Economics 620, Lecture 2:

Regression Mechanics (Simple Regression)

- Observed variables: \( y_i, x_i \quad i = 1, \ldots, n \)

- Hypothesized (model): \( E y_i = \alpha + \beta x_i \) or \( y_i = \alpha + \beta x_i + (y_i - E y_i) \); renaming we get: \( y_i = \alpha + \beta x_i + \varepsilon_i \)

- Unobserved: \( \alpha, \beta, \varepsilon_i \)

- EXAMPLE: ENGEL CURVES

- Utility function: \( u(z_1, \ldots, z_k) = \sum_{j=1}^{k} a_j \ln z_j. \)

- Budget constraint: \( m = \sum_{j=1}^{k} p_j z_j. \)

- FOC: \( \frac{a_j}{z_j} - \lambda p_j = 0 \quad j = 1, \ldots, k \)

\[ \Rightarrow \lambda = \frac{\sum_{j=1}^{k} a_j}{m} \]

\[ \Rightarrow z_j = \frac{a_j m}{p_j \sum_{\ell=1}^{k} a_{\ell}} \Rightarrow z_j p_j = \frac{a_j}{\sum_{\ell=1}^{k} a_{\ell}} m \]
Estimation

• We want to estimate: \[ E(y) = \alpha + \beta x \]

Where \( y \) is the expenditure on good \( j \) and \( x \) is income.

According to the model we also have:

\[ \beta = a_j / \sum a_{\ell}, \quad \alpha = 0 \]

• We would like to estimate the unknowns from a sample of \( n \) observations on \( y \) and \( x \).
The Least Squares Method

- The Least Squares criterion to estimate $\alpha$ and $\beta$ is to choose $\hat{\alpha}$ and $\hat{\beta}$ to minimize the sum of squared vertical distances between $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ and $y_i$.

- Why do we consider the vertical distances?

- Why do we square?

- Let $S(a, b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2$.

- Partial derivatives:

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n}(y_i - a - bx_i)$$

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} x_i(y_i - a - bx_i).$$
Normal Equations

• Normal equations:

\[ 0 = \sum_{i=1}^{n} (y_i - a - bx_i) \]
\[ 0 = \sum_{i=1}^{n} x_i (y_i - a - bx_i). \]

• \( \hat{\alpha} \) and \( \hat{\beta} \) are the Least Squares (LS) Estimators

\[ \sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i \]  \( \text{(1)} \)
\[ \sum_{i=1}^{n} x_i y_i = \hat{\alpha} \sum_{i=1}^{n} x_i + \hat{\beta} \sum_{i=1}^{n} x_i^2 \]  \( \text{(2)} \)

• From (1):

\[ \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \text{ where } \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} , \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]

• Substituting into (2):

\[ \sum x_i y_i = (\bar{y} - \hat{\beta} \bar{x}) \sum x_i + \hat{\beta} \sum x_i^2 \]
Normal Equations cont’d.

\[ \Rightarrow \sum x_i (y_i - \bar{y}) = \hat{\beta} \left( \sum x_i^2 - \bar{x} \sum x_i \right) \]
\[ = \hat{\beta} \left( \sum x_i^2 - n \bar{x}^2 \right) \]
\[ = \hat{\beta} \sum (x_i - \bar{x})^2 \]

\[ \Rightarrow \hat{\beta} = \frac{\sum x_i (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} \]

\[ \Rightarrow \hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \]

● Is this a minimum? Note that:

\[ \frac{\partial^2 S}{\partial a^2} = 2n; \quad \frac{\partial^2 S}{\partial a \partial b} = 2 \sum x_i; \quad \frac{\partial^2 S}{\partial b^2} = 2 \sum x_i^2 \]
Normal Equations cont’d.

• Is the Hessian p.d.?

• \( H = 2 \left[ \begin{array}{c} n \\ \sum x_i \\ \sum x_i^2 \end{array} \right] \)

• YES! Use Cauchy-Schwarz:
  \[
  \left( \sum x_i z_i \right)^2 \leq \left( \sum x_i^2 \right) \left( \sum z_i^2 \right)
  \]

• Here:
  \[
  \left( \sum x_i \right)^2 \leq \left( \sum x_i^2 \right) n
  \]

• Define the residuals as:
  \[
  e_i = y_i - \hat{\alpha} - \hat{\beta} x_i
  \]

• From the normal equations:
  \[
  \sum e_i = \sum x_i e_i = 0
  \]
Proof of Minimization

- Consider alternative estimators $a^*$ and $b^*$:

$$S(a^*, b^*) = \sum (y_i - a^* - b^* x_i)^2$$

$$= \sum [(y_i - \hat{\alpha} - \hat{\beta} x_i) + (\hat{\alpha} - a^*) + (\hat{\beta} - b^*) x_i]^2$$

$$= \sum e_i^2 + 2(\hat{\alpha} - a^*) \sum e_i + 2(\hat{\beta} - b^*) \sum x_i e_i$$

$$+ \sum [(\hat{\alpha} - a^*) + (\hat{\beta} - b^*) x_i]^2$$

$$\geq \sum e_i^2$$
Properties of Estimators

- LS estimators are unbiased:

\[
\hat{\beta} = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} = \frac{\alpha \sum (x_i - \bar{x}) + \beta \sum (x_i - \bar{x})x_i + \sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2} = \beta + \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2} \Rightarrow E\hat{\beta} = \beta,
\]

\[
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \alpha + (\beta - \hat{\beta})\bar{x} + \bar{\varepsilon} \Rightarrow E\hat{\alpha} = \alpha
\]
More Properties

• We cannot get more properties without further assumptions:

• Assume:

\[ V(y_i|x_i) = V(\varepsilon_i) = \sigma^2, \quad Cov(\varepsilon_i\varepsilon_j) = 0. \]

• Now:

\[
V(\hat{\beta}) = E(\hat{\beta} - \beta)^2 = E \left[ \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2} \right]^2
\]

\[
= \frac{\sum(x_i - \bar{x})^2\sigma^2}{(\sum(x_i - \bar{x})^2)^2},
\]

using \( E\varepsilon_i\varepsilon_j = 0 \). Thus:

\[
V(\hat{\beta}) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}
\]
More Properties cont’d.

- Now for $V(\hat{\alpha})$,

\[
\hat{\alpha} - \alpha = (\beta - \hat{\beta})\bar{x} + \bar{\varepsilon}
\]

\[
\Rightarrow V(\hat{\alpha}) = V(\hat{\beta})\bar{x}^2 + \frac{\sigma^2}{n}
\]

\[
\Rightarrow V(\hat{\alpha}) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]
\]

This requires $Cov(\hat{\beta}, \bar{\varepsilon}) = 0$. Why?

\[
E(\hat{\beta} - \beta)\bar{\varepsilon} = E \left[ \left( \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2} \right) \left( \frac{1}{n} \sum \varepsilon_j \right) \right]
\]

\[
= \frac{\sum (x_i - \bar{x})\sigma^2/n}{\sum (x_i - \bar{x})^2} = 0
\]
Engel Curve Example cont’d.

- We know: \( p_j z_j = \sum a_j m \).

- Is \( V(\varepsilon_j) = \sigma^2 \) plausible here?

- How about logs:

\[
\ln(p_j z_j) = \ln \left( \sum a_j \right) + \ln m.
\]

This implies the regression equation

\[
y = \alpha + \beta x
\]

where \( y \) is log expenditure on good \( j \) and \( x \) is log income.

- What are our expectations about the estimator values?

- Is this better?
Covariance of Estimators

\begin{align*}
\text{Cov}(\hat{\alpha}, \hat{\beta}) &= E[(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)] \\
&= E \left[ ((\beta - \hat{\beta})\bar{x} + \bar{\varepsilon}) \left( \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2} \right) \right] \\
&= -E \left[ \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2} \right]^2 \bar{x} \\
&= \frac{-\sigma^2 \bar{x}}{\sum (x_i - \bar{x})^2}.
\end{align*}
Gauss-Markov Theorem

- The LS estimator is the best linear unbiased estimator (BLUE).

- Proof:

  define \[ w_i = \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \]

  so \[ \hat{\beta} = \sum w_i y_i. \]

  Consider an alternative linear unbiased estimator:

  \[ \tilde{\beta} = \sum c_i y_i. \]

  Write \[ c_i = w_i + d_i. \]

  Note:

  \[ E\tilde{\beta} = \beta \Rightarrow E \sum c_i (\alpha + \beta x_i + \varepsilon_i) = \beta \]

  \[ E \sum c_i (\alpha + \beta x_i + \varepsilon_i) = \alpha \sum c_i + \beta \sum c_i x_i \]

  \[ \Rightarrow \sum c_i = 0; \quad \sum c_i x_i = 1 \]
Gauss-Markov Theorem proof cont’d.

- Note that

\[ w_i \text{ satisfies } \sum w_i = 0; \sum w_i x_i = 1, \text{ so } \sum d_i = 0 \text{ and } \sum d_i x_i = 0. \]

- so

\[
V(\tilde{\beta}) = E \left( \sum c_i \varepsilon_i \right)^2 = \sigma^2 \sum c_i^2 \\
= \sigma^2 \sum (w_i + d_i)^2 \\
= \sigma^2 \left[ \sum d_i^2 + 2 \sum w_i d_i + \sum w_i^2 \right]
\]

Now we have

\[
V(\tilde{\beta}) - V(\hat{\beta}) = \sigma^2 \left[ \sum d_i^2 + 2 \sum w_i d_i \right] \\
= \sigma^2 \sum d_i^2
\]

- WHY?

- This is minimized when the estimators are identical!

- A similar argument applies for \( \hat{\alpha} \) and any linear combination of \( \hat{\alpha} \) and \( \hat{\beta} \).
Estimation of Variance

- It is natural to use the sum of squared residuals to obtain information about the variance.

\[ e_i = y_i - \hat{\alpha} - \hat{\beta}x_i = (y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x}) \]
\[ = -(\hat{\beta} - \beta)(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon}) \]

\[ \Rightarrow \sum e_i^2 = (\hat{\beta} - \beta)^2 \sum (x_i - \bar{x})^2 \]
\[ + \sum(\varepsilon_i - \bar{\varepsilon})^2 - 2(\hat{\beta} - \beta) \sum (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) \]

- This will involve \( \sigma^2 \) in expectation - term by term.

- First term:

\[ E(\hat{\beta} - \beta)^2 \sum (x_i - \bar{x})^2 = \sigma^2 \]
Estimation of Variance cont’d.

- Second term:

\[
E \sum (\varepsilon_i - \bar{\varepsilon})^2 = E \left[ \sum \varepsilon_i^2 + n \left( \frac{1}{n} \sum \varepsilon_i \right)^2 - 2 \sum \varepsilon_i \bar{\varepsilon} \right] = n\sigma^2 + \sigma^2 - 2\sigma^2 = (n - 1)\sigma^2
\]

- Third term:

\[
E \left( \hat{\beta} - \beta \right) \sum (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) = 2E \left[ \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) \right] = 2E \frac{[\sum (x_i - \bar{x}) \varepsilon_i]^2}{\sum (x_i - \bar{x})^2} = 2\sigma^2
\]
Estimation of Variance cont’d.

- Adding the terms we get:
  \[ E \sum e_i^2 = (n - 2) \sigma^2 \]

- This suggest the estimator:
  \[ s^2 = \left( \sum e_i^2 \right) / (n - 2) \]

- This is an unbiased estimator

- It is a quadratic function of y

- This is all we can say without further assumptions
Summing Up

• With the assumption $Ey_i = \alpha + \beta x_i$, we can calculate unbiased estimates of $\alpha$ and $\beta$ (linear in $y_i$).

• Adding the assumption $V(y_i|x_i) = \sigma^2$ and $E\varepsilon_i\varepsilon_j = 0$, we can obtain sampling variance for $\hat{\alpha}$ and $\hat{\beta}$, get an optimality property and an unbiased estimate for $\sigma^2$.

• Note the the optimality property may not be that compelling and that we have very little information about the variance estimate.