

Econometrics Theory and Methods II

Exam 1

Suggested Solutions

1. (40 points)

- (a) We can see from the Stata output that the coefficient on income in the linear probability model is statistically insignificant (the t-statistic is 1.08). For the logit model, the z-statistic on income is 1.12, which is also insignificant.
- (b) Using the Stata output, we can test the null hypothesis that neither *income* nor *white* significantly affect the propensity to smoke. The maximized log-likelihood from a regression that constrains the coefficients on these variables to be zero is $\mathcal{L}_c = -525.09928$. The maximized log-likelihood on the unconstrained model (the one that includes *income* and *white* as well as *educ* and *age*) is $\mathcal{L}_u = -524.46796$. The likelihood ratio test statistic is

$$2(\mathcal{L}_u - \mathcal{L}_c) = 2[-524.46796 - (-525.09928)] = 1.26264.$$

The test statistic has an asymptotic χ^2 distribution with two degrees of freedom (one for each restriction). The $\alpha = 0.05$ critical value for a chi-square random variable with two degrees of freedom is 5.991, and our test statistic is well below this threshold, so we cannot reject the null hypothesis that $\beta_{income} = \beta_{white} = 0$.

- (c) For the linear probability model, the marginal effect is the same as the coefficient. For the logit model, the marginal effect (estimated at the sample means) is

$$\frac{\partial \widehat{\Pr}(Smoke = 1|x)}{\partial x} = \frac{\partial}{\partial x} \left(\frac{\exp(\bar{x}\hat{\beta})}{1 + \exp(\bar{x}\hat{\beta})} \right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1\bar{x})}{[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1\bar{x})]^2} \hat{\beta}_1.$$

Plugging in $\hat{\beta} = (-.0835693, .5627327)$ and $\bar{x} = 12.47088$, we find that the estimated marginal effect is $\partial \widehat{\Pr}(Smoke = 1|x)/\partial x = -.0197362$.

- (d) The variance of this marginal effect can be approximated using the Delta method. Define $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ as $h(\hat{\beta}_0, \hat{\beta}_1) = \exp(\hat{\beta}_0 + \hat{\beta}_1\bar{x})\hat{\beta}_1/[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1\bar{x})]^2$. (Note well that this is a function of both $\hat{\beta}_0$ and $\hat{\beta}_1$.) The Delta method tells us that, since $h(\cdot)$ is continuously differentiable,

$$h(\hat{\beta}_0, \hat{\beta}_1) \stackrel{a}{\sim} N \left(h(\hat{\beta}_0, \hat{\beta}_1), \nabla_{\hat{\beta}} V \nabla'_{\hat{\beta}} \right),$$

where

$$V = \begin{bmatrix} \sigma_0^2 & \rho_{01} \\ \rho_{01} & \sigma_1^2 \end{bmatrix}$$

is the variance-covariance matrix of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$. In this case, we have that $\nabla h(\hat{\beta}) = [h_0 \ h_1]$, where

$$h_0 = \nabla_{\hat{\beta}_0} h(\hat{\beta}) = \hat{\beta}_1 \frac{[1 + \exp(\bar{x}\hat{\beta})]^2 \exp(\bar{x}\hat{\beta}) - \exp(\bar{x}\hat{\beta}) 2 [1 + \exp(\bar{x}\hat{\beta})] \exp(\bar{x}\hat{\beta})}{[1 + \exp(\bar{x}\hat{\beta})]^4}$$

and

$$h_1 = \nabla_{\hat{\beta}_1} h(\hat{\beta}) = \frac{\exp(\bar{x}\hat{\beta})}{1 + \exp(\bar{x}\hat{\beta})} + \hat{\beta}_1 \frac{[1 + \exp(\bar{x}\hat{\beta})]^2 \exp(\bar{x}\hat{\beta})\bar{x} - \exp(\bar{x}\hat{\beta})2[1 + \exp(\bar{x}\hat{\beta})] \exp(\bar{x}\hat{\beta})\bar{x}}{[1 + \exp(\bar{x}\hat{\beta})]^4}.$$

Thus,

$$\begin{aligned} \text{Asy Var} \left(h(\hat{\beta}_0, \hat{\beta}_1) \right) &= \begin{bmatrix} h_0 & h_1 \end{bmatrix} \begin{bmatrix} \sigma_0^2 & \rho_{01} \\ \rho_{01} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} \\ &= h_0^2 \sigma_0^2 + 2h_0 h_1 \rho_{01} + h_1^2 \sigma_2^2. \end{aligned}$$

The standard error of the estimated marginal effect is the square root of this expression (replacing the unknown variances with estimated sample variances).¹ Since the estimated effect is asymptotically normal, we can reject the null hypothesis that it is equal to zero if

$$\left| \frac{\partial \widehat{\text{Pr}}(\text{Smoke} = 1|x)/\partial x}{\sqrt{\widehat{\text{Asy Var}} \left(h(\hat{\beta}_0, \hat{\beta}_1) \right)}} \right| > z_{\alpha/2}.$$

Carrying out the calculation gives a marginal effect of $\partial \widehat{\text{Pr}}(\text{Smoke} = 1|\bar{x})/\partial x = -.0197362$ with an estimated asymptotic standard error of .00569952, so the z-statistic is -3.46 , and we reject the null hypothesis that $\partial \text{Pr}(\text{Smoke} = 1|x)/\partial x = 0$ at any conventional significance level.

- (e) The estimated marginal effects are actually quite similar between the logit model and the linear probability model (-0.0197 and -0.0195 , respectively). However, since the logit model constrains the estimated probability of smoking to lie between 0 and 1, the logit model is somewhat preferred.
- (f) The variance of a Bernoulli random variable y with success probability p is $\text{Var}(y) = p(1 - p)$. In the case of a linear probability model, we model the success probability linearly as $\text{Pr}(y = 1|x) = x\beta + \varepsilon$. Thus $\text{Var}(y|x) = \text{Var}(\varepsilon|x) = \text{Pr}(y = 1|x)[1 - \text{Pr}(y = 1|x)]$. So the conditional variance is guaranteed to be a function of x , meaning that we are guaranteed to have heteroscedasticity. Robust standard errors are used to correct for this.

2. (40 points)

- (a) Since the number of packs smoked per day has a “natural” ordering, an ordered setup seems appropriate in this case. On the other hand, we might expect to see decreasing numbers as we move to higher categories, and this is not the case in our data—more people are in category 2 than in category 1. This aspect of the data might suggest that the natural ordering discussed above is not really appropriate.
- (b) Because ordered and unordered models are not nested, we cannot use the usual specification test. Instead we can use Vuong’s test. The numerator of the test statistic is

¹This can all be simplified quite a bit.

the difference in the maximized log-likelihoods between the two models. If we want to compare the ordered probit to the multinomial logit, then the difference is

$$\mathcal{L}_{OP} - \mathcal{L}_{MNL} = -844.15025 - (-827.65487) = -16.49538.$$

Thus the *direction* of Vuong's test is clear—the multinomial logit model appears better. However, in order to determine the statistical significance of this result, we need more information than is supplied in the Stata output.

- (c) Part (b) implies that our initial intuition—that the number of packs smoked per day has a natural ordering—may not be correct. In this case, an unordered model may be better suited to model our data.
- (d) Suppose the ordered probit model is the correct specification. Then if we truncate the categorical left-hand-side variable into a dichotomous variable and estimate a standard binary probit model, the estimated coefficients should be (asymptotically) the same as those for the ordered model (because both models are consistent and asymptotically normal). However, if the ordered probit is not the right specification, then we would expect to see the coefficients change between the probit and ordered probit models. One specification test that we could use would be to test for statistically significant differences between the estimated β 's for the probit and ordered probit models.

Furthermore, since the truncated ordered probit is consistent and asymptotically normal (CAN) while the full ordered probit model is consistent, asymptotically normal, and efficient, we could use a Hausman-type test for our specification.

However we test the changes in the coefficient estimates, it looks like we would not reject the ordered probit model since the actual changes are minimal.

3. (20 points)

- (a) In the ordered probit model we can usually only determine the effects of each of the regressors on the probability of the lowest and highest categories by examining the coefficients alone. Since $\partial \Pr(y = 0)/\partial x_j = -\beta_j \phi(\alpha_1 - x\beta)$ and $\phi(z) > 0$ for all z , we know that $\partial \Pr(y = 0)/\partial x_j$ is the opposite of the sign of β_j . Similarly, if J is the highest category, since $\partial \Pr(y = J)/\partial x_j = \beta_j \phi(\alpha_J - x\beta)$, we know that $\partial \Pr(y = J)/\partial x_j$ has the same sign as β_j .

For intermediate categories, the marginal effects are $\partial \Pr(y = k)/\partial x_j = \beta_j [\phi(\alpha_{k-1} - x\beta) - \phi(\alpha_k - x\beta)]$. Using the sample means, $\bar{x}\hat{\beta} = -81347.48$. Since $\max_j \{|\alpha_j|\} < |\bar{x}\hat{\beta}|$ in this case, the sign of $[\phi(\alpha_{k-1} - x\beta) - \phi(\alpha_k - x\beta)]$ is completely determined by the size of $|\alpha_{k-1} - x\beta|$ relative to $|\alpha_k - x\beta|$ (if $|\alpha_{k-1} - x\beta| < |\alpha_k - x\beta|$ then the marginal effect has the same sign as β_j , otherwise it has the opposite sign). Thus $|\hat{\alpha}_1 - \bar{x}\hat{\beta}| = |-.402 + 81347.48| < |\hat{\alpha}_2 - \bar{x}\hat{\beta}| = |-.044 + 81347.48|$, so the effect on the second category will have the *same* sign as β_j . For the third category, $|\hat{\alpha}_2 - \bar{x}\hat{\beta}| = |-.044 + 81347.48| < |\hat{\alpha}_3 - \bar{x}\hat{\beta}| = |.819 + 81347.48|$, so the effect will have the *same* sign as β_j . Putting this information together, we have the following directions for the marginal effects:

	$\Pr(y = 0)$	$\Pr(y = 1)$	$\Pr(y = 2)$	$\Pr(y = 3)$
<i>educ</i>	+	-	-	-
<i>white</i>	-	+	+	+
<i>age</i>	+	-	-	-
<i>income</i>	-	+	+	+

- (b) In the multinomial logit model, the marginal effect of x_m on the probability that an individual chooses option j is

$$\frac{\partial \Pr(y = j|x)}{\partial x_m} = \Pr(y = j|x) \left[\beta_{jm} - \sum_{k=0}^J \Pr(y = k) \beta_{km} \right].$$

These marginal effects cannot be computed (at least not very easily) without additional Stata output or a calculator. Furthermore, the coefficients from the multinomial logit and ordered probit models are not directly comparable. Hence no comparison can be drawn.