assumption 2.4-2.5. That is:

$$\mathrm{E}\left[\left(\bar{\tilde{\nu}}_{pg}^{\ell(a,c)}\right)\left(\bar{\tilde{\nu}}_{p'g'}^{\ell(a,c')}\right) \middle| W^{\ell(a,c,c')}=1\right] = 0. \tag{A.6}$$

Using results (A.4), (A.5), and (A.6) presented above, the conditional expectation of Equation (A.3), regardless of whether an advisor switch occurred, is equal to:

$$\begin{aligned}
&\mathrm{E}\left[\left\{\left(\bar{\nu}_{bd}^{\ell(a,c)} - \bar{\nu}_{am}^{\ell(a,c)}\right) - \left(\bar{\nu}_{ad}^{\ell(a,c')} - \bar{\nu}_{am}^{\ell(a,c')}\right)\right\}^2 \middle| W^{\ell(a,c,c')} = 1\right] \\
&= \left(\frac{1}{N^{\ell(a,c)}} + \frac{1}{N^{\ell(a,c')}}\right)\left\{\phi_d^2 + \phi_m^2 + 4\big(\psi_d + \psi_m\big) - 2\phi_{md}\right\}. \qquad \text{(A.7)}
\end{aligned}$$

Similar computation reveals that the third part of Equation (A.1), for the case in which turnover did not occur, $W^{\ell(a,c,c')} = 0$, is the right-hand side of Equation (A.7).

Based on Equations (A.2) and (A.7), we therefore have the following result:

$$\mathrm{E}\left[\left(DD\overline{\Delta Outcome}^{\ell(a,c,c')}\right)^2 \middle| W^{\ell(a,c,c')}\right] = \alpha\left(\frac{1}{N^{\ell(a,c)}} + \frac{1}{N^{\ell(a,c')}}\right) + \beta \cdot W^{\ell(a,c,c')},$$

where $\alpha = \phi_d^2 + \phi_m^2 + 4\big(\psi_d + \psi_m\big) - 2\phi_{md}$ and $\beta = 2\sigma_d^2(1-\rho_d)$.

B Identification of Students' Publications

B.1 A Score of Word Overlap in Titles

As described in Section 4, our measure of a graduate student's research achievement is based on the number of articles that he or she published in selected physics journals.

To identify the articles that were authored by each student in the sample, we compile physics papers from the Thomson Reuters WoS archive that satisfy the following three conditions: (1) the author names match the name of the student; (2) the publication dates are in the period from the year in which the student was enrolled in graduate school to four years after he or she received a doctoral degree; and (3) the words in the title overlap to some extent with those in the title of the student's master or doctorate thesis.

The first and second conditions can be easily verified because the authors' names and publication dates of articles are available from the WoS database, whereas the student names and the degree date of each student are found in the the master's and doctoral thesis catalogs of UTokyo's physics department.

To enforce the third condition, we define a score that assesses the degree of overlap in the words in titles. Let $\tilde{R}_i$ be the set of all physics articles that are associated with student $i \in \mathscr{I}$ after the first and second conditions presented above are satisfied. Note that, although all articles in the set $\tilde{R}_i$ include authors whose names are the same as student i, the student may or may not actually be the author of these articles. Such misidentification arises because of false positives in author name matching.

We use $t(r_{ij})$ to denote the *title* of article $r_{ij} \in \tilde{R}_i$ and use t_i to denote the title of student i's thesis (either master's or doctoral, depending on the context). Each title of an article or a thesis consists of *words*. For each article $r_{ij} \in \tilde{R}_i$, we compute the following score of word overlap in titles:

$$m_{ij} = \frac{\sum_{w \in \{t_i \cap t(r_{ij})\}} \phi(w)}{\max\left\{ \sum_{w \in t_i} \phi(w),\ \sum_{w \in t(r_{ij})} \phi(w) \right\}}, \tag{A.8}$$

where $\phi(w)$ is a weighting of word w that measures the rareness of the word.

Indeed, the frequency of words used in article titles varies substantially, some being common and others rare. Clearly, such information is potentially useful in deciding whether an article sharing the author name with a thesis is actually authored by the person who wrote the thesis. If the words included in both the titles of an article and thesis are relatively rare, there is a higher likelihood that the authors are the same, whereas the converse is true if the words are relatively common.

To utilize the intuition, $\phi(w)$ assigns high weight to relatively rare words and low weight to relatively common words. Following a similar approach to that proposed by Tang and Walsh (2010), we determine the weight, $\phi(w)$, based on the relative frequency of word w, which is computed by dividing its count frequency by the total counts of all technical terms that appear in all titles of the master's and doctoral theses of UTokyo's students. Specifically, we sort all words used in titles into five categories or quintiles based on their relative frequencies. For word w_k that is in the k-th quintile, the weight is given by $\phi(w_k) = (6 - k)^{-2/3}$ for $k = 1, 2, 3, 4, 5$.

One remaining issue concerns words referring to the same concept in physics that are rendered differently. For instance, words such as "energy", "energies", "energetics", and "energetic" are considered to represent the same notion. We address this issue by "standardizing" the words. Specifically, we undertake the following actions. First, we transform all non-letter, non-Greek characters and symbols into spaces. Second, we convert all words into lower case. Third, we reduce inflected (or derived) words to their word stem using a stemming algorithm.[45] For instance, the stemming algorithm reduces the words "energy", "energies", "energetics", and "energetic" to the unique root word, "energi". Fourth, we eliminate all of the non-informative "stopwords", that is, very high-frequency words such as *the*, *to*, *of*, and *study*. For example, consider an article with the title "*ENERGY-LEVEL STATISTICS OF METALLIC FINE PARTICLES*." In this case, the title is decomposed into the set of standardized root words as "energi", "level", "statist", "metal", "fine" and "particl".

We use the title word overlap score, given by Equation (A.8), when we identify that article $r_{ij} \in \tilde{R}_i$ is authored by student i, depending on whether the score, m_{ij}, exceeds the

[45] Specifically, we use Porter's stemming algorithm, which is the most commonly used algorithm for word stemming in English.

predetermined threshold, $\bar{m}$. Let $\hat{R}_i$ be the set of articles associated with student i by the word-overlapping-score method presented above such that $\hat{R}_i \subseteq \tilde{R}_i$.

B.2 An Optimal Threshold

How can we determine the threshold, $\bar{m}$, for the title word overlap score when matching articles and theses? Two types of matching errors are possible. We refer the first as a type 1 error, which occurs if we under-match articles, i.e., if we miss articles that are indeed authored by a student by regarding them as being written by another author. However, the second error, referred to as a type 2 error, arises when we include articles that are not authored by a target student. A type 1 error is likely to occur when we impose a threshold value, $\bar{m}$, that is too high, whereas a type 2 error will be more likely when we impose a low threshold, $\bar{m}$, and end up with spurious matches that actually belong to different authors.

One fundamental problem regarding the problem of identifying students' publications is that the true set, R_i, is unknown for student $i \in \mathscr{I}$, and therefore, the degrees of type 1 and type 2 errors cannot be assessed.

However, we might be able to obtain a reasonably accurate approximation set of published articles for certain students, especially for those who became academic researchers and published their CVs on the web. Let $\bar{\mathscr{I}} \subseteq \mathscr{I}$ be the set of such students/researchers. We acquired the CVs of 40 such researchers by a random web search and parsed the research publication information to create the benchmark set of articles. Our expectation is that the benchmark article set, $\bar{R}_i$, will contain reliable and comprehensive information on the true set, R_i, at least for student/researcher $i \in \bar{\mathscr{I}}$. Nevertheless, the set $\bar{R}_i$ might include some articles that are not directly related to their thesis projects. In this regard, the benchmark set should be close to but somewhat larger than the true set.

We use the benchmark article set to evaluate the performance of the matching procedure based on the word overlap score in titles. Specifically, to gauge the performance at each threshold value, we use two goodness-of-fit indices, $GOFI2a$ and $GOFI2b$, proposed by Trajtenberg et al. (2006). Let $\bar{R}_i$ be the benchmark set of student $i \in \bar{\mathscr{I}}$ and $\hat{R}_i(m)$ the corresponding set estimated by the matching procedure based on the word overlap score in titles, with m being the threshold value.

Those measures are defined as:

$$\begin{aligned} GOFI2a(m) &\equiv \text{Average}\left[\tfrac{|\bar{R}_i \cap \hat{R}_i(m)|}{|\bar{R}_i|}\right] \\ GOFI2b(m) &\equiv \text{Average}\left[\tfrac{|\bar{R}_i \cap \hat{R}_i(m)|}{|\hat{R}_i(m)|}\right], \end{aligned}$$

where the average is taken over all persons in the selected set $\bar{\mathscr{I}}$. In essence, if our matching procedure tends to under-match or over-match, $GOFI2a(m)$ or $GOFI2b(m)$ decrease, respectively. Therefore, we should seek to increase these indices to avoid type 1 and type 2 errors to the greatest extent possible, but a trade-off exists between the two goals.

Figure 9 presents those two indices for various values of m in increments of 0.05. $GOFI2b(m)$, which is presented as a solid blue line, increases in the range of a smaller threshold value, m, and reaches nearly 0.65 when $m = 0.25$ with no improvement being observed if $m > 0.25$. This leads to the implication that type 2 error will no longer be reduced dramatically if we set $m > 0.25$. Turning to $GOFI2a(m)$, which is presented as a dashed red line, it decreases consistently as the threshold value, m, rises, implying that type 1 error will be alleviated as the value of m decreases.

Accordingly, we consider the optimal threshold to be $\bar{m} = 0.25$, as this is the value that balances the two goodness-of-fit measures — $GOFI2a(m)$ is maximized (thus, type 1 error is minimized) on the *condition* that $GOFI2b(m)$ remains at a high level (thus, an increase in type 2 error is reduced as much as possible).

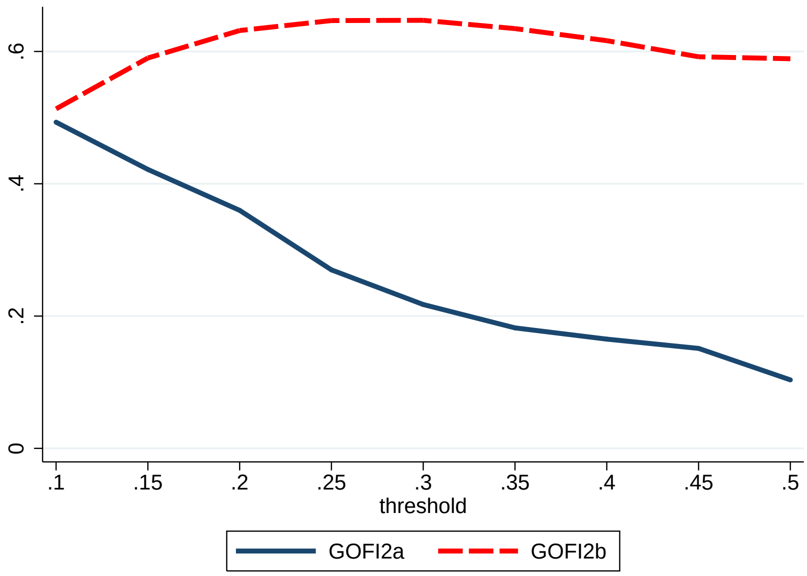

Figure 9: Comparison of Two Goodness-of-Fit Indices over Various Thresholds for the Word Overlap Score in Titles

C.3 Supplementary Materials for Section 5.2

Table C.1 : Estimation Results of Propensity Score

Adjacent Period	$\tau = 3$		$\tau = 4$		$\tau = 5$	
Age	-5.823	***	-6.494	***	-6.784	***
	[9.70]		[9.39]		[6.55]	
Num_Stud	21.900		-0.701	*	-0.212	
	[0.02]		[1.86]		[0.46]	
Rank_Assoc	0.832		1.109		-0.552	
	[0.76]		[1.00]		[0.26]	
Inst_Other	-0.208		-0.462		-0.308	
	[0.46]		[0.99]		[0.46]	
Period_90s	1.014	**	1.777	***	1.101	
	[1.99]		[4.11]		[1.49]	
Period_00s	-1.724	***	-1.375	**	-1.782	*
	[2.67]		[2.44]		[1.92]	
Age^2	0.063	***	0.072	***	0.073	***
	[10.11]		[9.58]		[6.84]	
Num_Stud^2	-7.263					
	[0.02]					
Age × *Period_80s*	1.465	**			1.157	
	[2.47]				[1.56]	
Num_Stude × *Period_80s*	-80.110	**			-63.27	
	[2.43]				[1.53]	
*Outcome*5 × *Inst_Other*	-1.816	*	-2.325	**	-3.059	**
	[1.76]		[2.06]		[1.99]	
Rank_Assoc × *Inst_Solid*			2.346	*	2.941	*
			[1.90]		[1.83]	
Inst_Solid × *Period_00s*	3.487	***	3.168	**		
	[3.02]		[2.29]			
Inst_Other × *Period_80s*	-3.700	***			-2.757	*
	[2.95]				[1.83]	
Constant	111.7		138.900	***	148.800	***
	[0.16]		[8.98]		[6.05]	
Sample Size	1446		1202		925	

Note:
1) The dependent variable is advisor switch indicator W
2) $***\ p < .01$, $**\ p < .05$, $*\ p < .10$
3) The p-values are in square brackets

C.4 Supplementary Materials for Section 6.2: Robustness Checks

Table C.2 : The Estimates of the Lower Bound of Advisor Quality Variance for Various Aggregation Periods with Over-Matching and Under-Matching Criteria for the Degree of Technical Term Overlap in Titles

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$ (1)	$\tau = 4$ (2)	$\tau = 5$ (3)	$\tau = 3$ (4)	$\tau = 4$ (5)	$\tau = 5$ (6)
Threshold 0.20 (overmatch)						
M1-D2/D1-P4†‡	0.0953 **	0.0790 **	0.0576 **	0.7246 **	0.6833 **	0.5492 **
	[0.0407]	[0.0184]	[0.0341]	[0.0460]	[0.0115]	[0.0213]
M1-D2/D1-P3†	0.0621 **	0.0425 **	0.0205	0.4672 **	0.4179 ***	0.2641 **
	[0.0323]	[0.0341]	[0.1505]	[0.0280]	[0.0071]	[0.0438]
M1-D1/D1-P4†	0.1659 *	0.1466 **	0.1214 **	1.1114 **	1.0467 **	2.2262 ***
	[0.0580]	[0.0231]	[0.0247]	[0.0490]	[0.0134]	[0.0085]
M1-D1/D1-P3†	0.1159 *	0.0913 **	0.0639 *	0.7805 **	0.7055 **	0.6066 **
	[0.0565]	[0.0349]	[0.0665]	[0.0364]	[0.0105]	[0.0127]
Threshold 0.30 (undermatch)						
M1-D2/D1-P4†‡	0.0509 *	0.0378 **	0.0191	0.3490 **	0.3135 ***	0.1795 *
	[0.0712]	[0.0476]	[0.1037]	[0.0222]	[0.0042]	[0.0548]
M1-D2/D1-P3†	0.0401 *	0.0282 *	0.0023	0.2662 **	0.2282 ***	0.0884
	[0.0719]	[0.0609]	[0.4395]	[0.0110]	[0.0019]	[0.1246]
M1-D1/D1-P4†	0.1123 *	0.0988 **	0.0683 *	0.6166 **	0.5541 ***	0.4385 **
	[0.0599]	[0.0237]	[0.0541]	[0.0345]	[0.0094]	[0.0220]
M1-D1/D1-P3†	0.0923 *	0.0783 **	0.0360	0.4971 **	1.0168 ***	0.3002 **
	[0.0582]	[0.0261]	[0.1507]	[0.0265]	[0.0094]	[0.0331]

Note:

1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$

2) The p-values are in square brackets

†)(master's level aggregation period) / (doctoral level aggregation period)

‡) The baseline cases.

Table C.3 : The Estimates of the Lower Bound of Advisor Quality Variance for Various Aggregation Periods: When Only 9 Top Journals Are Included in Student Research Outcomes

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
M1-D2/D1-P4†‡	0.0449 ***	0.0363 ***	0.0211 ***	0.0127 ***	0.1862 ***	0.1591
	[0.0005]	[0.0000]	[0.0079]	[0.3911]	[0.0039]	[0.0003]
M1-D2/D1-P3†	0.0377 ***	0.0262 ***	0.0070	0.0065 ***	0.1403 ***	0.1117 **
	[0.0027]	[0.0000]	[0.2448]	[0.4584]	[0.0010]	[0.0000]
M1-D1/D1-P4†	0.0829 ***	0.0675 ***	0.0355 **	0.1520 **	0.4216	0.3441 **
	[0.0080]	[0.0017]	[0.0431]	[0.0739]	[0.7579]	[0.0038]
M1-D1/D1-P3†	0.0680 ***	0.0495 ***	0.0115	0.0923 ***	0.3389 ***	0.2588 *
	[0.0085]	[0.0026]	[0.2533]	[0.1124]	[0.0084]	[0.0025]

Note:

1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$

2) The p-values are in square brackets

†)(master's level aggregation period) / (doctoral level aggregation period)

‡) The baseline case.

Table C.4 : The Estimates of the Lower Bound of Advisor Quality Variance for Various Aggregation Periods: When Change in Advisor Quality Variance Is Allowed in the Period Near Turnover

Dependent	Credit Share Weighted			First-authored-paper Based		
	$[DD\overline{\Delta outcome}]^2$			$[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
M1-D2/D1-P4†‡	0.0449 ***	0.0363 ***	0.0211 ***	0.0127 ***	0.1862 ***	0.1591
	[0.0005]	[0.0000]	[0.0079]	[0.3911]	[0.0039]	[0.0003]
M1-D2/D1-P3†	0.0377 ***	0.0262 ***	0.0070	0.0065 ***	0.1403 ***	0.1117 **
	[0.0027]	[0.0000]	[0.2448]	[0.4584]	[0.0010]	[0.0000]
M1-D1/D1-P4†	0.0829 ***	0.0675 ***	0.0355 **	0.1520 **	0.4216	0.3441 **
	[0.0080]	[0.0017]	[0.0431]	[0.0739]	[0.7579]	[0.0038]
M1-D1/D1-P3†	0.0680 ***	0.0495 ***	0.0115	0.0923 ***	0.3389 ***	0.2588 *
	[0.0085]	[0.0026]	[0.2533]	[0.1124]	[0.0084]	[0.0025]

Note:
1) $***\ p < .01,\ **\ p < .05,\ *\ p < .10$
2) The p-values are in square brackets
†)(master's level aggregation period) / (doctoral level aggregation period)
‡) The baseline case.

Table C.5 : The Estimates of the Lower Bound of Advisor Quality Variance for Various Aggregation Periods: The Student Proficiency Score is Set to Zero if the Student Is Coauthor with the Advisor

Dependent	Credit Share Weighted $[DD\overline{\Delta outcome}]^2$			First-authored-paper Based $[DD\overline{\Delta outcome}]^2$		
Adjacent Period	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 3$	$\tau = 4$	$\tau = 5$
	(1)	(2)	(3)	(4)	(5)	(6)
M1-D2/D1-P4†‡	0.0514	0.0416 *	0.0593 **	0.3444 *	0.3343 **	0.3422 ***
	[0.1168]	[0.0803]	[0.0101]	[0.0790]	[0.0247]	[0.0093]
M1-D2/D1-P3†	0.0359	0.0246	0.0496 **	0.2433 *	0.2301 **	0.2351 ***
	[0.1393]	[0.1413]	[0.0072]	[0.0645]	[0.0264]	[0.0097]
M1-D1/D1-P4†	0.0589	0.0422 *	0.0558 **	0.3712 *	0.3389 **	0.3547 ***
	[0.0878]	[0.0823]	[0.0155]	[0.0645]	[0.0236]	[0.0075]
M1-D1/D1-P3†	0.0434 *	0.0252	0.0461 **	0.2700 *	0.2347 **	0.2475 ***
	[0.0985]	[0.1445]	[0.0126]	[0.0580]	[0.0251]	[0.0073]

Note:

1) $***\ p < .01$, $**\ p < .05$, $*\ p < .10$

2) The p-values are in square brackets

†)(master's level aggregation period) / (doctoral level aggregation period)

‡) The baseline case.

References

Abadie, Alberto and Guido W Imbens, "Notes and Comments on the Failure of the Bootstrap," *Econometrica*, 2008, *76* (6), 1537–1557.

Abe, Yasumi and Satoshi P Watanabe, "Some Thoughts on Implementing US Physics Doctoral Education in Japanese Universities," *Asia Pacific Education Review*, 2012, *13* (3), 403–415.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, "Superstar Extinction," *Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Becher, Tony, "Graduate Education in Britain: The View from the Ground," in "The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan," University of California Press, 1993, pp. 115–156.

Borjas, George J. and Kirk B. Doran, "The Collapse of the Soviet Union and the Productivity of American Mathematicians," *Quarterly Journal of Economics*, 2012, *127* (3), 1143–1203.

_ and _ , "Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors," *Review of Economics and Statistics*, 2014. *forthcoming.*

Carrell, Scott E. and James E. West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 2010, *118* (3), 409–432.

Crump, Richard K., Joseph V. Hotz, Guido W. Imbens, and Oscar A. Mitnik, "Dealing with Limited Overlap in Estimation of Average Treatment Effects," *Biometrika*, 2009, *96* (1), 187–199.

Gellert, Claudius, "The Conditions of Research Training in Contemporary," in "The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan," University of California Press, 1993, pp. 45–68.

Grossman, Gene M. and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 1991, *58* (1), 43–61.

Gumport, Patricia J., "Graduate Education and Organized Research in the United States," in "The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan," University of California Press, 1993, pp. 225–260.

Hanushek, Eric A. and Steven G. Rivkin, "Teacher Quality," in "Handbook of the Economics of Education," Vol. 2, edited by Eric A. Hanushek and Finis Welch, Elsevier, 2006, chapter 18, pp. 1051–1078.

Hanushek, Eric A and Steven G Rivkin, "Generalizations about Using Value-added Measures of Teacher Quality," *The American Economic Review*, 2010, *100* (2), 267–271.

Hanushek, Eric A. and Steven G. Rivkin, "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics*, 2012, *4*, 131–157.

Hilmer, Michael J. and Christiana E. Hilmer, "Fishes, Ponds, and Productivity: Student-Advisor Matching and Early Career Publishing Success for Economics Phds," *Economic Inquiry*, 2009, *47* (2), 290–303.

Hoffmann, Florian and Philip Oreopoulos, "Professor Qualities and Student Achievement," *Review of Economics and Statistics*, 2009, *91* (1), 83–92.

Imbens, Guido W., "Matching methods in practice: Three examples," Technical Report, National Bureau of Economic Research 2014. NBER Working Paper #19959.

Imbens, Guido W and Donald B Rubin, *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, 2015.

Kawashima, Tatsuo and Fumihiro Maruyama, "The education of advanced students in Japan: engineering, physics, economics and history," in "The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan," University of California Press, 1993, pp. 326–353.

Liu, Xuan Zhen and Hui Fang, "Fairly Sharing the Credit of Multi-authored Papers and its Application in the Modification of h-index and g-index," *Scientometrics*, 2012, *91* (1), 37–49.

Low, Morris, *Science and the Building of a New Japan*, Palgrave Macmillan, 2005.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22*, 3–42.

Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German-Jewish Émigrés and US Invention," *American Economic Review*, 2014, *104* (10), 3222–3255.

Ogawa, Yoshikazu, "Challenging the Traditional Organization of Japanese Universities," *Higher Education*, 2002, *43* (1), 85–108.

Page, Marianne E and Gary Solon, “Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban,” *Journal of Labor Economics*, 2003, *21* (4), 831–855.

Polanyi, Michael, *Personal Knowledge: Towards a Post-Critical Philosophy*, University of Chicago Press, 1958.

— , *The Tacit Dimension*, Routledge, 1966.

Politis, Dimitris N and Joseph P Romano, “Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions,” *The Annals of Statistics*, 1994, *22* (4), 2031–2050.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, “Teachers, Schools, and Academic Achievement,” *Econometrica*, 2005, *73* (2), 417–458.

Romer, Paul M., “Endogenous Technological Change,” *Journal of Political Economy*, 1990, *98* (5), S71–S102.

Rosenbaum, Paul R and Donald B Rubin, “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *The American Statistician*, 1985, *39* (1), 33–38.

Tanabashi, Masaharu, “Shoichi Sakata: His Life and Physics, Collections of Materials in Sakata Memorial Archival Library,” *Progress of Theoretical Physics Supplement*, 2012, *197*, 1–14.

Tang, Li and John P. Walsh, “Bibliometric Fingerprints: Name Disambiguation Based on Approximate Structure Equivalence of Cognitive Maps,” *Scientometrics*, 2010, *84* (3), 763–784.

Trajtenberg, Manuel, Gil Shiff, and Ran Melamed, “The ”Names Game”: Harnessing inventors’ Patent Data for Economic Research,” Technical Report, National Bureau of Economic Research 2006. NBER Working Paper #12479.

Traweek, Sharon, *Beamtimes and Lifetimes: The World of High Energy Physicists*, Harvard University Press, 1988.

Ushiogi, Morikazu, “Graduate education and research organization in Japan,” in “The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan,” University of California Press, 1993, pp. 299–325.

Waldinger, Fabian, "Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany," *Journal of Political Economy*, 2010, *118* (4), 787–831.

_ , "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, 2012, *79* (2), 838–861.

Waltman, Ludo, "An Empirical Analysis of the Use of Alphabetical Authorship in Scientific Publishing," *Journal of Informetrics*, 2012, *6* (4), 700–711.

Weitzman, Martin L, "Recombinant Growth," *Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

Wisker, Gina and Gillian Robinson, "Doctoral "orphans": Nurturing and Supporting the Success of Postgraduates Who Have Lost Their Supervisors," *Higher Education Research & Development*, 2013, *32* (2), 300–313.

Zuckerman, Harriet, *Scientific elite: Nobel laureates in the United States*, Transaction Publishers, 1977.