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

Ninon Moreau-Kastler†

August 24, 2025

Abstract

I propose a counterfactual approach to estimate 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

Economists are often interested in studying variables which take only non-negative values and are non-normally distributed. Such outcomes can include 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 seek to estimate proportional treatment effects, or semi-elasticities: the percent change in the outcome generated by the treatment. In this case, 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 (Santos 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. The 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 average treatment effect over the treated (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. 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 main targeted quantity is not a proportional treatment effect and the paper does not discuss how to recover it at a higher level than at the cohort-time cell. He further provides evidence that

[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.

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): in prior work, estimates are computed in staggered treatment timing settings. They use Wooldridge (2023)'s fully saturated specification to estimate the change in trade caused by RTAs. The authors aggregate cohort-time proportional treatments effects by computing the average of cohort-year coefficients, weighting them by the share of treated observations in cohort-time cells. In this paper, I show that if treatment effect are small, and treatment heterogeneity occurs only across cohorts and time, this quantity has 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). In a more general case, the estimated quantity is closer to a weighted average of the total percent change within cohorts-time cells, and intermediate quantity between log-OLS and PPML. Compared to their strategy, my estimator is more suitable for cases where: (1) one estimates a proportional effect on the mean (quantity of interest of TWFE PPML), and TWFE PPML is different from TWFE log-OLS at baseline (2) one seek to test the extent to which TWFE PPML is biased (3) one suspect that the treatment effect is heterogeneous within cohort-year groups. Finally, I replicate Nagengast and Yotov (2025) estimates on trade with my estimator. In their context, the percent change of the average and the average percent change are close quantities, and my estimator confirm their results.[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

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

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 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 present an estimator recovering the ratio-of-ratios, analogous to the canonical setting. The estimator recovers a proportional treatment effect, and relies on an imputation approach to recover the ATT (the numerator), and to estimate its scaling quantity (the denominator). It can apply to a wide set of cases, such as conditional parallel trends, triple differences or time varying controls.

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 pre-trend coefficients in event studies. I show that my estimator 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 important empirical questions from the Economics literature. On top of replicating Nagengast and Yotov (2025) as described above, 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 wondered whether treaties facilitating international exchange of information between tax authorities decrease tax evasion and cross-border deposits owned in tax havens. The set-up of these studies motivates the use of a nonlinear estimator and the estimation of a proportional treatment effect. Bank deposits take positive values only, and country-pairs display very different foreigned owned deposits at baseline. Treaties are passed at different times is staggered and likely to be heterogeneous by time and country-pairs, providing the ideal setting.

The authors initially use a log-linear difference-in-difference strategy. Using Borusyak et al. (2024) estimator, I find that the author's estimate has a positive staggered treatment bias, i.e. slightly under estimate the effect of treaties. 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 decrease deposits held in tax havens *on average*, their effect is larger in country-pairs with small tax havens deposits at baseline. The change effect on the total treated tax haven deposits is weaker than the average change across tax havens. I show that in this case, compared to TWFE PPML, my proposed estimator confirms that staggered treatment biasunderestimates the true treatment effect. I show that using a aggregation strategy for PPML would strongly overestimate this bias.

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; Santos 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, and I propose a flexbile estimator that can be used in many settings. I contribute to the literature on the interpretation of models estimating semi-elasticities (Kennedy, 1981; van Garderen and Shah, 2002) and on the estimation of aggregate parameters using PPML (Breinlich et al., 2024; Tyazhelnikov and Zhou, 2021). 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 interpretations. 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, under the same conditions as for the linear case. I propose a new non-linear estimator robust to staggered treatment, and suitable for many settings. This last contribution extends to the litterature on counterfactual estimators (Gobillon and Magnac, 2016; Borusyak et al., 2024; Liu et al., 2024; Gardner, 2022).

Finally, I also contribute to the literature on the determinants of foreign-owned bank deposits in tax havens (Johannesen and Zucman, 2014; Menkhoff and Miethe, 2019; Andersen et al., 2022; Langenmayr and Zyska, 2023), by showing that small cross-border deposits in tax havens are more sensitive than large ones to changes in financial transparency.

The rest of the paper proceeds as follows. Section 2 presents the 2x2 canonical setting of multiplicative differences-in-differences. 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):

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

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.

[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.

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 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):

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

The expression of the ATT follows:

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

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

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

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

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

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

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

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:

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

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:

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

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:

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

The parallel trend assumption:

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

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 of multiplicative DiD 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 DiD, 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 Santos 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:

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

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]

[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.

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 $y_{igt} = \mathbb{1}\{g \leq t\}y_{igt}(1) + (1 - \mathbb{1}\{g \leq t\})y_{igt}(0)$.

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).

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

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 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 derivation 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:

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

[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.

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.

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 Issues of 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) use the fully interacted model from Wooldridge (2023) and combined with an aggregation strategy similar to the linear case. 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 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 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 the 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

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

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 and robust to any type of treatment heterogeneity in a staggered treatment setting. It is an imputation estimator based on the fact that one can specify the correct counterfactual model. Wooldridge (2023) shows that this approach is equivalent to the fully interacted model above, but the imputation approach simplifies computations and allows to integrate time-varying covariates.[12]

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 the parameters:

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

From Wooldridge (2023), 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. Contrary to coefficients $\widehat{\delta_{g,t}}$, this is a linear effect that can be averaged without loss of interpretability. The average treatment effect in

[12]Derivation for the equivalent interaction estimator are discussed in Online Appendix.

level on the treated sample is estimated by the difference between the observed outcome and the predicted one on the treated sample:

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

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, and yields the following estimator:

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

It 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. The ratio of the two should converge in probability to the true PTT when N grows, provided (i) that the denominator does not reach zero and (ii) the expectation of the counterfactual outcome is different from zero, i.e. the PTT is defined. Condition (i) is unlikely to take place: PPML only predicts strictly positive values. For (ii), if the researcher has reasons to believe that the treatment affects mainly the extensive margin, it means that PPML is not a relevant model to start with, and should rather use one for binary outcomes. 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).[13]

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.[14]

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:

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

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

[13]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.

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

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':

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

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 $E[y_{ict}|D_{it}] = exp(\alpha_i + \beta_{ct})$ on the treated sample for pre-treatment periods and the never-treated to get $\hat{\alpha}_i$ and $\hat{\beta_{ct}}$:

$$\widehat{RoR}_{imput} = \frac{\sum_{i,t} D_{igt} y_{igct}(1)}{\sum_{i,t} D_{igt} \widehat{y_{igct}(0)}} - 1 = \frac{\sum_{i,t} D_{igt} y_{igct}(1)}{\sum_{i,t} D_{igt} exp(\widehat{\alpha_i} + \widehat{\beta_{ct}})} - 1 \tag{3.19}$$

Control variables The imputation process allows for a flexible counterfactual model, as long as it is correctly specified.[15] The researcher assumes the true conditional mean to be:

$$E[y_{igt}|D_{igt}] = exp(\alpha_i + \beta_t + \delta_{it} D_{igt} + X'_{igt}\gamma) \tag{3.20}$$

[15]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:

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

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 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:

$$E[y_{ijpgt}|D_{igt}] = exp(\alpha_i + \beta_{jt} + \beta_{pt} + \delta_{it}D_{igt}) \tag{3.21}$$

The new parallel trend assumption becomes:

$$E[y_{ijpgt}(0)|D_{igt}] = exp(\alpha_i + \beta_{jt} + \beta_{pt}) = exp(\alpha_i) \times \underbrace{exp(\beta_{jt}) \times exp(\beta_{pt})}_{Relative\ growth\ rate} \tag{3.22}$$

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

4.1 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 $E[\eta_{igt}|D_{igt}] = 1$. 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 cases 1 and 2, heterogeneity depends on the time t. In cases 3 and 4, individual heterogeneity that is normally distributed across individuals, on top of time heterogeneity. In those later cases, within each each cell (g, t), the growth rate of the average is different from the average growth rate due to treatment. In cases 2 and 4, heteroskedasticity of the error term is correlated with treatment status, such that log-OLS is biased. The observed outcome y_{igt} is strictly positive (no difference driven by zeros). I simulate the data 1000 times.

4.2 Simulation results

Results of simulations are displayed in table 1. The first line displays the exponential of the average parameter δ_{it} minus one, the estimation target of the log-linear DiD. The second line displays the growth rate of the average treated outcome (PTT). 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.

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. 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 estimator is now taking negative values for a large number of simulations, and the mean is negative close to zero, even though the true value of the parameter and the growth rate of the average are large (0.233 and 0.776). 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 estimators is an unbiased estimator 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 1 – Simulations

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	$0.2D_{igt}$	$\delta_{it} = log(\|t-12.5\|) + \nu_i, \quad \nu_i \sim \mathcal{N}(0, 0.5)$

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

	Case 1 $exp(V(\eta_{it}\|.)) - 1 = \alpha_i$				Case 2 $exp(V(\eta_{it}\|.)) - 1 = 0.2D_i$			
	Mean	St.D.	Min	Max	Mean	St.D.	Min	Max
Qty of interest								
$exp(\overline{\delta_t}) - 1$	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
Estimator								
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

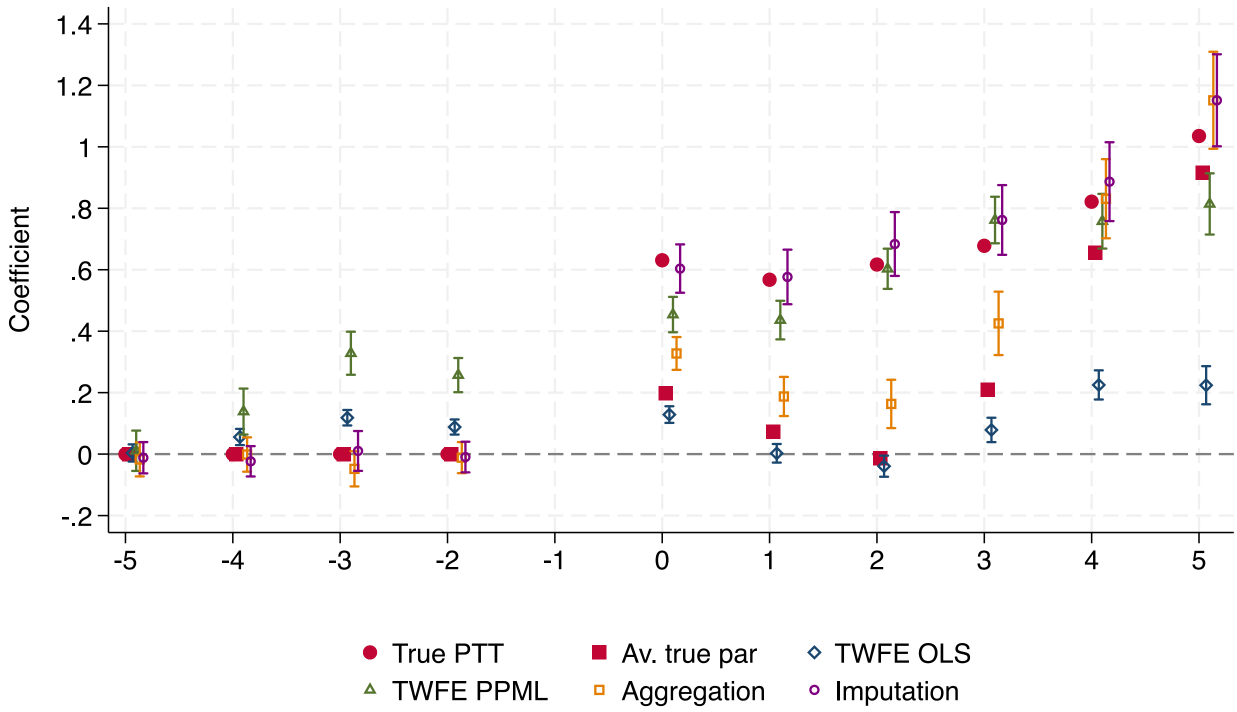
Heterogeneous treatment effect by time and individuals $exp(\delta_{it}) - 1$

	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
Qty of interest								
$exp(\overline{\delta_{it}}) - 1$	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
Estimator								
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 1). 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).[16] 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 of δ_l for TWFE estimators, with l the relative time. Red solid markers are set for the true value of the average parameter and the PTT_l.

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

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.

The imputation estimator is an unbiased estimator of time relative percentage changes in the average. It closely matches true PTT_l coefficients. 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.

5 Applications

5.1 Treaties of exchange of information

Set-up I apply my estimator to an recent and important question in Public Finance: does the exchange of information between countries reduce households' cross-border tax evasion? 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, first Johannesen and Zucman (2014) and Menkhoff and Miethe (2019) then 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.[17] The authors estimate the following model:

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

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.

[17] 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.

Table 2 – 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				
Fixed effects	Country pairs, Time					

Note: 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.

Results are presented in Table 2. 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%.[18] 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) 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%.[19] 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.

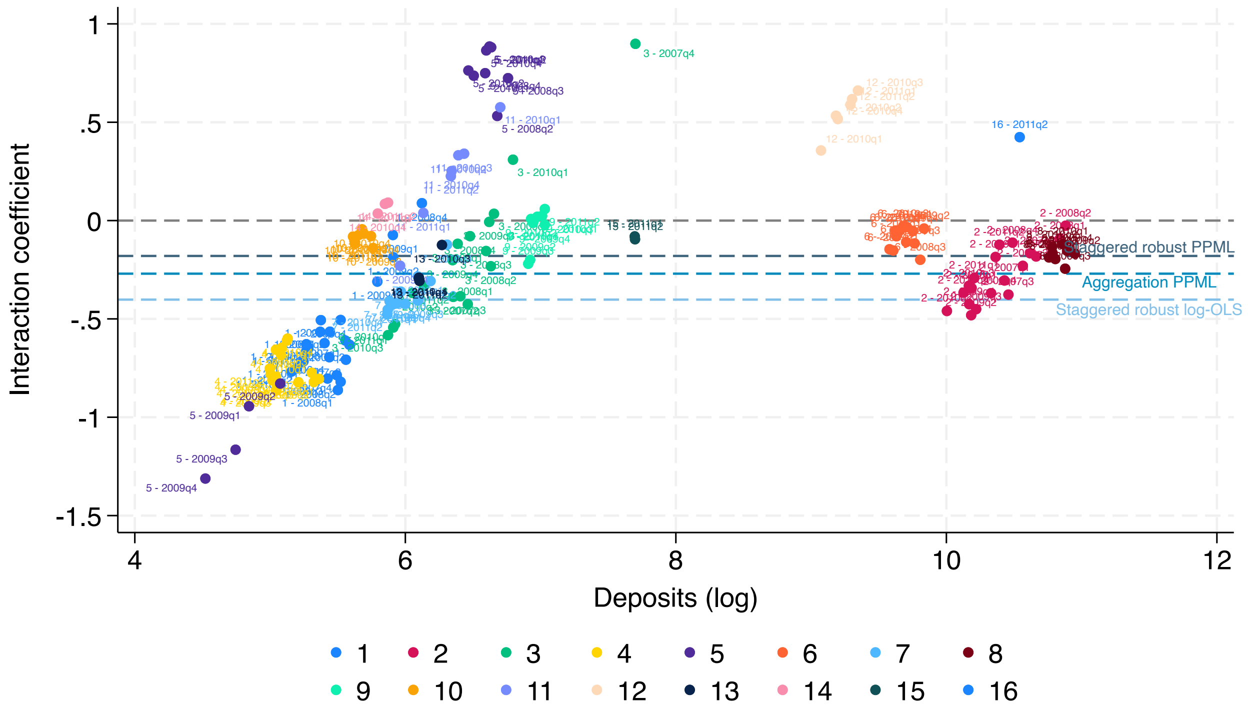
There is a large difference between the results derived from the log-linear difference-in-differences (OLS) and the ratio-of-ratios (PPML). The difference between columns (1) to (3) and (5)-(6) comes from the different causal interpretations of the estimates: the approximate average effect over country-pairs 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 2, 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

[18] $(exp(-0.384) - 1) \times 100 \approx -31.9$
[19] $(exp(-0.141) - 1) \times 100 \approx -13.2$

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*.

Figure 2 – Treaties of exchange of information: Interaction coefficients



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

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 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).[20] 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.

5.2 Gravity: the effect of RTAs

I revisit the results of Nagengast and Yotov (2025), who seek to recover unbiased estimates of the effect of RTAs on international trade. Up to their paper, the litterature has been estimating semi-elasticity of RTAs with potentially heterogeneous effects and staggered treatment timing, using PPML estimators. The estimation strategy and results are presented in Appendix and Table 3. I replicate their main results, finding that the standard TWFE PPML yields a semi-elasticity of 18%, while the corrected approach gives 46%. My imputation estimator produces a close estimate (43%), close to that obtained from the aggregation strategy. The staggered robust log-linear estimate is close as well (43.3%), while TWFE PPML and log-OLS present close estimates (18 and 19%). Overall, these results indicate that in this setting, the percent change of the average and the average percentage change are close. The targeted quantities my estimator and of Nagengast and Yotov (2025) are close, which explains the small difference between estimates.

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

6 Conclusion

This paper reconciles significant empirical issues encountered by applied economists when estimating treatment effects in non-linear models, using difference-in-differences methodologies. Researcher have been temped to use PPML estimators to account for multiplicative parallel trends, include zero observations or prevent estimates from heteroskedasticity bias. Yet, 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.

References

Andersen, J. J., Johannesen, N., and Rijkers, B. (2022). Elite capture of foreign aid: Evidence from offshore bank accounts. *Journal of Political Economy*, 130(2):388–425.

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.

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

Breinlich, H., Novy, D., and Santos Silva, J. (2024). Trade, gravity, and aggregation. *Review of Economics and Statistics*, 106(5):1418–1426.

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.

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*.

Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

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*.

Langenmayr, D. and Zyska, L. (2023). Escaping the exchange of information: Tax evasion via citizenship-by-investment. *Journal of Public Economics*, 221:104865.

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.

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

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*.

Santos Silva, J. 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.

Tyazhelnikov, V. and Zhou, X. (2021). Ppml, gravity, and heterogeneous trade elasticities. *Unpublished Working Paper*.

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

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 Supplementary results

A.1 TWFE PPML Maximum likelihood

Simple case We observe two individuals A and B, at time periods $t = 1, 2, 3$.

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

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

This yields:

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

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:

$$\mathcal{L}(\alpha,\beta,\delta)=\sum_i^N\sum_t^T y_{igt}(\alpha_i+\beta_t+\delta D_{igt})-\exp(\alpha_i+\beta_t+\delta D_{it}) \tag{A.1}$$

Which yelds the following first order conditions:

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

Finding a closed-form solution for δ implies solving a nonlinear system of αs, βs 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 an exact expression of the TWFE PPML estimator as a function of observables.

B Application

B.1 Gravity: the effect of RTAs

Set-up 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 determined 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 exporter and importer, Y_t total world production, t_{ijt} bilateral time varying trade costs, and Π_{it} and P_{jt} the multilateral resistance terms.[21] 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

[21]Solving for $\Pi_{it}^{1-\sigma}=\sum_j\left(\frac{t_{ijt}}{P_{jt}}\right)^{1-\sigma}\frac{E_{jt}}{Y_t}$ and $P_{jt}^{1-\sigma}=\sum_i\left(\frac{t_{ijt}}{\Pi_{it}}\right)^{1-\sigma}\frac{Y_{it}}{Y_t}$.

the content of the RTA and the trade elasticity.

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

$$y_{ij,t} = \exp\left\{\delta RTA_{ij,t} + \pi_{i,t} + \chi_{j,t} + \tau_{ij} + \theta_{ii,t}\right\} \times \epsilon_{ij,t}. \tag{B.1}$$

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. Border time fixed should be introduced when one includes internal trade flows.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.

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:

$$y_{ij,t} = \exp\left\{\sum_{g=q}^{T}\sum_{s=g}^{T} \delta_{gs} D_{gs} RTA_{ij(g),t(s)} + \pi_{i,t} + \chi_{j,t} + \tau_{ij} + \theta_{ii,t}\right\} \times \epsilon_{ij,t}. \tag{B.2}$$

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:

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

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 3. 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 B.1 and B.2: 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: 18% from TWFE PPML against 46.3% for the corrected estimate from Nagengast and Yotov (2025).[22] The authors interpret this gap 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 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. [23] The aggregation strategy and the imputation estimator then target quantities that are close to each other, explaining the comparable estimates.

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

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

Table 3 – 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.