CHAPTER 17

SOLUTIONS TO PROBLEMS

17.1 (i) Let m_0 denote the number (not the percent) correctly predicted when $y_i = 0$ (so the prediction is also zero) and let m_1 be the number correctly predicted when $y_i = 1$. Then the proportion correctly predicted is $(m_0 + m_1)/n$, where n is the sample size. By simple algebra, we can write this as $(n_0/n)(m_0/n_0) + (n_1/n)(m_1/n_1) = (1 - \bar{y})(m_0/n_0) + \bar{y}(m_1/n_1)$, where we have used the fact that $\bar{y} = n_1/n$ (the proportion of the sample with $y_i = 1$) and $1 - \bar{y} = n_0/n$ (the proportion of the sample with $y_i = 0$). But m_0/n_0 is the proportion correctly predicted when $y_i = 0$, and m_1/n_1 is the proportion correctly predicted when $y_i = 1$. Therefore, we have

$$(m_0 + m_1)/n = (1 - \bar{y})(m_0/n_0) + \bar{y}(m_1/n_1).$$

If we multiply through by 100 we obtain

$$\hat{p} = (1 - \bar{y})\hat{q}_0 + \bar{y}\cdot\hat{q}_1,$$

where we use the fact that, by definition, $\hat{p} = 100[(m_0 + m_1)/n]$, $\hat{q}_0 = 100(m_0/n_0)$, and $\hat{q}_1 = 100(m_1/n_1)$.

(ii) We just use the formula from part (i): $\hat{p} = .30(80) + .70(40) = 52$. Therefore, overall we correctly predict only 52% of the outcomes. This is because, while 80% of the time we correctly predict $y = 0$, $y_i = 0$ accounts for only 30 percent of the outcomes. More weight (.70) is given to the predictions when $y_i = 1$, and we do much less well predicting that outcome (getting it right only 40% of the time).

17.3 (i) We use the chain rule and equation (17.23). In particular, let $x_1 \equiv \log(z_1)$. Then, by the chain rule,

$$\frac{\partial E(y \mid y>0,\mathbf{x})}{\partial z_1} = \frac{\partial E(y \mid y>0,\mathbf{x})}{\partial x_1}\cdot\frac{\partial x_1}{\partial z_1} = \frac{\partial E(y \mid y>0,\mathbf{x})}{\partial x_1}\cdot\frac{1}{z_1},$$

where we use the fact that the derivative of $\log(z_1)$ is $1/z_1$. When we plug in (17.23) for $\partial E(y|y > 0,\mathbf{x})/\partial x_1$, we obtain the answer.

(ii) As in part (i), we use the chain rule, which is now more complicated:

$$\frac{\partial E(y \mid y>0,\mathbf{x})}{\partial z_1} = \frac{\partial E(y \mid y>0,\mathbf{x})}{\partial x_1}\cdot\frac{\partial x_1}{\partial z_1} + \frac{\partial E(y \mid y>0,\mathbf{x})}{\partial x_2}\cdot\frac{\partial x_2}{\partial z_1},$$

where $x_1 = z_1$ and $x_2 = z_1^2$. But $\partial E(y|y > 0,\mathbf{x})/\partial x_1 = \beta_1\{1 - \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\}$, $\partial E(y|y > 0,\mathbf{x})/\delta x_2 = \beta_2\{1 - \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\}$, $\partial x_1/\partial z_1 = 1$, and $\partial x_2/\partial z_1 = 2z_1$. Plugging these into the first formula and rearranging gives the answer.

17.5 (i) *patents* is a count variable, and so the Poisson regression model is appropriate.

(ii) Because β_1 is the coefficient on log(*sales*), β_1 is the elasticity of *patents* with respect to *sales*. (More precisely, β_1 is the elasticity of E(*patents*|*sales*,*RD*) with respect to *sales*.)

(iii) We use the chain rule to obtain the partial derivative of $\exp[\beta_0 + \beta_1\log(sales) + \beta_2 RD + \beta_3 RD^2]$ with respect to *RD*:

$$\frac{\partial E(patents\,|\,sales, RD)}{\partial RD} = (\beta_2 + 2\beta_3 RD)\exp[\beta_0 + \beta_1\log(sales) + \beta_2 RD + \beta_3 RD^2].$$

A simpler way to interpret this model is to take the log and then differentiate with respect to *RD*: this gives $\beta_2 + 2\beta_3 RD$, which shows that the semi-elasticity of *patents* with respect to *RD* is $100(\beta_2 + 2\beta_3 RD)$.

17.7 For the immediate purpose of determining the variables that explain whether accepted applicants choose to enroll, there is not a sample selection problem. The population of interest is applicants accepted by the particular university, and you have a random sample from this population. Therefore, it is perfectly appropriate to specify a model for this group, probably a linear probability model, a probit model, or a logit model, where the dependent variable is a binary variable called something like *enroll*, which is one of a student enrolling at the university. The model can be estimated, by OLS or maximum likelihood, using the random sample from accepted students, and the estimators will be consistent and asymptotically normal. This is a good example of where many data analysts' knee-jerk reaction might be to conclude that there is a sample selection problem, which is why it is important to be very precise about the purpose of the analysis, which requires one to clearly state the population of interest.

If the university is hoping the applicant pool changes in the near future, then there is a potential sample selection problem: the current students that apply may be systematically different from students that may apply in the future. As the nature of the pool of applicants is unlikely to change dramatically over one year, the sample selection problem can be mitigated, if not entirely eliminated, by updating the analysis after each first-year class has enrolled.

SOLUTIONS TO COMPUTER EXERCISES

C17.1 (i) If *spread* is zero, there is no favorite, and the probability that the team we (arbitrarily) label the favorite should have a 50% chance of winning.

(ii) The linear probability model estimated by OLS gives

$$\widehat{favwin} = .577 + .0194\, spread$$
$$\qquad\quad (.028) \quad (.0023)$$
$$\qquad\quad [.032] \quad [.0019]$$

$$n = 553, \quad R^2 = .111,$$

where the usual standard errors are in (·) and the heteroskedasticity-robust standard errors are in [·]. Using the usual standard error, the t statistic for H_0: $\beta_0 = .5$ is $(.577 - .5)/.028 = 2.75$, which leads to rejecting H_0 against a two-sided alternative at the 1% level (critical value ≈ 2.58). Using the robust standard error reduces the significance but nevertheless leads to strong rejection of H_0 at the 2% level against a two-sided alternative: $t = (.577 - .5)/.032 \approx 2.41$ (critical value ≈ 2.33).

(iii) As we expect, *spread* is very statistically significant using either standard error, with a t statistic greater than eight. If $spread = 10$, the estimated probability that the favored team wins is $.577 + .0194(10) = .771$.

(iv) The probit results are given in the following table:

Dependent Variable: *favwin*	
Independent Variable	Coefficient (Standard Error)
spread	.0925 (.0122)
constant	–.0106 (.1037)
Number of Observations	553
Log Likelihood Value	–263.56
Pseudo *R*-Squared	.129

In the probit model

$$P(favwin = 1|spread) = \Phi(\beta_0 + \beta_1 spread),$$

where $\Phi(\cdot)$ denotes the standard normal cdf, if $\beta_0 = 0$ then

$$P(favwin = 1|spread) = \Phi(\beta_1 spread)$$

and, in particular, $P(favwin = 1|spread = 0) = \Phi(0) = .5$. This is the analog of testing whether the intercept is .5 in the LPM. From the table, the t statistic for testing H_0: $\beta_0 = 0$ is only about -.102, so we do not reject H_0.

(v) When *spread* = 10 the predicted response probability from the estimated probit model is $\Phi[-.0106 + .0925(10)] = \Phi(.9144) \approx .820$. This is somewhat above the estimate for the LPM.

(vi) When *favhome*, *fav25*, and *und25* are added to the probit model, the value of the log-likelihood becomes –262.64. Therefore, the likelihood ratio statistic is 2[–262.64 – (–263.56)] = 2(263.56 – 262.64) = 1.84. The *p*-value from the χ_3^2 distribution is about .61, so *favhome*, *fav25*, and *und25* are jointly very insignificant. Once *spread* is controlled for, these other factors have no additional power for predicting the outcome.

C17.3 (i) Out of 616 workers, 172, or about 28%, have zero pension benefits. For the 444 workers reporting positive pension benefits, the range is from \$7.28 to \$2,880.27. Therefore, we have a nontrivial fraction of the sample with $pension_t = 0$, and the range of positive pension benefits is fairly wide. The Tobit model is well-suited to this kind of dependent variable.

(ii) The Tobit results are given in the following table:

Dependent Variable: *pension*		
Independent Variable	(1)	(2)
exper	5.20	4.39
	(6.01)	(5.83)
age	–4.64	–1.65
	(5.71)	(5.56)
tenure	36.02	28.78
	(4.56)	(4.50)
educ	93.21	106.83
	(10.89)	(10.77)
depends	(35.28	41.47
	(21.92)	(21.21)
married	(53.69	19.75
	(71.73)	(69.50)
white	144.09	159.30
	(102.08)	(98.97)
male	308.15	257.25
	(69.89)	(68.02)
union	———	439.05
		(62.49)
constant	–1,252.43	–1,571.51
	(219.07)	(218.54)
Number of Observations	616	616
Log Likelihood Value	–3,672.96	–3648.55
$\hat{\sigma}$	677.74	652.90

In column (1), which does not control for *union*, being white or male (or, of course, both) increases predicted pension benefits, although only *male* is statistically significant ($t \approx 4.41$).

(iii) We use equation (17.22) with *exper* = *tenure* = 10, *age* = 35, *educ* = 16, *depends* = 0, *married* = 0, *white* = 1, and *male* = 1 to estimate the expected benefit for a white male with the given characteristics. Using our shorthand, we have

$$\mathbf{x}\hat{\boldsymbol{\beta}} = -1{,}252.5 + 5.20(10) - 4.64(35) + 36.02(10) + 93.21(16) + 144.09 + 308.15 = 940.90.$$

Therefore, with $\hat{\sigma}$ = 677.74 we estimate E(*pension*|**x**) as

$$\Phi(940.9/677.74)\cdot(940.9) + (677.74)\cdot\phi(940.9/677.74) \approx 966.40.$$

For a nonwhite female with the same characteristics,

$$\mathbf{x}\hat{\boldsymbol{\beta}} = -1{,}252.5 + 5.20(10) - 4.64(35) + 36.02(10) + 93.21(16) = 488.66.$$

Therefore, her predicted pension benefit is

$$\Phi(488.66/677.74)\cdot(488.66) + (677.74)\cdot\phi(488.66/677.74) \approx 582.10.$$

The difference between the white male and nonwhite female is 966.40 – 582.10 = $384.30.

(iv) Column (2) in the previous table gives the results with *union* added. The coefficient is large, but to see exactly how large, we should use equation (17.22) to estimate E(*pension*|**x**) with *union* = 1 and *union* = 0, setting the other explanatory variables at interesting values. The t statistic on *union* is over seven.

(v) When *peratio* is used as the dependent variable in the Tobit model, *white* and *male* are individually and jointly insignificant. The p-value for the test of joint significance is about .74. Therefore, neither whites nor males seem to have different tastes for pension benefits as a fraction of earnings. White males have higher pension benefits because they have, on average, higher earnings.

C17.5 (i) The Poisson regression results are given in the following table:

Dependent Variable: *kids*		
Independent Variable	Coefficient	Standard Error
educ	–.048	.007
age	.204	.055
age^2	–.0022	.0006
black	.360	.061
east	.088	.053
northcen	.142	.048
west	.080	.066
farm	–.015	.058
othrural	–.057	.069
town	.031	.049
smcity	.074	.062
y74	.093	.063
y76	–.029	.068
y78	–.016	.069
y80	–.020	.069
y82	–.193	.067
y84	–.214	.069
constant	–3.060	1.211
n = 1,129		
$\mathcal{L}$ = –2,070.23		
$\hat{\sigma}$ = .944		

The coefficient on *y82* means that, other factors in the model fixed, a woman's fertility was about 19.3% lower in 1982 than in 1972.

(ii) Because the coefficient on *black* is so large, we obtain the estimated proportionate difference as exp(.36) – 1 $\approx$ .433, so a black woman has 43.3% more children than a comparable nonblack woman. (Notice also that *black* is very statistically significant.)

(iii) From the above table, $\hat{\sigma}$ = .944, which shows that there is actually underdispersion in the estimated model.

(iv) The sample correlation between $kids_i$ and $\widehat{kids_i}$ is about .348, which means the *R*-squared (or at least one version of it), is about $(.348)^2 \approx .121$. Interestingly, this is actually smaller than the *R*-squared for the linear model estimated by OLS. (However, remember that

OLS obtains the highest possible *R*-squared for a linear model, while Poisson regression does not obtain the highest possible *R*-squared for an exponential regression model.)

C17.7 (i) When log(*wage*) is regressed on *educ*, *exper*, $exper^2$, *nwifeinc*, *age*, *kidslt6*, and *kidsge6*, the coefficient and standard error on *educ* are .0999 (se = .0151).

(ii) The Heckit coefficient on *educ* is .1187 (se = .0341), where the standard error is just the usual OLS standard error. The estimated return to education is somewhat larger than without the Heckit corrections, but the Heckit standard error is over twice as large.

(iii) Regressing $\hat{\lambda}$ on *educ*, *exper*, $exper^2$, *nwifeinc*, *age*, *kidslt6*, and *kidsge6* (using only the selected sample of 428) produces $R^2 \approx .962$, which means that there is substantial multicollinearity among the regressors in the second stage regression. This is what leads to the large standard errors. Without an exclusion restriction in the log(*wage*) equation, $\hat{\lambda}$ is almost a linear function of the other explanatory variables in the sample.

C17.9 (i) Out of 660 families in the sample, 248 reported that they wanted none of the ecolabeled apples at set price.

(ii) The distribution is not continuous: there are clear focal points, and rounding. For example, many more people report one pound than either two-thirds of a pound or 1 1/3 pounds. This violates the latent variable formulation underlying the Tobit model, where the latent error has a normal distribution. Nevertheless, we should view Tobit in this context as a way to possibly improve functional form. It may work better than the linear model for estimating the expected demand function.

(iii) The following table contains the Tobit estimates and, for later comparison, OLS estimates of a linear model:

Dependent Variable: *ecolbs*		
Independent Variable	Tobit	OLS (Linear Model)
ecoprc	–5.82	–2.90
	(.89)	(.59)
regprc	5.66	3.03
	(1.06)	(.71)
faminc	.0066	.0028
	(.0040)	(.0027)
hhsize	.130	.054
	(.095)	(.064)
constant	1.00	1.63
	(.67)	(.45)
Number of Observations	660	660
Log Likelihood Value	–1,266.44	———
$\hat{\sigma}$	3.44	2.48
R-squared	.0369	.0393

Only the price variables, *ecoprc* and *regprc*, are statistically significant at the 1% level.

(iv) The variables *faminc* and *hhsize* are jointly insignificant at 1% and 5% level, with a *p*-value of 0.0774.

(v) The signs of the price coefficients accord with basic demand theory: the own-price effect is negative, and the cross price effect for the substitute good (regular apples) is positive.

(vi) The null hypothesis can be stated as H_0: $\beta_1 + \beta_2 = 0$. Define $\theta_1 = \beta_1 + \beta_2$. Then $\hat{\theta}_1 = -.16$. To obtain the *t* statistic, write $\beta_2 = \theta_1 - \beta_1$, plug in, and rearrange. This results in doing Tobit of *ecolbs* on (*ecoprc* – *regprc*), *regprc*, *faminc*, and *hhsize*. The coefficient on *regprc* is $\hat{\theta}_1$ and, of course, we get its standard error: about .59. Therefore, the *t* statistic is about –.27 and *p*-value = .78. We do not reject the null.

(vii) The smallest fitted value is -1.35, while the largest is 2.95.

(viii)The squared correlation between $ecolbs_i$ and $\widehat{ecolbs_i}$ is about .0389. This is one possible *R*-squared measure.

(ix) The linear model estimates are given in the table for part (ii). The OLS estimates are smaller than the Tobit estimates because the OLS estimates are estimated partial effects on

E(*ecolbs*|**x**), whereas the Tobit coefficients must be scaled by the term in equation (17.27). The scaling factor is always between zero and one, and often substantially less than one. The Tobit model does not fit better, at least in terms of estimating E(*ecolbs*|**x**): the linear model *R*-squared is a bit larger (.0393 versus .0389).

(x) This is not a correct statement. We have another case where we have confidence in the ceteris paribus price effects (because the price variables are exogenously set), yet we cannot explain much of the variation in *ecolbs*. The fact that demand for a fictitious product is hard to explain is not very surprising.

C17.11 (i) The fraction of women in the work force is 3,286/5,634 ≈ .583.

(ii) The OLS results using the selected sample are

$$\widehat{\log(wage)} = \underset{(.060)}{.649} + \underset{(.004)}{.099}\, educ + \underset{(.003)}{.020}\, exper - \underset{(.00008)}{.00035}\, exper^2$$

$$- \underset{(.034)}{.030}\, black + \underset{(.036)}{.014}\, hispanic$$

$n = 3{,}286$, $R^2 = .205$.

While the point estimates imply blacks earn, on average, about 3% less and Hispanics about 1.3% more than the base group (non-black, non-Hispanic), neither coefficient is statistically significant – or even very close to statistical significance at the usual levels. The joint *F* test gives a *p*-value of about .63. So, there is little evidence for differences by race and ethnicity once education and experience have been controlled for.

(iii) The coefficient on *nwifeinc* is −.0091, with $t = -13.47$, and the coefficient on *kidlt6* is −.500, with $t = -11.05$. We expect both coefficients to be negative. If a woman's spouse earns more, she is less likely to work. Having a young child in the family also reduces the probability that the woman works. Each variable is very statistically significant. (Not surprisingly, the joint test also yields a *p*-value of essentially zero.)

(iv) We need at least one variable affecting labor force participation that does not have a direct effect on the wage offer. So, we must assume that, controlling for education, experience, and the race/ethnicity variables, other income and the presence of a young children do not affect wage. These propositions could be false if, say, employers discriminate against women who have young children or whose husbands work. Further, if having a young child reduces productivity – through, say, having to take time off for sick children and appointments – then it would be inappropriate to exclude *kidlt6* from the wage equation.

(v) The *t* statistic on the inverse Mills ratio is 1.77, and the *p*-value against the two-sided alternative is .077. With 3,286 observations, this is not a very small *p*-value. The test on $\hat{\lambda}$ does not provide strong evidence against the null hypothesis of no selection bias.

(vi) Just as importantly, the slope coefficients do not change much when the inverse Mills ratio is added. For example, the coefficient on *educ* increases from .099 to .103 – a change within the 95% confidence interval for the original OLS estimate. [The 95% CI is (.092,.106.)]. The changes on the experience coefficients are also pretty small; the Heckman estimates are well within the 95% confidence intervals of the OLS estimates. Superficially, the *black* and *hispanic* coefficients change by larger amounts, but these estimates are statistically insignificant. Based on the wide confidence intervals, we expect rather wide changes in the estimates to even minor changes in the specification.

The most substantial change is in the intercept estimate – from .649 to .539 – but it is hard to know what to make of this. Remember, in this example, the intercept is the estimated value of log(*wage*) for a non-black, non-Hispanic woman with zero years of education and experience. No one in the full sample even comes close to this description. Because the slope coefficients do change somewhat, we cannot say that the Heckman estimates imply a lower average wage offer than the uncorrected estimates. Even if this were true, the estimated marginal effects of the explanatory variables are hardly affected.

C17.13 (i) Using the entire sample, the estimated coefficient on *educ* is .1037, with standard error = .0097.

(ii) 166 observations are lost when we restrict attention to the sample with *educ* < 16. This is about 13.5% of the original sample. The coefficient on *educ* becomes .1182, with standard error = .0126. This is a slight increase in the estimated return to education, and it is estimated less precisely (because we have reduced the sample variation in *educ*).

(iii) If we restrict attention to those with *wage* < 20, we lose 164 observations [about the same number in part (ii)]. But now the coefficient on *educ* is much smaller, .0579, with standard error = .0093.

(iv) If we use the sample in part (iii) but account for the known truncation point, log(20), the coefficient on *educ* is .1060 (standard error = .0168). This is very close to the estimate on the original sample. We obtain a less precise estimate because we have dropped 13.3% of the original sample.

C17.15 (i) The fraction of men employed is about .898 and the fraction abusing alcohol is about .099.

(ii) The simple regression, with heteroskedasticity-robust standard errors, is

$$\widehat{employ} = \underset{(.003)}{.901} - \underset{(.011)}{.028}\, abuse$$

$$n = 9{,}822, \quad R^2 = .0008$$

The intercept implies that if a man does not abuse alcohol, the chance of being employed is .901. If a man does abuse alcohol, the chance is smaller by .028, so it is .873. This makes sense although it may not be a good causal estimate. The two-sided p-value for the robust $t = -2.54$ is about .011, so it is a statistically significant relationship.

(iii) The probit coefficient on *abuse* is $-.148$ with $t = -2.72$ (p-value = .006), so the sign is the same and the statistical significance similar.

Because the only explanatory variable is binary, we have a very simple “saturated” model. Namely, each man is put into one and only one category: he abuses alcohol or does not. Thus, the APE for the probit model must be identical to the OLS estimate. A quick calculation shows that they agree to the seven reported decimal places: $-.0283046$.

(iv) The fitted values from the LPM and probit must be the same. In each case, there are only two fitted values, estimating the probability of being employed with *abuse* = 0 and *abuse* = 1. These estimates are the same no matter which model we use (linear, probit, logit, or some other binary response model). The two fitted values were given in the answer to part (ii) as .901 and .873, respectively (rounded to three decimal places).

(v) The full results are not reported here, but only the effect on *abuse* coefficient is reported. With the many controls added, the magnitude of the coefficient on *abuse* falls some somewhat to about $-.020$ (robust $t = -1.87$). So, the effect on alcohol abuse on the employment probability is smaller but it is still marginally statistically significant (robust two-sided p-value = .061).

(vi) Estimating a probit model with the same explanatory variables from (v) and computing the APE for *abuse* gives an APE of about $-.021$. It is no longer identical to the OLS estimate in the linear model because the model with many covariates is not saturated. But it is close. The probit estimate is somewhat more statistically significant with $t = -1.98$.

(vii) It is not clear that other health indicators should be controlled for. If they are included, it could be a case of “overcontrolling” because certain health problems may be the result of alcohol abuse. If we hold them fixed when making comparisons we could underestimate the total effect of alcohol abuse on employment. On the other hand, health problems not caused by alcohol abuse could be correlated with both employment and alcohol abuse, in which case we would be attributing too much to alcohol abuse.

(viii) The indicator of alcohol abuse may be correlated with unobserved factors that affect employment. Certain kinds of health issues were already mentioned in part (vii). Depression low self-esteem, and not being motivated are other examples. Indicators for whether one’s parents abused alcohol are reasonable attempts at finding instruments for *abuse*, but they likely are not completely unrelated to unobserved factors affecting employment. For example, if one’s mom abused alcohol it could have affected development as a child. We also need to check wither *mothalc* and *fathalc* predict *abuse* after controlling for the other variables. When we regress *abuse* on all of the other explanatory variables in part (v), and also include *mothalc* and *fathalc*, we get the following:

Coefficient (Robust t statistic):

mothalc: .042 (2.29)

fathalc: .048 (4.98)

Robust joint test of significant: F = 16.46, p-value = .0000.

So *mothalc* and *fathalc* are significant predictors of *abuse*, and the signs of the two coefficients make sense.

(ix) When we use *mothalc* and *fathalc* as IVs for *abuse* and estimate the LPM by 2SLS, the coefficient on *abuse* is –.355 (robust t = –2.18). This is a huge change from the OLS estimate of –.020, and if we take it literally, implies an unrealistically large effect: abusing alcohol lowers the probability of employment by .355. However, we should remember that the estimate, while statistically different from zero, is imprecise. The robust 95% confidence interval is –.673 to –.036 (which nevertheless excludes the OLS estimate).

(x) We need to obtain the residuals, $\hat{v}_2$, from the reduced form regression. When we add the residuals to the linear model estimated in part (v), its coefficient is .336 (robust t = 2.20). Thus, we conclude that OLS and 2SLS are statistically different, and the usual conclusion is that *abuse* is endogenous, and so we should rely on 2SLS. The problem here is that *mothalc* and *fathalc* seem like they are not completely exogenous to the employment equation. As discussed around equation (15.19), 2SLS could be worse than OLS in terms of inconsistency – loosely, "asymptotic bias." Unfortunately, we can never know for sure. What we can see is that the point estimates are very different for OLS and 2SLS.