Solutions for problem set 2

1. The equation

$$\ln w_i = 2.300 + 0.266 d_i + \widehat{u}_i \tag{1}$$

attempts to predict individual log-wage on the basis of whether or not she/he is a member of a trade-union. For a individual j who is not member of a trade union, we have $d_j = 0$ and therefore

$$\ln w_j = 2.300 + \widehat{u}_j,$$

whereas for a union member i, we have $d_i = 1$ and

$$\ln w_i = 2.300 + 0.266 + \widehat{u}_i = 2.566 + \widehat{u}_i.$$

Hence, the number 0.266 is the average difference between the log-wage of union members and non-members. This means that the approximative gap between wages of unionized and non-unionized workers is 26.6%.

An easier way to obtain these data is to simply compute the mean log-wages of union members and non-members and to calculate their difference.

2. The intercept term is estimated to be equal to 0.862 and corresponds to the mean log wage of an individual not belonging to a trade-union ($d_i = 0$), with no year of schooling ($s_i = 0$), no experience ($e_i = 0$) and male ($f_i = 0$). For those individuals, the expected log-wage is 0.862.

The value 0.168 is an estimate of the average increase in log-wage of a union member in comparison to non-union members, both having the same years of schooling, experience and gender. The wage gap between union members and non-members is about 16.8% *ceteris paribus*.

According to the model, an additional year of schooling yields in average a 10.2% increase in wage, all other (controlled) characteristics being held constant.

The estimated impact on average wage of an increase in labour market experience by 10 years is about 13.1% *ceteris paribus*.

The coefficient before the f_i dummy variable picks up the association between gender and log-wage, for in all other things identical individuals. The estimated gap is about 25.4%.

3. Let $f(x) = e^x/(1+e^x)$, then $f(0) = 1/2$, $0 < f(x) < 1$ and

$$\lim_{x \to -\infty} f(x) = 0,$$

and

$$\lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} \frac{1+e^x-1}{1+e^x} = \lim_{x \to +\infty} 1 - \frac{1}{1+e^x} = 1.$$

That is, the function f is into $(0;1)$ (and takes no other values outside this interval). This function f is also increasing in x:

$$\begin{aligned} f'(x) &= \frac{e^x}{1+e^x} - \frac{e^x}{(1+e^x)^2} e^x = \left(1 - \frac{e^x}{1+e^x}\right) \frac{e^x}{1+e^x} \\ &= \frac{1}{1+e^x} \frac{e^x}{1+e^x} > 0. \end{aligned}$$

These results for the function f can be applied directly to our logit model, as we can write

$$\begin{aligned} \Pr(d_i = 1 | s_i, e_i, f_i) &= \frac{\exp(\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i)}{1 + \exp(\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i)} \\ &= f(\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i). \end{aligned}$$

The analysis of the function f above implies that $0 < \Pr(d_i = 1 | s_i, e_i, f_i) < 1$ for all $\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i$. Also, $\Pr(d_i = 1 | s_i, e_i, f_i)$ is increasing in $\delta_0 + \delta_1 s_i + \delta_2 e_i + \delta_3 f_i$. The different parameters δ_k reflect how strongly the probability of being a union member depends on the k^{th} explanatory variable *ceteris paribus*. If $\delta_1 > 0$ then the probability of being a union member is increasing in schooling; if $\delta_1 < 0$ then $\Pr(d_i = 1 | s_i, e_i, f_i)$ is decreasing in s_i.

A linear regression model would specify

$$\Pr(d_i = 1 | s_i, e_i, f_i) = \alpha_0 + \alpha_1 s_i + \alpha_2 e_i + \alpha_3 f_i,$$

which is not necessarily confined in the $(0;1)$ interval and may therefore violate the properties of the conditional probability.

4. Equation (3) shows that experienced and educated workers are relatively likely to be union members, and that women are less likely to be union members than men. This empirical evidence may help to understand why the estimates of the coefficient associated with union-membership (d_i) are different between equations (1) and (2). Indeed, in (1) we did not control for schooling, sex and experience, but in (2) we did. It has been shown in Exercise 1 that the estimates in (1) are just the mean log-wage

of non-unionised workers (2.3) and the mean union log-wage premium (0.266). In equation (2) we again compute the log-wage effect of union membership (0.168), but this time for given characteristics (s_i, e_i, f_i). It can be seen from (2) that, *ceteris paribus*, schooled and experienced persons have a higher log-wage on average and women have a lower log-wage on average. As equation (1) does not control for schooling, experience and sex, the estimate 0.266 does not only reflect the effect of union membership itself on wages, but also the fact that union-members are mainly higher-educated, more experienced or/and male persons (see equation (3)) and have higher wages on average for those reasons. Thus, 0.266 in equation (1) overstates the effects of union membership on log wages (if indeed schooling, experience and sex are all the joint determinants of wages and union status) and 0.168 may be a better estimate.

5. In the extended model (4), we still find some of the results of model (2): union membership, schooling, experience have, *ceteris paribus*, a positive impact on wages, whereas women on average are paid less. The inclusion of two additional variables now provides some new information: nonwhite persons on average get a lower log-wage, whereas persons not paid by the hour get a higher log-wage.

   The extended logit model indicates that nonwhites are more likely to join a union whereas persons not paid by the hour are less likely to do so. The other estimates are rather similar to model (3).

   This evidence helps to understand why the estimate of union-membership are now greater than in (2). In (2) the variable d_i also captures the impacts of the omitted variables n_i and h_i. As the variables "nonwhite" and "paid by the hours" have a negative impact on log-wage, but persons with those characteristics are more likely to be member of a union, neglecting n_i and h_i as was the case in (2), leads to an underestimate of the wage gap due to union membership.

6. (a) We can isolate the effect of union membership as follow:

$$\begin{aligned}\ln w_i &= 1.014 \\ &\quad + (0.286 + 0.001 s_i - 0.011 e_i + 0.020 f_i + 0.075 n_i - 0.258 h_i)\, d_i \\ &\quad + 0.085 s_i + 0.113 e_i - 0.232 f_i - 0.143 n_i + 0.266 h_i + \widehat{u}_i.\end{aligned}$$

   For union menbers, $d_i = 1$ and

$$\ln w_i = 1.300 + 0.086 s_i + 0.102 e_i - 0.212 f_i - 0.068 n_i + 0.008 h_i + \widehat{u}_i.$$

   For non members, $d_i = 0$ and

$$\ln w_i = 1.014 + 0.085 s_i + 0.113 e_i - 0.232 f_i - 0.143 n_i + 0.266 h_i + \widehat{u}_i.$$

Now the impact of union membership on log-wages is

$$0.286 + 0.001 s_i - 0.011 e_i + 0.020 f_i + 0.075 n_i - 0.258 h_i. \quad (2)$$

It is not merely constant as in (1) and (2), but can depend on characteristics $(s_i, e_i, f_i, n_i, h_i)$. For two individuals identical in all respects but union membership, the difference in log-wages will be about equal to the number in (2). The variable h_i has a great impact on this difference. The other impacts are rather small in absolute value and not estimated very precisely: the reported standard error shows that the coefficients on s_i, e_i, f_i and n_i in (2) are not significantly different from zero.

(b) The estimated union effect is on average:

$$\begin{aligned} & 0.286 + 0.001 \times 13.145 - 0.011 \times 1.879 \\ & +0.020 \times 0.497 + 0.075 \times 0.153 - 0.258 \times 0.407 \\ = \; & 0.19489, \end{aligned}$$

which is roughly equal to the estimated impact in model (4).

(c) The fact that d_i, s_i, e_i, f_i, n_i and h_i explain less that 40% of the variation in wages suggests that there are important other sources of wage variation. Union status in particular explains only a small part of the observed wage differentials. From an economic point of view, we know that wages are different across industries, between small and large firms, across regions corresponding to different "labour markets", individual ability, etc. If we can measure these "omitted variables" and include them in our model, this could increase its explanatory power. More importantly, the inclusion of more explanatory variables could suppress (or at least reduce) the estimation bias due to omitted variables (as was the case in (1) in comparison to (2) and (4)).

In practice it has been found to be hard to explain much more than 40% of the individual variation in wages. Much of the variation is due to factors that remain unobserved to the economist.

7. We have found that the mean union wage gap is about 20% in the USA. This is consistent with the findings reported by Booth (1995, page 169). This is true even though our analysis has been simple (and not so careful) relative to some of the empirical literature. Obviously, a serious empirical study of union wage effects in the US would require more discussion of the various econometric problems we have discussed.