Roth and Sant'Anna (2023, Econometrica)

Student Presentation in Development Economics 1

Yasuyuki Matsumura (Kyoto University)

January 31st, 2025

https://yasu0704xx.github.io

Econometrica, Vol. 91, No. 2 (March, 2023), 737–747

WHEN IS PARALLEL TRENDS SENSITIVE TO FUNCTIONAL FORM?

JONATHAN ROTH
Department of Economics, Brown University

PEDRO H. C. SANT'ANNA
Microsoft Research and Department of Economics, Vanderbilt University

This paper assesses when the validity of difference-in-differences depends on functional form. We provide a novel characterization: the parallel trends assumption holds under all strictly monotonic transformations of the outcome if and only if a stronger "parallel trends"-type condition holds for the cumulative distribution function of untreated potential outcomes. This condition for parallel trends to be insensitive to functional form is satisfied if and essentially only if the population can be partitioned into a subgroup for which treatment is effectively randomly assigned and a remaining subgroup for which the distribution of untreated potential outcomes is stable over time. These conditions have testable implications, and we introduce falsification tests for the null that parallel trends is insensitive to functional form.

KEYWORDS: Difference-in-differences, functional form, robustness, testable implications.

This slide is available on
https://github.com/yasu0704xx/ArticleReview.

Contents

Introduction

Canonical Two-Period DiD

- Here we quickly review notations used in Roth and Sant'Anna (2023).
- Suppose that we have panel data $\{\{(Y_{it}, D_{it})\}_{t=0}^{T}\}_{i=1}^{n}$, where i and t index units and time periods, respectively. $Y_{it} \in \mathbb{R}$ and $D_{it} \in \{0, 1\}$ denote the outcome and treatment, respectively, for unit i in period t.
- Consider the simplest case where $T = 1$, no units are treated at $t = 0$ (i.e., $D_{i0} = 0$ for any i), and some but not all units become treated at $t = 1$.
- Then, only D_{i1} is relevant, and we simply write $D_i = D_{i1}$.
- Let $Y_{it}(d)$ denote the potential outcome given $D_i = d$.

Parallel Trends in Canonical DiD

- Parallel trends (PT) is a key assumption for point identification of ATT in Difference-in-Differences (DiD) designs. [1]
- In a canonical two-period DiD model, PT means

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|D_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|D_i = 0]$$

  In words, the untreated potential outcome is required to have the same average path over time between the treated and untreated groups.
- Under assumptions of no anticipation and PT, ATT is identified as

$$\begin{aligned}\tau_{\text{ATT}} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|D_i = 1] \\ &= \mathbb{E}[Y_{i1} - Y_{i0}|D_i = 1] - \mathbb{E}[Y_{i1} - Y_{i0}|D_i = 0]\end{aligned}$$

[1] Note that identification of ATT also requires another assumption, namely, "no anticipation," or "no anticipatory effects of treatment," which we do not focus on here.

PT and Functional Form

- The justification of PT may depend on the functional form. For example, which of the following should we expect to hold?

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|D_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|D_i = 0]?$$

$$\mathbb{E}[\log Y_{i1}(0) - \log Y_{i0}(0)|D_i = 1] = \mathbb{E}[\log Y_{i1}(0) - \log Y_{i0}(0)|D_i = 0]?$$

- Literatures note that the PT assumption may hold in log but not levels or vice versa (e.g., Meyer 1995; Athey and Imbens 2006; Kahn-Lang and Lang 2020).
- Roth and Sant'Anna (2023) show that, under a certain condition, PT holds for any strictly monotonic transformations:

$$\mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0))|D_i = 1] = \mathbb{E}[g(Y_{i1}(0)) - g(Y_{i0}(0))|D_i = 0]$$

  for all strictly monotonic g. The converse is also shown to be true. [2]

[2] This page and the next one quote a lecture slide by Yanagi, T. (2024).

- Roth and Sant'Anna (2023) show that PT is insensitive to functional form iff a "parallel-trends"-type condition holds for the entire cumulative distribution function (CDF) of $Y(0)$.
- They further provide the following characterizations: the above condition can be satisfied iff
    - treatment is as-if randomly assigned (i.e., the distribution of $Y_{it}(0)$ conditional on $D_i = d$ does not depend on d); or
    - the distribution of $Y_{it}(0)$ is stable over time for each treatment group (i.e., the distribution of $Y_{it}(0)$ conditional on $D_i = d$ does not depend on t); or
    - a hybrid of above two cases holds (i.e., θ fraction of the population satisfies the first one, and $(1 - \theta)$ fraction of the population satisfies the second one).
- In the setting where the treatment is not (as-if) randomly assigned, the assumptions needed for the insensitivity of PT to functional form will often be quite restrictive.

Literature and Contribution

- The seinsitivity of PT to functional form is pointed out by Meyer (1995); Athey and Imbens (2006); and Kahn-Lang and Lang (2020).
- Athey and Imbens (2006) introduce assumptions for identifying distributional treatment effects in DiD settings. This is related to, but distinct from, Roth and Sant'Anna (2023).
- Roth and Sant'Anna (2023) provide the first full characterizations of when PT is sensitive to functional form. They also provide testable implications of insensivity to transformations.

Invariance of Parallel Trends

Setup

- Consider the canonical DiD introduced above (for expositional simplicity).
  - Note that Roth and Sant'Anna (2023)'s results have immediate implications for DiD settings with multiple periods and staggered treatment timing. See Roth, Sant'Anna, Bilinski and Poe (2022) for a review.
  - Similarly, their resluts would go thorough in settings with conditional parallel trends. See Abadie (2005); and Sant'Anna and Zhao (2020) for details.
- The PT assumption imposes that

$$
\begin{aligned}
&\mathbb{E}[Y_{i1}(0)|D_i = 1] - \mathbb{E}[Y_{i0}(0)|D_i = 1] \\
&= \mathbb{E}[Y_{i1}(0)|D_i = 0] - \mathbb{E}[Y_{i0}(0)|D_i = 0]. \qquad (1)
\end{aligned}
$$

- We assume throughout that the four expectations in (1) exist and finite.

Definition (PT Insensitivity to Functional Form)

Definition 1

We say that the PT assumption is invariant to transformations if

$$\begin{aligned}&\mathbb{E}[g(Y_{i1}(0))|D_i = 1] - \mathbb{E}[g(Y_{i0}(0))|D_i = 1]\\&= \mathbb{E}[g(Y_{i1}(0))|D_i = 0] - \mathbb{E}[g(Y_{i0}(0))|D_i = 0],\end{aligned}$$

for strictly monotonic functions g such that the expectations above are finite.

PT-Type Condition

Proposition 3.1

PT is invariant to transformations <u>if and only if</u>

$$
\begin{aligned}
&F_{Y_{i1}(0)|D_i=1}(y) - F_{Y_{i0}(0)|D_i=1}(y) \\
&= F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y), \text{ for all } y \in \mathbb{R},
\end{aligned} \tag{2}
$$

where $F_{y_{it}(0)|D_i=d}(y)$ is the cumulative distribution function of $Y_{it}(0)|D_i = d$.

- The proof is given in the following pages, which is the same as the proof provided in p.739 of Roth and Sant'Anna (2023).
- Note that, if the outcome is continuous, the PT of CDFs is equivalent to the PT of PDFs (almost everywhere).

Proof (Prop. 3.1)

- ($\Longleftarrow$) Suppose that (2) holds. Integrating on both sides of the equation, it is easy to see that

$$\begin{aligned} &\int g(y)F_{Y_{i1}(0)|D_i=1} - \int g(y)F_{Y_{i0}(0)|D_i=1} \\ &= \int g(y)F_{Y_{i1}(0)|D_i=0} - \int g(y)F_{Y_{i0}(0)|D_i=0} \end{aligned} \tag{3}$$

  for any strictly monotonic g such that the integrals exist and are finite. Noting that (3) is equivalent to the invariance of PT to transformations, we obtain the desired result.

- ($\Longrightarrow$) Suppose that PT is invariant to transformations, that is, (3) holds for every strictly monotonic g such that the four expectations in Definition 1 exsit and are finite.
- Consider the following functions: $g_1(y) = y$, $g_2(y) = y - 1(y \leq \tilde{y})$ for any given $\tilde{y} \in \mathbb{R}$.
- Since g_1 is strictly monotonic, PT holds under transformation of g_1:

$$\mathbb{E}[Y_{i1}(0)|D_i = 1] - \mathbb{E}[Y_{i0}(0)|D_i = 1]$$
$$= \mathbb{E}[Y_{i1}(0)|D_i = 0] - \mathbb{E}[Y_{i0}(0)|D_i = 0],$$

which is equivalent to

$$\int y dF_{Y_{i1}(0)|D_i=1} - \int y dF_{Y_{i0}(0)|D_i=1}$$
$$= \int y dF_{Y_{i1}(0)|D_i=0} - \int y dF_{Y_{i0}(0)|D_i=0}.$$

- Similarly, PT also holds under the strictly monotonic transformation of g_2:

$$\mathbb{E}[Y_{i1}(0) - 1(Y_{i1}(0) \leq \tilde{y})|D_i = 1] - \mathbb{E}[Y_{i0}(0) - 1(Y_{i0}(0) \leq \tilde{y})|D_i = 1]$$
$$= \mathbb{E}[Y_{i1}(0) - 1(Y_{i1}(0) \leq \tilde{y})|D_i = 0] - \mathbb{E}[Y_{i0}(0) - 1(Y_{i0}(0) \leq \tilde{y})|D_i = 0],$$

which is equivalent to

$$\int \{y - 1(y \leq \tilde{y})\} dF_{Y_{i1}(0)|D_i=1} - \int \{y - 1(y \leq \tilde{y})\} dF_{Y_{i0}(0)|D_i=1}$$
$$= \int \{y - 1(y \leq \tilde{y})\} dF_{Y_{i1}(0)|D_i=0} - \int \{y - 1(y \leq \tilde{y})\} dF_{Y_{i0}(0)|D_i=0}.$$

- Combinig these results, we obtain

$$\int 1(y \leq \tilde{y}) dF_{Y_{i1}(0)|D_i=1} - \int 1(y \leq \tilde{y}) dF_{Y_{i0}(0)|D_i=1}$$
$$= \int 1(y \leq \tilde{y}) dF_{Y_{i1}(0)|D_i=0} - \int 1(y \leq \tilde{y}) dF_{Y_{i0}(0)|D_i=0}.$$

- Noting that $\tilde{y}$ is arbitary, the definition of CDF immediately leads to eqation (2), which is desired. $\square$

How Can Distributions Satisfying the PT-Type Condition be Generated?

Proposition 3.2

Suppose that the distributions of $Y_{it}(0)|D_i = d$ for all $d, t \in \{0, 1\}$ have a Radom-Nikodym density with respect to a common dominating, positive σ-finite measure. Then PT is invariant to transformations if and only if there exist

- $\theta \in [0, 1]$, and
- CDFs $G_t(\cdot)$ and $H_d(\cdot)$ depending only on time and group, respectively,

such that

$$F_{Y_{it}(0)|D_i=d}(y) = \theta G_t(y) + (1 - \theta)H_d(y),$$
$$\text{for all } y \in \mathbb{R} \text{ and } d, t \in \{0, 1\}. \tag{4}$$

Proof (Prop. 3.2)

- The following proof is the same as one provided in pp.745-746 of Roth and Sant'Anna (2023).
- By Prop. 3.1, it is sufficient to show that (2) $\iff$ (4).
- ($\Longleftarrow$) Note that $H_d(y)$ does not depend on time. Under (4), the LHS of (2) reduces to

$$(\text{LHS of } (2)) = \theta(G_1(y) - G_0(y)).$$

- By the same argument, the RHS of (2) also reduces to

$$(\text{RHS of } (2)) = \theta(G_1(y) - G_0(y)).$$

- Combining these results will immediately lead to (2), which is desired.

- ($\Longrightarrow$) Let $\mathcal{Y}$ denote the parameter space for $Y(0)$, and $\mathcal{Y}_y = \{\tilde{y} \in \mathcal{Y} | \tilde{y} \leq y\}$.
- Recall the assumption that the distributions of $Y_{it}(0)|D_i = d\,(d, t \in \{0, 1\})$ have a Radon-Nikodym density w.r.t. a common dominating, positive σ-finite measure (namely, λ).
- By this assumption, we can write

$$F_{Y_{it}(0)|D_i=d}(y) = \int_{\mathcal{Y}_y} f_{Y_{it}(0)|D_i=d} d\lambda,$$

where $f_{Y_{it}(0)|D_i=d}$ is the density (the Radom-Nikodym derivative).

- Under (2):

$$\begin{aligned} & F_{Y_{i1}(0)|D_i=1}(y) - F_{Y_{i0}(0)|D_i=1}(y) \\ & = F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y), \end{aligned}$$

it holds that

$$\begin{aligned} & f_{Y_{i1}(0)|D_i=1}(y) - f_{Y_{i0}(0)|D_i=1}(y) \\ & = f_{Y_{i1}(0)|D_i=0}(y) - f_{Y_{i0}(0)|D_i=0}(y), \; \lambda \text{ almost everywhere.} \end{aligned}$$

- To finalize the proof, we introduce the following lemma:

Lemma A.1

Suppose that the CDFs F_1 and F_2 are such that

$$F_j(y) = \int_{\mathcal{Y}_y} f_j d\lambda.$$

Then, we can decompose $F_j(y)$ as

$$F_j(y) = (1-\theta)F_{\mathsf{min}}(y) + \theta\tilde{F}_j(y) \text{ for } j = 1, 2,$$

where

- F_{min} and $\tilde{F}_j$ are CDFs,
- $\theta \in [0, 1]$, and
- θ and $\tilde{F}_j$ depend only on f_1 and f_2 through $f_1 - f_2$.

- Applying Lemma A.1 to the four CDFs on both sides of (2), we obtain the disired result (4).

Implications from Prop. 3.2

- Proposition 3.2 shows that PT of CDFs is satisfied if and only if the untreated potential outcomes for each group and time can be represented as a mixture of a common time-varying distribution that does not depend on group (with weight θ) and a group-specific distribution taht does not depend on time (with weight $1 - \theta$).
- This implies that there are three cases in which PT will be insensitive to functional form, depanding on the value of θ
  - Case 1 ($\theta = 1$): Random Assignment
  - Case 2 ($\theta = 0$): Stationary $Y(0)$
  - Case 3 ($0 < \theta < 1$): Non-random Assignment and Nonstationarity

Case 1: Random Assignment

- PT is insensitive to functional form when it holds that

$$F_{Y_{it}(0)|D_i=d}(y) = G_t(y) \text{ for all } y \in \mathbb{R} \text{ and } d, t \in \{0, 1\}.$$

  Note that the term $G_t(y)$ does not depend on group.

- This case corresponds with imposing that the distributions of untreated potential outcomes $Y(0)$ for treated and comparison groups are the same in each period, i.e.,

$$F_{Y_{it}(0)|D_i=1}(y) = F_{Y_{it}(0)|D_i=0}(y)$$

  for all $t = 0, 1$, and all y.

- This can occur under (as-if) random assignment of treatment.

Case 2: Stationarity of Untreated Potential Outcomes of Each Group

- PT is insensitive to functional form when it holds that

$$F_{Y_{it}(0)|D_i=d}(y) = H_d(y) \text{ for all } y \in \mathbb{R} \text{ and } d, t \in \{0, 1\}.$$

  Note that the term $H_d(y)$ does not depend on time.

- This case corresponds with imposing that the distribution of untreated potential outcome $Y(0)$ for both the treated and comparison popuations does not depend on time, i.e.,

$$F_{Y_{i1}(0)|D_i=d}(y) = F_{Y_{i0}(0)|D_i=d}(y)$$

  for all $d = 0, 1$, and all y.

Case 3: Non-random Assignment and Nonstationarity

- PT is insensitive to functional form when it holds for $0 < \theta < 1$ that

  $F_{Y_{it}(0)|D_i=d}(y) = \theta G_t(y) + (1-\theta) H_d(y)$ for all $y \in \mathbb{R}$ and $d, t \in \{0, 1\}$.

- This case corresponds with a hybrid of the above two cases.
- In each period, we can partition the treated and comparison groups so that
  - θ fraction of each group have the same distribution G_t, as if they were rondomly assigned, and
  - $1-\theta$ fraction of each group have a group-specific distribution H_d that does not depend on time (i.e, $1-\theta$ fraction have stationary potential outcomes).
- In principle, it is possible for the units in the θ and $1-\theta$ partitions to change across periods. However, it is difficult to imagine coresponding practical/empirical senarios. Thus, Roth and Sant'Anna (2023) mention in abstract only the case in which the population can be partitioned into the above two fractions.

Examples 1 and 2

- Example 1 provides a simplified illustration of the original PT (1), using the binary outcome $Y_i \in \{0, 1\}$.
- Example 2 describes the hypothetical case in which $Y_{it}(0)|D_i = d$ is normally distributed for all d and t, In this example, (2) can hold only if either
  - both treated and comparison groups have the same distribution of untreated potential outcome $Y(0)$ in each period; or
  - both groups have the group-specific distributions of $Y(0)$ which do not change over time.

Example 3

- Example 3 provides an illustration/simulation of Case 3.
- The distributions of untreated potential outcomes are generated by

$$F_{Y_{it}(0)|D_i=d}(y) = \frac{1}{2}G_t(y) + \frac{1}{2}H_d(y)$$
$$G_t \sim \text{lognormal}(2+t, 1), \quad H_d \sim \text{lognormal}(3+d, 1).$$

  for each $d = 0, 1$; $t = 0, 1$.

- As can be seen in (a) and (b) of Figure 1, the distributions of $Y(0)$ for the treated and comparison groups differ from each other in both time periods (pre-treatment and post-treatment periods).
- Figure 1-(c) shows that the change in the PDFs is the same for both groups.
- Table 1 also implies that PT invariance is considered to be plausible, computing sample means of each groups' $Y(0)$'s and $\log Y(0)$'s.

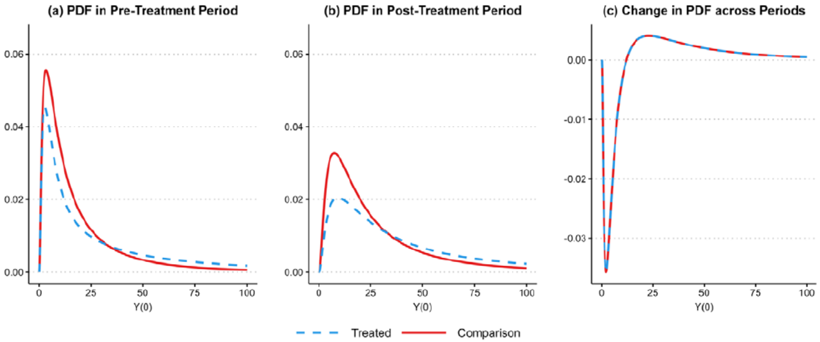

FIGURE 1.—Illustration of Case 3. Notes: Data generating process as discussed in Example 3.

TABLE I

MEAN OF $g(Y(0))$ BY GROUP.

g	Group	Pre-treatment	Post-treatment	Change
Levels	Comparison	22.65	33.12	10.47
Levels	Treated	51.10	61.57	10.47
Log	Comparison	2.50	3.00	0.50
Log	Treated	3.00	3.50	0.50

Note: Data generating process as discussed in Example 3.

Remarks 3 and 4

- Remark 3: The following empirical papers use a DiD design to estimate the effects of treatment on the distribution of an outcome: Almond, Hoynes, and Schanzenbach (2011); Cengiz, Dube, Lindner, and Zipperer (2019); and Stepner (2019).
- Such distributional analyses using DiD are required to ensure that one out of three scenarios (Cases 1, 2, and 3) disucussed above holds.
- Remark 4: Roth and Sant'Anna (2023) consider identification of the full counterfactual distribution of the treated group $Y_{i1}(0)|D_i = 1$, which can be recovered under condition (2). Related literature include Athey and Imbens (2006); Bonhomme and Sauder (2011); and Callaway and Li (2019).

Testable Implications of Invariance to Transformations

Falsification Test

- (2) can be rewritten as

$$F_{Y_{i1}(0)|D_i=1}(y) \\ = F_{Y_{i0}(0)|D_i=1}(y) + F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y). \tag{5}$$

- The LHS of (5) is a CDF, and thus must be weakly increasing.
- However, this is not guaranteed of the RHS of (5).
- Thus, we can falsify condition (5) if we reject the following null hypothesis:

$\mathbb{H}_0$: the RHS of (5) is weakly increasing in y.

Falsification Test for the Invariance of PT

- For simplicity, we focus on the case in which y has finite support $\mathcal{Y}$ and the null hypothesis is equivalent to testing that the implied distribution has nonnegative mass at all support points.
- That is, we are interested in testing that

$$f_{Y_{i1}(0)|D_i=1}(y) = f_{Y_{i0}(0)|D_i=1}(y) + f_{Y_{i1}(0)|D_i=0}(y) - f_{Y_{i0}(0)|D_i=0}(y) \geq 0 \text{ for all } y \in \mathcal{Y}, \tag{6}$$

  where $f_{Y_{it}(0)|D_i=d}(y) = \mathbb{E}[1(Y_{it}(0) = y)|D_i = d]$ is the probability mass function of $Y_{it}(0)|D_i = d$ at y.
- In practice, using the sample analogue of $f_{Y_{i1}(0)|D_i=1}(y)$, the null hypothesis can be renewed as

$$\mathbb{H}_0 : \mathbb{E}[\hat{f}_{Y_{i1}(0)|D_i=1}(y)] \geq 0 \text{ for all } y.$$

- This can be tested using methods of the moment inequality literature (Canay and Shaikh, 2017).

Caveats on Pre-Testing

- Note that failure to reject the null hypothesis does not necessarily imply that PT is insensitive to functional form.
- The falsification test is a pre-testing, which may induce certain problems. Roth (2022) considers such problems arising with pre-testing in DiD settings.

References

References i

- Abadie, A. (2005). "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies*, 72(1), 1–19.
- Almond, D., Hoynes, H. W., & Schanzenbach, D. W. (2011). "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes." *Review of Economics and Statistics*, 93(2), 387–403.
- Athey, S., & Imbens, G. (2006). "Identification and Inference in Nonlinear Difference-in-Differences Models." *Econometrica*, 74(2), 431–497.

- Bonhomme, S., & Sauder, U. (2011). "Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling." *Review of Economics and Statistics*, 93, 479–494.
- Callaway, B., & Li, T. (2019). "Quantile Treatment Effects in Difference in Differences Models With Panel Data." *Quantitative Economics*, 10(4), 1579–1618.
- Canay, I., & Shaikh, A. (2017). "Practical and Theoretical Advances in Inference for Partially Identified Models." In B. Honor et al. (Eds.), *Advances in Economics and Econometrics*, pp. 271–306.

- Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134(3), 1405–1454.
- Kahn-Lang, A., & Lang, K. (2020). "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications." *Journal of Business & Economic Statistics*, 38(3), 613–620.
- Meyer, B. D. (1995). "Natural and Quasi-Experiments in Economics." *Journal of Business & Economic Statistics*, 13(2), 151–161.

- Roth, J. (2022). "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights*, 4(3), 305–322.
- Roth, J., & Sant'Anna, P. H. C. (2021). "When Is Parallel Trends Sensitive to Functional Form?" *arXiv preprint arXiv:2010.04814v2*.
- Roth, J., & Sant'Anna, P. H. C. (2023). "When Is Parallel Trends Sensitive to Functional Form?" *Econometrica*, 91(2), 737–747.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2022). "What ' s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *arXiv preprint arXiv:2201.01194*.

Sant'Anna, P. H. C., & Zhao, J. (2020). "Doubly Robust Difference-in-Differences Estimators." *Journal of Econometrics*, 219(1), 101–122.

Stepner, M. (2019). "The Insurance Value of Redistributive Taxes and Transfers." Manuscript.