Suggested Answers for Problem Set 3

Nov. 22, 2002

1 In this case, $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} = 0$ since $\bar{y} = \frac{\sum y_i}{n} = 0$ and $\bar{x} = 0$. Thus, the regression line passes through the origin.

2 Note that the mean of X^* and Y^* are both zero.

$$\begin{aligned}
\hat{\alpha}_1 &= \bar{Y}^* - \hat{\alpha}_2 \bar{X}^* = 0 \\
\hat{\alpha}_2 &= \frac{\sum X_i^* Y_i^*}{\sum X_i^{*2}} = \frac{\sum \frac{x_i}{S_x}\frac{y_i}{S_y}}{\sum (\frac{x_i}{S_x})^2} = \frac{S_x}{S_y}\frac{\sum x_i y_i}{\sum x_i^2} = \frac{S_x}{S_y}\hat{\beta}_2 = \frac{\sqrt{\sum x_i^2/(n-1)}}{\sqrt{\sum y_i^2/(n-1)}}\hat{\beta}_2 \\
&= \sqrt{\frac{\sum \hat{\beta}_2^2 x_i^2}{\sum y_i^2}} = r
\end{aligned}$$

3 By definition, we have $\ln Y_i^* = \ln w_1 + \ln Y_i$ and $\ln X_i^* = \ln w_2 + \ln X_i$. Since $\ln w_1$ and $\ln w_2$ are constants, let them be c_1 and c_2. Now the first model becomes $(\ln Y_i + c_1) = \alpha_1 + \alpha_2(\ln X_i + c_2) + u_i^*$. Therefore,

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum(\ln X_i - \ln \bar{X}_i)(\ln Y_i - \ln \bar{Y}_i)}{\sum(\ln X_i - \ln \bar{X}_i)^2} \\
\hat{\alpha}_2 &= \frac{\sum[\ln X_i + c_2 - (\ln \bar{X}_i + c_2)][\ln Y_i + c_1 - (\ln \bar{Y}_i + c_1]}{\sum[\ln X_i + c_2 - (\ln \bar{X}_i + c_2)]^2} = \hat{\beta}_2 \\
\hat{\beta}_1 &= \ln \bar{Y} - \hat{\beta}_2 \ln \bar{X} \\
\hat{\alpha}_1 &= \ln \bar{Y} + c_1 - \hat{\alpha}_2(\ln \bar{X} + c_2) = \hat{\beta}_1 + c_1 - \hat{\beta}_2 c_2
\end{aligned}$$

4a 15,238 observations, 10,161 men and 5,077 women. Proportions of females in the three years are 25.94%, 32.86% and 38.26%.

4b men: 9.36%, significant at 1%.

women: 14.14%, significant at 1%.

4c For men, being married increases Y by 3.88% significantly. For women, being married decreases Y by 5.90%.

4d 40.86% and 61.36% for men and women respectively.

4e The coefficients of GE are 1.50%, 9.66% and 14.92% for 1980, 1990 and 2000 respectively. Only the coefficient in 1980 is not significant. It is increasing overtime.