# Chapter 9 Assessing Studies Based on Multiple Regression

2. (a) When $Y_i$ is measured with error, we have $\tilde{Y}_i = Y_i + w_i$, or $Y_i = \tilde{Y}_i - w_i$. Substituting the 2nd equation into the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ gives $\tilde{Y}_i - w_i = \beta_0 + \beta_1 X_i + u_i$, or $\tilde{Y}_i = \beta_0 + \beta_1 X_i + u_i + w_i$. Thus $v_i = u_i + w_i$.

   (b) (1) The error term $v_i$ has conditional mean zero given $X_i$:

$$E(v_i|X_i) = E(u_i + w_i|X_i) = E(u_i|X_i) + E(w_i|X_i) = 0 + 0 = 0.$$

   (2) $\tilde{Y}_i = Y_i + w_i$ is i.i.d since both $Y_i$ and $w_i$ are i.i.d. and mutually independent; $X_i$ and $\tilde{Y}_j (i \neq j)$ are independent since $X_i$ is independent of both $Y_j$ and $w_j$. Thus, $(X_i, \tilde{Y}_i)$, $i = 1, \ldots, n$ are i.i.d. draws from their joint distribution.

   (3) $v_i = u_i + w_i$ has a finite fourth moment given that both $u_i$ and $w_i$ have finite fourth moments and are mutually independent. So ($X_i$, $v_i$) have nonzero finite fourth moments.

   (c) The OLS estimators are consistent because the least squares assumptions hold.

   (d) Because of the validity of the least squares assumptions, we can construct the confidence intervals in the usual way.

   (e) The answer here is the economists' "On the one hand, and on the other hand." On the one hand, the statement is true: i.i.d. measurement error in $X$ means that the OLS estimators are inconsistent and inferences based on OLS are invalid. OLS estimators are consistent and OLS inference is valid when $Y$ has i.i.d. measurement error. On the other hand, even if the measurement error in $Y$ is i.i.d. and independent of $Y_i$ and $X_i$, it increases the variance of the regression error $(\sigma_v^2 = \sigma_u^2 + \sigma_w^2)$, and this will increase the variance of the OLS estimators. Also, measurement error that is not i.i.d. may change these results, although this would need to be studied on a case-by-case basis.

3. The key is that the selected sample contains only employed women. Consider two women, Beth and Julie. Beth has no children; Julie has one child. Beth and Julie are otherwise identical. Both can earn $25,000 per year in the labor market. Each must compare the $25,000 benefit to the costs of working. For Beth, the cost of working is forgone leisure. For Julie, it is forgone leisure and the costs (pecuniary and other) of child care. If Beth is just on the margin between working in the labor market or not, then Julie, who has a higher opportunity cost, will decide not to work in the labor market. Instead, Julie will work in "home production," caring for children, and so forth. Thus, on average, women with children who decide to work are women who earn higher wages in the labor market.

5 (a)

$$Q = \frac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1} + \frac{\gamma_1 u - \beta_1 v}{\gamma_1 - \beta_1}.$$

and

$$P = \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1} + \frac{u - v}{\gamma_1 - \beta_1}.$$

(b) $$E(Q) = \frac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1},\ \ E(P) = \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1}$$

(c)

$$Var(Q) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1^2\sigma_u^2 + \beta_1^2\sigma_v^2),\ Var(P) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\sigma_u^2 + \sigma_v^2), \text{ and}$$

$$Cov(P, Q) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1\sigma_u^2 + \beta_1\sigma_V^2)$$

(d) (i) $$\hat{\beta}_1 \xrightarrow{p} \frac{Cov(Q, P)}{Var(P)} = \frac{\gamma_1\sigma_u^2 + \beta_1\sigma_V^2}{\sigma_u^2 + \sigma_V^2},\quad \hat{\beta}_0 \xrightarrow{p} E(Q) - E(P)\frac{Cov(P, Q)}{Var(P)}$$

(ii) $\hat{\beta}_1 - \beta_1 \xrightarrow{p} \frac{\sigma_u^2(\gamma_1 - \beta_1)}{\sigma_u^2 + \sigma_V^2} > 0$, using the fact that $\gamma_1 > 0$ (supply curves slope up) and $\beta_1 < 0$ (demand curves slope down).

8. No, for two reasons. First, test scores in California and Massachusetts are for different tests and have different means and variances. [However, converting (9.5) into units for Massachusetts yields the implied regression to *TestScore*(*MA units*) = 740.9 − 1.80 × *STR*, which is similar to the regression using Massachusetts data shown in Column 1 of Table 9.2.] Second, the regression in Column 1 of Table 9.2 has a low $R^2$ suggesting that it will not provide a accurate forecast of test scores.

# Chapter 12 Instrumental Variables Regression

2. (a) When there is only one $X$, we only need to check that the instrument enters the first stage population regression. Since the instrument is $Z = X$, the regression of $X$ onto $Z$ will have a coefficient of 1.0 on $Z$, so that the instrument enters the first stage population regression. Key Concept 4.3 implies corr($X_i$, $u_i$) = 0, and this implies corr($Z_i$, $u_i$) = 0. Thus, the instrument is exogenous.

(b) Condition 1 is satisfied because there are no $W$'s. Key Concept 4.3 implies that condition 2 is satisfied because ($X_i$, $Z_i$, $Y_i$) are i.i.d. draws from their joint distribution. Condition 3 is also satisfied by applying assumption 3 in Key Concept 4.3. Condition 4 is satisfied because of conclusion in part (a).

(c) The TSLS estimator is $\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}}$ using Equation (10.4) in the text. Since $Z_i = X_i$, we have

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} = \frac{s_{XY}}{s_X^2} = \hat{\beta}_1^{OLS}.$$

5. (a) Instrument relevance. $Z_i$ does not enter the population regression for $X_i$

(b) $Z$ is not a valid instrument. $\hat{X}^*$ will be perfectly collinear with $W$. (Alternatively, the first stage regression suffers from perfect multicollinearity.)

(c) $W$ is perfectly collinear with the constant term.

(d) $Z$ is not a valid instrument because it is correlated with the error term.

8. (a) Solving for $P$ yields $P = \frac{\gamma_0 - \beta_0}{\beta_1} + \frac{u_i^d - u_i^s}{\beta_1}$; thus $Cov(P, u^s) = \frac{-\sigma_{u^s}^2}{\beta_1}$

(b) Because $Cov(P,u) \neq 0$, the OLS estimator is inconsistent (see (6.1)).

(c) We need a instrumental variable, something that is correlated with $P$ but uncorrelated with $u^s$. In this case $Q$ can serve as the instrument, because demand is completely inelastic (so that $Q$ is not affected by shifts in supply). $\gamma_0$ can be estimated by OLS (equivalently as the sample mean of $Q_i$).

9. (a) There are other factors that could affect both the choice to serve in the military and annual earnings. One example could be education, although this could be included in the regression as a control variable. Another variable is “ability” which is difficult to measure, and thus difficult to control for in the regression.

(b) The draft was determined by a national lottery so the choice of serving in the military was random. Because it was randomly selected, the lottery number is uncorrelated with individual characteristics that may affect earning and hence the instrument is exogenous. Because it affected the probability of serving in the military, the lottery number is relevant.