

Chiteji P&P Data Notes Appendix

1. The data used in this analysis come from the Panel Study of Income Dynamics (PSID). The PSID is a publicly available dataset. Data can be downloaded for free at the PSID's website:

<<http://psidonline.isr.umich.edu/data/>>

2. Our sample is limited to individuals who are heads of households. To identify household heads in any given year of the PSID, one must subset the PSID by choosing individuals who meet two criteria: (a) their relationship to head variable equals 10 (for the year or years of interest) and (b) their sequence number is equal to 1 (for the relevant year or years).

3. Dependent variables:

The results reported in the paper are for two separate regressions. Each analyzes a separate dependent variable.

(a) The first is drinking. This variable comes from the 1968 PSID's question about whether the individual frequents bars [v293].¹ Our drinking variable is a dichotomous variable that recodes the PSID's original responses as follows: Responses of "4" or "5" are treated as yeses; responses in the range of 1-3 are treated as "no."

(b) The second dependent variable that the paper examines is exercising, as reported in the 1999 wave of the survey. It is measured using the PSID's heavy physical activity variable [ER15541]. The PSID coded responses for the physical activity variable by using zeroes for "never," and numbers ranging from 1 to 100 for the number of times the household head engaged in heavy physical activity each month (for those who did so). Our regressor treats a "0" as a no, and a positive value as a "yes."

4. Noncognitive skills measures:

As noted in the paper, the results reported in the paper use the PSID's time horizon variable [v2946] and its self-efficacy index [v2939] as measures of individuals' socio-emotional attributes. These data come from the 1972 wave of the PSID.

5. Other regressors:

(a) To measure health insurance coverage we use the PSID variables "v158" for 1968 and "ER15779" for 1999. For the 1968 PSID variable, a response code of "4" denotes an instance of no health insurance, while responses in the range of 1-3 and 5-8 are coded as yeses. For the 1999 PSID variable, a response code of "5" indicates a "no," while a code of 1 indicates an instance in which the family does have health insurance.

(b) The PSID education variables used are v313 (for 1968) and ER16516 (for 1999).

(c) The current wage in 1968 is measured using the PSID's wage variable "v337," and the 1999 PSID variable "ER16514" is the measure of the wage in 1999.

¹ The official PSID variable names appear in brackets in this sentence.

(d) The occupation variable for 1968 is [v197]; and for 1999 the analysis reported in the paper uses the PSID variable that tracks the head's occupation on his main job [ER13215]. As noted in the PSID's on-line documentation, in the 1968 survey the PSID responses to the occupation question are single digit responses denoting different occupational groupings. For 1999, the PSID reports 3-digit occupational codes that correspond to those available in the 1970 Census.² As noted in the main text, we combine the PSID's information about the head's occupation with data on the average earnings in his or her occupation to create a proxy measure of expected future earnings.³ We are essentially using the average earnings in occupations to order occupations according to whether they are high-earning jobs or low-earning jobs (and everything in between). The earnings by occupation data come from the Bureau of Labor Statistic data reported in the *Statistical Abstract of the United States*, which is published annually by the federal government.⁴ It is available in print or on-line at www.census.gov. Note that for 1999 we had to treat individuals whose occupations were listed in the PSID as farm managers or private household employees as missing data because we were unable to obtain information about average earnings in these two occupations from the Statistical Abstracts tables. For the 1968 occupation data, individuals with PSID occupation response codes of 4 were dropped from the analysis because the data in Statistical Abstract did not allow us to easily identify clerical and sales workers.

² More specifically, the 1999 occupation measure in the PSID is "the 3-digit occupation code from 1970 Census of Population; Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the Bureau of Census" (PSID on-line documentation for variable ER13215). More detailed information about 1999 occupational codes is available in the PSID's Appendix V2, Wave XIV (1981) documentation.

³ For 1999, we use data about the average weekly earnings for men because that is what is available. For 1968, we have average yearly earnings for all full-time workers in the relevant occupational group.

⁴ The data for earnings across occupations in 1968 come from the column for the 1968 calendar year in Table 347 of the 1970 edition of *Statistical Abstract*. The 1999 earnings across occupations information is taken from the column for the 1999 calendar year in Table 696 of the 2000 edition of *Statistical Abstract*.