Competition Memo

Taking the con out of econometrics? New challenges to "Difference-in-Differences" analysis in competition cases

Economic analysis of merger and competition cases has been shaped by the *"credibility revolution"* that revolutionised economics as a discipline and won Angrist, Card and Imbens this year's Nobel Prize.[1] Tools like "difference-in-differences" are now widely used and antitrust authorities increasingly emphasise analysis that identifies causal relationships rather than mere correlations.

However, while these techniques are widely used, the economic toolkit continues to develop. The latest academic research shows that the sorts of analyses frequently used in competition circles can be subject to biases that need to be accounted for. Indeed, these biases can be so severe they can result in *diametrically opposite* results (e.g. a finding that market entry reduces prices when it does not or that it increases prices when it in fact reduces them).

This memo discusses the influence of the credibility revolution on antitrust and new techniques which need to be used when applying difference-in-differences analysis. We begin with the history of these methods and how they have been used in an antitrust context. We then explain the pitfalls which arise when applying difference-in-differences in a "staggered" setting (e.g. with multiple market entries and exits by merging parties in multiple locations over time). Finally, we explain how these issues can be corrected with careful application of the latest techniques.

1. The credibility revolution and difference-in-differences in competition analysis

The Nobel Prize winning contribution of Angrist et al. was to refocus economics as an *empirical* discipline and to popularise new tools which could distinguish *causation* from mere correlation.

Most famously, Card and Krueger used difference-in-differences ("**diff-in-diff**") analysis to look at the impact of minimum wage laws on employment.[2] They used a "natural experiment" in which New Jersey increased its minimum wage while Pennsylvania did not and looked at how low-wage employment at fast food restaurants varied across the two sides of the state border. Contrary to what one would expect based on a simple model of supply and demand, minimum wage laws did *not* reduce employment, a finding that has been corroborated by multiple studies.[3] The volume of empirical work in economics has since exploded relative to traditional theoretical work.[4]

Competition economics has mirrored these developments. The emphasis is on empirical data, not arguments based on high theory and, while it is still common to conduct correlation-based analysis (e.g. to look at how prices or margins differ across customers or markets according to the number of competitors), the drawbacks of such analysis are well recognised[5] and more plausibly causal relationships based on "natural experiments" are increasingly the gold standard.

For this reason, difference-in-differences has become a standard part of the competition economist's tool kit. In merger analysis it has been used to conduct retrospectives of past mergers in an industry (e.g. by the European Commission in INEOS/Solvay)[6] and to look at the impact of store opening and closures among merging retailers (e.g.

---

[1] See: https://www.nobelprize.org/prizes/economic-sciences/2021/press-release/. Other key contributors include Alan Krueger, and Edward Leamer from whose paper this memo takes its title. Leamer, E. E. (1983). Let's take the con out of econometrics, *American Economic Review.*

[2] Card, D, and Krueger, A. (1994). Minimum Wages and Employment: a Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review.*

[3] The hypothesis is that this counterintuitive result reflects either monopsony power on the part of restaurants, "general equilibrium" effects due to increased worker spending power, or that employment effects take longer to manifest through technological change rather than immediate layoffs. A caveat is that an increase in minimum wage to a higher level could have different effects from the increase from $4.25 to $5.05 seen in New Jersey.

[4] A 2013 study found the proportion of economic studies with no empirical content peaked in the 1980s and has declined ever since. Hamermesh, D. S. (2013). Six decades of top economics publishing: who and how?. *Journal of Economic Literature.*

[5] For example, the number of firms in a market could be endogenous and reflect the cost of serving customers. There may be more suppliers in low-cost markets, clouding the estimates.

[6] See Annex A to the Commission's decision. https://ec.europa.eu/competition/mergers/cases/decisions/m6905_20140508_20600_3967414_EN.pdf. See also Ormosi, P. Mariuzzo, F. Havell, R. Fletcher, A. and Lyons, B. (2015). A review of merger decisions in the EU: what can we learn from ex-post evaluations?. *Publication Office of the European Union*.

by the CMA in Ladbrokes/Coral or the EC in Ahold/Delhaize).[7] It has also been used in an antitrust setting (e.g. to look at the impact of the roll out of Google's "Shopping Unit" across Europe).[8]

2. How difference-in-differences can be implemented in competition analysis?

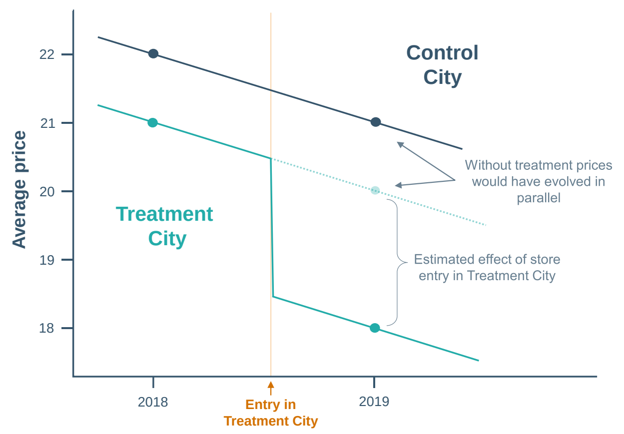
The diff-in-diff approach mirrors the Card and Krueger study discussed above: one compares the evolution of an outcome (e.g. prices) before and after a **"treatment"** (e.g. the entry of a new store or supplier) which affects one part of the sample **"the treatment group"**, but not the other **"the control group"**.[9]

For example, suppose that a supermarket opened a store in the "**Treatment City**" in 2019, while at the same time it was not yet present in a neighbouring "**Control City**". Diff-in-diff compares how prices in both cities evolved before and after the supermarket entered the Treatment City (e.g. by looking at price changes between December 2018 and December 2019 in both cities). If prices fell by more in the Treatment City than the Control City we would attribute this to the effect of entry.

The advantage of diff-in-diff is that it can "control for" other factors which might affect prices in both cities (e.g. changes in demand or input prices). The key assumption is that the price evolution in the Control City is a good proxy for how prices in the Treatment City would have performed if the supermarket did not open a store. The premise of this **"parallel trends assumption"** is that all other changes besides the store opening that might affect prices over time (e.g. changes in demand, weather, supply chain costs) affect both treatment and control group similarly.

In the next Figure, since the average prices in the Treatment City fell by 3 between 2018 and 2019 (from 21 to 18), and prices in the Control City fell by 1 (from 22 to 21), one would conclude the supermarket opening reduced prices by 2.[10]

Difference-in-differences



3. New research raises serious issues with diff-in-diff analysis based on "staggered entry"

The above was a simple example: one event at one point in time affecting one treatment group and one control group. While this will sometimes match reality (e.g. in INEOS/Solvay or Card and Krueger's minimum wage study), often we will have multiple "events" occurring on a "staggered" basis over time. For example, multiple store openings and closures in multiple cities over time.

While it seems natural to extend diff-in-diff to this setting (and indeed this has been done in multiple past competition cases), new research over the last 1-2 years shows such analysis can be seriously flawed,[11] casting doubt on a wealth of past analyses.

To understand why applying diff-in-diff to staggered settings could lead to erroneous conclusions, the Figure below builds on our prior example by having the Control City *also* experience entry, but a year after the Treatment City. Further, we allow for entry to take time to take effect (e.g. because consumers take a while to understand their new options or because the entrant's operations become more efficient over time) such that the 2019 entry in the Treatment City causes further price reductions in 2020.[12]

---

7 The CMA did not refer to difference-in-differences but ran an equivalent two-way fixed effects regression. See Appendix E: https://assets.publishing.service.gov.uk/media/579781a1e5274a31e0000002/ladbrokes-coral-final-report-appendices-and-glossary.pdf

8 See paragraph 479 of the Google Search (Shopping) decision. https://ec.europa.eu/competition/antitrust/cases/dec_docs/39740/39740_14996_3.pdf

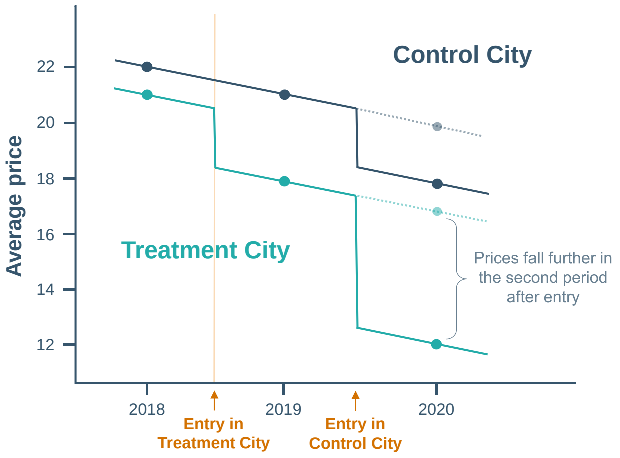
9 The intuition behind difference-in-differences is much older. For example, John Snow's famous analysis of the spread of cholera in the 1850s, which proved it to be a water-borne disease, was based on a similar (albeit less well formalised) approach in which he compared case volumes pre- and post-outbreak in areas around water pumps which had been cleaned vs. those which had not.

10 In the example, the effect is obtained as: (18 - 21) - (22 - 21) = (-3) - (-1) = -2.

11 Some papers that have pointed out issues with diff-in-diff in staggered settings are: De Chaisemartin, C., and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*; Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics;* Sun, L., and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics;* and Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

12 In practice one might expect prices in the Treatment City to fall in a continuous fashion over time rather than the two abrupt drops shown in the Figure. The exact same conceptual issues would arise, however.

“Staggered” difference-in-differences



Issues arise because, while a standard difference-in-differences will just estimate an effect in Treatment City, in staggered settings, the standard econometric approach estimates a combination of the effects on *both* cities.[13]

In the case of the Treatment City, the impact would be estimated as before, i.e. by looking at how prices changed between 2018 and 2019 in both cities. As before, the conclusion would be that prices in Treatment City fell by 2 units one year after the opening of the new store.

But, to estimate the effect of entry in the Control City, the standard econometric approach would compare the price changes in *both* cities between 2019 and 2020. This happens because only the Control City experiences an entry event between 2019 and 2020. Because the only thing that changes across cities is that a new supermarket opened in Control City, the standard approach attributes any difference in the price evolution to the supermarket's entry. Unfortunately, doing so in this case is wrong: prices in the Treatment City are still changing in response to the store that opened a year earlier.

In our example, the approach would mistakenly conclude that prices in the Control City *increased* following entry. This follows because, between 2019 and 2020, prices in Control City fell by 3 (from 21 to 18) while prices in Treatment City fell by 6 (from 18 to 12). As prices in Control City fell by less, one would conclude that the new supermarket caused prices to increase by 3.[14]

In practice, difference-in-differences does not estimate a separate effect for every individual city, but rather an overall average entry effect. But, the issues above affect this average estimate also. For instance, if one used the simple average of the estimated entry effect in the two cities, one would conclude entry caused prices to *increase* by 0.5.[15]

The essence of the issue is that the staggered diff-in diff will sometimes use as the “control group” cities that have already experienced market entry. This breaks the “parallel trends assumption” discussed above and results in potentially bogus comparisons and bogus results. Indeed, as shown above, the effects can be so pronounced that the estimated effect can have the *opposite sign* to the true effect (i.e. the analysis might say entry increases prices when it in fact reduces them or vice-versa).[16]

While this will not always be the case, the latest economic analysis tells us that these counterintuitive effects are most likely when the impact of the event becomes stronger with time or the effect size varies across locations. For example, if store openings have larger effects in some cities than others (e.g. because the stores are in a more attractive location or closer to the other merging party's stores or because of differences in the number of other competitors) or the effects of entry take time to manifest (e.g. because retailers invest to face the new competition and become more efficient with time).[17]

4. What's the solution?

The discussion shows that the traditional approach to staggered diff-in-diff can be seriously flawed, but solutions can be found.

As the problem is caused by using the wrong control group, i.e. observations that have already experienced the event of interest, the solution is intuitive: to exclude as controls observations that have been treated (e.g. have already experienced entry).

In the example above, one could just estimate the effect on Treatment City and ignore the entry effect in Control City. In practice, however, implementing the solution may not be as straightforward. We may have more than two cities in the data, with some experiencing entry at the same time and others seeing entry over a longer window. We may also be interested in estimating the effect of entry over longer time periods. In those cases, how to estimate the effect correctly may not be as obvious.

---

13 The standard econometric approach is to regress price on a variable equal to 1 in periods after the entry event while controlling for city and time specific factors using “fixed effects”.

14 In this case the treatment effect would be estimated as: (18-21)-(12-18)=(-3)-(-6)=3.

15 (-2+3)/2=0.5.

16 For example, Baker, Larcker and Wang (2021) revisited 3 papers previously published in the *Journal of Finance* and *Journal of Financial Economics*. They find that, once they account for the issues with staggered diff-in-diff the evidence no longer supported the papers' findings. See Baker, A., Larcker, D. F., and Wang, C. C. (2021). How Much Should We Trust Staggered Difference-In-Differences Estimates?. *Available at SSRN 3794018*.

17 Another factor that affects the size of the bias is whether or not there is a “clean” control group in the analysis. A clean control group would be a third city in which the new supermarket is never introduced during the period of analysis.

Fortunately, the same literature that has highlighted the issues with staggered diff-in-diff has proposed solutions. Some new methods build on the intuition of estimating diff-in-diff by excluding "treated" observations as controls,[18] or using as a control the last observations to be treated during the period it remains untreated.[19] Other approaches address the bias by, first, estimating time and city specific factors that affect prices using *only* untreated observations, and, then estimate the treatment effect after removing these confounding factors.[20] Finally, some authors suggested to just estimate a more detailed version of the traditional econometric approach.[21]

So, diff-in-diff remains a powerful tool. It just needs to be applied properly taking account of the latest techniques.

5. Conclusion

Last week's Nobel Prize salutes a profound shift in economics towards empirical analysis over theory and in favour of careful analysis of causality over correlation. Like all fields applying economic principles to the real world, antitrust analysis has been influenced and techniques like difference-in-differences are now widespread.

Competition economists should not rest on their laurels, however. Recent developments in the academic literature identify new pitfalls which could easily apply to the analysis done in the context of mergers and antitrust assessments. Solutions are at hand, but empirical analysis needs to be done with care and, just as antitrust economics took on board the lessons of the credibility revolution it needs to also incorporate the latest insights and methodologies.

Dr Oliver Latham, Dr Javier Brugués. The views here are those of the authors.

18 October 2021

---

[18] Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics.* A similar approach was proposed by Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics.*

[19] See Sun, L., and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics.*

[20] See Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419.*

[21] See Wooldridge, J. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *Available at SSRN 3906345.*