EU Tax Observatory Working Paper No. 31

Proportional Treatment Effects in Staggered Settings: An Approach for Poisson Pseudo-Maximum Likelihood

May 2025

Ninon Moreau-Kastler

EU Tax Observatory, Paris School of Economics

Proportional Treatment Effects in Staggered Settings: An Approach for Poisson Pseudo-Maximum Likelihood.*

Ninon Moreau-Kastler†

May 6, 2025

Abstract

I propose a counterfactual approach to measure proportional treatment effects for staggered multiplicative difference-in-differences (DiD) models with Poisson Pseudo-Maximum Likelihood (PPML). Two-way fixed effect (TWFE) linear estimators do not recover DiD estimates in the presence of a staggered treatment. I show that the wrong comparisons problem extends to TWFE PPML. I provide evidence that robust estimators for the linear case do not naturally extend to PPML, as aggregation of lower-level effects is challenging in the non-linear case. In these settings, my proposed estimator recovers a quantity analogous to that in the canonical 2-by-2 TWFE PPML model: the percent change of the average.

JEL codes: C21, C23, F14, H26

Keywords: PPML; difference-in-differences; ratio-of-ratios; staggered treatment

* I thank participants of the Paris Econometrics Seminar, PSE Labor and Public Econ Seminar, Kirill Borusyak, Clément De Chaisemartin, Xavier d'Haultfoeuille, Martin Mugnier, Jonathan Roth, Farid Toubal, and Morgan Ubeda for helpful comments and advice.

†EU Tax Observatory, Paris School of Economics.

1 Introduction

Applied economists are often interested in studying variables which take only non-negative values and are non-normally distributed. Such outcomes can include, for example, trade flows, sales or employment. Public policies or economic shocks generate changes in these outcomes whose magnitudes often vary across small or large countries, firms, or sectors. In such cases, researchers are interested in proportional treatment effects, or semi-elasticities: the percent change in the outcome generated by the treatment. The two-way fixed effects Poisson Pseudo-Maximum Likelihood (TWFE PPML) estimator presents several advantages over its log-linearized counterpart (TWFE log-OLS). It recovers a proportional treatment effect. It can include observations with zero in the outcome, and easily accommodate unit fixed effects without being subject to the incidental parameters problem (Wooldridge, 1999). It is suited for settings where the treatment changes the level of the outcome and the variance of the error term (Silva and Tenreyro, 2006).

When treatment effects are heterogeneous across units, TWFE PPML targets the percentage change of the average outcome in the treated group. It computes the multiplicative Difference-in-Differences, or Ratio-of-Ratios estimator: the ratio of the average outcome before and after treatment in the treated group, scaled by the change in the average outcome in the control group. The multiplicative Difference-in-Differences relies on the parallel trend assumption that the growth rate in the outcome of the two groups should have been the same without treatment.[1] In the presence of heterogeneous treatment effects, with staggered treatment timings, researchers have been recently concerned with the fact that two-way fixed effects estimators do not recover desired difference-in-differences estimates of the treatment effect. Linear estimators use "forbidden comparisons" of successively treated groups, and

[1] TWFE log-OLS targets the average individual log-points change, an approximation of the outcome percentage change across individuals, which implies a different parallel trend.

weight negatively some treatment effects, potentially yielding estimates of the wrong sign.[2] In this paper, I show that the same issue plagues the TWFE PPML estimator. Using a simple example with two indivuals treated at different times, I show that the estimated quantity differs significantly from the multiplicative DiD estimation target when there are heterogeneous treatment effects across cohorts and time.

Robust estimators have been developed in the linear case, recovering accurate DiD estimates for cohort and time cells and aggregating them at a higher scale (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Wooldridge, 2021). These approches do not suit non-linear estimators such as PPML: they average linear treatment effects, which is more challenging for non-linear settings. Imagine a researcher who observes employment changes in two firms A and B, which respectively employ 1 and 2 persons at baseline. If treatment increases employment by one person in each firm, the average change in percentage would be $(100\% + 50\%) \times (\frac{1}{2}) = +75\%$. The change of average (and total) employment would be different: $\frac{(2+3)-(1+2)}{1+2} \times 100 = +66\%$. [3] With linear effects, these quantities would all be the same, but they differ for non-linear effects such as percentage changes. If we observed instead 3 firms, A, B and C, with A and C being part of region 1 and B of region 2, one could now compute three quantities reflecting change in employment: the average firm change, the change of the average (total) employment, and a weighted average of regional employment changes.

Generalizing, because the multiplicative difference-in-differences model targets the change of the average treated outcome in percentage, it can usually not be recovered by a weighted sum of proportional changes at a lower level, lets say on groups g and periods t:

$$\sum_{g,t} \nu_{gt} \frac{E[y_{gt}(1)] - E[y_{gt}(0)]}{E[y_{gt}(0)]} \neq \frac{E[y(1)] - E[y(0)]}{E[y(0)]} \tag{1.1}$$

[2] See De Chaisemartin and D'Haultfoeuille (2023) for a review of this literature.
[3] TWFE log-OLS approximates the first quantity, and TWFE PPML targets the second one.

With ν_{gt} the weights associated to group g at time t, unless $\nu_{gt} = \frac{N_{gt}}{N}\frac{E[y_{gt}(0)]}{E[y(0)]}$, the relative size of cell g, t in total counterfactual outcome.

In this paper, I develop an estimator that recovers a proportional treatment effect that can be interpreted as a percentage change of the average outcome, even in staggered settings with heterogeneous treatment effects. This estimate can be interpreted as a semi-elasticity, and it corresponds to the simple setting quantity of interest: the percent change of the average, or the scaled ATT. This estimator rests on the idea that the TWFE PPML estimator recovers the ratio-of-ratios in the canonical 2 times and 2 groups setting (Angrist, 2001; Ciani and Fisher, 2019), and is equivalent to it in this simple setting. Using a parallel trend in growth rates, a counterfactual outcome can be estimated by multiplying the pre-treatment outcome of the treated group by the growth rate of the control group's outcome. My estimator then recovers a consistent estimated average treatment effect in level and scales it by the predicted counterfactual average outcome of the treated group. The estimated quantity corresponds to the growth rate of the average outcome caused by the treatment. I show that in cases without controls, this estimator can be computed using either a fully saturated model or an imputation estimator (Wooldridge, 2023; Borusyak et al., 2024).

This paper relates particularly to two recent papers on multiplicative DiD in staggered treatment settings. Wooldridge (2023) covers the more general case of non-linear DiD estimators in staggered designs. He explains that the model specification should allow for all margins of treatment heterogeneity that the data structure can identify to avoid the wrong comparison problem. Such a model corrects the underlying assumptions of the TWFE estimator, and recovers estimates of cohort-time treatment effects in level.[4] However, the targeted quantity is the treatment effect in level, and not the proportional treatment effect. Tthe paper does not discuss how to recover it at a higher level than at the cohort-time cell. He fur-

[4]The paper further shows that for a balanced panel, the TWFE and pooled estimation approaches are equivalent, requiring only to use cohort and time fixed effects which reduces considerably the incidental parameter problem of non-linear estimators.

ther provides evidence that a fully saturated model is equivalent to an imputation estimator, and that it allows to easily estimate linear treatment effects for non-linear models. In this paper, I clearly state the higher level quantity of interest of proportional treatment effect, and present a realiable approach to recover it in the staggered setting, building on this imputation result. My estimator is suited to recover a proportional treatment effect (semi-elasticity) at any aggregation scale.

Nagengast and Yotov (2025) revisit the semi-elasticity estimates of bilateral trade to regional trade agreement (RTAs): the authors note that most existing estimates are computed in staggered treatment timing settings. They use Wooldridge (2023)'s fully saturated specificaton to estimate partial equilibrium changes in trade caused by RTAs. They aggregate cohort-time proportional treatments effects by computing the weighted average of estimated model coefficients, weighting cohort-time cells by the share of treated observations. In this paper, I show that if treatment effect are small, and treatment heterogeneity occurs only across cohorts and time, this quantity approximates an average proportional treatment effect. It has then the same interpretation than the quantity of interest of the log-OLS estimator, but not of the PPML estimator (and should generally not be compared with it). I show that in a more general case, the estimated quantity is closer to a weighted average of the total percent change within cohorts-time celles, and intermediate quantity between log-OLS and PPML. Finally, I replicate their estimates on trade with my estimator, and provide close results: this caused by the percent change of the average and the average percent change are close quantities in their setting.[5]

After reminding the 2x2 canonical setting of the multiplicative difference-in-difference and TWFE PPML, I explore the multiperiod setting and heterogeneous treatment timing case. This paper is the first to provide a formal evidence of TWFE PPML bias, under the same conditions as TWFE OLS: heterogeneous treatment effect across time, and staggered

[5]The initial difference between TWFE log-OLS and TWFE PPML is small.

treatment timing. I show that in a simple setting with two individuals and three periods, TWFE PPML downscales a correct treatment effects with the ones of other cells, by analogy with the negative weights issue of the linear case.

I discuss potential estimates to recover multiplicative difference-in-differences (ratio-of-ratios) estimates in this setting. As discussed above, I show that estimators aggregating treatment effects estimated separately for each cohort-time cell, such as what is proposed by the litterature in the linear case, provide different quantities than the initial quantity of interest. I call this type of approaches "aggregation strategies". I further show that their causal interpretation can be difficult in more general cases. I then provide an estimator recovering the correct ratio-of-ratios analogous to the canonical setting. I show that this estimator can be recovered through an imputation process or fully saturated model, but that the imputation approach can apply to a wider set of cases.

I compare my estimator to the true quantity of interest, against alternative estimators in simulations of section 4. In staggered treatment timing case, I confirm that TWFE PPML is biased from the true quantity of interest, even for event study pre-trend coefficients. I show that the proposed estimator from this paper estimates the true percent change of the average of the treated sample, even when treatment is heterogeneous across time and individuals. In contrast, aggregation strategies estimator recovers the average parameter from the model only when there is no individual heterogeneity within treated cohorts.

I finally apply my estimator to empirical questions from the Economics literature. I revisit the effect of information exchange on request on bank deposits held in tax havens (Johannesen and Zucman, 2014; Menkhoff and Miethe, 2019), a significant public policy change at the begining of the 21st century. Researchers have tested whether treaties of exchange of information on request decrease cross-border deposits owned in tax havens and resulting from tax evasion. The set-up motivates the use of a nonlinear estimator and the estimation

of a proportional treatment effect. Bilateral deposits are censored to positive values only, and country pairs display very different baseline cross-border owned deposits. Treaties are passed at different times is staggered and likely to be heterogeneous by time and country-pairs, providing the ideal setting to test for robustness for recent bias of TWFE estimators. I find that the author's estimate have a small positive staggered treatment bias. However, the treated cohort display very large treatment effect heterogeneity, which causes the difference-in-difference estimates of the log-linearized model (log-OLS) to differ by a lot from the ratio-of-ratio (PPML) estimates. More precisely, even though treaties tend to cause a large a negative effect on tax havens deposit *on average*, their effect on the average volume of deposits held offshore is weaker as some large country-pairs react positively or do not react by much. I show that in this case, the proposed imputation estimator from this paper recovers a slightly smaller quantity than the TWFE PPML estimator, indicating that the later attenuates the magnitude true treatment effect. However, using an aggregation strategy by analogy with the linear case would strongly overestimate the bias due to staggered treatment. Finally, I replicate the empirical exercise of Nagengast and Yotov (2025) as described above.

This paper relates to several parts of the literature in applied econometrics. It relates first to a literature motivating the use of PPML estimators for multiplicative model estimation (Wooldridge, 1999; Silva and Tenreyro, 2006; Cohn et al., 2022; Chen and Roth, 2023) and to a literature on non-linear difference-in-differences (Angrist, 2001; Ciani and Fisher, 2019; Wooldridge, 2023). I show that PPML estimators can be easily extended to counterfactual estimators robust staggered treatment timings. I contribute to the literature on the interpretation of models estimating semi-elasticities (Kennedy, 1981; Jan van Garderen and Shah, 2002). I show that in the presence of heterogeneous non-linear treatment effects, different aggregation of individual or group level treatment effects yeld semi-elasticities with very different causal interpretations, some having more micro, intermediate or macro inter-

pretations. Applied researchers should keep in mind the desired interpretation they wish to recover. I further contribute to the literature on the estimation of treatment effects with difference-in-differences in the presence of heterogeneous treatment effects and binary treatment (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; De Chaisemartin and D'Haultfoeuille, 2023; Nagengast and Yotov, 2025). I show that the TWFE PPML estimator is biased in the staggered case, with some treatment effect scaling "negatively" the estimate. The conditions under which this estimator is biased are the same as for the linear case. I propose a new estimator robust to staggered treatment bias that recovers the ratio-of-ratios, a quantity similar to one yielded by the TWFE PPML estimator in the two-by-two canonical case. Finally, I contribute to the literature on counterfactual estimators (Borusyak et al., 2024; Liu et al., 2024), by developing a nonlinear counterfactual estimator.

The rest of the paper proceeds as follows. Section 2 presents the 2x2 canonical setting of multiplicative differences and the two-way fixed effect Poisson Pseudo-Maximum Likelihood estimator (TWFE PPML). Section 3 presents the staggered treatment case, the setting induced bias of TWFE PPML and a robust estimator to recover the ratio-of-ratios. Section 4 displays simulations comparing existing estimators in the canonical and staggered cases. Section 5 presents empirical applications. Section 6 concludes.

2 The 2x2 canonical setting

The researcher is interested in a policy or economic change affecting economic units denoted i observed through time t. There are N units observed. The change (from $D_{it} = 0$ to $D_{it+1} = 1$) affects a non-negative outcome of interest y_{it}. We have that the realized outcome $y_{it} = D_{it}y_{it}(1) + (1 - D_{it})y_{it}(0)$, with $y_{it}(1)$ and $y_{it}(0)$ the potential outcomes. In the canonical set-up, there are two groups of units $g = 0, 1$, at two periods $t = 0, 1$. Group 1 is treated at period

1 (i.e., the policy is implemented), and group 0 is never treated.

2.1 Quantity of interest and identification

In the case of multiplicative models, the researcher is often interested in the proportional treatment effect. The multiplicative difference-in-difference targets (Angrist, 2001):

$$\frac{E[y_{igt}(1)|D=1] - E[y_{igt}(0)|D=1]}{E[y_{igt}(0)|D=1]} = \frac{ATT}{E[y_1(0)|D=1]} = PTT \tag{2.1}$$

This quantity is the change in the outcome induced by the treatment among the treated, or the ATT, scaled by the non-treated outcome. It is the change of the expected outcome variable in percentage of the expected outcome in the absence of treatment: a semi-elasticity.[6]

2.1.1 Identifying assumptions

$E[y_1(1)|D=1]$ can be directly estimated from coresponding moments in the data, but not $E[y_1(0)|D=1]$ which is by definition never observed. Further assumptions allow estimating the ATT and PTT.

A1: No anticipation assumption On average, in the eventually treated group, there are no anticipatory changes that affect the potential outcomes before the intervention.

$$E[y_0(1) - y_0(0)|D=1] = 0 \tag{2.2}$$

A2: Multiplicative parallel trend assumption (MPT) This assumption states that in the absence of treatment, changes in percentages of expected outcomes should have been the same in the two groups. The averages of the two groups would have shown the same growth

[6]This is also a quantity that Chen and Roth (2023) advise to target when the researcher wants to include zeros and recover a proportional treatment effect.

in the absence of treatment.[7]

$$\frac{E[y_1(0)|D=1]}{E[y_0(0)|D=1]} = \frac{E[y_1(0)|D=0]}{E[y_0(0)|D=0]} \tag{2.3}$$

If it holds conditionally to some covariates X_{it}:

$$\frac{E[y_1(0)|D=1,X]}{E[y_0(0)|D=1,X]} = \frac{E[y_1(0)|D=0,X]}{E[y_0(0)|D=0,X]} \tag{2.4}$$

2.1.2 Identification

$E[y_1(0)|D=1]$ can be expressed as a function of terms that can be estimated using the multiplicative parallel trend assumption (A.2):

$$E[y_1(0)|D=1] = \frac{E[y_1(0)|D=0] \times E[y_0(0)|D=1]}{E[y_0(0)|D=0]}$$

We use (2.3) in (2.1) and recover the PTT expressed as a Ratio-of-Ratios (by analogy to a difference-in-differences in the linear case):

$$PTT = \frac{E[y_1(1)|D=1]}{E[y_0(0)|D=1]} \Big/ \frac{E[y_1(0)|D=0]}{E[y_0(0)|D=0]} - 1 \tag{2.5}$$

The expression of the ATT follows:

$$ATT = E[y_1(1) - y_1(0)|D=1] = E[y_1(1)|D=1] - \frac{E[y_1(0)|D=0] \times E[y_0(0)|D=1]}{E[y_0(0)|D=0]} \tag{2.6}$$

2.2 Estimation

2.2.1 Corresponding sample moments

The PTT and ATT can be estimated from their corresponding sample moments. With G_i a binary variable taking the value 1 if the individual i belongs to the treated group, and y_{it} the outcome of i at time t, the proportional treatment effect is estimated by computing a ratio of

[7] This assumption is also called the index parallel trend assumption by Wooldridge (2023).

ratios (RoR), relying on the multiplicative parallel trend assumption:

$$\begin{aligned}\widehat{RoR} &= \frac{\frac{\sum_{i=1}^{n} G_i(y_{i,1})}{\sum_{i=1}^{n} G_i}}{\frac{\sum_{i=1}^{n} G_i(y_{i,0})}{\sum_{i=1}^{n} G_i}} \Big/ \frac{\frac{\sum_{i=1}^{n} (1-G_i)(y_{i,1})}{\sum_{i=1}^{n} (1-G_i)}}{\frac{\sum_{i=1}^{n} (1-G_i)(y_{i,0})}{\sum_{i=1}^{n} (1-G_i)}} - 1 \\ &= \frac{\sum_{i=1}^{n} G_i(y_{i,1})}{\sum_{i=1}^{n} G_i(y_{i,0})} \Big/ \frac{\sum_{i=1}^{n} (1-G_i)(y_{i,1})}{\sum_{i=1}^{n} (1-G_i)(y_{i,0})} - 1\end{aligned} \tag{2.7}$$

When N grows, this is a consistent estimator of the PTT. And $\widehat{\tau}$ estimates the ATT:

$$\widehat{\tau} = \frac{\sum_{i=1}^{n} G_i(y_{i,1})}{\sum_{i=1}^{n} G_i} - \frac{\frac{\sum_{i=1}^{n} (1-G_i)(y_{i,1})}{\sum_{i=1}^{n} (1-G_i)} \times \frac{\sum_{i=1}^{n} G_i(y_{i,0})}{\sum_{i=1}^{n} G_i}}{\frac{\sum_{i=1}^{n} (1-G_i)(y_{i,0})}{\sum_{i=1}^{n} (1-G_i)}}$$

$$\widehat{\tau} = \frac{1}{\sum_{i=1}^{n} G_i} \Big(\sum_{i=1}^{n} G_i(y_{i,1}) - \frac{\sum_{i=1}^{n} (1-G_i)(y_{i,1})}{\sum_{i=1}^{n} (1-G_i)(y_{i,0})} \times \sum_{i=1}^{n} G_i(y_{i,0}) \Big) \tag{2.8}$$

In the right part of this expression, the average outcome of the treated group in period 0 is multiplied by the growth rate of the non-treated group between the two periods.

2.2.2 Equivalence of TWFE PPML and ROR estimator

In the linear canonical setting, there is a direct equivalence between the moments used and the recovered difference-in-differences and TWFE OLS estimated coefficients. A similar analogy holds in the multiplicative model case, and TWFE PPML, which recovers the RoR (Ciani and Fisher, 2019; Chen and Roth, 2023). The TWFE PPML estimator maximizes a quasi-log-likelihood based on the following conditional mean:

$$E[y_{it}|D_{it}] = exp(\alpha_i + \beta_t + \delta D_{it}) \tag{2.9}$$

In the canonical setting, we have:

$$exp(\widehat{\delta_{PPML}}) - 1 = \frac{\sum_{i=1}^{n} G_i(y_{i,1})}{\sum_{i=1}^{n} G_i(y_{i,0})} \Big/ \frac{\sum_{i=1}^{n} (1-G_i)(y_{i,1})}{\sum_{i=1}^{n} (1-G_i)(y_{i,0})} - 1 \tag{2.10}$$

The same quantity as in (2.7) that converges in probability, under the identification assumptions, to the quantity of interest (2.1).

2.2.3 Structural modeling approach

Structural modeling comes naturally from equation 2.9, potentially allowing for δ_i heterogeneous treatment effects. The researcher observes:

$$y_{it} = exp(\alpha_i + \beta_t + \delta_i D_{it})\eta_{it} \tag{2.11}$$

With η_{igt} captures remaining individual-time varying heterogeneity such that $E[\eta_{it}|D_{it}] = 1$. Using model 2.11 notations, another version ot the parallel trend assumption is:

$$E[y_{it}(0)|D_{it}] = exp(\alpha_i + \beta_t), \quad \forall(i,t) \tag{2.12}$$

In case of homogeneous treatment effect across individuals, the quantity of interest 2.1 corresponds to $exp(\delta) - 1$ from our model. The model can also be extended to include a vector of covariates X_{it}: $E[y_{it}|D_{it}] = exp(\alpha_i + \beta_t + \delta_i D_{it} + X'_{it}\gamma)$.

2.3 Difference with log-linear DiD

Researchers also rely on logarithm transformations of the outcome to estimate semi-elasticities. The log-linear DiD differs from the multiplicative DiD on several dimensions. Potential outcomes are now defined by $\ln y_{it} = D_{it} \ln y_{it}(1) + (1 - D_{it}) \ln y_{it}(0)$ and structural modeling follows: $\ln y_{it} = \alpha_i + \beta_t + \delta D_{it} + \ln \eta_{it}$.

2.3.1 Quantity of interest

The log-linear DiD targets a different quantity of interest when treatment effects are heterogeneous:

$$ATT = E[\ln y_1(1)|D = 1] - E[\ln y_1(0)|D = 1] = E[\delta_i|D = 1] \neq \ln E[exp(\delta_i)|D = 1] \tag{2.13}$$

The target of the linear model is the average log point change, which corresponds to the average parameter δ_i in the structural approach. This is the approximated average percentage

change when treatment effects δ_i are small, and not the average individual proportional effects (Jensen's inequality). The two model estimation targets are the same only when the treatment effect is homogeneous: $\delta_i = \delta, \forall i$.

2.3.2 Identification assumption

The no-anticipation assumption is:

$$E[\ln y_0(1) - \ln y_0(0)|D = 1] = 0 \tag{2.14}$$

The parallel trend assumption:

$$E[\ln y_1(0) - \ln y_0(0)|D = 1] = E[\ln y_1(0) - \ln y_0(0)|D = 0] \tag{2.15}$$

This assumption states that in the absence of treatment, the expected log of the outcome in the treated group should have changed by the same log points as the non-treated group. The parallel trend is on the growth of the averages and not on the average growths. There is not reason why the two should hold at the same time. With more pre-treatment time periods observed, one can undertake a visual exploration on pre-trends to check which assumption seems most plausible to hold. In the case of the multiplicative model, the pretrend should be similar when the researcher plots the logarithm of the average (or total with a balanced panel) outcome for treated and control groups. In case of the log-linear model, the pretrend should be similar when the researcher plots the average logarithm of the outcome for treated and control groups.

Finally, if one follows the structural modeling different assumptions rest on the error terms. The multiplicative approach assumes that $E[\eta_{it}|D_{it}] = 1$ while the log-OLS approach requires that $E[\varepsilon_{it}|D_{it}] = 0$. However $\varepsilon_{it} = \ln \eta_{it}$ if there η_{it} is heteroskedastic and its variance depends on treatement status, there will usually be that $E[\varepsilon_{it}|D_{it}] = f(D_{it})$ and log-OLS will miss its quantity of interest. Conceptually, this means that if the treatment affects both the

mean of y_{it} and its variance, the log-linear DiD will aggregate the two (potientially opposite) effects. This issue is already extensively discussed by Silva and Tenreyro (2006); Ciani and Fisher (2019); Cohn et al. (2022); Chen and Roth (2023) and I let the reader refer to their work for more details.

2.3.3 Zeros

Zeros in the outcome y_{it} are notoriously excluded from the estimation sample of the log-linear DiD. This exclusion is natural from the log-linear model because a proportional change for the extensive margin is not defined. TWFE PPML solves this issue by estimating a quantity that weights predicted individual proportional changes by their *predicted* counterfactual outcome share in total predicted counterfactual outcome:

$$PTT = \frac{E[y(1)|D=1] - E[y(0)|D=1]}{E[y(0)|D=1]} = \sum_{i,t,D_{it}=1} \frac{E(y_{it}(0))}{\sum_{i,t,D_{it}=1} E(y_{it}(0))} \Big(exp(\delta_{it}) - 1\Big) \quad (2.16)$$

Intuitivelly, TWFE PPML provides small weights to zeros or small observations, because they have a small contribution to the total outcome and the model predicts small counterfactual outcomes.[8]

3 Multiperiod setting and heterogeneous timing

I turn to the multiperiod and multicohort setting. There are now T time periods starting at $t = 1$, and G cohorts denoted g, of N units i treated at different times. Cohorts are groups of units treated at the same time, and g denotes the time of treatment. The never-treated cohort is denoted $g = \infty$. The potential outcomes are now defined by $y_{igt} = \mathbb{1}\{g \leq t\} y_{igt}(1) + (1 - \mathbb{1}\{g \leq t\}) y_{igt}(0)$.

[8] Observations with zero in the outcome will be included in the PPML estimation sample only if the observation for the same individual in the other period is strictly positive.

A1: No anticipation assumption On average, among the eventually treated group, there are no anticipatory changes that affect the potential outcomes before the intervention.

$$E[y_{gt}(1) - y_{gt}(0)|D_{gt}] = 0 \quad \forall t < g \tag{3.1}$$

A2: Multiplicative parallel trend assumption For $g \leq t$ and $g' > t$

$$\frac{E[y_{gt}(0)|D_{gt}]}{E[y_{gt-1}(0)|D_{gt-1}]} = \frac{E[y_{g't}(0)|D_{g't}]}{E[y_{g't-1}(0)|D_{g't-1}]} \tag{3.2}$$

This is equivalent to assuming that in the absence of treatment, the growth rate of the average outcome in the treated cohort between two time periods would have been the same as in the non-yet-treated and never-treated cohorts.[9]

3.1 TWFE PPML bias

The TWFE OLS estimators in a multiperiod multigroup setting can lead to biased estimates of the ATT because the model makes too strict assumptions on treatment homogeneity. When units are treated at different times and treatment effects are heterogeneous across time, the TWFE estimator makes wrong comparisons between treated and control groups, and estimates a quantity that averages treatment effects with negative weights (De Chaisemartin and D'Haultfoeuille, 2023).

In a simple example, this problem also arises with TWFE PPML and the multiplicative DiD. There are two individuals $i = A, B$ observed at three time periods $t = 1, 2, 3$. Individual A is treated in period $t = 2$ and individual B is treated in period $t = 3$, such that B is the control group for individual A in $t = 2$. The conditional mean of the outcome is:

$$E[y_{it}|D_{it}] = exp(\alpha_i + \beta_t + \delta_{it}D_{it})$$

If treatment effect is homogeneous, there is $\delta_{A2} = \delta_{A3} = \delta_{B3} = \delta$. If we have heterogeneous

[9] With structural modeling approach: $E[y_{igt}(0)|D_{igt}] = exp(\alpha_i + \beta_t), \quad \forall (i, g, t)$.

treatment effect then $\delta_{A2} \neq \delta_{A3} \neq \delta_{B3}$. The quantity of interest is then:

$$PTT = \frac{E[y(1)|D=1] - E[y(0)|D=1]}{E[y(0)|D=1]} = \sum_{i,t,D_{it}=1} \frac{E[y_{it}(0)|D=1]}{\sum_{i,t,D_{it}=1} E[y_{it}(0)|D=1]} \Big(exp(\delta_{it}) - 1\Big) \tag{3.3}$$

Which is a weighted sum of cohort and time-specific treatment effects $exp(\delta_{it}) - 1$. The weights ω_{it} correspond to the share of the counterfactual outcome in the total size of counterfactual observations.[10]

Solving the system from the log-likelihood first order conditions (see in Appendix) yields the TWFE PPML estimator for the proportional treatment effect $exp(\delta) - 1$:

$$exp(\widehat{\delta}_{PPML}) - 1 = \frac{y_{A2}(y_{B1} + y_{B3}) - y_{B2}(y_{A1} + y_{A3}))}{y_{B2}(y_{A1} + y_{A3}))} \tag{3.4}$$

With homogeneous treatment effect, using expected values of outcome realization, this quantity should yield:

$$\frac{E[y_{A2}(y_{B1} + y_{B3})|D_{it}] - E[y_{B2}(y_{A1} + y_{A3})|D_{it}]}{E[y_{B2}(y_{A1} + y_{A3})|D_{it}]} = exp(\delta) - 1 \tag{3.5}$$

With treatment heterogeneity, the quantity estimated by TWFE PPML becomes:

$$\frac{E[y_{A2}(y_{B1} + y_{B3})|D_{it}] - E[y_{B2}(y_{A1} + y_{A3})|D_{it}]}{E[y_{B2}(y_{A1} + y_{A3})|D_{it}]} = exp(\delta_{A2}) \times \frac{1 + exp(\delta_{B3} + \beta_3)}{1 + exp(\delta_{A3} + \beta_3)} - 1 \tag{3.6}$$

The TWFE PPML recovers here the growth rate of the only available "comparison period" $(t = 2)$, scaled by the differential in growth rate between the two groups in the second period. This scaling will be bigger if the common trend in this later period is large $(exp(\beta_3)$ is high$)$. There is an analogy with the problem encountered in the linear case, with some treatment effects scaling down the treatment effect, and potentially reverting the sign of the estimated effect.

[10] The PPML estimator weights more cells with large counterfactual outcomes and reduces weights associated with cells with smaller counterfactual outcomes, which are the most susceptible to display the most extreme proportional changes.

3.2 Robust estimators for TWFE PPML

Recent papers solve this issue in the linear case by allowing for the most flexible model given the data structure (Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2021). Wooldridge (2023) extends this idea to the non-linear case. With g_{iq} an indicator variable taking the value one if individual i is treated in period q, one can estimate the model corresponding to this conditional mean using PPML:

$$E[y_{it}|D_{it}] = exp\Big[\sum_{r=q}^{T}\sum_{l=0}^{T-r}\delta_{rs}\Big(D_{it}\times g_{ir}\times \mathbb{1}\{t-r=l\}\Big)+\alpha_i+\beta_t\Big] \tag{3.7}$$

In this model:

$$\begin{aligned}\delta_{gt} &= \ln(E[y_{gt}(1)|D=1]) - \ln(E[y_{gt}(0)|D=1]) \\ \Leftrightarrow exp(\delta_{gt}) - 1 &= \frac{E[y_{gt}(1)|D=1] - E[y_{igt}(0)|D=1]}{E[y_{igt}(0)|D=1]}\end{aligned} \tag{3.8}$$

So estimating δ_{gt} recovers the estimation target at the cohort-time level: the proportional treatment effect on cohort g and time t. The researcher is often interested in a more aggregated quantity of interest.

3.2.1 Aggregation estimators in the non-linear case

Robust estimators have been developed for the linear case to recover aggregate treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2021). These estimators rely on recovering treatment effects for correct building blocks (i.e. cohorts-time DiD) and aggregating them over the desired sample to recover an estimate of the ATT. Given that the models used are linear, the ATT can be easily retrieved by aggregating linear treatment effects.

Translated in the multiplicative setting, one could also compute the two-by-two estimates of $PTT_{g,t}$ by group and time-period, and average this effect to recover an aggregate treatment

effect. This would yield an estimator of the form:

$$\sum \nu_{g,t} \widehat{RoR}_{g,t} \tag{3.9}$$

With $\nu_{g,t}$ a weight associated with observations in g, t, chosen by the researcher depending on the estimation target.

Nagengast and Yotov (2025) uses the fully interacted model above and follows such an "aggregation" strategy. Coefficients δ_{gt} recover the multiplicative model estimation target for each cohort-time cell: $exp(\widehat{\delta_{gt}^{PPML}}) - 1$ is the multiplicative effect on the average of cohort g at time t. Their aggregation estimator with the same spirit as the linear case is:

$$exp\Big(\sum_{g}^{G}\sum_{t}^{T} \nu_{g,t}(\widehat{\delta_{g,t}^{PPML}})\Big) - 1 \tag{3.10}$$

Estimation can be easily implemented using the `ppmlhdfe` Stata command (Correia et al., 2020) when the number of parameters to estimate gets big: interaction coefficients can be estimated as fixed effect, appropriately rescaled and aggregated to recover (3.10). This quantity is a consistent estimator for:

$$\begin{aligned} &exp\Big(\sum_{g}^{G}\sum_{t}^{T} \nu_{g,t}(\delta_{g,t})\Big) - 1 \\ &= exp\Big(\sum_{g}^{G}\sum_{t}^{T} \nu_{g,t}(log\Big(\frac{E[y_{gt}(1)|D=1] - E[y_{gt}(0)|D=1]}{E[y_{gt}(0)|D=1]}\Big) + 1)\Big) - 1 \end{aligned} \tag{3.11}$$

If treatment is homogeneous within cohort-time cells, i.e. $\delta_{igt} = \delta_{gt} \quad \forall i, g, t$, this estimator approximates the average log-point change:

$$exp\Big(\sum_{g}^{G}\sum_{t}^{T} \nu_{g,t}\delta_{gt}\Big) - 1 = exp\Big(\sum_{g}^{G}\sum_{t}^{T} \nu_{g,t} E[\ln y_{gt}(1) - \ln y_{gt}(0)|D=1]\Big) - 1 \tag{3.12}$$

This is the estimation target of the log-linear model.[11] As treatment effects are heterogeneous across cohorts and time, it will be a different quantity than the percentage change in the average targeted in the canonical case. It should therefore not be compared to TWFE PPML

[11] The advantage is that it is robust to assuming $E[\eta_{igt}|D] = 1$.

to assess the bias caused by the staggered treatment, because the two are computing different quantities.

In the more general case, if we do not constraint treatment effects to be the same within cohort-time cells ($\delta_{igt} \neq \delta_{gt}$) the quantity recovered by this estimator might not have an interpretable meaning. In this case, the estimated coefficient $\widehat{\delta_{g,t}^{PPML}}$ will recover the proportional treatment effect on the average of cell g, t. The interpretation of (3.10) becomes the *approximate* average over cells of multiplicative treatment effect on the average of cells. This is an intermediate quantity between the estimated parameter (log-linear DiD) and the estimated growth rate of the average (ratio-of-ratios). The way those three quantities compare will depend on the correlation between treatment effects δ_{igt} and counterfactual outcomes $y_{igt}(0)$.

If $corr(\delta_{igt}, y_{igt}) > 0$, we will have that, in terms of estimation targets:

- log-linear DiD ($\sim$ average % change) < Aggregation PPML < Ratio-of-Ratios (% change of the average).

If $corr(\delta_{igt}, y_{igt}) < 0$, we will have that:

- log-linear DiD > Aggregation PPML > Ratio-of-Ratios

If the definition of cohort makes sense from an economic point of view (eg, a cohort is a region), one can be interested in targeting a quantity of interests that is an average over cohort-time effects: the average over yearly regional employment change for example. However inference might depend on the number of cohorts G now rather than on units N. If treatment cohorts groupings do not have a relevant economic meaning, the interpretation of the aggregation PPML estimate will arbitrarily depend on the structure of the panel and the treatment timings, and might lack causal interpretation.

3.2.2 Proposed imputation estimator

This section proposes a new estimator for proportional treatment effects, recovering a semi-elasticity derived from the ratio-of-ratios estimator. This estimator is robust to any type of treatment heterogeneity in a staggered treatment setting. It is an imputation estimator in the spirit Borusyak et al. (2024), based on the idea that one can specify the correct counterfactual model. Wooldridge (2023) shows that this approach, under some conditions, is equivalent to the fully interacted model and I derive the equivalent interaction estimator in appendix.

Under our identification assumptions, the expected conditional mean of the counterfactual outcome is: $E[y_{igt}(0)|D_{igt} = 1] = exp(\alpha_i + \beta_t)$. The parameters α_i and β_t can be estimated on the sample of never-treated and not-yet-treated observations. On this sample, the conditional mean is correctly specified and TWFE PPML consistently estimates each set of fixed effects. One can then predict the counterfactual outcomes for the treated sample, using estimates of these estimates:

$$\widehat{y_{igt}(0)} = exp(\widehat{\alpha_i} + \widehat{\beta_t})$$

Wooldridge (2023) states that, with $N_{g,t}$ the number of treated observations in cell (g, t):

$$\widehat{\tau_{g,t}} = \frac{1}{N_{g,t}} \sum_{i \in g} y_{igt}(1) - \widehat{y_{igt}(0)}$$

Estimates the ATT in level for cohort g and time t, and can also be recovered by predicting the treatment average partial effect for cell g, t. Contrary to coefficients $\widehat{\delta_{g,t}}$, this is a linear effect that can be aggregated linearly without loss of interpretability. I recover the average treatment effect in level on the full treated sample, which is equivalent to computing the difference between the observed outcome and the predicted one on the treated sample:

$$\widehat{\bar{\tau}} = \sum_{g,t,D=1} \frac{N_{g,t}}{N_D} \widehat{\tau_{g,t}} = \frac{1}{N_D} \sum_{i,t} D_{igt}(y_{igt}(1) - \widehat{y_{igt}(0)}) \tag{3.13}$$

With N_D the size of the total treated sample. To recover the proportional treatment effect, or

treatment semi-elasticity, this quantity can be scaled by the total counterfactual outcome to recover the following estimator:

$$\begin{aligned}\widehat{RoR}_{imput} &= \frac{\hat{\bar{\tau}}}{\frac{1}{N_D}\sum_{i,t} D_{igt}\widehat{y_{igt}(0)}} \\ &= \frac{\frac{1}{N_D}\sum_{i,t} D_{igt}y_{igt}(1)}{\frac{1}{N_D}\sum_{i,t} D_{igt}\widehat{y_{igt}(0)}} - 1 = \frac{\sum_{i,t} D_{igt}y_{igt}(1)}{\sum_{i,t} D_{igt}\widehat{y_{igt}(0)}} - 1\end{aligned} \tag{3.14}$$

This estimator is based on the ratio of the average of observed and counterfactual outcomes. Its interpretation is similar to the TWFE PPML and RoR estimator in the canonical setting: the percentage change in the average outcome due to treatment.

When N grows, the numerator converges in probability to the expected value of the treated outcome in the treated group. The denominator, under the parallel trend assumption, converges to the expected value of the untreated outcome in the treated group. It is obtained by multiplying the average outcome of the treated groups in the pre-treatment period by the growth rate of the non-treated group after treatment.

The ratio of the two should converge in probability to the true PTT when N grows, provided that the denominator does not reach zero. This is unlikely to take place: the PPML model only predicts strictly positive values. Moreover for the model to predict counterfactual outcomes very close to zero, it means that the researcher faces a DGP in which treatment affects mainly the extensive margin, and therefore is more suited for a binary outcome model. A panel bootstrap, with resampled units, can be used to recover standard errors for $\widehat{RoR}_{imput}$.

The estimator $\widehat{RoR}_{imput}$ can be easily computed for a less aggregated level, such as cohort or relative time. For example, the change in the average for cohort h would be:

$$\widehat{RoR}_{imput,h} = \frac{\sum_{i,t} D_{iht}y_{iht}(1)}{\sum_{i,t} D_{iht}\widehat{y_{iht}(0)}} - 1 \tag{3.15}$$

And the change on the average at relative time to treatment date l:

$$\widehat{RoR}_{imput,l} = \frac{\sum_t \frac{\mathbb{1}\{g=t-l\}}{N_{g,t}}\sum_i y_{igt}(1)}{\sum_t \frac{\mathbb{1}\{g=t-l\}}{N_{g,t}}\sum_i \widehat{y_{igt}(0)}} - 1 \tag{3.16}$$

To explore potential anticipation effects of the policy, one could compute the "leads" coefficients, either by gradually removing negative relative treatment years from the non treated sample and constructing $\widehat{RoR}_{imput,l}$ for $l = -1, -2, ...$ as in Borusyak et al. (2024).[12]

3.2.3 Special cases for the imputation estimator

Alternative weighting of quantities and scales The estimator $\widehat{RoR}_{imput}$ scales a quantity called the "simple-weighted ATT" by Callaway and Sant'Anna (2021) by a simple-weighted counterfactual outcome. One could consider different weighting schemes to apply to both the numerator and the denominator of the proportional treatment effect estimator. I discuss two weighting schemes in particular but this discussion can be extended according to the researcher's quantity of interest.[13]

The estimator in 3.14 scales the ATT by the average counterfactual; as such, it gives more weight to cohorts that are observed for the longest time. One possibility is to compute first ATT and counterfactual cohort averages, and aggregate cohort effects weighted by the size of each of them in terms of treated units:

$$\widehat{RoR}^{sel}_{imput} = \frac{\sum_g \frac{1}{N_{\overline{g}}} \left(\frac{1}{T-g+1} \sum_{t=g}^{T} \frac{1}{N_{g,t}} \sum_i y_{igt}(1) \right)}{\sum_g \frac{1}{N_{\overline{g}}} \left(\frac{1}{T-g+1} \sum_{t=g}^{T} \frac{1}{N_{g,t}} \sum_i \widehat{y_{igt}(0)} \right)} - 1 \tag{3.17}$$

With $N_{\overline{g}}$ the number of units i in cohort g.

Another issue, when plotting event-study types of estimates, is that the difference between coefficients for two relative times l and l' captures the treatment dynamic and the composition effect as some cohorts disappear. An alternative definition for $\widehat{RoR}_{imput,l}$ is to fix the cohort composition at relative time l with a comparison relatie time l':

$$\widehat{RoR}^{bal,l'}_{imput,l} = \frac{\sum_t \frac{\mathbb{1}\{g=t-l\} \times \mathbb{1}\{g+l' \leq T\}}{N_{g,t}} \sum_i y_{igt}(1)}{\sum_t \frac{\mathbb{1}\{g=t-l\} \times \mathbb{1}\{g+l' \leq T\}}{N_{g,t}} \sum_i \widehat{y_{igt}(0)}} - 1 \tag{3.18}$$

[12] If the goal is two compare to other estimator using $l = -1$ as a reference period, one can use this year of data to estimate α_i for the eventually treated group and the never treated cohort for β_t.

[13] I thank Jonathan Roth for his comments on this. See discussions in Callaway and Sant'Anna (2021) on other weighting strategies.

Categorial parallel trends Researchers often choose to specify categorical parallel trends, or parallel trends holding across some groups of the population. For example, if treated and control firms are compared over time within the same region or sector. In the TWFE model, this translates to specifying time fixed effects disaggregated by the desired categories denoted c:

$$y_{igct} = exp(\alpha_i + \beta_{ct} + \delta D_{igt})\eta_{igt}$$

The estimation of the treatment effect in level and counterfactual outcome requires slightly adjusting the counterfactual model and estimating more parameters. The imputation procedure only requires estimating $y_{ict} = exp(\alpha_i + \beta_{ct})\eta_{ict}$ on the treated sample for pre-treatment periods and the never-treated to get $\hat{\alpha_i}$ and $\hat{\beta_{ct}}$. The predicted counterfactual outcome on the treated sample:

$$\widehat{y_{igt}(0)} = exp(\widehat{\alpha_i} + \widehat{\beta_{ct}})$$

$$\widehat{\overline{\tau}} = \sum_{i \in \omega_1} y_{igt}(1) - \widehat{y_{igt}(0)}$$

Going through the imputation process is less computationally intensive than the interaction approach, especially when the number of categories c increases, and is numerically equivalent.

Control variables The imputation process allows for a flexible counterfactual model, as long as it is correctly specified.[14] The researcher assumes the true conditional mean to be:

$$E[y_{igt}|D_{igt}] = exp(\alpha_i + \beta_t + \delta_{it} D_{igt} + X'_{igt}\gamma) \tag{3.19}$$

[14]Note that the equivalence with the interaction approach breaks when introducing time-varying controls.

With X_{igt} a set of individual-specific, time-varying variables impacting the outcome y_{igt}. Under the conditional parallel trend (2.4), the counterfactual model can be estimated by:

$$\widehat{y_{igt}(0)} = exp(\widehat{\alpha_i} + \widehat{\beta_{ct}} + X'_{igt}\widehat{\gamma})$$

Which requires to estimate $\widehat{\alpha_i}$, $\widehat{\beta_{ct}}$ and $\widehat{\gamma}$ on the sample such that $D_{igt} = 0$. We recover the proportional treatment effect then as:

$$\widehat{RoR}_{imput} = \frac{\sum_{i\in\omega_1} y_{igt}(1) - \widehat{y_{igt}(0)}}{\sum_{i\in\omega_1} \widehat{y_{igt}(0)}} = \frac{\widehat{\bar{\tau}}_{imput}}{\sum_{i\in\omega_1} \widehat{y_{igt}(0)}}$$

Triple differences In a triple difference approach, researchers observe treated cohorts that differ along two additional dimensions, denoted as j and p, which are used to select control groups. These dimensions can represent sectors and products, or regions, and are used to correct the potential bias of a simple difference-in-differences estimator by cancelling out this bias using a supplementary dimension (Olden and Møen, 2022). The expected conditional mean takes the following form:

$$E[y_{i(jp)gt}|G, D_{it}] = exp(\alpha_i + \beta_{jt} + \beta_{pt} + \delta_{it}D_{igt}) \tag{3.20}$$

The new parallel trend assumption becomes:

$$E[y_{i(jp)gt}(0)|G] = exp(\alpha_i + \beta_{jt} + \beta_{pt}) = exp(\alpha_i) \times \underbrace{exp(\beta_{jt}) \times exp(\beta_{pt})}_{Relative\ growth\ rate} \tag{3.21}$$

Here, $exp(\beta_{jt})$ and $exp(\beta_{pt})$ denote the relative growth rates associated with the two dimensions. If j is a state and p a product, this assumption states that the relative growth rate between treated and non-treated products in the treated state should have been the same as in the non-treated states in the absence of treatment.

Using the imputation approach simplifies the analysis compared to interaction models, which require interacting all the cohort time interactions with the p and j dimensions to break down δ_{gs} coefficients. With the imputation approach, the expected outcome y_{ijpgt} is estimated

on the not-yet and never-treated samples as $exp(\alpha_i+\beta_{jt}+\beta_{pt})\eta_{ijt}$. The imputed counterfactual outcome $\widehat{y_{ijpgt}(0)}$ is then calculated as $exp(\widehat{\alpha_i}+\widehat{\beta_{jt}}+\widehat{\beta_{pt}})$. The estimate is recovered as above.

4 Simulations

I simulate data to compare the estimators presented in the previous sections.

4.1 Common treatment timing

4.1.1 Data generating process

I generate a panel of 10,000 individuals observed for three time periods $t = 1, 2, 3$. The outcome y_{it} follows a multiplicative data generating process:

$$y_{it} = exp(\alpha_i + \beta_t + \delta_i D_{it})\eta_{it}$$

With β_t the time effects, α_i the individual effects, D_{it} the treatment status and δ_i the treatment effect, and η_{it} a log-normal error term such that $E[\eta_{it}|D_{it}] = 1$. Individuals are treated in period 3, such that treatment timing is homogeneous. I generate some selection into treatment status, so that I need to implement a difference-in-differences strategy to recover the causal effect of treatment.

I also introduce heteroskedasticity in the error term as a function of observables: in one case, variance is a function of individual fixed effects, and in the other, it is a function of the treatment status. This second case jeopardizes retrieving a causal treatment effect via log-OLS. Finally, I simulate a homogeneous treatment effect and a case with heterogeneity in treatment effect across individuals. The average treatment effect is positive as $\delta > 0$ when it is homogeneous. Heterogeneity across individuals is normally distributed, such that the average growth rate corresponds to the growth rate of the average. Heterogeneity in treatment effect allows comparing estimators in the different quantity of interest estimated. The ob-

served outcome y_{it} is always strictly positive such that I abstract from the differences caused by including zeros in the sample. I simulate the data 1000 times.

Table 1 – Common timing: simulation cases

Case	$exp(V(\eta_{it}\|.)) - 1$	Treatment effect parameter
1	α_i	$\delta = 0.31$
2	$0.2D_{it}$	$\delta = 0.31$
3	α_i	$\delta_i = 0.31 + \nu_i, \quad \nu_i \sim \mathcal{N}(0, 0.5)$
4	$0.2D_{it}$	$\delta_i = 0.31 + \nu_i, \quad \nu_i \sim \mathcal{N}(0, 0.5)$

4.1.2 Simulations results

Table 2 displays the true distribution of the true treatment effects and of the estimators across simulations. The upper panel displays cases 1 and 2 with homogeneous treatment effect, and the lower panel displays cases 3 and 4 with heterogeneous treatment effect. In those later case, I provide the distribution of the average parameter $exp(\overline{\delta_i}) - 1$ and the true growth rate of the average $PTT = \frac{E[y_{igt}(1)|D=1] - E[y_{igt}(0)|D=1]}{E[y_{igt}(1)|D=1]}$. I estimate the treatment effect using TWFE PPML, the imputation estimator and TWFE log-OLS. Densities of estimators are displayed in Figure C1a in Appendix.

Two estimators are unbiased and strictly equivalent accross all cases: TWFE PPML and the imputation estimator. I compare TWFE PPML and TWFE log-OLS in cases 1 and 2. In case of homogeneous treatment and individual heteroskedasticity, in the left upper panel, log-OLS is the most efficient unbiased estimator, TWFE PPML displays twice the larger variance. When I introduce a change in the variance of the error term caused by a change in treatment, log-OLS estimates display a downward bias, while TWFE PPML and imputation are unbiased.

Cases 3 and 4 display heterogeneous treatment effects. In both cases, PPML estimates are quite close to the true growth rate of the average. Estimates are less precise when het-

Table 2 – 1000 simulations: canonical setting

Homogeneous treatment effect $exp(\delta) - 1$

	Case 1 $exp(V(\eta_{it}\|.)) - 1 = \alpha_i$				Case 2 $exp(V(\eta_{it}\|.)) - 1 = 0.2D_i$			
Estimator	Mean	St.D.	Min	Max	Mean	St.D.	Min	Max
$PTT = exp(\delta) - 1$	0.363	0	0.363	0.363	0.363	0	0.363	0.363
TWFE PPML/RoR	0.370	0.0810	0.112	0.658	0.362	0.0509	0.203	0.543
Imputation	0.370	0.0810	0.112	0.658	0.362	0.0509	0.203	0.543
TWFE log-OLS	0.364	0.0357	0.259	0.475	0.293	0.0266	0.200	0.367

Heterogeneous treatment effect $exp(\delta_i) - 1$

	Case 3 $exp(V(\eta_{it}\|.) - 1 = \alpha_i$				Case 4 $exp(V(\eta_{it}\|.)) - 1 = 0.2D_i$			
Estimator	Mean	St.D.	Min	Max	Mean	St.D.	Min	Max
$exp(\overline{\delta_i}) - 1$	0.363	0.0104	0.333	0.398	0.364	0.00991	0.334	0.396
True PTT	0.545	0.0125	0.508	0.585	0.545	0.0118	0.505	0.583
TWFE PPML/RoR	0.544	0.0961	0.233	0.981	0.545	0.0635	0.332	0.768
Imputation	0.544	0.0961	0.233	0.981	0.545	0.0635	0.332	0.768
TWFE log-OLS	0.363	0.0366	0.258	0.496	0.294	0.0281	0.195	0.379

eroskedasticity is correlated with individual effects (Case 3) rather than treatment (Case 4). Turning to TWFE log-OLS, we observe that the model recovers the exponential of the average parameter, a quantity that cannot be interpreted in terms of semi-elasticity. The magnitude of the treatment effect differs by a lot from the PPML estimates and the true growth rate of the average. Moreover, in case of treatment-related heteroskedasticity (Case 4 in the lower right panel) there is an added bias.

4.2 Staggered treatment

4.2.1 Data generating process

I generate a panel of 10,000 individuals observed for fifteen time periods $t = 1, ..., 15$. The outcome y_{igt} follows a multiplicative data generating process:

$$y_{igt} = exp(\alpha_i + \beta_t + \delta_{it} D_{igt})\eta_{igt}$$

With β_t the time effects, α_i the individual effects, D_{igt} the treatment status and δ_{it} the treatment effect. Finally η_{igt} a log-normal error term such that $E[\eta_{igt}|D_{igt}] = 1$. The setting gathers the conditions under which the TWFE bias arises. Individuals are treated in different periods starting at $t = 10$, such that treatment is staggered, and cohorts are indexed by g. Treatment effect is heterogeneous by time and individual.

I use two types of treatment heterogeneity. In the first case, heterogeneity depends on the time t. In the second case, I introduce individual heterogeneity that is normally distributed across individuals, on top of time heterogeneity. Heterogeneity in treatment effect is now distributed such that the growth rate of the average outcome is different from the average growth rate of the outcome due to the treatment for each cell (g, t) Again, I introduce heteroskedasticity in the error term as a function of observables: in one case variance is a function of individual fixed effects, and in the other, it is a function of the treatment status. The observed outcome y_{igt} is always strictly positive such that there is no difference between estimators driven by zeros in the outcome. I simulate the data 1000 times.

4.2.2 Simulation results

Results of simulations are displayed in table 4. The first two lines of each table panel represent a different quantity of interest based on the true model. The first line displays the exponential of the average parameter δ_{it} minus one, the estimation target of the log-linear DiD, which

Table 3 – Staggered treatement timing: simulation cases

Case	$exp(V(\eta_{it}\|.)) - 1$	Treatment effect parameter
1	α_i	$\delta_t = log(\|t - 12.5\|)$
2	$0.2D_{igt}$	$\delta_t = log(\|t - 12.5\|)$
3	α_i	$\delta_{it} = log(\|t - 12.5\|) + \nu_i, \quad \nu_i \sim \mathcal{N}(0, 0.5)$
4	$0.2D_{igt}$	$\delta_{it} = log(\|t - 12.5\|) + \nu_i, \quad \nu_i \sim \mathcal{N}(0, 0.5)$

approximates the average growth rate of the treated outcome. The second line displays the growth rate of the average treated outcome, which is the proportional treatment effect on the average (PTT). This is the quantity recovered by the ratio-of-ratios, to which PPML converges in the canonical setting. I compare five estimators: TWFE PPML, the proposed imputation estimator, an "aggregation" estimator for PPML based on equation 3.10, TWFE log-OLS and the log-linear estimator by Borusyak et al. (2024) robust to staggered settings. Densities of estimators are displayed in Figure C1b in Appendix.

The upper left panel presents the case with time constant heteroskedasticity and treatment heterogeneity by time. Even without treatment-induced heteroskedasticity, the TWFE log-OLS estimator falls behind all true quantities of interest. It is now lower than $exp(\overline{\delta_t}) - 1$ because of the staggered setting bias. I turn to the TWFE PPML estimator: results confirm that it also misses the multiplicative DiD or RoR estimation target in the staggered setting.

In contrast, the imputation estimator recovers a quantity close to 0.1 percentage points from the true PTT. In Figure C1b, the kernel density of the estimator over the 1000 simulations is centered around the true growth rate of the average. The aggregation estimator is below this value, and identifies the true average parameter, such as the estimator from Borusyak et al. (2024).

In the right upper panel, I introduce treatment-induced heteroskedasticity. The OLS es-

timator is now taking negative values for a large number of simulations, and the mean is at -0.00861. In case the true value of the parameter and the growth rate of the average are large (0.233 and 0.776), using the TWFE log-OLS estimator can lead to statistically non-significant parameters and potentially negative estimates. The imputation estimator recovers the true growth rate of the average. The aggregation estimator recovers the average parameter, contrary to the estimator from Borusyak et al. (2024) which suffers from the heteroskedasticity bias.

In the lower panels I introduce normally distributed individual heterogeneity on top of time heterogeneity. Individual treatment heterogeneity is centered around zero such that the true parameter average value stays the same. In both cases, only the imputation/interaction estimators are unbiased estimators of the PTT. Both TWFE estimators are biased. The aggregation estimator is now different from the average parameter. It targets an intermediate quantity between the average parameter and the growth rate. It averages estimates of true RoRs at the g, t level, weighting them by the relative size of the population of cell g, t in the treated sample. The estimator from Borusyak et al. (2024) identifies its quantity of interest only when there is no heteroskedasticity bias.

Table 4 – 1000 simulations: staggered treatment

Heterogeneous treatment effect by time $exp(\delta_t) - 1$

	Case 1 $exp(V(\eta_{it}\|.)) - 1 = \alpha_i$				Case 2 $exp(V(\eta_{it}\|.)) - 1 = 0.2D_i$			
Estimator	Mean	St.D.	Min	Max	Mean	St.D.	Min	Max
$exp(\overline{\delta_t}) - 1$	0.233	0.00354	0.220	0.247	0.233	0.00254	0.226	0.240
True PTT	0.776	0.0133	0.734	0.824	0.776	0.00941	0.748	0.809
Imputation	0.780	0.0879	0.398	1.096	0.775	0.0550	0.559	0.954
Aggregation	0.234	0.0471	0.0721	0.395	0.232	0.0291	0.141	0.332
Borusyak et al. (2024)	0.233	0.0161	0.178	0.282	0.170	0.0121	0.128	0.203
TWFE PPML	0.198	0.0567	0.0166	0.470	0.194	0.0360	0.0790	0.344
TWFE Log-OLS	0.0452	0.0125	-0.000729	0.0877	-0.00861	0.00968	-0.0364	0.0215

Heterogeneous treatment effect by time and individuals $exp(\delta_{it}) - 1$

	Case 3 $exp(V(\eta_{it}\|.)) - 1 = \alpha_i$				Case 4 $exp(V(\eta_{it}\|.)) - 1 = 0.2D_i$			
Estimator	Mean	St.D.	Min	Max	Mean	St.D.	Min	Max
$exp(\overline{\delta_{it}}) - 1$	0.233	0.00592	0.216	0.251	0.233	0.00440	0.221	0.247
True PTT	1.011	0.0307	0.909	1.113	1.013	0.0228	0.939	1.112
Imputation	1.010	0.102	0.432	1.359	1.016	0.0675	0.770	1.228
Aggregation	0.394	0.0529	0.185	0.592	0.397	0.0345	0.274	0.494
Borusyak et al. (2024)	0.233	0.0174	0.178	0.295	0.169	0.0146	0.121	0.217
TWFE PPML	0.340	0.0693	0.121	0.703	0.340	0.0472	0.164	0.499
TWFE Log-OLS	0.0448	0.0130	0.00833	0.0951	-0.00802	0.01000	-0.0350	0.0279

I turn to a dynamic specification to compare estimators. I study them in the full heterogeneity case, with treatment-induced heteroskedasticity and heterogeneous treatment effect across time and individuals (Case 4 of table 4). Figure 1 plots the results of the estimators in one simulation. All coefficients are expressed relative to $t-1$ to ensure that estimators have a similar reference point (Roth, 2024).[15] I estimate leads and lags for the TWFE log-OLS and PPML, the imputation and interaction estimators and the aggregation estimator. To derive confidence intervals more easily, I plot $log(\theta_l + 1)$ for each estimator with θ_l being an estimate of a non-linear treatment effect at l time periods of treatment. This corresponds to estimates

[15] I do not include the estimator from Borusyak et al. (2024) now as it has a different interpretation of leads coefficients.

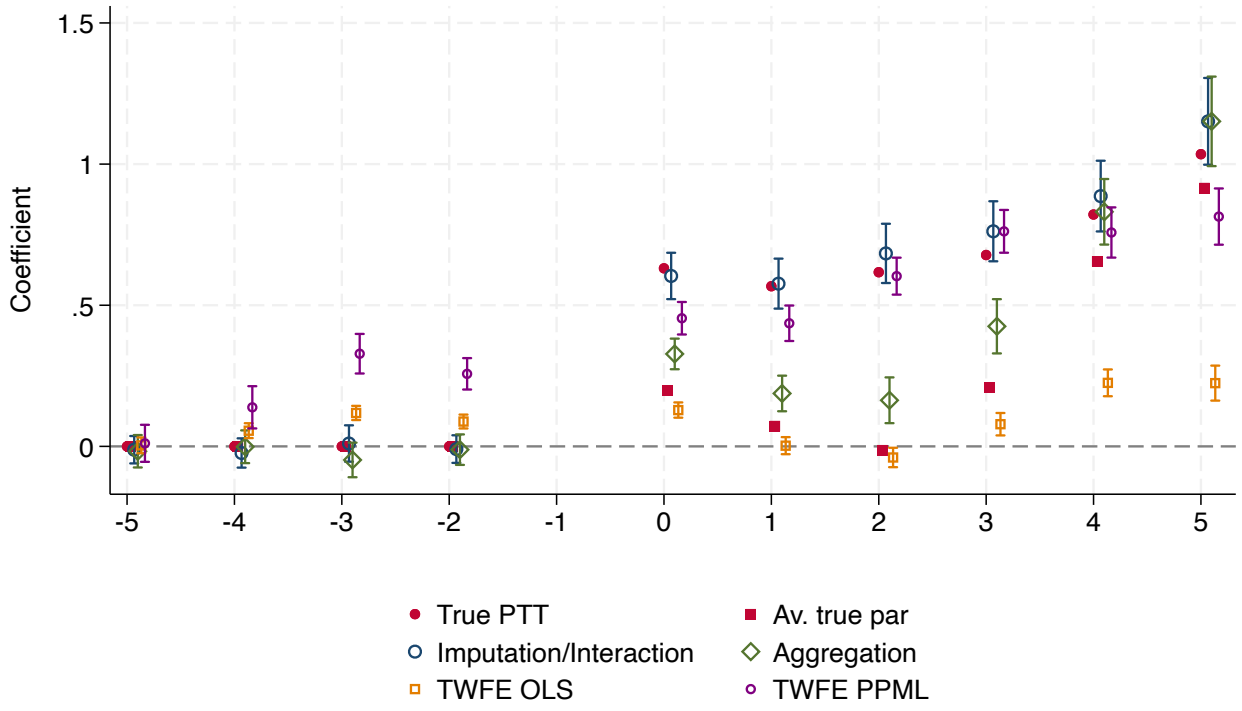
of δ_l for TWFE estimators, with l the relative time. Red markers are set for the true value of the average parameter and the PTT_l.

As expected, the imputation estimator is an unbiased estimator of time relative percentage changes in the average. It closely matches the true PTT_l. The TWFE PPML estimator is biased downward for first treatment periods. TWFE log-OLS cannot recover the true parameter for later time periods. But both TWFE PPML and TWFE log-OLS display false positive coefficients on pre-trend. The results of Sun and Abraham (2021) that staggered bias contaminates TWFE lead coefficients seem to hold for TWFE PPML. TWFE estimators point to non-existent pre-trends with staggered treatment, by displaying false positive coefficients.

The aggregation parameter recovers an intermediate quantity between the average parameter and the true treatment semi-elasticity (RoR). It converges to the true RoR in the later period, when there are fewer treated cohorts: in $t = 5$, when only the first cohort is treated, it computes the same quantity as the imputation estimator, because it covers only one (g, t) cell now. The imputation and aggregation estimators display close to zero and non statistically significant coefficients for leads.

Figures C2, C3 and C4 in the appendix display event studies for unique simulations generated in the other cases. In cases 1 and 2, when there is no heterogeneity in treatment effect within cohort-time cells (in simulations, driven by individual treatment heterogeneity), the aggregation estimator identifies the average true parameter. In case 3 it computes an intermediate quantity between the average true parameter. TWFE log-OLS always fails to identify the average true parameter, because of either the heteroskedasticity bias or the staggered treatment timing bias. Estimator from Borusyak et al. (2024) corrects for this bias and fails with treatment incuded heteroskedasticity. TWFE PPML displays large false-positive pre-trends and diverges from the true growth rate. Only the imputation estimator identifies the true growth rate before and after the policy change.

Figure 1 – Case 4: Event study



Note: 95% confidence intervals. Case 4: Heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. To ease the derivation of confidence intervals, I plot $log(\theta_l + 1)$ for each estimator.

5 Applications

5.1 Treaties of exchange of information

Set-up I apply my estimator to a major Public Economics research question: does the exchange of information between countries reduce households' cross-border tax evasion? Menkhoff and Miethe (2019) and Johannesen and Zucman (2014). Following the G20 2009 summit, many tax havens were compelled to sign bilateral treaties implementing exchange of information on request regarding bank account holders. These treaties, signed for example between France and Switzerland in 2009, make it mandatory for banks in both countries to report accounts held by each other's citizens to the tax authorities of their home countries, if the latter demand it. The signature and implementation of treaties vary across country pairs. Using data from the Bank of International Settlements (BIS) from 2003 to 2011 Menkhoff and Miethe (2019) and Johannesen and Zucman (2014) explore whether a treaty signed between

a tax haven and another country reduces deposits held by citizens of the home country in the tax haven. This is likely to be the case if those deposits are held for tax or regulation evasion purposes.

I replicate the findings of Menkhoff and Miethe (2019), as their replication package is publicly available (Johannesen and Zucman (2014) use a confidential version of the BIS data, including more Tax Havens). The two papers use the same identification strategy.[16] The authors estimate the following model:

$$\log(Deposit_{ijq}) = \alpha + \beta Signed_{ijq} + \gamma_{ij} + \theta_q + \epsilon_{ijq} \tag{5.1}$$

With $Deposit_{ijq}$ the deposits held by citizens of country i in tax haven j at time q. The treatment variable $Signed_{ijq}$ takes the value one when a treaty is signed between i and j at time q. Fixed effects for country pairs γ_{ij} and time θ_q are included. The authors use a two-way fixed effect log-linearized model, using as a control group all non-haven-to-haven dyads which did not sign a treaty during the time frame under study. The authors seek to estimate β which they interpret as the causal effect of treaties on deposits held in tax havens in percentage.

Results I replicate the strategy under equation (5.1) with five different estimators: the TWFE log-OLS estimator, the linear estimator from Borusyak et al. (2024), the TWFE PPML estimator, the proposed imputation estimator, and with a PPML aggregation strategy.

Results are displayed in Table 5. Column (1) presents the replication of Menkhoff and Miethe (2019) results using their methodology. On average, the signature of a treaty reduced deposits held in the partner tax havens by 31.9%.[17] In column (2), I restrict the sample and remove the few country-pairs that are always treated to avoid forbidden comparisons to this group. The results remain. In column (3), I use the estimator from Borusyak et al. (2024)

[16] Menkhoff and Miethe (2019) use a more conservative treatment, building on a few more years of perspective on these instruments: they only consider new TIEAs and DTCs implementing the OECD's banking transparency standards.

[17] $(exp(-0.384) - 1) \times 100 \approx -31.9$

to recover the log-linear DiD. The effect is slightly bigger than before, indicating that the staggered treatment biases the TWFE treatment effect estimate upward.

Columns (4), (5), and (6) display the non-linear estimations. In column (5), the TWFE PPML estimates that treaties signed decreased the average deposits held in tax havens by 13.2%.[18] Column (6) implements my proposed estimator robust to staggered bias: it recovers a drop in deposits by 16.5%. The comparison of columns (5)-(6) points to an upward bias because staggered treatment, as in columns (2)-(3) for the log-linear DiD.

There is a large difference between the results derived from the log-linear difference-in-differences (OLS) and the ratio-of-ratios (PPML). The difference between columns (1) to (3) and (5)-(6) comes from the different causal interpretations of the estimates: the approximate average effect over country-pairs and time (log-linear DiD) and the proportional change in the average (RoR). The average effect of treaties across country pairs and time is larger that the effect of the set of treaties on average deposits held in tax havens.

The joint distribution of treatment effects and deposit volumes across treatment cells illustrates the treatment effect heterogeneity causing this difference. In Figure D1 in the appendix, I plot a cohort-time specific coefficient from the full interaction model (such as in equation 3.7). Each coefficient recovers the RoR on the average of cell (ij, q). Cohorts (country-pairs ij treated at the same time) are displayed in the same color. We observe that even though cells (ij, q) display a large negative treatment effect *on average*, most of the cells exhibiting the strongest effects are small country-pairs in term of volume of tax haven deposits held. On the contrary, there are some cohorts exhibiting at the same time a weak or positive treatment effect, and a large volume of tax haven deposits, explaining the lower ratio-of-ratios, or change *in the average*.

The result of column (4) goes further in reconciling both results by showing that the aggregation estimator lies between them. It displays the estimate from an aggregation estimator

[18] $(exp(-0.141) - 1) \times 100 \approx -13.2$

used in Nagengast and Yotov (2025). As explained in section 3.2.1 the targeted quantity is the average $PTT_{g,q}$ over treated (g, q) cells. I estimate that on average, when a group of countries signs treaties with some tax havens on the same month, their deposits held in these tax havens drop by 23.9% (average change of the averages).[19] This interpretation depends on the group of country pairs treated together, which is not very informative in this case. I verify that when $corr(\beta_{ij,q}, y_{ij,q}) > 0$, the aggregation PPML provides a higher estimate than the log-linear DiD, and smaller than the multiplicative DiD.

Table 5 – Replication - Exchange of Information

	Linear estimators			Non-linear estimators		
	TWFE log-OLS (replication)	TWFE log-OLS	Borusyak et al. (2024)	Aggregation	TWFE PPML	Imputation
	(1)	(2)	(3)	(4)	(5)	(6)
Coef	-0.384***	-0.383***	-0.402***	-0.273**	-0.141**	-0.180**
S.e.	(0.09)	(0.09)	(0.074)	(0.11)	(0.078)	(0.091)
N	17267	16244	16244	16244	16244	16244
Control group	All	Never treated & Not yet treated				
Country-pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Column (1): Replication of Menkhoff and Miethe (2019). Standard errors adjusted for clustering by country-pairs. Standard errors for the imputation and aggregation estimators are computed through 500 bootstrap replications. No control variables included.

5.2 Gravity: the effect of RTAs

Set-up The second application revisits an important question in International Trade: how does bilateral trade respond to regional trade agreements (RTA)? Nagengast and Yotov (2025) seek to recover estimates robust to the two-way fixed effect negative weights problem, which could have plagued estimates from the previous literature.

The RTA estimates are usually based on the structural gravity theoretical framework, according to which bilateral trade in value between exporter i and importer j at time t is ruled

[19] $(exp(-0.273) - 1) \times 100 \approx -23.9$

by the following relationship:

$$X_{ijt} = \frac{Y_{it}E_{jt}}{Y_t}\left(\frac{t_{ijt}}{\Pi_{it}P_{jt}}\right)^{1-\sigma}$$

With Y_{it} and E_{jt} the output and the expenditure of respectively origin and destination, Y_t total world production, t_{ijt} bilateral time varying trade costs, and Π_{it} and P_{jt} the multilateral resistance terms. The latter are theoretically grounded terms that solve the system:

$$\Pi_{it}^{1-\sigma} = \sum_j \left(\frac{t_{ijt}}{P_{jt}}\right)^{1-\sigma}\frac{E_{jt}}{Y_t}$$

$$P_{jt}^{1-\sigma} = \sum_i \left(\frac{t_{ijt}}{\Pi_{it}}\right)^{1-\sigma}\frac{Y_{it}}{Y_t}$$

In this framework, the signature of an RTA between countries i and j will change the bilateral trade cost t_{ijt} at signature time t. The response of bilateral trade will depend on the content of the RTA and the trade elasticity.

The state-of-the-art specification derived from this framework is:

$$y_{ij,t} = \exp\left\{\delta RTA_{ij,t} + \pi_{i,t} + \chi_{j,t} + \tau_{ij} + \theta_{ii,t}\right\} \times \epsilon_{ij,t}. \tag{5.2}$$

With $y_{ij,t}$ bilateral trade, $RTA_{ij,t}$ a binary variable when an RTA is active between i and j and time t, $\pi_{i,t}$, $\chi_{j,t}$, τ_{ij} and $\theta_{ii,t}$ exporter-time, importer-time, dyad and border-time fixed effects. The exporter-time and importer-time fixed effects will control for multilateral resistance terms and the size of each economy. Bilateral fixed effects control for the time invariant part of bilateral trade cost, influenced by factors such as distance, language, historical ties, ... Border time fixed effects account for the international border effect when comparing internal to international trade flows.

In the trade literature, the model is usually estimated on a matrix of internal and international trade flows observed every year, with a PPML estimator, while clustering standard errors at the country-pair level. In this setting, we have (i) staggered RTAs implementation (ii) different RTAs negotiated, with potentially different content and effects on bilateral trade

(iii) an interest in expressing treatment as a semi-elasticity.

Nagengast and Yotov (2025) adapt the specification to account for potential bias caused by the staggered treatment timing. They remove the always treated country pairs from the control group, and follow Wooldridge (2023), by using a fully interacted specification for heterogeneous treatment effects:

$$y_{ij,t} = \exp\left\{\sum_{g=q}^{T}\sum_{s=g}^{T}\delta_{gs}D_{gs}RTA_{ij(g),t(s)} + \pi_{i,t} + \chi_{j,t} + \tau_{ij} + \theta_{ii,t}\right\} \times \epsilon_{ij,t}. \tag{5.3}$$

With g a group of countries pairs signing an RTA at the same time, s the relative time to treatment, D_{gs} a binary variable taking the value one for a cohort g treated s years ago, and δ_{gs} the cohort-time specific coefficient capturing treatment effect. To recover an aggregate treatment effect, the authors use an aggregation strategy:

$$\hat{\bar{\delta}} = \sum_{g=q}^{T}\sum_{s=g}^{T}\frac{N_{gs}}{N_D}\hat{\delta}_{gs},$$

With N_{gs} the number of treated observations from cohort g and time s and N_D the size of the treated sample.

Results I replicate the results of Nagengast and Yotov (2025), Table 1 columns (1) and (2), in columns (1) and (2) of Table 6. In column (3), I use my imputation estimator. In columns (4) and (5), I use linear estimators for the log-linear specification version of models 5.2 and 5.3: the fixed effects log-OLS and the interacted specification from Wooldridge (2021), which is robust to staggered treatment timing. The samples are smaller for these last columns because of zero observations being dropped.

Comparing columns (1) and (2), we observe that the difference between the estimated semi-elasticities is large: 0.18 from TWFE PPML against 0.463 for the corrected estimate from Nagengast and Yotov (2025), using an aggregation strategy.[20] The authors interpret this gap

[20] $exp(0.166) - 1 \simeq 0.18$ and $exp(0.381) - 1 \simeq 0.463$.

as caused by the negative weight issue of TWFE PPML. In column (3), my estimator provides a semi-elasticity of 0.434. This elasticity is close to the one provided by the aggregation strategy.

The close value provided by those two quantities can be rationalized by comparing estimates with the log-linear estimators. The TWFE log-OLS estimates a semi-elasticity of 0.19, close to TWFE PPML. Comparing the two quantities, the small initial difference indicates that the two estimation targets are close to each other. Comparing columns (3) and (5) and estimators robust to the bias caused by staggered treatment, we find again very close estimates. Taken together, this indicates that the average treatment semi-elasticity (over country pairs) and the semi-elasticity of the average bilateral trade flows are close quantities in this setting. [21] The aggregation strategy and the imputation estimator then target quantities that are close to each other, explaining the comparable estimates.

Table 6 – Replication - Gravity

	Non-linear estimators (replications)			Linear estimators	
	TWFE PPML	Aggregation	Imputation	TWFE log-OLS	Wooldridge (2021)
	(1)	(2)	(3)	(4)	(5)
Coef	0.166***	0.381***	0.361***	0.172***	0.36***
S.e.	(0.05)	(0.07)	(0.11)	(0.04)	(0.06)
N	105409	105409	105409	104802	104802
Control group	Never treated & Not yet treated				
Fixed effects	Exporter-Year, Importer-Year, Dyad, Border-Year				

Note: Column (1): Standard errors adjusted for clustering by country-pairs. Standard errors for the imputation, aggregation and Wooldridge (2021) estimators are computed through 500 bootstrap replications.

[21] Of course, part of the difference between the estimators can be due to either the missing zeros or the heteroskedasticity bias (Silva and Tenreyro, 2006).

6 Conclusion

This paper reconciles two significant empirical issues encountered by applied economists when estimating treatment effects in non-linear models, using difference-in-differences methodologies. First, the log-OLS estimator is biased in the presence of heteroskedasticity and treatment-induced changes in outcome variance. Even in the absence of bias, the researcher could prefer to use PPML because of the underlying parallel trend or the quantity targeted in the presence of treatment effect heterogeneity. Second, the traditional two-way fixed effects estimators do not accurately recover difference-in-differences estimates when treatment timing is staggered and the effect is heterogeneous. I show that this issue extends to two-way fixed effects PPML.

To reconcile both issues, I propose a novel estimator that recovers of a proportional treatment effect (semi-elasticity) even in cases of staggered treatment timing and heterogeneous treatment effects. Leveraging the interpretation of the TWFE PPML estimator in the canonical 2x2 setting, I develop an approach that accurately estimates the ratio-of-ratios, ensuring an interpretable treatment effect estimates similar to the canonical setting. The specified model can account for any kind of heterogeneity in the treatment effect, under the parallel trend and no anticipation assumption. Moreover, it can account for a parallel trend assumption conditional on some covariates, or the relative parallel trend of the triple difference setting.

Through empirical validation and simulations, I compare the proposed estimator to existing approaches. From simulations with staggered treatment timing and heterogeneous treatment effects, it appears that the interaction estimator proposed in this paper is the most suited to recover the correct treatment change in the average. In all studied cases, its density is centered around the true ratio-of-ratios, the quantity of interest of the 2-by-2 canonical setting. It is also robust to any type of treatment effect heterogeneity. Using an aggrega-

tion strategy for PPML, similar to the logic of solutions proposed for the linear case seems more suited to recover the average parameter when the treatment effect is homogeneous within cohort-time cells. This quantity is the same target as for the log-linear difference-in-difference. This aggregation strategy performs better than available log-linear estimators when there is treatment-induced heteroskedasticity. With more general patterns of treatment effect heteroskedasticity, this strategy will recover an intermediate quantity that might not have a clear interpretation. It is up to the empirical researcher to think about what is her preferred quantity to recover. The use of TWFE log-OLS and TWFE PPML is strongly discouraged in this setting, with the former potentially yielding negative estimates when the true treatment effect is positive and of a large magnitude.

I apply my estimator to two major empirical questions of the Public Economics and International Trade literatures. First, I study the setting of Johannesen and Zucman (2014) and Menkhoff and Miethe (2019). The authors investigate whether bilateral treaties of exchange of information decreased deposits held in tax havens' banks. I show that their results are close to the one estimated while correcting for staggered treatment and using a non-linear estimator. I show that using a TWFE PPML estimator in their set-up provides a smaller treatment effect, which is rationalized by the fact that it aggregates differently the strong heterogeneity in treatment effects across country pairs. My proposed estimator recovers a closer estimate to TWFE PPML than an aggregation strategy, showing that even though the treaties on average decreased deposits held offshore in tax havens, the average volume of deposits held in tax havens changed by a lower magnitude. Furthermore, by applying the proposed estimator to the empirical question of cross-border deposit behavior in response to exchange of information treaties, I showcase its practical relevance to answer empirical questions. Second, I revisit the RTA effect on bilateral trade in the setting of Nagengast and Yotov (2025). I show that in this setting, my estimator confirms the large bias due to staggered treatment

timing that the authors find, using an aggregation strategy. Moreover, this setting illustrates a case in which the average percentage change and the change of the average are close.

References

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of business & economic statistics*, 19(1):2–28.

Atal, J. P., Cuesta, J. I., González, F., and Otero, C. (2024). The economics of the public option: Evidence from local pharmaceutical markets. *American Economic Review*, 114(3):615–644.

Azoulay, P., Fons-Rosen, C., and Zivin, J. S. G. (2019). Does science advance one funeral at a time? *American Economic Review*, 109(8):2889–2920.

Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Bau, N. and Matray, A. (2023). Misallocation and capital market integration: Evidence from india. *Econometrica*, 91(1):67–106.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*.

Britto, D. G., Pinotti, P., and Sampaio, B. (2022). The effect of job loss and unemployment insurance on crime in brazil. *Econometrica*, 90(4):1393–1423.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chen, J. and Roth, J. (2023). Logs with Zeros? Some Problems and Solutions*. *The Quarterly Journal of Economics*, 139(2):891–936.

Ciani, E. and Fisher, P. (2019). Dif-in-dif estimators of multiplicative treatment effects. *Journal of Econometric Methods*, 8(1):20160011.

Cohn, J., Liu, Z., and Wardlaw, M. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*.

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast poisson estimation with high-dimensional fixed effects. *The Stata Journal: Promoting communications on statistics and Stata*, 20(1):95–115.

Cullen, Z. B. and Pakzad-Hurson, B. (2023). Equilibrium effects of pay transparency. *Econometrica*, 91(3):765–802.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

De Chaisemartin, C. and D'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econometrics Journal*.

Fetzer, T., Souza, P. C., Vanden Eynde, O., and Wright, A. L. (2021). Security transitions. *American Economic Review*, 111(7):2275–2308.

Giroud, X., Lenzu, S., Maingi, Q., and Mueller, H. (2024). Propagation and amplification of local productivity spillovers. *Econometrica*, 92(5):1589–1619.

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–1079.

Jan van Garderen, K. and Shah, C. (2002). Exact interpretation of dummy variables in semilogarithmic equations. *The Econometrics Journal*, 5(1):149–159.

Johannesen, N. and Zucman, G. (2014). The end of bank secrecy? an evaluation of the g20 tax haven crackdown. *American Economic Journal: Economic Policy*, 6(1):65–91.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semi logarithmic equations. *American Economic Review*.

Liu, L., Wang, Y., and Xu, Y. (2024). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 68(1):160–176.

Martinez, I. Z., Saez, E., and Siegenthaler, M. (2021). Intertemporal labor supply substitution? evidence from the swiss income tax holidays. *American Economic Review*, 111(2):506–546.

Menkhoff, L. and Miethe, J. (2019). Tax evasion in new disguise? examining tax havens' international bank deposits. *Journal of Public Economics*, 176:53–78.

Mirenda, L., Mocetti, S., and Rizzica, L. (2022). The economic effects of mafia: Firm level evidence. *American Economic Review*, 112(8):2748–2773.

Miyauchi, Y. (2024). Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade. *Econometrica*, 92(6):1869–1905.

Nagengast, A. and Yotov, Y. (2025). Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements. *American Economic Journal: Applied Economics*.

Olden, A. and Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*, 25(3):531–553.

Roth, J. (2024). Interpreting event-studies from recent difference-in-differences methods. *arXiv preprint arXiv:2401.12309*.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*.

Appendix

A. Literature review

Table A1 – Some top 5 papers with difference-in-differences strategy, a staggered treatment and a proportional treatment effect since 2018

Paper	Analysis	Outcome	Cells	Estimator
Azoulay et al. (2019)	Table 3	Number of publications, grant awards	Subfield × Year	TWFE PPML
Hjort and Poulsen (2019)	Tables 2, 3, 7, 8	Internet speed (asinh); Hour-sworked (asinh); Net firm entry (asinh); Employees (asinh); Value-dded (asinh)	Grid-cell × Year	TWFE OLS
Bailey et al. (2021)	Figue 6, Table 4	Income (log)	County × Year	TWFE OLS
Fetzer et al. (2021)	Table 1	Incidents (log+1)	District× Time	TWFE OLS
Martinez et al. (2021)	Figure 7.B, 10	Wage earnings of employees, Self-employment income per person	Canton × Time	TWFE OLS (scaled post estimation)
Mirenda et al. (2022)	Table 3, 7	Revenues (log); Tangible and intangible assets (log); Wage bill (log)	Firm × Time	TWFE OLS
Atal et al. (2024)	Figure 4, Table 2	Drug prices and sales (log)	Market × Time	TWFE OLS
Miyauchi (2024)	Table III	Sales Growth (Arc-Elasticity)	Firm × Time	Stacked DiD OLS

Cont'd on following page

Table A1, cont'd

Paper	Analysis	Outcome	Cells	Estimator
Bau and Matray (2023)	Table III, IV, V, Figure 2, 3, 4	Foreign debt, Foreign Spending, Revenues, Capital, Wages, MRPK	Firm × Year	TWFE log-OLS
Giroud et al. (2024)	Table II, III, IV, V, VI, VII, IX, Figure 1	TFP, Employment, Wages, Number of plants, Number of patents+1	County × Year	Three-way FE log-OLS
Britto et al. (2022)	Table I, III, IV, Figure 3	Employment, Income, Crime	Municipality-industry × Year	TWFE OLS, rescaled
Cullen and Pakzad-Hurson (2023)	Figure 3, 4, 5, Table I, II,	Employment, wages	State × Year	TWFE log-OLS, Staggered robust log-OLS

B. Supplementary results

B1 TWFE PPML Maximum likelihood

Simple case We observe two individuals A and B, at time periods $t = 1, 2, 3$. Treatment follows the pattern displayed by Table B1.

Table B1 – Staggered setting - Simple example

$\mathbb{E}[Y_{it}]$	$i = A$	$i = B$
$t = 1$	$exp(\alpha_A)$	$exp(\alpha_B)$
$t = 2$	$exp(\alpha_A + \beta_2 + \delta_{A2})$	$exp(\alpha_B + \beta_2)$
$t = 3$	$exp(\alpha_A + \beta_3 + \delta_{A3})$	$exp(\alpha_B + \beta_3 + \delta_{B3})$

The TWFE PPML estimation by maximum of log-likelihood implies the following first order conditions:

$$\begin{cases} \sum_{i,t,D_{it}=1}(y_{it} - \hat{y_{it}}) = 0 \\ \sum_{i=j,t}(y_{jt} - \hat{y_{jt}}) = 0 \\ \sum_{i,t=l}(y_{il} - \hat{y_{il}}) = 0 \end{cases}$$

Which translates in this system of equation in the simple case in table B1:

$$\begin{cases} y_{A2} + y_{A3} + y_{B3} = exp(\hat{\alpha_A} + \hat{\beta}_2 + \hat{\delta}) + exp(\hat{\alpha_A} + \hat{\beta}_3 + \hat{\delta}) + exp(\hat{\alpha_B} + \hat{\beta}_3 + \hat{\delta}) \\ y_{A1} + y_{A2} + y_{A3} = exp(\hat{\alpha_A}) + exp(\hat{\alpha_A} + \hat{\beta}_2 + \hat{\delta}) + exp(\hat{\alpha_A} + \hat{\beta}_3 + \hat{\delta}) \\ y_{B1} + y_{B2} + y_{B3} = exp(\hat{\alpha_B}) + exp(\hat{\alpha_B} + \hat{\beta}_2) + exp(\hat{\alpha_B} + \hat{\beta}_3 + \hat{\delta}) \\ y_{A1} + y_{B1} = exp(\hat{\alpha_A}) + exp(\hat{\alpha_B}) \\ y_{A2} + y_{B2} = exp(\hat{\alpha_A} + \hat{\beta}_2 + \hat{\delta}) + exp(\hat{\alpha_B} + \hat{\beta}_2) \\ y_{A3} + y_{B3} = exp(\hat{\alpha_A} + \hat{\beta}_3 + \hat{\delta}) + exp(\hat{\alpha_B} + \hat{\beta}_3 + \hat{\delta}) \end{cases}$$

This yields:

$$\begin{cases} exp(\hat{\beta}_2 + \hat{\delta}) = \frac{Y_{A2}}{exp(\hat{\alpha_A})} \\ exp(\hat{\alpha_A}) = \frac{(y_{A1}+y_{A3})\times(y_{A1}+y_{B1})}{y_{A1}+y_{B1}+y_{A3}+y_{B3}} \\ exp(\hat{\alpha_B}) = \frac{(y_{B1}+y_{B3})\times(y_{A1}+y_{B1})}{y_{A1}+y_{B1}+y_{A3}+y_{B3}} \\ y_{A1} + y_{B1} = exp(\hat{\alpha_A}) + exp(\hat{\alpha_B}) \\ y_{A2} + y_{B2} = exp(\hat{\alpha_A} + \hat{\beta}_2 + \hat{\delta}) + exp(\hat{\alpha_B} + \hat{\beta}_2) \\ exp(\hat{\beta}_3) = \frac{y_{A3}+y_{B3}}{(y_{A1}+y_{B1})exp(\hat{\delta})} \end{cases}$$

Quantity of interest If we have, heterogeneous treatment effect $\delta_{A2} \neq \delta_{A3} \neq \delta_{B3}$. The quantity of interest is then:

$$\begin{aligned} PTT &= \frac{E[y_{it}(1)|D=1] - E[y_{it}(0)|D=1]}{E[y_{it}(0)|D=1]} \\ &= \frac{(1/3)\Big(E(y_{A2}(1)) + E(y_{A3}(1)) + E(y_{B3}(1))\Big) - (1/3)\Big(E(y_{A2}(0)) + E(y_{A3}(0)) + E(y_{B3}(0))\Big)}{(1/3)\Big(E(y_{A2}(0)) + E(y_{A3}(0)) + E(y_{B3}(0))\Big)} \\ &= \frac{\Big(exp(\alpha_A + \beta_2 + \delta_{A2}) + exp(\alpha_A + \beta_3 + \delta_{A3}) + exp(\alpha_B + \beta_3 + \delta_{B3})\Big)}{\Big(exp(\alpha_A + \beta_2) + exp(\alpha_A + \beta_3) + exp(\alpha_B + \beta_3)\Big)} \\ &\quad - \frac{\Big(exp(\alpha_A + \beta_2) + exp(\alpha_A + \beta_3) + exp(\alpha_B + \beta_3)\Big)}{\Big(exp(\alpha_A + \beta_2) + exp(\alpha_A + \beta_3) + exp(\alpha_B + \beta_3)\Big)} \\ &= \sum_{i,t,D_{it}=1} \frac{E(y_{it}(0))}{\sum_{i,t,D_{it}=1} E(y_{it}(0))}\Big(exp(\delta_{it}) - 1\Big) = \sum_{i,t,D_{it}=1} \omega_{it}\Big(exp(\delta_{it}) - 1\Big) \end{aligned} \tag{.1}$$

General case In the general case there are N individuals, G cohorts and T time periods.

We estimate the parameters α, β and δ of the model with the correctly specified conditional mean, using TWFE PPML:

$$E[y_{igt}|D_{igt}] = exp(\alpha_i + \beta_t + \delta D_{igt})$$

The log-likelihod function is:

$$\mathcal{L}(\alpha, \beta, \delta) = \sum_{i}^{N} \sum_{t}^{T} y_{igt}(\alpha_i + \beta_t + \delta D_{igt}) - \exp(\alpha_i + \beta_t + \delta D_{it}) \tag{.2}$$

Which yelds the following first order conditions:

$$\begin{cases} \frac{\partial \mathcal{L}(\alpha,\beta,\delta)}{\partial \delta} = 0 \Leftrightarrow \sum_i^N \sum_t^T D_{igt} y_{igt} - \exp(\alpha_i + \beta t + \delta D_{igt}) = 0 \\ \frac{\partial \mathcal{L}(\alpha,\beta,\delta)}{\partial \alpha_j} = 0 \Leftrightarrow \sum_t^T \left(y_{jgt} - \exp(\alpha_j + \beta_t + \delta D_{jgt})\right) = 0 \\ \frac{\partial \mathcal{L}(\alpha,\beta,\delta)}{\partial \beta_l} = 0 \Leftrightarrow \sum_i^N \left(y_{igl} - \exp(\alpha_i + \beta_l + \delta D_{igl})\right) = 0 \end{cases}$$

From the first order condition on individual fixed effects, we obtain:

$$\exp(\alpha_j) = \frac{\sum_t y_{jgt}}{\sum_t \exp(\beta_t + \delta D_{jgt})} \tag{.3}$$

Individual fiexed effects can be replaced in the F.O.C., but finding a closed-form solution for δ implies solving a nonlinear system of β and δ. There is usually no closed form solution for this type of system (statistical softwares use iterative least squares methods to find estimates of δ), preventing us from deriving a more general proof on the bias of the TWFE PPML estimator.

B2 Equivalence of the imputation and saturated approaches

Wooldridge (2023) proposes to recover estimates of ATTs in level which converges to $\tau_{rs} = y_{rs}(1) - y_{rs}(0)$ by predicting the average partial effect of the treatment variable D_{it} over the desired treated sample, evaluated for the right value of cohort and time dummies. For time

period and cohorts r, s, it computes:

$$\begin{aligned}\widehat{\tau}_{inter,rs} =& E(\widehat{y}|D_{it}=1, g_{is}=1, f_{s_t}=1, \forall(k,l)\neq(r,s)\, g_{ik}=0; f_{l_t}=0) \\ & - E(\widehat{y}|D_{it}=0, g_{is}=1, f_{s_t}=1, \forall(k,l)\neq(r,s)\, g_{ik}=0; f_{l_t}=0) \\ =& N_{rs}^{-1}\sum_{i=1}^{N} D_{irs}[exp(\widehat{\alpha_i}+\widehat{\beta t}+\widehat{\delta_{rs}}) - exp(\widehat{\alpha_i}+\widehat{\beta t})]\end{aligned}$$

With N_{rs} the number of observations for cohort r at time s and D_{irs} an indicator variable if the observation belongs to cohort r observed at time s. Again, I can re-write the model to compute the average partial effect across the entire treated sample with $\mathbb{1}\{t-r=l\}$ to get the ATT l time periods after treatment:

$$\widehat{\tau}_{inter,rl} = N_{rl}^{-1}\sum_{i=1}^{N} D_{irl}[exp(\widehat{\alpha_i}+\widehat{\beta t}+\widehat{\delta_{rl}}) - exp(\widehat{\alpha_i}+\widehat{\beta t})] \tag{.4}$$

Interestingly, Wooldridge (2023) notes that this quantity is numerically equivalent to the imputation estimator from equation (3.13) on the same sample. It also has the advantage to have known analytical expressions for standard errors. As in the previously section, I propose to scale this quantity by the predicted counterfactual outcome in the absence of treatment on the same subsample:

$$\widehat{RoR}_{inter} = \frac{\widehat{\tau}_{inter}}{\sum_{i\in\omega_1}\widehat{y_{igt}(0)}} = \frac{\widehat{\tau}_{imput}}{\sum_{i\in\omega_1}\widehat{y_{igt}(0)}} = \widehat{RoR}_{imput} \tag{.5}$$

In the case with group specific parallel trends, the fully saturated model should write:

$$\begin{aligned}E[y_{ipt}|g_{iq},...,g_{iT}] = exp\Big[&\sum_{g=q}^{T}\beta_g g_{ig} + \sum_{s=2}^{T}\gamma_s fs_t + \sum_{p=1}^{P}\mathbb{1}\{c=p\}\kappa_p + \sum_{p=1}^{P}\sum_{s=2}^{T}(fs_t\times\mathbb{1}\{c=p\})\pi_{pt} \\ &+ \sum_{g=q}^{T}\sum_{l=0}^{T-r}(D_{it}\times g_{ig}\times\mathbb{1}\{t-g=l\})\delta_{gl} \\ &+ \sum_{p=1}^{P}\sum_{g=q}^{T}\sum_{l=0}^{T-r}\mathbb{1}\{c=p\}\times(D_{it}\times g_{ig}\times\mathbb{1}\{t-g=l\})\zeta_{pgl} \\ &+ \alpha_i + \beta_{ct}\Big]\end{aligned}$$

Coefficients γ_s and π_{pt} control for the group-specific parallel trend. Coefficients δ_{gs} and ζ_{pgs} control for the full heterogeneity of the treatment effect, by cohort, group and time. Variables $g_{ig}, fs_t, \mathbb{1}\{c=s\}, \mathbb{1}\{c=s\} \times fs_t$ are dropped because they are colinear with fixed effects α_i and β_{ct}, and we are left with the model to estimate:

$$\begin{aligned} E[Y_{ipt}|g_{iq},...,g_{iT}] = exp\Big[&\sum_{g=q}^{T}\sum_{l=0}^{T-r}(D_{it} \times g_{ig} \times \mathbb{1}\{t-g=l\})\delta_{gl} \\ &+ \sum_{p=1}^{P}\sum_{g=q}^{T}\sum_{l=0}^{T-r}\mathbb{1}\{c=p\} \times (D_{it} \times g_{ig} \times \mathbb{1}\{t-g=l\})\zeta_{pgl} + \alpha_i + \beta_{ct}\Big] \end{aligned} \tag{.6}$$

The average treatment effect ATT_{inter} is estimated as in the simple case, by predicting the treatement average partial effect on the treated sample. As before, the proportional treatment effect estimate is:

$$\widehat{RoR}_{inter} = \frac{\widehat{\tau}_{inter}}{\sum_{i\in\omega_1}\widehat{y_{igt}(0)}}$$

The mutiplicative parallel trend often holds conditionally on a set of control variables (assumption A2.B). Wooldridge (2023) explains how to include time-constant controls in the fully saturated model and keep the equivalence with the imputation estimator:

$$\begin{aligned} E(y_{ipt}|g_{it},...,g_{iT}) = exp\Big[&\sum_{s=2}^{T}(fs_tX_i)\pi_{Ct} + \sum_{g=q}^{T}\sum_{l=0}^{T-r}(D_{it} \times g_{ig} \times \mathbb{1}\{t-g=l\})\delta_{gl} \\ &+ \sum_{p=1}^{P}\sum_{g=q}^{T}\sum_{l=0}^{T-r}(\mathbb{1}\{c=s\} \times D_{it} \times g_{ig} \times \mathbb{1}\{t-g=l\})\zeta_{pgl} \\ &+ \sum_{p=1}^{P}\sum_{g=q}^{T}\sum_{l=0}^{T-r}(\dot{X_{ig}} \times D_{it} \times g_{ig} \times \mathbb{1}\{t-g=l\})\xi_{gl} + \alpha_{ij} + \beta_{jt}\Big] \end{aligned}$$

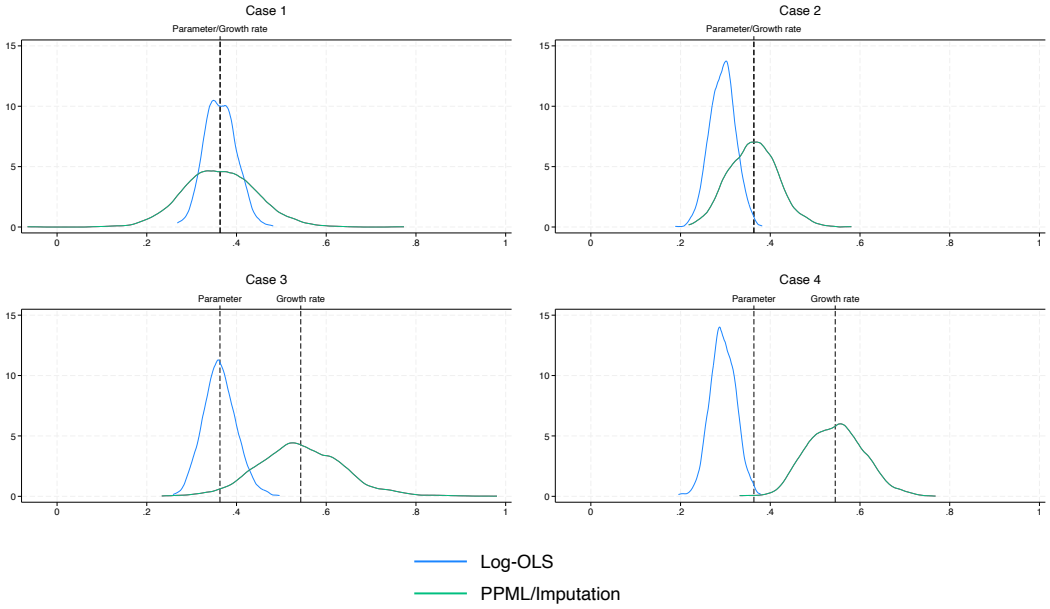
Coefficients π_{Ct} capture the divergence from the parallel trend due to control variable X_{ij}. Coefficients ξ_{gl} the divergence from the average treatment effect in cohort g at time l due to variables X_{ij}. Control variables are centered on the treated sample: $\dot{X} = X - E(X|D=1)$. This normalization ensures that δ_r has the desired interpretation $log(E[y_r(1)|g=1]) - log(E[y_r(0)|g=1])$ among the treated. The ATT is recovered as before by predicting average partial effects of treatment on the desired sample. The idea behind this model is to allow full heterogeneity in treatment effect across the level of control variables X_i in the sample.

Limitation to time constant covariates can be quite restrictive, especially if researchers wish to control for within cohorts time varying shocks that could counfound the treatment effect estimation. In this case the imputation approach becomes more tractable and general.

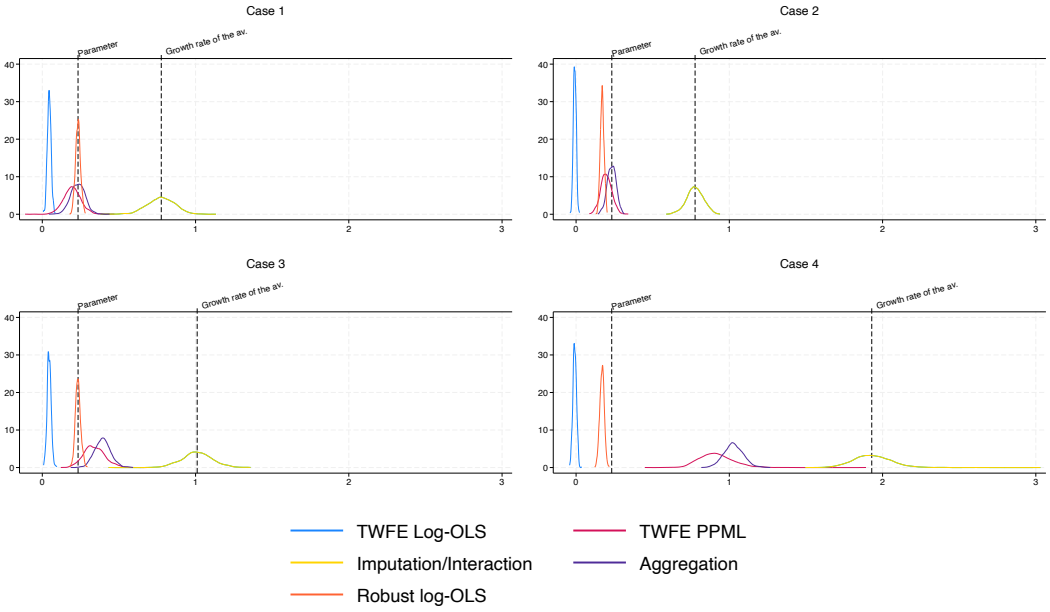
C. Simulations

Figure C1 – Simulations: density

(a) Common treatment timing

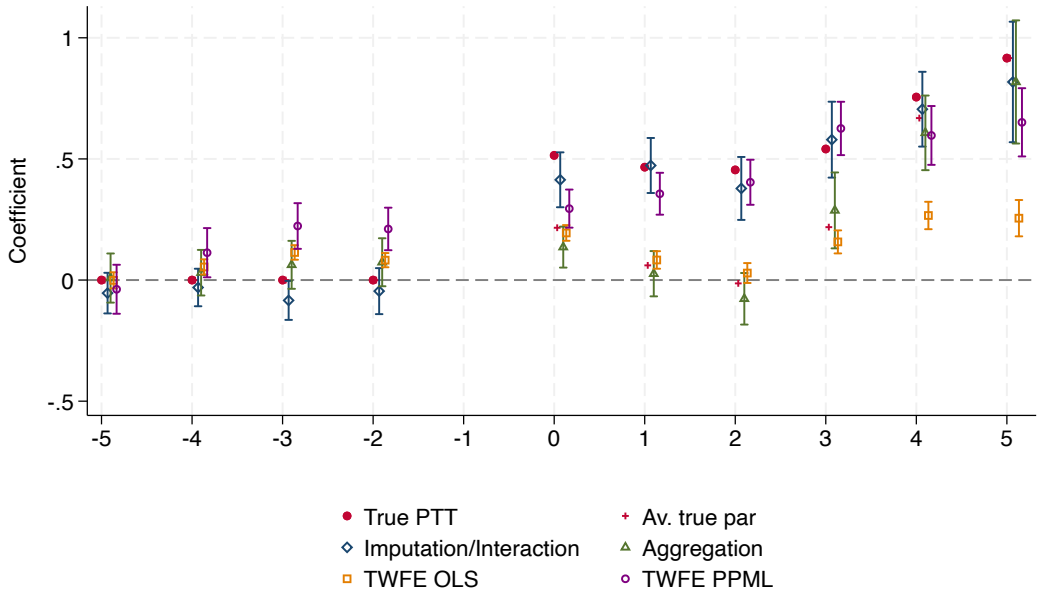


(b) Staggered treatment timing



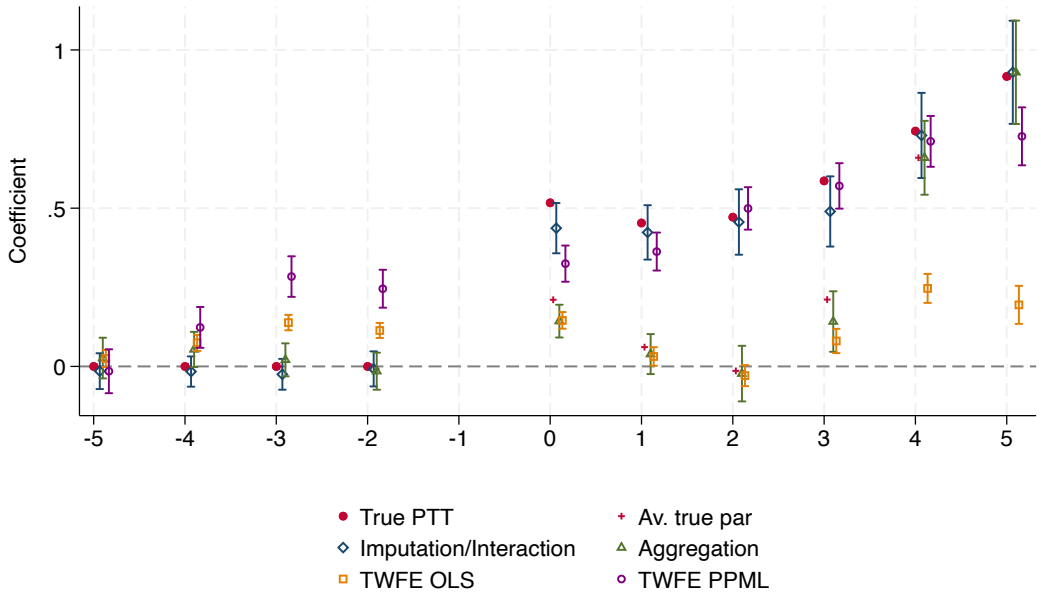
Note: Case 1: No heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. Case 2: Heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. Case 3: No heteroskedasticity function of treatment status, individual treatment effect heterogeneity. Case 4: Heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. Robust log-OLS from Borusyak et al. (2024).

Figure C2 – Case 1: Event study



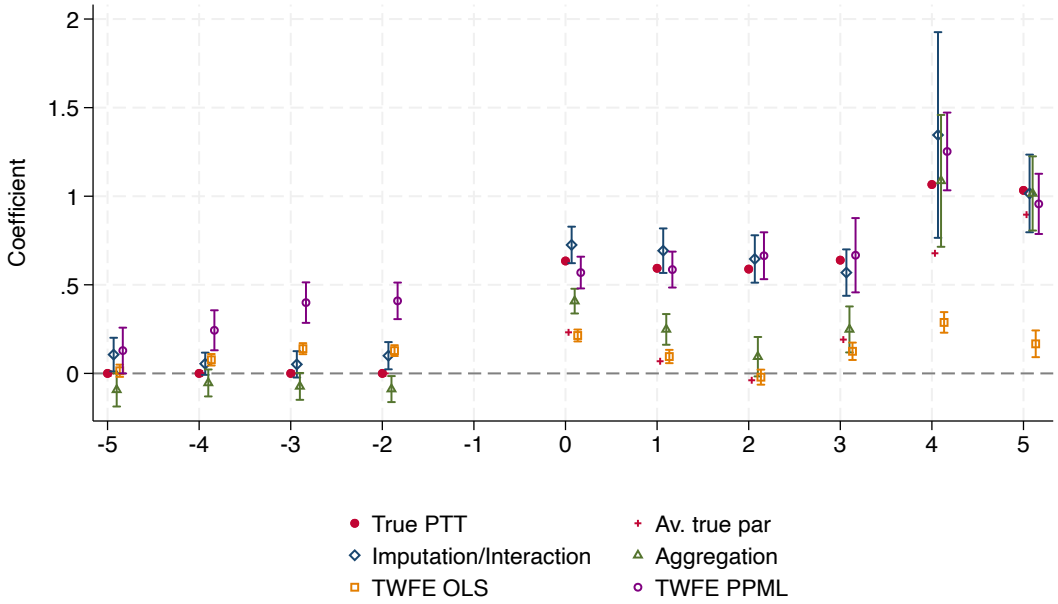
Note: 95% confidence intervals. Case 1: No heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. To ease the derivation of confidence intervals, I plot $log(PTT + 1)$ for each estimator.

Figure C3 – Case 2: Event study



Note: 95% confidence intervals. Case 2: Heteroskedasticity function of treatment status, no individual treatment effect heterogeneity. To ease the derivation of confidence intervals, I plot $log(PTT + 1)$ for each estimator.

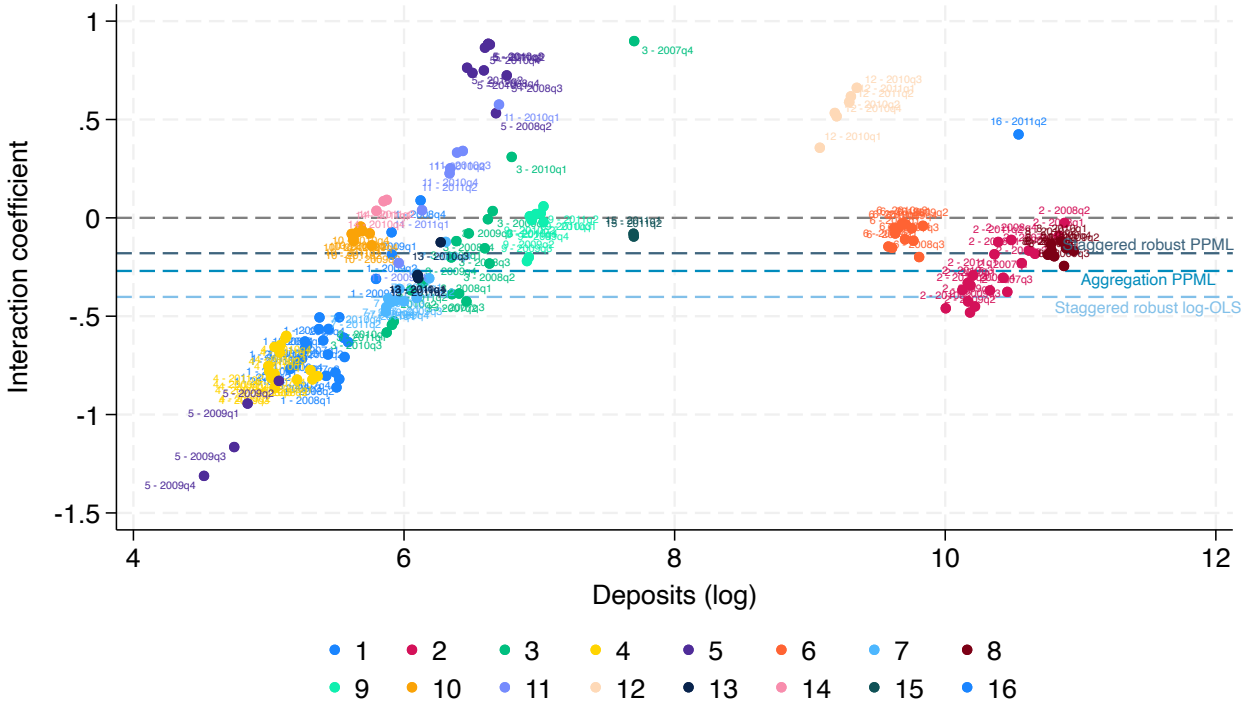
Figure C4 – Case 3: Event study



Note: 95% confidence intervals. Case 3: No heteroskedasticity function of treatment status, individual treatment effect heterogeneity. To ease the derivation of confidence intervals, I plot $log(PTT + 1)$ for each estimator.

A D. Application

Figure D1 – Interaction coefficients (Menkhoff and Miethe, 2019)



Note: Colors in the legend correspond to the different treated cohorts. Each dot correspond to a coefficient of the interaction from the aggregation estimator.