Same As it Ever Was: Gender, Race, and Ethnicity Differences in Promotion for Academic Economists

Donna K. Ginther, Shulamit Kahn, Daria Milakhina

Online Appendix

Data Linkage Methods and Supplementary Tables

We linked the Academic Analytics data to several other data sources in order to identify economists in non-economics departments as well as the race and gender faculty members. Academic Analytics identifies the Digital Object Identifiers (DOI) of publications of all faculty in their data set from 2004 to 2022. We linked these publication DOIs to the Crossref[1] data set to retrieve the ISSN of the journals publishing the papers. We then linked the ISSN of journal publications in Academic Analytics to the ISSNs of economics journals curated by the Australian Business Deans Council[2] (ABDC) to identify the share of publications by faculty members in economics journals. A person was defined as an economist if they were: a) employed in an economics, agricultural economics or applied economics department; b) if over their career they published an average of 59% of their publications appear in economics journals;[3] or c) they were identified as Black economists by Mixon & Upadhyaya (2024). Our measure of economists in non-economics departments will likely undercount economists employed in policy departments because policy journals such as *Research Policy* or the *Journal of Policy Analysis and Management* are not considered economics journals in the ABDC list. Our definition of economists working in non-economics departments was deliberately conservative. We also merged on the top 25% of economics departments using data from REPEC.[4]

---

1 https://www.crossref.org
2 https://abdc.edu.au/abdc-journal-quality-list/
3 We chose the 59% threshold since this reflected the publication share of Shulamit Kahn, an economist employed in a non-economics department.
4 https://ideas.repec.org/top/top.usecondept.html.

Academic Analytics does not collect data on race or ethnicity. We used several approaches to identify race. We began by imputing race based on the first and last names of the respondents. We estimated the race based on the probability that a name is statistically more likely to be a certain race. Last name race probabilities were taken from the data with frequently occurring surnames in 2010 Census data.[5] Census provides data on 162,253 names with a frequency of 100 or more and the probabilities of them being one of the race categories (Comenetz, 2016). When last name is missing from the Census data, we supplement it with the dataset from Rosenman, et al (2022). Their data is based on the voters' registration files and additionally includes race probabilities for first names. Last names that are not in the Census data are rare last names, and in overall population, they represent only a small number of individuals. In our faculty dataset, the voter registration data increased last name-race matches by about 10%. This supplementation approach for last names is recommended by Rosenman *et al* because voters' data from select states are not representative of national race distributions. Because Census has more race categories that Rosenman, et al (2022), we combine American Indian or Alaska Native and two races from Census with the "Other" race category from Rosenman et al into the "other race" category.

We also used race probabilities for the first names from Rosenman, et al (2022). To combine all race probabilities into an estimated race variable we used two methods. In the first method, we simply put equal weight on the last and fist name probabilities and picked the race with the highest average probability. For this method, if the probability of either name is not available, the race variable is still estimated based on the last or first name that is available.

---

[5] https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

In the second method, we adjusted last name probabilities to the first name probabilities using conditional probability formula. This method estimates race only when first and last names are matched to race probability. When the two algorithms agree on the race of the individual, we use that information to assign race. The race algorithms do a reasonable job of identifying Hispanic and Asian races. However, if the name is rare or if the name is "white sounding," the algorithms will misidentify Black economists. For example, the race algorithms identified Federal Reserve Governor, Lisa Cook as white. To address this problem, we used the list of the top 200 cited economists from Mixon & Upadhyaya (2024) and matched this information by first and last name. We also used the names of board members and past presidents of the National Economics Association to identify Black economists.

Approximately 3,500 of the observations of economists had disagreement in predicted race or were missing information on gender. Using a third method, we conducted web searches by name and used that information to assign race, and in approximately 1,000 cases, gender. Although the Census will start using the new race category of Middle Eastern or North African (MENA) to the 2030 Census, individuals from this set of countries are categorized as white. Individuals from the Asian subcontinent are categorized as Asian. People from Spanish- or Portuguese-speaking countries are categorized as Hispanic. During our web searches, we used country of the bachelor's degree to assign race. Photographs, biographies, and course reviews in "Rate My Professor" were used to categorize people by gender. Despite these extensive efforts, it is likely that race and to a lesser extent, gender is measured with some error.

References:

Rosenman, Evan; Olivella, Santiago; Imai, Kosuke, 2022, "Race and ethnicity data for first, middle, and last names", https://doi.org/10.7910/DVN/SGKW0K, Harvard Dataverse.

Comenetz, Joshua. 2016. Frequently Occuring Surnames in the 2010 Census. Technical Report. United States Census Bureau https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf

Appendix Table 1: Distribution of Economists by Department and Institution Type

	Institution Type		
	Research Very High	Not Research Very High	Total
Field Categories			
Field Missing			
Frequency	34	166	200
Percent	0.40	7.79	1.90
Economics			
Frequency	5,563	1,116	6,679
Percent	66.24	52.35	63.43
Finance			
Frequency	645	266	911
Percent	7.68	12.48	8.65
Business & Law			
Frequency	1,306	350	1,656
Percent	15.55	16.42	15.73
Social & Behavioral Science			
Frequency	486	110	596
Percent	5.79	5.16	5.66
Education			
Frequency	93	35	128
Percent	1.11	1.64	1.22
Life & Health Science			
Frequency	87	32	119
Percent	1.04	1.50	1.13
Physical Science & Engineering			
Frequency	71	28	99
Percent	0.85	1.31	0.94
Humanities			
Frequency	113	29	142
Percent	1.35	1.36	1.35
Total			
Frequency	8,398	2,132	10,530
Percent	100.00	100.00	100.00
Number of Institutions	145	169	314

Source: Academic Analytics 2009—2022. Counts are based on the first time an individual is observed in the sample.