DiD Estimator Maximum Entropy

To derive an estimator for the Difference-in-Differences (DiD) estimator using the principle of maximum entropy, we'll approach it by focusing on the concept of entropy in the context of information theory and how it can be used to derive estimators that make the least assumptions beyond the observed data.

1. Set Up the Problem

The DiD estimator is used in econometrics to estimate the treatment effect in observational studies where randomization is not possible. It compares the difference in outcomes before and after treatment for a treatment group and a control group.

Let's denote:

- Y_{it} as the outcome variable for unit i at time t.
- D_i as the treatment indicator, where $D_i = 1$ if the unit i is treated, and $D_i = 0$ otherwise.
- T_t as the time indicator, where $T_t = 1$ if the time period is after the treatment, and $T_t = 0$ if before.

The standard DiD estimator is given by:

$$\text{DiD} = \left(\overline{Y}_{1,1} - \overline{Y}_{1,0}\right) - \left(\overline{Y}_{0,1} - \overline{Y}_{0,0}\right)$$

Where:

- $\overline{Y}_{1,1}$ is the average outcome for the treated group after treatment.
- $\overline{Y}_{1,0}$ is the average outcome for the treated group before treatment.
- $\overline{Y}_{0,1}$ is the average outcome for the control group after treatment.
- $\overline{Y}_{0,0}$ is the average outcome for the control group before treatment.

2. Principle of Maximum Entropy

The principle of maximum entropy states that, given a set of constraints, the probability distribution which best represents the current state of knowledge is the one with the largest entropy.

To apply this principle to derive a DiD estimator, consider the joint distribution of the outcomes Y_{it}, the treatment assignment D_i, and the time periods T_t. The entropy of this distribution is:

$$H(P) = -\sum_{i,t} P(Y_{it}, D_i, T_t) \log P(Y_{it}, D_i, T_t)$$

The goal is to maximize this entropy subject to the known constraints of the problem.

3. Constraints

The constraints we consider are:

1. **Average Outcome Before and After Treatment**: The expectations of the outcomes before and after treatment for both treated and control groups are known (can be derived from the data):

$$\mathbb{E}[Y_{it} \mid D_i, T_t] = \overline{Y}_{D_i, T_t}$$

2. **Probability Distributions for Treatment and Time**: The marginal distributions of D_i and T_t should reflect the proportions observed in the data.

$$\sum_i P(D_i = d) = p_d, \quad \sum_t P(T_t = t) = p_t$$

3. **Independence Assumption**: A common assumption in DiD models is that the treatment assignment is independent of the time trends, i.e., D_i and T_t are independent:

$$P(D_i, T_t) = P(D_i)P(T_t)$$

4. Maximizing the Entropy

The problem now is to maximize the entropy:

$$H(P) = -\sum_{i,t} P(Y_{it}, D_i, T_t) \log P(Y_{it}, D_i, T_t)$$

Subject to the constraints on the expected values and the marginal distributions.

Using Lagrange multipliers, the solution to this maximization problem typically takes the form of a distribution that is exponential in the constraints:

$$P(Y_{it}, D_i, T_t) = \frac{1}{Z(\lambda)} \exp\left(\lambda_1 Y_{it} + \lambda_2 D_i + \lambda_3 T_t\right)$$

Where $Z(\lambda)$ is the normalization constant, and $\lambda_1, \lambda_2, \lambda_3$ are the Lagrange multipliers associated with the constraints.

5. Formulating the DiD Estimator

Given this exponential form, the expected difference in outcomes (analogous to the DiD estimate) can be derived by differentiating with respect to the corresponding Lagrange multipliers, leading to:

$$\text{DiD} = \frac{\partial \log Z(\lambda)}{\partial \lambda_1}$$

Given the independence assumption and linearity, this simplifies to:

$$\text{DiD} = (\mathbb{E}[Y \mid D = 1, T = 1] - \mathbb{E}[Y \mid D = 1, T = 0]) - (\mathbb{E}[Y \mid D = 0, T = 1] - \mathbb{E}[Y \mid D = 0, T = 0])$$

Which is exactly the standard Difference-in-Differences estimator.

Conclusion

The maximum entropy approach essentially reaffirms that the standard DiD estimator is consistent with the principle of maximum entropy under the assumption that the treatment assignment is independent of the time trend and that we have no additional information beyond the marginal distributions of the treatment and time indicators and their effect on the outcomes.