Prestige and Socioeconomic Scores for the 2010 Census Codes

GSS Methodological Report No. 124 (2015)

Michael Hout, *New York University*

Tom W. Smith, *NORC at the University of Chicago*

Peter V. Marsden, *Harvard University*

Introduction

The 2012 GSS included a popular prestige rating (Smith and Son 2014). A sample of 1,001 individuals, first interviewed in 2008 and included in the GSS panel, rated 90 occupations each; a rotation of occupations among respondents resulted in ratings for 860 occupational titles, most of which could be assigned to one of the 840 codes in the 2010 Standard Occupational Classification (SOC). This methodological report explains how we collected the ratings and converted them into prestige scores and a socioeconomic index for each of the 539 occupational categories of the Census Bureau's coding scheme now used in the GSS.

Occupational Titles and Occupational Categories

A broad sample of adults roughly representative of the U.S. household population rated occupational titles that correspond to the 2010 Standard Occupational Classification (SOC). The project was designed to update and extend past NORC prestige studies from 1947, 1963-1965, and 1989. We began with the 740 titles from the 1989 GSS prestige study (Nakao and Treas 1992), expecting to retain most of them. We dropped 99 titles, it turned out, for a variety of reasons. First, we dropped two made-up titles ("fooser" and "persologist"); although roughly half of the respondents asked to rate one of them in the 1989 study did so, few researchers have published results using these titles. Second, we altered several titles that were gendered pairs (e.g., "airline steward" and "airline stewardess"); we replaced most of them with gender-neutral titles (e.g., "flight attendant"). We retained "businessman," "landlord," and "policeman" to extend time series that go back at least to 1963-65. Third, we dropped seven titles that refer to activities that the Census Bureau does not classify as an occupation (e.g., "housewife," "retiree," and "prostitute") but kept two ("panhandler" and "street corner drug dealer"). Fourth, of the three titles that refer to an occupation that varies among respondents ("my own occupation," "the occupation my father had when I was growing up," and "the occupation of my spouse"), we kept "my own occupation" but dropped the other two. Fifth, when we coded the 723 titles still in consideration into the SOC. We found that they failed to cover 219 of the SOC codes. Therefore, we dropped 22 more titles that had been used in 1989 and added 219 new titles. We largely picked the new titles from the "illustrative examples" for the SOC (see **http://www.bls.gov/soc/2010/soc_alph.htm**).

At the end of this process we had 860 titles to be rated by 2012 GSS respondents. Each of the 539 categories of the 2010 scheme is covered by at least one title; 851 titles map onto the 2010 SOC codes. As previously noted, three titles from past studies do not map onto SOC codes: "my own occupation," "panhandler," "street corner drug dealer." Six other titles ("businessman," "supervisor of a skilled craftsman," "skilled craftsman in a factory," "semi-skilled worker in a factory," "unskilled worker in a factory," and "apprentice to a master craftsman") inherited from past studies are too general to be assigned to a single SOC code.

For the rating task, we divided the occupational titles into twelve batches of 90 titles. Each batch consisted of 70 occupational titles unique to that batch and 20 occupational titles that were common to all batches. Thus, each person rated 90 occupations (or less if they did not know enough about a title to rate it). With 1,001 people doing the rating, the occupations unique to a single batch were rated approximately 83 times while the common occupations could have been rated 1,001 times each if everyone gave every common occupation a rating. In practice, the rating task departed from this design to some extent, as we describe below.

The protocol for the rating task was first used by Hodge, Siegel, and Rossi (1965) and replicated by Nakao and Treas (1990). To start, the interviewer laid out a little board with boxes numbered from one to nine (reproduced in Nakao and Treas 1994). Box 9 was labeled "top" and box 1 was labeled "bottom." The words "top," "middle," and "bottom" were printed in the left margin, and the numbers 1 through 9 were printed in the right margin. The interviewer then handed the respondent a small card on which a job title was printed and read this statement to the respondent: "Please put the card in the box at the top of the ladder if you think that occupation has the highest possible social standing. Put it in the box of the bottom of the ladder if you think it has the lowest possible social standing. If it belongs somewhere in between, just put it in the box that matches the social standing of the occupation." The interviewer then handed the respondent 89 more cards and said, "Here are some more cards with names of occupations. Just put them on the ladder in the boxes that match the social standing they have. If you want to, you can change your mind about where an occupation belongs, and move its card to a different box.” After the respondent finished placing cards (or discarding ones that they could not place), the interviewer asked, “Would you like to change the placement of any occupation, or place a card which you couldn't place earlier?” When the respondent was finished, the interviewer collected the cards, putting those from box 1 into an envelope marked “1,” those from box 2 into an envelope marked “2,” and so forth up to 9; the discarded cards went into a tenth envelope.

Sample That Rated Occupations

The rating task was completed by people who were part of the GSS panel that was first interviewed in 2008. They rated occupations in 2012, near the end of their third (and last) interview. Attrition reduced original sample of 2,023 individuals to 1,295 by 2012. Of these 1,001 were interviewed in person; they are the sample that rated occupations. This subset of 1,001 of the original sample of 2,023 individuals is unweighted and is probably distinct in some ways from those who did not participate, but the sample is generally representative of the U.S. household-resident population for the ratings to be informative. In the past, ratings by professionals and educators correlated very highly with ratings by representative samples (Treiman 1977; Hauser 1992), so we expect that the ratings we have obtained are a valid representation of contemporary occupational prestige. Missing data and other problems further reduced the number of raters with usable data to 979.

From Ratings to Scores

We arranged the ratings into a dataset with one record for each combination of person and occupational title. In theory 1,001 raters doing 90 ratings each would yield 90,090 ratings, but some people rated less than 90 titles so we have 86,970 cases. Some of those were deemed to be invalid because the pattern suggests that either the rater or the interviewer reversed the codes (11 raters; 986 ratings)(Smith and Son, 2014). Others were dropped because the rater completed less than 20 ratings (11 raters; 147 ratings). Eight of the remaining raters completed the task by giving all their occupations the same score; we dropped those cases (8 raters; 719 ratings). Other raters used only 2-to-4 scores; we dropped all cases for which standard deviation of ratings was less than 0.9 (25 raters; 2,238 ratings). Our occupational scores come from this final dataset that consisted of 82,800 ratings provided by 946 raters.

Previous researchers (Nakao and Treas 1994) transformed the ratings so they range from 0 to 100 with the simple formula $Prestige = 12.5(Rating - 1)$; we follow that practice. We refer to this as the “standard prestige score.”

Duncan (1961), Hodge et al. (1964), and Hauser and Warren (1997) all focused on the percentage of ratings that were at or above a given threshold; "good" in Duncan's analysis and a rating of five in Hauser and Warren's (Hodge et al. had one dataset scored in four categories and two datasets scored 1-9 like ours). Following their lead, we calculated the percentage of ratings for each occupation that was greater than or equal to five. We refer to this as the "threshold prestige score."

To this point in the analysis, we have followed the practices of previous researchers with little deviation. Now we come to a point of more substantial departure. Previous researchers aggregated ratings to the census occupation level and averaged the standard and/or threshold prestige scores for each occupation category. They age-adjusted the averages to generate a score for that occupation (Hodge et al. 1965; Hauser and Warren 1997). Nakao and Treas (1990) made a couple of exceptions but mainly did that too. Averaging made sense when computing was time-consuming and expensive. But it does rely on a key assumption that raters do not differ. We take a more contemporary approach and remove the effect of each rater with a statistical adjustment based on a hierarchical linear model that uses the full dataset of 82,800 ratings.

Each rating reflects attributes of the occupations and of the raters. We want to capture the variation that reflects occupational differences and purge our prestige scores of variation related to differences among the raters. Our model for the standard prestige score is:

$$\mathit{Prestige}_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad (1)$$

for occupational title *i* and rater *j*. The α_i are the occupational differences of interest, and the β_j are differences among persons that we wish to control for in estimating the α_i. We estimate a hierarchical linear model (HLM) with raters *j* as the higher level. Expected values under (1) when $\beta_j = 0$ and $\varepsilon_{ij} = 0$ provide standard prestige scores purged of persistent differences among raters. Imagine two individuals who rated occupational titles from the same batch; each rated all 90 titles they saw. The first rater used the lower part of the scale, giving ratings that ranged from 1 to 7; the second used the upper part of the scale, giving ratings that ranged from 3 to 9. To keep the example simple, imagine that they both ranked the 90 titles in the same order; the second rater was just "more generous" in scoring. The HLM removes this difference between raters; each title would have the same adjusted score from these two raters because the range and order are identical. We take a similar approach to threshold prestige scores except that, as the observed data are binary, we fit a logistic HLM.[1]

The β_j terms allow us to see how large the differences among people are compared to differences among occupational titles. For the logit model of threshold scores, the variance of α_i is specified as a parameter of the model; the estimate from these data is 1.921. The standard deviation of the β_j is 1.906 when each occupation is given equal weight; when we calculate the expected logit from the fixed effects portion of the model, the core occupational titles get more weight and the variance of the expected values is slightly larger, 1.938.[2] The expected values from the HLM are on the logit scale; we invert the logit transformation to get threshold prestige scores that are purged of the person effects. They turn out to be very close to the means calculated in the usual way (r = 0.996), but some differences are as large as 10 percentage points. This result does not mean that rater variance is trivial; it just means that treating it as part of the total residual (as the usual approach does) leads to only a small amount of distortion in the scores.

In Figure 1 we summarize our results as histograms and kernel density plots for both the standard prestige variable and the threshold prestige score obtained after removing the rater-component from each. The standard prestige scoring resulted in substantial heaping around the mean prestige; both the histogram and the kernel density rise to a sharp peak around 44, then decline. Worse for the usability of the scale, 25 percent of occupations have a standard prestige score less than five points above or below the mean, and 48 percent have a rating less than ten points above or below the mean.[3] The threshold scoring approach resulted in a distribution that is far more uniform; the histogram and the kernel density rise quickly, decline a little, level off for most of the range, and diminish above 90 percent. Thus, the threshold approach provides much better discrimination among occupations for most of the data range. Further work will test our suspicion that the greater discrimination leads to better prediction. For now we rely on the histograms and densities to support our recommendation that researchers use the threshold measure.

(Figure 1 about here)

[1] We used the Stata routine -melogit- to obtain the estimates.
[2] The variance of the ε_{ij} is $\pi^2/3$ by definition.
[3] This is not a new difference. Similar plots based on the standard and threshold scores from the 1989 data show the same patterns. The standard score has a more sharply peaked histogram and kernel density plot when compared with those of the threshold score.

Socioeconomic Index

From its introduction in 1961, Duncan's socioeconomic index (SEI) has been a popular alternative to prestige scores. Combining information on the pay and credentials in an equation predicting prestige, Duncan's SEI and its successors remove some of the subjective aspects of popular ratings in a way that has proved to be better for estimating intergenerational correlations and many other correlations of interest (Hauser and Warren 1997).

Here we use our threshold-based prestige scores (from the HLM) as a criterion variable to generate SEI scores for the 2010 SOC. The original SEI and its successors were based on educational and income data (sometimes disaggregated by gender or self-employment) for each occupation in the census closest to the prestige study. The 2010 census had the fewest questions since 1870; the American Community Survey (ACS) was introduced to collect most of what used to be on the "long form" of the census, filled out by a sample of census households and group quarters. The education, occupation, and income data we need were long-form items that we now must get from the ACS. The ACS coded occupations to the 2010 SOC beginning in 2010.We used the three-year pooled public use sample for 2010-12 provided by IPUMS to estimate education and income for each occupation (Ruggles et al. 2010). There are fewer observations per occupation and, thus, more sampling error in three years of ACS data than in the census long-form samples used in previous studies because the three-year ACS sample is much smaller than the long-form samples from previous censuses were.[4]

The ACS data file available from IPUMS does not include all the occupational detail in the original data files. Forty-eight occupations were combined with others. For example, "sociologists" (code 1830) were combined with "miscellaneous social scientists and related workers" (code 1860) in the public-use file. Thus the public data file contains information on 491 occupational categories. The original and IPUMS ACS codes are shown in the Appendix.

We gathered data on income, usual hours worked, education, gender, race, and self-employment for each occupational category in the ACS. We selected people who were reported to be working in a

[4] According to the original ACS design, three years of ACS would have been 30 percent as large as the sample that would have filled out the census long form, but the Census Bureau reduced the size of the ACS samples in response to budget cuts.

given occupation at the time of the survey or, if not working, to have worked in that occupation when they last worked.

We turn first to pay. Hauser and Warren (1997) measured occupational pay both in terms of earnings (that is, the sum of wage, salary, and self-employed income in the previous year) and wages (that is, earnings per hour). They specified $25,000 (in 1990 dollars) as a threshold, with the aim of calculating the percentage of people working in each occupation who made that amount of money or more in the year before. They specified a second threshold of $14.30 per hour by dividing $25,000 by 50 weeks, then 35 hours per week. Adjusting $25,000 for inflation and rounding off, we get a new earnings threshold of $45,000; dividing by 50 weeks and 35 hours per week, we get a new wage threshold of $25.70.

Hauser and Warren (1997) focused on wages, presenting a latent variable model that showed hourly pay complemented credentials and threshold prestige best as an indicator of the pay component of social standing. We expected to follow their lead, but preliminary analyses using wages produced some anomalies. Most prominently, only 75 percent of physicians and surgeons had a wage over the threshold while 92 percent of nurse practitioners and 91 percent of pharmicists did. Physicians and surgeons had much higher annual earnings than the nurse practioners and pharmicists, but they also had higher hours — enough to drop one-fourth of them below the wage threshold. To avoid this prominent anomaly and some others, we used earnings as our pay measure, with a threshold of $45,000 per year.

For the educational component, Hauser and Warren used “some college” as their threshold. We considered both some college and the next major educational milestone, earning a college degree, as thresholds. In calculating the original SEI, Duncan (1961) used high school graduation as the threshold. Nakao and Treas (1994) moved the threshold up to some college, presumably because there was no longer enough variation in high school graduation rates among employed persons by 1980 (roughly 77 percent of the labor force had a high school diploma then by our calculation from Ruggles et al. (2010)). Exploratory analyses indicate that some college is still the best threshold. In the pooled 2010-2012 ACS we use here, 88 percent of 25-64 years olds have a high school diploma or more education, 57 percent have some college or more, and 30 percent have a college degree or more. Using some college as the threshold differentiates best among occupations for our analysis.

We fit five regressions that featured the threshold prestige measure (purged of rater-effects by the HLM method described above) as dependent variable and credentials and pay as independent variables. In these regressions we used all 539 occupational codes of interest; the 48 that were combined with others as described above have the same scores on the predictor variables as the category they were combined with. We used the 539 occupational categories so that we could get scores for all. Those scores with be the same for each pair that was combined in the public file. The results are in Table 1.

First we regressed percentages on percentages as the researchers prior to Hauser and Warren (1997) did (Model 0). In the 2012 data, credentials and pay had equal weight in the ratings. Finding equal weights for predictors echoes Duncan's (1961) result. Similar models for data from the 1960s and 1989 assigned more weight to education than to pay (Nakao and Treas 1994). Until we do more analysis, it will be hard to say whether the source of this change is in the categories the Census Bureau uses or in the public's assessments of social standing. We leave that for future research. We now turn to an alternative functional form introduced by Hauser and Warren (1997). They converted all percentages to "started logits":

$$\text{Started logit} = \ln((\text{Percentage} + 1) / (101 - \text{Percentage}))$$

where "percentage" refers to the percentage of interest — percentage rated 5 or more, percentage with some college or more, percentage earning $45,000 or more, etc. The usual logit transformation can result in undefined values for percentages of 0 or 100 and extreme values for percentages that approach those limits. The undefined logits fall out of the analysis while the extreme observations can have disproportionate leverage over regression results. Mosteller and Tukey (1979, pp. 109-115) proposed the started logit transformation as a method to keep all the cases in the analysis and reduce the leverage of the extreme observations. Hauser and Warren (1997) adopted the started logit approach, and we follow their lead here; Models 1-4 all use started logits instead of standard logits or percentages.

Using started logits for the regression of 2012 prestige (measured as percentage rated 5 or more with rater-effects removed) on some college or more education and earnings of $45,000 or more, we again find a near-equal weighting of credentials and pay. The started-logit functional form results in a better fit to the original data, as indicated by the scaled R^2.[5] A statistical test failed to reject (at the conventional 0.05 level) the null hypothesis that the coefficients for credentials and pay are equal. Hauser and Warren dropped seven influential observations (as indicated by a variety of post-estimation "influence" statistics they calculated) from their analysis; we drop the same occupations (now representing eight cases or six occupation categories) in Model 1.5. Our results are nearly identical with and without the cases that were influential in the Hauser and Warren's analysis. We replicated their search for influential observations in the 2012 data and found none to be "influential" by the definition Hauser and Warren used. In model 2 we add the racial and gender composition of occupations to our analysis. Neither percent black nor percent women significantly affect the ratings in 2012. We then add percent self-employed in model 3 and get another null result. We consider Model 1 to be our preferred model for these data.

In additional analyses (not shown) we performed the same analysis at the level of job titles (N = 851)[6] and nonredundant census categories (N = 491). The coefficients in those two analyses were identical to the ones in Table 1 because the most aggregated data — the 491 nonredundant census categories — contains all the information we have on credentials and pay. Less aggregation in the form of the full set of 539 occupational categories or 851 job titles introduces variation in the outcome variable (percentage rated 5 or more) but no more information about credentials or pay. So the R^2s and scaled R^2s for the alternative analyses differ; they are lower for the 851 job titles and higher for the 491 nonredundant census categories.

Conclusion

We have replicated and extended previous NORC prestige studies to generate occupational prestige and socioeconomic scores for the 539 occupational categories based on the 2010 Standard Occupational Classification (SOC) and U.S. Census Bureau's coding scheme. Respondents rated 860 occupational titles, 851 of which mapped onto 539 occupational categories and nine others that contribute to the

[5] What we are calling the "scaled R^2" is obtained by exponentiating the expected values from the started-logit regression, correlating those scores with the observed percentages for each occupational category, and squaring the result.

[6] As we noted on p. 2, nine of the 860 occupational titles did not correspond to a category in the SOC or the census, for example, "my own occupation" and "panhandler."

replication but do not map onto the SOC. We also used those ratings and data from the American Community Survey to generate a socioeconomic index (SEI) score for each occupational category.

The resulting occupational prestige and SEI scores can be linked to datasets like the GSS and CPS that report occupation using census codes. They can also be linked to other occupations coded that way, for example, the father's, mother's, and spouse's occupations in the GSS. We will merge the 2012 scores into the GSS cumulative data file and other GSS data products and provide an occupation-level data file others can use for their own analyses. Table 2 lists all the variables on the occupation-level data file.

REFERENCES

Duncan, Otis Dudley. 1961. “A Socioeconomic Index for All Occupations.” Pp. 109-138 in *Occupations and Social Status*, edited by Albert J. Reiss, Jr. Glencoe IL: Free Press.

Hauser, Robert M. 1982. "Occupational Status in the Nineteenth and Twentieth Centuries." *Historical Methods* 15: 11-26.

Hauser, Robert M., and John R. Warren. 1997. “Socioeconomic indexes for occupations: A review, update, and critique.” *Sociological Methodology* 27: 177-298.

Hodge, Robert W., Paul M. Siegel, and Peter H. Rossi. 1964. “Occupational Prestige in the United States, 1925-1963.” *American Journal of Sociology* 70: 286-302.

Hout, Michael, and Thomas A. Diprete. 2006. “RC28’s Contribution to Knowledge: An Assessment and an Agenda for the Future.” *Research in Social Stratification and Mobility* 24: 1-20.

Mosteller, Frederick, and John Tukey. 1979. *Data Analysis and Regression: A Second Course in Statistics.* Reading MA: Addison-Wesley.

Nakao, Keiko, and Judith Treas. 1990. “Computing 1989 Occupational Prestige Scores.” GSS Methodological Report MR70. http://publicdata.norc.org:41000/gss/DOCUMENTS/REPORTS/Methodological_Reports/MR070.pdf

Nakao, Keiko, and Judith Treas. 1994. “Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up.” *Sociological Methodology* 24: 1-72.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Smith, Tom W. and Jaesok Son. 2014. “Measuring Occupational Prestige on the 2012 General Social Survey.” GSS Methodological Report No. 122. Chicago: NORC.

Treiman, Donald J. 1977. Occupational Prestige in Comparative Perspective. New York: Academic Press.

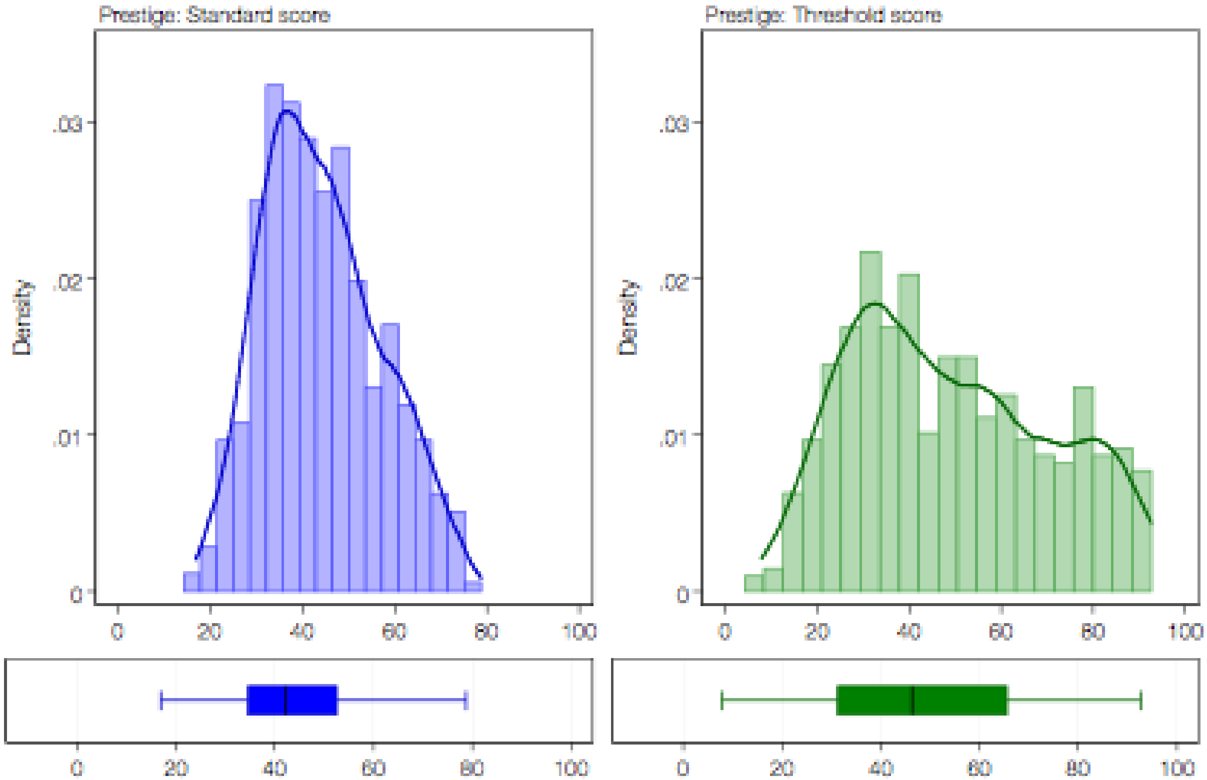

Figure 1. Histograms, kernel density estimates, and box plots for standard and threshold measures of occupational prestige, 2012.
Source: General Social Survey, 2012 prestige module.

Table 1. Regression coefficients for selected models of occupational prestige related to occupational education and earnings, 2012

	Model				
Independent variable	0	1	1.5	2	3
Some college or more	0.433	0.376	0.374	0.388	0.377
	(0.032)	(0.028)	(0.028)	(0.041)	(0.028)
Earned $45,000 or more	0.440	0.413	0.416	0.395	0.413
	(0.034)	(0.034)	(0.035)	(0.045)	(0.034)
Blacks				-0.032	
				(0.061)	
Women				-0.011	
				(0.030)	
Self-employed					-0.017
					(0.024)
Constant	2.977	-0.182	-0.179	-0.269	-0.228
	(1.427)	(0.036)	(0.037)	(0.129)	(0.074)
Functional form	Linear probability	Started logit	Started logit	Started logit	Started logit
R^2	0.702	0.689	0.689	0.689	0.689
Scaled R^2		0.705	0.706	0.705	0.705
N	539	539	533	539	539

Note: Standard errors in parentheses. The scaled R^2 is the squared correlation between the observed threshold score (purged of rater effects) and the exponentiated predicted score from the started logit regression.

Sources: General Social Survey prestige module, 2012 for dependent variable; American Community Surveys, 2010-2012 for independent variables.

Table 2. Variables in the Occupational Data File

Variable	Description
OCC10	2010 Census occupational category codes
OCC_IPUMS	2010 IPUMS occupational category codes. Please visit IPUMS (https://usa.ipums.org/usa/volii/c2ssoccup.shtml) for more information.
CENSUSTITLE10	Census occupation category names
COUNT_RATERS	Number of people who rated job titles in the occupation category
COUNT_JOBTITLES	Number of job titles rated in the occupation category
COUNT_AVE	Average number of raters per job title in the occupation category
RATING	Average rating of job titles in the occupation category. Raters rated each job title on a scale of 1 (bottom) to 9 (top). The rating here is aggregated at the level of occupation category.
PRESTG10	Prestige score for the 2010 occupation codes. This standard prestige score is a simple mean value of ratings for each occupation category, converted to a scale of 0 (bottom) to 100 (top). Please refer to GSS Methodological Report 70 for more information. This variable is included in the GSS public release.
PRESTG105	Threshold prestige score for the 2010 occupation codes. This prestige score is calculated using an alternative method, based on the percentage of ratings that was greater than or equal to a threshold (rating five). Please refer to GSS Methodological Report 124 for more information. This variable is not included in the GSS public release.
PRESTG105PLUS	Threshold prestige score for the 2010 occupation codes (person effect removed). This variable is also based on the threshold method as in PRESTG105; however, this variable takes one step further by removing rater effect using hierarchical linear modeling (HLM). Please refer to GSS Methodological Report 124 for more information. This variable is included in the GSS public release.
SEI10EDC	Percentage of those who had some college or more education in ACS 2010: 25-64 years old only. This variable was used to calculate SEI10. This variable is included in the GSS public release.
BAPLUS	Percentage of those who had college degree or more education in ACS 2010: 25-64 years old only
SOMECOLL_M	Percentage of those who had some college or more education in ACS 2010: 25-64 years old men only
SOMECOLL_W	Percentage of those who had some college or more education in ACS 2010: 25-64 years old women only
SOMECOLL_OTH	Percentage of those who had some college or more education in ACS 2010: 25-64 years old and those who work for others only
SOMECOLL_SELF	Percentage of those who had some college or more education in ACS 2010: 25-64 years old and self-employed only
SEI10INC	Percentage of those who earn $45k or more in ACS 2010: working full-year, full-time only. This variable was used to calculate SEI10. This variable is included in the GSS public release.
INCEARN45K_M	Percentage of those who earn $45k or more in ACS 2010: working full-year, full-time, men only

INCEARN45K_W	Percentage of those who earn $45k or more in ACS 2010: working full-year, full-time, women only
INCEARN45K_OTH	Percentage of those who earn $45k or more in ACS 2010: working full-year, full-time, and work for other only
INCEARN45K_SELF	Percentage of those who earn $45k or more in ACS 2010: working full-year, full-time, and self-employed only
BLACK_WEARN	Percentage of African-Americans in ACS 2010: all with earnings
WOMAN_WEARN	Percentage of women in ACS 2010: all with earnings
SELFEMP_WEARN	Percentage of self-employed in ACS 2010: all with earnings
SEI10	Socioeconoimc index for the 2010 occupation codes. It is estimated from 539 occupational categories, using PRESTG105PLUS. Please refer to GSS Methodological Report 124 for more information. This variable is included in the GSS public release.

Note: Variable names in bold indicate they are included in the GSS public data. You can find all these variables in a supplemental file at

http://gss.norc.org/Documents/other/PRESTG10SEI10_supplement.xls

Appendix Table: Original and IPUMS ACS codes for occupations

Occupational category	Original code	IPUMS -ACS code
Legislators	30	10
Funeral service managers	325	430
Postmasters and mail superintendents	400	430
Mathematicians	1210	1240
Statisticians	1230	1240
Biomedical engineers	1340	1330
Mining and geological engineers, including mining safety engineers	1500	1520
Nuclear engineers	1510	1530
Life scientists, all other	1660	1650
Survey researchers	1815	1860
Sociologists	1830	1860
Social science research assistants	1950	1965
Judges, magistrates, and other judicial workers	2110	2100
Media and communication equipment workers, all other	2960	2900
Exercise physiologists	3235	3245
Nurse midwives	3257	3258
Fish and game wardens	3830	3840
Transit and railroad police	3860	3850
Food preparation and serving related workers, all other	4160	4130
Correspondence clerks	5210	5350
Desktop publishers	5830	5940
Animal breeders	6020	6050
Hunters and trappers	6110	6100
Pile-driver operators	6310	6320
Solar photovoltaic installers	6540	6765
Septic tank servicers and sewer pipe cleaners	6750	6765
Roof bolters, mining	6910	6940
Roustabouts, oil and gas	6920	6800
Helpers--extraction workers	6930	6940
Electrical and electronics installers and repairers, transportation equipment	7050	7100
Wind turbine service technicians	7440	7630
Commercial divers	7520	7630

Signal and track switch repairers	7600	7630
Milling and planing machine setters, operators, and tenders, metal and plastic	8020	8220
Multiple machine tool setters, operators, and tenders, metal and plastic	8120	8220
Layout workers, metal and plastic	8160	8220
Textile bleaching and dyeing machine operators and tenders	8360	8400
Extruding and forming machine setters, operators, and tenders, synthetic and glass fibers	8430	8460
Fabric and apparel patternmakers	8440	8460
Model makers and patternmakers, wood	8520	8550
Semiconductor processors	8840	8965
Cooling and freezing equipment operators and tenders	8900	8965
Production workers, all other	8960	8965
Ship engineers	9330	9300
Bridge and lock tenders	9340	9420
Conveyor operators and tenders	9500	9560
Mine shuttle car operators	9730	9750
Tank car, truck, and ship loaders	9740	9750