How to account for (pre-)trends?

Camila Steffens[*] and Jan Stuhler[§]

October 31, 2025

Work in Progress

Click here for the latest version

Abstract

This paper assesses the performance of classical strategies to account for (pre-)trends in difference-in-differences designs. We focus on three main approaches: controlling for or matching on pre-trends, extrapolating linear trends, and controlling for group-by-time fixed effects. Through Monte Carlo simulations using real labor market data, we examine incidental trends that may emerge due to correlations between treatment and unobserved characteristics. Drawing on these simulations and supporting analytical results, we provide intuitive insights into the performance of these methods and further formalize the conditions that justify their application.

Keywords: Difference-in-Differences, Event Study, Pre-trends, Parallel Trends.

1 Introduction

An increasing number of studies rely on (generalized) difference-in-differences designs, often visualizing the effects of a shock over time using an *event study* plot (Currie et al., 2020).[1] In these settings, researchers estimate causal effects of a treatment by comparing the evolution of the outcomes in exposed units (e.g. regions or individuals, the treated group) with a comparison group that was not exposed to the treatment. The key underlying identification assumption is that the post-treatment outcomes of both groups would have followed parallel trajectories in the absence of treatment.

Because the parallel trends assumption is not directly testable, researchers resort to assessing pre-treatment trajectories – *pre-trends* – to provide indirect evidence in its support.[2] In some cases, evidence of non-parallel pre-trends may discourage researchers from pursuing certain projects, potentially leading to publication bias. But commonly, researchers attempt to control or adjust for differential trajectories that could otherwise bias the estimates. Although a variety of techniques are used, to our knowledge – and as also noted by Miller (2023, Appendix F) – there has been no systematic comparison of their effectiveness.

In this paper, we study the performance of the most popular approaches to account for (pre-)trends in difference-in-differences designs. We start by surveying recent event study papers published at *AER* or *AER-Insights*, showing that three methods accounting for differential pre-treatment trajectories are customary practices among applied researchers: (i) incorporating pre-trends as a covariate, (ii) extrapolating linear trends, and (iii) controlling for broader group (often, macro-region-by-time) fixed effects. About 81% of the surveyed papers employed at least one of these methods.

After formalizing the assumptions underlying these methods, we provide intuitive insights into their behavior by simulating *placebo treatments* in real labor market data. Similarly as Bertrand et al. (2004), we use real data to closely mimic practical research sce-

---

[1]"Over time, it has become rare to use difference-in-differences without showing an event study graph, and conversely it is rare to show event studies without a control group." (Currie et al., 2020, p.45)

[2]Roth (2022) shows that pre-testing may suffer from low power, consequently leading to under-rejection of the null hypothesis. Ghanem et al. (2022) further shows that the parallel trends assumption imposes strong restrictions on the time-series properties of the outcomes. These papers highlight the need to address possible deviations from this identifying assumption.

narios. Our simulations draw on a 5% sample of Brazilian matched employer–employee records (1995–2019), aggregated to 558 regional labor markets. Rather than imposing artificial trends, we simulate treatment allocations that correlate with baseline characteristics of the labor markets, resulting in *incidental* differences in pre-trends between treated and untreated regions. We then estimate "treatment effects" before and after the event across many specifications and simulations, and summarize each method's performance in terms of bias, standard deviation and root mean squared errors. Finally, we provide analytical results and auxiliary evidence to assess whether (and explain why) our empirical findings may generalize to other settings.

Our main finding is that incorporating pre-trends as a covariate ("trends-as-control") is unlikely to be a reliable strategy. While this approach eliminates bias in the pre-treatment period – creating a misleading impression that the method "works" – in simulations it yields post-treatment estimates that are nearly identical to the unconditional (and biased) baseline estimates. We show empirically and analytically that this occurs because outcome growth is (much) less serially correlated at the unit than at the treatment level. For example, (treated) labor markets with a greater share of skilled workers might experience stronger employment growth on average, but the specific markets experiencing high growth fluctuate significantly from one period to the next. Building on this observation, we propose a simple test that researchers can use to evaluate the viability of trend-as-control in their own setting.

Importantly, this problem generalizes to related empirical strategies, such as matching on pre-trends. In some applied literatures, it has become common to match control units based on pre-trends in the outcome of interest. One example is the literature on mass layoffs, where researchers compare displaced workers with a group of non-displaced workers who are often selected based on pre-treatment characteristics and outcomes. For instance, Fackler et al. (2019) includes earnings and days per year in employment for all pre-treatment years, while Schmieder et al. (2023) and Bertheau et al. (2023) select non-displaced workers based on wages or earnings in two pre-treatment years. Our findings suggest a cautionary note: if the outcome we match on is only weakly correlated over time, the selected control group may fail to provide an accurate counterfactual after treatment, even if they exhibit similar

pre-trends.[3]

The key insight is that pre-trends contain both valuable structural information (the "factor structure") and transitory noise. We show that when the variance of the idiosyncratic shock is large relative to the systematic component, controlling for pre-trends can lead researchers to significantly underestimate the persistence of common shocks. This concern relates to the risk of overfitting in the synthetic control method.[4] The synthetic control literature highlights the need for a sufficiently large number of pre-treatment periods relative to the scale of transitory shocks in order to accurately recover the unobservable factors (Abadie et al., 2010; Ferman and Pinto, 2021; Arkhangelsky and Hirshberg, 2023). Furthermore, trends tend to be more persistent at higher levels of aggregation, such as macro-regions, states, or treatment versus control groups, which supports the use of synthetic control in contexts involving aggregated time series data (Abadie and Vives-i Bastida, 2022). However, this issue has received little attention in the difference-in-differences literature, even though applications often focus on microdata settings, where time series are often more volatile, and the number of observations typically exceeds the number of periods.

An alternative strategy is to extrapolate linear trends, which reduces bias more effectively in our applications. However, this improvement comes at the cost of increased variance, partly due to the tendency of linear extrapolations to "overshoot": as differential trends often diminish over time due to regression to the mean, linear extrapolation tends to over-correct, leading to estimates biased in the opposite direction. We show that this over-correction becomes more pronounced when extrapolating linear trends from a short pre-treatment to a longer post-treatment period. The performance depends therefore strongly on the horizon over which linear trends are fitted. Another practical question is whether researchers should account for differential linear trends on the treatment level (i.e., a single trend) or if they should allow trends to differ for each unit at the level at which the treatment is assigned

---

[3]In our analysis, we focus on controlling for or matching on pre-treatment outcomes. However, the mentioned studies also control for other pre-determined characteristics, which may attenuate the concerns raised in our paper. Ideally, researchers should evaluate the sensitivity of their findings to the inclusion of additional covariates. Related, Kaul et al. (2022) highlights that when all pre-treatment outcomes are included in a synthetic control algorithm aimed at optimizing their fit, all other covariates become irrelevant.

[4]The underlying assumption behind synthetic control is that post-treatment potential outcomes are independent of the treatment "conditional on permanent unobserved characteristics and observed pre-treatment information" (Arkhangelsky and Hirshberg, 2023). Similarly, controlling for pre-trends within difference-in-differences designs relies on an equivalent assumption.

(e.g., for each region). We demonstrate that both strategies are equivalent in expectations and perform similarly in simulations, implying that researchers need not be concerned about the choice between treatment- and unit-specific trends.

Finally, controlling for group-by-time fixed effects leads to partial bias elimination in our simulations, offering a modest improvement over the unconditional (biased) baseline. This approach effectively mitigates differential trends driven by confounds at the aggregate group level, but is only feasible when there is variation in treatment allocation within groups. For instance, in our simulations, treatment is assigned at the local labor market level, allowing us to include state-by-year fixed effects. We analytically demonstrate that when group- and unit-level confounds move in the same direction, including group-by-time fixed effects can eliminate bias if the weighted average of within groups' trends is less pronounced than the overall (unconditional) differential trend. This is more likely when there is greater variation in treatment within the groups that exhibit weaker trends.

Our paper contributes to the growing literature on difference-in-differences and event study designs by offering intuitive insights into the performance of common practices to account for (pre-)trends. A significant focus of recent research has been on the limitations of the standard two-way fixed effects estimator in staggered treatment designs (Goodman-Bacon, 2021; Roth et al., 2023). Multiple studies have proposed new difference-in-differences estimators that are robust to heterogeneous effects over time and across cohorts with different treatment timings (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Wooldridge, 2021; Borusyak et al., 2024). However, comparatively less attention has been given to how practitioners should address (pre-)trends, despite its critical importance in applied research.

Recent efforts employ partial identification approaches to assess the sensitivity of estimates to potential violations of the parallel trends assumption, drawing on pre-treatment data and contextual insights (Freyaldenhoven et al., 2021; Rambachan and Roth, 2023; Ghanem et al., 2022). Instead of yielding a single point estimate, these methods construct confidence sets that contain the causal effect of interest under a specified set of assumptions. In a different approach, Freyaldenhoven et al. (2019) propose using a covariate as a proxy

for unobserved confounding. However, they show that including this covariate directly as a control is only valid if it is a perfect proxy; otherwise, they recommend an instrumental variables strategy. Similarly, Brown and Butts (2025) develop a GMM estimator to address heterogeneous exposure to unobserved common shocks, combining imputation using never-treated units with instruments to recover interactive fixed effects. The method remains consistent even in short panels. In contrast to these studies, rather than introducing new methods, we evaluate the performance of classical strategies commonly used in the literature. Despite their widespread application, there is limited understanding of their effectiveness and best practices.[5]

In the following section, we introduce our setting, detailing the underlying specifications and our approach. In Section 3 we describe how we simulate "fake" treatments in administrative data on employment and wages from Brazil. The main simulation results are discussed in Section 4, where Sections 4.1-4.3 provide additional details and analytical insights for each respective method. Section 5 concludes, summarizing the key takeaways. Throughout this paper, our aim is to present our main findings in an intuitive manner, with additional formalization and proofs available in Appendix A.

2 Setting

The setting in this paper is a difference-in-differences (DID) design where the researchers observe data for $i = 1, ..., N$ units during $\underline{T}$ periods prior to and $\overline{T}$ periods after some intervention. For simplicity, we consider a setting in which N_1 units are treated and N_0 units are untreated (such that $N = N_0 + N_1$), and treatment time is homogeneous between units. The treatment hits at $t = 0$, so $t = 1, ..., \overline{T}$ are post-treatment periods and $t = \underline{T}, ..., -1$ represent pre-treatment periods. For simplicity, we interchangeably refer to t as the year or

---

[5]There exist accessible papers addressing other relevant questions for practitioners, such as when to weight sample data (Solon et al., 2015) or adjust standard errors for clustering (Abadie et al., 2023). Additionally, there are guides for best practices in the estimation and visualization of event studies (Freyaldenhoven et al., 2021; Miller, 2023; Schmidheiny and Siegloch, 2023) or difference-in-differences designs in general (Baker et al., 2025), but we are not aware of any study that systematically compares classical strategies to control for trends.

the event time.[6]

In a standard event study, researchers often estimate dynamic effects over time with respect to a reference year (Miller, 2023; Schmidheiny and Siegloch, 2023; Baker et al., 2025). We normalize the effect to be zero in the year of treatment (i.e., $t = 0$), and estimate pre- and post-treatment coefficients according to the following specification:

$$y_{i,t} = \alpha_i + \tau_t + \sum_{e=\underline{T}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{\overline{T}} \beta_e \times D_i \times \mathbb{I}[e = t] + \epsilon_{i,t} \tag{S1}$$

where α_i are the unit fixed effects, τ_t are year fixed effects, and $D_i = 1$ for treated units, zero otherwise. Parameters β_t identify the difference in the (average) variation of the outcome from the event time $e = 0$ to $e = t > 0$ in treated units compared to untreated units. In what follows, we refer to equation (S1) as the *"baseline"* specification.

The causal interpretation of β_t requires that, in the absence of treatment, the evolution of the average outcome in the treated units would be the same as the untreated units (see Appendix A for a formal demonstration). In principle, the parallel trends assumption is required for the post-treatment period only (i.e., for $t > 0$). However, since counterfactual trends for the treated group are not observed after the treatment, researchers usually assess parallel trends in the pre-treatment period. Thus, in the absence of anticipation effects, testing the null hypotheses $H_0 : \gamma_t = 0$ for all $t < 0$ is equivalent to testing for pre-treatment parallel trends. A standard approach, then, is to use the pre-trends test to support the validity of the parallel trends assumption in the post-treatment period. Nevertheless, due to low-power, this test tends to under-reject the null hypothesis, and undetected trajectories can significantly affect post-treatment estimates (Roth, 2022).

We searched all the papers published at *AER* or *AER-Insights* in 2022 and 2023 that included the term *"event study"*, resulting in 31 studies using linear regressions to estimate dynamic treatment effects. Our examination reveals that researches usually rely on the lack of pre-trends, or statistically insignificant pre-trends, as supporting evidence for the parallel trends assumption. However, they also regularly extend the difference-in-differences

---

[6]In our setting, year and event time are collinear because the timing of treatment is uniform across all units. However, this is not the case in settings with staggered treatment adoption.

estimator to address non-parallel trends. The main methods used by the researchers are summarized in Table 1.

Table 1: Common Approaches to Address Pre-trends

	Unit Level	Covariates	Macro FEs	Linear Trends	Trend-as-control
Agha et al. (2022)	Drug class	✓	.	.	✓[a]
Allen et al. (2023)	Grid	✓	.	.	.
Ang (2023)	County	✓	State x Time	Unit-specific	.
Bacher-Hicks et al. (2022)	State	.	.	Unit-specific	.
Barwick et al. (2023)	Individual/Firm	✓	.	.	✓
Bertheau et al. (2023)	Individual	✓	.	.	✓[a]
Best et al. (2023)	Good	✓	Product x Region x Time	.	.
Black et al. (2023)	School and Cohort	✓	Region	.	.
Braghieri et al. (2022)	College	✓	.	Unit-specific	.
Brot-Goldberg et al. (2023)	Individual	.	Region/Experiment x Time	.	.
Butters et al. (2022)	Store	✓	Retail Chain x State	.	✓[b]
Cabral et al. (2022)	Industry- occupation	.	.	Unit-specific	.
Cantoni and Pons (2022)	Individual	.	Relative Election	.	.
Cao and Chen (2022)	County	✓	Province x Time	Prefecture	✓
Cicala (2022)	Power Control Areas	✓	Broader Time	.	.
Cullen and Perez-Truglia (2023)	Individual	✓	.	.	.
Currie et al. (2023)	County	✓	State x Time	.	.
Dahl et al. (2022)	County	✓	.	.	.
Dillender (2023)	City	✓	State x Time	Unit-specific	✓[a]
Duquennois (2022)	Individual-Question	✓	Question Group x School	.	.
East et al. (2023)	State	✓	Region x Time	Unit-specific	✓[b]
Esposito et al. (2023)	County	✓	State x Time	.	.
Greenstone et al. (2022)	City	✓	.	.	.
Gross et al. (2022)	Birthday group	.	.	.	.
Hansen and Wingender (2023)	Crop-country	✓	DDD	Trend-break	.
Janssen and Zhang (2023)	Pharmacy	.	Zip-code x Time	.	.
Porzio et al. (2022)	Country-Ind. Cohort	.	.	Cohort	.
Schmieder et al. (2023)	Individual	✓	.	.	✓[a]
Smith et al. (2022)	Firm	.	.	.	.
Sodini et al. (2023)	Household	.	DDD	.	.
Sviatschi (2022)	Municipality	✓	.	Unit-specific	✓[a]

Notes: We searched on Google Scholar for papers published in the *American Economic Review – AER* and *AER-Insights* – in 2022 and 2023 that were mentioning the term *"event study"*. Our search delivered 50 papers, from which 40 were empirical studies. We further removed nine papers: five macroeconomic or structural papers, two using instrumental variables, one regression discontinuity design, and a fully synthetic control study. This table shows our findings for the remaining 31 papers.

---

[a]Matching on pre-treatment outcome
[b]Synthetic control

2.1 Trend-as-control and Matching on Pre-trends

A commonly employed technique to address (pre-)trends is to control for, or condition on, differences in pre-treatment outcome growth (i.e., pre-trends). A straightforward way to implement this strategy is by directly incorporating outcome growth over k pre-treatment periods as a covariate $PreTrend_{i,k}$. The intuition is that, in the absence of treatment, the outcome would have followed the same trend among units with equal pre-treatment

trajectories.

Since this covariate is constant over time, in standard applications, it must be interacted with year indicators, which requires the omission of one term to avoid perfect multicollinearity. A particular feature of the trend-as-control approach is that if the interaction of the pre-trend covariate with the baseline year ($t = 0$) is omitted, then, due to the variable's construction, the interaction with $t = -k$ becomes redundant and should also be excluded.[7]

The linear specification is presented in equation (S2), hereafter referred to as the *"trend-as-control"* specification:

$$y_{i,t} = \alpha_i + \tau_t + \sum_{\substack{e=\underline{T} \\ e \neq -k}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{\overline{T}} \beta_e \times D_i \times \mathbb{I}[e = t] + \sum_{\substack{e=\underline{T} \\ e \neq 0 \\ e \neq -k}}^{\overline{T}} \phi_e \times PreTrend_{i,k} \times \mathbb{I}[e = t] + \epsilon_{i,t} \quad \text{(S2)}$$

where $PreTrend_{i,k} := y_{i,0} - y_{i,-k}$ is outcome growth over k pre-treatment years. For outcomes in log, it measures pre-trends in percentage change.

Specification (S2) assumes that the evolution of the outcomes for both treated and untreated units does not depend on the covariates. As such, it is not recommended when average treatment effects are expected to vary between covariate subgroups. Since practitioners typically estimate simplified specifications, such as in equation (S2), our main simulation results are based on this approach. However, as shown in Appendix Figure C.1, our conclusions also hold with more flexible methods.[8]

Remark 1. *The approach proposed by Callaway and Sant'Anna (2021) allows researchers to control for pre-trends in settings with staggered treatment adoption.*

The survey summarized in Table 1 suggests that researchers often employ the trend-

---

[7]Additionally, conditional on $PreTrend_{i,k}$, the expected value of $\hat{\gamma}_{-k}$ is equal to zero by construction. For this reason, we also recommend omitting the interaction of event time $t = -k$ with the treatment indicator when controlling for pre-treatment outcome growth of length k. Failing to omit these terms can lead to issues in the estimation of standard errors, depending on the program/software used.

[8]Wooldridge (2021) proposes a more flexible approach by incorporating interactions between the (recentered) covariates and the treatment indicator. However, in addition to assuming a linear functional form for the covariates, the interactions may become infeasible when the number of covariates is large. To overcome these limitations, Sant'Anna and Zhao (2020) introduces a doubly robust approach that yields consistent estimates as long as either the propensity score model or the outcome model is correctly specified, providing robustness to misspecification in one of them.

as-control approach by matching on pre-treatment outcomes. We highlight that, under common support, matching and regression are comparable methods, differing primarily in how covariate-specific effects are weighted to obtain average treatment effects (Angrist and Pischke, 2009). Therefore, our findings extend to matching on pre-trends, as we show in Section 4.1.2. We further analyze the relationship between post- and pre-treatment trends, showing that the effectiveness of trend-as-control in addressing bias depends on the persistence of outcome growth at the unit level.

2.2 Linear Trends

A popular alternative is to extrapolate pre-treatment trajectories into the post-treatment period. This approach is also considered by Roth (2022) and Wooldridge (2021) for addressing differential trends. As Roth (2022) explains, one might control for linear trends even when no strong pre-trend is evident. This is typically implemented by controlling for a variable capturing unitary changes in time ($Trend_t$) interacted with unit-specific indicators, thereby fitting linear trends slopes for each unit of analysis (Miller, 2023). We demonstrate that controlling for such "unit-specific linear trends" at the level of treatment assignment is equivalent, in expectations, to controlling for "treatment-specific linear trends".[9] In our main analysis, we focus on the latter, as it provides an intuitive interpretation in terms of differential trends between the treated and control units (Wooldridge, 2021).

In our review of the AER papers, we found that researchers often state concerns that the linear trends coefficient(s) could be contaminated with post-treatment trajectories. We highlight that this will not be the case as long as we thoughtfully omit the "right" event times from the regression, as also discussed by Miller (2023) and further illustrated in Appendix A.3. In summary, we understand as best practices the following:

- First, as in the baseline specification, researchers should include year fixed effects (τ_t) when controlling for linear trends to account for common shocks that affect both treated and control units;

---

[9]Goodman-Bacon (2021) also states the equivalence of group and unit-specific linear trends estimator in settings with heterogeneous treatment timing.

- Second, rather than using a single event-time indicator for several post-treatment years, researchers should include event-time indicators for each period individually;

- Third, due to multicollinearity of the linear trends interactions, practitioners must impose an alternative restriction. We recommend omitting an additional (pre-treatment) event time indicator. For instance, in the *"linear trends"* specification below, in addition to the baseline, we also omit the interaction with pre-event time $-k$:

$$y_{i,t} = \alpha_i + \tau_t + \sum_{\substack{e=\underline{T}\\e\neq -k}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e=t] + \sum_{e=1}^{\overline{T}} \beta_e \times D_i \times \mathbb{I}[e=t] + \delta_k \times D_i \times Trend_t + \epsilon_{i,t} \quad \text{(S3)}$$

Specification (S3) provides a nonparametric identification: δ_k is the slope of the differential trajectory of the treated group compared to the untreated group over k pre-treatment periods. As a result, the parameters β_t have causal interpretation if the deviation from the parallel trend is equivalent to a linear extrapolation of the pre-treatment trajectory. We formalize the identification of the parameters as well as the underlying parallel trend assumption in Appendix A.3.

Remark 2. *Wooldridge (2021) omits all the pre-treatment interactions. Instead, we opt to omit only one additional pre-treatment interaction ($t = -k$ in our specification), using the remaining pre-treatment coefficients as a diagnostic for the linearity assumption. If the pre-treatment coefficients γ_t for $-k < t < 0$ are jointly close to zero, there is supporting evidence that the differential trends are well approximated by a linear extrapolation in the pre-treatment period. Otherwise, researchers should consider alternative methods, such as higher-order polynomials (e.g., quadratic trends).*

As to which pre-treatment interaction one should omit, Miller (2023) recommends spacing the omitted coefficients further apart. The reason is that statistical noise tends to have a greater impact when the omitted coefficients are closer together.

A drawback of the linear trends strategy is that trends are unlikely to remain linear indefinitely; they may eventually diminish or even reverse, consistently with regression to the mean. We discuss the implications of the linear trends extrapolation in these contexts in Section 4.2.1.

2.3 Group-by-time Fixed Effects

A third popular strategy is to control for more flexible time fixed effects (FEs), allowing shocks to vary across broader groups by interacting them with group-level indicators. For instance, when studying labor markets, unit i may belong to a broader regional area, such as states. We can therefore control for *"state-by-year FEs"* as in the following specification:

$$y_{i,t} = \alpha_i + \tau_{st} + \sum_{e=\underline{T}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{\overline{T}} \beta_e \times D_i \times \mathbb{I}[e = t] + \epsilon_{i,t} \qquad \text{(S4)}$$

where τ_{st} are time fixed effects specific for each state s. Group-by-time FEs require the treatment allocation to happen at a nested level (i.e., treatment variation within the group). For instance, treatment at county level allows for state-by-year FEs, and individual level interventions allow for county or school-by-year FEs, and so on.

The widespread adoption of this approach (Table 1) may be attributed to its flexibility and simplicity. By controlling for group-by-time FEs, this method allows yearly shocks to vary across the groups (e.g., macro-regions or states) while still requiring the parallel trends assumption to hold within each group. As a result, this strategy effectively mitigates differential trends driven by group-level confounds. However, given the treatment variation within groups, unobserved factors at more granular levels, which are not addressed by the fixed effects, may persist. If the (unconditional) bias is predominantly driven by unit-level confounds, group-by-time FEs offer limited potential for bias reduction. In Appendix A.4, we analytically demonstrate that the effectiveness of group-by-time FEs depends on the extent to which the overall (unconditional) trend is more pronounced than the weighted average of within groups' trends, where the weights are proportional to the size of each group and the variation in treatment within them (Goodman-Bacon, 2021).

In this paper, we study the performance of the strategies commonly used by researches to control for pre-trends. We focus on the specifications discussed throughout this section. Next, we describe the dataset and the simulation procedure used to evaluate the behavior of these different methods. Our main simulation results are discussed in Section 4.

3 Simulations

In order to compare the performance of the different strategies, we use real regional level data on employment with simulated treatment assignments. We estimate the dynamic effects of a binary treatment using the specifications discussed in the previous section. Our main outcome $y_{i,t}$ is log of total employment in region i and event time t. We also provide simulation results for log of average monthly wage in Appendix D.

3.1 Data

Our data is based on a 5% sample of all workers in the Brazilian matched employer-employee administrative records between 1995 and 2019.[10] The sample contains 4,015,280 workers, with more than 28.7 million single-employment spells for individuals from 18 to 65 years old, representing 5% of yearly formal employment in Brazil. We observe the worker's year of birth, gender, and education. In addition, we observe monthly wage, municipality of the job, industry, and other characteristics of the establishment.

We aggregate the worker-level data at the microregion level to obtain employment information for each of the 558 local labor markets (LLMs) in Brazil. In addition, we compiled detailed statistics of the LLMs: gender composition (i.e., share of male workers), industry composition (share of workers in 10 industry's groups), skill composition (share of workers with low, medium and high skill), age composition (share of young, adults and older workers), GDP, GDP per capita, population, and population density.

3.2 Treatment Assignment

Each LLM i draws a treatment status (D_i) from a random binomial distribution, where the probability of being treated is correlated with their baseline characteristics (X_i). We

---

[10]Sampled from more than 80 million workers that were full-time employed (i.e., at least 30 hours per week) in December of any year. Our main results are based on a sample designed to mimic typical data availability. The findings are similar when using the full dataset.

simulate two alternative treatment assignments, as follows:

$$P_{1,i}(X_i) := P[D_i = 1|\ high.school_i] \tag{1}$$

$$P_{2,i}(X_i) := P[D_i = 1|\ high.school_i,\ manufacturing_i,\ pop.density_i] \tag{2}$$

where $high.school_i$ is the share of workers in region i with a high school degree; $manufacturing_i$ is the share of employment in manufacturing; and $pop.density_i$ is population (1,000) per km^2. All these characteristics are with respect to the year 2000. *Treatment 1* ($P_{1,i}$) is negatively correlated with the baseline share of high-school workers, and applies to 309 LLMs on average, while the remaining 249 LLMs serve as the comparison group. In contrast, *Treatment 2* ($P_{2,i}$) is positively correlated with the baseline share of workers with high school education and population density but is negatively correlated with the baseline share of manufacturing in the LLM. On average, there are 219 treated LLMs, and 339 serving as control.

Of course, the most effective approach would be to control for the observable characteristics that determine the treatment allocation. However, we aim to study the performance of alternative strategies to account for differential trends when correlates of the treatment are unobserved; we therefore implement our analysis as if these correlates are not observed by the researcher. One obvious concern with this strategy is that the confounder that we chose here may not be sufficiently representative of the type of confounders that cause non-parallel trends in other settings. However, the limitations (and potential advantages) of each method will turn out to be quite systematic and intuitive, and we will discuss to what extent they are expected to generalize. As we also provide analytical results, the empirical applications serve primarily as illustrations of our conceptual results.

3.3 Random Selection of Event Time

In order to avoid results that might arise due to patterns in specific periods, we also randomize the year of treatment (T_0) across the simulations. We draw T_0 from a random uniform (integer) distribution from 2005 and 2010. Then we restrict the period of analysis so as to keep nine event times before and after the treatment (year $\in [T_0 - 9, T_0 + 9]$). For simplicity

in the exposition of the results, we set event time $T_0 = 0$ as the baseline.

Therefore, both the baseline (T_0) and the assignment of treatment status D_i vary across simulations. Within simulations, T_0 is homogeneous: all units for which $D_i = 1$ are treated at the same time.[11] We estimate dynamic treatment effects for each event time by comparing the treated regions ($D_i = 1$) with the untreated regions ($D_i = 0$) using the strategies described in the previous section.

3.4 Monte Carlo Simulations

We evaluate the performance of the strategies that are commonly applied to deal with differential trends by implementing the specifications discussed in Section 2 across 1,000 Monte Carlo simulations (separately for each treatment). In each simulation $\omega = 1, .., 1,000$, we estimate γ_t^ω (for $-9 \le t \le -1$), and the post-treatment parameters β_t^ω (for $1 \le t \le 9$), as well as their respective standard errors. We weight each observation by the size of the labor market in $t = 0$, measured by log of workers, and the standard errors are clustered at the LLM level.

Since the true effects are zero, we quantify the bias arising from incidental trends by averaging the estimates across all simulations. Estimators might be biased in different directions across simulations, resulting in a small average bias. Further, imprecise estimators often result in wider confidence intervals, increasing the likelihood of encompassing the true parameter value, even in the presence of systematic bias. We measure dispersion using the standard deviation (SD) of the estimated coefficients across simulations, and summarize the trade-off between bias and variance with the root mean squared error (RMSE).

Appendix B contains additional explanation of the simulations and the main specifications, as well as a formalization of the performance measures. We also present a summary of the performance of the different methods considered in this paper in Table B.1, including their coverage (proportion of simulations where the true value falls within the 95% confidence interval).

---

[11]We expect our main findings to extend to staggered treatment settings, under proper estimation approaches. See Wooldridge (2021) for linear specifications that are suitable for staggered treatment designs, and Callaway and Sant'Anna (2021) for a more flexible approach.

4 Main Results

Here we analyze the simulation results for the main specifications described in Section 2. The top panels of Figure 1 plot the average coefficients across simulations, illustrating the bias associated with deviations from the parallel trends assumption. The "baseline" specification (S1) serves as the benchmark that does not account for existing differential trends between treated and untreated regions. The baseline coefficients will therefore capture trends that arise due to the correlation of the simulated treatment with regional confounders. We consider two treatment definitions, as described in Section 3.2.

The baseline trends depicted in Figure 1a show that *Treatment 1* happens to be associated with a gradual decline in employment, resulting in a negative bias in the impact estimates (recall the true treatment effect is zero). On average (across simulations, see Section 3.4), this bias reaches -2.7% at event time t=9. In contrast, *Treatment 2* is associated with an increasing and stronger trend in employment, accumulating a gap of almost 5% at t = 9 (Figure 1b). In line with these patterns, we are more likely to (falsely) reject the null hypothesis of a zero treatment effect for *Treatment 2*. Strategies that aim to account for differential trajectories based on pre-trends have therefore a greater potential of improvement in the second scenario.

How well do standard methods perform in correcting for such (pre-)trends? Figure 1 summarizes the performance of three commonly applied methods (see Section 2): (i) trend-as-control, (ii) linear trends, and (iii) group-by-time (e.g., state-by-year) FEs. A first observation is that the methods perform very differently. A second observation is that the methods have different strengths; for example, while the linear trends extrapolation reduces the bias more effectively, it also increases the variance of the coefficient estimates more than the other methods. A third important observation is that trend-as-control does not work well; the bias post-treatment remains nearly as large as in our baseline estimator. We provide a detailed analysis of these patterns in the next sections.

(a) Bias (*Treatment 1*)

(b) Bias (*Treatment 2*)

(c) Standard Deviation (*Treatment 1*)

(d) Standard Deviation (*Treatment 2*)

(e) RMSE (*Treatment 1*)
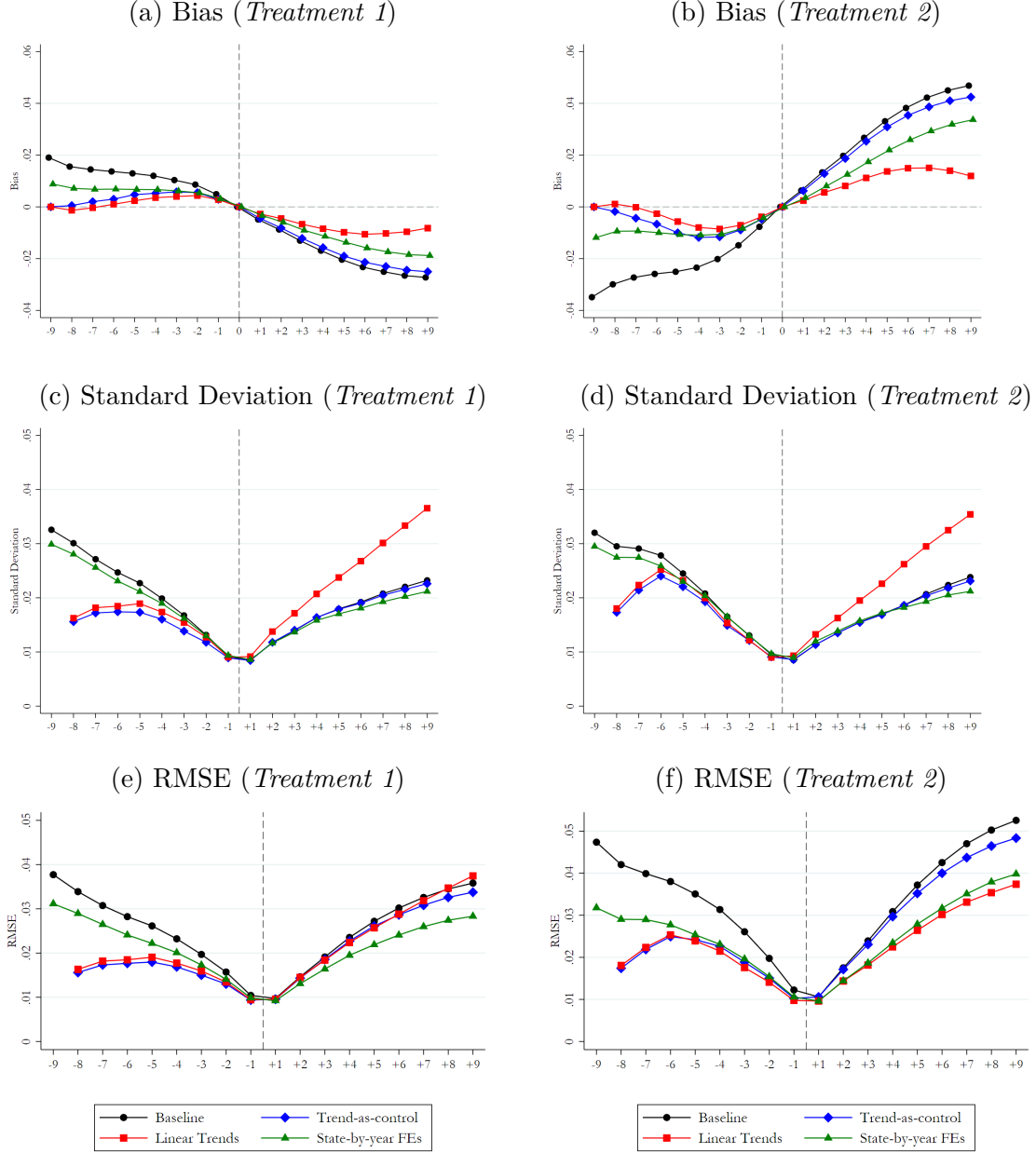
(f) RMSE (*Treatment 2*)

Figure 1: Simulation Results for Main Strategies

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of employment. See Section 2 for definition of each estimator and Section 3.2 for treatment definitions. Average number of treated LLMs: 309 in *Treatment 1* and 219 in *Treatment 2*.

4.1 Trend-as-control and Matching on Pre-Trends

The coefficients from the trend-as-control specification displayed in the top panels of Figure 1 capture the differences in trends after conditioning on pre-treatment trends over $k = 9$ pre-treatment years (interacted with event time, see eq. (S2)). As a consequence, there is a mechanical improvement in the pre-treatment period, with a nearly full elimination of pre-trends. But strikingly, the post-treatment estimates remain largely unchanged, offering

16

hardly any improvement over the baseline without controls. These pre- and post-treatment patterns make for a deceptive combination: the strong bias reduction in the pre-period could lead researchers to incorrectly conclude that potential threats to parallel trends have been eliminated, creating a misleading sense of the effectiveness of this method.

The remaining panels report performance measures that account for precision. Consistent with the previous findings, trend-as-control reduces both the SD and the RMSE in the pre-period, but has little effects on the precision of the estimates in the post-period. We find this same pattern across treatments and outcome variables. In Appendix Figure C.2 we show that while the trends-as-control approach performs better when measured over longer periods, it remains ineffective regardless of whether the pre-trend is measured over short or long periods.[12]

Notably, these findings extend to alternative specifications and to related methods, such as interacting the covariate with the treatment dummy or the doubly robust approach from Sant'Anna and Zhao (2020) (Appendix Figure C.1). We also show that the coefficient estimates remain similar when matching on pre-trends rather than including those pre-trends as a regressor in the event study. This result is intuitive, as matching and regression methods are fundamentally similar under common support, differing primarily in how covariate-specific effects are weighted to obtain average treatment effects (Angrist and Pischke, 2009).

Prior literature (Chabé-Ferret, 2015, 2017; Ham and Miratrix, 2024) has highlighted that matching on pre-treatment outcomes can *introduce* bias in contexts where selection depends on these outcomes, as in the case of Ashenfelter's dip. The intuition is that participants who select into a program often experience a transitory shock in their outcomes prior to treatment. Thus, matching (or controlling) for pre-treatment outcomes effectively means also matching on this shock, resulting in biased estimates. Differently from these studies, we focus on controlling for or matching on pre-treatment *growth*, extending the analysis to alternative sources of selection. As we show below, even without selection on transitory shocks, the trend-as-control approach can still suffer from attenuation bias.

---

[12]We show that pre-treatment coefficients exhibit some bias when controlling for shorter trends (i.e., growth over $k = 1$ and $k = 3$ pre-treatment periods). Nevertheless, in both cases, the post-treatment estimates remain largely unchanged.

4.1.1 Why Do Trend-as-control and Matching Tend to Fail?

Why does the trend-as-control approach perform so poorly in our simulations, and why do we expect this pattern to generalize to other settings? We provide an intuitive explanation here which we then formalize in the next subsection. Since the analytical results are derived in terms of conditional expectation, they broadly generalize, extending, for instance, to matching on pre-trends.

In a nutshell, the trend-as-control approach mitigates bias driven by serial correlation in the outcome growth at the unit (e.g., regional) level but does not properly account for trends that arise instead on the treatment level. This is problematic because serial correlation at the unit level is often low, even in settings where there are highly persistent shocks at the treatment level. For example, even though the treated and untreated groups experience systematically different employment trends in our simulations (Figure 1), individual regions show only weak serial correlation in employment growth. Trend-as-control only captures the latter, and thus fails to address the systematic difference in trends between the treatment and control groups.

An alternative and perhaps useful way to understand this argument is to view trend-as-control as a "noisy regressor". Unit-level trends are only a noisy proxy for treatment-level trends. Therefore, by including region-level pre-trends on the right side of the regression we introduce attenuation bias: the coefficient on these pre-trends will be biased towards zero, implying that they fail to adequately correct for differences in trends at the treatment level. The greater the fluctuations or "noise" at the region-level outcome, the more severe the attenuation bias and the less effective the trend-as-control approach.[13] One obvious source for such noise at the unit level is sampling error, implying that trends-as-control will do worse when applied to subsamples in which fewer workers are observed per region.

As we show below, trend-as-control can be effective when deviations from parallel trends are primarily driven by persistent outcome growth at the regional level. However, this

---

[13]Our argument here is related to Freyaldenhoven et al. (2019) and Brown and Butts (2025). They show that incorporating covariates X_i as proxies for the unobserved confounding factors directly as controls only works when X_i are not noisy. Otherwise, Freyaldenhoven et al. (2019) propose to use an instrument for X_i (e.g., leads of the treatment/policy), while Brown and Butts (2025) employ these covariates as instruments for the factor loadings within an interactive fixed effects model.

condition might often not hold. Outcomes such as wage or employment growth tend to vary more for the individual regions (such as labor markets) that constitute a treatment group than for the group overall. We are therefore skeptical on how effective trend-as-control or matching on pre-trends can be in typical applications. On a positive note, it is straightforward to test whether differential pre-trends are due to serial correlation at the unit level (which the method can address) or at the treatment level (which it cannot).

We provide an example in Figure 2, which plots the coefficient estimates from regressions of employment growth on lagged trends (normalized by their respective duration) at the treatment and regional levels. As already suggested by the event study estimates, employment growth is highly persistent at the treatment level, with a first-lag coefficient around 0.75.[14] The figure also illustrate that these trends extend over longer periods, consistent with our earlier findings. In contrast, there is little or even *negative* persistence at the regional level, with a small negative first-lag auto-correlation (around -0.14), implying that LLMs experiencing expansion in one year tend to see slower growth in the following period. Over longer pre-trends, there is essentially no serial correlation at the regional level.

These findings illustrate that differential trends on the treatment level can emerge even when there is only weak persistence at the level of analysis. While the specific LLMs experiencing high growth may vary in any given period, regions with particular baseline characteristics – i.e., the treated group – may exhibit systematically different growth rates. We next formalize this relationship between growth and pre-trends using a linear factor model for potential outcomes, which allows for time-varying unobservables.

4.1.2 When Could Trend-as-control Work?

We illustrate the behavior of the trend-as-control approach from a setting where parallel trends deviations arise from differential exposure to common shocks (i.e., based on a factor model). Let $y_{i,t}^0$ be the potential outcome of region i in the untreated scenario in each period

---

[14]We measured the persistence of growth at the aggregate level using the simulations for the two treatment scenarios described in Section 3. Specifically, we calculated the yearly (weighted) average number of workers for both treated and control groups, then regressed average employment growth on lagged trends. The coefficients and the confidence intervals are the average across simulations.
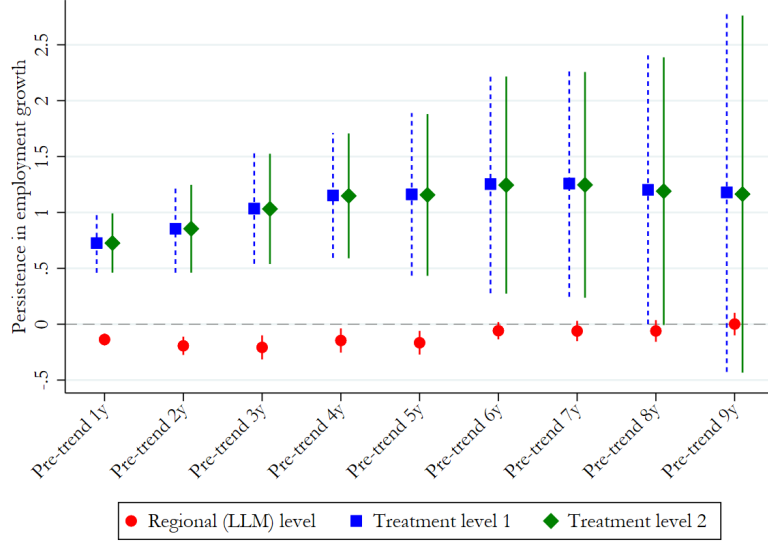
Figure 2: Persistence of Employment Growth

Notes: Coefficients from separate regressions of employment growth on lagged trends, where each trend is normalized by its respective duration. For instance, for "pre-trend 5y" the regressor is cumulative employment growth over five years, divided by five. The *regional level* regressions are weighted by the log number of workers in the region and include year FEs. The *treatment level* coefficients and standard errors are averaged across 1,000 simulation draws within *Treatment 1* and *Treatment 2* scenarios.

t, defined as follows:

$$y_{i,t}^0 = \tau_t + \mu_i + f_t \alpha_i + \varepsilon_{i,t} \tag{3}$$

where f_t is a $(1 \times F)$ vector of unobservable common factors, α_i is a $(F \times 1)$ vector of factor loadings (i.e., unit-specific exposure to common shocks), and $\epsilon_{i,t}$ is an idiosyncratic shock. The linear factor model presented in equation (3) is a standard approach in the synthetic control literature and a natural extension of the two-way fixed effects model commonly used in the difference-in-differences literature.[15] By incorporating the interactive fixed effects $f_t \alpha_i$, it allows for time-varying unobservables that may affect units differently.

The demeaned potential outcome then is given by: $\dot{y}_{it}^0 = \dot{\mu}_i + f_t \dot{\alpha}_i + \dot{\varepsilon}_{it}$, where $\dot{y}_{it}^0 :=$ $y_{it}^0 - \mathbb{E}[y_{it}^0]$. Taking the long-differences between t and the baseline and defining $\lambda_t := f_t - f_0$, we obtain the potential outcome growth as follows:

$$\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0 = \lambda_t \dot{\alpha}_i + \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0} \tag{4}$$

---

[15]This model is employed in the seminal synthetic control paper by Abadie et al. (2010). Additional examples include, among others, Gobillon and Magnac (2016), Ferman and Pinto (2021), Arkhangelsky et al. (2021), and Brown and Butts (2025).

Therefore, we obtain the observed outcome growth for each period t with respect to the baseline as follows:

$$\Delta \dot{y}_{i,t} = \beta_t^0 D_i + \underbrace{\lambda_t \dot{\alpha}_i + \Delta \dot{\varepsilon}_{i,t}}_{\Delta \dot{y}_{i,t}^0} \tag{5}$$

where $\Delta \dot{y}_{i,t} := \dot{y}_{i,t} - \dot{y}_{i,0}$, $\Delta \dot{\varepsilon}_{i,t} := \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}$, $\lambda_t := f_t - f_0$, and $D_i = 1$ if unit i if treated, zero otherwise. Note that β_t^0 in equation (5) is not, a priori, the same at the β_t parameter in the baseline specification (S1). While β_t^0 represents the treatment effect, β_t identifies the average treatment effect on the treated (ATT_t) in equation (S1) only under the parallel assumption (see Appendix A.1).

For simplicity, we derive our main argument under a set of assumptions that are standard (no anticipation, sampling) or innocuous (uncorrelated idiosyncratic shocks). We focus on the selection on unobservables that arises from the correlation between treatment assignment (D_i) and unit-specific exposure to common shocks (α_i).[16] As shown in Appendix A.1, the asymptotic **unconditional bias** ($\tilde{\Delta}_t$) from a regression that omits the factor structure is the following (Proposition A.1):

$$\tilde{\Delta}_t := plim \ \hat{\beta}_t - \beta_t^0 = \lambda_t \frac{Cov(\alpha_i, D_i)}{Var(D_i)} \tag{6}$$

The asymptotic bias in equation (6) states that the differential trend observed in the outcome results from differential exposure to common shocks varying in time. Since the vector α_i is unobservable, one potential alternative to mitigate the associated bias is the trend-as-control approach. Let, for instance, the pre-trend over k pre-treatment periods be defined as $PreTrend_{i,k} := \dot{y}_{i,-1} - \dot{y}_{i,-(k+1)}$.[17] Under no anticipation, we can write it using our

---

[16]As we assume no selection in the idiosyncratic shocks, we abstract from the source of bias in DiD designs as discussed in Chabé-Ferret (2015, 2017). This literature has shown that, when selection is based on pre-treatment outcomes (e.g., Ashenfelter's dip), matching on these outcomes can *introduce* bias. In contrast, we show that even in the absence of selection on transitory shocks, the trend-as-control approach can still suffer from attenuation bias. In Appendix A.2.2 we derive our argument under more general conditions, with the same substantive implications.

[17]To abstract from a mechanical correlation between the "measurement error" in the pre-trend and the error term in equation (5), we define $PreTrend_{i,k}$ over k pre-treatment periods using $t = -1$ as the baseline rather than $t = 0$. In Appendix A.2.2, we explicitly account for this correlation when $t = 0$ is used as the baseline, and we also relax the assumption of serially uncorrelated idiosyncratic shocks.

factor model as follows:

$$PreTrend_{i,k} = \underbrace{\lambda_{-k}\dot{\alpha}_i}_{\text{Systematic component}} + \underbrace{\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)}}_{\text{Noise}} \qquad (7)$$

where $\lambda_{-k} := f_{-1} - f_{-(k+1)}$.

In order to express the outcome variation in equation (5) as a function of observed pre-trends, we must rewrite the factor structure in terms of past shocks. However, in a high-dimensional factor structure, there is no strong reason to expect that the trend-as-control approach will eliminate bias. For instance, post-trends may be driven by different factors than pre-trends, creating no direct link between the two observables. To address this, we consider scenarios where the trend-as-control approach and DID are more likely to succeed: one where the source of bias is one-dimensional, or alternatively, where the multiple factors evolve at the same rate. It corresponds to the assumption that there is a specific source of bias (e.g., industry structure) that plays a similar role in the pre- and post-treatment period (e.g., manufacturing shows a stronger business cycle pattern than other sectors, as in Blanchard et al. (1992)).

Formally, we assume that $\lambda_t = \rho_{t,k}\lambda_{-k}$, where $\rho_{t,k}$ represents the parameters capturing the persistence of shocks across k pre-treatment and t post-treatment periods.[18] This assumption accommodates, for example, linear trends where $\rho_{t,k} = \frac{t}{k}$ captures proportional growth over t post-treatment and k pre-treatment periods. Moreover, it allows for deviations from linearity, such as accelerating trends ($\rho_{t,k} > \frac{t}{k}$) or decelerating trends ($\rho_{t,k} < \frac{t}{k}$). This assumption implies that outcome growth in equation (5) can be rewritten as a function of observed pre-trends as follows (see Appendix A.2.1 for details):

$$\Delta\dot{y}_{i,t} = \beta_t D_i + \rho_{t,k}PreTrend_{i,k} + \underbrace{\left(\Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)})\right)}_{\text{error term}} \qquad (8)$$

Note that $PreTrend_{i,k}$ is a noisy proxy for the systematic component driving the differen-

---

[18]When multiple factors are present and their dynamics are not linearly structured, it becomes infeasible to summarize the factor structure using a single covariate. This represents a key distinction from the synthetic control method, which fits the factor loadings of the treated unit(s) using a long set of pre-treatment outcomes without imposing strong assumptions on the evolution of common shocks over time.

tial trends. As usual, the implication of the "measurement error" in the pre-trend covariate is that the estimate of the coefficient $\rho_{t,k}$ from equation (8) is biased towards zero (Proposition A.2 in Appendix A.2.1). The reliability of $PreTrend_{i,k}$ as a proxy of the structure of the unobserved factors can be assessed by the proportion of its variance explained by the variation in the systematic component (that is, the signal) rather than noise. Under the assumptions on sampling and the properties of the error terms, the *signal to total variance ratio* is:

$$\zeta := \frac{Var(\lambda_{-k}\dot{\alpha}_i)}{Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2} \tag{9}$$

Conditioning on trend-as-control will reduce bias more effectively when differential trends are primarily explained by correlations in common shocks over time, rather than idiosyncratic shocks, i.e., when the ζ is closer to 1. In this case, outcome growth is more persistent at the unit level, implying that pre-trend provides more information about the unobserved exposure α_i, acting as a better proxy for the systematic component of the bias. This result is summarized in the following relation between the unconditional and conditional asymptotic biases (Proposition A.3 in Appendix A.2.1):

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \tilde{\Delta}_t \left(\frac{1 - \zeta}{1 - \tilde{R}^2} \right) \tag{10}$$

where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i, ζ is the signal to total variance ratio, $1 - \zeta$ is the noise to total variance ratio, and $\left(\frac{1-\zeta}{1-\tilde{R}^2} \right) \in (0,1)$. Trend-as-control will partially eliminate bias from the unobserved factor structure by a proportion $\left(\frac{\zeta - \tilde{R}^2}{1-\tilde{R}^2} \right) \in (0,1)$.

4.1.3 Illustration with Simulated Data

To illustrate our analytical result, we simulate outcome growth based on equation (5) as follows: $y_{i,t} = y_{i,t-1} + \beta_t D_i + (f_t - f_{t-1})\alpha_i + \Delta\epsilon_{i,t}$. We model the common shock as a linear trend, such that $(f_t - f_{t-1}) = 1$. In addition, $\Delta\epsilon_{i,t} \sim N(0, \sqrt{2}\sigma)$ i.i.d. for all (i,t) and $\alpha_i \sim \text{Uniform}(0,1)$ i.i.d. Furthermore, we simulate a random treatment allocation based on α_i, where $P(D_i = 1) = \alpha_i$, which implies that units following an increasing employment trend over time are more likely to be treated. We impose that the true treatment effect is

zero (i.e., $\beta_t = 0$). The baseline outcome is the log of workers in Brazilian LLMs in 2009. We generate five pre and five post-treatment event times, in addition to the baseline (set as $t = 0$) and measure the pre-trends from $t = -5$ to $t = 0$. Note that the *signal to total variance ratio* is given by:

$$\zeta = \frac{Var(\alpha_i)}{Var(\alpha_i) + Var(\Delta\epsilon_{i,t})} = \frac{1/12}{1/12 + 2\sigma^2}$$

Figure 3a displays the unconditional bias, measured by $\hat{\beta}_t$. In panel (a), we also provide the bias conditional on trend-as-control in two different scenarios: (i) large noise, with $\sigma = 1$ and $\zeta = 0.04$; (ii) small noise, with $\sigma = 0.03$ and $\zeta = 0.98$. In the trend-as-control specification, we linearly control for outcome growth over five pre-treatment years as in equation (S2). In panel (b), we consider alternative "matching on pre-trends" strategies in the first scenario (i.e., $\zeta = 0.04$).

The baseline coefficients (unconditional bias) in Figure 3a display upward trends, which are equivalent across the different scenarios. Trend-as-control (controlling for growth over five pre-treatment years) yields unbiased *pre*-treatment estimates regardless of the signal to total variance ratio. However, the post-treatment coefficients reveal that the effectiveness of this approach depends on the time-series properties of the data. While trend-as-control removes all the bias when the noise is significantly small relative to the signal (i.e., $\zeta = 0.98$), it has limited success in eliminating bias in the noisier scenario ($\zeta = 0.04$). Panel 3b illustrates that this underwhelming performance extends to matching on pre-treatment outcomes or trends.[19]

4.1.4 Diagnostics for Trend-as-control

Our analytical results demonstrate that the performance of pre-trends as control depends on how informative they are about differential trends that arise at the treatment level. In particular, pre-trends perform better when the variance of idiosyncratic shocks is small

---

[19]A previous study using simulated data has already discussed the limited benefits of matching on pre-trends (Daw and Hatfield, 2018), noting that improvements are smaller when outcomes exhibit weak serial correlation. We refine this argument by showing that what matters is the serial correlation in outcome *growth*, not in levels.

(a) Performance with Different Levels of Noise   (b) Controlling vs. Matching on Pre-trends
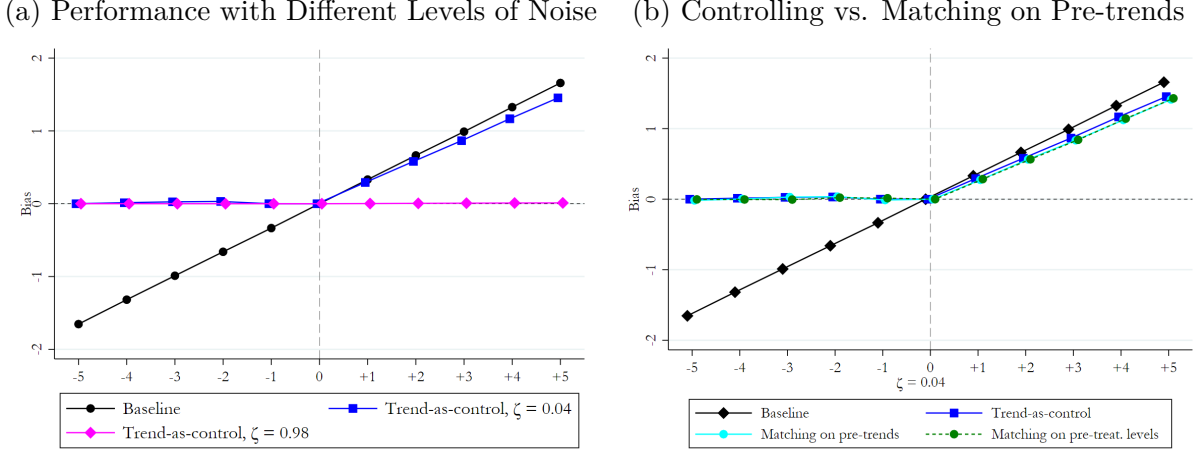
Figure 3: Illustration of the Trend-as-control Approach

Notes: Simulations using 2009 employment data from Brazilian LLMs as the baseline. The outcome is $y_{i,t} = y_{i,t-1} + growth_{i,t}$, where $y_{i,0}$ is log of workers in 2009. Based on equation (5) and setting $f_t = t$, $growth_{i,t} = \alpha_i + \Delta\epsilon_{i,t}$, where $\Delta\epsilon_{i,t} \sim N(0, \sqrt{2}\sigma)$ i.i.d. for all (i,t) and $\alpha_i \sim$ Uniform$(0,1)$ i.i.d.. We randomly allocate a treatment status across Brazilian LLMs that depends on the factor loadings: $P(D_i = 1) = \alpha_i$. In panel (a), we consider two scenarios: (i) large noise ($\sigma = 1$) and *small signal to total variance ratio* ($\zeta = 0.04$); (ii) small noise ($\sigma = 0.03$) and *large signal to total variance ratio* ($\zeta = 0.98$). In the trend-as-control specification, we linearly control for outcome growth over five pre-treatment years as in equation (S2). In panel (b), we consider alternative "matching on pre-trends" strategies in the first scenario. The treatment occurs in $t = 0$ and the true treatment effects are zero.

relative to the variance of systematic factors driving deviations from parallel trends. This result is related to the synthetic control literature, which highlights the need for a sufficiently large number of pre-treatment periods relative to the scale of transitory shocks in order to accurately recover unobservable factors (Abadie et al., 2010; Ferman and Pinto, 2021; Arkhangelsky and Hirshberg, 2023). This condition is often not met in typical difference-in-differences applications, where the number of observations is much larger than the number of periods. Additionally, trends tend to be more persistent at higher levels of aggregation, supporting the use of synthetic control in these contexts (Abadie and Vives-i Bastida, 2022).

In contrast, difference-in-differences applications often deal with microdata, where time series tend to be more volatile. When the variance of the idiosyncratic shock is large relative to the systematic component, controlling for $PreTrend_{i,k}$ can lead researchers to significantly underestimate the persistence of the common shocks. The implication of our findings is that the trend-as-control approach may be of limited value – or even actively misleading – in many applications. Fortunately, it is fairly straightforward to assess whether the approach is promising in a given setting.

Diagnostic 1. *Placebo-in-time*

One diagnostic approach, inspired by the synthetic control literature (Abadie, 2021), involves backdating the reference period for pre-trends and examining the event study estimates during the resulting placebo periods (i.e., from the new backdated baseline up to the original baseline). The absence of treatment effects in these placebo periods provides supportive evidence that controlling for pre-trends effectively mitigates bias.

For illustration, we use the same simulations described in the previous section (Figure 3), where the benchmark trend-as-control is measured over the entire pre-treatment period. In Figure 4, alongside the benchmark trend-as-control estimates, we also present two placebo estimates, where the reference period is shifted to $t = -3$: the "baseline" estimate without controls, used to measure unconditional bias; and the "placebo trend-as-control", which controls for pre-trends measured from $t = -5$ to $t = -3$. Therefore, we can assess the coefficients between $t = -3$ and $t = 0$ for diagnostic purposes.

In the scenario with noisy pre-trends (and a low signal to total variance ratio), while the coefficients of the benchmark trend-as-control approach are misleading, the "placebo trend-as-control" estimates in Figure 4.a reveal the weak performance of this approach in the placebo period, with coefficients close to the baseline. In contrast, when the pre-trends are informative (Figure 4.b), the placebo-in-time estimates provide supporting evidence for the effectiveness of this approach.

However, this approach may still produce misleading conclusions in situations where selection into treatment is driven by transitory shocks that occur close to the treatment date (Arkhangelsky and Hirshberg, 2023). We thus propose an alternative diagnostic, based on the time series properties of the data.

Diagnostic 2. *Time series properties of the data*

When multiple periods are available, researchers can assess the features of the time-series by regressing outcome growth on lags <u>using pre-treatment data</u> at the unit level. Let $\Delta y_{i,-k} = y_{i,0} - y_{i,-k}$ and $\Delta y_{i,-s} = y_{i,-(k)} - y_{i,-(k+s)}$, and consider a regression of pre-trends on lags as follows:

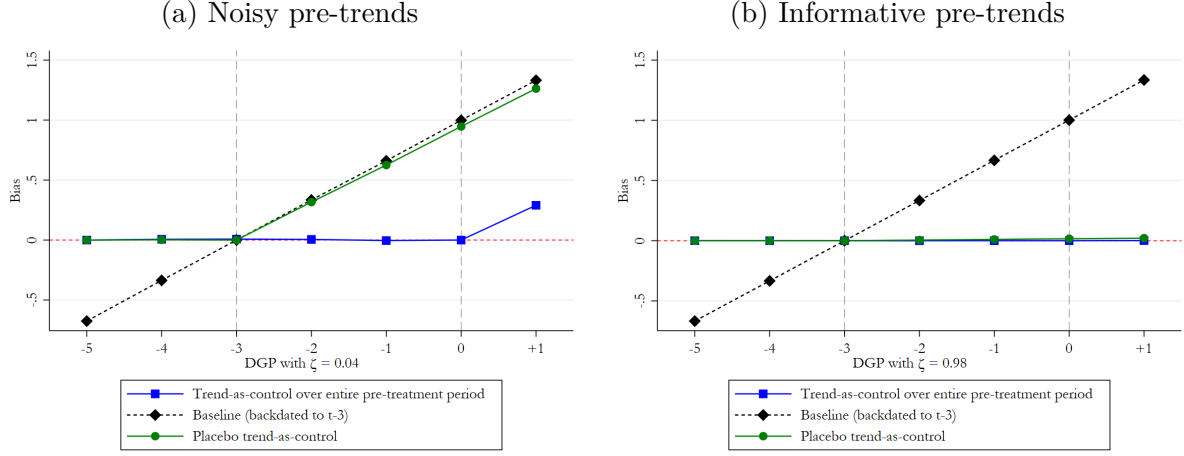$$\Delta y_{i,-k} = \pi_{k,s} \Delta y_{i,-s} + \nu_{i,-k}$$

26

Figure 4: Placebo-in-time

Notes: Simulations using 2009 employment data from Brazilian LLMs as the baseline. The outcome is $y_{i,t} = y_{i,t-1} + growth_{i,t}$, where $y_{i,0}$ is log of workers in 2009. Based on equation (5) and setting $f_t = t$, $growth_{i,t} = \alpha_i + \Delta\epsilon_{i,t}$, where $\Delta\epsilon_{i,t} \sim N(0, \sqrt{2}\sigma)$ i.i.d. for all (i,t) and $\alpha_i \sim$ Uniform(0,1) i.i.d.. We randomly allocate a treatment status across Brazilian LLMs that depends on the factor loadings: $P(D_i = 1) = \alpha_i$. We consider two scenarios: Panel (a) large noise ($\sigma = 1$) and *small signal to total variance ratio* ($\zeta = 0.04$); and Panel (b) small noise ($\sigma = 0.03$) and *large signal to total variance ratio* ($\zeta = 0.98$). The treatment occurs in $t = 0$ and the true treatment effects are zero. In the standard trend-as-control specification, we linearly control for outcome growth over five pre-treatment years as in equation (S2). For the placebo-in-time, we backdate the baseline to $t - 3$, thus omitting the interaction of this year with the treatment dummy instead of $t = 0$. The placebo trend-as-control includes outcome growth from $t = -5$ to $t = -3$.

In Appendix A.2.1, we show that the signal to total variance ratio can be approximate by the size of the serial correlation at the unit level ($\pi_{k,s}$) relative to the persistence of the observed pre-trends at the treatment level (i.e., relative to $\rho_{k,s} = \frac{\tilde{\Delta}_{-k}}{\tilde{\Delta}_{-s}}$).[20] In particular, we show that controlling for $PreTrend_{i,s} := \Delta y_{i,-s}$ when estimating "treatment effects" in period $t = -k$ will mitigate bias as follows:

$$\frac{\tilde{\Delta}_{-k} - \tilde{\Delta}_{-k}(\textbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} = \frac{plim \; \hat{\pi}_{k,s}}{\rho_{k,s}} - \tilde{R}_s^2 \frac{\tilde{\Delta}_{-k}(\textbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} \tag{11}$$

All the elements of equation (11) can be estimated using pre-treatment data. If one expect the persistence of growth at the unit level ($\pi_{k,s}$) and the persistence of common shocks ($\rho_{k,s}$) to follow a similar pattern in the post-treatment data, the relative magnitude of these elements can provide insight into the effectiveness of the trend-as-control strategy. For instance, the weak performance of this strategy in our setting can be attributed to the relatively low serial correlation of employment growth within regions (e.g., $\pi_{9,-9} = 0.12$), in

---

[20]Note that, since $\tilde{\Delta}_{-k}$ measures the bias over k pre-treatment periods using $t = -k$ as the "baseline", it is equal to *minus* the coefficient of the event study (where the baseline is $t = 0$).

contrast to the persistent trend at the treatment level (e.g., $\rho_{9,-9} \approx 1$; see Figure 2).

In Appendix A.2.2, we further allow serial correlation in the idiosyncratic shocks. As a result, trend-as-control can potentially *introduce bias.* Equation (11) remains unchanged, but since the bias expressions contain a term that depends on the shocks' autocorrelation, the left-hand side can assume negative values. This happens because now $\text{sign}(plim \; \hat{\pi}_{k,s})$ is not necessarily equal to $\text{sign}(\rho_{ks})$. In fact, we highlight that $\text{sign}(plim \; \hat{\pi}_{k,s}) \neq sign(\rho_{ks})$ is a sufficient condition for *bias introduction.*[21]

4.2 Linear Trends

As shown in the top panels of Figure 1, the linear trends extrapolation (S3) exhibits much better performance in terms of bias and coverage than the trend-as-control approach. Linear trends do not fully eliminate the bias in our simulations, as the underlying trends are non-linear, amplifying over time in both *Treatment 1* and *2* (cf. baseline estimator). Still, they eliminate most of the bias, and the confidence intervals cover the true null effects almost 95% of the times across all post-treatment years (Appendix Table B.1).[22]

These results illustrate that linear trend extrapolations may reduce bias even in settings in which the true underlying trends are not linear. However, in our applications, the approach also leads to a larger dispersion of estimates compared to other strategies (Figures 1c and 1d). This trade-off between variance and bias is summarized by the RMSE in the bottom panels of Figure 1. Since the underlying trends in *Treatment 1* are not strong, the large variance associated with the linear trends strategy overcompensates its bias mitigation, resulting in a larger RMSE compared to alternative strategies. Conversely, Figure 1f illustrates that in scenarios with stronger underlying trends, the benefits of employing linear trends for bias reduction can outweigh the increased dispersion, such that the RMSE decreases.

The performance of the linear trends correction in terms of bias and RMSE depends

---

[21]This pattern is evident in the Brazilian data, where employment growth is negatively correlated with its first lag (Figure 2). Consequently, in Appendix Table B.1, we show that controlling for growth in the year preceding treatment ("Short Trend-as-control") introduces bias into the estimates at $t+1$.

[22]In our main analysis, we fitted the linear trends slope by omitting the indicator for event time $t = -9$ from our regressions. Thus, the respective post-treatment estimates represent the differential trend net of the linear extrapolation of growth over nine pre-treatment years.

therefore on the strength of underlying trends. Linear trends should not be included by default, as they can substantially worsen the quality of the resulting estimates in some settings. Most obviously, the benefits of the linear trends correction depends on the extent to which the underlying trends are indeed linear. To illustrate these points further, we provide alternative simulations with log of wage as the outcome in Appendix D. As the underlying trends are much weaker in this case, the linear trends extrapolation performs worse in all dimensions.

Linear trends will thus perform better in scenarios in which there are indeed strong underlying trends, and while reducing bias, they may increase dispersion. In the next sections, we provide additional results that are perhaps less intuitive. In Section 4.2.1, we demonstrate that if the data exhibits mean reversion, the linear trends extrapolation tends to *over-correct*, leading to a bias in the opposite direction. This problem is less apparent when considering the average bias across simulations (as in the top panels of Figure 1), but becomes evident in a more targeted analysis. This over-correction also contributes to the wider dispersion associated with the linear trends approach.

Finally, we ask *how* researchers should control for pre-trends. In Section 4.2.2, we show that fitting treatment-specific linear trends is equivalent, in expectation, to unit-specific linear trends, as long as the units for fitting the specific trends are at the level of assignment of the treatment. Therefore, the performance is not improved when multiple linear trends are included, instead of a single slope. In Section 4.2.3, we discuss the performance of long vs. short linear trends. If the outcome is observed for many pre-periods, researchers have the choice to extrapolate trends from shorter or longer pre-treatment intervals. We illustrate the potential trade-off in this choice: while extrapolating from short (and therefore recent) pre-trends may reduce bias more effectively, it also amplifies issues related to noise and over-correction.

4.2.1 When Could Linear Trends Work?

When controlling for linear trends, the resulting post-treatment coefficient β_t from specification (S3) can be derived as follows:

$$\beta_t = DID_t - t \times \delta_k \tag{12}$$

where $DID_t = ATT_t + \Delta_t$ is the unconditional difference-in-differences (i.e., from the baseline specification), Δ_t is the unconditional bias, and δ_k is the slope of differential linear trends fitted in the pre-treatment period of length k.

The linear trends approach can be analyzed using the linear factor model for outcome growth explained in the previous section. The bias conditional on a linear trend extrapolation of length k is derived as follows (see Proposition A.4 in Appendix A.3):

$$\Delta_t(LT_k) = \left(\rho_{t,k} - \frac{t}{k} \right) (\lambda_{-k} \Delta\alpha) \tag{13}$$

Comparing the size of the bias net of the linear trends extrapolation ($\Delta_t(LT_k)$) with the bias of the baseline estimator (Δ_t) from equation (6), we can express their relative size as:

$$\frac{\Delta_t(LT_k)}{\Delta_t} = 1 - \frac{1}{\frac{k}{t} \times \rho_{t,k}} \tag{14}$$

As is intuitive, the effectiveness of linear trend extrapolation depends on how the common shocks evolve over time, specifically their linearity, which is determined by whether $\frac{k}{t} \times \rho_{t,k} \approx 1$. On average, in our simulations, we observed that the trends were amplifying over time (i.e., $\frac{k}{t} \times \rho_{t,k} > 1$). This explains the partial elimination of bias and the superior performance associated with the linear trends approach in our simulations. Nevertheless, the large dispersion of the estimates, along with a weak performance in terms of RMSE, might be suggesting patterns of over-correction across simulations. In other words, although the average bias is closer to zero, the wide dispersion might suggest that the linear trends approach may be over-correcting to different directions across the simulations.

In Figure 5, we present the coefficients of a regression of the post-treatment estimates obtained from the different strategies studied in this paper on pre-trends of symmetrical lengths. In both treatment allocations, the trajectories move in the same direction before and after the treatment, resulting in a positive correlation of the remaining trends after controlling for state-by-year FEs and trend-as-control. This is consistent with these methods only partially accounting for trends. In contrast, in scenarios with weaker post-treatment trends (*Treatment 1*), the coefficients of the linear trends extrapolation tend to be negatively correlated with pre-trends (Figure 5a). This negative correlation is due to "overshooting", where larger pre-trends are associated with stronger bias in the opposite direction in the post-treatment period. The correlation is closer to zero when the underlying trends amplify more strongly in the post-treatment period, as in *Treatment 2* (Figure 5b). However, even in this case, there is a tendency towards overshooting in the long run.
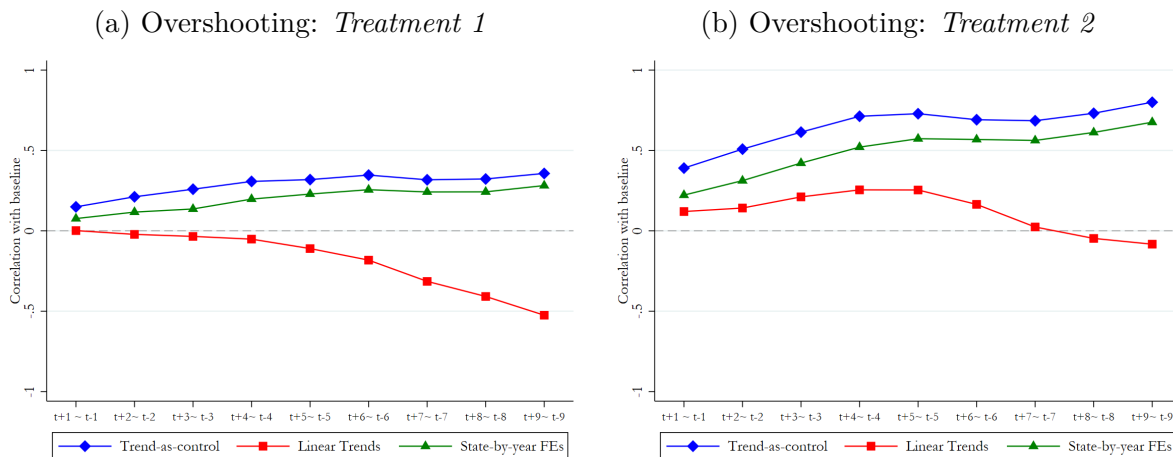

Figure 5: Overshooting of the Linear Extrapolation

Notes: Coefficients are obtained from regressions of the estimates across 1,000 simulations for each treatment allocation. We regress the post-treatment estimates from three main strategies on unconditional pre-trends of equal lengths.

4.2.2 Treatment vs. Unit-specific Linear Trends

The linear trends discussion was based on an approach that fits a single parameter for the treatment group, as recommended by Wooldridge (2021). This yields the simple interpretation in terms of differential trends between the treated and control groups. In practice, however, researchers often control for "unit-specific linear trends" (see Miller 2023, and

Table 1). Therefore, instead of a single parameter for the slope of the differential linear trends (δ_k), this approach allows a specific slope ($\delta_{i,k}$) for each unit i. One obvious question is whether allowing for such unit-specific trends provides any advantages compared to the simpler treatment-specific trend.

We show that both strategies are equivalent in expectations, as long as the units for fitting the specific trends are at the level of assignment of the treatment. In short, the treatment-specific linear trend slope (δ_k) represents the difference in the expected $\delta_{i,k}$ between the treated and untreated units.[23] A formal demonstration is available in Appendix A.3.2.

Figure 6 illustrates this equivalence using our simulation results. In our setting, the treatment is allocated at the unit (e.g., local labor market) level. In the top panels, we can observe that treatment-specific and unit-specific linear trends yield almost identical average coefficients. As noted ealier, one caveat with the linear trends approach is that it introduces noise. We here show that its performance is not improved when including multiple (unit-specific) linear trends, instead of just the treatment-specific one. This is illustrated in the bottom panels of Figure 6, showing that both specifications yield similar RMSE. Of course, when selecting the regional level at which to fit the linear trends, researchers must be aware that equivalence does not hold if the linear trends are fitted at a more aggregated unit than the treatment assignment. For example, in Figure 6 we present results for "state-specific linear trends" in a setting where the treatment is assigned at a finer (e.g., local labor market) level. Although more precise in terms of standard deviation, linear trends fitted at a more aggregate regional level exhibit lower performance in terms of bias and RMSE.

4.2.3 Long vs. Short Linear Trends

Our results indicate that linear extrapolations are more susceptible to overshooting when extrapolating over wider event windows, consistent with the idea that the assumptions behind linear trend extrapolation are "more plausible in a short panel" (Freyaldenhoven

---

[23]We highlight that the equivalence holds in expectations(i.e., $\delta_k \equiv \mathbb{E}[\delta_{i,k}|D_i = 1] - \mathbb{E}[\delta_{i,k}|D_i = 0]$), while some differences might arise in finite samples. In fact, our simulations (Figure 6) reveal that estimates from unit-specific linear trends are marginally smaller (thus less biased), but at the cost of marginally larger standard errors.

| (a) Bias (*Treatment 1*) | (b) Bias (*Treatment 2*) |
| (c) RMSE (*Treatment 1*) | (d) RMSE (*Treatment 2*) |

Legend:
- Treatment-specific Linear Trends
- Unit-specific Linear Trends
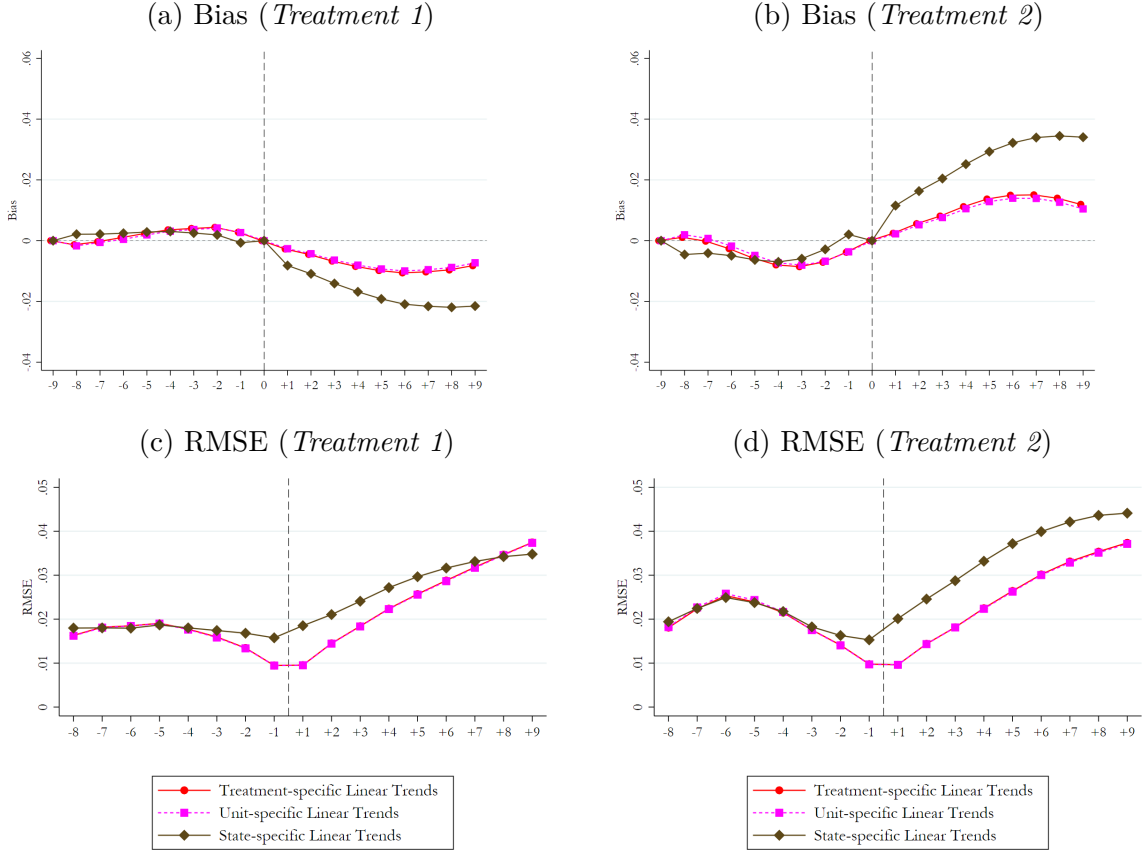- State-specific Linear Trends

Figure 6: Equivalence of Treatment-specific and Unit-specific Linear Trends

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of employment. See Section 2.2 for definition of the "treatment-specific linear trends" estimator and Section 3.2 for treatment definitions. The unit- and state-specific linear trends specification interacts the $Trend_t$ variable with indicators for each unit or state, respectively (see Appendix Section A.3.2).

et al., 2021). Another key consideration when controlling for linear trends is the decision of whether to extrapolate those trends from shorter or longer pre-treatment intervals, which essentially involves choosing which extra event time to exclude.[24]

To illustrate the potential trade-offs between short and long trends, we revisit our simulations. In our main analysis, we extrapolated "long linear trends" (over 9 pre-treatment years) up to a symmetrical length of time. In Figure 7, we compare this specification to a variant in which we fit the linear trends over a shorter interval (of just three pre-treatment years), labeled as "short linear trends". As shown in panels (a) and (b), the average bias in the post-treatment period is closer to zero for the "short" extrapolation, implying stronger

---

[24]As discussed above (and illustrated in Appendix A.3.1), due to multicolinearity, we need to omit an extra event time from the regression. For instance, if both the baseline and the coefficient for event time $-k$ are restricted to zero, the linear trend slope will capture differential growth over k pre-treatment periods. See Miller (2023) for alternative restrictions.

bias elimination. However, in panels (c) and (d), the RMSE associated with the short linear trends extrapolation exceed that of the longer one, indicating a much wider variability of the estimates. This illustrates an important trade-off in choosing between "short" or "long" pre-trends. On the one hand, recent pre-trends may be more informative about the expected strength of trends in the post-treatment period than pre-trends that occurred many years ago.[25] On the other hand, there tends to be higher variability in estimates based on "short" as compared to "long" trends. By extrapolating from these noisy pre-trends we therefore end up with noisy estimates of the post-treatment effects (i.e., high RMSE), even though the estimates have little average bias in either direction.

Moreover, even though short linear trends show little systematic bias on average across all simulations, they will tend to generate systematic distortions due to their greater susceptibility to regression to the mean.[26] This potential issue is already visible in panels (a) and (b) of Figure 7, as the short linear trends specification tends to overcorrect in the long term, such that the bias flips sign (cf. baseline and short linear trends in the last periods). But these panels only show the average bias across many simulations, and regression to the mean creates much larger distortions in any given application, in particular when pre-trends are pronounced. To illustrate this, Panels (e) and (f) display the average bias for a subset of simulation draws where pre-testing suggests potential rejection of the parallel trends assumption (i.e., p-value of the joint test of pre-treatment baseline coefficients is less than or equal to 10%). Consistently with our predictions, (short) linear trends extrapolation is related to bias in the opposite direction (i.e., overshooting), which increases in the length of the post-treatment period t. Importantly, this pattern is much stronger for the "short" linear trends approach, illustrating that over-correction tends to be more problematic when extrapolating linear trends from short pre-treatment periods.

What we cannot answer is the question of what is the optimal length of the pre-treatment period over which linear trends should be extrapolated. Intuitively, this should depend on the shape of trends in the pre-period, the length of the post-period over which we want to

---

[25]For example, in our applications trends amplify over time, so recent pre-trends are on average more similar in size to the post-treatment trends.

[26]The assumption that different units grow at different deterministic rates is implausible over long time horizons. This is obvious for outcomes that are bounded, but regression to the mean is pervasive even for outcomes that are not.

extrapolate those trends, and the objective function (e.g., the extent to which we are concerned about the bias or the dispersion of the resulting estimates). However, our discussion here provides insights about the trade-offs that researchers face in this choice.

4.3 Group-by-time FEs

A third popular strategy to address differential trends is to control for broader group-by-time (often, macro-region-by-time) FEs. As our treatment is assigned at the local labor market level, we study how the inclusion of state-by-year FEs affects the performance of the event study estimator.

As shown in panels (a) and (b) of Figure 1, incorporating state-by-year FEs (equation (S4)) decreases bias in our applications more effectively than trend-as-control, but not as effectively as linear trends. The remaining bias stems from differential trajectories that arise between labor markets within states. In our simulations, a significant portion of the bias persists within states, particularly in *Treatment 2* (Figure 1b). The potential of state-by-year FEs to address differential trends depends therefore on the extent to which these arise at the state vs. unit-level, a point that we formalize in Appendix A.4.

While the bias reduction is modest, the estimator with state-by-year FEs leads to the smallest standard deviation among all estimators, including the baseline estimator. For *Treatment 1* it also leads to the smallest RMSE. For *Treatment 2*, the estimator with linear trends has a smaller RMSE, owing to stronger underlying trends in this scenario and the more effective bias reduction of this estimator. One potential takeaway from these results is that linear trends is the more attractive option if the pre-trends are strong and bias rather than precision is the main concern, while state-by-year FEs is more suitable if the underlying trends are not very strong (or highly non-linear), and/or the precision of the estimates is a primary concern.

Still, the performance of group-by-time FEs will depend critically on the extent to which differential trends arise at the unit vs. group (e.g., state) level, and how much of the variation in the treatment itself occurs within vs. between groups. As shown in Appendix A.4.1, this

35

(a) Bias (*Treatment 1*)　　　　(b) Bias (*Treatment 2*)

(c) RMSE (*Treatment 1*)　　　　(d) RMSE (*Treatment 2*)

(e) Bias Short Linear Trends (*Treatment 1*)　　　　(f) Bias Short Linear Trends (*Treatment 2*)
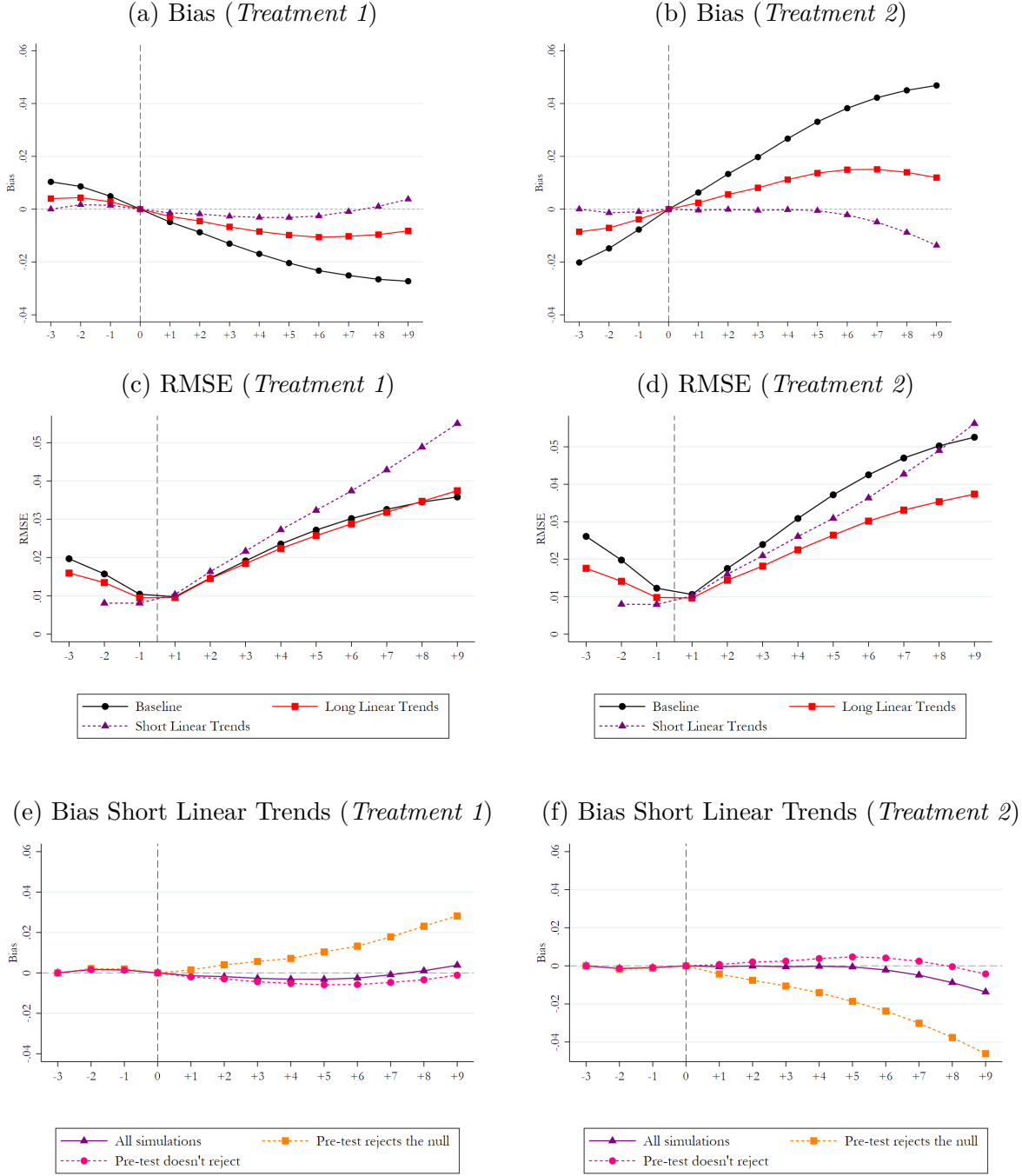
Figure 7: Performance of Short vs. Long Linear Trends

Notes: See Section 2.2 for definition of the linear trends' estimator and Section 3.2 for treatment definitions. The outcome is log of employment at the regional (i.e., LLM level). Panels (a), (b), (c) and (d) present summary statistics across 1,000 simulations, where "long" and "short" linear trends are from growth over $k = 9$ and $k = 3$ pre-treatment periods, respectively. Panels (e) and (f) also show the bias of the "short linear trends" extrapolation across two subsets of simulations. "Pre-test rejects the null" includes simulation draws where the p-value for the joint significance of the pre-treatment coefficients in the baseline specification is less than or equal to 10%, encompassing 168 and 225 draws in *Treatment 1* and *Treatment 2*, respectively. "Pre-test doesn't reject" corresponds to the remaining simulations with a p-value > 10%.

strategy effectively mitigates differential trends driven by confounds at the aggregate group level, while still relying on the parallel trends assumption within each group. The resulting

aggregate bias is a weighted average of within groups' deviations, where the weights are proportional to the size of each group and the within-treatment variation (Goodman-Bacon, 2021). If the (unconditional) bias is predominantly driven by unit-level confounds, group-by-time FEs offer limited potential for bias reduction. Additionally, this approach could introduce bias if it leads to larger weights being assigned to groups with stronger within differential trends.[27] Conversely, if smaller weights are assigned to group with stronger within differential trends, this approach will lead to partial bias elimination, as long as confounds at the group level move in the same direction as unit-level ones. This occurs because, in addition to eliminating between-group confounds, it also results in weaker aggregate within-group trends.[28]

5 Conclusion

The difference-in-differences literature has experienced exponential growth in recent years, with substantial focus on the limitations of the standard TWFE estimator and the development of new methods that are robust to heterogeneous effects over time and across treatment cohorts. However, comparatively less attention has been given to how practitioners should handle (pre-)trends, despite its critical importance in applied research. This paper addresses that gap by evaluating the effectiveness of classical strategies commonly used in the literature. We conducted Monte Carlo simulations to assess the performance of the three most common approaches: (i) incorporating pre-trends as a covariate (trend-as-control); (ii) extrapolating linear trends; and (ii) controlling for group-by-time FEs (e.g., state-by-year FEs). We also formalized the assumptions motivating their application and provided intuitive insights into their behavior based on simulation and analytical results.

Our main finding is that trend-as control is unlikely to be a reliable strategy. In settings

---

[27]This is a necessary, but not sufficient, condition for bias introduction. See Appendix A.4.1 for a detailed explanation.

[28]We highlight that if a significant portion of treatment variation occurs between groups, incorporating specific time fixed effects may inflate the variance of the estimates. Consequently, if the relative increase in variance outweighs the bias reduction, this could result in a larger RMSE. In the extreme case where all variation arises at the group level (e.g., constant treatment allocation within groups or states), this approach becomes infeasible. In such a scenario, group-by-time FEs would absorb all the variation in the data, leaving no scope for identifying treatment effects.

where the serial correlation of outcome growth tends to be weak on the unit level, this approach fails to eliminate bias driven by differential trends between treatment and control groups. Importantly, this method may – automatically – mitigate bias in the pre-treatment period, while yielding post-treatment estimates that are similar to unconditional (and biased) baseline estimates. As a consequence, pre-testing conditional on this approach may mislead researchers into believing that the parallel trends assumption is satisfied. This issue generalizes across several specifications, in particular to matching on pre-trends.

To evaluate whether the trend-as-control method is appropriate in a given context, we propose a straightforward exercise: when strong pre-trends are observed in event study estimates, researchers should assess the persistence of outcome growth at the unit level using pre-treatment data. If weak serial correlation is detected, it suggests that the trend-as-control strategy is unlikely to correct for differential trends in the post-treatment period.

If the trend-as-control approach is expected to fail but differential trends are persistent at the treatment level, linear trend extrapolation may be a more effective alternative for reducing bias. However, this comes at the cost of increased variance and risk of over-correction. Since differential trends often diminish over time due to regression to the mean, linear extrapolation may result in estimates biased in the opposite direction. Importantly, researchers need not be worried about the choice between treatment- and unit-specific trends, since they yield very similar estimates and exhibit comparable performance, as long as there is no treatment variation within the unit-level.

Controlling for group-by-time fixed effects can address differential trends driven by confounding factors at the group level but does not resolve deviations from parallel trends that occur within those groups. In our simulations, although this approach performed better than the unconditional baseline and trend-as-control, it achieved only modest bias reduction. One takeaway from our results is that linear trend extrapolation may be a better option when pre-trends are strong and bias is the primary concern. Conversely, group-by-time fixed effects may be more appropriate when underlying trends are weaker (or highly non-linear), and when precision is the main priority.

In this paper, we addressed the issue of incidental trends that may arise from correlations

between treatment and pre-determined characteristics of the exposed units. When these characteristics are unobserved by the researcher, a key question is whether observed pre-treatment growth (i.e., pre-trends) provides sufficient information to recover the confounding factors driving differential trends at the treatment level. This concern may also extend to related methods that rely on pre-treatment data, such as synthetic control. Further research is needed to explore the distinctions between difference-in-differences conditional on "pre-trends" and the synthetic control approach.

References

A. Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of economic literature*, 59(2):391–425, 2021.

A. Abadie and J. Spiess. Robust post-matching inference. *Journal of the American Statistical Association*, 117(538):983–995, 2022.

A. Abadie and J. Vives-i Bastida. Synthetic Controls in Action. *arXiv preprint n.2203.06279*, 2022.

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35, 2023.

L. Agha, S. Kim, and D. Li. Insurance design and pharmaceutical innovation. *American Economic Review: Insights*, 4(2):191–208, 2022.

R. C. Allen, M. C. Bertazzini, and L. Heldring. The economic origins of government. *American Economic Review*, 113(10):2507–2545, 2023.

D. Ang. The birth of a nation: Media and racial hate. *American Economic Review*, 113(6): 1424–1460, 2023.

J. D. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, 2009.

D. Arkhangelsky and D. Hirshberg. Large-sample properties of the synthetic control method under selection on unobservables. *arXiv preprint n. 2311.13575*, 2, 2023.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021.

A. Bacher-Hicks, J. Goodman, J. G. Green, and M. K. Holt. The COVID-19 pandemic disrupted both school bullying and cyberbullying. *American Economic Review: Insights*, 4(3):353–370, 2022.

A. Baker, B. Callaway, S. Cunningham, A. Goodman-Bacon, and P. H. Sant'Anna. Difference-in-differences designs: A practitioner's guide. *arXiv preprint arXiv:2503.13323*, 2025.

P. J. Barwick, Y. Liu, E. Patacchini, and Q. Wu. Information, mobile communication, and referral effects. *American Economic Review*, 113(5):1170–1207, 2023.

A. Bertheau, E. M. Acabbi, C. Barceló, A. Gulyas, S. Lombardi, and R. Saggio. The unequal consequences of job loss across countries. *American Economic Review: Insights*, 5(3):393–408, 2023.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.

M. C. Best, J. Hjort, and D. Szakonyi. Individuals and organizations as sources of state effectiveness. *American Economic Review*, 113(8):2121–2167, 2023.

S. E. Black, J. T. Denning, L. J. Dettling, S. Goodman, and L. J. Turner. Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well-being. *American Economic Review*, 113(12):3357–3400, 2023.

O. J. Blanchard, L. F. Katz, R. E. Hall, and B. Eichengreen. Regional evolutions. *Brookings papers on economic activity*, 1992(1):1–75, 1992.

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

L. Braghieri, R. Levy, and A. Makarin. Social media and mental health. *American Economic Review*, 112(11):3660–3693, 2022.

Z. Brot-Goldberg, T. Layton, B. Vabson, and A. Y. Wang. The behavioral foundations of default effects: theory and evidence from Medicare Part D. *American Economic Review*, 113(10):2718–2758, 2023.

N. L. Brown and K. Butts. Dynamic treatment effect estimation with interactive fixed effects and short panels. *Journal of Econometrics*, 250:106013, 2025.

R. A. Butters, D. W. Sacks, and B. Seo. How do national firms respond to local cost shocks? *American Economic Review*, 112(5):1737–1772, 2022.

M. Cabral, C. Cui, and M. Dworsky. The demand for insurance and rationale for a mandate: Evidence from workers' compensation insurance. *American Economic Review*, 112(5): 1621–1668, 2022.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 2021.

E. Cantoni and V. Pons. Does context outweigh individual characteristics in driving voting behavior? Evidence from relocations within the United States. *American Economic Review*, 112(4):1226–1272, 2022.

Y. Cao and S. Chen. Rebel on the canal: Disrupted trade access and social conflict in China, 1650–1911. *American Economic Review*, 112(5):1555–1590, 2022.

S. Chabé-Ferret. Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1):110–123, 2015.

S. Chabé-Ferret. Should we combine difference in differences with conditioning on pre-treatment outcomes? 2017.

S. Cicala. Imperfect markets versus imperfect regulation in US electricity generation. *American Economic Review*, 112(2):409–441, 2022.

Z. Cullen and R. Perez-Truglia. The old boys' club: Schmoozing and the gender gap. *American Economic Review*, 113(7):1703–1740, 2023.

J. Currie, H. Kleven, and E. Zwiers. Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings*, 110:42–48, 2020.

J. Currie, J. Voorheis, and R. Walker. What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality. *American Economic Review*, 113(1):71–97, 2023.

G. B. Dahl, R. Lu, and W. Mullins. Partisan fertility and presidential elections. *American Economic Review: Insights*, 4(4):473–490, 2022.

J. R. Daw and L. A. Hatfield. Matching and regression to the mean in difference-in-differences analysis. *Health services research*, 53(6):4138–4156, 2018.

C. De Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

M. Dillender. Evidence and lessons on the health impacts of public health funding from the fight against HIV/AIDS. *American Economic Review*, 113(7):1825–1887, 2023.

C. Duquennois. Fictional money, real costs: Impacts of financial salience on disadvantaged students. *American Economic Review*, 112(3):798–826, 2022.

C. N. East, S. Miller, M. Page, and L. R. Wherry. Multigenerational impacts of childhood access to the safety net: Early life exposure to Medicaid and the next generation's health. *American Economic Review*, 113(1):98–135, 2023.

E. Esposito, T. Rotesi, A. Saia, and M. Thoenig. Reconciliation narratives: The birth of a nation after the US Civil War. *American Economic Review*, 113(6):1461–1504, 2023.

D. Fackler, J. Stegmaier, and E. Weigt. Does extended unemployment benefit duration ameliorate the negative employment effects of job loss? *Labour Economics*, 59:123–138, 2019.

B. Ferman and C. Pinto. Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221, 2021.

S. Freyaldenhoven, C. Hansen, and J. M. Shapiro. Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–3338, 2019.

S. Freyaldenhoven, C. Hansen, J. P. Pérez, and J. M. Shapiro. Visualization, identification, and estimation in the linear panel event-study design. Technical report, National Bureau of Economic Research, 2021.

D. Ghanem, P. H. Sant'Anna, and K. Wüthrich. Selection and Parallel Trends. *arXiv preprint n. 2203.09001*, 2022.

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

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

M. Greenstone, G. He, R. Jia, and T. Liu. Can technology solve the principal-agent problem? Evidence from China's war on air pollution. *American Economic Review: Insights*, 4(1): 54–70, 2022.

T. Gross, T. J. Layton, and D. Prinz. The liquidity sensitivity of healthcare consumption: Evidence from social security payments. *American Economic Review: Insights*, 4(2): 175–190, 2022.

D. W. Ham and L. Miratrix. Benefits and costs of matching prior to a difference in differences analysis when parallel trends does not hold. *The Annals of Applied Statistics*, 18(3):2096–2122, 2024.

C. W. Hansen and A. M. Wingender. National and global impacts of genetically modified crops. *American Economic Review: Insights*, 5(2):224–240, 2023.

A. Janssen and X. Zhang. Retail pharmacies and drug diversion during the opioid epidemic. *American Economic Review*, 113(1):1–33, 2023.

A. Kaul, S. Klößner, G. Pfeifer, and M. Schieler. Standard synthetic control methods: The case of using all preintervention outcomes together with covariates. *Journal of Business & Economic Statistics*, 40(3):1362–1376, 2022.

D. L. Miller. An introductory guide to event study models. *Journal of Economic Perspectives*, 37(2):203–230, 2023.

T. Porzio, F. Rossi, and G. Santangelo. The human side of structural transformation. *American Economic Review*, 112(8):2774–2814, 2022.

A. Rambachan and J. Roth. A more credible approach to parallel trends. *The Review of Economic Studies*, 2023.

J. Roth. Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322, 2022.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

P. H. Sant'Anna and J. Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020.

K. Schmidheiny and S. Siegloch. On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Journal of Applied Econometrics*, 38(5):695–713, 2023.

J. F. Schmieder, T. Von Wachter, and J. Heining. The costs of job displacement over the business cycle and its sources: evidence from Germany. *American Economic Review*, 113 (5):1208–1254, 2023.

M. Smith, D. Yagan, O. Zidar, and E. Zwick. The rise of pass-throughs and the decline of the labor share. *American Economic Review: Insights*, 4(3):323–340, 2022.

P. Sodini, S. Van Nieuwerburgh, R. Vestman, and U. von Lilienfeld-Toal. Identifying the benefits from homeownership: A Swedish experiment. *American Economic Review*, 113 (12):3173–3212, 2023.

G. Solon, S. J. Haider, and J. M. Wooldridge. What are we weighting for? *Journal of Human Resources*, 50(2):301–316, 2015.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2021.

M. M. Sviatschi. Spreading gangs: Exporting US criminal capital to El Salvador. *American Economic Review*, 112(6):1985–2024, 2022.

J. Wooldridge. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *Available at SSRN 3906345*, 2021.

A   Proofs

A.1   Baseline: Unconditional DID

Let the observed outcome be defined in terms of potential outcomes:

$$y_{i,t} = y_{i,t}^0 + (y_{i,t}^1 - y_{i,t}^0)D_i$$

where $y_{i,t}^0$ is the potential outcome in the untreated state and $y_{i,t}^1$ is the potential outcome in the treated state. We introduce a common assumption in difference-in-differences designs, and a necessary condition for the use of pre-treatment outcomes to construct counterfactuals:

Assumption A.1. *(**No Anticipation**)* $\dot{y}_{i,s}^0 = \dot{y}_{i,s}$ *for all $s \leq 0$*

The *No Anticipation* assumption requires that the treatment has no effect in the pre-treatment period (i.e., $\beta_t = 0$ for all $t \leq 0$ in our specifications).

Proof. **Identification of the β_t parameters in the baseline specification (S1)**
In the baseline specification (S1), the β_t coefficients are identified as the difference-in-differences (DID) estimands for each $t > 0$, as follows:

$$\beta_t \equiv DID_t = \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 1] - \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 0]$$

Based on Assumption A.1, we can write this expression in terms of potential outcomes:

$$\beta_t = \mathbb{E}[y_{i,t}^1 - y_{i,0}^0|D_i = 1] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 0] \ \ \forall t > 0$$

Summing zero in the right-hand-side and reordering the terms:

$$\beta_t = \underbrace{\mathbb{E}[y_{i,t}^1 - y_{i,t}^0|D_i = 1]}_{\mathbf{ATT}_t} + \left\{\underbrace{\mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 1] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 0]}_{\Delta_t}\right\} \ \ \forall t > 0 \quad \text{(A.1)}$$

where Δ_t is the bias from the violation of the Parallel Trends assumption.[29] $\qquad\square$

---

[29]The Parallel Trends assumption can be stated as follows:

$$\Delta_t := \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 1] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 0] = 0 \ \ \forall t > 0$$

Equation (A.1) implies that, under the assumptions of *No Anticipation* and *Parallel Trends*, the β_t coefficients in the baseline specification (S1) identify the ATT_t, regardless of the functional form of the potential outcomes. However, in order to understand how different methods can address deviations from the Parallel Trends assumption, we adopt a linear factor model for the potential outcomes, as defined in equation (3). We then obtain the (demeaned) potential outcome growth as follows:

$$\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0 = \lambda_t \dot{\alpha}_i + \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0} \tag{A.2}$$

where $\dot{y}_{i,t}^0 := y_{i,t}^0 - \mathbb{E}(y_{i,t}^0)$, $\lambda_t := (f_t - f_0)$, f_t is a $(1 \times F)$ vector of unobservable common factors, α_i is a $(F \times 1)$ vector of factor loadings (i.e., unit-specific exposure to common shocks), and $\epsilon_{i,t}$ is an idiosyncratic shock. Therefore, we obtain the observed outcome growth for each period t with respect to the baseline as follows:

$$\Delta \dot{y}_{i,t} = \beta_t^0 D_i + \underbrace{\lambda_t \dot{\alpha}_i + \Delta \dot{\varepsilon}_{i,t}}_{\Delta \dot{y}_{i,t}^0} \tag{A.3}$$

where $\Delta \dot{y}_{i,t} := \dot{y}_{i,t} - \dot{y}_{i,0}$, $\Delta \dot{\varepsilon}_{i,t} := \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}$, $\lambda_t := f_t - f_0$, and $D_i = 1$ if unit i if treated, zero otherwise.[30]

We further introduce a set of standard assumptions regarding the data-generating process, focusing in particular on selection into treatment as relevant to our analysis. We adopt a sampling assumption commonly used in the literature on interactive fixed effects and factor models (see, for example, Brown and Butts (2025)). Specifically, we treat the common time-varying factors (f_t) as fixed, while the heterogeneity in unit-specific factor loadings (α_i) and idiosyncratic errors ($\varepsilon_{i,t}$) is considered as stochastic and independently distributed across units. This allows us to focus on these sources of variation as the primary drivers of sampling uncertainty.

Assumption A.2. *(**Sampling**) The vector of the common factors (f_t) is considered as*

---

[30]Note that β_t^0 is not, a priori, the same at the β_t parameter in the baseline specification (S1). While β_t^0 represents the treatment effect, β_t identifies the average treatment effect on the treated (ATT_t) only under the parallel assumption, as shown above.

fixed. Unknown factor loadings (α_i) and the idiosyncratic shocks ($\varepsilon_{i,t}$) are stochastic and i.i.d. across i.

Finally, in our illustration, we focus on the selection of unobservables that arises from the correlation between treatment assignment (D_i) and unit-specific exposure to common shocks (α_i). We assume that, conditional on the factor loadings, treatment assignment is independent of the idiosyncratic shocks.

Assumption A.3. *(**Properties of the idiosyncratic shock**)*

$$\mathbb{E}[\varepsilon_{i,t}|D_i, \alpha_i] = \mathbb{E}[\varepsilon_{i,t}|\alpha_i] \ \ and \ \mathbb{E}[\varepsilon_{i,t}|\alpha_i] = 0$$

Further, for simplicity, we assume that the error term is uncorrelated over time, and $Var(\varepsilon_{i,t}) = \sigma^2$ for all i and t.[31]

Proposition A.1. *(**Unconditional Bias**) Under Assumptions A.1-A.3:*

$$\Delta_t = \lambda_t \left(\mathbf{E}[\alpha_i|D_i = 1] - \mathbf{E}[\alpha_i|D_i = 0] \right)$$

where $\Delta_t := \mathbf{E}[\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0|D_i = 1] - \mathbf{E}[\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0|D_i = 0]$ represent the bias arising from deviations from the parallel trends assumption.

*Further, considering that the outcome is defined as in equation (A.3), the **asymptotic bias** ($\tilde{\Delta}_t$) from a regression that omits the factor structure is the following:*

$$\tilde{\Delta}_t := plim \ \hat{\beta}_t^0 - \beta_t^0 = \lambda_t \frac{Cov(\alpha_i, D_i)}{Var(D_i)} \tag{A.4}$$

Proof. Let the unconditional bias (Δ_t) be defined in terms of demeaned potential outcomes as follows:

$$\Delta_t = \mathbb{E}[\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0|D_i = 1] - \mathbb{E}[\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0|D_i = 0]$$

---

[31]Assuming homoskedasticity and no serial correlation in the idiosyncratic shocks simplifies the exposition of the attenuation bias associated with the trend-as-control approach, but these assumptions are not essential to our main argument. We provide an extension introducing serial correlation in Appendix A.2.2.

We can derive Δ_t by taking the conditional expectations of equation (A.2). Assumption A.3 implies that $\dot{\varepsilon}_{i,t} = \varepsilon_{i,t}$. In addition, from the Law of Iterated Expectations (L.I.E) and Assumption A.3, $\mathbf{E}[\varepsilon_{i,t} - \varepsilon_{i,0}|D_i] = \mathbf{E}\left[\mathbf{E}[\varepsilon_{i,t} - \varepsilon_{i,0}|D_i, \alpha_i]|D_i\right] = 0$. In addition, from Assumption A.2 and given that $\mathbf{E}[\dot{\alpha}_i|D_i] = \mathbf{E}[\alpha_i|D_i]$, we obtain the unconditional bias as follows:

$$\Delta_t = \lambda_t \left(\mathbf{E}[\alpha_i|D_i = 1] - \mathbf{E}[\alpha_i|D_i = 0]\right)$$

In an unconditional regression, one would estimate β_t^0 from equation (A.3) without accounting for the factor structure, i.e., using the specification $\Delta \ddot{y}_{it} = \beta_t^0 D_i + \nu_{i,t}$, where the error term is $\nu_{i,t} = \lambda_t \dot{\alpha}_i + \Delta \dot{\varepsilon}_{i,t}$.

$$plim\ \hat{\beta}_t^0 = \frac{Cov(\Delta \ddot{y}_{i,t}, D_i)}{Var(D_i)} = \beta_t^0 + \frac{Cov(\nu_{i,t}, D_i)}{Var(D_i)}$$

Let $\tilde{\Delta}_t := plim\ \hat{\beta}_t^0 - \beta_t^0$ be the asymptotic bias. Then:

$$\tilde{\Delta}_t = \frac{Cov(\lambda_t \dot{\alpha}_i, D_i)}{Var(D_i)} + \frac{Cov(\Delta \dot{\epsilon}_{i,t}, D_i)}{Var(D_i)}$$

From the L.I.E and Assumption A.3, $Cov(\Delta \dot{\epsilon}_{i,t}, D_i) = 0$. Further, from Assumptions A.2 and since $Cov(\dot{\alpha}_i, D_i) = Cov(\alpha_i, D_i)$, we obtain the asymptotic bias as follows:

$$\tilde{\Delta}_t = \lambda_t \frac{Cov(\alpha_i, D_i)}{Var(D_i)}$$

$\square$

A.2 Conditional DID

When controlling for a pre-determined variable X_i, the DID estimand consists of the average difference in the outcome trajectory conditional on X_i:[32]

$$DID_t(X_i) = \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 1, X_i] - \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 0, X_i]$$

As in the unconditional case, we can write this expression in terms of potential outcomes, which implies that $DID_t(X_i)$ can be decomposed as follows:

$$DID_t(X_i) = ATT_t(X_i) + \Delta_t(X_i)$$

where $\Delta_t(X_i) := \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 1, X_i] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 0, X_i]$ is the bias from the violation of the *Conditional Parallel Trends* assumption, which is a function of the covariate X_i. Therefore, to obtain the aggregate bias, we must take the expectation of this function over the distribution of X_i among the treated:

$$\Delta_t(\boldsymbol{X}) := \mathbb{E}[\Delta_t(X_i)|D_i = 1] \tag{A.5}$$

Further, $ATT_t(X_i) := \mathbb{E}[y_{i,t}^1 - y_{i,t}^0|D_i = 1, X_i]$ is the average treatment effect on the treated for each value of X_i. From the L.I.E, the ATT_t is the average of conditional effects $ATT_t(X_i)$ on the treated group:

$$ATT_t = \mathbb{E}[ATT_t(X_i)|D_i = 1] \tag{A.6}$$

A.2.1 Trend-as-control

In the trend-as-control approach, the covariate X_i is the pre-treatment growth of the outcome. Let, for instance, the pre-trend over k pre-treatment periods be defined as $PreTrend_{i,k} := \dot{y}_{i,0} - \dot{y}_{i,k}$. Under no anticipation, we can write it using our factor model as follows:

$$PreTrend_{i,k} = \lambda_{-k}\dot{\alpha}_i + \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-(k)} \tag{A.7}$$

---

[32]Differently from the unconditional specification, the equivalence between parameters β_t in specification (S2) and the DID_t estimand depends on functional form assumptions. We consider different specifications in Appendix C.

where $\lambda_{-k} := f_0 - f_{-k}$.

To reconcile the trend-as-control approach within a model featuring interactive fixed effects, we impose additional structure by assuming that the common shocks follow a deterministic trend over time. Further, the presence of multiple factors can be accommodated under the assumption that they evolve at the same rate, as formalized below:

Assumption A.4. *(Deterministic trend assumption) The common shocks follow a deterministic trend with persistence parameters $\rho_{t,k}$ over t post-treatment and k pre-treatment periods, which are identical across all factors:*

$$\lambda_t = \rho_{t,k}\lambda_{-k}$$

Under Assumption A.4, we can rewrite equation (A.3) in terms of observed pre-trends as follows:

$$\Delta\dot{y}_{i,t} = \beta_t^0 D_i + \rho_{t,k} PreTrend_{i,k} + \underbrace{\left(\Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-(k)})\right)}_{\text{error term}} \tag{A.8}$$

where $\rho_{t,k}$ is the parameter that measures persistence in deterministic common shocks. Note that $PreTrend_{i,k}$ is a noisy proxy for the systematic component driving the differential trend shown in equation (A.4). The reliability of $PreTrend_{i,k}$ as a proxy of the structure of the unobserved factors can be assessed by the proportion of its variance explained by the variation in the systematic component (that is, the signal). Under the assumptions on sampling and the properties of the error terms, the *signal to total variance ratio* is:

$$\zeta := \frac{Var(\lambda_{-k}\dot{\alpha}_i)}{Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2} \tag{A.9}$$

Proposition A.2. *(Attenuation Bias in Trend-as-control)*
Under Assumptions A.1-A.3, when estimating $\rho_{t,k}$ from equation (A.8), we obtain the following expression:

$$plim\ \hat{\rho}_{t,k} = \rho_{t,k}\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right) + \frac{Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})} \tag{A.10}$$

Further, if we let $PreTrend_{i,k} := \dot{y}_{i,-1} - \dot{y}_{i,-(k+1)}$, and $\lambda_{-k} := f_{-1} - f_{-(k+1)}$, then $Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)}) = 0$ and can represent the conditional bias as follows:

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k}\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right) \tag{A.11}$$

where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i, and $\rho_{t,k}$ is the parameter measuring the persistence in the deterministic common shocks. Since $\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right) \in (0,1)$, the trend-as-control coefficient is biased towards zero, resulting from the fact that pre-trend is a noisy proxy for the factor structure.

Proof.

$$plim \ \hat{\rho}_{t,k} = \frac{Var(D_i)Cov(\Delta\dot{y}_{i,t}, PreTrend_{i,k}) - Cov(D_i, PreTrend_{i,k})Cov(D_i, \Delta\dot{y}_{i,t})}{Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2}$$

where

$$Cov(\Delta\dot{y}_{i,t}, PreTrend_{i,k}) = \beta_t^0 Cov(D_i, PreTrend_{i,k}) + \rho_{t,k}Var(PreTrend_{i,k})$$
$$+ Cov(PreTrend_{i,k}, \Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k}))$$
$$\text{and } Cov(\Delta\dot{y}_{i,t}, D_i) = \beta_t^0 Var(D_i) + \rho_{t,k}Cov(D_i, PreTrend_{i,k})$$

Since the terms multiplying β_t^0 cancel-out:

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k} - Var(D_i)\frac{\rho_{t,k}Var(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k}) - Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2}$$

Note that $Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2 = (1-\tilde{R}^2) \times Var(PreTrend_{i,k})Var(D_i)$, where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i.

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k} - \rho_{t,k}\frac{Var(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})} + \frac{Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})}$$

where $Var(PreTrend_{i,k}) = Var(\lambda_k\dot{\alpha}_i) + Var(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})$ and, under Assumption A.3, $Var(\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k}) = 2\sigma^2$. Let $\zeta := \frac{Var(\lambda_k\dot{\alpha}_i)}{Var(\lambda_k\dot{\alpha}_i)+2\sigma^2}$ be the signal to total variance ratio, thus

$\frac{Var(\dot{\varepsilon}_{i,0}-\dot{\varepsilon}_{i,-k})}{Var(PreTrend_{i,k})} = 1 - \zeta$. Then we can replace the first term in the right-hand-side to obtain the bias expression from equation (A.10), reproduced below:

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k}\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right) + \frac{Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})}$$

To simplify the exposition of the attenuation bias, in our main derivations, we defined $PreTrend_{i,k}$ over k pre-treatment period using $t = 1$ as the baseline instead of $t = 0$. This choice avoids having $\dot{\varepsilon}_{i,0}$ appear in both the pre-trend variable and the post-treatment idiosyncratic shock, which would otherwise introduce a mechanical correlation between the "measurement error" in the pre-trend and the error term in outcome growth. Therefore, let $PreTrend_{i,k} := \dot{y}_{i,-1} - \dot{y}_{i,-(k+1)}$, and $\lambda_{-k} := f_{-1} - f_{-(k+1)}$, then $Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)}) = 0$ we obtain the expression in equation (A.11), reproduced below:

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k}\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right)$$

where and $\zeta := \frac{Var(\lambda_{-k}\dot{\alpha}_i)}{Var(\lambda_{-k}\dot{\alpha}_i) + Var(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)})} \equiv \frac{Var(\lambda_{-k}\dot{\alpha}_i)}{Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2}$.

Note that $\tilde{R}^2 = \frac{Cov(D_i, PreTrend_{i,k})^2}{Var(PreTrend_{i,k})Var(D_i)} = \frac{Cov(D_i, \lambda_{-k}\dot{\alpha}_i)^2}{(Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2)Var(D_i)} < \frac{Var(D_i)Var(\lambda_{-k}\dot{\alpha}_i)}{(Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2)Var(D_i)} \equiv \zeta$

$$\tilde{R}^2 < \zeta, \zeta \in (0,1) \text{ and } \tilde{R}^2 \in (0,1) \implies \left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2}\right) \in (0,1)$$

$\square$

Proposition A.3. *(Conditional Bias)*

Under Assumptions A.1-A.4, when controlling for $PreTrend_{i,k} := \dot{y}_{i,-1} - \dot{y}_{i,-(k+1)}$, the asymptotic bias of the DID_t estimator is proportional to the asymptotic bias in the persistence parameter $\rho_{t,k}$:

$$\tilde{\Delta}_t(\boldsymbol{PreTrend}_k) := plim \ \hat{\beta}_t^0 - \beta_t^0 = -\frac{\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)}\left(plim \ \hat{\rho}_{t,k} - \rho_{t,k}\right) \qquad (A.12)$$

This implies the following relation between the unconditional and conditional asymptotic

bias expressions:

$$\tilde{\Delta}_t(\boldsymbol{PreTrend}_k) = \tilde{\Delta}_t \left(\frac{1 - \zeta}{1 - \tilde{R}^2} \right) \tag{A.13}$$

where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i, ζ is the signal to total variance ratio, $1 - \zeta$ is the noise to total variance ratio, and $\left(\frac{1-\zeta}{1-\tilde{R}^2} \right) \in (0,1)$.

Trend-as-control will partially eliminate bias from the unobserved factor structure by a proportion $\left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2} \right) \in (0,1)$.

Proof. When controlling for $PreTrend_{i,k}$ measured from $t = -(k+1)$ to $t = -1$ to estimate β_t^0 in equation (A.8) we get the following:

$$plim \ \hat{\beta}_t^0 = \frac{Var(PreTrend_{i,k})Cov(\Delta\dot{y}_{i,t}, D_i) - Cov(D_i, PreTrend_{i,k})Cov(PreTrend_{i,k}, \Delta\dot{y}_{i,t})}{Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2}$$

where $Cov(\Delta\dot{y}_{i,t}, D_i) = \beta_t^0 Var(D_i) + \rho_{t,k} Cov(D_i, PreTrend_{i,k})$

and $Cov(\Delta\dot{y}_{i,t}, PreTrend_{i,k}) = \beta_t^0 Cov(D_i, PreTrend_{i,k}) + \rho_{t,k} Var(PreTrend_{i,k})$

$$+ Cov(PreTrend_{i,k}, \Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)}))$$

$$plim \ \hat{\beta}_t^0 = \beta_t^0 - \frac{Cov(D_i, PreTrend_{i,k})Cov(PreTrend_{i,k}, \Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)}))}{Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2}$$

where, under Assumption A.3:

$$Cov(PreTrend_{i,k}, \Delta\dot{\varepsilon}_{i,t} - \rho_{t,k}(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)})) = -\rho_{t,k}Var(\dot{\varepsilon}_{i,-1} - \dot{\varepsilon}_{i,-(k+1)})) = -\rho_{t,k}2\sigma^2$$

Let $\tilde{\Delta}_t(\boldsymbol{PreTrend}_k) := plim \ \hat{\beta}_t^0 - \beta_t^0$.

Further, $Var(PreTrend_{i,k})Var(D_i) - Cov(D_i, PreTrend_{i,k})^2 = (1 - \tilde{R}^2) \times Var(PreTrend_{i,k})Var(D_i)$, where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i. Then:

$$\tilde{\Delta}_t(\boldsymbol{PreTrend}_k) = \rho_{t,k} \frac{2\sigma^2 Cov(D_i, PreTrend_{i,k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})Var(D_i)} = \rho_{t,k} \left(\frac{1-\zeta}{1-\tilde{R}^2} \right) \frac{Cov(D_i, PreTrend_{i,k})}{Var(D_i)}$$

Under Assumptions A.2 and A.3, $Cov(D_i, PreTrend_{i,k}) = \lambda_{-k}Cov(\alpha_i, D_i)$. Further, as shown above:

$$plim \ \hat{\rho}_{t,k} = \rho_{t,k} \left(\frac{\zeta - \tilde{R}^2}{1 - \tilde{R}^2} \right) \implies plim \ \hat{\rho}_{t,k} - \rho_{t,k} = -\rho_{t,k} \left(\frac{1-\zeta}{1-\tilde{R}^2} \right)$$

Therefore, we obtain the expression of the conditional bias as a function of the attenuation bias in the trend-as-control parameter shown in equation (A.12):

$$\tilde{\Delta}_t(\textbf{PreTrend}_k) = -\left(plim\ \hat{\rho}_{t,k} - \rho_{t,k}\right)\frac{\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)}$$

The asymptotic unconditional bias (Appendix A.1) can be written as follows:

$$\tilde{\Delta}_t = \rho_{t,k}\frac{\lambda_{-k}Cov(\alpha_i,D_i)}{Var(D_i)}$$

This implies the relation between the unconditional and conditional asymptotic bias expressions shown in equation (A.13) and reproduced below:

$$\tilde{\Delta}_t(\textbf{PreTrend}_k) = \tilde{\Delta}_t\left(\frac{1-\zeta}{1-\tilde{R}^2}\right)$$

where $\tilde{R}^2$ is the R^2 of a regression of $PreTrend_{i,k}$ on D_i, ζ is the signal to total variance ratio, $1-\zeta$ is the noise to total variance ratio, and $\left(\frac{1-\zeta}{1-\tilde{R}^2}\right) \in (0,1)$.

$$\frac{\tilde{\Delta}_t - \tilde{\Delta}_t(\textbf{PreTrend}_k)}{\tilde{\Delta}_t} = \left(\frac{\zeta - \tilde{R}^2}{1-\tilde{R}^2}\right)$$

Therefore, trend-as-control will partially eliminate bias from the unobserved factor structure by a proportion $\left(\frac{\zeta-\tilde{R}^2}{1-\tilde{R}^2}\right) \in (0,1)$. $\qquad\square$

Proof. **Diagnostic 2**

Let $\Delta y_{i,-k} = y_{i,0} - y_{i,-k}$ and $\Delta y_{i,-s} = y_{i,-(k+1)} - y_{i,-(k+1+s)}$, and consider a regression of pre-trends on lags as follows:

$$\Delta y_{i,-k} = \pi_{k,s}\Delta y_{i,-s} + \nu_{i,-k}$$

Based on the linear factor model and Assumption A.1, we have the following:

$$\Delta \dot{y}_{i,-k} = \lambda_{-k} \dot{\alpha}_i + (\dot{\epsilon}_{i,0} - \dot{\epsilon}_{i,-k})$$

$$\Delta \dot{y}_{i,-s} = \lambda_{-s} \dot{\alpha}_i + (\dot{\epsilon}_{i,-(k+1)} - \dot{\epsilon}_{i,-(k+1+s)})$$

Assumption A.4 implies $\lambda_{-k} = \rho_{ks} \lambda_{-s}$. Therefore:

$$plim \ \hat{\pi}_{k,s} = \frac{Cov(\Delta y_{i,-k}, \Delta y_{i,-s})}{Var(\Delta y_{i,-s})} = \rho_{ks} \frac{Var(\lambda_{-s} \alpha_i)}{Var(\Delta y_{i,-s})} \equiv \rho_{ks} \times \zeta_{(s)}$$

This implies that the *signal to total variance ratio* can be expressed as follows:

$$\zeta_{(s)} = \frac{plim \ \hat{\pi}_{k,s}}{\rho_{ks}} \tag{A.14}$$

Also note that the unconditional bias in $t = \{-k, -s\}$ due to a regression of pre-trends on D_i, omitting the factor structure (e.g., $\Delta \dot{y}_{i,-k} = \beta_{-k} D_i + \nu_{i,-k}$ where $\beta_{-k} = \beta_{-s} = 0$) is given by $\tilde{\Delta}_{-k} = \lambda_{-k} \frac{Cov(\alpha_i, D_i)}{Var(D_i)}$ and $\tilde{\Delta}_{-s} = \lambda_{-s} \frac{Cov(\alpha_i, D_i)}{Var(D_i)}$ respectively.

Finally, assume that we employ trend-as-control to reduce the unconditional bias by controlling for $PreTrend_{i,s} := \Delta \dot{y}_{i,-s}$. Therefore, the true model is:

$$\Delta \dot{y}_{i,-k} = \rho_{ks} PreTrend_{i,s} + (\dot{\epsilon}_{i,0} - \dot{\epsilon}_{i,-k}) - \rho_{ks}(\dot{\epsilon}_{i,-(k+1)} - \dot{\epsilon}_{i,-(k+1+s)})$$

But we regress: $\Delta \dot{y}_{i,-k} = \tilde{\beta}_{-k} D_i + \rho_{ks} PreTrend_{i,s} + \nu_{i,-(k,s)}$, where $\tilde{\beta}_{-k} = 0$ and $\nu_{i,-(k,s)} = (\dot{\epsilon}_{i,0} - \dot{\epsilon}_{i,-k}) - \rho_{ks}(\dot{\epsilon}_{i,-(k+1)} - \dot{\epsilon}_{i,-(k+1+s)})$. Following the same steps as before, the conditional bias can be expressed as shown below:

$$\tilde{\Delta}_{-k}(\mathbf{PreTrend}_s) = \left(\frac{1 - \zeta_{(s)}}{1 - \tilde{R}_s^2} \right) \tilde{\Delta}_{-k}$$

where $\tilde{R}_s^2$ is the R^2 of a regression of $PreTrend_{i,s}$ on D_i.

Reordering the terms:

$$\tilde{\Delta}_{-k}(\mathbf{PreTrend}_s) \times (1 - \tilde{R}_s^2) = (1 - \zeta_{(s)}) \times \tilde{\Delta}_{-k}$$

$$\tilde{\Delta}_{-k}(\mathbf{PreTrend}_s) - \tilde{R}_s^2 \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s) = \tilde{\Delta}_{-k} - \zeta_{(s)} \tilde{\Delta}_{-k}$$

$$\left(\tilde{\Delta}_{-k} - \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s) = \zeta_{(s)} \tilde{\Delta}_{-k} - \tilde{R}_s^2 \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)\right) \div \tilde{\Delta}_{-k}$$

$$\frac{\tilde{\Delta}_{-k} - \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} = \zeta_{(s)} - \tilde{R}_s^2 \frac{\tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}}$$

Therefore, replacing the expression from equation (A.14):

$$\frac{\tilde{\Delta}_{-k} - \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} = \frac{plim \; \hat{\pi}_{k,s}}{\rho_{ks}} - \tilde{R}_s^2 \frac{\tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} \qquad (A.15)$$

Further, under Assumption A.4, we obtain $\rho_{ks} = \frac{\tilde{\Delta}_{-k}}{\tilde{\Delta}_{-s}}$, which implies the following expression:

$$\frac{\tilde{\Delta}_{-k} - \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}} = \frac{plim \; \hat{\pi}_{k,s} \tilde{\Delta}_{-s} - \tilde{R}_s^2 \tilde{\Delta}_{-k}(\mathbf{PreTrend}_s)}{\tilde{\Delta}_{-k}}$$

$\square$

A.2.2 Trend-as-control with serial correlation in idiosyncratic shocks

In the previous section, we simplified our illustration by measuring pre-trends using $t = -1$ as the baseline rather than $t = 0$. If we instead use the same baseline as the post-treatment estimates, we must account for the correlation of the idiosyncratic errors in equation (A.10). Under the assumption of no serial correlation in shocks (Assumption A.3), we have that $Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k}) = -Var(\dot{\varepsilon}_{i,0}) = -\sigma^2$. To generalize the analysis, we next relax this assumption and allow the shocks to be serially correlated, as detailed below:

Assumption A.5. *(Properties of the idiosyncratic shock)*

$$\mathbb{E}[\varepsilon_{i,t}|D_i, \alpha_i] = \mathbb{E}[\varepsilon_{i,t}|\alpha_i]$$

$$\mathbb{E}[\varepsilon_{i,t}|\alpha_i] = 0$$

$$Var(\varepsilon_{i,t}) = \sigma^2 \ for \ all \ (i,t)$$

$$Cov(\varepsilon_{i,t}, \varepsilon_{i,t-k}) = r^k \sigma^2 \ for \ all \ (i,t)$$

The first term on the right-hand side of equation (A.10) remains unchanged, despite the fact that $Var(PreTrend_{i,k}) = Var(\lambda_{-k}\dot{\alpha}_i) + 2\sigma^2(1 - r^k)$. Further, from Assumption A.5:

$$Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k}) = -\sigma^2(1 - r^k)(1 - r^t)$$

This implies that:

$$\frac{Cov(\dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0}, \dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})}{(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})} = \frac{-2\sigma^2(1 - r^k)(1 - r^t)}{2(1 - \tilde{R}^2) \times Var(PreTrend_{i,k})} \equiv \frac{-(1 - r^t)}{2}\left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right)$$

Therefore, the bias in the trend-as-control estimator would include an additional term, as follows:

$$plim \ \hat{\rho}_{t,k} - \rho_{t,k} = -\left(\rho_{t,k} + \frac{(1 - r^t)}{2}\right) \times \left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) \tag{A.16}$$

Replacing equation (A.16) into the conditional bias expression from equation (A.12), the resulting conditional bias becomes:[33]

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \tilde{\Delta}_t\left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) + \frac{(1 - r^t)\tilde{\Delta}_{-k}}{2}\left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) \tag{A.17}$$

Note that the first term on the right-hand side of equation (A.17) is equivalent to our main

---

[33]Demonstration:

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = -(plim \ \hat{\rho}_{t,k} - \rho_{t,k})\frac{\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)} \ \text{[equation (A.12)]}$$

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \left(\rho_{t,k} + \frac{(1 - r^t)}{2}\right) \times \left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) \times \frac{\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)}$$

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \underbrace{\frac{\rho_{t,k}\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)}}_{\tilde{\Delta}_t} \times \left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) + \frac{(1 - r^t)}{2} \times \left(\frac{1 - \zeta}{1 - \tilde{R}^2}\right) \times \underbrace{\frac{\lambda_{-k}Cov(\alpha_i, D_i)}{Var(D_i)}}_{\tilde{\Delta}_{-k}}$$

expression of bias reduction conditional on trend-as-control. The second term represents an additional bias due to the serial correlation of the idiosyncratic shock. However, even without autocorrelation (i.e., $r = 0$), we would still get a similar expression when measuring the pre-trend using the same baseline as the post-treatment growth. In the following, we consider two scenarios for evaluating the potential of trend-as-control for bias reduction.

1. Maintaining the assumptions of selection on α_i and persistent common shocks, we substitute $\tilde{\Delta}_{-k} = \tilde{\Delta}_t / \rho_{t,k}$ into equation (A.17). The conditional bias can be rewritten as:

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \tilde{\Delta}_t \left(\frac{1 - \zeta}{1 - \tilde{R}^2} \right) \times \left(1 + \frac{(1 - r^t)}{2\rho_{t,k}} \right) \qquad (A.17.1)$$

As a result, the trend-as-control approach becomes even less effective. However, as long as $\zeta > \frac{(1-r^t)+2\rho_{t,k}\tilde{R}^2}{(1-r^t)+2\rho_{t,k}}$, it still reduces the bias.

We highlight that the expression for regression using only pre-trends obtained in equation (A.15) remains unchanged. The main difference now is that, due to the correlations of the idiosyncratic shocks, it is possible that $\text{sign}(plim\ \hat{\pi}_{k,s}) \neq \text{sign}(\rho_{ks})$:

$$plim\ \hat{\pi}_{k,s} = \rho_{ks}\zeta_{(s)} - \frac{(1 - \zeta_{(s)})(1 - r^k)}{2} \qquad (A.18)$$

As a result, the expression in equation (A.15) can also assume negative values. Therefore, trend-as-control will reduce bias in the pre-treatment period as long as:

$$\frac{plim\ \hat{\pi}_{k,s}}{\rho_{ks}} > \tilde{R}_s^2 \frac{\tilde{\Delta}_k(\mathbf{PreTrend}_s)}{\tilde{\Delta}_k} \qquad (A.19)$$

All the elements in (A.19) can be estimated using pre-treatment data. If one expect the persistence of growth at the unit level (measured by $\pi_{k,s}$) and the persistence of common shocks (measured by ρ_{ks}) to follow a similar pattern in post-treatment data, then the relative size of these elements can be informative for how well trend-as-control would work. In particular, we highlight that $\text{sign}(plim\ \hat{\pi}_{k,s}) \neq \text{sign}(\rho_{ks})$ is a sufficient condition for *bias introduction* in the pre-treatment data.

2. Assume that there are pre-trends (i.e., $\tilde{\Delta}_{-k} \neq 0$), but there are no common shocks

after treatment (i.e., $\lambda_t = 0 \implies \rho_{t,k} = 0$). In this case, the parallel assumption would hold unconditionally ($\tilde{\Delta}_t = 0$). However, when conditioning on trend-as-control, one would in fact *introduce bias* proportional to the size of the observed pre-trends:

$$\tilde{\Delta}_t(\mathbf{PreTrend}_k) = \frac{(1 - r^t)\tilde{\Delta}_{-k}}{2} \left(\frac{1 - \zeta}{1 - \tilde{R}^2} \right) \tag{A.17.2}$$

A.3 Linear Trends

Because of collinearity, when controlling for linear trends, there is the need to omit an additional event-time indicator. Therefore, δ_k captures the slope of differential linear trends from the omitted period $e = -k$ up to the baseline.[34] For example, in our main simulations, we omit the event-time interaction for the most distant pre-treatment year, allowing δ_k to represent differential growth over $k = 9$ pre-treatment periods.

Proposition A.4. *(Bias Net of Linear Trend Extrapolation)*
Under Assumptions A.1-A.3, the bias of the DID_t conditional on a linear trend extrapolation of length k is equal to the unconditional bias minus $\frac{t}{k}$ times the differential pre-trend:

$$\Delta_t(LT_k) = \lambda_t \Delta\alpha - \frac{t}{k}\left(\lambda_{-k}\Delta\alpha\right) \tag{A.20}$$

where $\Delta\alpha := \mathbf{E}[\alpha_i | D_i = 1] - \mathbf{E}[\alpha_i | D_i = 0]$ are the differential trajectories that arise from treated and untreated units being differentially exposed to common shocks (f_t), $\lambda_t := f_t - f_0$, and $\lambda_{-k} := f_0 - f_{-k}$.

Further, under Assumption A.4, we can simplify the bias expression to:

$$\Delta_t(LT_k) = \left(\rho_{t,k} - \frac{t}{k}\right)\left(\lambda_{-k}\Delta\alpha\right) \tag{A.21}$$

Comparing the size of the bias net of the linear trends extrapolation $(\Delta_t(LT_k))$ with the bias of the baseline estimator (Δ_t) expressed in Proposition A.1, we can express their relative size as:

$$\frac{\Delta_t(LT_k)}{\Delta_t} = 1 - \frac{1}{\frac{k}{t} \times \rho_{t,k}} \tag{A.22}$$

The linear trend extrapolation will result in a bias variation by a proportion $\frac{1}{\frac{k}{t} \times \rho_{t,k}}$.

---

[34]The remaining pre-treatment coefficients can be used to assess the linear specification (Borusyak et al., 2024). Coefficients close to zero support the linear approximation, whereas deviating coefficients suggest non-linear trajectories. If more pre-treatment periods are available, we could control for additional polynomials (Wooldridge, 2021). We focus on the linear specification throughout this paper. In settings with staggered treatment adoption, Goodman-Bacon (2021) recommends detrending the outcome by subtracting group-(unit-)specific linear trends estimated solely from pre-treatment data to avoid the "contamination" of treatment effects. However, the linear trends specification can be extended to include group-specific linear trends interactions, as proposed by Wooldridge (2021).

Proof. Considering a linear TWFE regression controlling for the linear trend extrapolation of length k as shown in specification (S3), the post-treatment parameter of interest, β_t, is identified as follows:

$$\beta_t = \underbrace{\mathbb{E}[y_{i,t} - y_{i,0}|D_i = 1] - \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 0]}_{DID_t} - \delta_k \times t$$

resulting from $\mathbb{E}[y_{i,t} - y_{i,0}|D_i = 1] = \tau_t - \tau_0 + \delta_k \times (Trend_t - Trend_0) + \beta_t$, where $Trend_t - Trend_0 = t$, and $\mathbb{E}[y_{i,t} - y_{i,0}|D_i = 0] = \tau_t - \tau_0$, and δ_k is the (linear) slope of the differential pre-trend of length k.

In Appendix $A.1$, we showed that $DID_t = ATT_t + \Delta_t$. Considering the linear factor model for potential outcome, the unconditional bias was derived (Proposition A.1) under Assumptions A.1-A.3 as $\Delta_t = \lambda_t \Delta\alpha$, where $\Delta\alpha := \mathbb{E}[\alpha_i|D_i = 1] - \mathbb{E}[\alpha_i|D_i = 0]$.

Let the bias net of the linear trend extrapolation be defined as $\Delta_t(LT_k) := \beta_t - ATT_t$. Therefore, Assumptions A.1-A.3 imply that:

$$\Delta_t(LT_k) = \lambda_t \Delta\alpha - \delta_k \times t$$

where, taking into account that $\mathbb{E}[y_{i,0} = y_{i,-k}|D_i = 1] = \tau_0 - \tau_{-k} + \delta_k(Trend_0 - Trend_{-k})$, for $Trend_0 - Trend_{-k} = k$, and $\mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 0] = \tau_0 - \tau_{-k}$, implies:

$$\delta_k = \frac{\mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 1] - \mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 0]}{k}$$

Further, rewriting δ_k in terms of demeaned outcomes, and considering the pre-trends model from equation (A.7) under Ass. A.1: $\dot{y}_{i,0} - \dot{y}_{i,-k} = \lambda_{-k}\dot{\alpha}_i + (\dot{\varepsilon}_{i,0} - \dot{\varepsilon}_{i,-k})$. This implies that, under Assumptions A.2 and A.3:

$$\delta_k = \tfrac{1}{k}\lambda_{-k}\Delta\alpha$$

where $\Delta\alpha := \mathbb{E}[\alpha_i|D_i = 1] - \mathbb{E}[\alpha_i|D_i = 0]$

Then we get the following expression for the bias:

$$\Delta_t(LT_k) = \lambda_t \Delta\alpha - \frac{t}{k}\lambda_{-k}\Delta\alpha$$

Further, under Assumption A.4, the bias net of the linear trend extrapolation can be written as follows:

$$\Delta_t(LT_k) = \left(\rho_{t,k} - \frac{t}{k}\right)\lambda_{-k}\Delta\alpha$$

Therefore, the relationship between $\Delta_t(LT_k)$ and Δ_t is given by:

$$\frac{\Delta_t(LT_k)}{\Delta_t} = \left(\frac{\rho_{t,k} - \frac{t}{k}}{\rho_{t,k}}\right)[\lambda_{-k}\Delta\alpha][\lambda_{-k}\Delta\alpha]^{-1}$$

$$\frac{\Delta_t(LT_k)}{\Delta_t} = \left(\frac{\rho_{t,k} - \frac{t}{k}}{\rho_{t,k}}\right)$$

$$\implies \frac{\Delta_t(LT_k)}{\Delta_t} = 1 - \left(\frac{1}{\frac{k}{t} \times \rho_{t,k}}\right)$$

Which implies in the following percent change in bias:

$$\frac{\Delta_t - \Delta_t(LT_k)}{\Delta_t} = \left(\frac{1}{\frac{k}{t} \times \rho_{t,k}}\right)$$

□

Assumption A.6. *Parallel Trends with Linear Trends:*

$$\mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 1] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0|D_i = 0] = \frac{t}{k} \times \{\mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 1] - \mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 0]\} \quad \forall t > 0$$

Proposition A.5. *Under Assumptions A.1 and A.6, the parameter β_t in equation (S3) identifies the Average Treatment Effect on the Treated in period $t > 0$ (ATT_t).*

Proof. From the definitions of Δ_t and the identification of δ_k, Assumption A.6 implies that $\Delta_t = \delta_k \times t \ \forall t > 0 \implies \beta_t = ATT_t$ □

Remark 3. *For a symmetric time change such that $Trend_k - Trend_0 = Trend_0 - Trend_{-k}$ for some $k > 0$, Assumption A.6 states that the difference in the trajectories of the potential*

outcomes from $t = 0$ to $t = k$ in the absence of treatment is equal to the pre-treatment difference in the trajectories between treated and control groups:

$$\mathbb{E}[y_{i,k}^0 - y_{i,0}^0 | D_i = 1] - \mathbb{E}[y_{i,k}^0 - y_{i,0}^0 | D_i = 0] = \mathbb{E}[y_{i,0} - y_{i,-k} | D_i = 1] - \mathbb{E}[y_{i,0} - y_{i,-k} | D_i = 0]$$

A.3.1 Illustration of Linear Trends

In the surveyed papers, we found that researchers often state concerns that the linear trends coefficient(s) could capture post-treatment trajectories. Miller (2023) discusses how to avoid this issue. First, researchers must exercise caution when applying post-treatment restrictions. For example, rather than using a single event-time indicator for several post-treatment years, they should include event-time indicators for each period individually, as we did in our specifications. Second, there are alternative restrictions that can address multicollinearity due to linear trends, with a common approach being to omit an additional (pre-treatment) event-time indicator. As we demonstrated above, the parameter(s) of the differential linear trends will not capture post-treatment changes as long as we thoughtfully exclude the "right" event times from the regression.

For illustration, let us consider a simple example where we observe five years before and five years after the treatment ($t = -5, -4, ..., 0, +1, .., +5$). For this purpose, we simulated a scenario where the outcome follows an increasing trend among treated units at 5% per year. We also simulated a constant treatment effect of 10%. The baseline estimates displayed in Figure A.1 capture both the linear trends and the treatment effect. When controlling for linear trends, we further omit event time $t = -5$. Therefore, δ identifies the slope of the differential (linear) trajectory of the outcome in the treated group compared to the untreated group from $t = -5$ to $t = 0$, which is 0.05. δ <u>does not</u> reflect outcome growth after the treatment. Instead, it linearly extrapolates the pre-trend over the entire period, as shown by the red line in Figure A.1. As a consequence, the estimates of β_t when controlling for linear trends are the standard difference-in-differences (i.e., the "baseline" estimates) *minus* the "Linear Extrapolation". They are depicted in the blue dots in Figure A.1, showing the 10% "treatment effect".
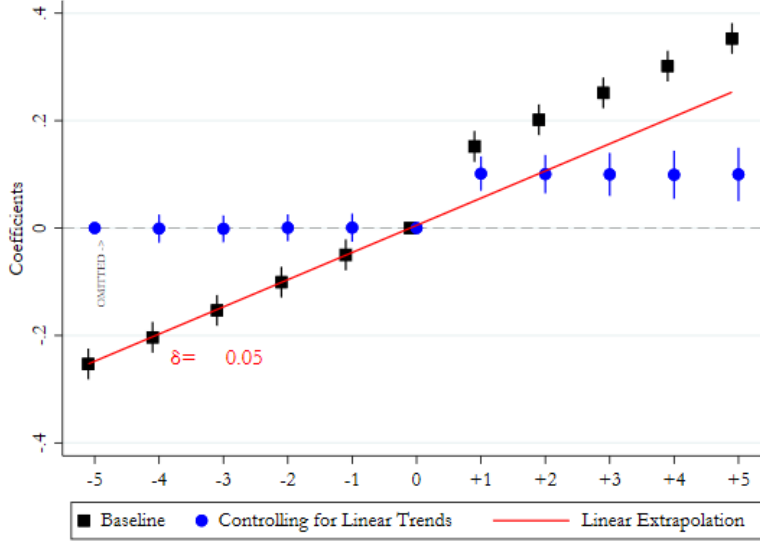
Figure A.1: Illustration of the Linear Trends Approach

Notes: We randomly allocate a treatment status (D_i) across Brazilian LLMs. The outcome is log of workers. The treatment hits in 2014, and we estimate dynamic effects for five pre- and five post-treatment years. We imputed an increasing trend for the treated group: $y_{i,t} = y_{i,t} + 0.05 \times Trend_t \times D_i$. Further, we added 0.10 to the outcome of the treated units from $t+1$ to $t+5$ to represent a 10% treatment effect. The linear trends specification is shown in equation (S3), where $\underline{T} = -5$ and $\overline{T} = +5$.

Importantly, when applying similar restrictions (i.e., normalization of two pre-treatment parameters to zero), Miller (2023) recommends to space the omitted coefficients further apart. The reason is that statistical noise tends to have a greater impact when the omitted coefficients are closer together. We further discuss long vs. short trends in Section 4.2.3.

A.3.2 Equivalence of Treatment-specific and Unit-specific Linear Trends

Our main linear trends simulations fit a single parameter for the treatment group, as formalized in Wooldridge (2021). This yields the simple interpretation in terms of differential trends between the treated and control groups. In practice, however, researchers often control for "unit-specific linear trends". Therefore, instead of a single parameter for the slope of the differential linear trends (δ_k), this approach allows a specific slope ($\delta_{i,k}$) for each

unit i, as follows:

$$log(y_{i,t}) = \alpha_i + \tau_t + \sum_{\substack{e=\underline{T} \\ e \neq -k}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{T} \beta_e \times D_i \times \mathbb{I}[e = t]$$

(S3.1)

$$+ \delta_{i,k} Trend_t \times \mathbb{I}[region = i] + \epsilon_{i,t}$$

for linear trends extrapolated over k pre-treatment years.

In this section, we show that controlling for unit-specific is similar to treatment-specific linear trends. In fact, both approaches are equivalent in expectations:

$$\mathbb{E}[\delta_{i,k}|D_i = 1] - \mathbb{E}[\delta_{i,k}|D_i = 0] = \delta_k$$

(A.23)

Proof.

$y_{i,-k} = \alpha_i + \tau_{-k} + \delta_{i,k} Trend_{-k} + \epsilon_{i,-k}$

$y_{i,0} = \alpha_i + \tau_0 + \delta_{i,k} Trend_0 + \epsilon_{i,0}$

Let the omitted unit be $j \neq i$:

$y_{j,-k} = \alpha_j + \tau_{-k} + \epsilon_{j,-k}$

$y_{j,0} = \alpha_j + \tau_0 + \epsilon_{j,0}$

Therefore:

$\delta_{i,k} \times k = \{[y_{i,0} - y_{i,-k}] - [y_{j,0} - y_{j,-k}]\} + \{[\epsilon_{j,0} - \epsilon_{j,-k}] - [\epsilon_{i,0} - \epsilon_{i,-k}]\}$

In expectations:

$\mathbb{E}[\delta_{i,k}|D_i] \times k = \mathbb{E}[y_{i,0} - y_{i,-k}|D_i] - \mathbb{E}[y_{j,0} - y_{j,-k}] + \mathbb{E}[\epsilon_{j,0} - \epsilon_{j,-k}] - \mathbb{E}[\epsilon_{i,0} - \epsilon_{i,-k}|D_i]$

Taking the difference for treated and untreated:

$k \times \{\mathbb{E}[\delta_{i,k}|D_i = 1] - \mathbb{E}[\delta_{i,k}|D_i = 0]\} = \mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 1] - \mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 0]$

$\implies \mathbb{E}[\delta_{i,k}|D_i = 1] - \mathbb{E}[\delta_{i,k}|D_i = 0] = \dfrac{\mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 1] - \mathbb{E}[y_{i,0} - y_{i,-k}|D_i = 0]}{k} \equiv \delta_k$

$\square$

Our simulation results discussed in Section 4.2.2 indicate that some differences might

arise in finite samples. We further show that the equivalence only holds if unit-specific linear trends are fitted at the level of treatment allocation (or nested). In contrast, it does not hold if the linear trends are fitted at a more aggregated unit than the treatment assignment. For instance, our simulations results differ when controlling for "state-specific linear trends" as detailed in equation (S3.2).

$$log(y_{i,t}) = \alpha_i + \tau_t + \sum_{\substack{e=\underline{T} \\ e \neq -k}}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e \times D_i \times \mathbb{I}[e = t]$$
$$+ \delta_{s,k} Trend_t \times \mathbb{I}[i \in s] + \epsilon_{i,t}$$
(S3.2)

for linear trends extrapolated over k pre-treatment years, allowing for a specific slope for each state s.

A.4 Group-by-time FEs

When controlling for group-by-time (e.g., state-by-year) FEs, the $DID(s)$ consists of the average difference in the outcome trajectory between treated and untreated units within each group (e.g., state) s:

$$DID_t(s) = \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 1, group = s] - \mathbb{E}[y_{i,t} - y_{i,0}|D_i = 0, group = s] \quad (A.24)$$

Which can be decomposed as follows:

$$DID_t(s) = ATT_t(s) + \Delta_t(s) \quad (A.25)$$

where $\Delta_t(s)$ is the bias from the violation of the group-specific parallel trends assumption formalized below, and $ATT_t(s)$ is the average treatment effect on the treated for each group s (the proof follows as in the conditional DID).

Assumption A.7. ***Group-specific Parallel Trends:*** *For each group (e.g., state) s, the*

following must hold:

$$\Delta_t(s) := \mathbb{E}[y_{i,t}^0 - y_{i,0}^0 | D_i = 1, group = s] - \mathbb{E}[y_{i,t}^0 - y_{i,0}^0 | D_i = 0, group = s] = 0 \ \ \forall t \neq 0$$

Finally, based on the Law of Iterated Expectation, the ATT_t is the average of the group-specific effects $ATT_t(s)$ among the treated:

$$ATT_t = \mathbb{E}[ATT_t(s)|D_i = 1] \tag{A.26}$$

It is straightforward to observe that the inclusion of group-by-time FEs requires variation in treatment allocation within groups. In other words, the identification of $DID(s)$ requires the presence of observations for which $D_i = 1$ and $D_i = 0$ within each group s.

A.4.1 Conditional Bias

We study the potential of group-by-time FEs to eliminate bias by rewriting the linear model for the untreated potential outcome introduced in equation (3) by replacing the year-specific effects τ_t by group-specific shocks τ_{st}. Note that now the time effects $\tau_{s,t}$ will not be eliminated by demeaning the outcome, unless it is done at the group (e.g., state) level. The demeaned potential outcome is then given by:

$$\dot{y}_{i,t}^0 = \dot{\mu}_i + \dot{\tau}_{s,t} + f_t \dot{\alpha}_i + \dot{\varepsilon}_{i,t} \tag{A.27}$$

where f_t is a vector of common shock; $\dot{\alpha}_i = \alpha_i - \mathbb{E}[\alpha_i]$ is the unit-specific exposure to common shocks; and $\dot{\varepsilon}_{i,t} = \varepsilon_{i,t} - \mathbb{E}[\varepsilon_{i,t}]$ is the (demeaned) idiosyncratic shock. Therefore, potential outcome growth can be expressed as follows:

$$\dot{y}_{i,t}^0 - \dot{y}_{i,0}^0 = (\dot{\tau}_{s,t} - \dot{\tau}_{s,0}) + \lambda_t \dot{\alpha}_i + \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0} \tag{A.28}$$

where $\lambda_t := f_t - f_0$

And the observed outcome can be written as follows:

$$\Delta \dot{y}_{i,t} = \beta_t^0 D_i + (\dot{\tau}_{s,t} - \dot{\tau}_{s,0}) + \lambda_t \dot{\alpha}_i + \dot{\varepsilon}_{i,t} - \dot{\varepsilon}_{i,0} \tag{A.29}$$

Proposition A.6. *(**Unconditional Bias with Group-by-time FEs**)*

Under Assumptions A.1-A.3 (see Section 4.1.2), the unconditional bias (i.e., without controlling for group-by-time FEs) can be expressed as:

$$\tilde{\Delta}_t^S = \underbrace{\lambda_t \frac{Cov(\alpha_i, D_i)}{Var(D_i)}}_{\tilde{\Delta}_t} + \Delta\tau_t \tag{A.30}$$

where the first term on the right-hand side represents the unconditional bias in the model without variation in time shocks across groups (when $\tau_{s,t} = \tau_t \; \forall s$, see Proposition A.1). The sector term represents the difference in the evolution of group-specific shocks between groups:
$\Delta\tau_t := \mathbb{E}[\dot{\tau}_{s,0} - \dot{\tau}_{s,-1}|D_i = 1] - \mathbb{E}[\dot{\tau}_{s,0} - \dot{\tau}_{s,-1}|D_i = 0] \equiv \frac{Cov(\dot{\tau}_{s,t} - \dot{\tau}_{s,0}, D_i)}{Var(D_i)}.$

Note that the estimator for β_t^0 from equation (A.29) is a weighted average of $\hat{\beta}_t^s$, the effects estimated within each group (Angrist and Pischke, 2009; Goodman-Bacon, 2021):

$$plim \; \hat{\beta}_t^0 = \sum_s \frac{n_s Var(D_i|S = s)}{\mathbb{E}[Var(D_i|S)]} \times (plim \; \hat{\beta}_t^s),$$
$$\text{where } plim \; \hat{\beta}_t^s = \frac{Cov(\Delta\dot{y}_{i,t}, D_i|S = s)}{Var(D_i|S = s)} \tag{A.31}$$

and $n_s := \frac{N_s}{N}$ is the share of observations in group s, and $Var(D_i|S)$ is the conditional variance of the treatment. Equation (A.31) shows that the regression with group-by-time FE weights each group-specific effect by the relative size of the group and the variance of treatment within the group.

Proposition A.7. *(**Conditional Bias with Group-by-time FEs**)*

Under Assumptions A.1-A.3 (see Section 4.1.2), and considering homogeneous treatment

effects across groups the asymptotic bias of the DID_t estimator with Group-by-time FEs is:

$$\tilde{\Delta}_t^S(\boldsymbol{S}) := plim \ \hat{\beta}_t^0 - \beta_t^0 = \lambda_t \frac{\mathbb{E}[Cov(\alpha_i, D_i|S)]}{(1 - \tilde{R}_s^2) \times V(D_i)} \tag{A.32}$$

where $\mathbb{E}[Cov(\alpha_i, D_i|S)] := \sum_s n_s Cov(\alpha_i, D_i|S = s)$ is the average of the conditional covariance between the factor exposure and the treatment allocation; and $\tilde{R}_s^2$ is the fraction of the variance in D_i explained by the group fixed effects.

Proof. Considering homogeneous treatment effects across groups and Assumptions A.1-A.3,

$$plim \ \hat{\beta}_t^s = \beta_t^0 + \lambda_t \frac{Cov(\alpha_i, D_i|S = s)}{Var(D_i|S = s)}$$

This implies that equation (A.31) can be written as follows:

$$plim \ \hat{\beta}_t - \beta_t^0 = \sum_s \frac{n_s Var(D_i|S = s)}{\mathbb{E}[Var(D_i|S)]} \times \lambda_t \frac{Cov(\alpha_i, D_i|S = s)}{Var(D_i|S = s)}$$

$$plim \ \hat{\beta}_t - \beta_t^0 = \lambda_t \frac{1}{\mathbb{E}[Var(D_i|S)]} \sum_s n_s Cov(\alpha_i, D_i|S = s)$$

$$plim \ \hat{\beta}_t - \beta_t^0 = \lambda_t \frac{\mathbb{E}[Cov(\alpha_i, D_i|S)]}{\mathbb{E}[Var(D_i|S)]}$$

By the law of total variance, $Var(D_i) = \underbrace{\mathbb{E}[Var(D_i|S)]}_{\text{avg. within groups variance}} + \underbrace{Var[\mathbb{E}(D_i|S)]}_{\text{between groups variance}}$. The between groups variance is eliminated with the inclusion of group-by-time FEs. Let the fraction of the variance explained by the group fixed effects be defined as $\tilde{R}_s^2 := \frac{Var[\mathbb{E}(D_i|S)]}{V(D_i)}$. Then $\mathbb{E}[Var(D_i|S)] = (1 - \tilde{R}_s^2) \times V(D_i)$. $\qquad\square$

A comparison of equations (A.30) and (A.32) shows that, in addition to eliminating the differential evolution of group-specific shocks ($\Delta\tau_t$), controlling for group-by-time fixed effects also alters the bias arising from differential factor exposure.

Comparing the overall conditional bias with the unconditional version:

$$\frac{\tilde{\Delta}_t^S(\mathbf{S})}{\tilde{\Delta}_t^S} = \frac{\lambda_t}{(1 - \tilde{R}_s^2)} \frac{\mathbb{E}[Cov(\alpha_i, D_i | S)]}{\lambda_t Cov(\alpha_i, D_i) + Cov(\dot{\tau}_{s,t} - \dot{\tau}_{s,0}, D_i)}$$

$$\frac{\tilde{\Delta}_t^S(\mathbf{S})}{\tilde{\Delta}_t^S} = \frac{1}{(1 - \tilde{R}_s^2)} \frac{1}{\frac{Cov(\alpha_i, D_i)}{\mathbb{E}[Cov(\alpha_i, D_i | S)]} + \frac{Cov(\dot{\tau}_{s,t} - \dot{\tau}_{s,0}, D_i)}{\lambda_t \mathbb{E}[Cov(\alpha_i, D_i | S)]}}$$

$$= \frac{1}{(1 - \tilde{R}_s^2)} \frac{1}{\left(1 + \underbrace{\frac{Cov(\mathbb{E}[\alpha_i | S], \mathbb{E}[D_i | S])}{\mathbb{E}[Cov(\alpha_i, D_i | S)]}}_{A} + \underbrace{\frac{Cov(\dot{\tau}_{s,t} - \dot{\tau}_{s,0}, D_i)}{\lambda_t \mathbb{E}[Cov(\alpha_i, D_i | S)]}}_{B}\right)}$$

Condition for bias reduction: $A + B \geq \frac{\tilde{R}_s^2}{1 - \tilde{R}_s^2}$

In general, a **necessary condition** for bias reduction is that $B \geq 0$ (i.e., state-specific shocks ($\tau_{s,t}$) and the common shocks f_t move in the same direction).

Bias reduction is inversely proportional to the proportion of treatment variance from *between groups* (i.e., $\tilde{R}_s^2$).

B Monte Carlo Simulations

We study the performance of different strategies based on real data with simulated treatment assignments. Our data contains 558 regions (e.g., local labor markets) i, spanning from 1995 to 2019, from the Brazilian linked employer-employee administrative information. The performance measures are obtained from the estimates across 1,000 Monte Carlo simulations.

In each simulation, treatment is assigned (D_i) from a random binomial distribution, where the probability of being treated (P_i) is correlated with baseline (i.e., in 2000) characteristics of the region. We set $P_i(X_i)$ to range between 0 and 1, as follows:

$$P_{1,i}(X_i) = \max\{0, 0.8 - high.school_i\} \tag{B.1}$$

$$P_{2,i}(X_i) = \frac{\exp(2 \times high.school_i - 3 \times manufacturing_i + 0.1 \times pop.density_i)}{1 + \exp(2 \times high.school_i - 3 \times manufacturing_i + 0.1 \times pop.density_i)} \tag{B.2}$$

where $high.school_i$ is the share of workers in region i with a high school degree in the baseline; $manufacturing_i$ is the share of employment in manufacturing; and $pop.density_i$ is population (1,000) per km^2. Because of the random aspect in the treatment allocation, the number of treated and untreated units changes in each simulation.

The event time $T_0 = 0$ is set as the baseline, and is drawn from a random uniform (integer) distribution from 2005 and 2010. Each simulation contains 19 years of data, randomly selecting between six alternative spans: 1996-2014; 1997-2015; 1998-2016; 1999-2017; 2000-2018; 2001-2019. Therefore, there are nine periods before and nine periods after the treatment. The year of treatment is fixed within simulations and constant across units i: all units for which $D_i = 1$ are treated at the same time.

B.1 Specifications

We estimate dynamic treatment effects for each event time by comparing the treated regions ($D_i = 1$) with the untreated regions ($D_i = 0$), where the outcomes is the log of workers in region i and event time t. In each simulation $\omega = 1, .., 1000$, we estimate γ_t^ω and β_t^ω, as well as their respective standard errors σ_t^ω, using the specifications summarized in this section.

B.1.1 Baseline

$$log(y_{i,t}) = \alpha_i^\omega + \tau_t^\omega + \sum_{e \leq -1} \gamma_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \epsilon_{i,t}^\omega \qquad \text{(S1$'$)}$$

where α_i^ω are region fixed effects, τ_t^ω are year fixed effects, and D_i^ω is a binary variable equal to 1 if region i is treated in simulation ω, zero otherwise. The subscript e refers to event time.

B.1.2 Trend-as-control

The trend-as-control approach consists in controlling for a covariate $PreTrend_{i,k}$ that measures outcome growth over k pre-treatment years. Since this covariate is fixed over time, we also interact with year indicators.

$$log(y_{i,t}) = \alpha_i^\omega + \tau_t^\omega + \sum_{\substack{e=\underline{T} \\ e \neq -k}}^{-1} \gamma_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \sum_e \phi_e^\omega PreTrend_{i,k} \times \mathbb{I}[e = t] + \epsilon_{i,t}^\omega$$

$$\text{(S2$'$)}$$

where $PreTrend_{i,k} = log(y_{i,0}) - log(y_{i,-k})$ is outcome growth over k pre-treatment years, and event time $e = -k$ is omitted in addition to the baseline.

B.1.3 Linear Trends

$$log(y_{i,t}) = \alpha_i^\omega + \tau_t^\omega + \sum_{\substack{e=\underline{T} \\ e \neq -k}}^{-1} \gamma_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \delta_k^\omega Trend_t \times D_i^\omega + \epsilon_{i,t}^\omega \quad \text{(S3$'$)}$$

where δ_k^ω is the slope of the differential linear trend over k pre-treatment years.

B.1.4 Group-by-time FEs

$$log(y_{i,t}) = \alpha_i^\omega + \tau_{st}^\omega + \sum_{e=\underline{T}}^{-1} \gamma_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e^\omega \times D_i^\omega \times \mathbb{I}[e = t] + \epsilon_{i,t}^\omega \qquad \text{(S4$'$)}$$

where τ_{st}^ω are group-by-time (e.g., state-by-year) fixed effects.

B.2 Performance Measures

The simulated coefficients $\hat{\theta} = (\hat{\gamma}, \hat{\beta})$ are the average across all the simulations ω:

$$\hat{\theta}_t = \frac{\sum_{\omega=1}^{1,000} \hat{\theta}_t^{\omega}}{1,000} \tag{B.3}$$

The true parameter $\theta_t = 0$ for all t. However, since the treatment allocation is correlated with characteristics of the labor markets, incidental trends may lead to non-zero simulated effects. Therefore, $\hat{\theta}_t$ obtained from equation (B.3) gives us a measure of bias.

Nevertheless, estimators might be biased in different directions across simulations, resulting in a small average bias. Further, imprecise estimators often result in wider confidence intervals, increasing the likelihood of encompassing the true parameter value, even in the presence of systematic bias. The variance is obtained as follows:

$$V(\hat{\theta}_t) = \frac{\sum_{\omega=1}^{1,000} \left(\hat{\theta}_t^{\omega} - \hat{\theta}_t\right)^2}{1,000 - 1} \tag{B.4}$$

We then summarize the trade-off between bias and variance by the Root Mean Squared Errors (RMSE). Finally, in order to disentangle the dispersion of the estimates from the RMSE, we also report the Standard Deviation (SD).

$$RMSE(\hat{\theta}_t) = \sqrt{\left(V(\hat{\theta}_t) + \hat{\theta}_t^2\right)} \tag{B.5}$$

$$SD(\hat{\theta}_t) = \sqrt{\hat{V}(\hat{\theta}_t)} \tag{B.6}$$

In Table B.1, we also report the coverage, which indicates the proportion of times that the true value $\theta_t = 0$ falls within the 95% confidence interval (CI). For instance, a coverage of 95% means that the confidence interval includes the (true) null effect in 950 out of 1000 simulations. Based on robust standard errors clustered at the unit (LLM) level ($\hat{\sigma}_t^{\omega}$), we measure the coverage as follows:

$$Coverage(\hat{\theta}_t) = \frac{\sum_{\omega=1}^{1,000} \mathbb{I}\left\{0 \in \left[\hat{\theta}_t^\omega - 1.96 \cdot \hat{\sigma}_t^\omega, \hat{\theta}_t^\omega + 1.96 \cdot \hat{\sigma}_t^\omega\right]\right\}}{1,000} \tag{B.7}$$

Table B.1: Summary of performance measures for main specifications: LLM level

	$t = +1$				$t = +9$			
	Bias	Coverage	SD	RMSE	Bias	Coverage	SD	RMSE
Baseline	0.006	0.906	0.009	0.011	0.047	0.456	0.024	0.053
Trend-as-control ($k = 9$)	0.006	0.907	0.009	0.011	0.042	0.513	0.023	0.048
Doubly robust[a]	0.006	0.915	0.009	0.011	0.042	0.560	0.023	0.048
Interacted[b]	0.005	0.923	0.009	0.011	0.040	0.602	0.024	0.047
Matching on pre-trend	0.006	0.929	0.012	0.013	0.033	0.840	0.030	0.045
Short Trend-as-control ($k = 1$)	0.007	0.900	0.008	0.011	0.046	0.470	0.023	0.052
Long Linear Trends (LT, $k = 9$)	0.002	0.943	0.009	0.010	0.012	0.946	0.035	0.037
Unit-specific LT	0.002	0.951	0.009	0.010	0.010	0.958	0.036	0.037
LT (pre-testing fails)	0.001	0.933	0.005	0.005	-0.001	0.920	0.020	0.020
LT (pre-testing passes)	0.003	0.946	0.008	0.008	0.016	0.954	0.029	0.033
Short Linear Trends (LT, $k = 3$)	-0.000	0.954	0.010	0.010	-0.014	0.927	0.055	0.056
Short LT (pre-testing fails)	-0.004	0.929	0.005	0.007	-0.046	0.858	0.030	0.055
Short LT (pre-testing passes)	0.001	0.961	0.009	0.009	-0.004	0.947	0.045	0.045
Group× time FEs	0.004	0.943	0.009	0.010	0.034	0.656	0.021	0.040
Covariates[c]	-0.000	0.955	0.009	0.009	-0.001	0.946	0.023	0.023

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of workers. See Section 2 for definition of each estimator.
[a] Doubly robust difference-in-differences based on Sant'Anna and Zhao (2020), implemented with the *csdid* command for Stata (Callaway and Sant'Anna, 2021).
[b] See Appendix C for details, in particular equation (S2.2).
[c] We linearly control for the treatment probability $P_i(X_i)$ interacted with years. See Section 3.2 for treatment definitions.

C Alternative Trend-as-control Specifications

In our survey of papers published in the *AER* (Table 1), we observed that researchers usually employ the trend-as-control approach by matching on pre-treatment outcomes. We highlight that the key findings of this paper also extend to matching methods, as well as to more flexible specifications. First, under common support, matching and regression are fundamentally similar, differing mainly in how they weight covariate-specific effects to estimate average treatment effects (Angrist and Pischke, 2009). Second, the insights from the model in Section 4.1.2 are derived in terms of conditional expectations, making them applicable beyond specific functional forms.

As shown in Figure C.1, our simulation results remain consistent to matching on the pre-trend, as well as across alternative specifications that allow for greater flexibility in handling covariates. For the matching approach, we implement a single nearest-neighbor matching (with replacement) based on the covariate $PreTrend_{i,-9}$. We then estimate the baseline equation (S1) using only the matched controls as the comparison group ("matching on pre-trend").[35]

In our main specification (S2) we linearly control for $PreTrend_{i,k}$ interacted with time indicators. Besides the standard linearity assumption, equation (S2) assumes that the treatment effect are homogeneous across all covariate subgroups. As an alternative, practitioners can allow for more flexibility in this relationship by interacting $PreTrend_{i,k}$ with the treatment dummy (Wooldridge, 2021), as shown in the "interacted trend-as-control" equation (S2.2).

$$
\begin{aligned}
log(y_{i,t}) = \alpha_i + \tau_t + \sum_{e=-8}^{-1} \gamma_e \times D_i \times \mathbb{I}[e = t] + \sum_{e=1}^{9} \beta_e \times D_i \times \mathbb{I}[e = t] \\
+ \sum_e \phi_{1,e}(PreTrend_{i,-9} - \overline{PreTrend}_{-9}) \times D_i \times \mathbb{I}[e = t] + \sum_e \phi_{2,e}(PreTrend_{i,-9}) \times \mathbb{I}[e = t] + \epsilon_{i,t}
\end{aligned}
$$

$$\text{(S2.2)}$$

---

[35] When restricting the sample using a matching algorithm, researchers need to adjust the standard errors in the estimation stage to account for correlations created in the matching step (Abadie and Spiess, 2022).

where $PreTrend_{i,-9} = log(y_{i,-9}) - log(y_{i,0})$ is growth over $k = 9$ pre-treatment periods. We recenter $PreTrend_{i,-9}$ around its average value within the treated group $(\overline{PreTrend}_{-9})$, allowing β_e to be interpreted as the average treatment effect on the treated at the average pre-treatment level. Note that $PreTrend_{i,-9}$ should be recentered only when it is interacting with the treatment dummy. Additionally, for accurate inference, researchers must account for the uncertainty in estimating $\overline{PreTrend}_{-9}$ (Wooldridge, 2021).

However, besides still assuming a linear functional form on the covariates, the interacted specification may become infeasible when the number of covariates is large. To address these limitations, we recommend using the doubly robust approach proposed by Sant'Anna and Zhao (2020). This method yields consistent estimates if either the propensity score model or the outcome model is correctly specified, providing robustness to misspecification in one of them. We implement this approach using the *csdid* command in *Stata* (Callaway and Sant'Anna, 2021).[36]

---

[36]Callaway and Sant'Anna (2021) also allows researchers to control for pre-trends with the doubly robust in settings with staggered treatment adoption.

(a) Bias (*Treatment 1*)   (b) Bias (*Treatment 2*)

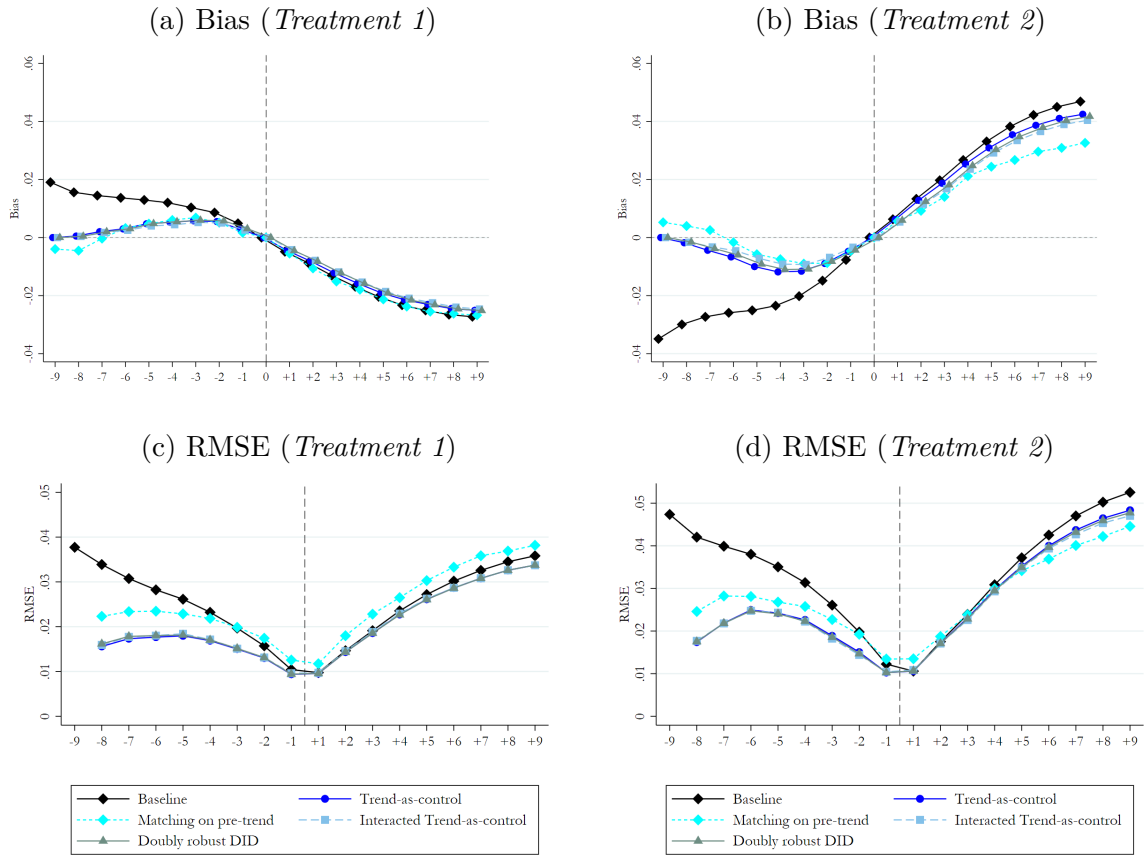(c) RMSE (*Treatment 1*)   (d) RMSE (*Treatment 2*)

Figure C.1: Alternative Specifications for Trend-as-control

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of employment. See Section 2 for definition of each estimator and Section 3.2 for treatment definitions.

In our main simulations, we control for growth from the earliest event time to the baseline (i.e., over $k = 9$ pre-treatment years). In Figure C.2, we also show that our findings are robust to alternative lengths of trend-as-control (e.g., growth over $k = 1$ or $k = 3$ pre-treatment periods).

(a) Bias (*Treatment 1*)

(b) Bias (*Treatment 2*)

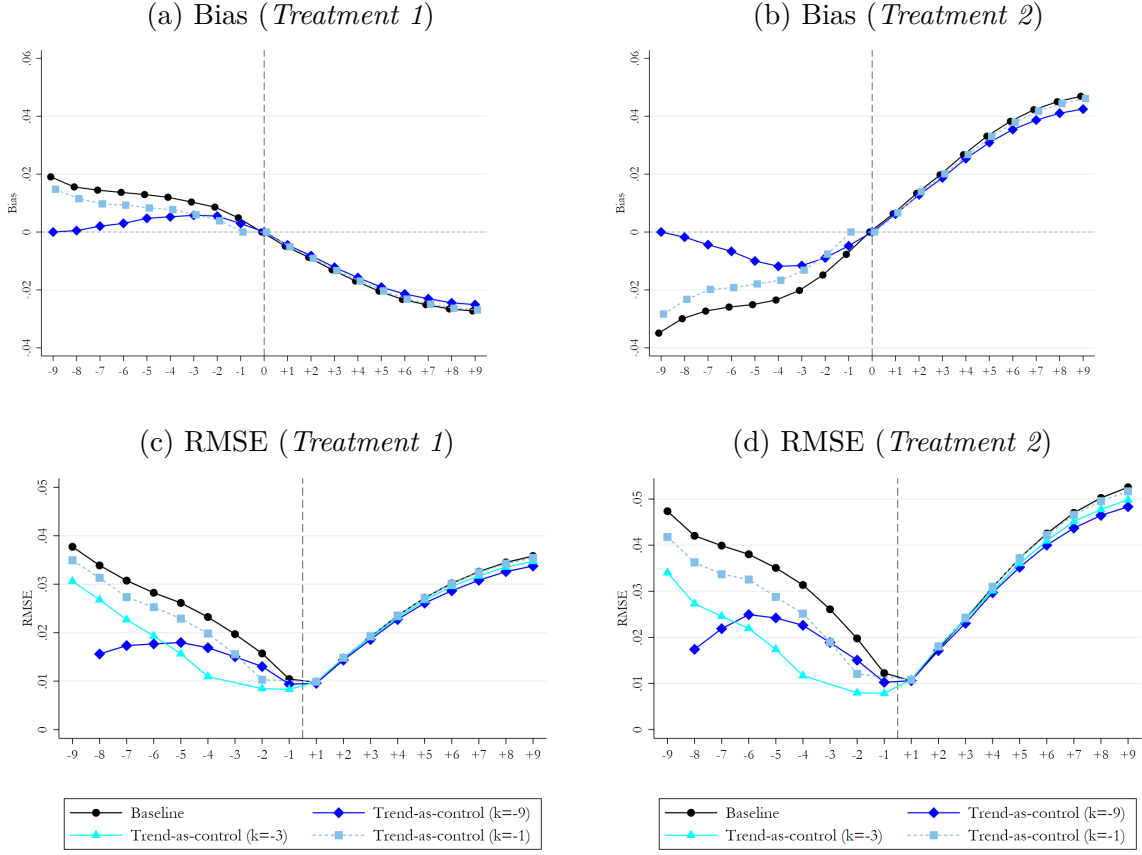(c) RMSE (*Treatment 1*)

(d) RMSE (*Treatment 2*)

Figure C.2: Performance of Short vs. Long Trend-as-control

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of employment. See Section 2 for definition of each estimator and Section 3.2 for treatment definitions.

D   Simulation Results for Log of Wage

(a) Bias (*Treatment 1*)   (b) Bias (*Treatment 2*)

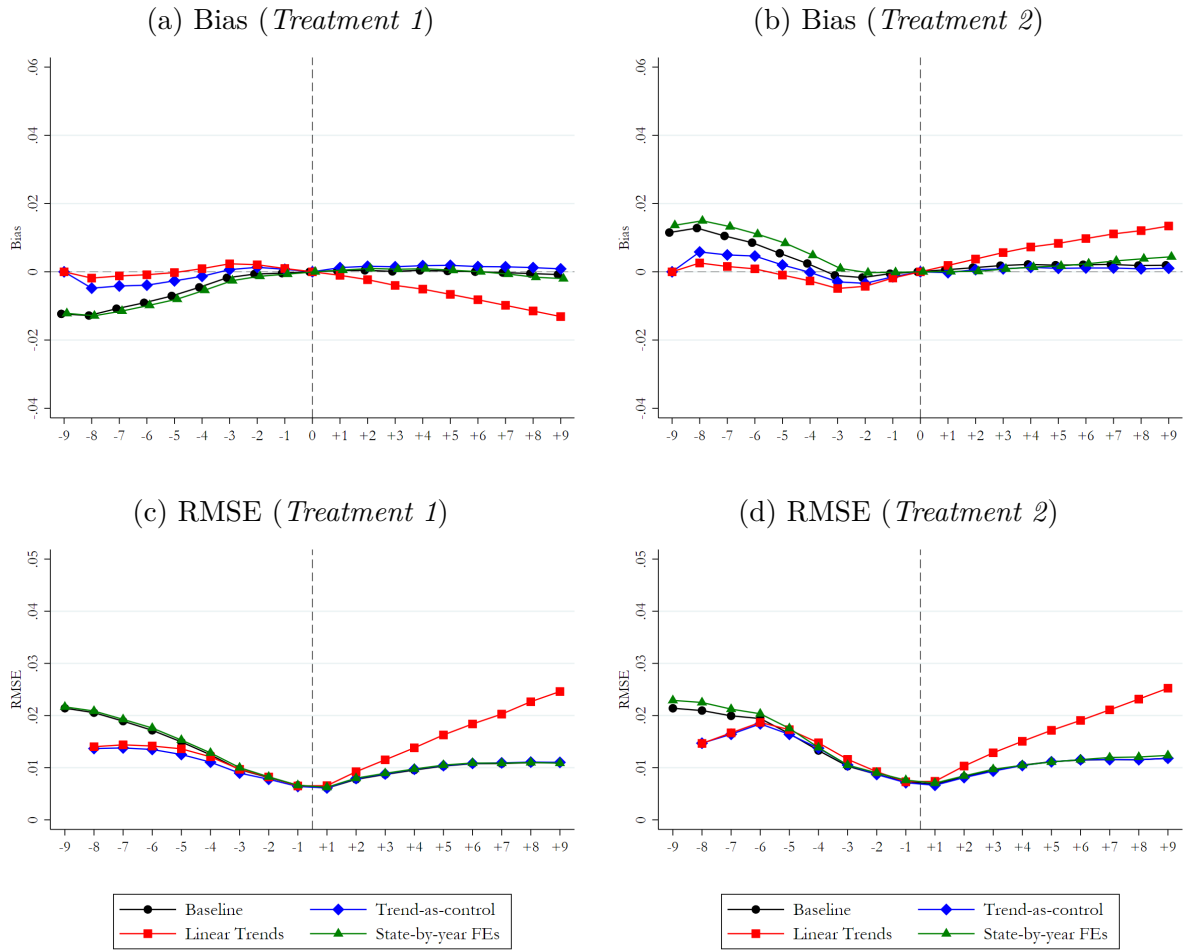(c) RMSE (*Treatment 1*)   (d) RMSE (*Treatment 2*)

Figure D.1: Performance of Main Strategies for Log of Wage

Notes: Summary statistics across 1,000 simulations at the regional (i.e., LLM level), where the outcome is log of average wage. See Section 2 for definition of each estimator and Section 3.2 for treatment definitions.