You did not make the constraints noisy

You're right—I overlooked introducing noise into the constraints. Let's incorporate noisy constraints and derive the maximum entropy estimators for b_0 and b_1.

Problem Setup with Noisy Constraints

We begin with the linear model:

$$y_i = b_0 + b_1 x_i + \epsilon_i$$

The original constraints are:

1. $E[\epsilon_i] = 0$ (errors have zero mean)
2. $E[\epsilon_i^2] = \sigma^2$ (errors have constant variance)
3. $E[x_i \cdot \epsilon_i] = 0$ (errors are orthogonal to the covariates)

To introduce noise into these constraints, we modify them to:

1. $E[\epsilon_i] = \mu_\epsilon$ (where μ_ϵ is a small noise term around zero)
2. $E[\epsilon_i^2] = \sigma^2 + \delta_\epsilon^2$ (where δ_ϵ^2 is a small noise term around σ^2)
3. $E[x_i \cdot \epsilon_i] = \gamma$ (where γ is a small noise term around zero)

Step 1: Maximum Entropy Framework

Given these noisy constraints, we aim to find the probability distribution $p(\epsilon)$ for the errors ϵ_i that maximizes entropy. The entropy of the distribution is:

$$H(p) = -\int p(\epsilon) \log p(\epsilon)\, d\epsilon$$

Step 2: Introduce Lagrange Multipliers

We introduce Lagrange multipliers λ_1, λ_2, and λ_3 for the noisy constraints. The Lagrange function is:

$$\mathcal{L} = -\int p(\epsilon) \log p(\epsilon)\, d\epsilon + \lambda_1 \left(\int p(\epsilon)\epsilon\, d\epsilon - \mu_\epsilon \right) + \lambda_2 \left(\int p(\epsilon)\epsilon^2\, d\epsilon - (\sigma^2 + \delta_\epsilon^2) \right.$$