ECON 203, CHALLENGE QUIZ 7, PART ONE

CHARLES ANCEL

Consider a random sample (Y_i, X_i), $i = 1, \dots, n$, from a population such that $m(x) := \mathbb{E}(Y_i|X_i = x) = \beta_0 + \beta_1 x$. Write the linear regression model as $Y_i = \mathbb{E}(Y_i|X_i = x) + U_i = \beta_0 + \beta_1 x + U_i$, where U_i is the random error with $\mathbb{E}(U_i^2) = \sigma^2$. Show that:

$$\widehat{m}(x) = \beta_0 + \beta_1 x + \frac{1}{n}\sum_{i=1}^{n}\left[1 + (x - \bar{X})\frac{(X_i - \bar{X})}{s_X^2}\right]U_i,$$

where $s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Proof. Consider the linear regression model $Y_i = \beta_0 + \beta_1 X_i + U_i$, where:

- Y_i is the dependent variable.
- X_i is the independent variable.
- U_i is the error term with $E(U_i) = 0$ and $Var(U_i) = \sigma^2$.
- β_0 and β_1 are parameters.

The OLS estimators are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.$$

The predicted value for Y_i is:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

The residual U_i is defined as:

$$U_i = Y_i - \hat{Y}_i.$$

To express $\widehat{m}(x)$, substitute $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\widehat{m}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = (\bar{Y} - \hat{\beta}_1\bar{X}) + \hat{\beta}_1 x.$$

Now, consider the expression for $\hat{\beta}_1$ and substitute it to obtain $\widehat{m}(x)$:

$$\widehat{m}(x) = \bar{Y} + \left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{s_X^2}\right)(x - \bar{X}).$$

Express $Y_i - \bar{Y}$ as $\beta_1(X_i - \bar{X}) + U_i$ and substitute back to get:

$$\widehat{m}(x) = \bar{Y} + \left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})(\beta_1 (X_i - \bar{X}) + U_i)}{s_X^2} \right) (x - \bar{X}),$$
$$= \bar{Y} + \beta_1 (x - \bar{X}) + \frac{(x - \bar{X})}{s_X^2} \sum_{i=1}^{n} (X_i - \bar{X}) U_i.$$

Finally, incorporate β_0 and β_1 to isolate U_i:

$$\widehat{m}(x) = \beta_0 + \beta_1 x + \frac{1}{n} \sum_{i=1}^{n} \left[1 + (x - \bar{X}) \frac{(X_i - \bar{X})}{s_X^2} \right] U_i.$$

This equation highlights the adjustment for each observation's residual, factoring in the distance from the mean and normalized by the sample variance, thus showing the detailed derivation of $\widehat{m}(x)$. $\square$