Data Appendix for "Will The University Endowment Tax Slow Scientific Progress? Evidence from Elite Economics PhD Programs"

Joshua Angrist, Marc Diederichs, and Glenn Ellison*
with Melinda Liu

May 2026

*e-mail: angrist@mit.edu, marc.diederichs@uni-passau.de, and gellison@mit.edu

Contents

1 Data Set Construction

This paper uses data from five sources, starting with a sample of top-ranked MIT applicants.

1.1 The MIT Applicant Sample

Our sample of MIT Economics Ph.D. applicants includes applicant cohorts from 1996–1997, 2005–2010, and 2012–2019, for a total of 3766 applicants. Data for omitted cohorts are lost or incomplete. Applicant records include application year, GRE scores, and admissions committee ratings, from which we generate a cohort-specific applicant rank. In most years, the dataset includes the 190–200 highest-rated applicants per cohort. In some years, fewer applicants were advanced to the next round; data for some applicants in a few years are missing.

One or two faculty members initially evaluated all applicants, with additional readers assessing high-rank files; grades were typically assigned on a 12-point scale. We work with applicants' within-cohort rank, computed by averaging all available scores and assigning tied applicants the mean rank. These ranks summarize independent evaluations from the initial admissions stage and do not perfectly predict final admission or funding decisions.

1.2 School and Graduation Year Information

For each applicant completing a PhD, we aim to identify the degree-granting institution. This information comes from EconLit, the data set compiled for Angrist and Diederichs (2024), and manual searches of department websites and applicant CVs.

December issues of the Journal of Economic Literature (JEL) list economics doctorates awarded by U.S. and Canadian institutions in the previous academic year. We obtained these data from EconLit, the AEA's online bibliographic database. The JEL graduate list reports each graduate's name, degree-granting institution and year, dissertation title, and a JEL subject code for the thesis. Economics or other department staff typically submit degree lists to the JEL; submissions that miss the print deadline nevertheless appear in digital EconLit. We merge applicants to EconLit dissertation records based on student names and an expected graduation window of four to nine years after the application year, matching to EconLit dissertations from 1999–2024.

EconLit misses some dissertations. Most MIT applicants who complete a PhD do so at one of the elite eight. For graduates of these schools, school information is augmented with the Angrist and Diederichs (2024) data set, which sources information from ProQuest, administrative records provided by department heads, and MIT DSpace, as well as EconLit.

For applicants for whom EconLit and the citeangrist2024dissertation data contain no school or degree year information, we search department websites and applicant CVs. This process recovers school information for students who have not yet graduated or who graduated too recently to appear in the EconLit dissertation data through 2024. In addition to economics PhDs, we identified schools for applicants earning PhDs in economics-adjacent fields such as finance, business, or public policy. This additional information coomes from online CVs and websites.

1.3 Applicant Publications

The research output of students is measured by counting articles in EconLit. Our copy of EconLit lists more than one million journal articles in nearly 2000 economic and business outlets, around 300,000 collective volume articles such as the Handbook of Labor Economics, and includes variables indicating author names and affiliations.

We merge students in our sample to the EconLit author variable based on author last name and first name initial. This allows us to include all publications even if the first name is abbreviated. For example, most of James Heckman's publications are listed under "Heckman, James" but some are listed under

"Heckman, J". Common last names are disambiguated using first names, middle names (where available), and the affiliation indicated on EconLit publication records. Publication counts exclude collective volume articles that appear under the same title as a journal article by the same author.

Common names sometimes generate implausible publication records. These are flagged when applicants have more than one publication prior to application, appear to have many affiliation changes across publications, fall in the top 3 for total EconLit publications in their cohort with no top five publications, and have very common names. Flagged cases are validated by comparing EconLit with CVs found online.

Some students appear under different names in MIT applicant records and EconLit, resulting in a false attribution of zero publications in EconLit. As a check on this, we search CVs and other online sources for students with no EconLit publications. This process disambiguated names for 126 applicants. These cases are typically one of the following:

1. The applicant has multiple last names, and the combination of names in our dataset differs from that in EconLit. For instance, "Benitez, Hugo A" in our dataset equals "Benitez Silva, Hugo" in EconLit because the full name is Hugo A. Benitez-Silva.

2. The student goes by a middle name. For instance, "Tenreyro, Maria S" in our dataset equals "Tenreyro, Silvana" in EconLit. Online searches show that this applicant's full name is Maria Silvana Tenreyro.

3. One or two letters of the name differ across sources, but there is reason to suspect a match (i.e. same last name, with an application year matching the online profile).

4. The applicant has an extra middle or last name in EconLit. For instance, "Villaverde, Jesus" in our dataset but "Fernandez Villaverde, Jesus" in EconLit, while the full name is Jesus Fernandez-Villaverde).

5. The applicant goes by a nickname. Example: "Andrikogiannopoulou, Anastasia A" in our dataset equals "Andrikogiannopoulou, Angie" in EconLit. This is resolved by an applicant CV for Anastasia (Angie) Andrikogiannopoulou.

1.4 NSF Funding Data

The NSF Graduate Research Fellowship Program (GRFP) website provides annual lists of applicants who were offered awards and applicants receiving an honorable mention. We downloaded award lists from 1994-2025 and used these list to determine which MIT applicants won an NSF. Our NSF award list is limited to applicants in economics and adjacent fields of study (specifically, fields including the terms "econ", "math", "statistics", "computer science", "comp/IS/eng", "data science", and "business").

1.5 Applicant Gender

MIT applicant records identify gender for 89% of students. This leaves 412 students with missing gender information.

For students without gender data in the admissions records, gender is inferred using first-name frequencies from the U.S. Social Security Administration registry. This procedure assigns gender to an additional 354 students (9% of the sample).

For remaining unclassified applicants, we determine whether the relevant name is more common among men or women as represented in online sources.

26.3% of applicants in our sample are classified as female.

1.6 Merging Sources

After downloading, cleaning, and formatting the data, sources are merged as follows. Applicant data for 2,587 applicants from 1996-2019 are merged with a file recording MIT admissions status. School and year are then added, followed by EconLit research output. Finally, NSF award status is added and gender classified to complete the match.

2 Journal Lists and Weights

EconLit indexes nearly 2,000 journals, covering the entire spectrum of economics outlets as well as many from adjacent fields. Econlit also includes articles from collective volumes like Elsevier handbooks and the NBER Macro Annual. To avoid overcounting publications by journal editors, we exclude forewords, editorial comments, and editorial errata.

Publication information is collated for three sets of journals. The first is the full EconLit journal list, which changes over time and contains many rarely-cited publications. The second is based on a list generated for the (Angrist et al., 2020) study of the impact of economics scholarship on other disciplines, referred to here as the "Deep Impact" (DI) journal list after the paper's title. For each discipline studied, Angrist et al. (2020) selects a "trunk journal" (e.g., the *American Economic Review* for economics) and retains the fifty journals most cited by the trunk journal in any decade from 1970 to 2015 (non-economics discipline lists tend to be shorter than 50). The DI list used here combines discipline-specific lists for economics and related fields such as accounting, finance, management, marketing, multidisciplinary science, operations research, political science, sociology, and statistics. To this, we add the following relatively new journals: AEJ Microeconomics, Theoretical Economics, Quantitative Economics, JPE: Microeconomics, and JPE: Macroeconomics. The updated DI list contains 137 journals.

The third list consists of the usual top five journals: the QJE, AER, JPE, Econometrica, and the Review of Economic Studies. The AEA Papers And Proceedings was published as the May AER from 1911 to 2017. We remove P&P publications from the top five using AER issue numbers for this period.

Our weighted DI outcome applies citation-based journal influence weights to DI publications and gives each author $1/n$ credit for an n-authored paper. The journal influence weights used here are similar to those used for Angrist et al. (2020); these weights are proportional to the frequency with which articles in each journal are cited by top five journals. More precisely, influence weights are a weighted average of the ratio of the fraction of each top five journal's recent citations that go to each journal divided by the number of recent articles in the cited journal.

The weights that each top five journal gets in this count come from a page-rank style formula. We use influence weights computed for 2015 (weights are mostly stable over time). Weights for new second-tier journals (AEJ: Microeconomics, AER: Insights, JPE: Macroeconomics, JPE: Microeconomics, Quantitative Economics, and Theoretical Economics) are set to the average for AEJ: Applied Economics, AEJ: Policy, and AEJ: Macroeconomics. The NBER Macroeconomics Annual, which does not have a 2015 weight, is assigned its 2014 citation-based weight. A weight for the AEA Papers and Proceedings weight is derived from the regular AER weight using an adjustment based on the ratio of pages per article in the two journals. Journal influence weights are normalized so that the average article in a top five journal gets a weight of one.

The calculation of weights uses Web of Science data on citation counts, organized by ISSN. Weights are matched to Econlit by journal ISSN. The exceptions are *IMF Staff Papers*, the *NBER Macroeconomics Annual*, and the *Journal of Economic and Social Measurement*, which are found by journal name. Our weighted DI publication measure allocates each author on coauthored publications $\frac{1}{n}$ credit for an n-authored publication. For instance, for an applicant with two publications in top five journals, one with three authors and one with two authors, the coauthorship-adjusted count is $\frac{1}{3} + \frac{1}{2} = \frac{5}{6}$.

References

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2020). Inside job or deep impact? extramural citations and the influence of economic scholarship. *Journal of Economic Literature*, 58(1):3–52.

Angrist, J. and Diederichs, M. (2024). Dissertation paths: Advisors and students in the economics research production function. Working Paper 33281, National Bureau of Economic Research.