

Solving endogeneity problems - diff in diff

November 17, 2021

Introduction

In modern empirical finance studies, particularly in corporate finance studies, one tries to get around endogeneity problems by utilizing “natural experiments.” Many such natural experiments lead to estimations using a “diff in diff” formulation. We will give the intuition behind such formulations.

1 Endogeneity

In many finance settings, one has problems of endogeneity.

For example, consider the *corporate governance* problem.

One assumes that there is an optimal way for a corporation to be governed. This optimal governance structure will lead to the best possible economic results for the corporation (corporate profits).

How can one think about estimating the optimal governance relation.

Well, how about estimating the link between output (results) and the function of the inputs that produce that output

$$\text{Economic results} = f(\text{Inputs})$$

where inputs are governance variables such as the compensation scheme of the CEO, board structure, leverage structure, industry structure, etc.¹

There are however a number of issues with such formulations that can be summarized as endogeneity.

For example, Demsetz (1983) argues that such a formulation is problematic, because it assumes that we observe “out of equilibrium” solutions to $f()$. However, what if the firm period by period chooses its optimal governance structure? Then our empirical search for the optimal $f()$ is illusory.

On a less philosophical note, it is probably easier to realize that there is a “chicken and egg” problem in such a formulation, since we are usually observing a panel (firms observed year by year). The governance one year is input to the economic results. The economic results one year may lead to changes in governance next year (sacking the CEO). When most of the relevant governance variables are highly persistent, we may have a problem.

These kinds of endogeneity problems are endemic in economic analysis.

One way to solve them is to look for “exogenous” events, where the economic environment changes in ways that was not foreseen by participants in the economy.

So for example, when Norway introduced a 40% “quota” of women in corporate boards, this is an exogenous change to on governance variable, the composition of corporate boards.

How participants react following such exogenous events can be used to tease out properties of the functional relationships of interest (such as the governance relationship $f()$).

2 Differences of Differences (“Diff in Diff”)

A often used tool to look at data around exogenous events is a “diff in diff” analysis.

A “diff in diff” formulation is a standard type of analysis discussed in standard textbooks.²

¹For some US papers doing such estimations, see Demsetz and Lehn (1985), Morck, Shleifer, and Vishny (1988). For Norway, see Bøhren and Ødegaard (2001) (summarized in Bøhren and Ødegaard (2006)).

²E.g. (Angrist and Pischke, 2008, Section 5.2) or (Greene, 2018, Ch 19).

Let us use the notation in (Greene, 2018, pg 933).

$$y_{it} = \theta_t + \mathbf{X}'_{it}\beta + \gamma C_i + u_i + \epsilon_{it}$$

Here y_{it} is the variable to be explained. The dependent variable can be explained by a set of controls \mathbf{X}_{it} . The treatment variable is indicated by the dummy variable C_i . Here u_t is an unobserved individual effect. To get rid of this individual effect analysis is done on the differenced version of this regression

$$\Delta y_{it} = (\theta_1 - \theta_0) + (\Delta \mathbf{X}_{it})'\beta + \gamma \Delta C_i + \Delta \epsilon_{it}$$

The difference of sample averages across the two groups

$$\hat{\gamma} = [\overline{\Delta y} | (\Delta C = 1)] - [\overline{\Delta y} | (\Delta C = 0)]$$

is the *simple difference in differences* estimator.

2.1 Implementing the diff in diff analysis

There is a common way of implementing a diff in diff analysis. The idea is that the observations can be grouped into

- One group affected by an intervention (treated)
- One group not affected by the intervention (nontreated)

Estimation of the diff in diff is based on regressions of the type

$$y = \beta_0 + \beta_1 d_{treated} + \beta_2 d_{time} + \delta d_{treated} \times d_{time} + \alpha \mathbf{X} + \varepsilon \quad (1)$$

where y is the variable of interest (i.e. liquidity), $d_{treated}$ is a dummy variable for whether an element belongs to the treatment or the control group, and d_{time} a time dummy for the second period. The coefficient of interest, δ , multiplies the interaction term, which is the same as a dummy variable equal to one for the observations in the treatment group in the second period. The coefficient δ measures the direct effect of the intervention.

In the regression we allow for additional covariates \mathbf{X} .

One will also typically adjust for the panel data nature of the data by including fixed date and stock effects, and adjusting the standard errors in the panel for clustering.

Note for R users: The panel data adjustments can be easily done using the R library `plm`. The calculation of standard errors in a panel setting in R is described Croissant and Millo (2008).

3 Summarizing

Diff in diff

Potential solution to *endogeneity* problems in economic analysis.

Observed date when an exogenous event happens

Investigate changes in differences around date, including controls.

Rely on

- The event is truly exogenous
- The event is important for the economic decisions of interest

References

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics*. Princeton University Press, 2008.

Øyvind Bøhren and Bernt Arne Ødegaard. Corporate governance and economic performance in Norwegian listed firms. Research Report Nr 11/2001, Norwegian School of Management, November 2001.

Øyvind Bøhren and Bernt Arne Ødegaard. Governance and performance revisited. In Paul Ali and Greg Grogouriu, editors, *International Corporate Governance after Sarbanes-Oxley*, pages 27–64. Wiley, February 2006. doi: 10.1002/9781119201885.ch3.

Yves Croissant and Giovanni Millo. Panel data econometrics in R: the `plm` package, 2008. R vignette, available at CRAN.

Harold Demsetz. The structure of ownership and the theory of the firm. *Journal of Law and Economics*, 26:375–390, 1983.

Harold Demsetz and Kenneth Lehn. The structure of corporate ownership: Causes and consequences. *Journal of Political Economy*, 93:1155–77, 1985.

William H Greene. *Econometric Analysis*. Pearson, eight edition, 2018.

Randall Morck, Andrei Shleifer, and Robert W Vishny. Mangement ownership and market valuation: An empirical analysis. *Journal of Financial Economics*, 20:293–315, 1988.