SOLUTIONS TO PROBLEMS

15.1 (i) It has been fairly well established that socioeconomic status affects student performance. The error term $u$ contains, among other things, family income, which has a positive effect on GPA and is also very likely to be correlated with PC ownership.

(ii) Families with higher incomes can afford to buy computers for their children. Therefore, family income certainly satisfies the second requirement for an instrumental variable: it is correlated with the endogenous explanatory variable (see (15.5) with $x = PC$ and $z = faminc$). But as we suggested in part (i), $faminc$ has a positive affect on GPA, so the first requirement for a good IV, (15.4), fails for $faminc$. If we had $faminc$ we would include it as an explanatory variable in the equation; if it is the only important omitted variable correlated with PC, we could then estimate the expanded equation by OLS.

(iii) This is a natural experiment that affects whether some students own computers. Some students who buy computers when given the grant would not have without the grant. (Students who did not receive the grants might still own computers.) Define a dummy variable, $grant$, equal to one if the student received a grant, and zero otherwise. Then, if $grant$ was randomly assigned, it is uncorrelated with $u$. In particular, it is uncorrelated with family income and other socioeconomic factors in $u$. Further, $grant$ should clearly be correlated with PC: the probability of owning a PC should be significantly higher for students receiving grants. Incidentally, if the university gave grant priority to low income students, $grant$ would be negatively correlated with $u$, and IV would be inconsistent.

15.2 (i) It seems reasonable to assume that $dist$ and $u$ are uncorrelated because classrooms are not usually assigned with convenience for particular students in mind.

(ii) The variable $dist$ must be partially correlated with $atndrte$. More precisely, in the reduced form

\[ atndrte = \pi_0 + \pi_1 priGPA + \pi_2 ACT + \pi_3 dist + \nu. \]

we must have $\pi_3 \neq 0$. Given a sample of data we can test $H_0: \pi_3 = 0$ against $H_1: \pi_3 \neq 0$ using a $t$ test.

(iii) We now need instrumental variables for $atndrte$ and the interaction term, $priGPA \cdot atndrte$. (Even though $priGPA$ is exogenous, $atndrte$ is not, and so $priGPA \cdot atndrte$ is generally correlated with $u$.) Under the exogeneity assumption that $E(u|priGPA, ACT, dist) = 0$, any function of $priGPA$, $ACT$, and $dist$ is uncorrelated with $u$. In particular, the interaction $priGPA \cdot dist$ is uncorrelated with $u$. If $dist$ is partially correlated with $atndrte$ then $priGPA \cdot dist$ is partially correlated with $priGPA \cdot atndrte$. So, we can estimate the equation.
\[ \text{stdfnl} = \beta_0 + \beta_1 \text{atndrte} + \beta_2 \text{priGPA} + \beta_3 \text{ACT} + \beta_4 \text{priGPA} \cdot \text{atndrte} + u \]

by 2SLS using IVs \( \text{dist} \), \( \text{priGPA} \), \( \text{ACT} \), and \( \text{priGPA} \cdot \text{dist} \). It turns out this is not generally optimal. It may be better to add \( \text{priGPA}^2 \) and \( \text{priGPA} \cdot \text{ACT} \) to the instrument list. This would give us overidentifying restrictions to test. See Wooldridge (1999, Chapters 5 and 9) for further discussion.

15.3 It is easiest to use (15.10) but where we drop \( \bar{z} \). Remember, this is allowed because \( \sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x}) = \sum_{i=1}^{n} z_i(x_i - \bar{x}) \) and similarly when we replace \( x \) with \( y \). So the numerator in the formula for \( \hat{\beta}_1 \) is

\[ \sum_{i=1}^{n} z_i(y_i - \bar{y}) = \sum_{i=1}^{n} z_i y_i - \left( \sum_{i=1}^{n} z_i \right) \bar{y} = n_1 \bar{y}_1 - n_1 \bar{y}, \]

where \( n_1 = \sum_{i=1}^{n} z_i \) is the number of observations with \( z_i = 1 \) and we have used the fact that \( \left( \sum_{i=1}^{n} z_i y_i \right)/n_1 = \bar{y}_1 \), the average of the \( y_i \) over the \( i \) with \( z_i = 1 \). So far, we have shown that the numerator in \( \hat{\beta}_1 \) is \( n_1(\bar{y}_1 - \bar{y}) \). Next, write \( \bar{y} \) as a weighted average of the averages over the two subgroups:

\[ \bar{y} = (n_0/n)\bar{y}_0 + (n_1/n)\bar{y}_1, \]

where \( n_0 = n - n_1 \). Therefore,

\[ \bar{y}_1 - \bar{y} = ((n - n_1)/n)\bar{y}_1 - (n_0/n)\bar{y}_0 = (n_0/n)(\bar{y}_1 - \bar{y}_0). \]

Therefore, the numerator of \( \hat{\beta}_1 \) can be written as

\[ (n_0n_1/n)(\bar{y}_1 - \bar{y}_0). \]

By simply replacing \( y \) with \( x \), the denominator in \( \hat{\beta}_1 \) can be expressed as \( (n_0n_1/n)(\bar{x}_1 - \bar{x}_0) \). When we take the ratio of these, the terms involving \( n_0 \), \( n_1 \), and \( n \), cancel, leaving

\[ \hat{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0). \]

15.4 (i) The state may set the level of its minimum wage at least partly based on past or expected current economic activity, and this could certainly be part of \( u_t \). Then \( gMIN_t \) and \( u_t \) are correlated, which causes OLS to be biased and inconsistent.

(ii) Because \( gGDP_t \) controls for the overall performance of the U.S. economy, it seems reasonable that \( gGUSMIN_t \) is uncorrelated with the disturbances to employment growth for a particular state.
(iii) In some years, the U.S. minimum wage will increase in such way so that it exceeds the state minimum wage, and then the state minimum wage will also increase. Even if the U.S. minimum wage is never binding, it may be that the state increases its minimum wage in response to an increase in the U.S. minimum. If the state minimum is always the U.S. minimum, then $gMIN_t$ is exogenous in this equation, and we would just use OLS.

15.5 (i) From equation (15.19) with \( \sigma_u = \sigma_x \), plim \( \bar{\beta}_1 = \beta_1 + (.1/.2) = \beta_1 + .5 \), where \( \bar{\beta}_1 \) is the IV estimator. So the asymptotic bias is .5.

(ii) From equation (15.20) with \( \sigma_u = \sigma_x \), plim \( \bar{\beta}_1 = \beta_1 + \text{Corr}(x,u) \), where \( \bar{\beta}_1 \) is the OLS estimator. So we would have to have Corr\((x,u) > .5 \) before the asymptotic bias in OLS exceeds that of IV. This is a simple illustration of how a seemingly small correlation (.1 in this case) between the IV (z) and error (u) can still result in IV being more biased than OLS if the correlation between z and x is weak (.2).

15.6 (i) Plugging (15.26) into (15.22) and rearranging gives

\[
y_1 = \beta_0 + \beta_1(\pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2) + \beta_2 z_1 + u_1
\]

\[
= (\beta_0 + \beta_1 \pi_0) + (\beta_1 \pi_1 + \beta_2) z_1 + \beta_1 \pi_2 z_2 + u_1 + \beta_1 v_2,
\]

and so \( \alpha_0 = \beta_0 + \beta_1 \pi_0 \), \( \alpha_1 = \beta_1 \pi_1 + \beta_2 \), and \( \alpha_2 = \beta_1 \pi_2 \).

(ii) From the equation in part (i), \( \nu_1 = u_1 + \beta_1 v_2 \).

(iii) By assumption, \( u_1 \) has zero mean and is uncorrelated with \( z_1 \) and \( z_2 \), and \( v_2 \) has these properties by definition. So \( \nu_1 \) has zero mean and is uncorrelated with \( z_1 \) and \( z_2 \), which means that OLS consistently estimates the \( \alpha_j \). (OLS would only be unbiased if we add the stronger assumptions \( E(u_1|z_1,z_2) = E(v_2|z_1,z_2) = 0.\))

15.7 (i) Even at a given income level, some students are more motivated and more able than others, and their families are more supportive (say, in terms of providing transportation) and enthusiastic about education. Therefore, there is likely to be a self-selection problem: students that would do better anyway were also more likely to attend a choice school.

(ii) Assuming we have the functional form for \textit{faminf} correct, the answer is yes. Since \( u_1 \) does not contain income, random assignment of grants means that grant designation is not correlated with unobservables such as student ability, motivation, and family support.

(iii) The reduced form is

\[
\text{choice} = \pi_0 + \pi_1 \text{faminf} + \pi_2 \text{grant} + v_2,
\]

and we need \( \pi_2 \neq 0 \). In other words, after accounting for income, the grant
amount must have some affect on choice. This seems reasonable, provided the grant amounts differ within each income class.

(iv) The reduced form for score is just a linear function of the exogenous variables (see Problem 15.6):

\[ \text{score} = \alpha_0 + \alpha_1 \text{faminc} + \alpha_2 \text{grant} + \nu_1. \]

This equation allows us to directly estimate the effect of increasing the grant amount on the test score, holding family income fixed. From a policy perspective this is itself of some interest.

15.8 (i) Family income and background variables, such as parents' education.

(ii) The population model is

\[ \text{score} = \beta_0 + \beta_1 \text{girlhs} + \beta_2 \text{faminc} + \beta_3 \text{meduc} + \beta_4 \text{feduc} + u_1, \]

where the variables are self-explanatory.

(iii) Parents who are supportive and motivated to have their daughters do well in school may also be more likely to enroll their daughters in a girls' high school. It seems likely that girlhs and \( u_1 \) are correlated.

(iv) Let \( \text{numghs} \) be the number of girls' high schools within a 20 mile radius of a girl's home. To be a valid IV for girlhs, numghs must satisfy two requirements: it must be uncorrelated with \( u_1 \) and it must be partially correlated with \( \text{girlhs} \). The second requirement probably holds, and can be tested by estimating the reduced form

\[ \text{girlhs} = \pi_0 + \pi_1 \text{faminc} + \pi_2 \text{meduc} + \pi_3 \text{feduc} + \pi_4 \text{numghs} + \nu_2 \]

and testing \( \text{numghs} \) for statistical significance. The first requirement is more problematical. Girls' high schools tend to locate in areas where there is a demand, and this demand can reflect the seriousness with which people in the community view education. Some areas of a state have better students on average for reasons unrelated to family income and parents' education, and these reasons might be correlated with \( \text{numghs} \). One possibility is to include community-level variables that can control for differences across communities.

15.9 Just use OLS on an expanded equation, where SAT and \( \text{cumGPA} \) are added as proxy variables for student ability and motivation; see Chapter 9.

15.10 (i) Better and more serious students tend to go to college, and these same kinds of students may be attracted to a private and, in particular, a Catholic, high school. This is another example of the self-selection problem.
(ii) A standardized score is a measure of student ability, so this can be used as a proxy variable in an OLS regression. Having this measure in an OLS regression certainly represents an improvement over having no proxies for student ability.

(iii) The first requirement is that CathRel must be uncorrelated with unobserved student motivation and ability (whatever is not captured by any proxies) and other factors in the error term. This holds if growing up as a Catholic (as opposed to attending a Catholic high school) does not make you a better student. It seems reasonable to assume that Catholics do not have more innate ability than non-Catholics. Whether being Catholic is unrelated to student motivation, or preparation for high school, is a thornier issue.

The second requirement is that being Catholic has an effect on attending a Catholic high school, controlling for the other exogenous factors that appear in the structural model. This can be tested.

(iv) Evans and Schwab (1995) find that being Catholic substantially increases the probability of attending a Catholic high school. Further, it seems that assuming CathRel is exogenous in the structural equation is reasonable. See Evans and Schwab (1995) for an in-depth discussion.

15.11 (i) We plug \( x_t^* = x_t - e_t \) into \( y_t = \beta_0 + \beta_1 x_t^* + u_t \):

\[
y_t = \beta_0 + \beta_1 (x_t - e_t) + u_t = \beta_0 + \beta_1 x_t + u_t - \beta_1 e_t
\]

\[= \beta_0 + \beta_1 x_t + v_t,
\]

where \( v_t = u_t - \beta_1 e_t \). By assumption, \( u_t \) is uncorrelated with \( x_t^* \) and \( e_t \); therefore, \( u_t \) is uncorrelated with \( x_t \). Since \( e_t \) is uncorrelated with \( x_t^* \),

\[
E(x_t e_t) = E[(x_t^* + e_t) e_t] = E(x_t^* e_t) + E(e_t^2) = \sigma_e^2
\]

Therefore, with \( v_t \) defined as above, \( Cov(x_t, v_t) = Cov(x_t, u_t) - \beta_1 Cov(x_t, e_t) = -\beta_1 \sigma_e^2 < 0 \) when \( \beta_1 > 0 \).

Because the explanatory variable and the error have negative covariance, the OLS estimator of \( \beta_1 \) has a downward bias [see equation (5.4)].

(ii) By assumption \( E(x_{t-1}^* u_t) = E(e_{t-1} u_t) = E(x_{t-1}^* e_t) = E(e_{t-1} e_t) = 0 \), and so \( E(x_{t-1} u_t) = E(x_{t-1} e_t) = 0 \) because \( x_t = x_t^* + e_t \). Therefore, \( E(x_{t-1} v_t) = E(x_{t-1} u_t) - \beta_1 E(x_{t-1} e_t) = 0. \)

(iii) Most economic time series, unless they represent the first difference of a series or the percentage change, are positively correlated over time. If the initial equation is in levels or logs, \( x_t \) and \( x_{t-1} \) are likely to be positively correlated. If the model is for first differences or percentage changes, there still may be positive or negative correlation between \( x_t \) and \( x_{t-1} \).

(iv) Under the assumptions made, \( x_{t-1} \) is exogenous in

\[
y_t = \beta_0 + \beta_1 x_t + v_t,
\]

as we should in part (ii): \( Cov(x_{t-1}, v_t) = E(x_{t-1} v_t) = 0 \). Second, \( x_{t-1} \) will often be correlated with \( x_t \), and we can check this easily enough by running a regression of \( x_t \) on \( x_{t-1} \). This suggests estimating the equation by instrumental variables, where \( x_{t-1} \) is the IV for \( x_t \). The IV estimator will be consistent for \( \beta_1 \) (and \( \beta_0 \)), and asymptotically normally distributed.
SOLUTIONS TO COMPUTER EXERCISES

15.12 (i) The regression of log(wage) on sibs gives

$$\log(\hat{\text{wage}}) = 6.861 - 0.0279 \text{ sibs}$$

$$(0.022) \quad (0.0059)$$

$$n = 935, \ R^2 = .023.$$

This is a reduced form simple regression equation. It shows that, controlling for no other factors, one more sibling in the family is associated with monthly salary that is about 2.8% lower. The $t$ statistic on sibs is about -4.73. Of course sibs can be correlated with many things that should have a bearing on wage including, as we already saw, years of education.

(ii) It could be that older children are given priority for higher education, and families may hit budget constraints and may not be able to afford as much education for children born later. The simple regression of educ on brthord gives

$$\hat{\text{educ}} = 14.15 - 0.283 \text{ brthord}$$

$$(0.13) \quad (0.046)$$

$$n = 852, \ R^2 = .042.$$

(Note that brthord is missing for 83 observations.) The equation predicts that every one unit increase in brthord reduces predicted education by about .28 years. In particular, the difference in predicted education for a first-born and fourth-born child is about .85 years.

(iii) When brthord is used as an IV for educ in the simple wage equation we get

$$\log(\hat{\text{wage}}) = 5.03 + 0.131 \text{ educ}$$

$$(0.43) \quad (0.032)$$

$$n = 852.$$

(The $R$-squared is negative.) This is much higher than the OLS estimate (.060) and even above the estimate when sibs is used as an IV for educ (.122). Because of missing data on brthord, we are using fewer observations than in the previous analyses.

(iv) In the reduced form

$$\text{educ} = \pi_0 + \pi_1 \text{sibs} + \pi_2 \text{brthord} + \nu,$$

we need $\pi_2 \neq 0$ in order for the $\beta_j$ to be identified. We take the null to be
\[ n = 2,061, \ R^2 = .0059, \]

which shows that predicted IQ score is about 2.6 points higher for a man who grew up near a four-year college. The difference is statistically significant (t statistic \( \approx 3.51 \)).

(iii) When we add \textit{smsa66}, \textit{reg662}, ..., \textit{reg669} to the regression in part (ii), we obtain

\[
\hat{\text{IQ}} = 104.77 \ + \ .348 \ \text{nearc4} + 1.09 \ \text{smsa66} + ... \\
( 1.62) \quad ( .814) \quad ( 0.81)
\]

\[ n = 2,061, \ R^2 = .0626, \]

where, for brevity, the coefficients on the regional dummies are not reported. Now, the relationship between IQ and \textit{nearc4} is much weaker and statistically insignificant. In other words, once we control for region and environment while growing up, there is no apparent link between IQ score and living near a four-year college.

(iv) The findings from parts (ii) and (iii) show that it is important to include \textit{smsa66}, \textit{reg662}, ..., \textit{reg669} in the wage equation to control for differences in access to colleges that might also be correlated with ability.

15.15 (i) The equation estimated by OLS, omitting the first observation, is

\[
\hat{\beta}_1 = 2.37 \ + \ .692 \ \text{inf}_t \\
(0.47) \quad (0.091)
\]

\[ n = 48, \ R^2 = .555. \]

(ii) The IV estimates, where \textit{inf}_{t-1} is an instrument for \textit{inf}_t, are

\[
\hat{\beta}_1 = 1.50 \ + \ .907 \ \text{inf}_t \\
(0.65) \quad ( 1.43)
\]

\[ n = 48, \ R^2 = .501. \]

The estimate on \textit{inf}_t is no longer statistically different from one. (If \( \beta_1 = 1 \), then one percentage point increase in inflation leads to a one percentage point increase in the three-month T-bill rate.)

(iii) In first differences, the equation estimated by OLS is

\[
\Delta \hat{\beta}_1 = .105 \ + \ .211 \ \Delta \text{inf}_t \\
(0.186) \quad (0.073)
\]

\[ n = 48, \ R^2 = .154. \]
This is a much lower estimate than in part (i) or part (ii).

(iv) If we regress $\Delta \ln f_t$ on $\Delta \ln f_{t-1}$ we obtain

$$\Delta \ln f_t = \hat{\beta}_0 + \hat{\beta}_1 \Delta \ln f_{t-1}$$

$$= .088 - .0096 \Delta \ln f_{t-1}$$

$$(.325) \quad (.1266)$$

$n = 47$, $R^2 = .0001$.

Therefore, $\Delta \ln f_t$ and $\Delta \ln f_{t-1}$ are virtually uncorrelated, which means that $\Delta \ln f_{t-1}$ cannot be used as an IV for $\Delta \ln f_t$.

15.16 (i) When we add $\hat{\beta}_2$ to the original equation and estimate it by OLS, the coefficient on $\hat{\beta}_2$ is about $-.057$ with a t statistic of about $-1.08$. Therefore, while the difference in the estimates of the return to education is practically large, it is not statistically significant.

(ii) We now add $\text{nearc2}$ as an IV along with $\text{nearc4}$. (Although, in the reduced form for $\text{educ}$, $\text{nearc2}$ is not significant.) The 2SLS estimate of $\beta_1$ is now $\hat{\beta}_1 = .157$, se($\hat{\beta}_1$) = .053. So the estimate is even larger.

(iii) Let $\hat{\beta}_1$ be the 2SLS residuals. We regress these on all exogenous variables, including $\text{nearc2}$ and $\text{nearc4}$. The $n$-R-squared statistic is $R^2 = .20$. There is one overidentifying restriction, so we compute the p-value from the $\chi^2$ distribution: $p$-value $= P(\chi^2 > 1.20) \approx .55$, so the overidentifying restriction is not rejected.

15.17 (i) Sixteen states executed at least one prisoner in 1991, 1992, or 1993. (That is, for 1993, exec is greater than zero for 16 observations.) Texas had by far the most executions with 34.

(ii) The results of the pooled OLS regression are

$$\hat{\text{mrt}}te = -5.28 - 2.07 \ d93 + .128 \ \text{exec} + 2.53 \ \text{unem}$$

$$= (4.43) \quad (2.14) \quad (.263) \quad (0.78)$$

$n = 102$, $R^2 = .102$, $R^2 = .074$.

The positive coefficient on exec is no evidence of a deterrent effect. Statistically, the coefficient is not different from zero. The coefficient on $\text{unem}$ implies that higher unemployment rates are associated with higher murder rates.

(iii) When we difference (and use only the changes from 1990 to 1993), we obtain

$$\Delta \hat{\text{mrt}}te = .413 - .104 \ \Delta \text{exec} - .067 \ \Delta \text{unem}$$

$$= (.209) \quad (.043) \quad (.159)$$

125
\[ n = 51, R^2 = .110, \bar{R}^2 = .073. \]

The coefficient on \( \Delta \text{exec} \) is negative and statistically significant (p-value \( \approx .02 \) against a two-sided alternative), suggesting a deterrent effect. One more execution reduces the murder rate by about .1, so 10 more executions reduces the murder rate by one (which means one murder per 100,000 people). The unemployment rate variable is no longer significant.

(iv) The regression \( \Delta \text{exec} \) on \( \Delta \text{exec}_{-1} \) yields
\[
\Delta \text{exec} = .350 - 1.08 \Delta \text{exec}_{-1} \\
(0.370) 
(0.17)
\]

\[ n = 51, R^2 = .456, \bar{R}^2 = .444, \]

which shows a strong negative correlation in the change in executions. This means that, apparently, states follow policies whereby if executions were high in the preceding three-year period, they are lower, one-for-one, in the next three-year period.

Technically, to test the identification condition, we should add \( \Delta \text{unem} \) to the regression. But its coefficient is small and statistically very insignificant, and adding it does not change the outcome at all.

(v) When the differenced equation is estimated using \( \Delta \text{exec}_{-1} \) as an IV for \( \Delta \text{exec} \), we obtain
\[
\Delta \text{mr} = .411 - .100 \Delta \text{exec} - .067 \Delta \text{unem} \\
(0.211) 
(0.064) 
(0.159)
\]

\[ n = 51, R^2 = .110, \bar{R}^2 = .073. \]

This is very similar to when we estimate the differenced equation by OLS. Not surprisingly, the most important change is that the standard error on \( \beta_1 \) is now larger and reduces the statistically significance of \( \beta_1 \).

[Instructor's Note: As an illustration of how important a single observation can be, you might want the students to redo this exercise dropping Texas, which accounts for a large fraction of executions. The results are not nearly as significant. Does this mean Texas is an "outlier"? Not necessarily, especially given that we have differenced to remove the state effect. But we reduce the variation in the explanatory variable, \( \Delta \text{exec} \), substantially by dropping Texas.]

15.18 (i) As usual, if \( \text{unem}_t \) is correlated with \( e_t \), OLS will be biased and inconsistent for estimating \( \beta_1 \).

(ii) If \( E(e_t|\text{inf}_{t-1}, \text{unem}_{t-1}, \ldots) = 0 \) then \( \text{unem}_{t-1} \) is uncorrelated with \( e_t \), which means \( \text{unem}_{t-1} \) satisfies the first requirement for an IV in
\[
\Delta \text{inf}_t = \beta_0 + \beta_1 \text{unem}_t + e_t.
\]
(iii) The second requirement for unem_{t-1} to be a valid IV for unem_t is that unem_{t-1} and unem_t must be sufficiently correlated. The regression unem_t on unem_{t-1} yields

\[
\hat{unem}_t = 1.57 + 0.732 unem_{t-1} \\
(0.58) (0.097)
\]

\[ n = 48, R^2 = .554. \]

Therefore, there is a strong, positive correlation between unem_t and unem_{t-1}.

(iv) The expectations-augmented Phillips curve estimated by IV is

\[
\Delta \hat{infr}_t = 0.694 - 0.138 unem_t \\
(1.883) (0.319)
\]

\[ n = 48, R^2 = .048. \]

The IV estimate of \( \beta_1 \) is much less in magnitude than the OLS estimate (-0.543), and \( \hat{\beta}_1 \) is not statistically different from zero. The OLS estimate had a t statistic of about -2.36 [see equation (11.19)].