

Difference-in-Differences: How It is Used in Retrospective Merger Review and Marketing Analysis

Rui Huang* and Pian Chen** | June 24, 2026

I. Introduction

This article is the second part of our data science crash courses — *Causal Methods in the Courtroom and the C-Suite*. The courses introduce fundamental methods that economists, statisticians, and data scientists use to make causal inference. We write for two groups: (1) lawyers who want to understand whether challenged conduct caused harm and how to quantify damages and (2) business leaders who make product, pricing, and marketing decisions.

When we cannot run a random experiment (discussed in the first part of the data science crash courses), we need methods to construct a counterfactual world from data that we already have. This article focuses on a causal inference method called difference-in-differences (DiD). DiD seeks to identify the impact of conduct (e.g., a merger or an ad campaign) by comparing the change in outcomes (e.g., prices or sales) over time between the treatment group exposed to the conduct and the control group not exposed to the conduct.

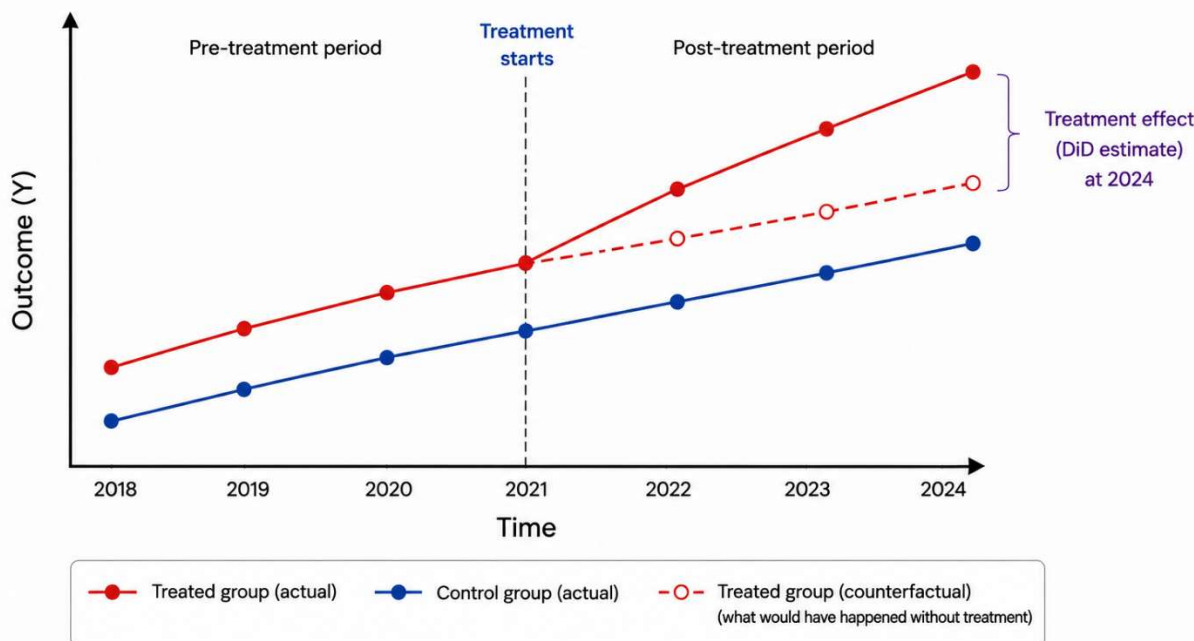
II. The DiD Methodology and The Parallel-Trends Assumption

The idea behind DiD is simple: (1) for a treatment group we measure the difference in the outcome before and after a treatment; (2) for a control group we measure the difference in the outcome over the same period; (3) we then calculate the difference between (1) and (2), i.e., difference-in-differences. The first difference is the difference in outcomes of the treatment group over time, and it captures everything that happened, including the treatment plus everything else moving in the background. The second difference is the difference in the outcomes of the control group over the same period, and it captures everything else moving in the background. When we subtract the second difference from the first difference, what happens in the background cancels out, and what's left is the treatment effect.^[1]

The validity of DiD rests on one critical assumption: **parallel trends**. It means that, absent the treatment, the treated and control groups would have experienced the same change in their average outcomes from before to after the treatment (Angrist and Pischke, 2009; Cunningham, 2021).^[2] The assumption concerns the unobserved post-treatment counterfactual; pre-period parallelism is only the evidence for it, not the assumption itself.^[3]

Parallel Trends Assumption

In the absence of treatment, the treated group would have followed the same trend as the control group.



While one can't prove the assumption because the post-treatment world where the treated group didn't experience treatment doesn't exist, one may attempt to make it plausible by showing that the two groups did move in parallel for some meaningful period prior to the treatment. However, parallel pre-trends are necessary but **not** sufficient. The literature warns specifically against treating "pre-trends look parallel" as proof that they would have continued in parallel.^[4] Rambachan and Roth (2023) recommend treating parallel pre-trends as one input, formally analyzing the result's sensitivity to plausible violations, and triangulating with economic knowledge and alternative control groups.^[5]

III. How DiD is Used in Retrospective Merger Review and Daubert Risk

DiD is a workhorse of empirical antitrust economics, and its flagship application is the merger retrospective review. It is a useful tool when a clean, unaffected benchmark market is available to difference out common shocks. A substantial literature applies DiD to merger analysis: Orley Ashenfelter, Daniel Hosken, Matthew Weinberg, John Kwoka and others have applied DiD to beer, airlines, consumer products, appliances, hospitals, and gasoline mergers where they compared prices in markets the merger affected to prices in markets it didn't.^{[6] [7]}

A concrete example comes from hospital-merger litigation. In February 2004, the FTC challenged Evanston Northwestern Healthcare's acquisition of Highland Park Hospital, which was consummated on January 1, 2000. The agency's expert measured the merger's price effect using DiD: the treatment group was the merged hospitals, and the control group was a set of comparable hospitals that had not merged. Using insurers' transaction-price data, the analysis found that prices at the merged hospitals rose roughly 11 to 18 percentage points more than at

the control hospitals after the merger, and the expert showed the result was robust across alternative control groups, data sources, and case-mix adjustments.^[8] In August 2007, the Commission ruled the acquisition anticompetitive under Section 7 of the Clayton Act. It imposed a conduct remedy (separate contract-negotiating teams) rather than divestiture.^[9]

DiD is admissible and widely accepted by federal courts. The Federal Judicial Center's *Reference Manual on Scientific Evidence* — whose Reference Guide on Multiple Regression is the standard reference courts use to evaluate regression-based economic testimony — covers the regression framework on which DiD is built, and its Fourth Edition (2025) updates that guidance to reflect current empirical practice.^{[10][11]}

Antitrust is the legal domain where economists face Daubert challenges most often. By one recent analysis of Daubert challenges, more than a third of challenged antitrust economic experts fail to meet the standard.^[12] Expert testimony built on DiD passes Daubert routinely when properly executed. But “properly executed” carries a lot of weight. The parties’ experts would dispute at length which non-merging entities were similar enough to the merged entities — in size, services, and especially pre-merger price trends — to serve as a valid counterfactual.

An opposing expert often argues, in some combination, that:

- **The treated and control groups were already diverging before the treatment.** If prices in the affected market were already rising faster than in the unaffected market before the merger, attributing the *full* post-merger gap to the merger may overestimate the merger effect. Plotting the pre-period and showing a visible divergence is often enough to seriously wound a DiD damages estimate.
- **An independent shock hit the treated and control groups disproportionately.** If a regulatory change, a natural disaster, or a major customer’s bankruptcy happened in the post-period and hit the treated and control groups disproportionately, DiD can’t separate their effects from the treatment effect. A defense challenge may show that an independent shock depressed the control group in the post-period and inflates the measured gap, biasing damages upward.
- **Results are sensitive to specification choices, such as the time window or which units serve as controls.** Courts usually treat such disagreements as going to the weight of the testimony — a battle of experts — rather than its admissibility. The estimate’s fragility becomes a reliability problem under *Daubert* only when reasonable alternative specifications move the damages figure by a factor of two and the expert never tested or disclosed that sensitivity.

These attacks reduce to two linked questions: is the comparison (control) group well chosen, and would the groups have satisfied parallel trends? The two are inseparable — the choice of control group is precisely what makes parallel trends plausible, and a control that fails the assumption can sometimes be replaced by one that satisfies it.

A defensible DiD analysis therefore rests on three pillars: pre-period parallelism over a meaningful window, robustness across reasonable alternative control groups, and honest treatment of any confounding event that touched one group but not the other. In practice, the credible workflow is to use a staggered-robust estimator when treatment timing varies across units, present the pre-period event-study plot as evidence of parallel movement, and report a formal sensitivity analysis showing how large a violation of parallel trends would have to be to overturn the result.^{[5][13][14]}

IV. How DiD Is Used in Marketing Analysis

The same method underlies a large share of marketing measurement work, even when it is not labeled DiD. When a campaign runs in some regions but not others, comparing the change in sales across the two sets of regions is a difference-in-differences estimate; the same logic applies when an analyst compares sales growth in markets that received a new campaign with markets that did not. DiD became the default tool for retrospective marketing analysis because marketing data fit it naturally: sales are observed repeatedly across many regions over time, and campaigns switch on and off at known dates — exactly the panel data structure DiD requires.

The cleanest application is a **regional rollout retrospective**. Suppose a retailer launched a new advertising campaign in five major metropolitan areas in the western United States in January and held off in the rest of the country. By July, sales in the campaign metros were up 8% relative to the same month a year earlier; nationally, the increase was 5%. A DiD comparison attributes the 3-percentage-point gap to the campaign. The strength of this analysis depends on the same parallel-trends assumption as it does in court: were the Western markets moving with the rest of the country before January? If yes, the answer is credible. If Western markets were already pulling ahead (perhaps because the retailer had placed its strongest stores there, or because the regional economy was running hotter there), then DiD would overstate the campaign's effect.

A second application is **measuring the lift of a brand campaign that can't be A/B tested**. National TV campaigns reach whole households; users cannot be randomized into “saw the Super Bowl ad” and “didn't see it.” But one can compare states that received heavy TV weight to states that received light weight, before and after the campaign aired. This is how a large fraction of national brand measurement actually gets done.

The same failure modes that threaten a courtroom analysis arise in marketing analysis as well. For example, consumers in the treated and control markets may have held very different preferences before the campaign ever ran. A confounding event, such as a competitor's product recall or a category-wide demand spike, may hit the treated and control markets disproportionately. The post-period may coincide with a holiday season that is more pronounced in the treated markets. Or the control markets may be quietly affected by spillover (residents of “control” Charlotte drive to “treated” Atlanta and see the billboard there). Every one of these maps to a parallel-trends failure or a contamination of the control group, and every one of these requires the same defenses: pre-period evidence, identification of contemporaneous shocks, and robustness across alternative controls.

V. Conclusion and What's Coming Up Next

DiD is a well-accepted method for causal inference with one critical assumption — parallel trends. Whether the audience is a federal judge weighing a Daubert motion or a CMO deciding whether the campaign worked, the question that determines whether a DiD estimate can be trusted is the same: would the treated and control groups have continued to move in parallel if the treatment hadn't happened? And the evidence that builds credibility is the same in both: showing visible pre-period parallelism over a meaningful window, addressing any confounding event that touched the treated and control groups disproportionately, and performing robustness checks across reasonable alternative controls. The audiences differ; the discipline doesn't.

Coming up next: the synthetic control method, the natural complement to DiD. Where DiD compares a treated group to one or more pre-chosen control groups and assumes parallel

trends, synthetic control builds a single “synthetic” comparison as a weighted average of available control units, choosing the weights so that the synthetic unit closely tracks the treated unit’s pre-treatment outcome and characteristics.^{[15][16]} It is most useful when only one unit is treated — a single firm, market, or jurisdiction — and no off-the-shelf control group is convincing. Synthetic control does not escape the identifying problem: it relies on its own version of parallel trends (that the treated unit and its synthetic counterpart would have moved together absent the treatment), but it makes that assumption more transparent and more credible by matching the pre-treatment path directly. Its trade-offs are a need for a long pre-period and a good pool of donor units, sensitivity of the weights, and inference based on placebo (permutation) tests rather than conventional standard errors. The next article takes up these details and discusses when to prefer each method.

Sources

- [1] Card, D., & Krueger, A. B. (1994). “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania.” *American Economic Review*, 84(4): 772–793. <https://www.jstor.org/stable/2118030>
- [2] Angrist, J. D., & Pischke, J-S. (2009). *Mostly Harmless Econometrics*, ch. 5. <https://press.princeton.edu/books/paperback/9780691120355/mostly-harmless-econometrics>
- [3] Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. <https://mixtape.scunning.com/>
- [4] Roth, J. (2022). “Pretest with Caution: Event-Study Estimates After Testing for Parallel Trends.” *American Economic Review: Insights*, 4(3): 305–322. <https://www.aeaweb.org/articles?id=10.1257/aeri.20210236>
- [5] Rambachan, A., & Roth, J. (2023). “A More Credible Approach to Parallel Trends.” *Review of Economic Studies*, 90(5): 2555–2591. <https://academic.oup.com/restud/article-abstract/90/5/2555/7039335?redirectedFrom=fulltext>
- [6] Ashenfelter, O., Hosken, D., & Weinberg, M. (2015). “Efficiencies Brewed: Pricing and Consolidation in the U.S. Beer Industry.” *RAND Journal of Economics*, 46(2): 328–361. <https://onlinelibrary.wiley.com/doi/10.1111/1756-2171.12092>
- [7] Kwoka, J. (2014). *Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy*. MIT Press. <https://direct.mit.edu/books/monograph/3084/Mergers-Merger-Control-and-RemediesA-Retrospective>
- [8] Haas-Wilson, D., & Garmon, C. (2011). “Hospital Mergers and Competitive Effects: Two Retrospective Analyses.” *International Journal of the Economics of Business*, 18(1): 17–32. <https://www.tandfonline.com/doi/full/10.1080/13571516.2011.542952>
- [9] In re Evanston Northwestern Healthcare Corp., FTC Docket No. 9315, Opinion of the Commission (Aug. 6, 2007). <https://www.ftc.gov/sites/default/files/documents/cases/2007/08/070806opinion.pdf>
- [10] Federal Judicial Center & National Research Council, Reference Manual on Scientific Evidence, 3rd ed. (2011). <https://nap.nationalacademies.org/catalog/13163>
- [11] Federal Judicial Center and the National Academies of Sciences, Engineering, and Medicine, Reference Manual on Scientific Evidence, 4th ed. (2025). https://www.fjc.gov/sites/default/files/materials/15/Reference%20Manual_Vol_I_March_2026.pdf

[12] ProMarket, “Expert Economist Testimonies Are Challenged More Often in Antitrust Cases” (Jan. 28, 2025). <https://www.promarket.org/2025/01/28/expert-economist-testimonies-are-challenged-more-often-in-antitrust-cases/>

[13] Callaway, B., & Sant’Anna, P. H. C. (2021). “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics*, 225(2): 200–230. <https://www.sciencedirect.com/science/article/abs/pii/S0304407620303948?via%3Dihub>

[14] Roth, J., Sant’Anna, P. H. C., Bilinski, A., & Poe, J. (2023). “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.” *Journal of Econometrics*, 235(2): 2218–2244. <https://arxiv.org/abs/2201.01194>

[15] Abadie, A., Diamond, A., & Hainmueller, J. (2010). “Synthetic Control Methods for Comparative Case Studies.” *Journal of the American Statistical Association*, 105(490): 493–505. <https://www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08746>

[16] Abadie, A. (2021). “Using Synthetic Controls.” *Journal of Economic Literature*, 59(2): 391–425. <https://www.aeaweb.org/articles?id=10.1257/jel.20191450>

About the Authors

* Rui Huang is Principal Economist and Data Science Expert at Nutcracker Economics. She has over 20 years of experience in antitrust economics, applied econometrics, and data science, with expertise spanning causal inference, experimentation design, and marketing ROI measurement. She spent over eight years leading science teams at major technology companies, served as a Staff Economist at the U.S. Department of Justice Antitrust Division, and was a tenure-track professor at the University of Connecticut with 10+ peer-reviewed publications. She holds a PhD in Economics from UC Berkeley.

** Pian Chen is Founder and Lead Economist at Nutcracker Economics. She has over 15 years of experience in litigation consulting and government oversight, with deep expertise in antitrust economics and financial fraud investigation. She previously held senior positions at leading litigation consulting firms and served as Associate Director of Economic Modeling at the Public Company Accounting Oversight Board (PCAOB). She holds a PhD in Agricultural and Resource Economics from UC Davis.