Appendix

The "LS" estimator as a solution to a Maximum Entropy Problem

Consider the linear model $y = b_0 + b_1 x + \epsilon$. We make the following assumptions:

(1) $\mathbb{E}[\epsilon] = 0$ Errors have zero mean

(2) $\mathbb{E}[\epsilon^2] = \sigma^2$ Errors have constant variance

(3) $\mathbb{E}[x\epsilon] = 0$ Orthogonality

We define the entropy of a probability distribution $p(\epsilon)$ as:

$$H[p(\epsilon)] = -\int d\epsilon \, [p(\epsilon) \ln p(\epsilon)]$$

where the integration is over the support of ϵ (usually $\mathbb{R}$)

We seek a function $p(\epsilon)$ that maximize H. Using the method of Lagrange multipliers:

$$\mathcal{L} = -\int d\epsilon [p(\epsilon) \ln p(\epsilon)] + \lambda_0 \left[\int d\epsilon \, p(\epsilon) - 1\right] + \lambda_1 \left[\int d\epsilon [\epsilon p(\epsilon)]\right] + \lambda_2 \left[\int d\epsilon [\epsilon^2 p(\epsilon)] - \sigma^2\right] + \lambda_3 \left[\int d\epsilon [x \epsilon p(\epsilon)]\right]$$

(where we have added the constraint that the probability distribution integrate to 1).

Taking derivative of $\mathcal{L}$ w.r.t. $p(\epsilon)$:

$$\frac{\partial L}{\partial p(\epsilon)} = -[1 + \ln p(\epsilon)] + \lambda_0 + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon = 0$$

$-1 - \ln p(\epsilon) + \lambda_0 + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon$

$$\ln p(\epsilon) = \lambda_0 - 1 + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon$$

$$p(\epsilon) = \exp[(\lambda_0 - 1) + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon]$$

Taking derivatives w.r.t. $\lambda_0 \ldots \lambda_3$:

$$\frac{\partial \mathcal{J}}{\partial \lambda_0} = \int d\epsilon [p(\epsilon)] - 1 = 0 \rightarrow \int d\epsilon [p(\epsilon)] = 1$$

$$\frac{\partial \mathcal{J}}{\partial \lambda_1} = \int d\epsilon [\epsilon p(\epsilon)] = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = \int d\epsilon [\epsilon^2 p(\epsilon)] - \sigma^2 = 0 \rightarrow \int d\epsilon [\epsilon^2 p(\epsilon)] = \sigma^2$$

$$\frac{\partial L}{\partial \lambda_3} = \int d\epsilon [x \epsilon] = 0$$

Solving:

$$\int_{-\infty}^{\infty} d\epsilon \left[\exp\left([\lambda_0 - 1] + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon\right)\right] = 1$$

$$\int_{-\infty}^{\infty} \epsilon \exp\left([\lambda_0 - 1] + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon\right) = 0$$

$$\int_{-\infty}^{\infty} \epsilon^2 \exp\left([\lambda_0 - 1] + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon\right) = \sigma^2$$

$$\int_{-\infty}^{\infty} x \exp\left[[\lambda_0 - 1] + \lambda_1 \epsilon + \lambda_2 \epsilon^2 + \lambda_3 x \epsilon\right] = 0$$

Solving this yields:

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\epsilon^2 / 2\sigma^2\right)$$

Note that $\epsilon = y - b_0 - b_1 x$. Thus we can rewrite $p(\epsilon)$ as:

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(y - b_0 - b_1 x)^2}{2\sigma^2}\right]$$

We can now find the values of b_0 and b_1 that maximize the entropy