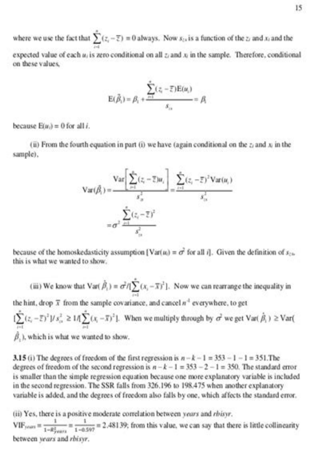

15

where we use the fact that $\sum_{i=1}^{n}(z_i-\bar{z})=0$ always. Now s_{zx} is a function of the z_i and x_i and the expected value of each u_i is zero conditional on all z_i and x_i in the sample. Therefore, conditional on these values,

$$E(\tilde{\beta}_1)=\beta_1+\frac{\sum_{i=1}^{n}(z_i-\bar{z})E(u_i)}{s_{zx}}=\beta_1$$

because $E(u_i)=0$ for all i.

(ii) From the fourth equation in part (i) we have (again conditional on the z_i and x_i in the sample),

$$\text{Var}(\tilde{\beta}_1)=\frac{\text{Var}\left[\sum_{i=1}^{n}(z_i-\bar{z})u_i\right]}{s_{zx}^2}=\frac{\sum_{i=1}^{n}(z_i-\bar{z})^2\text{Var}(u_i)}{s_{zx}^2}$$
$$=\sigma^2\frac{\sum_{i=1}^{n}(z_i-\bar{z})^2}{s_{zx}^2}$$

because of the homoskedasticity assumption [$\text{Var}(u_i)=\sigma^2$ for all i]. Given the definition of s_{zx}, this is what we wanted to show.

(iii) We know that $\text{Var}(\hat{\beta}_1)=\sigma^2/[\sum_{i=1}^{n}(x_i-\bar{x})^2]$. Now we can rearrange the inequality in the hint, drop $\bar{x}$ from the sample covariance, and cancel n^{-1} everywhere, to get $[\sum_{i=1}^{n}(z_i-\bar{z})^2]/s_{zx}^2\geq 1/[\sum_{i=1}^{n}(x_i-\bar{x})^2]$. When we multiply through by σ^2 we get $\text{Var}(\tilde{\beta}_1)\geq\text{Var}(\hat{\beta}_1)$, which is what we wanted to show.

3.15 (i) The degrees of freedom of the first regression is $n-k-1=353-1-1=351$.The degrees of freedom of the second regression is $n-k-1=353-2-1=350$. The standard error is smaller than the simple regression equation because one more explanatory variable is included in the second regression. The SSR falls from 326.196 to 198.475 when another explanatory variable is added, and the degrees of freedom also falls by one, which affects the standard error.

(ii) Yes, there is a positive moderate correlation between *years* and *rbisyr*. $\text{VIF}_{years}=\frac{1}{1-R^2_{years}}=\frac{1}{1-0.597}=2.48139$; from this value, we can say that there is little collinearity between *years* and *rbisyr*.