



Here,  $s_i$  are individual  $i$ 's years of schooling,  $e_i$  his/her work experience in 10s of years (i.e.,  $e_i = 1$  means 10 years of experience, etc.) and  $f_i$  is a sex dummy ( $f_i = 1$  if  $i$  is a female; 0 otherwise).

- Give an economic interpretation of each the coefficients in equation (2).

3. I would like to relate these results to the relation between union status ( $d_i$ ) and the individual characteristics  $s_i$ ,  $e_i$  and  $f_i$ . To this end, I use a “logit” model, which specifies the probability that an individual with characteristics  $s_i$ ,  $e_i$  and  $f_i$  is a union member as

$$\frac{\exp(\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i)}{1 + \exp(\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i)},$$

where  $\delta_0$ ,  $\delta_1$ ,  $\delta_2$  and  $\delta_3$  are unknown parameters (like  $\alpha$  and  $\beta$  in the linear regression model). We can alternatively interpret this probability as the share of the subpopulation with characteristics  $s_i$ ,  $e_i$  and  $f_i$  that are union members.

- Check that this probability or share lies between 0 and 1 (as it should) and that it is increasing in the “index”  $\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i$ . What is the interpretation of the parameters  $\delta_0$ ,  $\delta_1$ ,  $\delta_2$  and  $\delta_3$  (you can omit details in this case)? Why wouldn't I just specify a linear regression model for union status instead?

4. I have estimated the logit model, i.e. computed estimates of  $\delta_0$ ,  $\delta_1$ ,  $\delta_2$  and  $\delta_3$  such that the model fits the data best (in some way we will not discuss). This gives the following predicted union membership probability for given characteristics  $s_i$ ,  $e_i$  and  $f_i$ ,

$$\frac{\exp(-2.603 + 0.030 s_i + 0.370 e_i - 0.483 f_i)}{1 + \exp(-2.603 + 0.030 s_i + 0.370 e_i - 0.483 f_i)}. \quad (3)$$

- Discuss (3) in relation to (1) and (2). In particular, discuss the differences in estimated union effects on wages between (1) and (2).

5. I am still somewhat uncomfortable with the model and have added a race dummy  $n_i$  ( $n_i = 1$  if  $i$  is nonwhite according to the CPS definition and  $n_i = 0$  otherwise) and an indicator  $h_i$  that is 1 if  $i$  is *not* paid by the hour and 0 if he/she is. This gives

$$\begin{aligned} \ln(w_i) = & 1.038 & + & 0.208 & d_i & + & 0.084 & s_i & + & 0.112 & e_i \\ & (0.074) & & (0.036) & & & (0.005) & & & (0.012) \\ & & & - & 0.235 & f_i & - & 0.124 & n_i & + & 0.229 & h_i & (4) \\ & & & & (0.026) & & (0.036) & & & (0.029) \\ & & & + & \hat{u}_i & & & & & & & & \end{aligned}$$

and an estimated probability of union membership

$$\frac{\exp(-3.237 + 0.081s_i + 0.424e_i - 0.557f_i + 0.652n_i - 0.599h_i)}{1 + \exp(-3.237 + 0.081s_i + 0.424e_i - 0.557f_i + 0.652n_i - 0.599h_i)}$$

- Briefly discuss.

6. Finally, I am worried that the effect of union membership on wages may be heterogeneous between workers. So, I have estimated a further extended model, giving

$$\begin{aligned} \ln(w_i) = & 1.014 + 0.286 d_i + 0.085 s_i + 0.113 e_i \\ & (0.079) \quad (0.231) \quad (0.005) \quad (0.013) \\ & - 0.232 f_i - 0.143 n_i + 0.266 h_i \\ & (0.028) \quad (0.041) \quad (0.031) \\ & + 0.001 s_i \cdot d_i - 0.011 e_i \cdot d_i + 0.020 f_i \cdot d_i \\ & (0.015) \quad (0.033) \quad (0.072) \\ & + 0.075 n_i \cdot d_i - 0.258 h_i \cdot d_i + \hat{u}_i \\ & (0.089) \quad (0.085) \end{aligned}$$

- What is the estimated union effect on wages as a function of the other characteristics of the individual? Discuss.
  - The averages of  $s_i$ ,  $e_i$ ,  $f_i$ ,  $n_i$  and  $h_i$  in the sample are respectively 13.145, 1.879 (i.e. 18.79 years), 0.497 (i.e. 49.7% females), 0.153 and 0.407. What is the estimated union effect on wages at these sample average values of  $s_i$ ,  $e_i$ ,  $f_i$ ,  $n_i$  and  $h_i$ ? Compare to the estimated union effect in (4).
  - Even in the last, richest model the characteristics  $d_i$ ,  $s_i$ ,  $e_i$ ,  $f_i$ ,  $n_i$  and  $h_i$  together explain less than 40% of the variance in log wages. The residuals account for the remaining variation. Discuss.
7. Based on all these results, if you had to report on the effect of unions on wages what would you say? How do the results compare to those in the literature (check a review in a text book).
8. *Optional:* If you can handle regression software (in particular: STATA) and would like to replicate some of my regressions, or try some of your own, you can download the data as a STATA file (`wagetest.dta`) from the course web page. I have also posted a STATA batch file (`probb442.do`) that reproduces my work for this problem set.