1 Multiple Regression

- Most econometric models fall into this category. Bi-variate case is very restrictive.
- Once you consider the case for 2 independent variables, then you can generalize to n independent variables.
- Now looking at the least squares regression line:
  \[ Y_i = a + b_1 X_{1i} + b_2 X_{2i} + e_i \]
  or
  \[ \hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} \]
  or
  \[ Y_i = \hat{a} + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} \]
  where \( \hat{\cdot} \) indicates an estimate.
- Same basic procedure as the bi-variate case. Need to find a minimum for:
  \[ g = g(a, b_1, b_2) = \sum_i e_i^2 \]
  where \( e_i = Y_i - \hat{Y}_i \).
- First order conditions: (see Goldberger)
  \[ \frac{\partial g}{\partial a} = -2 \sum_i e_i = 0 \]
  \[ \frac{\partial g}{\partial b_1} = -2 \sum_i e_i X_{1i} = 0 \]
  \[ \frac{\partial g}{\partial b_2} = -2 \sum_i e_i X_{2i} = 0 \]
- Implies: \( \sum_i e_i = 0, \sum_i e_i X_{1i} = 0, \sum_i e_i X_{2i} = 0. \)
- From 1) we get:
  \[ \hat{a} = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2 \]
From 2) we get:

\[
\hat{b}_1 = \frac{\sum_i x_{1i}y_i - \hat{b}_2 \sum_i x_{1i}x_{2i}}{\sum_i x_{1i}^2}
\]

where \(x_i = \sum_i (X_i - \bar{X})\).

From 3) we get:

\[
\hat{b}_2 = \frac{\sum_i x_{2i}y_i - \hat{b}_1 \sum_i x_{1i}x_{2i}}{\sum_i x_{2i}^2}
\]

Two equations and two unknowns; solving gives:

\[
\hat{b}_1 = \frac{\sum_i x_{1i}y_i \sum_i x_{1i}^2 - \sum_i x_{1i}x_{2i} \sum_i x_{2i}y_i}{\sum_i x_{1i}^2 \sum_i x_{2i}^2 - (\sum_i x_{1i}x_{2i})^2}
\]

\[
\hat{b}_2 = \frac{\sum_i x_{2i}y_i \sum_i x_{1i}^2 - \sum_i x_{1i}x_{2i} \sum_i x_{1i}y_i}{\sum_i x_{1i}^2 \sum_i x_{2i}^2 - (\sum_i x_{1i}x_{2i})^2}
\]

\(\hat{b}_1\) and \(\hat{b}_2\) are known as partial regression coefficients.

A consideration of functional form.

<table>
<thead>
<tr>
<th>Model</th>
<th>Form</th>
<th>Slope = (\frac{dY}{dX})</th>
<th>Elasticity = (\frac{dY}{dX} \cdot \frac{X}{Y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>(Y = a + bX)</td>
<td>(b)</td>
<td>(b)</td>
</tr>
<tr>
<td>Log-linear</td>
<td>(\ln Y = a + b \ln X)</td>
<td>(b\left(\frac{Y}{X}\right))</td>
<td>(b)</td>
</tr>
<tr>
<td>Semi-log</td>
<td>(\ln Y = a + X)</td>
<td>(b\left(\frac{Y}{X}\right))</td>
<td>(\frac{b(X)^*}{X})</td>
</tr>
<tr>
<td>Reciprocal</td>
<td>(Y = a + b\frac{1}{X})</td>
<td>(-b\left(\frac{1}{X^2}\right))</td>
<td>(-b\left(\frac{1}{X^2}\right)^*)</td>
</tr>
</tbody>
</table>

Where * indicates that the elasticity coefficient is variable depending on the value taken by either \(Y\) or \(X\). Usually use the sample mean.