

Analyzing the Diversity of Invited Speakers at Economic Seminars from 2014 to 2019*

Replication of Jennifer L. Doleac, Erin Hengel, and Elizabeth Pancotti's
'Diversity in Economics Seminars: Who Gives Invited Talks?'

Sima Shmuylovich

February 15, 2024

This paper investigates the diversity of speakers in economic seminars from 2014 to 2019, looking at gender identity and whether or not a speaker belongs to an ethnically unrepresented minority. 1) what was done, 2) what was found, and 3) why this matters (all at a high level).

Table of contents

1	Introduction	1
2	Data	3
2.1	Raw Data	3
2.2	Cleaned Data	4
2.3	Measurement	5
3	Results	5
	References	6

1 Introduction

The Importance of Analyzing Diversity of Invited Speakers at Economic Seminars
In the landscape of academia, the need to foster diversity and inclusion has emerged as a central pillar in the pursuit of equity, scholarly excellence, and societal relevance.

*Code and data are available at: <https://github.com/sshmuylovich/diversity-in-economic-seminars.git>; Replication on Social Science Reproduction platform available at: ...

Firstly, the commitment to diversity and inclusion within academic settings mirrors broader societal values that emphasize fairness, equity, and justice. By scrutinizing the composition of invited speakers at economic seminars, academic institutions can assess their alignment with these values, ensuring they do not perpetuate historical biases or exclusionary practices. This reflective process is crucial for fostering an academic culture that not only respects but actively promotes diversity.

Secondly, diversity of seminar speakers significantly enriches academic discourse. Diversity of perspective introduces new viewpoints, methodologies, and inquiries all of which are instrumental in driving innovative, robust, and comprehensive research outcomes. The exchange of diverse ideas catalyzes critical thinking and creativity, which are essential for the advancement of knowledge and the solving of complex economic problems.

Thirdly, when individuals from underrepresented groups see successful figures who resemble them, it reinforces the belief that academic and research success is attainable for everyone, irrespective of their gender, race, or geographic origin. This aspect of diversity analysis is paramount for cultivating an inclusive academic environment where all members feel valued and motivated.

Invited seminar talks are a primary way that economists: 1) get feedback on their work, 2) disseminate their research, and 3) build their professional networks. Not only are there benefits to the faculties who extend the invitations, but those invitations also have positive externalities to the speaker. Current gender and racial disparities in the field of economics may be exacerbated by disparities in who is invited to give seminar talks. Thus far, data limitations have prevented an informed discussion of this possibility. This paper aims to provide descriptive evidence on who gives invited seminar talks in economics, based on an extensive, ongoing data collection effort conducted by Jennifer L. Doleac, Erin Hengel, and Elizabeth Pancotti (Jennifer L. Doleac and Pancotti 2021).

This paper presents an analysis of the diversity of invited speakers at economic seminars across a range of universities over a six-year period, from 2014 to 2019. Specifically, the paper focused on three critical dimensions of diversity: gender, whether speakers are from underrepresented minorities (URMs), and the geographic region of the hosting institutions.

The structure of this paper comprises three sections: Data, Results, and Discussion. In the Data section (Section 2), we discuss the data source and the process of measuring and cleaning the datasets from the replicated paper (Jennifer L. Doleac and Pancotti 2021). The Results section (Section 3) summarizes the findings, presenting graphs to illustrate trends. The Discussion section (**Discussion**) delves into the analysis of our findings, exploring Additionally, we discuss limitations and propose avenues for further research, concluding the paper.

2 Data

The dataset used in this study is sourced from the replication package linked to Doleac, Hengel, and Pancotti’s paper (Jennifer L. Doleac and Pancotti 2021). Using the R programming language (R Core Team 2023), data cleaning and analysis procedures were performed, making use of various R libraries and packages including tidyverse (Wickham et al. 2019), janitor (Firke 2023), ggplot2 (Wickham 2016), knitr (Xie 2023), readr (Wickham, Hester, and Bryan 2023), and dplyr (Wickham et al. 2023).

The data cleaning and analysis procedures were replicated from the original paper to maintain consistency in the datasets used. The primary aim of this analysis is to evaluate diversity of economics seminar speakers, with a particular focus on gender and URM status and across universities in the United States, identifying demographic, geographical, and temporal trends.

The following sections will offer a thorough examination of the raw data in (Section 2.1), explore the data cleansing process in (Section 2.2), and then delve into the measurement considerations in (Section 2.3).

2.1 Raw Data

Doleac, Hengel, and Pancotti developed a list of economics and economics-adjacent departments in the United States and abroad, using a variety of sources. As a part of their data collection process, Doleac, Hengel, and Pancotti webscraped information on seminar schedules from university department websites (Jennifer L. Doleac and Pancotti 2021). Their raw data was then stored in their replication package as a csv file, ‘data.csv’ and acts as this study’s raw dataset.

Doleac, Hengel, and Pancotti recorded the demographic characteristics (gender and URM status) of the seminar speakers based on information listed on their website or CV and the location of their undergraduate institution. In their paper, they acknowledge that these are imperfect proxies for actual gender and URM status, and are best interpreted as indicating how that person is perceived by others. Furthermore, Doleac, Hengel, and Pancotti interpreted URM status as whether someone grew up as an underrepresented minority, facing the various disadvantages that URM status implies (fewer role models in economics who look like you, and bias from peers, educators, and supervisors). In their paper, they used the following definition of URM status: black, Latinx, or Native American, and grew up in U.S (Jennifer L. Doleac and Pancotti 2021).

Presented in Table 1 are the first six speakers that appear in the raw data. The subsequent section (Section 2.2) will show the data cleaning process, bringing clarity to the structure of the refined dataset used for analysis.

Table 1: Raw data showing the first 6 invited speakers

school	id	semester	urm_status	years_from_phd	repec_owndept	repec_semdept	phd_program	census_region
5b50fcc8154085b41a73d57f820251e7	88ac1392e9e0c2ac81e1a9d72fc648c	2018-01-01	Non-URM female	18	11-25	11-25	1	NA
5b50fcc8154085b41a73d57f820251e7	116599d35a0fad4843502ef61d5b24d5	2019-01-01	Non-URM male	19	26-50	11-25	1	NA
5b50fcc8154085b41a73d57f820251e7	05ac062b03fcfb983d8c9b0c59e78e7	2014-08-01	Non-URM female	12	1-10	11-25	1	NA
060ed1d6669f9354439f02f85fb02035	2f89c7cd6227ce766a38d435d8a053d5	2018-08-01	Non-URM male	3	101-289	101-289	1	NA
060ed1d6669f9354439f02f85fb02035	1aba91f655e4a122d1f44691c51cd354	2016-01-01	Non-URM female	18	11-25	101-289	1	NA
060ed1d6669f9354439f02f85fb02035	5319a62ec563cb5f4d95259fabfc9094	2017-01-01	Non-URM male	0	11-25	101-289	1	NA

2.2 Cleaned Data

While Doleac, Hengel, and Pancotti restricted their sample to include only talks by scholars who had a Ph.D. at the time (excluding graduate students) and external speakers (excluding speakers from the same institution as the host department), this study does. Nor does this study restrict its sample to talks by scholars who are economists (defined by Doleac, Hengel, and Pancotti as having an economics or economics-adjacent Ph.D., or being affiliated with an economics department).

The dataset has been refined to include only five variables: Seminar ID, Semester, URM Status, Gender, and Geographical Region of Host Department. Cleaning splitting the raw Demographic Variable into URM Status and Gender and replacing NA values in the geographical region column with “International” as explained in Doleac, Hengel, and Pancotti’s data collection process.

Presented in Table 2 are the first six speakers along with the semester the seminar took place in, the speaker’s URM status, the speaker’s gender, and the geographical region of the seminar’s host department. I have excluded the seminar ID the speaker’s talk was a part of from the table because it is long and unreadable. However, it is important to note that it does exist in the actual cleaned data.

Table 2: Cleaned Data showing the first attributed of the first 6 economics seminar speakers

seminar_id	Semester	URM Status	Gender	demographic	Geographical Region of Host Department
88ac1392e9e0c2ac81e1a9d72fc648c	2018-01-01	Non-URM	Female	Non-URM	International
116599d35a0fad4843502ef61d5b24d5	2019-01-01	Non-URM	Male	Non-URM	International
05ac062b03fcfb983d8c9b0c59e78e7	2014-08-01	Non-URM	Female	Non-URM	International
2f89c7cd6227ce766a38d435d8a053d5	2018-08-01	Non-URM	Male	Non-URM	International
1aba91f655e4a122d1f44691c51cd354	2016-01-01	Non-URM	Female	Non-URM	International

5319a62ee563cb5f4d9525920b7e9094	Non-URM	Male	Non-URM Male	International
01-01	URM			

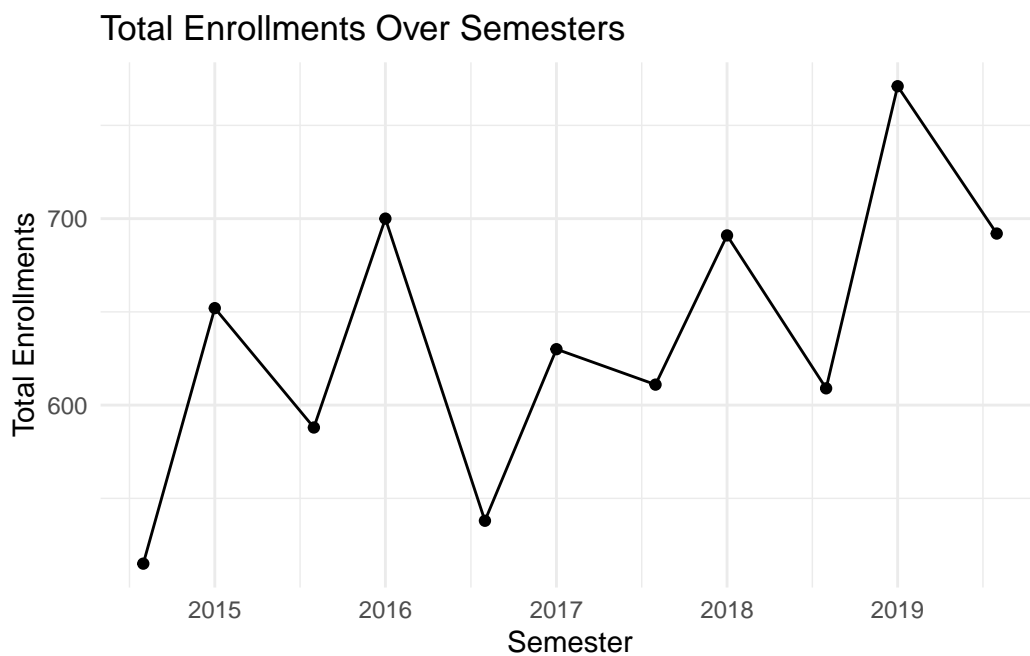
2.3 Measurement

TODO...

The subsequent section (Section 3) will present the results of our analysis, offering insights into ..., and the implications of the findings.

3 Results

```
ggplot(analysed_data, aes(x = semester, y = total)) +
  geom_line() + # or geom_col() for a bar plot
  geom_point() +
  theme_minimal() +
  labs(title = "Total Enrollments Over Semesters", x = "Semester", y = "Total Enrollments")
```



References

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Jennifer L. Doleac, Erin Hengel, and Elizabeth Pancotti. 2021. “Diversity in Economics Seminars: Who Gives Invited Talks?” *AEA Papers and Proceedings* 111: 55–59. <https://doi.org/10.1257/pandp.20211084>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.