Difference-in-differences (DiD) is a popular approach to estimate causal relationships when randomization is not feasible. In the basic set up, a group (the treatment group) is exposed to an intervention (treatment) at some period while another group (the control group) is never exposed to the treatment. Under the assumption that the two groups would have followed “parallel trends” in the absence of the treatment, the causal impact of the treatment can be estimated by comparing the difference in the outcome between the treatment and control groups before and after the intervention.

In practical work, employing the DiD methodology can pose several challenges:

1. Parallel trends assumption violation: If the two groups would not have followed parallel trends without the intervention, the DiD estimation is invalid.
2. Anticipation effects: If units in the treatment group anticipate the intervention and change their behavior beforehand, it can bias the results.
3. Spillover effects: The intervention might indirectly affect the control group, which would bias the results.
4. Time-varying confounders: Factors that change over time and affect the treatment and control group differently can bias the estimates.
5. Heterogenous treatment effects: The impact of the intervention might vary across units or over time, and this variation might be correlated with membership to the control or treatment group.
6. Staggered treatment timing: if the treatment is rolled out in several phases, the standard two-period DiD may not be appropriate.
7. Endogenous selection: Assignment into the treatment group might be associated with unobserved factors affecting the outcome.

This appendix develops an Info-Metrics (Generalized Maximum Entropy) based estimator for DiD that can correct for endogenous selection and heterogenous treatment effects.

# Theory

Consider the standard DiD model:

where: is outcome for individual at time . are parameters (is the parameter of interest), is a group indicator ( for the treatment group, for the control group), is a period indicator ( for the post-treatment period, for the pre-treatment time period) and is an error term. Under traditional DiD assumptions, this equation can be estimated using OLS. However, if the treatment effect varies with characteristics of the treatment unit, is no longer a consistent estimator of the treatment effect. However, in this case the traditional DiD model can be expanded so that:

where: is a vector of covariates, is a vector of parameters representing the direct effect of the covariates on the outcome and is a vector of parameters representing the effect of the covariates through the treatment channel.

In cases where treatment is endogenous (that is the probability of assignment to is correlated with factors affecting the outcome ), the above approaches lead to biased estimates of the treatment effect because the error terms are correlated with the treatment indicator.

We can model the selection into treatment equation through a latent variable approach:

where: is the latent selection variable, is a vector of covariates affecting selection into treatment (which could include elements of as well as additional covariates), is a vector of parameters and is an error term.

We can assume a correlation structure between the selection and treatment effect error terms. For example:

where: and are the standard deviation of the subscripted error terms and is the correlation coefficient.

We can rewrite the above equations in the standard matrix form used in info-metrics:

where:

is a vector of outcomes

is a matrix of regressors (including intercept, treatment group indicator, time period indicator and covariates)

is a vector of parameters

is a vector of errors

For each element in we can define a support space . Then the parameter

where: is the probability associated with the -th support point of .

Similarly, we can define a support space for the error such that:

We wish to pick values of and such that

Subject to constraints

However, the above approach does not correct for selection effects. We can correct for selection using a two-step process. Assume that the error terms associated with the selection equation are normally distributed. We can estimate the selection equation using a probit model and compute an Inverse Mills Ratio for each unit. That is, assume

And estimate this model using probit. Then we can compute the inverse-Mills Ratio:

The DiD model can now be consistently estimated by augmenting the vector of regressors with the IMR.

Simulation Results

I compare the performance of the GME-based DiD estimator vs. the traditional OLS-based DiD estimator, as well as an adjusted OLS-based estimator that takes into account treatment heterogeneity and selection effects. I begin by simulating a dataset with no treatment heterogeneity and no selection. I then simulate a dataset with treatment heterogeneity, data with selection and data with both treatment heterogeneity and endogenous selection. Table 1 shows the details of each dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| Base dataset: no heterogenous treatment effect and no endogenous selection | Dataset 1: with heterogenous treatment effect and no endogenous selection | Dataset 2: with no heterogenous treatment effect and endogenous selection | Dataset 3: with heterogenous treatment effect and endogenous selection |
| True base treatment effect | True base treatment effect | True base treatment effect | True base treatment effect |
| Selection effect | Selection effect | Selection effect | Selection effect |
| Treatment heterogeneity strength | Treatment heterogeneity strength | Treatment heterogeneity strength | Treatment heterogeneity strength |

Table 2 shows the performance of the three different estimators for each model (recall true treatment effect is 2.0)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Base Dataset | Dataset 1 | Dataset 2 | Dataset 3 |
| GME | 1.955 (0.172) | 1.919 (0.187) | 1.949 (0.1768) | 1.827 (0.199) |
| Traditional DiD | 1.814 (0.527) | 1.773 (0.499) | 2.275 (0.531) | 2.119 (0.499) |
| Correctly Specified | 1.971 (0.025) | 1.971 (0.025) | 1.946 (0.014) | 1.946 (0.014) |

If the “correct” specification is available to the researcher, then that DiD estimator can outperform all other DiD estimators. However, when this information is not available, the GME based DiD outperforms traditional OLS-based estimators.