Financial Econometrics
Shangai Jiao Tong University
Summer School – 2025-2026
Problem Set - Lecture 4
With Solutions

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Chapter 6: Linear Regression with Multiple Regressors

Question 1. Data were collected from a random sample of 200 home sales from a community in 2013. Let *Price* denote the selling price (in \$1000s), *BDR* denote the number of bedrooms, *Bath* denote the number of bathrooms, *Hsize* denote the size of the house (in square feet), *Lsize* denote the lot size (in square feet), *Age* denote the age of the house (in years), and *Poor* denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields:

$$\begin{aligned}\widehat{Price} &= 109.7 + 0.567 \cdot BDR + 26.9 \cdot Bath + 0.239 \cdot Hsize \\ &\quad + 0.005 \cdot Lsize + 0.1 \cdot Age - 56.9 \cdot Poor\end{aligned}$$

with $R^2 = 0.85$ and $SER = 45.8$.

(a) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

(b) Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 80 square feet. What is the expected increase in the value of the house?

(c) What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

(d) Compute the $\bar{R}^2$ for the regression.

Solution:

(a) **Converting existing space to bathroom:**

If the bathroom is created from existing space, only $Bath$ changes ($\Delta Bath = 1$), while $Hsize$ remains constant.

Expected increase:

$$\Delta\widehat{Price} = 26.9 \times \Delta Bath = 26.9 \times 1 = \boxed{\$26,900}$$

(Remember: $Price$ is in thousands of dollars.)

(b) **Adding bathroom with new construction (80 sq ft):**

Now both $Bath$ and $Hsize$ change: $\Delta Bath = 1$, $\Delta Hsize = 80$.

Expected increase:

$$\begin{aligned}\Delta\widehat{Price} &= 26.9 \times \Delta Bath + 0.239 \times \Delta Hsize \\ &= 26.9 \times 1 + 0.239 \times 80 \\ &= 26.9 + 19.12 = 46.02\end{aligned}$$

$$\boxed{\$46,020}$$

Interpretation: Adding both the bathroom and square footage provides significantly more value than converting existing space.

(c) **Loss from "poor" condition:**

If condition changes from good to poor: $\Delta Poor = 1$.

Expected change in price:

$$\Delta\widehat{Price} = -56.9 \times \Delta Poor = -56.9 \times 1 = -56.9$$

$$\text{Loss} = \boxed{\$56,900}$$

(d) **Adjusted R^2:**

The formula for adjusted R^2 is:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2)$$

where $n = 200$ (observations) and $k = 6$ (regressors: BDR, Bath, Hsize, Lsize, Age, Poor).

$$\begin{aligned}
\bar{R}^2 &= 1 - \frac{200-1}{200-6-1}(1 - 0.85) \\
&= 1 - \frac{199}{193}(0.15) \\
&= 1 - 1.031 \times 0.15 \\
&= 1 - 0.1547 = \boxed{0.845}
\end{aligned}$$

Alternatively, using the relationship:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{TSS/(n-1)}$$

Note: The adjusted R^2 (0.845) is slightly lower than R^2 (0.85), reflecting the penalty for including multiple regressors. The model explains about 84.5% of the variation in home prices after adjusting for degrees of freedom.

Question 2. (Requires calculus) Consider the regression model:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

for $i = 1, \ldots, n$. (Notice that there is no constant term in the regression.) Following analysis like that used in Appendix 4.2:

(a) Specify the least squares function that is minimized by OLS.

(b) Compute the partial derivatives of the objective function with respect to β_1 and β_2.

(c) Suppose that $\sum_{i=1}^{n} X_{1i}X_{2i} = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^{n} X_{1i}Y_i / \sum_{i=1}^{n} X_{1i}^2$.

(d) Suppose that $\sum_{i=1}^{n} X_{1i}X_{2i} \neq 0$. Derive an expression for $\hat{\beta}_1$ as a function of the data (Y_i, X_{1i}, X_{2i}), $i = 1, \ldots, n$.

(e) Suppose that the model includes an intercept: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Show that the least squares estimators satisfy $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$.

(f) As in (e), suppose that the model contains an intercept. Also suppose that $\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^n (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) / \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2$. How does this compare to the OLS estimator of β_1 from the regression that omits X_2?

Solution:

(a) **Least squares objective function:**

OLS minimizes the sum of squared residuals:

$$S(\beta_1, \beta_2) = \sum_{i=1}^n (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i})^2$$

(b) **First-order conditions:**

Taking partial derivatives and setting equal to zero:

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n X_{1i}(Y_i - \beta_1 X_{1i} - \beta_2 X_{2i}) = 0$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum_{i=1}^n X_{2i}(Y_i - \beta_1 X_{1i} - \beta_2 X_{2i}) = 0$$

Simplifying (dividing by -2):

$$\sum_{i=1}^n X_{1i} Y_i = \beta_1 \sum_{i=1}^n X_{1i}^2 + \beta_2 \sum_{i=1}^n X_{1i} X_{2i} \qquad (1)$$

$$\sum_{i=1}^n X_{2i} Y_i = \beta_1 \sum_{i=1}^n X_{1i} X_{2i} + \beta_2 \sum_{i=1}^n X_{2i}^2 \qquad (2)$$

(c) **When $\sum X_{1i} X_{2i} = 0$ (orthogonal regressors):**

From equation (1) with $\sum X_{1i} X_{2i} = 0$:

$$\sum_{i=1}^n X_{1i} Y_i = \hat{\beta}_1 \sum_{i=1}^n X_{1i}^2 + \hat{\beta}_2 \cdot 0$$

Solving:

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^n X_{1i} Y_i}{\sum_{i=1}^n X_{1i}^2}}$$

Key insight: When X_1 and X_2 are orthogonal (uncorrelated when centered), the multiple regression coefficient equals the simple regression coefficient.

(d) **General case ($\sum X_{1i}X_{2i} \neq 0$):**

From the normal equations (1) and (2), solve simultaneously.

From (2):

$$\hat{\beta}_2 = \frac{\sum X_{2i}Y_i - \hat{\beta}_1 \sum X_{1i}X_{2i}}{\sum X_{2i}^2}$$

Substituting into (1):

$$\sum X_{1i}Y_i = \hat{\beta}_1 \sum X_{1i}^2 + \frac{\sum X_{1i}X_{2i}}{\sum X_{2i}^2}\left(\sum X_{2i}Y_i - \hat{\beta}_1 \sum X_{1i}X_{2i}\right)$$

Solving for $\hat{\beta}_1$:

$$\boxed{\hat{\beta}_1 = \frac{\sum X_{1i}Y_i \cdot \sum X_{2i}^2 - \sum X_{2i}Y_i \cdot \sum X_{1i}X_{2i}}{\sum X_{1i}^2 \cdot \sum X_{2i}^2 - (\sum X_{1i}X_{2i})^2}}$$

(e) **Model with intercept:**

For $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, the first-order condition for β_0 is:

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}) = 0$$

This gives:

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{1i} + \hat{\beta}_2 \sum X_{2i}$$

Dividing by n:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2$$

Therefore:

$$\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2}$$

Interpretation: The regression line passes through the point of means $(\bar{X}_1, \bar{X}_2, \bar{Y})$.

(f) **Intercept model with orthogonal demeaned regressors:**

When $\sum(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$, the demeaned regressors are orthogonal.

From part (e), we can write the model in deviations from means. The OLS estimator becomes:

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2}}$$

Comparison with simple regression:

This is **identical** to the OLS estimator from regressing Y on X_1 alone (omitting X_2).

Key insight: When X_1 and X_2 are **uncorrelated** in the sample, including or excluding X_2 does not change the estimated coefficient on X_1. This is because there is no omitted variable bias when the omitted variable is uncorrelated with the included regressors.