Applied Econometrics II
Dept of Economics, Carnegie Mellon University
73-360, Fall 2000

Solution #4

For this assignment we want to calculate the difference in the probability of being employed according to whether a person lives in the city of pittsburgh, the mon valley, or the rest of Allegheny county. So, please calculate the difference in prob of being employed between the city and rest of county and between the mon valley and rest of county. In all cases, please hold constant age and years of schooling and calculate the quantities at sample means.

Please use the probit, logit and linear probability models. Discuss which of the model(s) you prefer and why. In addition, please test the theory that location has no effect on the probability of being employed.

Notice the question is about effect sizes, so we will calculate 95% confidence intervals throughout.

Let's begin with the linear probability model. Notice the use of heteroskedasticity corrected standard errors!

For the difference between employment probabilities in Pittsburgh and the rest of Allegheny County (a 90% CI):

$$
\begin{aligned}
\beta_{\text{pgh}} &= -0.054 \pm 1.645 \left(\sqrt{0.000437} \right) \\
&= -0.054 \pm 0.03
\end{aligned}
\tag{1}
$$

For the difference between employment probabilities in the Mon Valley and the rest of Allegheny County (a 90% CI):

$$
\begin{aligned}
\beta_{\text{mon}} &= -0.022 \pm 1.645 \left(\sqrt{0.000611} \right) \\
&= -0.022 \pm 0.04
\end{aligned}
\tag{2}
$$

Now, we will use the logit model. For the logit model, it would not be appropriate here to use the approximate method, since the change we are thinking about (pgh going from 0 to 1 or mon going from 0 to 1) is large — neither pgh nor mon can ever change more than 1! So, since we do not know how to calculate CI for the exact method, we will not calculate CI. Also, for all calcultions, we will evaluate age and years of schooling at their sample means.

Using the logit model and the output on page 3 and the sample means on page 1:

$$
\begin{aligned}
P\left\{emp = 1|\text{pittsburgh}\right\} &= \frac{exp\left(-2.805 - 0.843 + 0.026(37.78) + 0.368(13.84)\right)}{1 + exp\left(-2.805 - 0.843 + 0.026(37.78) + 0.368(13.84)\right)} \\
&= \frac{exp(2.427)}{1 + exp(2.427)} \\
&= 0.919
\end{aligned}
$$

$$
\begin{aligned}
P\left\{emp = 1|\text{monvalley}\right\} &= \frac{exp\left(-2.805 - 0.446 + 0.026(37.78) + 0.368(13.84)\right)}{1 + exp\left(-2.805 - 0.446 + 0.026(37.78) + 0.368(13.84)\right)} \\
&= \frac{exp(2.824)}{1 + exp(2.824)} \\
&= 0.944
\end{aligned}
$$

$$
\begin{aligned}
P\left\{emp = 1|\text{rest}\right\} &= \frac{exp\left(-2.805 + 0.026(37.78) + 0.368(13.84)\right)}{1 + exp\left(-2.805 + 0.026(37.78) + 0.368(13.84)\right)} \\
&= \frac{exp(3.27)}{1 + exp(3.27)} \\
&= 0.963
\end{aligned}
$$

So, the estimate of the difference between the employment rates for an average person, controlling for age and schooling, between Pittsburgh and the rest of Allegheny County is 4.4%. The similar figure for the difference between the Mon Valley and the rest of Allegheny County is 1.9%.

Now, to test the importance of location, we must test the null hypothesis: $H_0 : \beta_{\text{pgh}} = \beta_{\text{mon}} = 0$. To do this, we compare the logits on pages 3 and 4. The likelihood ratio statistic is:

$$
\begin{aligned}
LR &= -2(lnL_R - lnL_{UR}) \\
&= 426.6 - 418.6 \\
&= 8
\end{aligned}
$$

Taking 8 to the χ_2^2 table, we see that we can reject the null that location has no effect at the 5% but not the 1therefore at least 95% but not 99% confident that location affects employment rates.

Next, we go to the probit model, appearing on page 7. Recall that the signs of the coefficients are reversed!

$$
\begin{aligned}
P\left\{emp = 1|\text{pittsburgh}\right\} &= \Phi(-1.067 - 0.415 + 0.0122(37.78) + 0.172(13.84)) \\
&= \Phi(1.359) \\
&= 0.913
\end{aligned}
$$

$$P\{emp = 1|\text{monvalley}\} = \Phi(-1.067 - 0.200 + 0.0122(37.78) + 0.172(13.84))$$
$$= \Phi(1.574)$$
$$= 0.942$$

$$P\{emp = 1|\text{rest}\} = \Phi(-1.067 + 0.0122(37.78) + 0.172(13.84))$$
$$= \Phi(1.774)$$
$$= 0.963$$

So, the estimate of the difference between the employment rates for an average person, controlling for age and schooling, between Pittsburgh and the rest of Allegheny County is 5.0%. The similar figure for the difference between the Mon Valley and the rest of Allegheny County is 2.1%.

Now, to test the importance of location, we must test the null hypothesis: $H_0 : \beta_{\text{pgh}} = \beta_{\text{mon}} = 0$. To do this, we compare the probits on pages 6 and 8. The likelihood ratio statistic is:

$$LR = -2(lnL_R - lnL_{UR})$$
$$= -2(-214.11 - (-210.14)))$$
$$= 7.94$$

Taking 7.94 to the χ^2_2 table, we see that we can reject the null that location has no effect at the 5% but not the 1therefore at least 95% but not 99% confident that location affects employment rates.

The probit and/or logit results are superior to the linear probability model results since the lpm is an incoherent model (ie it can make predictions outside the 0,1 interval). Notice that the three models give pretty similar results, though.