

Economics 718: Topics in Applied Econometrics
Problem Set 1
DUE: Tuesday November 26, 2013

Is there a marriage male–wage premium?

The object of this problem set is to determine whether males earn a wage premium while married. To be precise let $W_i^{nm}(a)$ denote the wage for male i age a when never–married (single) and $W_i^m(a)$ when married. The claimed premium is $\pi(a, b) = W_i^m(b) - W_i^{nm}(a) > 0$, for ages $b > a$.

Weekly wages and marriage are positively correlated. The substantive issue is whether the correlation reflects *selectivity* of males who get marriage (males who earn higher wages are more likely to get married). Or does it reflect *state dependence*? Marriage allows a division of labor that may permit the male to specialize, exert more effort to the market and thus secure higher wages.

I have put `E718.malePS1.dta` on the course home page. This is an extract of males from the NLSY79. Documentation for the file is in self–contained within the file (variable labels and value labels). I have created a number of variables to facilitate your analysis.

- `Urban` is an indicator variable equal to one if the respondent lives in a MSA, and zero otherwise.
- `South` is an indicator if the respondent lives in one of twelve southern states.¹
- `rwkearn` is real weekly earnings, defined as (total wage and salary earnings per annum)/weeks worked per year deflated by the CPI 1982–1984=100.
- `lrwkearn` is the natural logarithm of `rwkearn`.
- `White`, `Black`, and `Hispanic` are coded from `race`.
- `s90` is an indicator variable equal to 1 if `year == 1990, 1991, 1992`, and zero otherwise.

A few words on the NLSY

Missing value codes used in the NLS program are: -1 Refused; -2 Don't know; -3 Invalid skip;² -4 Valid Skip; -5 Not Interviewed. For many variables but not all I have reset missing value codes (-1,...,-5) to missing (.)

Respondents of the NLSY79 report on earnings the calendar year *before* the interview date. For example, an individual interviewed in February 1993 reports his or her salary and wage income for calendar year 1992. In 1994, the NLSY79 changed from an annual interview cycle to a biennial interview cycle. Thus, information from the 1993 is the last interview round using an annual instrument. Given the birth years of the cohort (1957–1964) most of the first marriages occur before 1994.

The variable `cps` contains the CPS activity code, in English, employment status. NLSY respondents answered the CPS questions on employment to enable comparisons between the NLSY and CPS. The CPS questions reference the week before the interview to mimic the CPS survey. The NLSY obtains information on the work status week–by–week (and by job) since the date of last

¹The states, by postal abbreviation, are: AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, TX, VA.

²These are interviewer errors or programming errors — the question should have been asked but was not.

interview (or January 1, 1978 in the initial interview). Because of the timing differences, do not be confused by respondents who report themselves `cps = 6` “unable to work” (last week) but who report positive hours (`hrs`) and weeks (`wks`) of work (last year). There are a number of these “oddities”.

Online documentation for the NLSY79 and other NLS cohorts is available at www.bls.gov/nls.

The assignment is to perform the requested analyses below (and others as you deem necessary and appropriate) to determine whether there is a male marriage–premium in the NLSY79. Your answer must include.

1. An economic framework to interpret your empirical analyses. THIS FRAMEWORK NEED NOT BE A FORMAL MATHEMATICAL MODEL.
2. A discussion of your reasoning regarding your analysis sample (what groups are included and why). Your data includes all male respondents of the NLSY79.
3. Please present three analyses of the marriage–wage premium:
 - (a) A descriptive analysis: means, cross–tabs, regression (ols).
 - (b) An analysis that controls for selection on the observables.
 - (c) An analysis that controls for selection on the unobservables.

Finally, I ask that you type up your answers, but do *NOT* take the time to reformat statistical output into tables—just annotate the STATA or SAS or R or ... output.