Causal Inferences on Young Economics Professors' Salaries*

Yong Bian[†1] and Wei Kong[‡2]

[1]Department of Economics, University of Missouri, Columbia, MO 65201, USA
[2]School of Business, Shanghai University of International Business and Economics, Shanghai 201620, China

Jun. 29, 2020

Abstract

In this paper, we analyze the causal effects on young economics professors' salaries of three interesting factors: gender, PhD graduation school rank, and undergraduate major. The dataset used is novel, contains detailed research productivity measures and other demographic information of young economics professors from top 50 public research universities in the U.S. We apply double/debiased machine learning (DML) to the models and obtain consistent estimators under high-dimensional control variable space. By tracking the first 10 years of their professional working experience we find that there barely exist causal effects on salaries from the above three factors in most of the years when controlling for research productivity and background, neither economically nor statistically. Exceptionally, the gender effect in experience year 7 is both statistically and economically significant. For both PhD graduation school rank and undergraduate major, although lack of statistical significance in almost all the experience years, the estimates for experience years 7–9 are large in magnitude. Then, based on what have been found, we discuss possible economic mechanisms and reasons.

1 Introduction and Motivation

The past 40 years' literature contains many studies of salary in the academic field. However, not many focus specifically on economics, and most lack either depth or breadth. There are studies concentrating on some particular factors and their relationships with salary,

*Thanks to my advisor Dr. David M. Kaplan's initial suggestions on studying young economics professor's salary

†Email address: ybzkf@mail.missouri.edu. Address: University Avenue, Columbia MO 65201, USA.

‡Email address: kongwei@suibe.edu.cn. Address: 1900 Wenxiang Rd, Songjiang District, Shanghai 201620, China.

such as gender, race, seniority, and/or administrative service (Bratsberg, Ragan, and Warren, 2003; Hamermesh, Johnson, and Weisbrod, 1982; Johnson and Stafford, 1974; Li and Koedel, 2017; Megdal and Ransom, 1985; Moore, Newman, and Turnbull, 1998). For example, Moore et al. (1998) point out that with research productivity held constant, the negative relationship between seniority and earnings disappears. Johnson and Stafford (1974) give an overall description of salary differences between female and male faculty. Others focus on mechanisms. For example, Kahn and Lange (2014) consider employers learning employees' true productivity over time as well as true productivity itself changing over time.

A good measure of research productivity is important. As Katz (1973) found in interviews, "It became clear that research ability, publication record, and national reputation were the most important factors influencing salary and promotion decisions." However, many researchers treat each person's productivity as constant over time (e.g., Claypool, Janssen, Kim, and Mitchell, 2017; Ginther and Hayes, 2003) or only use a few rough or self-reported productivity measures (e.g., Sax, Hagedorn, Arredondo, and Dicrisi, 2002; Yang and Webber, 2015). In such cases, there is reason to doubt the causality of the coefficient estimates of the treatment variables.

Another concern is model specification. Specifically, most studies use linear regression (e.g., Altonji and Pierret, 2001; Ginther and Hayes, 2003; Yang and Webber, 2015). There can be misspecification bias even if conditional independence holds.

Moreover, results in the literature are limited by a small number of universities and/or use of survey data. Many include data from only a single university; the broadest (non-survey) dataset we found in the literature spanned nine unspecified state universities (Moore et al., 1998). If there is significant heterogeneity across universities, then external validity is a concern. As usual, use of self-reported survey data may cause self-selection bias.

Unlike full professors who already have many publications and an established reputation, it is hard to precisely value a young economics professor's research ability, which is important in determining salary. In economics, it takes longer to get a paper published than in other scientific disciplines. Even after acceptance, it may take one year (or longer) to get officially published. Therefore, young economics professors may have fewer publications compared to professors in other disciplines at the same career stage.

Because publication quantity is so limited for most young economics professors, employers might look into more detailed research and publication information. For example, employers may examine numbers of citations, coauthors, and pages. Such measures may further interact with journal quality; e.g., number of pages published in "tier one" journals, pages in "tier two" journals, etc. This intuition is further supported by research; e.g., Liebowitz and Palmer (1984) write, "Where articles are published can affect one's promotion, tenure, and salary at one's present job." As Gibson, Anderson, and Tressler (2017) mention, different tiers of journal rank have different meanings in determine salary. We include journal tier-interacted publication information to better control for research productivity and to improve previous studies' estimations. Also, Gibson et al. (2017) argue that number of citations has a different impact on salary in different levels of universities. Our focus on public research universities could help reduce one sort of heterogeneity from the university level.

The above facts led us to construct a novel dataset that contains detailed productivity measures along with other background and demographic characteristics. We contribute to the literature by using this dataset to estimate causal effects on salary of gender, PhD graduation school rank, and undergraduate major (economics, STEM). We have collected data from most of the top 50 public research universities in the United States, from calendar year 2008 to 2014, focusing on younger economics faculty. We collected personal educational and demographic information as well as detailed measurements of research productivity: publications, citations, coauthors, pages, and journal rank.[1]

In this paper, we study how the three factors of interest causally affect young economics professors' salaries when controlling for research productivity and background variables. The analysis is based on professional working experience years (Altonji and Pierret, 2001; Johnson and Stafford, 1974; Perna, 2001; Toutkoushian, Bellas, and Moore, 2007) instead of calendar years. In each experience year sub-dataset, we have collected individuals from calendar year 2008 to 2014, capturing approximately 80 variables. In order to estimate causal effects under this high-dimensional control variable set, we apply the double/debiased machine learning (DML) method (Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, and Newey, 2017; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018). Without setting specific functional forms, we want to take the advantage of machine learning techniques nested in the DML method in dealing with high-dimensional data, letting the method choose which and how to use variables from the high-dimensional control set to get consistent causal effect estimators.

Then we explore possible mechanisms behind the possible causal effects. Altonji and Pierret (2001) and Altonji (2005) model how employers may statistically discriminate based on group types. Negotiation power (Claypool et al., 2017; Gerhart and Rynes, 1991) might be another source of causality as well as mobility differences (Blackaby, Booth, and Frank, 2005). There could be many other explanations as well.

2 Literature Review

We found many papers discuss salary disparity problem in the literature, but particularly, we are interested in those related to young economics faculty's salary payment. We collect those relate to our interests on how young economics professor's salary is determined and what's the mechanism behind that.

People might be discriminated on salary. Altonji and Pierret (2001) shows that employers perhaps start with statistical discrimination on people who newly enter the labor force based on easy to observe characteristics which are highly correlated with productivity, and the employers then gradually learn the workers' true productivity and adjust the salaries accordingly. In their paper, they use experience time variable t to interact with variables such as education, black and Armed Forces Qualification Test (AFQT) tracking how those factors affect salary payment over time. They state that coefficients on easy to observe variables decrease whereas those on hard to observe variables increase with years of working experience growing; employers have limited information on early year labor force people and there is statistical discrimination on the basis of education. This study relates to our research in the aspect of how we look at the problem, we also want to figure out how the determines of salary change and track the effects over years; but we will discuss causality and instead

[1]Full list of variables shown in Table 3.

of their just using one unchanged variable to represent productivity we rely on time varying productivity measurements under control.

In a later paper, Altonji (2005) expend the previous paper, discuss and specifies on different type of jobs. One of the advantages to study academic labor market is that the market is well defined; each person under study has a PhD degree, which will exclude many kinds of heterogeneity (Johnson and Stafford (1974)). Altonji (2005) mentioned employers learn more about workers in high skill jobs each time period. Research related academic jobs are belong to high skilled job types, so according to Altonji (2005), a research university faculty's true productivity will be gradually learned by employers, and their changing research productivity affects salary.

Besides employer learning process, we keep an eye on productivity change and how they affect salary. In a more recent paper Kahn and Lange (2014) shows how a worker's productivity change over time and emphasis both mechanisms are important: they use a dynamic structure which nests both employer learning and productivity heterogeneity in the model to capture the idea. They also point out "The pure Employer Learning model predicts that salary payment correlates more with past than with future performance measures because firms rely on past but not future performance measures to set current pay. In contrast, the pure dynamic productivity heterogeneity model implies that pay correlates similarly with past and future performance." This means employers will not only rely on what they observe of an individual's past productivity but also what they expect on his future performance to decide a salary. That's the reason we collect current and cumulative productivity measures. Employers actually can observe forth coming information of a paper being published in future time, therefore doing a robustness check is needed. By also including next years' productivity information into each experience year model we could verify our model specification.[2]

However, most studies do not take productivity change into consideration and lack of good measures on it. Our dataset can make up this defect. Because of lacking a good measure of productivity people have to treat each person's productivity as constant over life time like Ginther and Hayes (2003) and Claypool et al. (2017). An important drawback of the most commonly used approach in literature as noted by Kahn and Lange (2014) is to treat Armed Forces Qualifying Test (AFQT) as a proxy for true productivity. The score does not change over time whereas the true productivity may. Kahn and Lange (2014) also point out that traditional data sources make it difficult to distinguish the two because they lack independent measures of productivity. That's why our new dataset is valuable. Unlike Kahn and Lange (2014) focusing on a labor market with single employer or with the productivity measure being subjective and discrete taking only values 1, 2, 3, or 4, we include detailed measures in our dataset of young professor productivity such as page of paper, number of coauthor, number of citation; those information by journal rank tiers; current and cumulative infomation, etc. Our dataset contains multi-dimensional and time varying productivity information.

There are many other studies discussing how salary is influenced by particular factors such as seniority, citation and department rank. Those give us a guide on what information we should include in our dataset how we construct it. Moore et al. (1998) estimate the relation between seniority (years at the same university) and salary in research universities

[2]More details will be showed in Section 7.

after controlling for individuals' publication, education, and they've found that seniority is negatively related to salary. Bratsberg et al. (2003) present a study on a panel of professors from five universities that further assures the negative relation between seniority and faculty pay by controlling for research productivity. However, Barbezat and Donihue (1998) holds an opposite opinion on seniority with a detailed study on different subgroup of people. Given those information, we include seniority information as control in our dataset. Hamermesh et al. (1982) concentrate on citation and provide evidence of citation being an effective productivity measure. But the limitation is that they only use data from 7 large universities. Ehrenberg, Pieper, and Willis (1998) and Formby and Hoover (2002) prove that department ranking is correlated with salary payment, higher rank of their PhD program usually related to higher pay. As a contrast and extension, we collect data from top 50 research universities in the U.S. including detailed information on citation, seniority and department rank as well as other demographic information(Claypool et al. (2017)) such as visiting status, geometric location, age, school[3] those may correlate with salary payment.

We notice that many of the current literature are quite limited in either breadth or depth. Most of them are conducted by linear regression(Moore et al. (1998), Claypool et al. (2017),etc) and usually only several universities are involved. Many include data from only a single university; the broadest (non-survey) dataset we found in the literature spanned nine unspecified state universities (Moore et al. (1998)).

Survey data are widely used. Ehrenberg et al. (1998) use survey data from 1974/75 to 1980/81 periods studying whether lower tenure probability universities would pay higher to new assistant professors. They admit that due to the limitation of available data their results might encounter biases. Because of the use of survey data there could be self selection problems as well. Sax et al. (2002) use a national survey data of college and university faculties. The productivity measure they use is only consist of self reported number of past 2 years published article papers. The possibility that high productive person happens to have published less than usual or none papers during the 2 survey years makes their result doubtful. Ehrenberg (2002) apply institutional data to analyze the salary difference between public universities and their private counterparts. However, they use aggregated data which may not be able to reflect individual level effects. Li and Koedel (2017) use self collected data to analyze the wage difference in race-ethic and gender among public university faculties in 6 different disciplines. Similar to Claypool et al. (2017), they use one year cross sectional data with one-time wage decomposition.

Next, let's look at specifically how people study gender, graduate school rank, undergraduate major in the literature.

Gender. The disparity of salary between male and female faculties in academic field has been studied quite a lot. People may have an impression on women professors that they have less incentive to take part in research works , taking on more service and have more teaching loads (Smart (1991), Mitchell and Hesli (2013), Sax et al. (2002)). But we want to find whether the difference is causal or just simply statistical correlation. Many studies show gender may be correlated with productivity measures. Smart (1991) proofs gender could influence salary through different factors such as academic rank, working age and male domination discipline. Being female negatively related to publishing referred papers and

[3]More details will be showed in section 3.

total publication(Sax et al. (2002), Yang and Webber (2015)). There are unexplained gander salary gaps even with control for education, productivity and other background information (Barbezat (1987b), Perna (2001), Toutkoushian and Hoffman (2002), Toutkoushian and Conley (2005), Blackaby et al. (2005), Claypool et al. (2017)). Other than that, Johnson and Stafford (1974) focus on the gender difference in academic salaries and give a reason of combination of statistical discrimination and human capital differences over work life time. Sax et al. (2002) discuss the gender gap on productivity. Recently, Claypool et al. (2017) show that among political science faculty male instead of female has negotiation power that can help improve their salary. Li and Koedel (2017) point out academic field, experience and research productivity are three major factors that drive wage differences. However, those cannot fully explain the gender difference. Ginther and Hayes (2003) use SDR data to analyze gender difference in humanities professors salary and promotion. They showed that the gender gap in salaries are not significant, if there is discrimination on salaries it comes through promotion.

PhD Graduation School Rank. Evidences show graduation rank correlates with salary. The graduate school rank is a comprehensive index that can reflect quality of education a student receives. So people graduate from higher ranked universities may also obtain higher potential human capital, there is positive effect on salary from graduation school rank (Claypool et al. (2017)). Graves, Marchand, and Thompson (1982) proof there is association among department ranking, publication and salary. But they only show that full professor's salary is strongly related to publication and ranking. Stock and Alston (2000) indicate that there is a positive correlation between program rank and initial salaries, even after controlling for information on qualifications such as research and teaching. Those findings help us to construct one of our research goal: finding more precise causal estimates of graduation rank and we will expend the analysis more deeply by tracking 10 years of experience.

Undergraduate Major. Undergraduate major is another interesting factor we want to figure out its correlation with salary. It's always been the case that an economics professor holds a PhD degree in economics. When admission by their PhD receiving school, they were selected by observable economics talent, in which quantitative ability is an important item to be considered. And also, quantitative ability has been perceived to be very important in determine research productivity (Grove and Wu (2007)). Also, Zhang (2005) showed that social science, business and some other art undergraduate majors have negative coefficients on the probability of graduate enrollment whereas math, biological and other science majors have positive coefficients on the probability of graduate enrollment. We wonder if undergraduate major would also affect research productivity in a professor's prolonged after graduation research life.

Since the numbers of control variables is quite large in our new dataset, applying Double/Debiased Machine Learning (DML) method is needed in consistently estimating causal/treatment effects. Chernozhukov et al. (2018) and Chernozhukov et al. (2017) prove by constructing Neyman-Orthogonal scores, and sample splitting, the estimates are consistent to the true causal parameters under high dimensional control space.

3 Background and Data

In this study, we use a novel data which aggregates salary data from top public research universities in the United States, focusing on young economics faculty members, matched to personal educational and demographic characteristics and a detailed account of research productivity. The data includes young faculty members if in academic year 2008-2009, they were assistant professors or still graduate students, because our target is a full list of the young economics faculty members [4] Thus, anyone who was an associate or full professor in the initial year was excluded. Then, we track the full list of young faculty members forward to 2014-2015 academic year as long as they were still employed in the system of public universities.[5].

For each faculty member, the information of salary, education background, demographic characteristics, work experience and research productivity is collected. The available salary data for the full list of young faculty members each year is from the state's webpage. The time unit of analysis is one year. [6] And all salaries are in 2008 dollars adjusting by the CPI index.

The individual educational and demographic characteristics were accessed by the most recent CV of faculty members. We vastly describe these characteristics by the background of their undergraduate and graduate education. For the background of undergraduate education, the main measures are the school rank and the major. As the feature of economics profession, we consider the contribution of undergraduate education to the foundation of economic sense and math. Thus, it includes two variables of major, whether the major is economics and whether the major is STEM. For the background of graduate education, in addition to the school rank and the major, the total years of graduate education, including masters degrees, degrees in ECON/non-ECON fields and the total years of getting PhD degree.

In addition, we have the information of work experience, including the rank of school faculty members served as, the number of years faculty members started the position of assistant professor, the number of years they were promoted to associate/full professor, the number of years faculty members have worked at the current school, the year faculty members visited other schools.

The journal publication is the main measure for the research productivity in the economics profession. How to value a journal publication is a question of interest in the academics of economics. It is the important criteria for the economics faculties to be promoted. To evaluate the research productivity, we collect the journal publication information on their CV and the Google scholar database.[7] To measure the worth of a journal publication, we include journal rank, number of pages, number of co-authors and number of citations. To

[4]We only focus on the tenure-track faculty members. All the non-tenure-track, including "teaching faculty" or "instructors", emeritus, adjunct, and visiting faculty are excluded.

[5]For the job-hopper, if she transferred to a public university, we still track her performance until 2014-2015 academic year. Otherwise, if she entered to a private school or the industry, she was out of the sample.

[6]For some schools, the salary data was reported per academic year. For others, the salary data was reported per calendar year. No matter what cases are, the length of time for each observation was adjusted to one year.

[7]We only take journal articles into account, not "working papers" (or "discussion papers," etc.), not book chapters (unless it's NBER something).

calculate the publication value for each faculty member per year, the trade off between quality and quantity is another difficulty we encountered. For the quality of publication, we adopt the journal rankings from the IDEAS 2015[8] and divide the rankings into five tiers. Publishing in a top ranked journal is quite different from a lower ranked journal. Thus, we set the corresponding variables to value the publication for different tiers separately. To consider the effects of both quality and quantity of publication, it is valued by number of publication pages * $log((1 + r(max))/Journal\ rank) * log(2 + number\ of\ citation)/(1 + number\ of\ coauthors)$, where $r(max)$ is max of journal rank. Besides, we also consider the cumulative, average and the best values of publication are important to capture the research productivity of a faculty member. "Average productivity of publication" is calculated by $Cumulative\ publication\ value/(Year - PhD\ graduation\ year + 1)$. And the best value of publication measures the best performance of research productivity which is the best rank among the accumulative publications for each faculty member.

Table 1: Public Universities Faculty Members Worked in

University of Arizona	University of Illinois at Chicago
University of California-Berkeley	University of Illinois at Urbana-Champaign
University of California-Davis	Purdue University
University of California-Irvine	University of Maryland
University of California-Los Angeles	University of Michigan
University of California-Riverside	Michigan State University
University of California-Santa Barbara	University of Minnesota
University of California-Santa Cruz	University of Missouri
University of California-San Diego	Rutgers University-New Brunswick
University of Florida	Ohio State University
Florida State University	University of Texas-Austin
Georgia State University	University of Virginia
George Washington University	University of Washington
Iowa State University	University of Wisconsin-Madison

Table 1 lists the 28 public universities faculty members worked in from 17 different states. These universities are diverse from the aspect of school quality. The criteria we used to measure the school quality is the ranking from IDEAS during 2011-2014. Around 93% universities ranked at top 200.[9] The sample includes 363 faculty members from the department of economics.[10] All the faculty members are employed by the top public universities in US. The data include the information of all the young faculties from the year 2008 [11]or the year of staring to be an assistant professor [12]. And we track them forward to 2014 unless

[8]https://ideas.repec.org/top/top.journals.all.html

[9]Among these universities, 21.4% universities ranked at top 30, 25% universities ranked at top 30-50, 25% universities ranked at top 50-100, 21.4% universities ranked at top 100-200.

[10]Considering the gap among different fields, this study examines the faculty members from department of economics, excluding the business school.

[11]For the faculty members if in academic year 2008-2009, they were assistant professors

[12]For the faculty members if in academic year 2008-2009, they were still graduate students

they leave the system of public universities. We focus on professors with experience less and equal to 10 years and take the faculties with the same years of experience as one cohort. The statistics of each cohort are summarized in the Table 2.

Table 2: Summary of Statistics, by Experience

Experience	Number of Faculties	Male	Graduate School Rank	Undergraduate Econ
0	99	0.74	36.40	0.71
1	118	0.72	39.94	0.69
2	120	0.72	45.83	0.70
3	125	0.72	42.10	0.70
4	135	0.72	45.10	0.70
5	135	0.76	50.84	0.70
6	134	0.73	51.53	0.69
7	111	0.74	51.86	0.64
8	80	0.71	59.65	0.64
9	70	0.76	54.95	0.67
10	51	0.76	50.41	0.76

Table 3 shows information collected in our dataset. We collected the data by individual, and keep track of them within the year range of 2008 to 2014 as long as they are working in the population pool showed in Table 1.

4 Methodology

As shown above, we have high dimensional control variables included in our datasets and relatively small sample size by experience year. In order to avoid overfitting, multicollinearity and model crush problems, before the data is put into the model, principal component analysis (PCA) on control variable set is conducted. PCA can help reduce the dimension and kill the multicollinearity in the explanatory variables. Then the Double/Debiased Machine Learning[13] (DML) method is applied:

$$Y = D\theta_0 + g_0(\mathbf{X}) + U, \quad \mathbb{E}[U \mid \mathbf{X}, D] = 0 \tag{4.1}$$

$$D = m_0(\mathbf{X}) + V, \quad \mathbb{E}[V \mid \mathbf{X}] = 0 \tag{4.2}$$

Equation (4.1) represents a partial linear Conditional Expectation Function (CEF), with variables we are interested in (could be gender, school rank ... in each time of estimation) linearly involved, denoted by D and all the other characteristic variables lump summed in a nonparametric function $g_0(\mathbf{X})$. Characteristic variables are allowed to be correlated with treatment variables, as is summarized by (4.2), another CEF. We call parameters in $g_0(\mathbf{X})$ and $m_0(\mathbf{X})$ nuisance parameters, since we don't care much about them. The nuisance parameter space is allowed to be increasing with sample size and could be infinite

[13] This method is a Chernozhukov et al. (2018) work.

Table 3: Variables in the dataset

Dummy Variable (=1 if Yes)	***Continuous variable***	***Continuous variable***
Visiting	PhD Graduation School Score(USnews_2013)	Current number of publication in tier 3
South school	School rank in current year	Current number of publication in tier 4
Graduate major in Economics	Speed measure of promoting to associate professor	Current number of publication in tier 5
Undergraduate major in Economics	Average productivity of publication	Cumulative publication value
Undergraduate in STEM	Best rank of publication in current year	Cumulative publication value in tier 1
Male	Best rank of publication till current year	Cumulative publication value in tier 2
Associate Professor	Current publication value	Cumulative publication value in tier 3
Full Professor	Current publication value in tier 1	Cumulative publication value in tier 4
Associate and Full Professor	Current publication value in tier 2	Cumulative publication value in tier 5
Salary total	Current publication value in tier 3	Cumulative number of total publication paper pages
Salary base	Current publication value in tier 4	Cumulative number of publication paper pages in tier 1
	Current publication value in tier 5	Cumulative number of publication paper pages in tier 2
	Current number of total publication papers pages	Cumulative number of publication paper pages in tier 3
Factor Variable	Current number of publication papers pages in tier 1	Cumulative number of publication paper pages in tier 4
Year	Current number of publication papers pages in tier 2	Cumulative number of publication paper pages in tier 5
PhD graduation year	Current number of publication papers pages in tier 3	Cumulative number of total citation
Assistant Professor starting year	Current number of publication papers pages in tier 4	Cumulative number of citation in tier 1
Associate Professor starting year	Current number of publication papers pages in tier 5	Cumulative number of citation in tier 2
Full Professor starting year	Current number of total citation	Cumulative number of citation in tier 3
School code	Current number of citation in tier 1	Cumulative number of citation in tier 4
	Current number of citation in tier 2	Cumulative number of citation in tier 5
	Current number of citation in tier 3	Cumulative number of total coauthors
Continuous variable	Current number of citation in tier 4	Cumulative number of coauthors in tier 1
Experience year	Current number of citation in tier 5	Cumulative number of coauthors in tier 2
Age	Current number of total coauthors	Cumulative number of coauthors in tier 3
Salary	Current number of coauthors in tier 1	Cumulative number of coauthors in tier 4
CPI adjusted salary (base=2008)	Current number of coauthors in tier 2	Cumulative number of coauthors in tier 5
Seniority (total)	Current number of coauthors in tier 3	Cumulative number of total publication
Seniority (continuous)	Current number of coauthors in tier 4	Cumulative number of publication in tier 1
Total years of graduate education	Current number of coauthors in tier 5	Cumulative number of publication in tier 2
Total Years of PhD education	Current number of total publication	Cumulative number of publication in tier 3
PhD Graduation School Rank(IDEAS_2013)	Current number of publication in tier 1	Cumulative number of publication in tier 4
PhD Graduation School Rank(USnews_2013)	Current number of publication in tier 2	Cumulative number of publication in tier 5

Note: "Associate Professor starting year" and "Full Professor starting year" has been set at 3000 if they were not an associate/full professor as of CV date. "Seniority (total)" is calculated by current year minus the year they first started at the current school. "Seniority (continuous)" is considered as years of the most recent continuous period. "Total years of graduate education" includes masters degrees, degrees in non-ECON fields, etc. "Speed measure of promoting to associate professor" is calculated by $1/(\textit{Associate Professor starting year} - \textit{Assistant Professor starting year} + 1)$ if $\textit{Associate professor} == 1$. "Average productivity of publication" is calculated by $\textit{Cumulative publication value}/(\textit{Year} - \textit{PhD graduation year} + 1)$. "Current publication value" is publication value for each person in current year: base on the value of each publication. Publication value is calculated by number of publication pages * $log((1+r(max))/\textit{Journal rank}) * log(2+\textit{number of citation})/(1+\textit{number of coauthors})$, where $r(max)$ is max of journal rank. Tiered variables are continuous and partitioned by journal rank. Tier 1 include journals ranked from 0 to 10; tier 2 include journals ranked from 11 to 50; tier 3 include journals ranked from 51 to 150; tier 4 include journals ranked from 151 to 300; tier 5 include journals ranked 301 and after. We take IDEAS 2015 journal rank here as criterion.

in dimension. We neither know nor specify function forms for g_0 or m_0. They could be nonlinear and/or complicated. $D\theta_0$ in (4.1) is a linear restriction, assuming that the effect is homogeneous to every case in the dataset. It can be used to both binary and continuous treatment variables.

More generally, let treatment D in (4.1) go to the nonparametric function together with $\mathbf{X}$,

$$Y = g_0(D, \mathbf{X}) + U, \quad \mathbb{E}[U \mid \mathbf{X}, D] = 0 \tag{4.3}$$

allowing fully interactions between treatment and controls so that heterogeneous effects can be obtained from this model. According to Chernozhukov et al. (2018) Patial linear model can be applied to both dummy treatment viriable and continuous variable, general nonparametric model can only be applied to dummy treatment variable. So both models are used in the analysis of gender and undergraduate economic major effects, only partial linear model is used in analysis of graduate school rank effect because the rank variable is continuous.

Because it's hard to know what the true CEFs are in reality, the flexibility of $g_0(\mathbf{X})$ and $m_0(\mathbf{X})$ may help us getting close to the true CEFs as much as possible. However, since machine learning techniques are used here in the estimation of $g_0(\mathbf{X})$ and $m_0(\mathbf{X})$, another important point worth to note here is that we care about unconfoundedness or conditional independence assumption (CIA)[14] We are interested in causal meanings so we need CIA to hold. Only with CIA holds, the DML estimate of θ_0 has causal meaning.

The essential idea of DML is (1) to build a Neyman Orthogonal score[15]; the score is a function which needs to satisfy not only a moment condition but also an orthogonality condition to overcome the regularization bias; (2) to remove the bias caused by overfitting by doing sample splitting.

The story starts with a general moment condition as Chernozhukov et al. (2018) equation (2.9),

$$\mathbb{E}(\psi(D, \mathbf{X}; \theta_0, \eta_0)) = 0 \tag{4.4}$$

where ψ is a vector of score functions, it could be in any form, a maximum likelihood score function, a GMM moment function and so on; η_0 denote the true value of nuisance parameters included in g_0 and m_0, $\eta_0 \in \tau$ where τ is the nuisance parameter space. The score function must satisfy an additional condition which is its Gateaux derivative $D_r[\eta - \eta_0]$ exists and being non-sensitive to to the change of nuisance parameters η towards any direction. The Gateaux derivative is

$$D_r[\eta - \eta_0] := \partial_r\{\mathbb{E}[\psi(D, \mathbf{X}; \theta_0, \eta_0 + r(\eta - \eta_0))]\}, \quad \eta \in \tau$$

where $r \in [0, 1)$. The Neyman orthogonality condition is defined as Chernozhukov et al. (2018) definition (2.1), $D_r[\eta - \eta_0]$ exists for all $r \in [0, 1)$ and $\eta \in \tau$ and at $r = 0$,

$$\partial_\eta \, \mathbb{E}\, \psi(D, \mathbf{X}; \theta_0, \eta_0)[\eta - \eta_0] = 0 \tag{4.5}$$

[14]Unconfoundedness means conditioning on $\mathbf{X}$, the counterfactuals $Y(0)$ and $Y(1)$ are uncorrelated with treatment D. CIA means conditioning on $\mathbf{X}$ the choice of D is statistically independent with U, the model error term. They are similar concepts, We treat them as the same in this paper.

[15]Introduced by Neyman (1959)

The two conditions (4.4) and (4.5) together make the DML method different from others[16] and being able to consistently estimate causal effects.

It is necessary to make sure observations from $(D, \mathbf{X})$ is i.i.d.(independent, identically distributed). Also, sample splitting plays an important role. Before any estimation, we should divide the sample into K folds randomly, such that each subsample I_k contains N/K number of observations, where $k \in 1, ..., K$. For each subsample I_k, construct an ML estimator of $\hat{\eta}_{0k}$ and the final DML estimator of interest is solved by

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{nk}[\psi(D, \mathbf{X}; \tilde{\theta}_0, \hat{\eta}_{0k}] = 0 \tag{4.6}$$

When the kth fold as main subsample is used to obtain a DML estimator $\check{\theta}_{0k}$, all the rest subsamples (except the kth fold) is correspondingly an auxiliary sample. So for a K fold splitting, we do have K different main samples and K different auxiliary samples. For each time of estimation, we use the kth subsample I_k as main sample to estimate θ_0 and its corresponding auxiliary sample to estimate m_0 and g_0. Then take average over K estimates. In this partial linear model, a rough estimation procedure would be: (1) split the sample to several folds, treat one fold as a main sample and the rest as auxiliary sample; (2) use ML (in this paper, random forest) to predict D given $\mathbf{X}$ and estimate $g_0(\mathbf{X})$ using auxiliary sample; (3) estimate parameter of interest θ_0 using main sample. (4) do this procedure K times and take average of K estimates, that will be the DML estimator.

5 Empirical Strategy

The empirical study focuses on effects on salary of three interesting factors: Gender, a binary variable as "Male"[17]; young economics professor's Graduation School Rank, a continuous variable as "PhD graduation school rank"[18]; and their Undergraduate Majors, binary variables as "Undergraduate major in Economics" and "Undergraduate major in Economics"[19]. We want to figure out how causally those factors affect salary. We denote salary as Y, its potential outcomes as (Y_0, Y_1) when treatment factor is a binary variable; denote the treatment/causal variable as D and the control variables as $\mathbf{X}$. Once a factor being analyzed (as treatment variable), we treat all the other variables in Table 3 as controls. For both binary and continuous treatment/causal variable, a partial linear structural model

$$Y = D\theta_0 + g_0(\mathbf{X}) + U \tag{5.1}$$

can be applied. Apart from (5.1), a more general structural model

$$Y = g_0(D, \mathbf{X}) + U \tag{5.2}$$

[16]Traditional methods like OLS only need one moment condition (4.4).

[17]A variable name in Table 3

[18]A variable name in Table 3

[19]Variable names in Table 3

can be used when D is a binary variable, where g_0 denotes an unknown/nonparametric function. Model (5.2) is more general than (5.1) since it allows treatment variable to interact with all other variables. When Conditional Independence Assumption (CIA)[20] and Overlap Assumption[21] are satisfied, the average causal effect of a continuous variable is $ACE \equiv \theta_0$; the average treatment effect of a binary variable is $ATE \equiv \mathbb{E}[Y_1 - Y_0]$.[22] Also, according to Wooldridge (2010), when the set of control variables is richer, CIA has more chance to hold. Our dataset provides multi dimensional productivity measures on publication information and other educational and demographic information up to 80 categories, so we believe CIA could hold in our dataset. Meanwhile, doubts from people cannot be avoided, we know we cannot guarantee the variables in this dataset are plenty enough to represent or be able to proxy any unobserved information that affects the determination on salary.[23]

Gender We want to find the salary difference of a male young economics professor relative to a female young economics professor once holding personal educational and productivity information unchanged. Gender difference in salary is proved to be affected by many factors, we believe that as long as we control for enough productivity and background information, ATE of gender is identified by estimating both CEF model (4.1) and (4.3) consistently. We assume

$$(Y_1, Y_0) \perp\!\!\!\perp Gender \mid \mathbf{X} \tag{5.3}$$

here $\mathbf{X}$ includes all the other information except "Male" and "Salary" in Table 3. However, as being mentioned before, it is reasonable to doubt that the CIA might fail. For example, we don't have teaching information (teaching load, teaching ability, etc) in our dataset. Teaching time has been proved shorten people's research time spending, inhibit publications per faculty (Graves et al. (1982)) and related to lower salary (Perna (2001)). People would tend to believe female professors might have higher teaching ability and being payed more, which bias the salary gender gap to be smaller than it actual is. Even though it seems safer to have additional teaching information under control, it won't be a huge concern for us if not. Because we are focusing on public research universities and we presume those schools should assign similar teaching loads to young economics professors. Also, we don't have reputation information included in our dataset which has been stated in the beginning of the paper to be influential as well on salary. That also won't be a big problem since we are looking at young economics professors most of whom have not yet built up solid reputation in the field. As being said, we cannot exclude the situation that CIA fails, but we believe our dataset has already gathered plenty of research and productivity information richer than current studies and essential in studying our research problem. We are confident about showing interesting and informative estimation results.

Graduate School Rank People may be curious about whether PhD graduation school rank has causal effect on young economics professor's salary and how the effect evolves with more working experience years gained. Specifically, when estimating graduate school rank effect, we use "PhD Graduation School rank(IDEAS_2013)" only and omit "PhD Gradua-

[20]Wooldridge (2010) Assumption ATE.1 and a weaker version ATE.1'

[21]Wooldridge (2010) Assumption ATE.2

[22]Wooldridge (2010) Proposition 21.1

[23]According to Wooldridge (2010), section 21.3, unobservables are allowed to be correlated with D only when the unobservables are not correlated with Y_0 and Y_1.

tion School Rank(USnews_2013)" and "PhD graduation School Score(USnews_2013)" from the dataset and put everything else in the control set $\mathbf{X}$. "PhD graduation school rank" is a continuous variable, so only partial linear structural model (5.1) is estimable by DML. For the purpose of being comparable with partial linear model, we create a binary variable as "Rank Dummy": being 1 if rank in the first half of the ranking list; being 0 if rank in the second half of ranking list as to be used in structural model (5.2). Again, we argue that CIA

$$\textit{Graduate Shool Rank} \perp\!\!\!\perp U \mid \mathbf{X} \tag{5.4}$$

is satisfied in our dataset, because the rank of graduation school is mainly related to research and teaching ability of the PhD receiving school, which may influence the young professor's research ability, and we have controlled for the young faculties research ability.

Undergraduate Major Undergraduate education is an important factor for a person's career path so we want to find out how professor's salary are affected by undergraduate major. Undergraduate major may influence salary through many paths, indirectly through graduate school rank, indirectly through human/capital productivity, etc. As we control those information in our model, we are able to check on the causality between Undergraduate majors and salary. Other than simply tracking on economics undergraduate majors, we suspect that STEM majors would also have an influence on the salary and we want to look at how different the ECON major effect is from STEM majors, whether they interact with each other.[24] Because we wonder if an economics faculties' research ability and productivity is related with their science education background. Employers might have the impression that people who have better math, statistics and/or other science major background seems to be more productive in their later on research life even if they don't have a difference in actually paper productivity with others. In order to test it, same as dealing with gender effects, when estimating the average treatment effect (ATE) of undergraduate being economics major on salary of young economics professors we put everything else in the control set $\mathbf{X}$; and when estimate the effect of undergraduate STEM major we switch the position of "Undergraduate major in Economics" and "Undergraduate major in STEM". For each structural model (5.1) and (5.2) we report both "Economics" results and "STEM" results. Because we controlled for "STEM" in $\mathbf{X}$, we believe that CIA

$$(Y_1, Y_0) \perp\!\!\!\perp \textit{Undergraduate major in Economics} \mid \mathbf{X} \tag{5.5}$$

holds and when controlled for "Undergraduate major in Economics" in $\mathbf{X}$

$$(Y_1, Y_0) \perp\!\!\!\perp \textit{Undergraduate in STEM} \mid \mathbf{X} \tag{5.6}$$

holds. Actually, (5.5) and (5.6) seems easier to hold than the previous two. Undergraduate major choices could be correlated with such as years of education, research ability. Thus the undergraduate majors may affect the salary through many paths. Once we control for those demographic and research information, we consider the choice made on undergraduate majors are independent with the potential outcomes.

We believe that within each experience year the observations are independent and iden-

[24]the two variables are not exclusive from each other.

tically distributed (iid). That is because people within each experience year are collected from year 2008/2009 to 2014/2015 who work at research public universities in Table 1. We separately estimate causal effects with respect to different experience years. You can think for each model we are using cross sectional data with iid observations. This iid assumption can rule out the situations when the treatment on one individual affects others' outcomes (Wooldridge (2010)). Again, it is never a bad thing to be critical. Because here iid is a very strong assumption[25], people may doubt on its validity. For example, for newly recruited young economics professors, their salary within one department might be close to each other. Also, when promotion and/or salary increase decisions are made the fundings between departments might be different and for those with limited fundings might have only some qualified but not all professors' salary increased, so young faculties from same department may correlates with each other on salary payments. But at least we can assume iid holds between departments. Since we don't have too many individuals coming from one department in one year, that won't be a huge concern. However, if a lot of people are drew from same department, then we may worry about the potential problem it might cause. Another reason we separately estimate the effects by experience year is that people can come and leave the population pool of research public universities; also, we care about average effect on people who actually work at those research public universities instead of on a fix group of people, so we don't specifically track anyone's career path. we don't care who is in the pool or for how long they are in the pool but we want to make sure within each experience year pool we have iid samples.

Because we separate the data by different experience years, each year's data contains relatively large number of variables (p) compared to the number of observations (n), [26] and the nature of DML procedure of splitting the data into many folds onto each fold the estimation are conducted.[27] we take one more step of doing principle component analysis on the control variables each time before DML estimation and treat all principal components as control variables in the model. This procedure helps to reduce multicollinearity without losing any information. In DML, we use this PCA pre-processed data in the estimation accompany with machine learning methods random forest, neural network and decision tree to build a DML+PCA model, which you will see in the report tables. Also, we use the original data in DML estimation accompany with Post Lasso[28] which is a method more robust to $p > n$ cases as another set of models.

[25]In Wooldridge (2010) chapter 21, the author makes a stronger assumption of iid sample than what is needed Stable Unit Treatment Value Assumption (SUTVA) in treatment literature. Because random sampling implies SUTVA.

[26]More than 80 variables; approximately 90 to 100 observations, only 40 50 in the last years dataset.

[27]In each of estimation the actually used data are less than what it really contained within an experience year.

[28]Post Lasso is a method originated by Belloni and Chernozhukov (2013). Lasso is not applicable at the same time with PCA, because after PCA the principle components are already orthogonal with each other.

6 Results

Table 4: Means and Standard Deviations of Salaries

		Exp0	Exp1	Exp2	Exp3	Exp4
Mean		112,987	119,930	116,417	110,126	110,911
SD		16,014	22,372	18,951	15,460	19,916
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
Mean	111,102	118,783	127,603	133,948	138,424	145,388
SD	23,630	36,333	46,649	54,620	51,163	54,557

Note: Means are in year 2008 dollar value.

Table 5: Gender effects on log salary (DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		0.01	0.045	0.027	0.011	0.108
se(median)		(0.084)	(0.244)	(0.071)	(0.048)	(0.079)
se		(0.072)	(0.207)	(0.071)	(0.038)	(0.067)
B. Partial Linear Model						
ATE		0.035	0.018	0.048	0.037	0.091
se(median)		(0.023)	(0.037)	(0.026)	(0.022)	(0.03)
se		(0.022)	(0.034)	(0.026)	(0.022)	(0.03)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.046	0.068	0.161	0.093	0.085	0.005
se(median)	(0.06)	(0.094)	(0.074)	(0.127)	(0.162)	(0.151)
se	(0.057)	(0.075)	(0.064)	(0.113)	(0.162)	(0.124)
B. Partial Linear Model						
ATE	0.111	0.08	0.161	0.066	0.094	-0.063
se(median)	(0.034)	(0.041)	(0.061)	(0.077)	(0.091)	(0.136)
se	(0.032)	(0.038)	(0.054)	(0.075)	(0.087)	(0.096)

Note: "Exp" is short for experience years. "ATE" reports "best" median treatment effect estimations across splits, here "best" is among Trees, Random Forest and Neural network methods. How "best" are calculated is refered to Chernozhukov et al. (2018). "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. "Splits" means how many times we randomly separate the data into different (pre-setted) folds.

Table 6: Graduate school rank effects on log salary (DML+PCA)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE	-0.166	-0.048	0.018	0.004	0.053
se(median)	(0.17)	(0.148)	(0.087)	(0.046)	(0.095)
se	(0.17)	(0.148)	(0.085)	(0.045)	(0.087)
B. Partial Linear Model					
ATE	0.07	0.084	0.043	0.039	0.165
se(median)	(0.057)	(0.088)	(0.064)	(0.049)	(0.083)
se	(0.051)	(0.074)	(0.059)	(0.048)	(0.076)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.023	0.043	0.151	0.118	0.141	0.092
se(median)	(0.074)	(0.079)	(0.08)	(0.082)	(0.112)	(0.125)
se	(0.074)	(0.064)	(0.072)	(0.078)	(0.105)	(0.125)
B. Partial Linear Model						
ATE	0.132	0.21	0.294	0.041	0.19	0.289
se(median)	(0.099)	(0.131)	(0.187)	(0.187)	(0.147)	(0.211)
se	(0.092)	(0.128)	(0.173)	(0.187)	(0.134)	(0.206)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Estimation on interactive model effects are obtained by creating a dummy variable which is denoted 1 when rank is less or equal to the rank median; 0 if rank is greater than the rank median. Estimation on partial linear model effects are obtained by creating a transformed rank variable, which is calculated by $1 - rank/max(rank)$. In this way, the rank range is between 0 and 1. Better ranked schools are assigned values close to 1, worse ranked schools are assigned values close to 0.

Table 7: Undergraduate ECON effects on log salary(DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		-0.023	-0.039	-0.06	-0.049	-0.03
se(median)		(0.095)	(0.114)	(0.081)	(0.086)	(0.11)
se		(0.092)	(0.107)	(0.081)	(0.067)	(0.08)
B. Partially Linear Model						
ATE		0.004	-0.023	-0.057	-0.036	-0.03
se(median)		(0.035)	(0.037)	(0.035)	(0.031)	(0.04)
se		(0.03)	(0.035)	(0.034)	(0.031)	(0.036)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	-0.058	-0.024	-0.117	-0.17	-0.106	-0.159
se(median)	(0.103)	(0.077)	(0.087)	(0.1)	(0.138)	(0.266)
se	(0.085)	(0.062)	(0.082)	(0.091)	(0.135)	(0.246)
B. Partially Linear Model						
ATE	-0.027	-0.021	-0.127	-0.129	-0.137	-0.098
se(median)	(0.038)	(0.044)	(0.058)	(0.099)	(0.092)	(0.133)
se	(0.038)	(0.043)	(0.055)	(0.087)	(0.091)	(0.121)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp8", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

Table 8: Undergraduate STEM effects on log salary (DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		0.106	-0.062	-0.042	0.014	0.124
se(median)		(0.175)	(0.247)	(0.077)	(0.048)	(0.117)
se		(0.131)	(0.13)	(0.074)	(0.048)	(0.099)
B. Partially Linear Model						
ATE		-0.02	-0.015	0.004	0.031	0.098
se(median)		(0.028)	(0.044)	(0.032)	(0.029)	(0.033)
se		(0.027)	(0.035)	(0.032)	(0.028)	(0.033)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.091	0.138	0.142	0.096	0.101	0.177
se(median)	(0.131)	(0.153)	(0.113)	(0.168)	(0.197)	(0.172)
se	(0.105)	(0.114)	(0.113)	(0.162)	(0.178)	(0.157)
B. Partially Linear Model						
ATE	0.054	0.097	0.089	0.08	0.116	0.119
se(median)	(0.04)	(0.053)	(0.063)	(0.101)	(0.109)	(0.101)
se	(0.039)	(0.048)	(0.058)	(0.098)	(0.106)	(0.093)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp6", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

Table 9: Gender effects on log salary (DML with PostLasso)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE(PostLasso)	0.038	0.022	0.006	0.04	0.071
se(median)	(0.042)	(0.124)	(0.255)	(0.025)	(0.076)
se	(0.032)	(0.104)	(0.255)	(0.025)	(0.066)
B. Partial Linear Model					
ATE(PostLasso)	0.029	0.009	0.045	0.024	0.06
se(median)	(0.024)	(0.034)	(0.028)	(0.026)	(0.03)
se	(0.022)	(0.031)	(0.026)	(0.02)	(0.028)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.117	0.111	0.047	0.168	0.112	0.353
se(median)	(0.117)	(0.08)	(0.262)	(0.167)	(0.26)	(14.5)
se	(0.099)	(0.072)	(0.23)	(0.115)	(0.178)	(7.663)
B. Partial Linear Model						
ATE	0.064	0.084	0.041	0.074	-0.017	-0.035
se(median)	(0.048)	(0.037)	(0.062)	(0.088)	(0.072)	(0.099)
se	(0.027)	(0.036)	(0.049)	(0.059)	(0.07)	(0.097)

Note: "Exp" is short for experience years. "ATE" reports median treatment effect estimations across splits. "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. "Splits" means how many times we randomly separate the data into different (pre-setted) folds.

Table 10: Graduate school rank effects on log salary (DML with PostLasso)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		0.022	0.011	-0.029	-0.032	-0.046
se(median)		(0.042)	(0.049)	(0.149)	(0.083)	(0.124)
se		(0.037)	(0.048)	(0.122)	(0.064)	(0.123)
B. Partial Linear Model						
ATE		0.004	0.016	0.002	0.004	0.027
se(median)		(0.052)	(0.11)	(0.09)	(0.065)	(0.1)
se		(0.048)	(0.083)	(0.049)	(0.051)	(0.076)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	-0.035	0.048	-0.19	-0.939	-0.22	-0.033
se(median)	(0.118)	(0.178)	(0.292)	(1.034)	(0.308)	(0.483)
se	(0.085)	(0.134)	(0.292)	(0.811)	(0.307)	(0.434)
B. Partial Linear Model						
ATE	0.069	0.056	0.087	-0.109	0.009	0.134
se(median)	(0.087)	(0.083)	(0.174)	(0.195)	(0.136)	(0.204)
se	(0.079)	(0.08)	(0.156)	(0.131)	(0.136)	(0.163)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Estimation on interactive model effects are obtained by creating a dummy variable which is denoted 1 when rank is less or equal to the rank median; 0 if rank is greater than the rank median. Estimation on partial linear model effects are obtained by creating a transformed rank variable, which is calculated by $1 - rank/max(rank)$. In this way, the rank range is between 0 and 1. Better ranked schools are assigned values close to 1, worse ranked schools are assigned values close to 0.

Table 11: Undergraduate ECON effects on log salary(DML with PostLasso)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		-0.053	-0.086	-0.052	0.013	-0.006
se(median)		(0.065)	(0.075)	(0.084)	(0.054)	(0.055)
se		(0.052)	(0.056)	(0.076)	(0.043)	(0.052)
B. Partially Linear Model						
ATE		0.024	0.011	-0.05	-0.005	0.034
se(median)		(0.033)	(0.037)	(0.033)	(0.028)	(0.042)
se		(0.028)	(0.035)	(0.029)	(0.027)	(0.031)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.011	0.038	-0.031	-0.45	-0.112	-0.797
se(median)	(0.094)	(0.267)	(0.273)	(0.398)	(0.655)	(2.567)
se	(0.094)	(0.218)	(0.27)	(0.314)	(0.358)	(1.316)
B. Partially Linear Model						
ATE	0.019	0.015	-0.021	-0.135	-0.085	-0.129
se(median)	(0.035)	(0.053)	(0.061)	(0.068)	(0.076)	(0.112)
se	(0.035)	(0.04)	(0.046)	(0.067)	(0.069)	(0.112)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp8", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

Table 12: Undergraduate STEM effects on log salary (DML with PostLasso)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		0.009	-0.021	0.009	0.016	0.098
se(median)		(0.083)	(0.162)	(0.057)	(0.052)	(0.228)
se		(0.065)	(0.093)	(0.048)	(0.046)	(0.135)
B. Partially Linear Model						
ATE		0.001	-0.018	-0.008	0.005	0.044
se(median)		(0.026)	(0.034)	(0.03)	(0.03)	(0.03)
se		(0.025)	(0.034)	(0.028)	(0.027)	(0.03)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.093	0.008	-0.085	0.218	0.308	0.564
se(median)	(0.268)	(0.122)	(0.191)	(1.03)	(0.935)	(1.304)
se	(0.114)	(0.093)	(0.137)	(0.951)	(0.569)	(0.857)
B. Partially Linear Model						
ATE	0.031	0.057	-0.011	0.08	-0.101	0.03
se(median)	(0.039)	(0.044)	(0.069)	(0.076)	(0.094)	(0.108)
se	(0.032)	(0.041)	(0.064)	(0.075)	(0.083)	(0.088)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp6", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

Table 4 gives the average and standard deviation of each experience years CPI adjusted salary. Tables 5–8 show DML with PCA pre-processed datasets estimation results. Tables 9–12 show DML accompany with post Lasso method using raw data results. Specifically, in DML we set fold=2, which means treating half sample as auxiliary sample to estimate $g(.)$ and $m(.)$, the other half as main sample to estimate treatment\causal effect then swap the main and auxiliary sample and do this procedure again; calculate the average of the two, that's one time of DML estimation. In order to make the estimator more robust to outliers, by setting split=5, we run 5 times of DML estimation procedure and take the medium. We take "best"[29] columns from Chernozhukov et al. (2018) result reporting system as our reported ATE estimators in Tables 5–8. We also include a more robust version of standard error which is defined as median methods adjusted standard error[30]. The coefficients are point estimates of causal effects on log CPI adjusted salary. Also, it is worth to note that all the the effects are based on people in the population pool which is young economics professors who work at top 50 research university economics department and are on the

[29]A result getting from choosing best methods for $g(.)$ and $m(.)$ separately.

[30]Chernozhukov et al. (2018) Definition 3.3.

tenure track.[31]

Gender The gender effect estimators of both interactive and partial linear models in Table 5 are in a rough range of 0.01 to 0.1 across all the experience years. That means male economics professors are estimated earning 1% to 10% more on average than female economics professors during their early years in career when holding research productivity and demographic variables constant. However, most of them are neither statistically nor economically significant. When experience is less than 4 years, in both partial linear and interactive models the point estimates are less than 4% without statistical significance. Specifically, in experience year zero, the more general interactive model gives point estimates of 1% which is neither statistically significant nor economically meaningful. There is not enough evidence to show that gender causally make a difference in salary payment once we control for research and productivity measures and other demographic variables.

The coefficients of gender effect goes up and then down with most of them non statistically and economically significant. When working experience years come close to promotion dates, the salary difference caused by gender becomes severe. In years of 6 to 9, the effects are around 10%, which is a significant amount in magnitude and are higher than the previous few years' and peaked in year 7. In experience year 7, the gender effects are quite strong and both statistically and economically significant. Also in Table 9, there exists a similar trend with experience 5 to 7 years' effects much greater than the other years'.

Overall, the positive male and female salary gap are consistent with other gender salary gap studies in the literature such as Perna (2001), Toutkoushian and Conley (2005). However, the lack of statistical significance in most of the coefficients makes our result different from them and the values are smaller than Toutkoushian and Conley (2005) reported of an overall 26% salary difference between full time male and female four-year institution faculties. There might be problems from data collection, and limitation of sample size, which makes the sample standard error large. But according to our study, there is no gender causal effect both economically and statistically on young economics professors salary in their first few career years (before 5 experience years) and slight\moderate gender gap in experience years from 5 to 7 once controlling for research productivity and other demographic variables. Also, this outcome is different from the conclusion of Altonji and Pierret (2001), which asserts, by doing linear regressions on wage, coefficients of easy to observe variables that are highly related to productivity goes down with year of experience years.

The reasons for causality could come from many sources. Statistical discrimination is a highly discussed one.(Altonji and Pierret (2001), Altonji (2005)) Even with similar productivity, employers might still have the stereotype impression that young female researchers have less potential of being as productive as male. Thus providing them less salary even though they have same productivity record as their male counterparts up to the moment when salary has to be decided. People may want to say, especially for recent a few years, there are more protection for women from being discriminated. That's in accord with our empirical results: in the first few experience years there is no gender discrimination between male and female However, in experience year 5 to 7 which is also near most people's

[31] We are not interested in people who are not in this pool, for example people who has been offered a position from any of these universities but not taking it and people who leave out of the tenure track or move to private school after a few years working in the system.

promotion date the gender difference or discrimination is obvious. Another source is negotiation power difference between men and women.(Gerhart and Rynes (1991), Claypool et al. (2017)) Some research show that women negotiate over salaries less than men because they believe women do not have the incentive to move if their demand is not fully satisfied. It has also been suggested that women may not negotiate less frequently than men. But their bargaining power is not as good, so the negotiation results differ. Men get more return from negotiation. Moreover, female usually takes more family burdens like taking care of children and other household duties, which make them less likely to move once settle down. Women are less likely to seek outside offers or searching outside options. (Blackaby et al. (2005))

Those reasons mentioned above might also work together causing a gender difference in 5 to 7 years of their early stage in career. Men have more bargaining power might because they have more options with outside offers and have more incentive to move if obtained better options. However, women may be less likely to move once offered other choices. Even though women negotiate, the power is not strong. Woman's personal characteristics of being more tolerant than man. All of these may consequently lead to the difference reflected on their salaries among these years.

PhD graduation school rank The original graduate school rank variable is continuous and only applicable to partial linear model, in order to make the model results easier to interpret and comparable with each other, we do two transformations on it. First, we create a dummy variable for graduate school rank. The rank dummy is assigned 1 when the actual PhD graduation school rank is less or equal to the rank median over all individuals in that experience year; 0 if rank is greater than the rank median. Panel A in Table 6 and Table 10 reports interactive model results estimated with the rank dummy. Second, we transfer the original rank variable to be in a range of 0 to 1 values, letting the smallest number representing the least ranked school and highest number representing the highest ranked school. The transferred continuous rank variable is calculated by $1 - rank/max(rankofthatyear)$. Panel B in Table 6 and Table 10 report the partial linear model results estimated with a transferred continuous rank variable.

There are explanation difference between the two models. Take the effect of experience year 5 as an example: in the interactive model, the estimate of ATE 0.023 means a change from a worse than median ranked school to a better than median ranked school would increase the average salary of young economics professor with 5 working experience years by 2.3%; in partial linear model, the school rank increase by 10 percentile the the average salary of young economics faculty with 5 working experience year would increase by 1.32%. Partial linear model provides more insights on the quantiles but it has been restricted to a linear effect. Interactive model is more general on the function form, but it conveys less information, because it only separate all schools into two category. Obviously the highest ranked school would have much difference from the lowest ranked one with in each category.

Overall, the partial linear Rank effect model gives larger in the magnitude estimates across all experience years than those from interactive model. In interactive model, Rank effect estimates take a range roughly from 0.01 to 0.15 without statistical significance. In partial linear model, it takes a range approximately from 0.07 to 0.2 not statistically significant either.[32] Except seeing negative signs on the first two experience years estimates in

[32]We would prefer to focus on Table 6 results because Table 10 gives too many negative point estimates

interactive model, most of the rank effect estimates are positive. That means graduating from higher ranked school could help a young economics professor earning more money, even if with other conditions (productivity, other background) being hold constant. This result is accord with Stock and Alston (2000), saying there is a positive correlation between program rank and initial salaries, even after controlling for information on qualifications such as research and teaching. The effects become larger from year 6 to 9. However, we notice that even though in years like 6 and 9 the point estimates are large, there is not enough evidence to show those are significantly different from zero. In some years, the estimates fall on the upper bound of its 95% confidence interval but that includes zero. Therefore, it is hard to say there is PhD graduation school rank casual effect on young economics professors' salary in most of the early working experience years. However, we must be careful with rank effects in year 7, 8, and 9, even if those are not statistically significant, they are indeed economically significant.

Reasonably, the employers might have discriminated them in salary based on PhD graduation rank during those years because those are years near promotion. The discrimination could associate with the facts that people graduate from higher ranked universities also obtain higher potential human capital(Claypool et al. (2017)). Moreover, perhaps coming from a higher-ranked PhD program is helpful for getting positive letters for promotion and tenure, which usually correspond to a salary increase.

Undergraduate Major Tables 7 and 8, and Tables 11 and 12 show how the two related and interesting factors, undergraduate major in economics and undergraduate major in STEM majors, affect young economics professors salary. They are not exclusive to each other, because there exists cases that one person owns a dual degree with one of them economics the other a STEM major. When estimating one effect we take the other under control.

The economics undergraduate major negatively influences salary. The range of the effects is approximately from -0.17 to -0.02. That's a 2% to 17% decrease in salary. Economics undergraduate major may not be a good sign if you want to earn a higher salary. And similarly, the estimates are quite small and not statistically significant in the first few experience years (0 to 6), such that there is no difference in salary caused by whether you own an undergraduate economics major in these years once control for research productivity and demographic information. However, we should pay attention that in both interactive and partial linear model, the effect estimates become larger from experience year 7. Again, since those point estimates are not statistically significant, even with large point estimates cannot exclude them from zero effects.

Other than on first few experience years (0 to 2) where the estimates take negative sign, the STEM undergraduate majors seems positively affect salary with controls. Similar to economics undergraduate major effects, the STEM undergraduate major effects are not statistically significant either. We cannot confidently distinguish them from zero. The point estimates take a range approximately from -0.04 to 0.14. STEM undergraduate majors would tend to positively affect salary compared to economics undergraduate major. However, we should be careful that we cannot exclude them from zero effect because of the absence of statistical significance.

with very large standard errors which is less informative than Table 6.

Overall, holding an economics undergraduate degree may not be as helpful as holding a STEM major undergraduate degree for a better salary even with similar research and demographic background. As mentioned in Section 2, observable economics talent is a key determinants in the PhD admission process. Among them quantitative ability is an important factor to be considered. Because quantitative ability has been proved to be highly correlated with research productivity(Grove and Wu (2007)). Also, Zhang (2005) showed that social science, business and some other art undergraduate majors have negative coefficients on the probability of graduate enrollment whereas math, biological and other science majors have positive coefficients on the probability of graduate enrollment. That explains what we observe from the estimation.

Post Lasso Selected Variables Figures 1 to 9 give an idea on the variables selected by post lasso method in "DML with PostLasso" interactive models. This part is trying to show what DML did with machine learning. In interactive models, DML estimates $g(1, \mathbf{X})$ and $g(0, \mathbf{X})$ separately so we can see there are two different series representing each. The vertical axis shows the frequencies of each variable being selected over all $(fold = 2) * (split = 5)$ times the horizontal axis shows the variables being used in that specific experience year's model. Some variables are selected more than once and by both functions; some are selected only by one function with very few times. Since the reported ATE are aggregating over all folds and splits, we consider them as "selected" as long as they have ever been selected in the DML estimation process. The frequencies difference might be caused by sample size, which leaves more randomness to the results. You may also wonder why some variables are not selected at all. That is because the mechanism of Lasso push hard on some of the variable coefficients to be exactly zero, resulting them to be not "selected" by the model. That is not to say they are not meaningful under control. That is exactly where the difference between traditional methods and DML come from. The L1 regularization terms in Lasso (Tibshirani (1996)) will give exactly zero solutions on regression coefficients for the purpose of seeking minimum MSE under bias and variance trade off. Remember as mentioned in section 4, the regularization terms in machine learning methods could cause regularization bias, even so, DML can reduce the regularization bias and give consistent causal estimators. But if you simply put everything into a OLS model, the estimation would crush because of the singularity problem let alone finding causal estimators. We also notice that those variables selected by different effect and year models are quite similar to each other. They all select year, schoolcode and research productivity by journal rank tiers variables as a subset of control.

The graphs provide insights on how DML works, also provoke more thinking. The good thing is that we know which variables are important in the estimation of $g(.)$ and $m(.)$ functions for prediction purpose and those variables convince that it's worth to control for a detailed research productivity measures as well as background information. The bad side is that because of the randomness we can hardly see a variable being selected with high frequency (more than 5/10).

Discussion Since economics departments experiencing a hard time in learning the true research ability of their young faculties, we suspect that in determining young economics professors' salaries employers would not only rely on publication information but also other easier to observe information that they may think related to research ability. With our novel dataset we are able to control for more detailed research productivity measures, then the estimation of causal effects of Gender, PhD graduation school rank and Undergraduate majors on salary could be conducted successfully. we wonder if there are some other reason than research ability that affects the determination in salary. But like has been said by Ginther and Hayes (2003) we cannot make a sure conclusion that the difference is only come from discrimination or any other single source. For some of the early career year, we can not say there is an effect from all three aspects when control for research productivity and other information; for some later year in career, the gender causal effect are strong and significant and may come from multiple reasons. Statistical discrimination based on PhD graduation school rank and Undergraduate majors could happen during later experience years like 6, 7 or 8. But we must be careful, because those estimates are not statistically significant in DML models. Overall, we notice a jump on the point estimates of almost all models from experience year 6 to 8. That could be an evidence showing that during those years (near promotion dates) young economics professors may experience discrimination based on gender, PhD background as well as Undergraduate major.

7 Robustness Check

Table 13: Robustness Check: Gender Effects (DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		0.059	0.052	0.057	0.056	0.105
se(median)		(0.035)	(0.056)	(0.047)	(0.044)	(0.057)
se		(0.031)	(0.05)	(0.045)	(0.032)	(0.054)
B. Partial Linear Model						
ATE		0.078	0.064	0.055	0.036	0.094
se(median)		(0.027)	(0.038)	(0.032)	(0.024)	(0.033)
se		(0.026)	(0.037)	(0.032)	(0.023)	(0.032)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.11	0.117	0.089	0.203	0.061	NA
se(median)	(0.059)	(0.083)	(0.097)	(0.086)	(0.131)	(NA)
se	(0.048)	(0.073)	(0.094)	(0.086)	(0.129)	(NA)
B. Partial Linear Model						
ATE	0.092	0.109	0.126	0.161	0.066	-0.129
se(median)	(0.035)	(0.047)	(0.07)	(0.073)	(0.09)	(0.092)
se	(0.033)	(0.046)	(0.065)	(0.068)	(0.087)	(0.09)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. NA shows up in the whole column of Exp10, meaning interactive model does not work on this specific dataset and no result reported.

Table 14: Robustness Check: Graduate School Rank Effects (DML+PCA)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE	0	0.076	0.008	0.014	0.036
se(median)	(0.057)	(0.072)	(0.046)	(0.041)	(0.044)
se	(0.046)	(0.056)	(0.044)	(0.041)	(0.044)
B. Partial Linear Model					
ATE	0.111	0.171	0.063	0.068	0.149
se(median)	(0.057)	(0.09)	(0.057)	(0.06)	(0.084)
se	(0.056)	(0.075)	(0.056)	(0.055)	(0.081)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.064	0.145	0.187	0.128	0.079	0.114
se(median)	(0.056)	(0.056)	(0.08)	(0.133)	(0.149)	(0.143)
se	(0.046)	(0.056)	(0.075)	(0.132)	(0.136)	(0.141)
B. Partial Linear Model						
ATE	0.113	0.26	0.216	0.143	0.338	0.035
se(median)	(0.061)	(0.133)	(0.221)	(0.147)	(0.221)	(0.279)
se	(0.059)	(0.124)	(0.208)	(0.142)	(0.206)	(0.253)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Estimation on interactive model effects are obtained by creating a dummy variable which is denoted 1 when rank is less or equal to the rank median; 0 if rank is greater than the rank median. Estimation on partial linear model effects are obtained by creating a transformed rank variable, which is calculated by $1 - rank/max(rank)$. In this way, the rank range is between 0 and 1. Better ranked schools are assigned values close to 1, worse ranked schools are assigned values close to 0.

Table 15: Robustness Check: Undergraduate Econ Effects (DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		-0.016	-0.002	-0.069	-0.034	-0.007
se(median)		(0.072)	(0.101)	(0.064)	(0.056)	(0.067)
se		(0.068)	(0.099)	(0.064)	(0.056)	(0.065)
B. Partially Linear Model						
ATE		-0.001	-0.025	-0.068	-0.048	-0.016
se(median)		(0.039)	(0.047)	(0.047)	(0.036)	(0.038)
se		(0.038)	(0.045)	(0.045)	(0.035)	(0.036)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	-0.037	-0.11	-0.159	-0.163	-0.307	NA
se(median)	(0.068)	(0.1)	(0.073)	(0.116)	(0.218)	(NA)
se	(0.062)	(0.084)	(0.072)	(0.112)	(0.215)	(NA)
B. Partially Linear Model						
ATE	-0.025	-0.066	-0.17	-0.172	-0.292	-0.148
se(median)	(0.042)	(0.054)	(0.067)	(0.109)	(0.127)	(0.222)
se	(0.041)	(0.05)	(0.065)	(0.105)	(0.119)	(0.2)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. NA shows up in the whole column of Exp10, meaning interactive model does not work on this specific dataset and no result reported.

Table 16: Robustness Check: Undergraduate STEM Effects (DML+PCA)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		-0.029	-0.018	0.038	0.046	0.106
se(median)		(0.063)	(0.059)	(0.058)	(0.042)	(0.066)
se		(0.06)	(0.058)	(0.055)	(0.042)	(0.057)
B. Partially Linear Model						
ATE		-0.019	-0.008	0.043	0.049	0.089
se(median)		(0.035)	(0.043)	(0.039)	(0.03)	(0.037)
se		(0.035)	(0.042)	(0.037)	(0.029)	(0.036)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.089	0.111	0.116	0.146	0.168	0.068
se(median)	(0.067)	(0.137)	(0.105)	(0.134)	(0.204)	(0.22)
se	(0.061)	(0.128)	(0.101)	(0.133)	(0.202)	(0.19)
B. Partially Linear Model						
ATE	0.053	0.08	0.104	0.138	0.174	0.01
se(median)	(0.044)	(0.062)	(0.072)	(0.112)	(0.122)	(0.198)
se	(0.042)	(0.059)	(0.068)	(0.11)	(0.118)	(0.131)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits.

In order to test the model specification in DML estimation, we do another set of models as a robustness check. They are different from the original DML models in the control variables we use. The robustness check models will do exactly the four previous DML models as in Tables 5–8 except we change the control variables in X. The robustness check models include next years' current and cumulative productivity measures ,current year current productivity measures and all the other demographic measures.[33] As has been mentioned previously, Because economics papers usually take a while to be officially published after being accept, people may update their resumes by a "forth coming" paper in some year if the paper was officially published the year after. Employers can observe those changing information with time goes by, whereas our data set is not able to catch "forth coming" information. Therefore, we want to see how it look like to include next years productivity information and whether this change is robust to the model performance or not. Tables 13–16 reports the robustness check model results.[34] There is not big difference in point estimates and overall trend between the main DML models and robustness check models. You may notice that the

[33] We do not include current year cumulative productivity information, because we believe that next year's cumulative productivity measures have already contain the current year cumulative information. But current productivity is not representable by next years' information.

[34] In the result tables, if NA shows up in the whole column, meaning this model does not work on this specific dataset and no result reported; if NA shows up in se(median), meaning no "best" result reported and we report a median value among all ML estimators and its corresponding se.

point estimate range change a little bit but overall they are pretty similar to the main DML models. For example, the range of Gender effect in Robust check models is approximately 0.03 to 0.16, slightly shifting higher compared to the main DML model and the estimates are mostly statistically insignificant; the range of graduate school rank effect in robust check interactive model is about 0.01 to 0.18, in partial linear model is about 0.03 to 0.2, those are about the same as in main DML model.

8 Conclusion

This paper makes the following contributions. First, we created a novel dataset that has never been studied before in the literature. This dataset can be helpful in many research questions. Second, we applied a newly boarded method in consistently estimating causal effect with high dimensional control variables. We are able to compare them with traditional causal effect models, to show the variable selected and to discuss the advantages and disadvantages. Last but not least, we look into how easy-to-observe variables that are correlated with productivity influence a young economics professor's salary once holding constant the productivity measurements.

Meanwhile, the results could be reasonably doubted. Why are the larger effect estimates not statistically significant? Are they because the true effect is indeed zero or because we don't have enough data to reduce the randomness? Note that our data is collected from public website, different states may have difference criterion in reporting salaries. And because of the limit of sample size, the noise from the data might affect the result quite a lot.

In the future, it is useful to expand the current work in two directions, one is expanding the current dataset, try to give more precise confidence intervals for point estimates; the other is trying to dig the DML model even further to make it more visible showing the estimation process and results and make it more flexible to combine with other deep learning methods.

9 Appendix

9.1 Supplementary Results

People may be curious about how OLS performs. Would they be different or similar with DML estimation? Tables 17–23 show the OLS estimation results. They are conducted using the same data as Tables 5–8. We report three kinds of models. The first one is a simple OLS model with only the treatment variable itself, as reported in panel A; the second one is an OLS model with only background information as control variables and report with 5 principal components and with 10 principal component results as showed in panel B; the third one, we include all (background and productivity) information as control variables and also report 5 principal and 10 principal components results.[35] DML results are thus more comparable with

[35]Like how we process the data in DML estimation, we do principal component analysis on all the control variables we want to include into the OLS model, and put the principal components together with treatment variables into OLS model. 5 principal components are the top 5 ones explaining the variance of the original variable set; 10 principal components are the top 10 ones explaining the variance of the original variable set.

OLS panel C. We notice that when adding more information to the control the coefficients are decreasing towards zero, and within each panel, when include more principal components the coefficients are also decreasing. These are consistent with our perception, because people would suspect there is a positive bias due to the unobservant in the simple OLS model. And again, with the thoery from Wooldridge (2010) saying that with more controls in the model the treatment coefficient is more likely to have causal meaning, people would believe more in the OLS results with more information controlled. In Gender effect models, compared with OLS model, DML (interactive) is a totally functional form relaxed model, gives near zero estimates in first few experience years estimates and greater estimates in experience year 6 to 7 than the OLS. In graduate school rank effect models, OLS panel C gives overall smaller point estimates than DML and even several negative point estimates. We are afraid that if it is OLS that put too much restriction onto the strcuture that make the estimates towards zero. In undergraduate major models, if we compare panel C of table 21 and 23 and DML results, we may find that the overall signs are the same, undergraduate ECON major positive and undergraduate STEM major negative. But OLS results might still be more close to zero than DML results.

Table 17: Gender effects on log salary (OLS)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. only gender dummy					
coefficient	0.053	0.027	0.051	0.042	0.098
se	(0.024)	(0.038)	(0.029)	(0.023)	(0.032)
B. add controls W/O productivity					
(5 principal components)					
coefficient	0.064	0.038	0.07	0.068	0.103
se	(0.026)	(0.04)	(0.027)	(0.02)	(0.032)
(10 principal component)					
coefficient	0.068	0.022	0.062	0.043	0.047
se	(0.028)	(0.037)	(0.027)	(0.022)	(0.028)
C. add all controls					
(5 principal components)					
coefficient	0.055	0.024	0.044	0.037	0.075
se	(0.026)	(0.041)	(0.029)	(0.022)	(0.029)
(10 principal component)					
coefficient	0.067	0.018	0.056	0.055	0.082
se	(0.029)	(0.042)	(0.03)	(0.023)	(0.03)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. only gender dummy						
coefficient	0.116	0.138	0.168	0.135	0.082	0.067
se	(0.035)	(0.044)	(0.06)	(0.077)	(0.098)	(0.106)
B. add controls W/O productivity						
(5 principal component)						
coefficient	0.114	0.153	0.132	0.111	0.006	0.141
se	(0.037)	(0.04)	(0.06)	(0.069)	(0.01)	(0.144)
(10 principal component)						
coefficient	0.074	0.055	0.066	0.061	0.067	0.006
se	(0.033)	(0.038)	(0.052)	(0.078)	(0.105)	(0.143)
C. add all controls						
(5 principal component)						
coefficient	0.082	0.07	0.055	0.048	-0.03	-0.061
se	(0.031)	(0.038)	(0.045)	(0.063)	(0.08)	(0.092)
(10 principal component)						
coefficient	0.08	0.078	0.085	0.038	-0.029	-0.093
se	(0.031)	(0.038)	(0.049)	(0.056)	(0.076)	(0.094)

Note: Panel A shows simple regression of salary on gender dummy. Panel B includes demographic, educational and year fix effect, school fix effect but not productivity measures. Panel C includes all control variables. All OLS result tables report heteroscedasticity robust standard errors.

Table 18: Graduate school rank effects on log salary (OLS)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. only school rank					
coefficient	0.102	0.172	0.078	0.07	0.191
se	(0.059)	(0.101)	(0.072)	(0.057)	(0.092)
B. add controls W/O productivity					
(5 principal component)					
coefficient	0.056	0.175	0.022	0.064	0.184
se	(0.055)	(0.105)	(0.089)	(0.072)	(0.092)
(10 principal component)					
coefficient	0.014	-0.142	-0.027	-0.008	0.012
se	(0.077)	(0.137)	(0.097)	(0.082)	(0.112)
C. add all controls					
(5 principal component)					
coefficient	0.082	0.184	0.024	0.059	0.09
se	(0.052)	(0.082)	(0.073)	(0.054)	(0.075)
(10 principal component)					
coefficient	0.032	0.161	-0.043	0.033	0.101
se	(0.073)	(0.104)	(0.097)	(0.063)	(0.071)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. only school rank						
coefficient	0.21	0.293	0.288	0.014	0.175	0.279
se	(0.094)	(0.151)	(0.254)	(0.249)	(0.141)	(0.389)
B. add control W/O productivity						
(5 principal component)						
coefficient	0.136	0.17	0.191	-0.041	0.14	0.315
se	(0.085)	(0.13)	(0.214)	(0.229)	(0.147)	(0.39)
(10 principal component)						
coefficient	-0.026	0.026	0.07	-0.136	0.031	0.174
se	(0.1)	(0.107)	(0.138)	(0.152)	(0.128)	(0.25)
C. add all controls						
(5 principal component)						
coefficient	0.108	0.058	0.125	-0.118	-0.009	0.291
se	(0.081)	(0.08)	(0.144)	(0.166)	(0.158)	(0.285)
(10 principal component)						
coefficient	0.078	0.037	0.142	-0.073	-0.016	0.308
se	(0.076)	(0.086)	(0.175)	(0.145)	(0.167)	(0.292)

Note: Transformed rank variable is being used, calculated by $1 - rank/max(rank)$. Panel A shows simple regression of salary on rank. Panel B includes demographic, educational and year fix effect, school fix effect. Panel C includes all.

Table 19: Graduate school rank (dummy) effects on log salary (OLS)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. only school rank					
coefficient	0.033	0.049	0.019	0.016	0.046
se	(0.029)	(0.034)	(0.03)	(0.027)	(0.033)
B. add controls W/O productivity					
(5 principal component)					
coefficient	0.017	0.043	0.015	0.007	0.041
se	(0.031)	(0.037)	(0.032)	(0.027)	(0.036)
(10 principal component)					
coefficient	0.013	0.009	0.004	-0.023	-0.027
se	(0.035)	(0.038)	(0.027)	(0.027)	(0.032)
C. add all controls					
(5 principal component)					
coefficient	0.018	0.043	0.024	0.003	0.01
se	(0.032)	(0.038)	(0.031)	(0.26)	(0.031)
(10 principal component)					
coefficient	0.014	0.037	0.007	-0.012	0.005
se	(0.037)	(0.04)	(0.034)	(0.027)	(0.033)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. only school rank						
coefficient	0.053	0.111	0.186	0.14	0.105	0.109
se	(0.037)	(0.045)	(0.057)	(0.077)	(0.084)	(0.103)
B. add control W/O productivity						
(5 principal component)						
coefficient	0.04	0.085	0.151	0.042	0.068	0.161
se	(0.038)	(0.046)	(0.054)	(0.079)	(0.099)	(0.098)
(10 principal component)						
coefficient	-0.017	0.01	0.028	-0.048	0.037	0.058
se	(0.033)	(0.046)	(0.058)	(0.068)	(0.08)	(0.087)
C. add all controls						
(5 principal component)						
coefficient	0.023	0.038	0.079	0.014	-0.036	0.076
se	(0.034)	(0.041)	(0.045)	(0.06)	(0.067)	(0.08)
(10 principal component)						
coefficient	0.016	0.026	0.09	-0.043	-0.049	0.083
se	(0.034)	(0.043)	(0.051)	(0.054)	(0.076)	(0.067)

Note: Dummy rank variable is being used, equal to 1 when rank is less or equal to the rank median, 0 if rank is greater than the rank median. Panel A shows simple regression of salary on dummy rank. Panel B includes demographic, educational and year fix effect, school fix effect. Panel C includes all.

Table 20: Undergraduate major effects on log salary (OLS coefficients)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. only dummy					
Ugrad_ECON	0.005	-0.015	-0.071	-0.038	-0.038
se	(0.033)	(0.04)	(0.037)	(0.032)	(0.039)
Ugrad_STEM	-0.024	-0.035	0.017	0.03	0.102
se	(0.03)	(0.039)	(0.036)	(0.03)	(0.034)
B. add controls W/O productivity					
Ugrad_ECON	0.004	-0.023	-0.019	-0.013	0.014
se	(0.034)	(0.041)	(0.034)	(0.036)	(0.047)
Ugrad_STEM	0.023	-0.021	0.067	0.047	0.111
se	(0.054)	(0.069)	(0.07)	(0.06)	(0.064)
Ugrad_ECON*Ugrad_STEM	-0.069	0.034	-0.097	-0.039	-0.032
se	(0.06)	(0.088)	(0.081)	(0.066)	(0.078)
C. add all controls					
Ugrad_ECON	0.018	0	-0.046	-0.033	0.002
se	(0.035)	(0.045)	(0.037)	(0.038)	(0.053)
Ugrad_STEM	0.027	0.009	0.06	0.032	0.095
se	(0.053)	(0.08)	(0.072)	(0.059)	(0.063)
Ugrad_ECON*Ugrad_STEM	-0.077	-0.001	-0.099	-0.007	-0.028
se	(0.062)	(0.088)	(0.081)	(0.065)	(0.072)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. only dummy						
Ugrad_ECON	-0.021	-0.051	-0.133	-0.174	-0.15	-0.192
se	(0.042)	(0.048)	(0.06)	(0.094)	(0.103)	(0.169)
Ugrad_STEM	0.054	0.107	0.109	0.096	0.136	0.18
se	(0.042)	(0.053)	(0.063)	(0.11)	(0.115)	(0.122)
B. add controls W/O productivity						
Ugrad_ECON	-0.032	-0.073	-0.15	-0.223	-0.233	0.008
se	(0.062)	(0.074)	(0.096)	(0.089)	(0.114)	(0.289)
Ugrad_STEM	0.001	-0.034	-0.026	-0.065	-0.061	0.439
se	(0.081)	(0.085)	(0.103)	(0.17)	(0.227)	(0.319)
Ugrad_ECON*Ugrad_STEM	0.048	0.208	0.169	0.134	0.209	-0.342
se	(0.096)	(0.112)	(0.56)	(0.229)	(0.282)	(0.37)
C. add all controls						
Ugrad_ECON	-0.013	-0.044	-0.147	-0.196	-0.194	0.028
se	(0.061)	(0.048)	(0.06)	(0.078)	(0.095)	(0.103)
Ugrad_STEM	0.03	-0.055	-0.084	0.038	-0.146	0.282
se	(0.077)	(0.068)	(0.074)	(0.163)	(0.162)	(0.194)
Ugrad_ECON*Ugrad_STEM	0.017	0.171	0.192	0.074	0.216	-0.258
se	(0.092)	(0.09)	(0.107)	(0.191)	(0.2)	(0.254)

Note: Panel A shows simple regression of log salary on undergraduate major dummy separately. Panel B controls demographic, educational and year fix effect, school fix effect as being principle components. Panel C controls all information as being principle components. Both B and C are using 5 principal components as control variables.

Table 21: Undergraduate major effects on log salary (OLS ATE)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. only dummy					
Ugrad_ECON	0.005	-0.015	-0.071	-0.038	-0.038
se	(0.033)	(0.04)	(0.037)	(0.032)	(0.039)
Ugrad_STEM	-0.024	-0.035	0.017	0.03	0.102
se	(0.03)	(0.039)	(0.036)	(0.03)	(0.034)
B. add controls W/O productivity					
ATE (Ugrad_ECON)	-0.02	-0.013	-0.046	-0.024	0.004
se	(0.031)	(0.038)	(0.033)	(0.031)	(0.037)
ATE (Ugrad_STEM)	-0.027	0.004	-0.005	0.019	0.088
se	(0.03)	(0.043)	(0.032)	(0.027)	(0.037)
C. add all controls					
ATE (Ugrad_ECON)	-0.009	0	-0.074	-0.035	-0.007
se	(0.032)	(0.04)	(0.033)	(0.032)	(0.04)
ATE (Ugrad_STEM)	-0.028	0.008	-0.013	0.027	0.074
se	(0.028)	(0.04)	(0.031)	(0.025)	(0.033)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. only dummy						
Ugrad_ECON	-0.021	-0.051	-0.133	-0.174	-0.15	-0.192
se	(0.042)	(0.048)	(0.06)	(0.094)	(0.103)	(0.169)
Ugrad_STEM	0.054	0.107	0.109	0.096	0.136	0.18
se	(0.042)	(0.053)	(0.063)	(0.11)	(0.115)	(0.122)
B. add control W/O productivity						
ATE (Ugrad_ECON)	-0.017	-0.016	-0.103	-0.19	-0.184	-0.091
se	(0.05)	(0.061)	(0.074)	(0.086)	(0.098)	(0.208)
ATE (Ugrad_STEM)	0.036	0.114	0.084	0.026	0.095	0.158
se	(0.047)	(0.059)	(0.074)	(0.107)	(0.119)	(0.139)
C. add all controls						
ATE (Ugrad_ECON)	-0.008	0.003	-0.093	-0.177	-0.143	-0.046
se	(0.048)	(0.042)	(0.05)	(0.078)	(0.082)	(0.065)
ATE (Ugrad_STEM)	0.043	0.068	0.041	0.088	0.015	0.07
se	(0.044)	(0.05)	(0.058)	(0.105)	(0.093)	(0.087)

Note: Panel A shows simple regression of log salary on undergraduate major dummy separately. Panel B controls demographic, educational and year fix effect, school fix effect as being principle components. Panel C controls all information as being principle components. Both B and C are using 5 principal components as control variables.

Table 22: Undergraduate major effects on log salary (OLS coefficients) II

	Exp0	Exp1	Exp2	Exp3	Exp4
B. add controls W/O productivity					
Ugrad_ECON	0.011	0.007	-0.041	-0.037	0.023
se	(0.04)	(0.036)	(0.038)	(0.035)	(0.053)
Ugrad_STEM	0.052	-0.007	0.027	0.002	0.056
se	(0.054)	(0.077)	(0.06)	(0.058)	(0.06)
Ugrad_ECON*Ugrad_STEM	-0.096	0.009	-0.067	-0.001	-0.006
se	(0.062)	(0.089)	(0.071)	(0.063)	(0.072)
C. add all controls					
Ugrad_ECON	0.002	-0.002	-0.038	-0.043	0.03
se	(0.039)	(0.047)	(0.037)	(0.039)	(0.054)
Ugrad_STEM	0.013	0.012	0.067	0.022	0.125
se	(0.057)	(0.083)	(0.073)	(0.063)	(0.064)
Ugrad_ECON*Ugrad_STEM	-0.052	0.001	-0.104	0.003	-0.07
se	(0.061)	(0.09)	(0.085)	(0.068)	(0.075)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
B. add control W/O productivity						
Ugrad_ECON	-0.006	-0.044	-0.102	-0.153	-0.108	0.062
se	(0.057)	(0.057)	(0.088)	(0.101)	(0.134)	(0.103)
Ugrad_STEM	0.044	-0.011	-0.079	-0.04	0.038	0.385
se	(0.071)	(0.064)	(0.094)	(0.172)	(0.207)	(0.197)
Ugrad_ECON*Ugrad_STEM	-0.032	0.139	0.209	0.178	0.108	-0.189
se	(0.09)	(0.09)	(0.126)	(0.21)	(0.261)	(0.272)
C. add all controls						
Ugrad_ECON	-0.003	-0.05	-0.164	-0.161	-0.19	0.041
se	(0.064)	(0.055)	(0.071)	(0.09)	(0.119)	(0.132)
Ugrad_STEM	0.022	-0.07	-0.114	0.031	-0.194	0.338
se	(0.079)	(0.072)	(0.086)	(0.173)	(0.2)	(0.227)
Ugrad_ECON*Ugrad_STEM	0.006	0.185	0.199	0.037	0.241	-0.363
se	(0.092)	(0.089)	(0.118)	(0.204)	(0.248)	(0.291)

Note: Panel B controls demographic, educational and year fix effect, school fix effect as being principle components. Panel C controls all information as being principle components. Both B and C are using 10 principal components as control variables.

Table 23: Undergraduate major effects on log salary (OLS ATE) II

	Exp0	Exp1	Exp2	Exp3	Exp4
B. add controls no productivity					
ATE (Ugrad_ECON)	-0.022	0.01	-0.06	-0.037	0.021
se	(0.033)	(0.034)	(0.033)	(0.029)	(0.039)
ATE (Ugrad_STEM)	-0.017	-0.001	-0.022	0.002	0.051
se	(0.031)	(0.041)	(0.028)	(0.029)	(0.033)
C. add all controls					
ATE (Ugrad_ECON)	-0.016	-0.003	-0.067	-0.042	0.008
se	(0.034)	(0.04)	(0.035)	(0.033)	(0.04)
ATE (Ugrad_STEM)	-0.025	0.013	-0.01	0.024	0.073
se	(0.032)	(0.043)	(0.031)	(0.03)	(0.034)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
B. add control no productivity						
ATE (Ugrad_ECON)	-0.016	-0.006	-0.043	-0.109	-0.083	0.008
se	(0.044)	(0.046)	(0.069)	(0.093)	(0.121)	(0.091)
ATE (Ugrad_STEM)	0.02	0.088	0.056	0.081	0.118	0.229
se	(0.043)	(0.05)	(0.062)	(0.095)	(0.138)	(0.102)
C. add all controls						
ATE (Ugrad_ECON)	-0.001	0.001	-0.108	-0.152	-0.133	-0.064
se	(0.05)	(0.048)	(0.057)	(0.079)	(0.093)	(0.091)
ATE (Ugrad_STEM)	0.027	0.062	0.015	0.057	-0.015	0.039
se	(0.043)	(0.051)	(0.06)	(0.1)	(0.096)	(0.095)

Note: Panel B controls demographic, educational and year fix effect, school fix effect as being principle components. Panel C controls all information as being principle components. Both B and C are using 10 principal components as control variables.

Table 24: Gender effects on standardized salary (DML)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE	0.355	0.145	0.364	0.327	0.511
se(median)	(0.564)	(0.228)	(0.235)	(0.298)	(0.296)
se	(0.251)	(0.216)	(0.2)	(0.279)	(0.216)
B. Partial Linear Model					
ATE	0.385	0.015	0.35	0.416	0.527
se(median)	(0.181)	(0.231)	(0.174)	(0.182)	(0.168)
se	(0.166)	(0.203)	(0.163)	(0.162)	(0.162)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.518	0.458	0.563	0.406	0.179	0.225
se(median)	(0.337)	(0.253)	(NA)	(0.253)	(0.347)	(0.344)
se	(0.296)	(0.252)	(0.198)	(0.232)	(0.314)	(0.343)
B. Partial Linear Model						
ATE	0.524	0.413	0.386	0.321	0.168	0.22
se(median)	(0.168)	(0.165)	(0.199)	(0.217)	(0.285)	(0.293)
se	(0.144)	(0.155)	(0.186)	(0.202)	(0.256)	(0.277)

Note: "Exp" is short for experience years. "ATE" means median treatment effect estimations across splits, here it reports "best" estimation results among Trees, Random Forest and Neural network methods. how "best" are calculated is refered to Chernozhukov et al. (2018). "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. "Splits" means how many times we randomly separate the data into different (pre-setted) folds.

Table 25: Graduate school rank effects on standardized salary (DML)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE	0.164	0.219	0.137	-0.003	0.263
se(median)	(0.371)	(0.209)	(0.541)	(0.317)	(0.205)
se	(0.221)	(0.179)	(0.388)	(0.317)	(0.204)
B. Partial Linear Model					
ATE	0.684	0.546	0.529	0.421	1.068
se(median)	(0.465)	(0.476)	(0.381)	(0.495)	(0.463)
se	(0.43)	(0.404)	(0.38)	(0.356)	(0.444)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.203	0.152	0.465	0.363	0.333	0.155
se(median)	(0.196)	(0.28)	(0.245)	(0.292)	(0.274)	(0.357)
se	(0.187)	(0.26)	(0.211)	(0.29)	(0.263)	(0.355)
B. Partial Linear Model						
ATE	0.923	0.612	0.688	0.214	0.528	0.686
se(median)	(0.506)	(0.438)	(0.554)	(0.553)	(0.406)	(0.593)
se	(0.422)	(0.41)	(0.55)	(0.519)	(0.354)	(0.56)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Estimation on interactive model effects are obtained by creating a dummy variable which is denoted 1 when rank is less or equal to the rank median; 0 if rank is greater than the rank median. Estimation on partial linear model effects are obtained by creating a transformed rank variable, which is calculated by $1 - rank/max(rank)$. In this way, the rank range is between 0 and 1. Better ranked schools are assigned values close to 1, worse ranked schools are assigned values close to 0.

Table 26: Undergraduate ECON effects on standardized salary(DML)

	Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model					
ATE	-0.3	-0.113	-0.475	-0.231	0.196
se(median)	(0.791)	(0.327)	(0.279)	(0.27)	(0.377)
se	(0.791)	(0.235)	(0.246)	(0.258)	(0.304)
B. Partially Linear Model					
ATE	-0.007	0.049	-0.411	-0.338	0.132
se(median)	(0.235)	(0.243)	(0.253)	(0.235)	(0.224)
se	(0.233)	(0.211)	(0.238)	(0.232)	(0.21)

	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	-0.173	-0.132	-0.286	-0.41	-0.431	NA
se(median)	(0.254)	(0.274)	(0.29)	(NA)	(0.543)	(NA)
se	(0.235)	(0.272)	(0.244)	(0.31)	(0.532)	(NA)
B. Partially Linear Model						
ATE	-0.056	-0.075	-0.218	-0.489	-0.333	-0.465
se(median)	(0.23)	(0.177)	(0.22)	(0.29)	(0.307)	(0.492)
se	(0.208)	(0.176)	(0.217)	(0.286)	(0.307)	(0.469)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp8", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

Table 27: Undergraduate STEM effects on standardized salary (DML)

		Exp0	Exp1	Exp2	Exp3	Exp4
A. Interactive Model						
ATE		-0.079	-0.031	0.062	0.198	0.547
se(median)		(0.429)	(0.311)	(0.308)	(0.261)	(0.231)
se		(0.429)	(0.308)	(0.269)	(0.25)	(0.212)
B. Partially Linear Model						
ATE		-0.08	-0.035	0.106	0.239	0.549
se(median)		(0.217)	(0.209)	(0.227)	(0.215)	(0.215)
se		(0.216)	(0.209)	(0.227)	(0.207)	(0.191)
	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
A. Interactive Model						
ATE	0.407	0.366	0.166	0.311	0.249	NA
se(median)	(0.546)	(0.437)	(0.343)	(0.566)	(0.888)	(NA)
se	(0.544)	(0.34)	(0.326)	(0.566)	(0.668)	(NA)
B. Partially Linear Model						
ATE	0.278	0.253	-0.009	0.304	0.321	0.316
se(median)	(0.209)	(0.19)	(0.229)	(0.345)	(0.335)	(0.28)
se	(0.207)	(0.187)	(0.195)	(0.341)	(0.329)	(0.254)

Note: "Exp" is short for experience years. "ATE" reports median average treatment effect across splits; "se(median)" reports median methods adjusted standard errors; "se" reports median standard error across splits. Under "Exp6", because there is not a "best" result reported, what listed is chosen from a medium value of coefficients among all reported machine learning methods. No results reported under "Exp10".

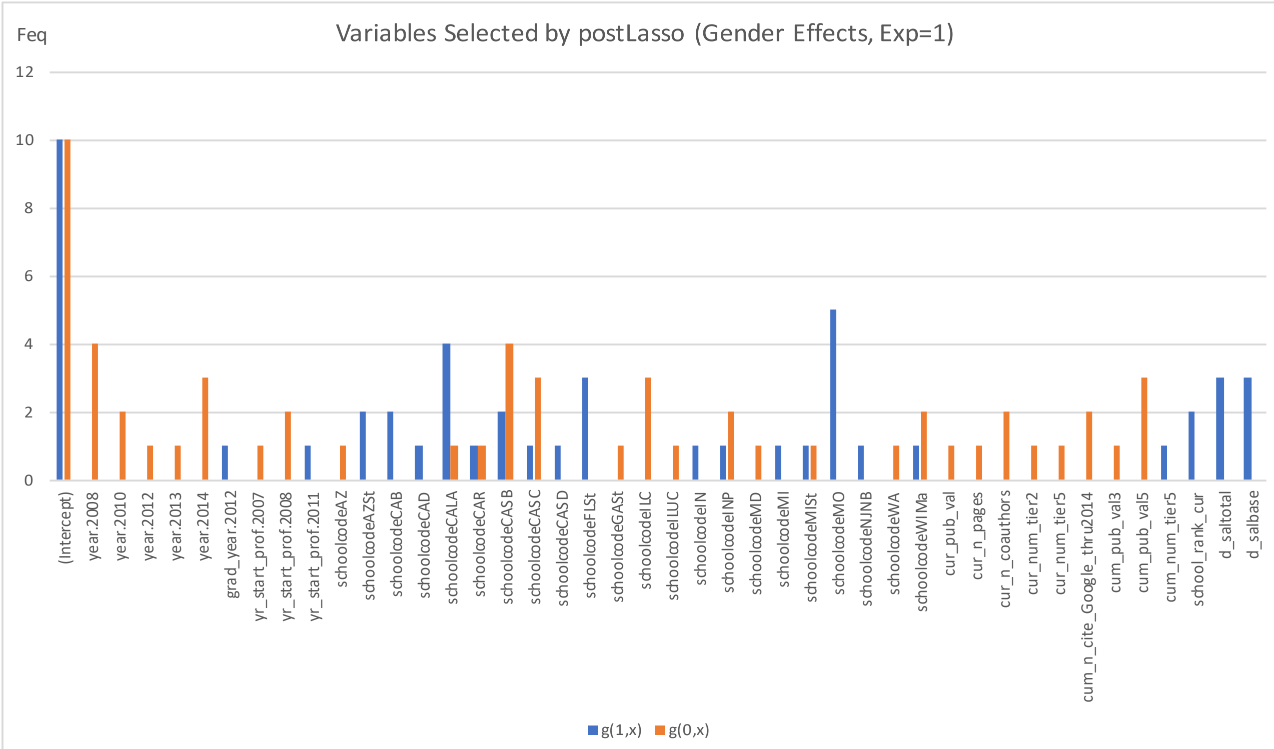

Figure 1: Post Lasso Selected variables in interactive model: Gender Effect, EXP=1

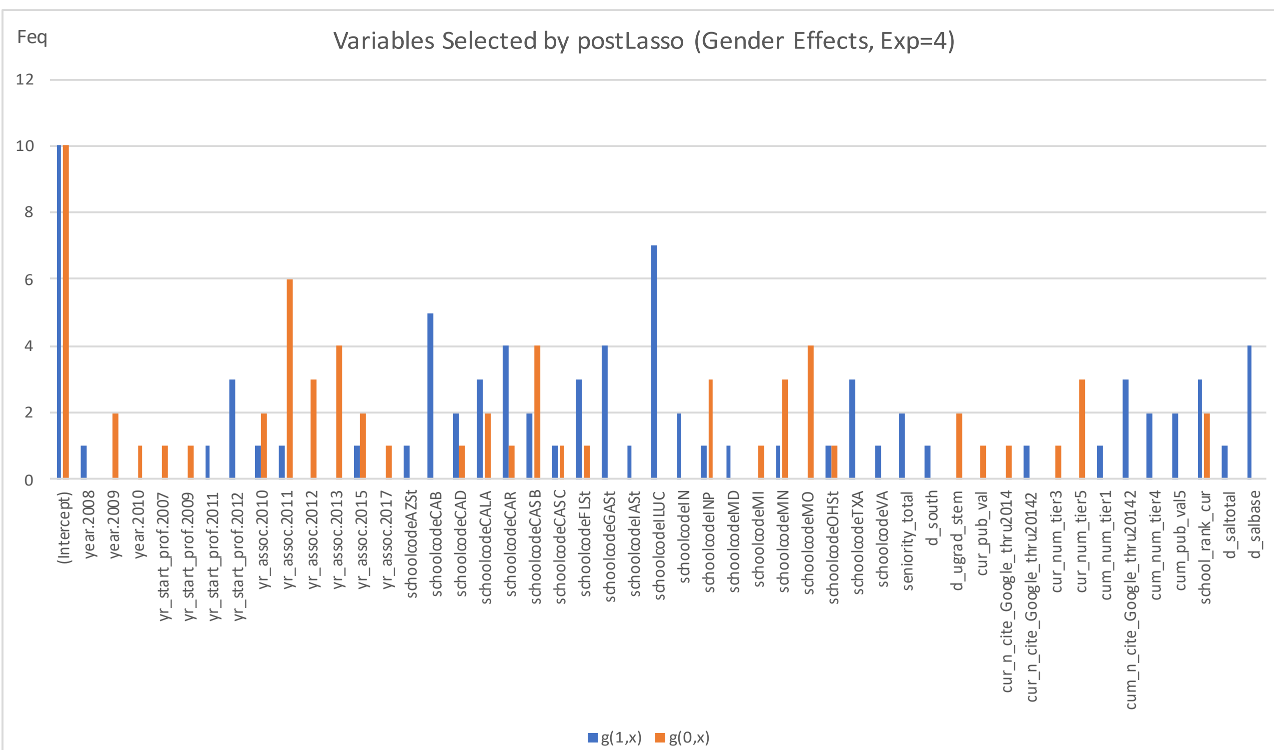

Figure 2: Post Lasso Selected variables in interactive model: Gender Effect, EXP=4

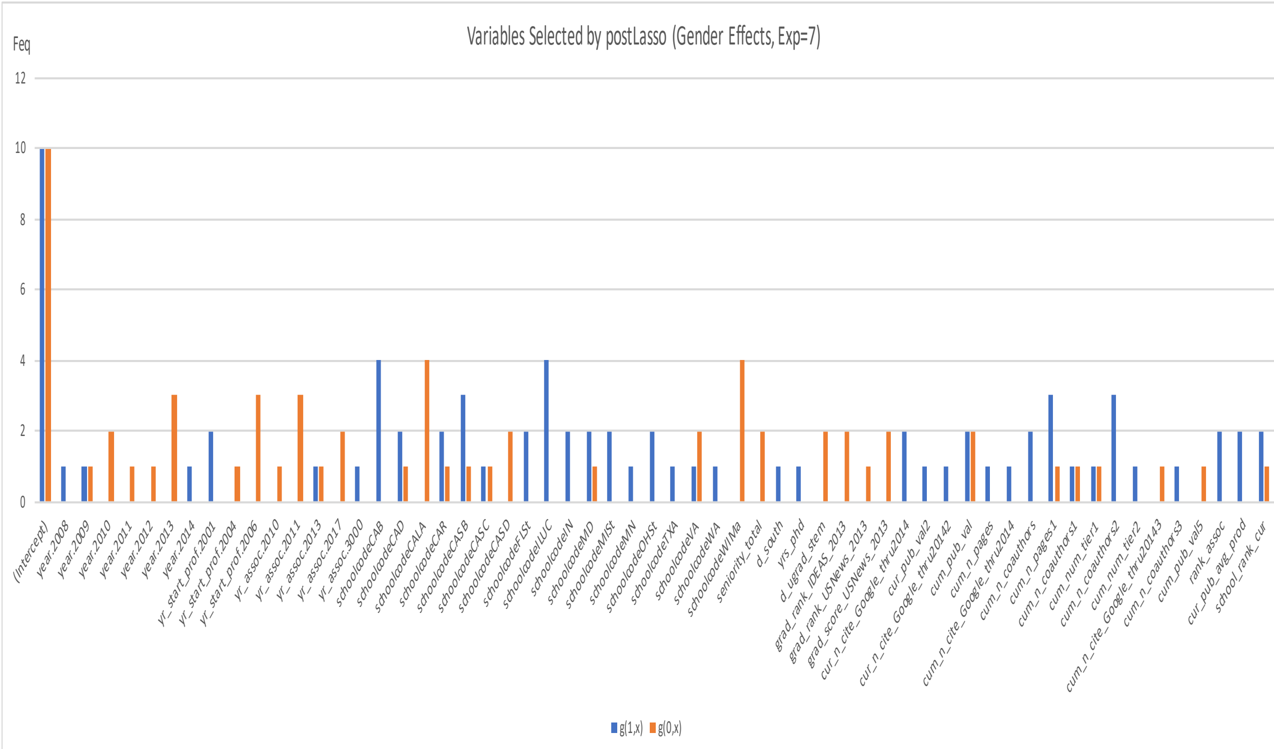

Figure 3: Post Lasso Selected variables in interactive model: Gender Effect, EXP=7

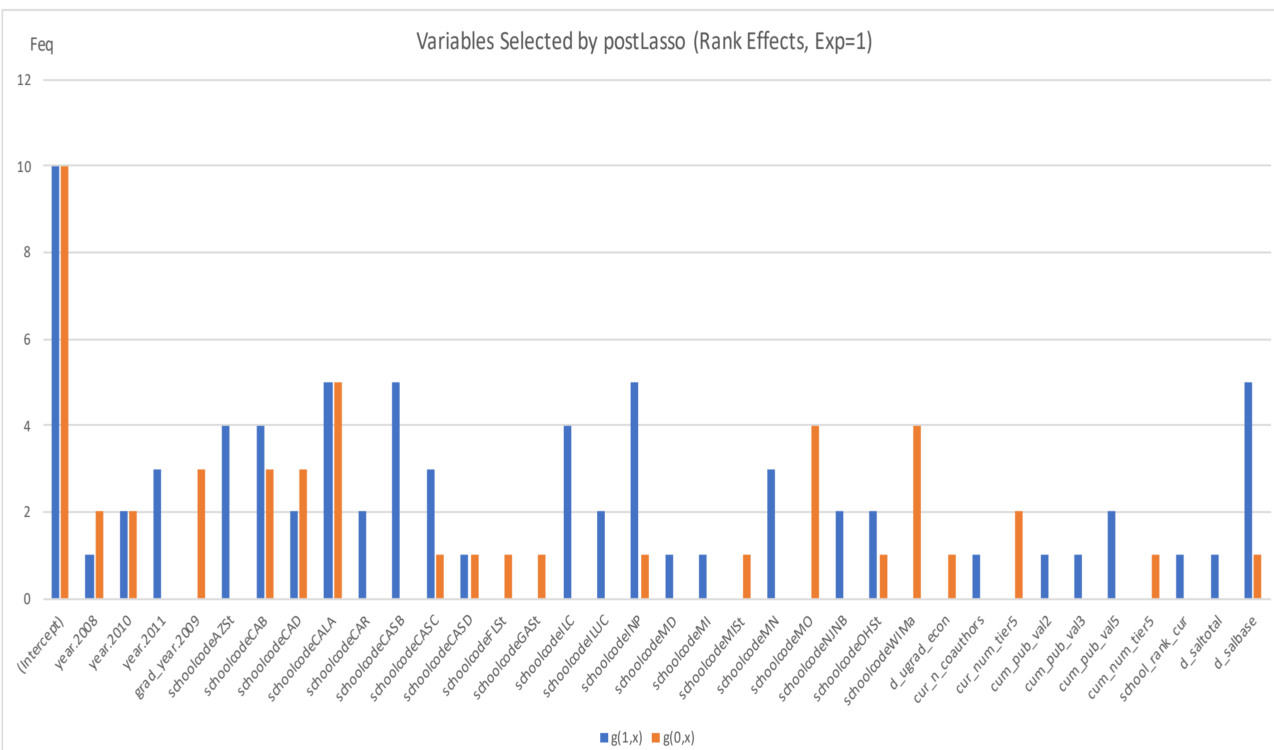

Figure 4: Post Lasso Selected variables in interactive model: Rank Effect, EXP=1

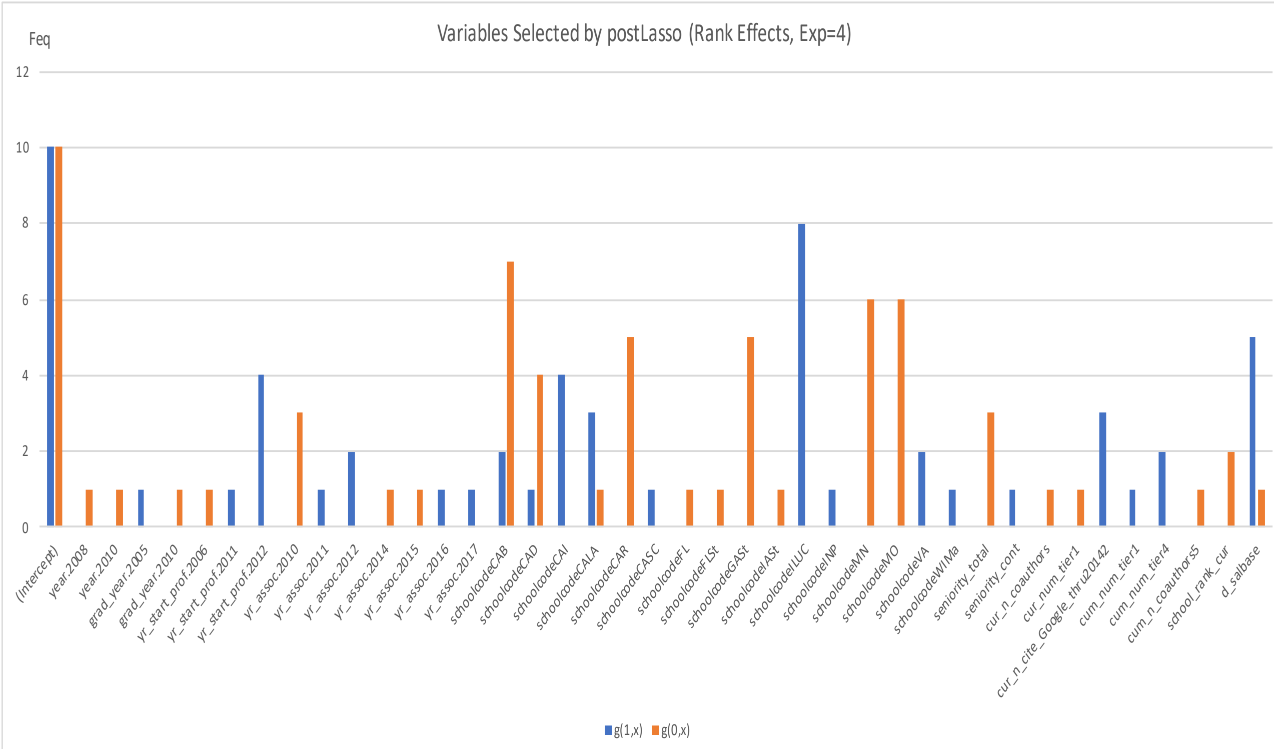

Figure 5: Post Lasso Selected variables in interactive model: Rank Effect, EXP=4

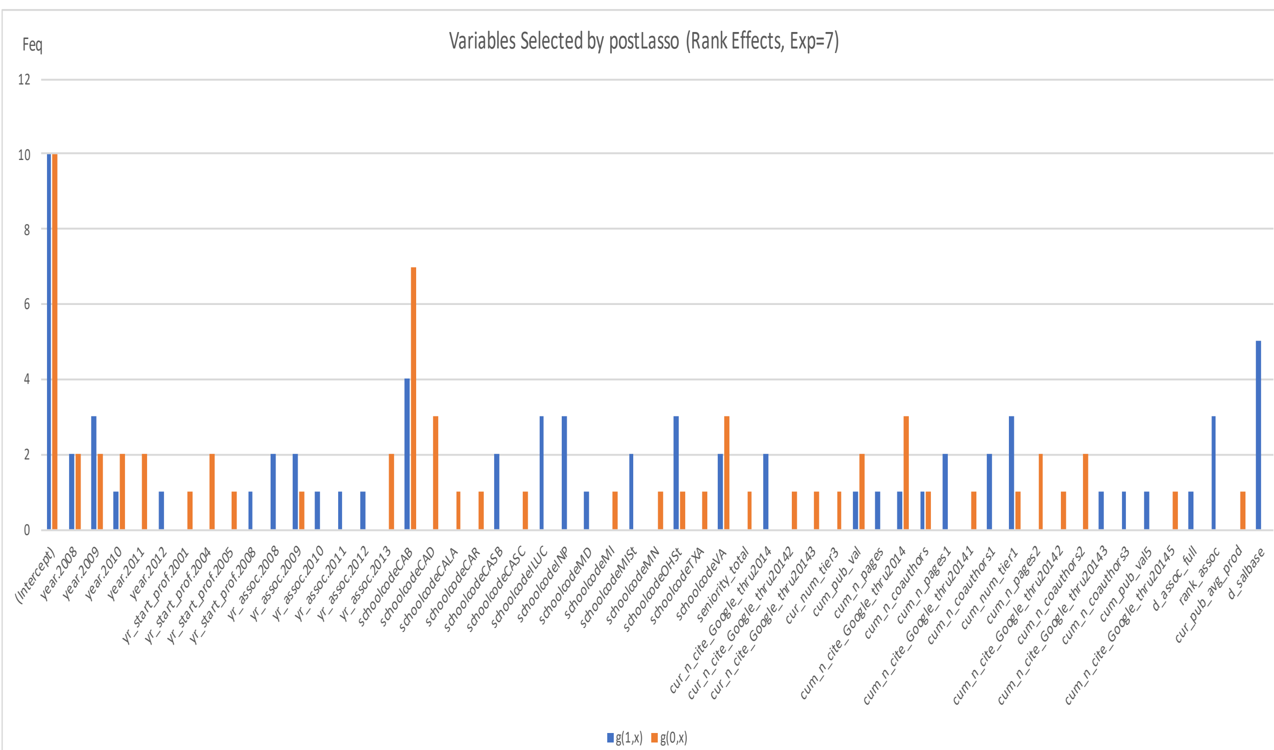

Figure 6: Post Lasso Selected variables in interactive model: Rank Effect, EXP=7

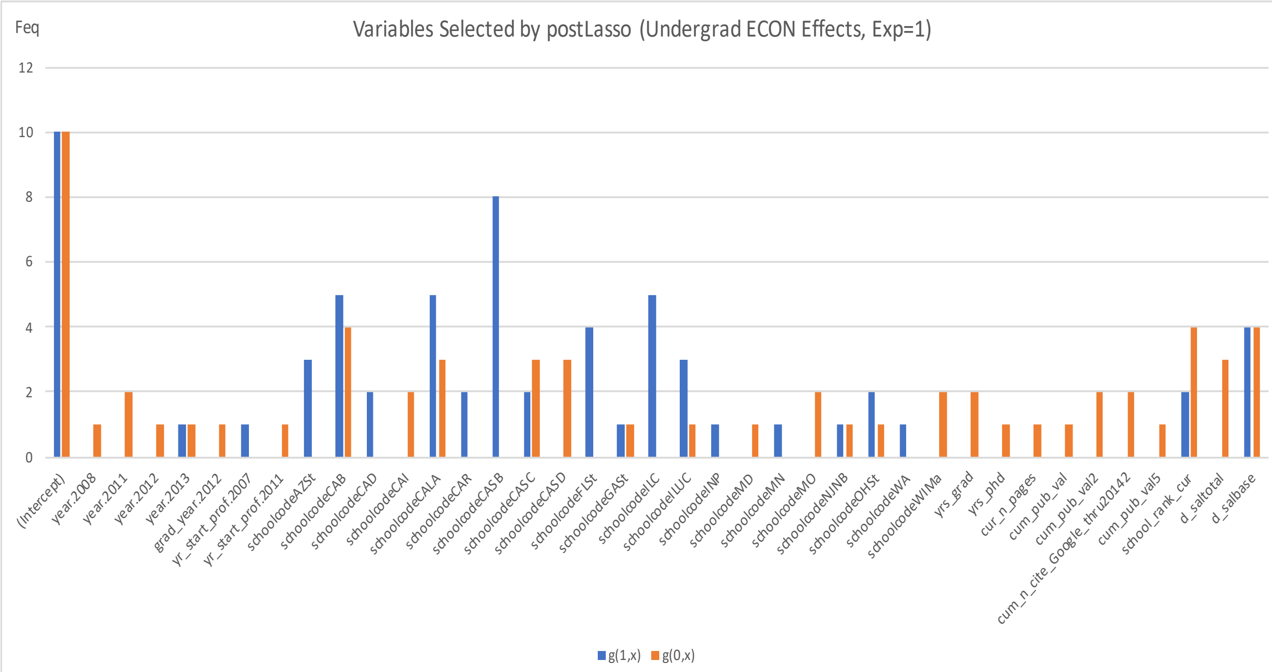

Figure 7: Post Lasso Selected variables in interactive model: Undergraduate ECON Effect, EXP=1

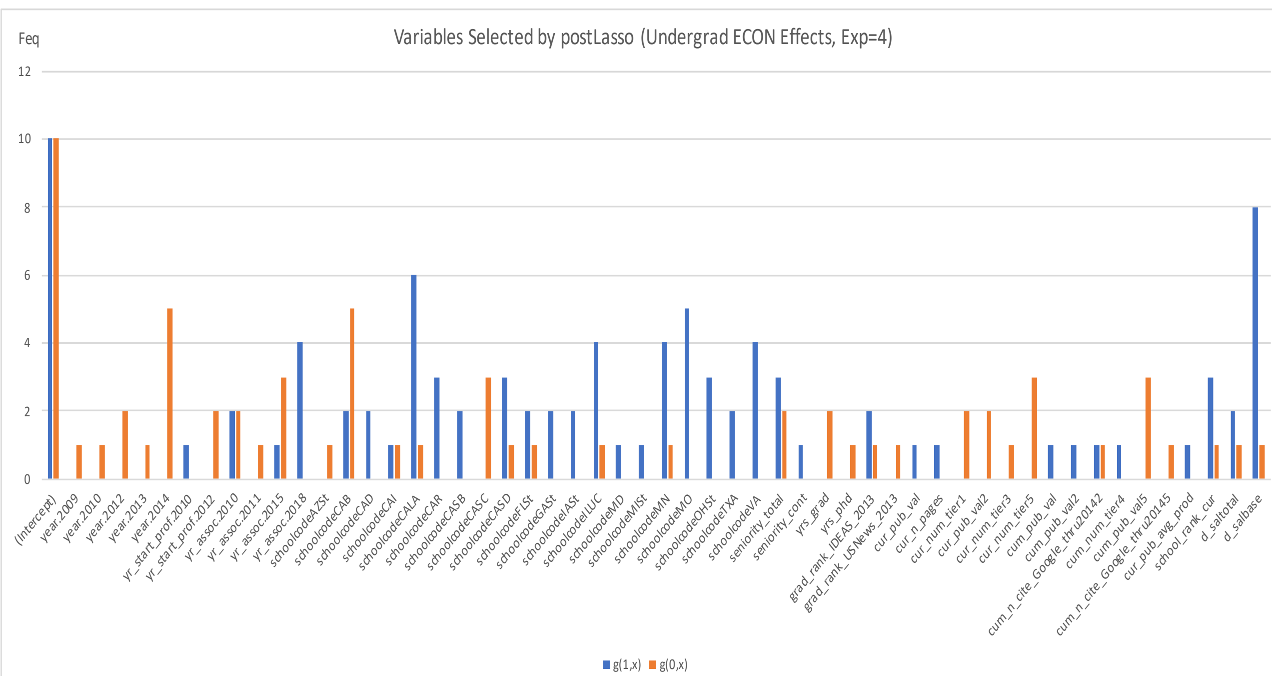

Figure 8: Post Lasso Selected variables in interactive model: Undergraduate ECON Effect, EXP=4

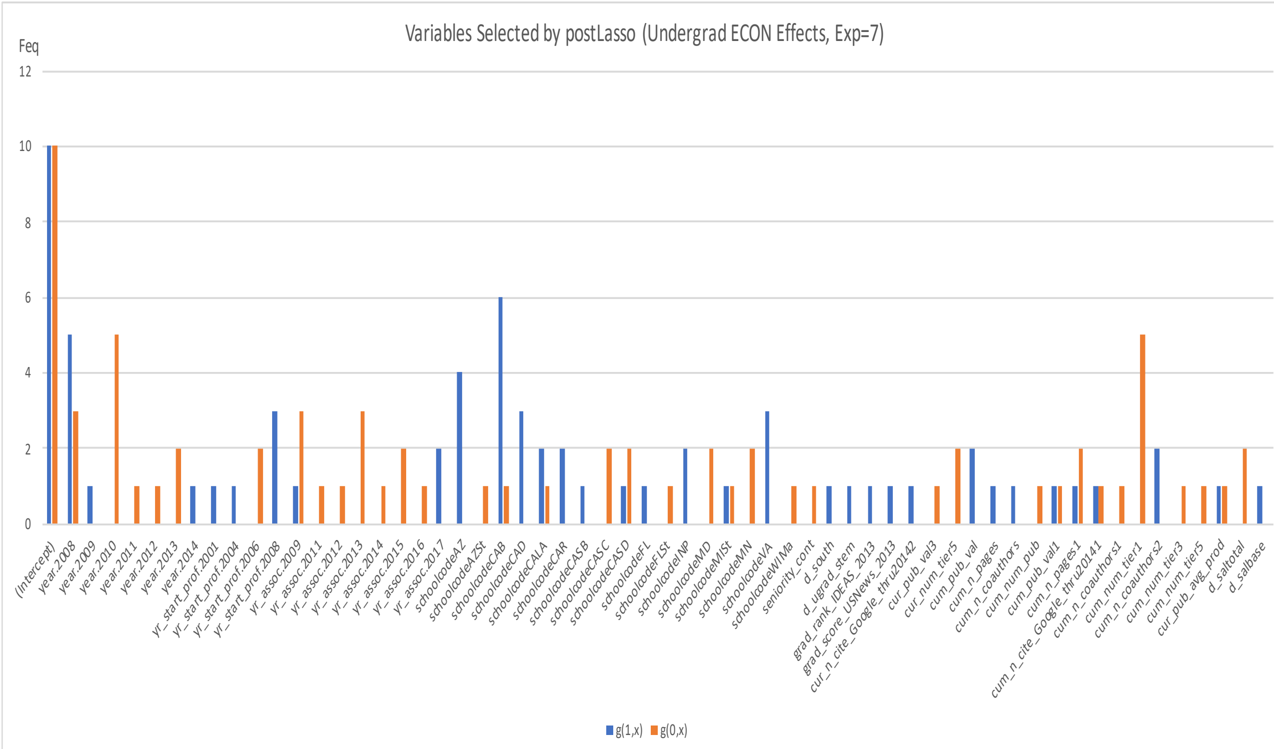

Figure 9: Post Lasso Selected variables in interactive model: Undergraduate ECON Effect, EXP=7

9.2 A short proof of partial linear model consistency

Now let's see how DML estimators being consistent in a partial linear model. We use $\check{\theta}_0$ to represent the consistent DML estimator. The inference on θ_0 relays on the score function defined as in Chernozhukov et al. (2018) equation (4.3):

$$\psi(D, \mathbf{X}; \theta, \eta) := (Y - D\theta - g(\mathbf{X}))(D - m(\mathbf{X})) \tag{9.1}$$

which satisfies the moment condition $\mathbb{E}(VU) = 0$ and the orthogonality condition (4.5). After some algebra, a DML estimator of $\check{\theta}_0$ is given by,

$$\check{\theta}_0 = (\frac{1}{n}\sum_{i\in I} \hat{V}_i D_i)^{-1} \frac{1}{n}\sum_{i\in I} \hat{V}_i (Y_i - \hat{g}_0(X_i)) \tag{9.2}$$

where $\hat{V}$ is the residual from ML estimation of (4.2), $\hat{V} = D - \hat{m}_0(X)$. Note that $\hat{m}_0$ and $\hat{g}_0$ are obtained by auxiliary sample and $\hat{\theta}_0$ is obtained by main sample. The scaled decomposed estimation error is then,

$$\begin{aligned}\sqrt{n}(\check{\theta}_0 - \theta_0) &= (\mathbb{E}\, V^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} V_i U_i \\ &+ (\mathbb{E}\, V^2)^{-1}(\frac{1}{\sqrt{n}}\sum_{i\in I}(m_0(X_i) - \hat{m}_0(X_i))(g_0(X_i) - \hat{g}_0(X_i))) \\ &+ o_p(1)\end{aligned} \tag{9.3}$$

To be consistent with Chernozhukov et al. (2018), let's separate this equation into three parts

$$a^* = (\mathbb{E}\, V^2)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} V_i U_i$$

$$b^* = (\mathbb{E}\, V^2)^{-1}(\frac{1}{\sqrt{n}}\sum_{i\in I}(m_0(X_i) - \hat{m}_0(X_i))(g_0(X_i) - \hat{g}_0(X_i)))$$

$$c^* = o_P(1)$$

The first term converges to a normal distribution under mild condition, $a^* \rightsquigarrow N(0, \Sigma)$. The second term b^* now is determined by estimation errors of $m_0(\mathbf{X})$ and $g_0(\mathbf{X})$. It contains regularization bias from both of them. The converge rate depends on specific Machine Learning methods used (usually is slower than a square root rate). Scornet, Biau, and Vert (2015) prove the consistency of Random Forests. Biau (2012) show that the converge rate of Random Forest estimators depends on its strong features, the rate order has a form of $n^{\frac{-0.75}{Slog2+0.75}}$ where S is a subset of features. The converge rate is $n^{\frac{-p}{2p+d}}$, where d is the dimension of the raw explanatory variables, and p is the assumed degree of smoothness of the CEF (like number of derivatives). We can control the bound by choosing proper p's for any given d. We now know $m_0(\mathbf{X})$ and $g_0(\mathbf{X})$ are estimated with a slower converge rate to their true value, but the product of the two makes the whole term converge within a vanishing upper bound. This upper bound is $\sqrt{n}n^{-(\varphi_m+\varphi_g)}$, where φ_m is the converge rate of $\hat{m}_0$ to m_0; φ_g is the converge rate of $\hat{g}_0$ to g_0. Thus b^* vanishes eventually if $\varphi_m + \varphi_g > 1/2$.[36] Another requirement for $\check{\theta}_0$ to be consistent is to control the remainder items in c^* and make sure $c^* = o_P(1)$. In partial linear model, terms like

$$\frac{1}{\sqrt{n}}\sum_{i\in I} V_i(\hat{g}_0(\mathbf{X}_i) - g_0(X_i)) \tag{9.4}$$

are included in c^*. Without sample splitting, model error terms V_i and estimation errors $\hat{g}_0(\mathbf{X}_i) - g_0(\mathbf{X}_i)$ are generally related. The reason is that in estimating $\hat{g}_0$ information contained in observation i has already been used, whereas V_i also has information from observation i, the relation between them will cause poor performance of c^*. Conditional on

[36]Chernozhukov et al. (2018) made this claim in their paper. They assume if the specific machine learning method used in the model has this property, then b^* will converge properly. They also show good simulation results. In practice it's hard to find theoretical justifications for how each method converge, but here we use Random Forest and there are papers as mentioned earlier showing its consistency. I strongly believe DML outperforms naively picking some variables and running a simple regression.

auxiliary sample and with $\mathbb{E}(V_i|\mathbf{X}_i) = 0$, equation (9.4) has mean zero and variance with order $\frac{1}{n}\sum_{i\in I}(\hat{g}_0(\mathbf{X}_i) - g_0(\mathbf{X}_i) \rightarrow_P 0$.

References

Altonji, J. G., 2005. Employer learning, statistical discrimination and occupational attainment. American Economic Review 95 (2), 112–117.

Altonji, J. G., Pierret, C. R., 2001. Employer learning and statistical discrimination. Quarterly Journal of Economics 116 (1), 313–350.

Barbezat, D. A., 1987b. Salary differentials by sex in the academic labor market. The Journal of Human Resources 22 (3), 422–428.

Barbezat, D. A., Donihue, M. R., 1998. Do faculty salaries rise with job seniority? Economics Letters 58 (2), 239–244.

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19 (2), 521–547.

Biau, G., 2012. Analysis of a random forests model. Journal of Machine Learning Research 2012 (13), 1063–1095.

Blackaby, D., Booth, A. L., Frank, J., 2005. Outside offers and the gender pay gap: Empirical evidence from the uk academic labour market. The Economic Journal 115 (501), F81–F107. URL `https://www.jstor.org/stable/3590464`

Bratsberg, B., Ragan, Jr., J. F., Warren, J. T., 2003. Negative returns to seniority: New evidence in academic markets. Industrial and Labor Relations Review 56 (2), 306–323.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., 2017. Double/debiased/neyman machine learning of treatment effects. American Economic Review 107 (5), 261–65.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econometrics Journal 21 (1), C1–C68, arXiv:1608.00060.
URL `https://arxiv.org/abs/1608.00060`

Claypool, V. H., Janssen, B. D., Kim, D., Mitchell, S. M., 2017. Determinants of salary dispersion among political science faculty: The differential effects of where you work (institutional characteristics) and what you do (negotiate and publish). Political Science and Politics 50 (1), 146–156.

Ehrenberg, R. G., 2002. Studying ourselves: The academic labor market. Tech. rep., National Bureau of Economic Research.

Ehrenberg, R. G., Pieper, P. J., Willis, R. A., 1998. Do economics departments with lower tenure probabilities pay higher faculty salaries? Review of Economics and Statistics 80 (4), 503–512.

Formby, J. P., Hoover, G. A., 2002. Salary determinants of entry-level academic economists and the charactertics of those hired on the tenure track. Eastern Economic Journal 28 (4), 509–522.

Gerhart, B. A., Rynes, S. L., 1991. Determinants and consequences of salary negotiations by male and female mba graduates. Journal of Applied Psychology 76 (2), 256–262.

Gibson, J., Anderson, D. L., Tressler, J., 2017. Citations or journal quality: Which is rewarded more in the academic labor market? Economic Inquiry 55 (4), 1945–1965.

Ginther, D. K., Hayes, K. J., 2003. Gender differences in salary and promotion for faculty in the humanities 1977-95. The Journal of Human Resources 38 (1), 34–73.

Graves, P. E., Marchand, J. R., Thompson, R., 1982. Economics departmental rankings: Research incentives, constraints and efficiency. The American Economic Review 72 (5), 1131–1141.
URL https://www.jstor.org/stable/1812028

Grove, W. A., Wu, S., 2007. The search for economics talent: doctoral completion and research productivity. American Economic Review 97 (2).

Hamermesh, D. S., Johnson, G. E., Weisbrod, B. A., 1982. Scholarship, citations and salaries: Economic rewards in economics. Southern Economic Journal 49 (2), 472–481.

Johnson, G. E., Stafford, F. P., 1974. The earnings and promotion of women faculty. American Economic Review 64 (6), 888–903.

Kahn, L. B., Lange, F., 2014. Employer learning, productivity, and the earnings distribution: Evidence from performance measures. Review of Economic Studies 81 (4), 1575–1613.

Katz, D. A., 1973. Faculty salaries, promotions, and productivity at a large university. American Economic Review 63 (3), 469–477.

Li, D., Koedel, C., 2017. Representation and salary gaps by race-ethnicity and gender at selective public universities. Educational Researcher 46 (7), 343–354.

Liebowitz, S. J., Palmer, J. P., 1984. Assessing the relative impacts of economics journals. Journal of Economic Literature 22 (1), 77–88.

Megdal, S. B., Ransom, M. R., 1985. Longitudinal changes in salary at a large public university: what response to equal pay legislation? American Economic Review 75 (2), 271–274.

Mitchell, S. M., Hesli, V. L., 2013. Women don't ask? women don't say no? bargaining and service in the political science profession. Political Science Politics 46 (2013), 355–369.

Moore, W. J., Newman, R. J., Turnbull, G. K., 1998. Do academic salaries decline with seniority? Journal of Labor Economics 16 (2), 352–366.

Perna, L. W., 2001. Sex differences in faculty salaries: A cohort analysis. The Review of Higher Education 24 (3), 283–307.

Sax, L. J., Hagedorn, L. S., Arredondo, M., Dicrisi, F. A., 2002. Faculty research productivity: Exploring the role of gender and family-related factors. Research in Higher Education 43 (4), 423–446.

Scornet, E., Biau, G., Vert, J.-P., 2015. Consistency of random forests. The Annals of Statistics, Institute of Mathematical Statistics 43 (4), 1716–1741.

Smart, J. C., 1991. Gender equity in academic rank and salary. The Review of Higher Education 14 (4), 511–525.

Stock, W. A., Alston, R. M., 2000. Effect of graduate program rank on success in the job market. The Journal of Economic Education 31 (4), 389–401.

Tibshirani, R. J., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 58 (1), 267–288.

Toutkoushian, R. K., Bellas, M. L., Moore, J. V., 2007. The interaction effects of gender, race, and marital status on faculty salaries. The Journal of Higher Education 78 (5), 572–601.

Toutkoushian, R. K., Conley, V. M., 2005. Progress for women in academe, yet inequities

persist: Evidence from nsopf: 99. Research in Higher Education 46 (1), 1–28.
URL http://www.jstor.org/stable/40197383

Toutkoushian, R. K., Hoffman, E. P., 2002. Alternatives for measuring the unexplained wage gap. New Directions for Institutional Research, 71–89.

Wooldridge, J. M., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd Edition. MIT Press.

Yang, L., Webber, K. L., 2015. A decade beyond the doctorate: the influence of a us post-doctoral appointment on faculty career, productivity, and salary. Higher Education 70 (4), 667–687.

Zhang, L., 2005. Advance to graduate education: the effect of college quality and undergraduate majors. The Review of Higher Education 28 (3), 313–338.