On the (Mis) Use of the Fixed Effects Estimator[*]

Daniel L. Millimet[†]
Southern Methodist University & IZA

Marc F. Bellemare[‡]
University of Minnesota

November 2, 2025

Word Count: 10459 (Excludes Appendix)

Abstract

Data that span multiple units and time periods allow controlling for time-invariant heterogeneity correlated with the covariates. While researchers can do this in different ways, the fixed effects estimator—also known as the within estimator, and equivalent to the least squares dummy variable approach—has become the default choice. But when time-invariant attributes are not invariant to time—that is, when they are not invariant to the length of the panel—the fixed effects estimator can be considerably biased as researchers incorporate additional time periods. We show that, in finite samples, first-differencing and novel rolling estimators can offer researchers a practical alternative to the fixed effects estimator in this case. These estimators are simple to implement and can significantly reduce bias relative to the fixed effects estimator under certain data-generating processes. Most importantly, researchers should always provide results from multiple estimators. We illustrate this with simulations and four replications.

JEL: C23, C51, C52

Keywords: Panel Data, Fixed Effects, First Differences, Rolling Estimation

[*]We thank Debopam Bhattacharya as well as three anonymous reviewers for thoughtful comments which led to a considerably improved manuscript. We are also grateful to Badi Baltagi, Jeff Bloem, Colin Cameron, Bernhard Dalheimer, Raffaella Giancomini, Paul Glewwe, Hugo Jales, Jason Kerwin, Ivan Korolev, Carlos Lamarche, James MacKinnon, John Mullahy, Chris Muris, Christian Vossler, and Le Wang for comments, as well as to seminar and conference participants at the University of Sydney, Queensland University of Technology, Universidad del Pacífico, the Econometric Society Australasia Meetings, the New York Econometrics Camp, and the Southern Economic Association Meetings. All remaining errors are ours.

[†]Corresponding author. Department of Economics, Box 0496, Southern Methodist University, Dallas, TX 75275-0496, United States. E-mail: millimet@smu.edu. Tel: 214.768.3269.

[‡]Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, Saint Paul, MN 55108, United States. E-mail: mbellema@umn.edu.

"You keep using that word. I do not think it means what you think it means."

– Inigo Montoya, *The Princess Bride* (1987)

1 Introduction

Ever since Mundlak (1961) tried to explain agricultural productivity differentials by including a separate dummy variable for each farmer in his data to eliminate the bias stemming from between-farm differences in "management," the fixed effects (FE) estimator has been used by researchers to control for time-invariant, unit-specific heterogeneity (see, e.g., Balestra and Nerlove, 1966; Bellemare and Millimet, 2025; Hsiao, 2007). Nowadays, researchers continue to rely on FE for causal inference in panel data settings (Imai and Kim, 2019). Imai and Kim (2021, p. 405) even refer to FE as the "default methodology for estimating causal effects with panel data."

As is well known, the advantage of FE over pooled ordinary least squares (POLS) is that by removing both observed *and* unobserved time-invariant, unit-specific heterogeneity, FE requires an arguably more palatable assumption to be unbiased than POLS does. Whereas unbiasedness of FE requires strict exogeneity with respect to the time-varying error, POLS requires exogeneity with respect to a composite error that includes all period- and unit-specific unobserved heterogeneity.

One important limitation of the FE estimator is rarely acknowledged, however: In practice, the amount of time-invariant unobserved heterogeneity *is not invariant to time*. That is, the amount of time-invariant unobserved heterogeneity changes with the length of the panel. While this limitation could be ignored in the past, this is no longer true given the ever-increasing availability of longer panel data sets.

We explore the consequences in finite samples of ignoring this limitation in empirical research and discuss computationally simple alternatives for applied researchers. To do so, we begin by documenting the bias-variance trade-off that arises when the time dimension of a panel increases in the presence of time-varying unit-specific fixed effects. Intuitively, as T increases, the variance of the FE estimator will shrink, but the estimator may become biased due to increasingly likely violation of the strict exogeneity assumption.[1]

We then turn to comparing the FE estimator with a common yet oft-overlooked alternative, viz. the first-difference (FD) estimator. Starting from a general data-generating process, we characterize the conditions under which (i) the FD estimate will be less biased than the POLS estimate, and (ii) the FD estimate will be less biased than the FE estimate. By decomposing the variable of interest into two unit- and time-specific components—one correlated with time-varying unit-specific heterogeneity, and one not—we show that the relative biases of POLS, FE, and FD hinges upon

[1] We are not referring here to $T \to \infty$ in the usual asymptotic sense, where the underlying data-generating process is fixed as T grows. Rather, we are referring to a situation where the data-generating process itself changes as T grows because the definition of the fixed effect changes. We are also not referring to situations where T increases because the researcher uses higher-frequency data for the same length of time (e.g., T goes from 5 to 10 because data over a five-year period are collected semi-annually rather than annually). We restrict attention to cases where T increases in a researcher's sample because additional rounds of data become available, but for which the frequency remains the same.

the degree of serial correlation in these two components. In particular, the more persistent is the unit-specific unobserved heterogeneity, the more likely that FD will have the smallest bias. When the unit-specific unobserved heterogeneity is perfectly persistent (i.e., time invariant), than FE and FD are both unbiased. Strikingly, FE only achieves a smaller bias than FD in the data-generating processes we consider when the unit-specific unobserved heterogeneity is sufficiently less persistent (i.e., time-varying). This stands in stark contrast with current research practice: use FE because the unit-specific unobserved heterogeneity is assumed to be perfectly persistent.[2]

Finally, we present additional estimators that not only improve on FE in certain situations with large T, but are also simple to implement. While the proposed alternatives to FE are inefficient under the assumptions of the standard fixed effects model (including homoskedastic and serially uncorrelated errors), they are more robust to the types of misspecification discussed here under certain data-generating processes as the amount of unobserved heterogeneity removed by these alternatives is (weakly) greater. Our most important conclusion, however, is that practitioners should always provide results from multiple estimators, as differences are indicative of model misspecification.[3]

Beyond the FD estimator, the alternative estimators we discuss include the twice first-differenced (TFD) estimator as well as novel estimators which we dub the rolling first-differenced (RFD) estimator and the rolling twice first-differenced (RTFD) estimator. The RFD estimator proceeds by estimating a series of rolling regressions by FD where only two consecutive time periods are retained at each step. If desired, the individual two-period-specific estimates can be combined into a final summary estimate using minimum-distance estimation (MDE). The RTFD estimator proceeds similarly; a series of rolling regressions by TFD where three consecutive time periods are retained at each step and the estimates are combined if desired using MDE.[4] Our consideration of rolling estimators is analogous to the proposed use of rolling regressions for time series models under local stationarity wherein parameters are assumed to evolve gradually over time and asymptotic theory relies on obtaining data on a finer and finer grid rather than $T \to \infty$ (e.g., Dahlhaus, 1997; Cai and Juhl, 2023). Similarly, Wang *et al.* (2023) propose a similar solution to the problem of forecasting an ultra-long time series. In our case, RFD and RTFD will be superior (in terms of bias) to FE under local time-invariance of the relevant unobserved heterogeneity.

There is also an additional advantage to our rolling estimators. In many situations, not only might there be temporal heterogeneity in the unit fixed effects but in the slope coefficients as well. While this is not our primary focus, our novel RFD and RTFD estimators easily accommodate temporal heterogeneity in both the slope coefficients and the unit fixed effects. Both estimators entail a series of rolling regressions, estimating separate slope parameters for each two- (for RFD) or three- (for RTFD) period interval. Rather than assuming common parameters over time and aggregating

[2]Intuitively, this is similar to Bellemare *et al.* (2017) who show that one can only exogenize an explanatory variable x_{it} by replacing it with its lagged value x_{it-1} when the unobserved confounders u causing both x and the outcome variable y are not serially correlated—what they dub the assumption of "no dynamics among unobservables."

[3]We are not the first to make this point. A similar argument is put forth in Laporte and Windmeijer (2005) and McKinnish (2008) albeit in different contexts than the one considered here.

[4]Alternatively, the rolling estimators can be estimated via the generalized method of moments (GMM) by stacking the relevant moment conditions.

using MDE or GMM, one can examine the various period-specific coefficient estimates and simply not aggregate (i.e., not constrain the coefficients to be the same over time). Moreover, our rolling estimators can incorporate recent methods to allow cross-sectional, group-specific heterogeneity in the the slope parameters in a straightforward manner.

Through simulations we find that FD and RFD both outperform FE when the standard linear panel data model is misspecified and the unit-specific unobserved heterogeneity is sufficiently persistent, but generally not when it is correctly specified. Thus, the estimators we discuss may sacrifice (some) efficiency for (much) robustness when the unit-specific unobserved heterogeneity is likely to be highly persistent, but not time invariant. This is also the case in Aquaro and Čížek (2013) when dealing with outliers, and in Pesaran and Zhou (2018) when considering the asymptotic efficiency of POLS relative to FE in the presence of sparse unit fixed effects. Moreover, we find that the RFD estimator outperforms the FD estimator in some settings. Finally, we find that other estimators—TFD, RTFD, and the interactive fixed effects (IFE) estimator from Bai (2009)—work well in some situations. Through replications of Rose (2004), Tomz *et al.* (2007), Leipziger (2024), James (2015), and Djankov and Reynal-Querol (2010), we then find that the choice of estimator can matter greatly in practice. Overall, we recommend that researchers report FD, RFD, TFD, RTFD, and IFE estimates in addition to FE when using panel data with more than two time periods.

Our contribution is thus fourfold. First, we clarify the trade-off involved with ever-longer panel data when using the standard FE estimator—a trade-off that pits efficiency gains against potentially greater bias as the length of a panel increases. Second, we characterize the conditions under which the FE estimator will perform relatively worse than the FD estimator in terms of bias—the conditions, as it were, that can lead applied researchers to misuse the FE estimator. Third, we provide guidance to empirical researchers regarding simple alternatives to the mechanical application of the FE estimator in static TWFE models, namely the TFD, RFD, and RTFD estimators.[5] Fourth, through simulations and the replication of four studies, we find that the FD, TFD, RFD, RFE, and IFE estimators will frequently lead to conclusions that are qualitatively different from those obtained using FE.

Lest there be confusion, our focus in this paper is *not* in difference-in-differences (DID) designs. Moreover, the time-varying heterogeneity we are concerned with is not synonymous with the heterogeneity at issue with staggered treatment adoption as introduced in the recent DID literature. While the number of studies relying on DID is immense (and, seemingly, never-ending), our focus is on the simpler case—as in Mundlak (1961)—wherein a researcher relies on unit-specific fixed effects to purge a regression error term of its correlation with all right-hand side variables included in the regression, and not just a (binary) treatment.

In particular, for expositional purposes, we consider the case of a single continuous, real-valued right-hand side variable. Thus, we do not rely on the potential outcomes model. But should one be inclined to do so, one could easily cast our analysis within the potential outcomes framework as

[5] Ishimaru (2025) shows that the two-way fixed effects estimator is equivalent to an FD estimator pooling all possible period-pair differences (what he refers to as "between-period gaps"). Our RFD estimator only uses one-period differences, and therefore is an improvement.

one can view our analysis as focusing on the specification of potential outcomes in a panel context where the intercept for the potential outcomes may not be time invariant (e.g., Liu *et al.*, 2024).[6] Throughout the paper, we briefly discuss our work in relation to the DID and broader two-way fixed effects (TWFE) literatures where necessary.

The remainder of this paper is organized as follows. Section 2 discusses the relevant econometric literature. Section 3 introduces the static linear panel data setup, clarifies what is at stake as T increases, and compares FE, FD, RFD, and others in the presence of time-varying unit-specific fixed effects as well as temporal or group-specific heterogeneity in the slopes. In section 4, we illustrate these estimators via simulation and four replications. Section 5 concludes.

2 Literature Review

We are not the first to consider time-varying unit-specific fixed effects, particularly as T increases. In fact, the point was explicitly made in Mundlak (1961, p. 44):[7]

> "Instead of beginning by conceptualizing what we mean by management we shall assume that whatever management is, it does not change considerably over time; and for short periods, say a few years, it can be assumed to remain constant."

and in Mundlak (1978, p. 82):

> "[I]t would be unrealistic to assume that the individuals do not change in a differential way as the model assumes ... [I]t is more realistic to assume that individuals do change differentially but at a pace that can be ignored for short time intervals."

Similarly, Plümper and Troeger (2019, p. 21) (emphasis added) write:

> "Our findings caution applied researchers to not overlook the potential drawbacks of relying on the fixed effects estimator as a default ... [S]cholars ought to devote much more attention to modeling dynamics appropriately *instead of relying on a default solution before they control for potentially omitted variables with constant effects using a fixed effects specification*."

Yet, the point made by these authors is often forgotten by researchers despite the growing availability of data sets with large T. McKenzie (2012) even advocates for "more T" when researchers collect their own experimental data. The fact that some or all of what is time-invariant when, say, $T = 5$ may no longer be time-invariant when $T = 10$ must be acknowledged. Nevertheless, researchers typically neither alter their interpretation of the fixed effects nor acknowledge the ever-stronger assumption required for unbiasedness as T grows. Hill *et al.* (2020, p. 363) state:

[6] For example, with a multi-valued treatment, $d_{it} \in \{0, 1, ..., D\}$ for unit $i = 1, ..., N$ at time $t = 1, ..., T$, potential outcomes may be specified as $Y_{it}(d) = \alpha_{di} + x_{it}\beta_d + \varepsilon_{it}$, $d = 0, 1, ..., D$. Our focus is on α_{di}, the unit- and potential outcome-specific intercept, and what it captures as T increases.

[7] See Bellemare and Millimet (2025) for a historical account of Mundlak's contribution to panel data econometrics.

> "[W]e often assume (without direct confirmation) that certain characteristics (e.g., biology, personality, or culture) do not change over time ... [I]t is often unclear whether some characteristics are actually fixed or variable. So what exactly are fixed-effects models controlling? ... [T]he answer to this question is often ambiguous. Uncertainty along these lines could introduce a host of theoretical and empirical problems."

Worse, the issue also arises when relevant attributes remain time-invariant as T grows but the effects of these attributes are time-varying.[8] Hill *et al.* (2020, p. 364) continue:

> "Researchers typically state and restate how fixed-effects models control for time-invariant characteristics, but this is only true if those variables have the same effects at each point in time. If the coefficients for supposed time-invariant characteristics vary with time, they become equivalent to time-varying characteristics."

Wooldridge (2010, p. 282) makes the same point, but in a more understated manner:

> "The assumption that [the fixed effect] is constant over time, and has a constant partial effect over time, is crucial..."

As stated previously, we are not the first to consider time-varying unit-specific fixed effects, and there is small literature on this topic. But existing models in this literature impose restrictions on how the unit fixed effects vary over time. Moreover, these models have yet to gain traction among applied researchers, presumably in part due to their computational intensity. Mundlak (1978), for his part, introduces unit-specific time trends into the standard panel data model; a practice that is now widespread (e.g., Autor, 2003). More recently, IFE models allow for common time-varying factor loadings on the unit-specific fixed effects (e.g., Ahn *et al.*, 2001, 2013; Bai, 2009). This allows the time-invariant, unit-specific heterogeneity to have homogeneous but time-varying effects as well as time-invariant, unit-specific responses to common period-specific shocks. Gobillon and Magnac (2016, p. 537) show that a "single-dimensional additive local effect as in the setup underlying difference-in-differences estimation" will generally be biased in the presence of more complex unobserved heterogeneity. While higher-dimensional factors are possible, the computational intensity increases greatly. Nevertheless, IFE does relax the traditional FE estimator in a particular way.

Another class of models extends the standard specification by allowing for (unknown) group membership or structural breaks. Bonhomme and Manresa (2015) allow for time-varying, group-specific effects. This entails aggregating units into a few large groups and then incorporating time-varying group-specific fixed effects. Papke and Wooldridge (2023) consider the case of time-invariant group-specific fixed effects with a focus on testing whether unit or group fixed effects are more appropriate. Li *et al.* (2016) use LASSO to estimate an IFE model with multiple structural breaks in the slope coefficients. Boldea *et al.* (2020) examine a model with multiple structural breaks

[8] We use the term "relevant" to refer to unobserved heterogeneity that is not independent of the right-hand side variables.

where each regime has unique slope parameters and unit fixed effects. Lumsdaine *et al.* (2023) also allow for multiple structural breaks, where units are aggregated into groups with common slope coefficients and fixed effects before and after the break date. Moreover, their setup permits units to change groups at the structural break. Kaddoura and Westerlund (2023) extends the IFE model to allow for structural breaks in the slope coefficients. Wang *et al.* (2024) combines Lumsdaine *et al.* (2023) and Kaddoura and Westerlund (2023) where multiple structural breaks in the slope coefficients are allowed, units are aggregated into groups with common slope coefficients before and after the break dates, units may change groups at the structural break, and the intercepts follow an IFE model.

While these studies improve upon FE, they have a practical limitation. Specifically, models based on structural breaks assume a common break date for all units for estimation to be feasible as estimation typically relies on a grid search over all possible break dates. Assuming a common break date, this entails searching over $T-1$ possible break dates. If each unit $i = 1, ..., N$ is allowed a unique break date, this entails searching over $(T-1)^N$ possible combinations. Even allowing for multiple, common structural breaks can quickly become computationally intensive.

The bias-efficiency trade-off that emerges as T increases when unit fixed effects capture less unobserved heterogeneity also relates to other literatures exploring failures of the standard linear panel data model. Bramati and Croux (2007) and Aquaro and Čížek (2013) discuss the bias of FE in the presence of outliers and consider robust alternatives, particularly ones using temporal differencing. Ignoring time variation in the unit-specific fixed effects may behave similarly to the outliers considered in this literature.

Additionally, many studies examine heterogeneity of various forms in the slope coefficients. Sarafidis and Weber (2015) and Gibbons *et al.* (2019) consider a panel data model with group-specific slope coefficients. Campello *et al.* (2019) considers the specification testing in a model with unit-specific coefficients and time-invariant unit fixed effects. Cai and Juhl (2023) consider a time series model with time-varying slope coefficients and, using rolling regressions, develop a test for such heterogeneity. Keane and Neal (2020) allow for spatial and temporal slope heterogeneity and propose a new estimation algorithm referred to as the mean observation OLS procedure. See also Su and Chen (2013), Ando and Bai (2016), Baltagi *et al.* (2016), Qian and Su (2016), Su *et al.* (2016), Liu *et al.* (2020), Okui and Wang (2021), Mehrabani (2023), among others.[9]

We contribute to this literature by (i) clarifying the trade-off involved with ever-longer panel data when using FE as well as (ii) providing guidance to empirical researchers concerning simple alternatives.

3 Panel Data Estimation

Our objective in this section is twofold. First, we want to make abundantly clear the role that T plays in the critical assumption required for unbiased and consistent estimation of linear panel data

[9] Sun and Shapiro (2022) discusses slope heterogeneity in the context of linear panel data models assessing treatment exposure and connects this to the recent literature on difference-in-differences with staggered adoption.

models with unobserved heterogeneity. Second, when this assumption fails, we want to emphasize that the choice of estimator can be consequential.

To proceed, in Section 3.1 we present the usual panel setup and the popular FE estimator. We demonstrate how the strict exogeneity assumption becomes more demanding as T increases. When the FE estimator is biased due to a failure of the strict exogeneity assumption in the presence of time-varying, unit-specific unobserved heterogeneity, alternative estimators may perform better. In Section 3.2 we theoretically compare several popular linear panel data model estimators. Our goal is to focus on estimators that are simple and general—and thus more likely to be adopted by applied researchers. In Section 3.3 we discuss differencing estimators in more detail and also present a novel estimation algorithm based on rolling regressions. Section 3.4 briefly discusses the incorporation of slope heterogeneity. Section 3.5 presents two simple tests of strict exogeneity. Finally, in Section 3.6 we discuss interactive fixed effects estimators.

3.1 Setup

Consider the following data-generating process (DGP):

$$y_{it} = x_{it}\beta + \alpha_{it} + \varepsilon_{it}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \tag{1}$$

where $i \in \mathcal{N}$, $\mathcal{N} = \{1, ..., N\}$ denotes cross-sectional units, $t \in \mathcal{T}$, $\mathcal{T} = \{1, ..., T\}$ denotes time periods, x_{it} is a $1 \times K$ vector of time-varying observed attributes (which may include period fixed effects), α_{it} is a possibly time-varying, unit-specific fixed effect, and ε_{it} is a mean-zero error.[10] Researchers often arrive at Equation (1) using the potential outcomes framework, where x includes a binary treatment indicator, and the corresponding coefficient is the average treatment effect on the treated. Equation (1) is also the standard DID model with non-staggered adoption (e.g., Gobillon and Magnac, 2016). With staggered adoption, or other sources of treatment effect heterogeneity, the assumption of homogeneous slopes, β, is rejected. We return to this in Section 3.4. For simplicity, we start with the model in Equation (1) as our primary focus is on the role of α_{it}.

As β is not identified in Equation (1), researchers assume $\alpha_{it} = \alpha_i$ for all (i, t) in the majority of panel data studies. In this case, α_i is described as capturing all time-invariant attributes for unit i, whether observed or unobserved, and ε_{it} as capturing all time-varying, unobserved determinants of y (that vary across units if x includes period fixed effects). Our focus here is on what exactly is meant by "time-invariant" and "time-varying" in this context—terms that are often casually (if not mechanically) invoked by researchers when claiming to obtain an unbiased estimate of β.

Estimation of Equation (1) under the assumption that $\alpha_{it} = \alpha_i$ for all i is most often accomplished using the FE estimator, also known as the within estimator and equivalent to the least

[10] It is standard practice by researchers to include time fixed effects. We allow x_{it} to subsume them to save on notation since the time effects do not play an explicit role in our analysis. Equivalently, one can re-write the DGP in Equation (1) as $\widetilde{y}_{it} = \lambda_t + \widetilde{x}_{it}\beta + \alpha_{it} + \widetilde{\varepsilon}_{it}$ and then define $z_{it} := \widetilde{z}_{it} - \overline{\widetilde{z}}_{\cdot t}$, $z \in \{y, x, \varepsilon\}$. In this case, the DGP in Equation (1) can be viewed as a transformed model—after cross-sectionally mean-differencing—to remove the time fixed effects, which is an application of the Frisch-Waugh-Lovell Theorem. The reader should interpret our statements about x_{it} as implicit statements about $\widetilde{x}_{it} - \overline{\widetilde{x}}_{\cdot t}$.

squares dummy variables (LSDV) estimator (Mundlak, 1961). This entails using POLS after mean-differencing to estimate the model given by

$$\begin{aligned} y_{it} - \overline{y}_i &= (x_{it} - \overline{x}_i)\,\beta + (\varepsilon_{it} - \overline{\varepsilon}_i) \\ \ddot{y}_{it} &= \ddot{x}_{it}\beta + \ddot{\varepsilon}_{it}, \end{aligned} \tag{2}$$

where two dots over a variable indicate deviation from its unit-specific mean. Given a random sample, $\{y_{it}, x_{it}\}_{(i,t)} \in \mathcal{N} \times \mathcal{T}$, the FE estimator requires $\mathtt{E}[\varepsilon_{it}|\mathbf{x}_i, \alpha_i] = 0$ for all t for unbiasedness, where $\mathbf{x}_i$ is a $T \times K$ matrix of covariates for unit i. This condition, known as strict exogeneity, implies that x_{it} is independent of the time-varying error term ε_{is}, in every period $s = 1, ..., T$ conditional on α_i.[11]

The presumption often made by applied researchers is that larger T is "better" than smaller T. In studies with $N >> T$, which is typically the case in applied microeconomics, asymptotic results rely on $N \to \infty$ with T fixed. As such, the only advantage of increasing T is efficiency; sample size grows by N with each additional time period assuming a balanced panel, yet only one degree of freedom is lost if time effects are included (Hsiao, 2007).[12]

An overlooked consequence of larger T, however, is how it impacts the strict exogeneity assumption.[13] To formalize the role of T, we introduce some additional notation. Define $\alpha_i^{t,t'}$ as unobserved attributes of unit i that are time-invariant over the period spanning t to t', where $t, t' \in \mathcal{T}$, $t \geq 1$, $t' \leq T$, and $t < t'$. Without loss of generality, we can rewrite the time-varying, unit-specific fixed effect in Equation (1) as

$$\alpha_{it} \equiv \alpha_i^{1,T} + \eta_{it}^{1,T}, \tag{3}$$

where $\eta_{it}^{1,T}$ captures the deviation between α_{it} and $\alpha_i^{1,T}$. Substituting Equation (3) into Equation (1), the strict exogeneity assumption required by the FE estimator becomes $\mathtt{E}\left[\eta_{it}^{1,T} + \varepsilon_{it}|\mathbf{x}_i, \alpha_i^{1,T}\right] = c$, $t = 1, ..., T$, for some scalar c.[14]

Suppose a researcher obtains access to additional periods of data such that now the sample includes $t \in \mathcal{T}'$, $\mathcal{T}' = \{1, ..., T'\}$ and $T' > T$. As a result of the longer panel, the DGP becomes

$$y_{it} = x_{it}\beta + \alpha_i^{1,T'} + \eta_{it}^{1,T'} + \varepsilon_{it}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T}'.$$

The strict exogeneity assumption required by the FE estimator becomes $\mathtt{E}\left[\eta_{it}^{1,T'} + \varepsilon_{it}|\widetilde{\mathbf{x}}_i, \alpha_i^{1,T'}\right] = c$, $t = 1, ..., T'$, where $\widetilde{\mathbf{x}}_i$ is a $T' \times K$ matrix of covariates for unit i. Critically, this is a stronger condition than that required in the original sample period $\mathcal{T}$ for two reasons. First, and most

[11] In contrast, estimation of Equation (1) by POLS requires (contemporaneous) exogeneity, given by $\mathtt{E}[\alpha_{it} + \varepsilon_{it}|x_{it}] = 0$ for all t (or $\mathtt{E}[\alpha_i + \varepsilon_{it}|x_{it}] = 0$ for all t if the assumption that $\alpha_{it} = \alpha_i$ is invoked).

[12] In dynamic panels, there are gains to larger T beyond efficiency as the bias of FE estimates diminishes with T (Nickell, 1981).

[13] Wooldridge (2010) provides a detailed discussion of the exogeneity assumptions required by POLS, FE, FD, and the Random Effects (RE) estimator.

[14] Note, we do not impose $\mathtt{E}\left[\alpha_i^{1,T}\right] = 0$ or $\mathtt{E}\left[\eta_{it}^{1,T}\right] = 0$. As a result, strict exogeneity holds if $\mathtt{E}\left[\eta_{it}^{1,T} + \varepsilon_{it}|\mathbf{x}_i, \alpha_i^{1,T}\right]$ is constant. If the constant is not assumed to be zero, then one should augment the mean-differenced model in Equation (2) with an intercept. We return to this below.

importantly, the conditioning set changes from $\alpha_i^{1,T}$ to $\alpha_i^{1,T'}$. Specifically, as T increases, it is likely that some previously unobserved, time-invariant attributes now become time-varying. Or, that that the effects of time-invariant attributes change over the sample period. In either case, $\alpha_i^{1,T'}$ will capture fewer attributes than $\alpha_i^{1,T}$, such that $\alpha_i^{1,T'} \subset \alpha_i^{1,T}$. For instance, Murayama and Gfrörer (forthcoming, p. 2) state: "If participants were repeatedly assessed over three months, for example, any characteristics that are stable for this duration can be viewed as time-invariant confounders, even if they may change for a longer period."

Second, the conditioning set changes from $\mathbf{x}_i$ to $\widetilde{\mathbf{x}}_i$, implying that covariates from the additional time periods must also be strictly exogenous. This makes it clear that researchers must balance the efficiency gain from larger T against the stronger strict exogeneity assumption needed for unbiasedness.[15]

Remark 1. *With sufficiently large T, researchers must think critically about what unobserved heterogeneity is time invariant, both the heterogeneity itself and the return to that heterogeneity. The definition of 'sufficiently' depends on the context.*

3.2 Failure of Strict Exogeneity

When the unobserved heterogeneity, α, is not time invariant, x is no longer strictly exogenous even if ε is white noise (WN) unless $\texttt{Cov}(x_{it}, \eta_{is}^{1,T}) = 0$ for all t and s. When this condition does not hold, the FE estimator is biased and inconsistent. It is possible that alternative estimator(s) may achieve a smaller bias (in finite samples and/or asymptotically), however. Before proceeding, it is important to note that our analysis in this paper considers a particular violation of strict exogeneity. While we believe this type of violation to be an important and frequent one in empirical research, other violations of strict exogeneity may lead to different conclusions than those presented here. We return to this in Section 3.3. Nonetheless, the larger point remains: alternatives to the FE estimator may achieve a smaller bias when strict exogeneity is violated.

[15] The same potentially holds for changes in N. Increasing N implies that period fixed effects, if they are included in the model, likely control for fewer period-specific attributes. There is a distinction between increasing N versus increasing T, however, as the former entails drawing additional observations from the same underlying population while the latter entails expanding the population to include additional time periods. We leave this discussion for future research.

To continue, we augment the DGP in Equation (1) with the following structure

$$\begin{aligned}
x_{it} &= \lambda\alpha_{it} + z_{it} \\
\alpha_{it} &= \delta\alpha_{it-1} + \varepsilon_{\alpha,it} \\
z_{it} &= \rho z_{it-1} + \varepsilon_{z,it} \\
\varepsilon_{it} &\sim WN(0, \sigma^2_{\varepsilon}) \\
\varepsilon_{\alpha,it} &\sim WN(0, \sigma^2_{\varepsilon_\alpha}) \\
\varepsilon_{z,it} &\sim WN(0, \sigma^2_{\varepsilon_z}) \\
\mathtt{Cov}\left(z_{it}, \varepsilon_{is}\right) = \mathtt{Cov}\left(z_{it}, \varepsilon_{\alpha,is}\right) &= 0 \quad \forall\, t, s \\
\mathtt{Cov}\left(\varepsilon_{\alpha,it}, \varepsilon_{is}\right) = \mathtt{Cov}\left(\varepsilon_{\alpha,it}, \varepsilon_{z,is}\right) = \mathtt{Cov}\left(\varepsilon_{it}, \varepsilon_{z,is}\right) &= 0 \quad \forall\, t, s
\end{aligned} \tag{4}$$

and focus on the case where $K = 1$ so that x_{it} is a scalar for simplicity.[16] Here, x is decomposed into two parts, one capturing the unobserved heterogeneity and the other, denoted by z, capturing all other determinants of x. The coefficient λ allows α to have different effects on x and y. The persistence in α and z are modeled as AR(1) processes, where δ and ρ, respectively, reflect the persistence. All error terms are white noise. As such, z is strictly exogenous with respect to α and ε, but x is strictly exogenous with respect to only ε. We assume that $|\delta|, |\rho| < 1$ such that stationarity holds.[17]

The DGP in Equations (1) and (4) likely characterizes many real world applications. First, it nests the usual linear panel data model with only time invariant heterogeneity; set $\delta = 1$ and $\sigma^2_{\varepsilon_\alpha} = 0$. Second, while modeling persistence as an AR(1) process places specific structure on the data-generating process, it captures the essence of Mundlak (1961, 1978) and others that the unobserved heterogeneity envisioned by researchers is thought to evolve slowly over time. It is also consistent with most applications of dynamic panel data models that model the outcome as following an AR(1) process. Third, it is nearly identical to the DGP in McKinnish (2008) where z is interpreted as the unobserved 'true' regressor, x its observed counterpart, and α reflecting measurement error.

We wish to compare the bias of the three most common estimators of linear panel data models: POLS, FD, and FE.[18] The POLS estimator is obtained by estimating

$$y_{it} = \beta x_{it} + \widetilde{\varepsilon}_{it} \tag{5}$$

[16] The model can be derived from a multiple regression where the other covariates—and time fixed effects—have been partialled out, invoking the Frisch-Waugh-Lovell Theorem. One must be careful to interpret the results of our analysis in terms of the residualized y and x, however.

[17] Under this structure, x_{it} follows an ARMA(2,1) process and is also stationary if $|\delta|, |\rho| < 1$. Proof is provided in Appendix A.2.

[18] All proofs are provided in Appendix A.3 and A.4.

by OLS where $\widetilde{\varepsilon}_{it} \equiv \alpha_{it} + \varepsilon_{it}$. The probability limit of $\widehat{\beta}$ is

$$\texttt{plim}\,\widehat{\beta} = \frac{\texttt{Cov}(y,x)}{\texttt{Var}(x)} = \beta + \frac{\lambda\texttt{Var}(\alpha)}{\texttt{Var}(x)} = \beta + \left[\frac{\lambda\sigma^2_{\varepsilon_\alpha}}{1-\delta^2}\right]\left[\frac{\lambda^2\sigma^2_{\varepsilon_\alpha}}{1-\delta^2} + \frac{\sigma^2_{\varepsilon_z}}{1-\rho^2}\right]^{-1}. \tag{6}$$

The bias is proportional to the share of the variance of x that is due to α.

The FD estimator is obtained by estimating

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta\varepsilon_{it}, \tag{7}$$

by OLS where Δ is the first-difference operator. When $T = 2$ the FD estimator is algebraically identical to the FE estimator. With $T > 2$ FE and FD differ, although both are unbiased and consistent under strict exogeneity. The probability limit of $\widehat{\beta}_{fd}$ is

$$\texttt{plim}\,\widehat{\beta}_{fd} = \frac{\texttt{Cov}(\Delta y,\Delta x)}{\texttt{Var}(\Delta x)} = \beta + \frac{\lambda\texttt{Var}(\Delta\alpha)}{\texttt{Var}(\Delta x)} = \beta + \left[\frac{\lambda\sigma^2_{\varepsilon_\alpha}}{1+\delta}\right]\left[\frac{\lambda^2\sigma^2_{\varepsilon_\alpha}}{1+\delta} + \frac{\sigma^2_{\varepsilon_z}}{1+\rho}\right]^{-1}. \tag{8}$$

The bias is now proportional to the share of the variance of Δx that is due to $\Delta\alpha$.

Finally, for FE, the probability limit of $\widehat{\beta}_{fe}$ is

$$\begin{aligned}
\texttt{plim}\,\widehat{\beta}_{fe} &= \frac{\texttt{Cov}(\ddot{y},\ddot{x})}{\texttt{Var}(\ddot{x})} = \beta + \frac{\lambda\texttt{Var}(\ddot{\alpha})}{\texttt{Var}(\ddot{x})} && (9)\\
&\approx \beta + \left\{\frac{\lambda\sigma^2_{\varepsilon_\alpha}}{1-\delta^2}\left[1 - \frac{1}{T}\left(\frac{1+\delta}{1-\delta} - \frac{2\delta(1-\delta^T)}{T(1-\delta)^2}\right)\right]\right\} && (10)\\
&\qquad \left\{\frac{\lambda^2\sigma^2_{\varepsilon_\alpha}}{1-\delta^2}\left[1 - \frac{1}{T}\left(\frac{1+\delta}{1-\delta} - \frac{2\delta(1-\delta^T)}{T(1-\delta)^2}\right)\right] + \right.\\
&\qquad\qquad \left.\frac{\sigma^2_{\varepsilon_z}}{1-\rho^2}\left[1 - \frac{1}{T}\left(\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^T)}{T(1-\rho)^2}\right)\right]\right\}^{-1}.
\end{aligned}$$

The bias is proportional to the share of the variance of $\ddot{x}$ that is due to $\ddot{\alpha}$ and is an approximation based on large T, which is acceptable since the motivation for our analysis is that the unobserved heterogeneity becomes time-varying with large T. Note, as T becomes very large, $\texttt{plim}\,\widehat{\beta}_{fe} \approx \texttt{plim}\,\widehat{\beta}$ since $\texttt{Var}(\ddot{\alpha}) \approx \texttt{Var}(\alpha)$. For smaller T, $\texttt{Var}(\ddot{\alpha}) < \texttt{Var}(\alpha)$ assuming $\delta \in (0,1)$, with the reduction increasing in δ. If $\delta = 0$, implying that α_{it} is white noise, then $\texttt{Var}(\ddot{\alpha})$ and $\texttt{Var}(\alpha)$ are equivalent.

The relative magnitude of the biases (in absolute value) depend crucially on δ and ρ and are summarized in the following proposition.

Proposition 1. *If $\delta > \rho$, implying that α is more persistent than z, then $|Bias(\widehat{\beta})| \gtrsim |Bias(\widehat{\beta}_{fe})| > |Bias(\widehat{\beta}_{fd})|$. If $\delta < \rho$, implying that z is more persistent than α, then $|Bias(\widehat{\beta}_{fd})| > |Bias(\widehat{\beta}_{fe})| \gtrsim |Bias(\widehat{\beta})|$. If $\delta = \rho$, then $|Bias(\widehat{\beta})| = |Bias(\widehat{\beta}_{fe})| = |Bias(\widehat{\beta}_{fd})|$.*

Proof See Appendix A.3 and A.4. ■

Thus, when strict exogeneity fails because the unobserved heterogeneity is time-varying and follows an AR(1) process, the FE estimator is *never* optimal among this set of estimators in terms of minimizing the bias. With large T, however, the bias of POLS and FE will be quite close. Whether this ranking extends to other models of persistent, but time-varying, unobserved heterogeneity is unclear and left for future research. The AR(1) assumption, however, seems like a reasonable approximation as discussed above. Moreover, researchers currently assume that α is so persistent that it is time invariant, implying that the case where $\delta > \rho$ is apt to be more common, making FD the bias-minimizing estimator among the set. A similar finding appears in McKinnish (2008) for the case where the covariate suffers from serially correlated measurement error; see also Wooldridge (2010, Section 11.5). If the measurement error is more persistent than the covariate, then FD minimizes the bias.[19]

Remark 2. *If strict exogeneity fails due to time-varying unobserved heterogeneity as posited in Equation (4), researchers must think critically about the relative persistence of the unobserved heterogeneity and the covariates net of the unobserved heterogeneity. While the ranking of POLS, FD, and FE depends on which is more persistent, FE is never optimal in terms of minimizing the bias.*

3.3 Differencing Estimators in More Detail

To appreciate why the FD estimator is less biased than the FE estimator when the time-varying, unit-specific heterogeneity is more persistent than the remaining variation in the covariates, consider the following estimating equation

$$y_{it} = x_{it}\beta + \alpha_i^{s,s'} + \eta_{it}^{s,s'} + \varepsilon_{it}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{S} \tag{11}$$

where $\mathcal{S} = \{s, ..., s'\}$ and $s < s'$. This is identical to Equation (1) except for a generic period defined by $\mathcal{S}$. Given the definition of $\alpha_i^{s,s'}$, the unit fixed effect will capture (weakly) fewer attributes and $\eta_{it}^{s,s'}$ (weakly) more attributes as $s' - s \to \infty$. Thus, in our setup, setting $s' = s + 1$ yields an unbiased and consistent estimate of β under a weaker strict exogeneity condition than when $s' > s+1$. Moreover, with only two time periods, s and s', the FE and FD estimators are identical. This is a simple yet powerful insight that is typically overlooked in empirical research. Retaining only periods s and $s+1$, the FD estimator applies OLS to the following estimating equation

$$\Delta y_{i,s+1} = \Delta x_{i,s+1}\beta + \Delta\eta_{i,s+1}^{s,s+1} + \Delta\varepsilon_{i,s+1}. \tag{12}$$

Unbiasedness requires $\mathtt{E}\left[\Delta\eta_{i,s+1}^{s,s'} + \Delta\varepsilon_{i,s+1} | x_{i,s}, x_{i,s+1}\right] = c$.[20]

When this strict exogeneity condition fails, however, the bias depends on the share of the variance of Δx that is due to $\Delta\eta$. On the one hand, this share will increase if the portion of x that is orthogonal to α (such as z in Section 3.2) is more persistent than the part that is not orthogonal. In this case, while first-differencing removes more unobserved heterogeneity than mean-differencing

[19] McKinnish (2008) did not consider POLS, but did analyze long-differences along with FD and FE.

[20] As discussed above, since there is no reason to assume $\mathtt{E}\left[\eta^{s,s+1}\right] = 0$, one should augment the first-differenced model in Equation (12) with an intercept.

(or longer differences), it removes even more of the strictly exogenous variation in x. As such, something akin to the signal-to-noise ratio falls. Here it is the "exogenous signal-to-time-varying unobserved heterogeneity ratio" that declines. In such cases, the bias is minimized using the longest difference (i.e., $y_{iT} - y_{i1}$) and the bias of POLS and FE fall somewhere in between that of the longest difference and the first difference. We do not place much confidence in the longest difference estimator, however, since we are concerned with applications where T is large, thus making the efficiency loss from only using two time periods quite severe. Nonetheless, it is an estimator that researchers might consider, particularly if N is large.[21]

On the other hand, the bias will decrease if the portion of x that is orthogonal to α (such as z in Section 3.2) is less persistent than the part that is not orthogonal. In this case, first-differencing removes more unobserved heterogeneity than strictly exogenous variation in x, reducing the bias. The longest difference will have the largest bias and, again, the bias of POLS and FE fall somewhere in between that of the first difference and longest difference.

When first-differencing improves the "exogenous signal-to-time-varying unobserved heterogeneity ratio," it would be inefficient to use only two time periods—as suggested by Equation (12)—if a researcher has access to a panel of length T. Stacking (12) using $s = 1, ..., T-1$ and estimating the model using POLS yields the FD estimator applied to the full sample.

As an alternative to stacking the estimating equations and performing a single estimation, we introduce a new estimator which we refer to as rolling first-differences (RFD). The estimator can be implemented in two ways. First, one can separately estimate Equation (12) for $s = 1, ..., T-1$ with $s' = s+1$. This yields $T-1$ estimates of β, denoted by $\widehat{\beta}_{FD}(s, s+1)$, where $(s, s+1)$ refers to the sample periods used in the estimation. A final estimate can be obtained via model averaging

$$\widehat{\beta}_{RFD} = \sum_{s=1}^{T-1} \nu_s \widehat{\beta}_{FD}(s, s+1), \tag{13}$$

where ν_s is the weight given to $\widehat{\beta}_{FD}(s, s+1)$ and the weights sum to one. See Algorithm 1.

The estimator in Equation (13) is an example of MDE. The preferred approach to combining the $T-1$ estimates uses the optimal weight matrix (Wooldridge, 2010, Section 14.5). In this case, ν_s is a function of the variances and covariances of the individual estimates. In Algorithm 1 we simplify this computation by ignoring the covariances (see, e.g. Mullahy, 2016).[22] Moreover, an anonymous reviewer also pointed out that the FD estimator is a special case of the rolling FD when the two-period FD estimators have equal variance.

The second approach to implementation of RFD is to stack the $T-1$ equations and estimate

[21] Hahn *et al.* (2007) propose the use of the longest difference estimator in the context of dynamic panel data models with high persistence in the dependent variable as it minimizes the bias associated with weak instruments.

[22] Alternatively, one might use a clustered bootstrap to obtain the empirical variances and covariances of the estimates.

the model by GMM. The moment conditions are given by

$$\begin{aligned} \mathtt{E}\left[\Delta x_{12}'\left(\Delta y_{i2} - \Delta x_{i2}\beta\right)\right] &= 0 \\ &\vdots \\ \mathtt{E}\left[\Delta x_{1T}'\left(\Delta y_{iT} - \Delta x_{iT}\beta\right)\right] &= 0 \end{aligned} \tag{14}$$

This allows the usual GMM inference and properties to apply.

Algorithm 1 Rolling First Differences (RFD) Estimator

1: **while** $s = 1, ..., T-1$ **do**

2: Apply the FD estimator (with or without an intercept in the differenced equation) using data from two periods, $\{y_{it}, x_{it}\}_{\forall (i,t)\in \mathcal{N}\times\mathcal{S}}$, $\mathcal{S} = \{s, s+1\}$.

3: Collect the coefficient estimates and robust variance-covariance matrix: $\widehat{\beta}_{FD}(s, s+1)$, $\widehat{\Sigma}_{\beta_{FD}}(s, s+1)$, where $(s, s+1)$ refers to the sample periods used in the estimation.

4: **end while**

5: Estimate β using the minimum distance estimator

$$\hat{\beta}_{RFD} = \arg\min_{\beta} \quad \left(\hat{\boldsymbol{\beta}} - \beta\right)' W^{-1} \left(\hat{\boldsymbol{\beta}} - \beta\right),$$

where

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_{FD}(1,2) \\ \vdots \\ \widehat{\beta}_{FD}(T-1,T) \end{bmatrix}; \quad W = \begin{bmatrix} \widehat{\Sigma}_{\beta_{FD}}(1,2) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \widehat{\Sigma}_{\beta_{FD}}(T-1,T) \end{bmatrix}.$$

Or, alternatively, estimate via OLS

$$\hat{\beta}_{FD}(s, s+1) = \beta_{RFD} + \zeta_s$$

weighted by the inverse of $\sqrt{\mathtt{Var}\left(\widehat{\beta}_{FD}(s, s+1)\right)}$.

The RFD approach raises a question: Is there an advantage to RFD over FD, since RFD and FD eliminate the same unobserved attributes? As mentioned previously, FE is asymptotically efficient among the class of estimators relying on strict exogeneity when ε_{it} is conditionally homoskedastic and not serially correlated. RFD (in Algorithm 1) is based on separate FD estimates over two time periods rather than a single FD estimate over $T-1$ time periods, where FD and FE are equivalent when $T = 2$. This suggests a possible efficiency gain when using RFD instead of FD. Moreover, since RFD (in Algorithm 1) is based on a series of $T-1$ cross-sectional regressions, there is no need to contemplate the use of clustered standard errors. As discussed in Abadie *et al.* (2022), decisions over clustering when using longitudinal data are not trivial. Finally, RFD lends itself more naturally to situations where researchers wish to allow for temporal heterogeneity in β. We return to this in Section 3.4.

Offsetting these advantages of the RFD estimator is the inefficiency that results from using a sub-optimal weight matrix in the MDE step and/or the use of rolling regressions if the unit fixed effects are constant for all units over the full sample period (Cai and Juhl, 2023). These are potentially addressed using the GMM implementation. We explore this in Section 4.

The logic here can be extended *ad infinitum*. For instance, if the portion of Δx that is orthogonal to $\Delta\alpha$ (such as Δz in Section 3.2) is less persistent than the part that is not orthogonal, first-differencing again will further reduce the bias. This estimator, known as twice first-differencing (TFD), is obtained by estimating

$$\Delta^2 y_{it} = \Delta^2 x_{it}\beta + \Delta^2\varepsilon_{it}, \tag{15}$$

by POLS, where $\Delta^2 y_{it} = \Delta y_{it} - \Delta y_{it-1}$ is the twice-difference operator. In fact,

$$\begin{aligned} \texttt{plim}\,\widehat{\beta}_{tfd} &= \frac{\texttt{Cov}(\Delta^2 y, \Delta^2 x)}{\texttt{Var}(\Delta^2 x)} = \beta + \frac{\lambda\texttt{Var}(\Delta^2\alpha)}{\texttt{Var}(\Delta^2 x)} \\ &= \beta + \left[\frac{\lambda(3-\delta)\sigma^2_{\varepsilon_\alpha}}{1+\delta}\right]\left[\frac{\lambda^2(3-\delta)\sigma^2_{\varepsilon_\alpha}}{1+\delta} + \frac{(3-\rho)\sigma^2_{\varepsilon_z}}{1+\rho}\right]^{-1}. \end{aligned} \tag{16}$$

Under the DGP in Section 3.2, we summarize the performance of TFD in the following proposition.

Proposition 2. *If $\delta > \rho$, implying that α is more persistent than z, then $|Bias(\widehat{\beta}_{tfd})|$ is smaller than that for POLS, FD, and FE. If $\delta < \rho$, implying that z is more persistent than α, then $|Bias(\widehat{\beta}_{tfd})|$ is larger than that for POLS, FD, and FE. If $\delta = \rho$, then $|Bias(\widehat{\beta}_{tfd})|$ is equal to that for POLS, FD, and FE.*

Proof See Appendix A.3 and A.4. ■

Thus, when the unobserved heterogeneity is more persistent than the remaining variation in the covariate—which is the more likely case as argued previously—further differencing reduces the bias (at the expense of efficiency). Moreover, it is trivial to alter Algorithm 1 to obtain the RTFD estimator, where $T-2$ TFD estimates, $\widehat{\beta}_{TFD}(s, s+2)$, are combined. GMM can also be used to estimate RTFD.

Before continuing, it is worth considering two additional DGPs that may be empirically relevant, but where the unobserved heterogeneity is nonstationary. In the first case, the estimating equation is given by Equation (1) where α_{it} is

$$\alpha_{it} = \alpha_i^0 + \alpha_i^1 t. \tag{17}$$

This is a situation of unit-specific linear time trends and is discussed in Mundlak (1978). Here, the unobserved heterogeneity is never constant across time periods, and the expectation of α_{it} varies over time. First-differencing yields

$$\Delta y_{it} = \Delta x_{it}\beta + \alpha_i^1 + \Delta\varepsilon_{it}, \tag{18}$$

and mean-differencing yields

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \alpha_i^1\ddot{t} + \ddot{\varepsilon}_{it}. \tag{19}$$

Both FE and FD will be biased unless x is strictly exogenous conditional on α_i^0 alone. TFD removes both α_i^0 and α_i^1 and thus requires only that x is strictly exogenous conditional on α_i^0 and α_i^1 for unbiasedness. Thus, assuming ε is white noise, TFD has the smallest bias—which is zero—regardless of the persistence in x.

In the second case, the DGP is identical to that in Section 3.2 except that $\delta = 1$, implying that α_{it} follows a random walk. A random walk may be a reasonable approximation in many situations as it assumes unobserved heterogeneity and its effects are time-invariant in expectation rather than in actuality; unobserved heterogeneity is time invariant except for period-specific, random shocks. First-differencing yields

$$\Delta y_{it} = \Delta x_{it}\beta + \varepsilon_{\alpha,it} + \Delta\varepsilon_{it} \tag{20}$$

and mean-differencing yields

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{\alpha}_{i,t-1} + \ddot{\varepsilon}_{\alpha,it} + \ddot{\varepsilon}_{it}. \tag{21}$$

In this case, the FE estimator requires strict exogeneity of x with respect to ε_α and ε. In contrast, the FD estimator requires strict exogeneity of x with respect to ε, but only predetermined with respect to ε_α. The TFD estimator, however, also requires strict exogeneity of x with respect to ε_α.[23] If these conditions fail, TFD will have the smallest bias as noted in Proposition 2 since $\delta > \rho$. FE continues to have a larger bias than both, but a smaller bias than POLS.

Remark 3. *If strict exogeneity fails due to time-varying unobserved heterogeneity that is more persistent than the covariates net of the unobserved heterogeneity, then differencing beyond first-differencing will further reduce the bias at the expense of inefficiency. Researchers should think critically about this bias-variance trade-off. In addition, if the unobserved heterogeneity follows a (unit-specific) linear time trend or a random walk, differencing beyond first-differences will minimize the bias and convert the model to one that is stationary. The FE estimator will not.*

In light of this, one might wonder why the FE estimator is the default choice in empirical studies. We believe the answer lies in the fact that the FE estimator is more efficient than differencing estimators when ε_{it} is conditionally homoskedastic and not serially correlated (Aquaro and Čížek, 2013). In fact, it is the asymptotically efficient estimator among the class of estimators relying on strict exogeneity under these conditions (Wooldridge, 2010, Section 10.6). The widespread use of clustered standard errors nowadays, however, belies this assumption. Moreover, the focus on efficiency comes at the neglect of the potential for bias and inconsistency. Lastly, in situations

[23] In this case, the estimating equation is $\Delta^2 y_{it} = \Delta^2 x_{it}\beta + \Delta\varepsilon_{\alpha,it} + \Delta^2\varepsilon_{it}$. Unbiasedness requires that x_{it} (x_{it-1}) is orthogonal to $\varepsilon_{\alpha,it-1}$ ($\varepsilon_{\alpha,it}$).

with large T, where the difference between the FE and FD (and TFD) estimators are likely to be greatest, efficiency is less likely to be a concern.

That said, researchers should be aware that there are also alternative sources of misspecification where FE is more robust than FD. One example is when a covariate suffers from classical measurement and the true covariate is serially correlated with a declining correlogram (Griliches and Hausman, 1986). Another example is when, say, y_{it} and x_{it} depend on an interactive fixed effects structure, $\alpha_{it} = \lambda_t' \alpha_i$, where λ_t is a vector of common factors. If $\lambda_t \overset{\text{iid}}{\sim} F$, for some distribution F that does not depend on α_i, then FE is preferred to FD and TFD since α_{it} is not persistent.[24] A final example is when a covariate responds to lagged—rather than contemporaneous—unobserved heterogeneity.[25] For instance, if the DGP in Section 3.2 is modified such that

$$x_{it} = \lambda \alpha_{it-1} + z_{it}, \tag{22}$$

then the relative (absolute value of the) biases become more complex.[26] With large T, POLS and FE are preferred to FD and TFD if δ is close to zero. If $\delta = 0$, POLS is unbiased and consistent and FE is consistent. If δ is sufficiently greater than zero, then FD has the smallest bias, followed by TFD and then POLS and FE. If $\delta = 1$, then FD is unbiased and consistent.

3.4 Slope Heterogeneity

Efficiency issues aside, the primary advantage of RFD relative to FD (or RTFD relative to TFD) arises with slope heterogeneity. As discussed in Section 2, in many empirical applications there is reason to suspect that not only might unobserved heterogeneity be time-varying as T increases, but that the slopes will be as well. The slopes may also vary cross-sectionally following a group structure. Both types of heterogeneity are easily incorporated into the RFD and RTFD estimators.

To allow for both types of heterogeneity, we can generalize Equation (1) to

$$y_{it} = x_{it}\beta_{g_{it},t} + \alpha_{it} + \varepsilon_{it}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T}, \tag{23}$$

where $\beta_{g_{it},t}$ is a $K \times 1$ vector of slope parameters that depends on the time period t and the group membership of unit i in period t, g_{it}. For identification, the number of groups, G, should be sufficiently smaller than N. Equation (23) nests the models in Bonhomme and Manresa (2015), Boldea *et al.* (2020), Lumsdaine *et al.* (2023), and Wang *et al.* (2024).

Bonhomme and Manresa (2015) considers the case where $\beta_{g_{it},t} = \beta_{g_i}$ and $\alpha_{it} = \alpha_{g_i,t}$ where g_i denotes the time-invariant group of unit i. This setup allows for cross-sectional (at the group level) heterogeneity in the slopes and cross-sectional (at the group level) and temporal (at the period level) heterogeneity in the intercepts. Boldea *et al.* (2020) considers the case where $\beta_{g_{it},t} = \beta_{m(t)}$ and $\alpha_{it} = \alpha_{i,m(t)}$, where period t belongs to regime m, $m = 1, ..., M$, and $M < T$. Thus, temporal

[24] We are thankful to an anonymous referee for suggesting this example.
[25] We are thankful to Chris Muris for highlighting this point.
[26] See Appendix A.5 and A.6.

(at the regime level) heterogeneity in the slopes is allowed and cross-sectional (at the unit level) and temporal (at the regime level) heterogeneity in the intercepts is allowed. Moreover, the timing of changes in β and α is assumed to be identical. Lumsdaine *et al.* (2023) considers the case where $\beta_{g_{it},t} = \beta_{g_{i,m(t)},m(t)}$ and $\alpha_{it} = \alpha_{i,m(t)}$. Hence, cross-sectional (at the group level) and temporal (at the regime level) heterogeneity in the slopes is allowed and cross-sectional (at the unit level) and temporal (at the regime level) heterogeneity in the intercepts is allowed. In addition, group membership is allowed to vary across regimes. The timing of changes in β, α, and group membership, however, are assumed to be the same and common to all units. Wang *et al.* (2024) is identical to Lumsdaine *et al.* (2023) in that $\beta_{g_{it},t} = \beta_{g_{i,m(t)},m(t)}$, but they consider the case where α_{it} follows an IFE structure. Hence, cross-sectional (at the group level) and temporal (at the regime level) heterogeneity in the slopes is allowed, but some restrictions are placed on the heterogeneity in the intercepts; we discuss IFE further in Section 3.6.

In contrast to these prior approaches, RFD and RTFD automatically allow for temporal (at the period level) heterogeneity in β while also removing as much unobserved, unit-specific heterogeneity as possible. One simply need not combine the $T-1$ estimates of β via MDE (i.e., skip line 5 in Algorithm 1) or impose equality of β across the moment conditions in Equation (14). This imposes less structure than in Boldea *et al.* (2020) and Lumsdaine *et al.* (2023) and removes the need to perform a grid search over all possible break dates.[27] Moreover, we can further amend the algorithm to include cross-sectional (at the group level) heterogeneity in the slopes building on Bonhomme and Manresa (2015) and Lumsdaine *et al.* (2023). Algorithm 2 outlines this approach.

A few remarks are warranted. First, Algorithm 2 should be implemented many times with different starting assignments for the group structure. Second, extending Algorithm 2 to RTFD is trivial; simply replace each first-difference with the corresponding twice first-difference. Third, the theoretical results in Lumsdaine *et al.* (2023) imply that the group structure can be consistently estimated as $N \to \infty$ under certain assumptions.[28] Specifically, Lumsdaine *et al.* (2023) prove that the group structure and the structural break dates can be consistently estimated as $N, T \to \infty$ and $NT^{-\delta} \to 0$ for some $\delta > 0$ under their assumptions. They also show that the asymptotic behavior of their estimator is as if the group structure and structural break dates are known *a priori*, implying that standard statistical inference can be used. In our case, we are allowing for a structural break at every period and estimating the group structure using only cross-sectional data. That said, there is a fundamental difference between the current setup and Lumsdaine *et al.* (2023). In their case, there is no overlap in the sample before and after the estimated break date. In our case, our rolling regressions overlap unless one revises line 1 in Algorithm 2 to $s = 2, 4, 6, ..., T$ if T is even or ending at $T-1$ if T is odd. With overlap, inference becomes nonstandard as in Cai and Juhl (2023) and new critical values are needed. An alternative is to use the fuzzy clustering approach in Lewis *et al.* (2023). This approximates the steps in Algorithm 2 with an objective function that is continuous.

[27] Alternatively, one might allow for temporal heterogeneity in the slopes, but smooth the changes over time using, for example, kernel smoothing (e.g., Cai and Juhl, 2023).

[28] See Assumption 1 in Lumsdaine *et al.* (2023)

The slope parameters and group assignments can be estimated in a single step by GMM.[29] This allows the usual GMM inference to apply.

Algorithm 2 Rolling First Differences (RFD) Estimator with Slope Heterogeneity

1: **while** $s = 2, ..., T$ **do**

2: **while** $G = 2, ..., \overline{G}$ **do**

3: Let G be the number of groups. Let $g_s^{(0)}$ be a $1 \times N$ vector assigning each unit to a group $g \in 1, ..., G$. Set $m = 0$.

4: Compute

$$\beta_s^{(m+1)} = \underset{\beta_s \in \mathcal{R}^{\mathtt{G} \times \mathtt{K}}}{\texttt{arg min}} \sum_i \left[\Delta y_{is} - \Delta x_{is} \beta_{g_{is}^{(m)}, s}\right]^2$$

which can be done by applying OLS separately to each group.

5: Given $\beta_s^{(m+1)}$ compute the optimal group assignment for each unit

$$g_s^{(m+1)} = \underset{\mathbf{\Gamma_G}}{\texttt{arg min}} \sum_i \left[\Delta y_{is} - \Delta x_{is} \beta_{g_{is},s}^{(m+1)}\right]^2$$

where $\mathbf{\Gamma_G}$ is the set of all possible partitions of N units into G mutually exclusive groups and Δx_{is} may or may not include an intercept. The solution can be obtained by computing for each $i \in \mathcal{N}$

$$g_{is}^{(m+1)} = \underset{g \in \{1,...,G\}}{\texttt{arg min}} \left[\Delta y_{is} - \Delta x_{is} \beta_{g,s}^{(m+1)}\right]^2$$

6: Set $m = m + 1$ and go to line 4 (until numerical convergence).

7: **end while**

8: Compute

$$G_s^* = \underset{G \in \left\{2,...,\overline{G}\right\}}{\texttt{arg min}} \sum_i \left[\Delta y_{is} - \Delta x_{is} \beta_{g_{is},s}^{G}\right]^2 + \Pi$$

where $\beta_{g_{is},s}^G$ is final estimate of $\beta_{g_{is},s}$ when the number of groups is G and Π is a penalty that is a function of N and T and the number of parameters $K \times G$ (e.g., Aikake or Bayesian Information Criterion). The maximum possible number of groups, $\overline{G}$, must be sufficiently smaller than N.

9: **end while**

3.5 Testing Strict Exogeneity

Although the strict exogeneity assumption may fail for reasons other than time-varying unobserved heterogeneity, testing this assumption is crucial regardless. Unfortunately, empirical researchers

[29] Lewis *et al.* (2023) impose time-invariant group assignments. If one wishes to do that here, their estimator is directly applicable. Instead, Algorithm 2 loops over periods, allowing group structure to change across periods, but at the expense of estimating group assignment in each period using only a single cross-section.

rarely perform such a test despite the ease of doing so.[30] There are two simple ways to test the strict exogeneity assumption. First, as discussed in Wooldridge (2010, p. 325), one can estimate

$$\Delta y_{it} = \Delta x_{it}\beta + w_{it}\gamma + \Delta\alpha_{it} + \Delta\varepsilon_{it}, \tag{24}$$

where w is a subset of x (that excludes period fixed effects), by POLS and test $\mathtt{H}_o : \gamma = 0$ using a two-sided alternative. Alternatively, one can estimate

$$y_{it} = x_{it}\beta + w_{i,t+1}\gamma + \alpha_{it} + \varepsilon_{it} \tag{25}$$

using the FE estimator and again test $\mathtt{H}_o : \gamma = 0$ using a two-sided alternative.

Second, following Laporte and Windmeijer (2005), under the null hypothesis that the stronger strict exogeneity required by the FE estimator is true, β_{FE} and β_{FD} (and β_{TFD}) are unbiased and consistent for β and thus the estimates should be sufficiently close. If the estimates diverge in a statistically meaningful way, then this provides evidence against the null. Wooldridge (2010, p. 321) states that "if FE and FD estimates differ in ways that cannot be attributable to sampling error, we should worry about the violations of the strict exogeneity assumption." The test can be accomplished by estimating the following stacked regression via POLS

$$\begin{pmatrix} \ddot{y}_{it} \\ \Delta y_{it} \end{pmatrix} = \begin{pmatrix} \ddot{x}_{it} \\ \Delta x_{it} \end{pmatrix}\beta + \begin{pmatrix} 0 \\ \Delta x_{it} \end{pmatrix}\theta + u_{it} \tag{26}$$

and testing $\mathtt{H}_o : \theta = 0$ using a two-sided alternative and a cluster-robust estimator of the covariance matrix (Papke and Wooldridge, 2023; Spierdijk, 2023). Of course, rejection of the null may also occur under other sources of misspecification that differentially affect FE and FD. Moreover, choosing an estimator on the basis of this test risks introducing problems associated with pre-testing (Papke and Wooldridge, 2023).

Remark 4. *When estimating linear panel data models with unobserved heterogeneity, the FE and differencing estimators are asymptotically equivalent if strict exogeneity holds. All estimators relying on strict exogeneity will be biased and inconsistent if this assumption fails, with the direction and magnitude of the bias varying across estimators. Testing this assumption is straightforward yet rarely done in practice. Researchers should present the results of such tests and think critically about the implications.*

3.6 Interactive Fixed Effects

Our analysis of estimators that are simple and general—and thus more likely to be adopted by applied researchers—would be incomplete without considering a new addition to the empirical toolkit: the interactive fixed effects estimator (IFE).[31] The IFE estimator originates in Bai (2009),

[30]This is likely because failure to reject the null of exogeneity when performing such tests is generally thought to be less convincing than a rejection of the same null.

[31]The estimator is available in Stata (`regife`), R (`interFE`), and Python (`pyInteractiveFixedEffects`).

with the specification given by

$$y_{it} = x_{it}\beta + \lambda_t'\alpha_i + \varepsilon_{it}, \;\; \forall (i,t) \in \mathcal{N} \times \mathcal{T} \tag{27}$$

where λ_t is a vector of common factors. This specification allows unit- and time-varying unobserved heterogeneity, but with a particular functional form. It posits unit-specific responses to common temporal shocks as well as common time-varying amplification of time invariant unit-specific unobserved heterogeneity. The additional flexibility of the IFE estimator may offer advantages over the traditional FE or FD estimators in certain situations such as the inclusion of unit-specific time trends (Mundlak, 1978) or higher-dimensional unobserved heterogeneity (Gobillon and Magnac, 2016). In a setting with time-varying, unit-specific unobserved heterogeneity with unknown form, however, the common factor structure for all units may be less robust than FD or TFD. That said, the estimator can accommodate multiple factors; we consider one- and two-factor IFE estimators in Section 4.

4 Examples

We begin by briefly discussing the results of Monte Carlo simulations. We then discuss the results of four replications. As our primary focus is on the effects of time-varying fixed effects, we (mostly) abstract from slope heterogeneity.

4.1 Monte Carlo Study

Details and full results for our Monte Carlo study are relegated to Appendix B. We compare several estimators: (i) POLS, (ii) FE, (iii) FD, (iv) TFD, (v) RFD, (vi) RTFD, (vii) GMM FD, (viii) GMM TFD, and (ix) IFE with one and two factors, respectively denoted IFE1 and IFE2. All FD and TFD estimators exclude an intercept in the differenced equations.[32] For simplicity, we do not consider cases of slope heterogeneity.

We evaluate the estimators in terms of the bias, absolute bias, and root mean squared error (RMSE) of the slope coefficient on a single unit- and time-varying covariate from 200 replications for each experimental design and configuration.

The results are plotted in Figures B.1 to B.24 in Appendix B. Because there are a lot of figures, we provide a roadmap for the reader. Specifically,

- In each figure, Panel A displays the performance of each estimator with $N = 100$ observations, and Panel B displays the performance of each estimator with $N = 500$ observations.

- In Figures B.1 to B.16, the five vertical panels in each of Panels A and B vary the degree of serial correlation δ in the unit-specific fixed effect α_{it}, ranging from no serial correlation to

[32]The performance of the estimators with and without a constant are virtually identical given the experimental designs. In practice, we recommend researchers include a constant. Our simulation results show there is no cost to its inclusion even if the true value is zero.

perfect serial correlation, or such that $\delta \in \{0, 0.25, 0.5, 0.75, 1\}$. The degree of serial correlation ρ in the part of the covariate orthogonal to α_{it} is fixed at 0.5.

- We vary the variance $\sigma^2_{\varepsilon_\alpha}$ of the fixed effect α_{it} across Figures B.1 to B.16. Specifically, Figures B.1 to B.4 set $\sigma^2_{\varepsilon_\alpha} = 0.25$, Figures B.5 to B.8 set $\sigma^2_{\varepsilon_\alpha} = 0.50$, Figures B.9 to B.12 set $\sigma^2_{\varepsilon_\alpha} = 0.75$, and Figures B.13 to B.16 set $\sigma^2_{\varepsilon_\alpha} = 1$.

- In Figures B.17 to B.24, x_{it} is a function of $\alpha_{i,t-1}$ following the discussion at the end of Section 3.3. We vary the persistence in the data by varying both δ and ρ. We set $\delta = \rho = 0.1$ in Figures B.17 to B.20, and set $\delta = \rho = 0.9$ in Figures B.21 to B.24.

- Finally, we present the RMSE of two simple model averaging (MA) estimators in Figures B.4, B.8, B.12, and B.16, where the label MA1 and MA2 refer to the simple averaging of the FE, FD, and TFD estimators and of the FE, RFD, and RTFD estimators, respectively.[33]

- In each figure, the x-axis shows the number of time periods, or panel length. Since our focus is on what happens as a panel encompasses longer time periods, we consider $T \in \{5, 10, 20, 30\}$.

For illustrative purposes, Figure 1 is identical to Figure B.15, but serves to showcase the pattern of results that we find.

There are several important takeaways:

1. All estimators considered are biased when x_{it} depends on α_{it}.

2. The disparity in the point estimates between the estimators is increasing in $|\delta - \rho|$ and T. However, the bias of the best performing estimator changes only modestly as $|\delta - \rho|$. Thus, the magnitude of the difference between, say, the FE and FD estimates provides no information on which estimator may be preferred or the magnitude of the biases.

3. Consistent with the theoretical analysis, POLS performs best in terms of bias and RMSE when $\delta = 0$ or 0.25 (and hence is smaller than $\rho = 0.5$. The performances of FE, IFE1, and IFE2 are close to POLS, particularly as T increases.

4. The rankings are reversed when $\delta = 0.75$ or 1 (and hence is larger than $\rho = 0.5$). TFD and RTFD perform the best and the performances of FD and RFD are close. FE, IFE1, IFE2, and POLS perform considerably worse, particularly as T increases.

5. When $\delta = \rho = 0.5$ (and hence α and z are equally persistent), the performance of all estimators is nearly indistinguishable.

6. The results improve only marginally with N and the qualitative conclusions are unaffected by N.

[33] For a more complete introduction to model averaging in linear regression models with or without structural breaks, see, e.g., Hansen (2007, 2009). Antonelli and Cefalu (2020) consider averaging many causal estimates in high-dimensional settings.

7. Altering the variance of α_{it} changes the magnitude of the bias and RMSE, but not the relative performance of estimators.

8. When x_{it} depends on $\alpha_{i,t-1}$ rather than α_{it}, the relative performance of the estimators depends on whether δ and ρ are close to zero or close to one. When both are close to zero, FE performs best, followed closely by POLS, IFE1, and IFE2. When both are close to one, FD and RFD perform best and POLS is always the worst. The relative performance of the remaining estimators depends on T; when T is small, FE, IFE1, and IFE2 outperform TFD and RTFD, while the reverse is true with large T. As before, the MA estimators offer a robust middle option regardless of the degree of persistence.

In sum, the simulation results make clear that (i) time-varying unobserved heterogeneity can be consequential, (ii) FE is rarely the optimal estimator, and (iii) researchers must think carefully about the persistence of the covariates relative to the unobserved heterogeneity to justify their estimator choice, particularly with large T. Thus, reporting several estimators along with tests of strict exogeneity should become standard practice.

4.2 Replications

To illustrate the importance of estimator choice when unobserved heterogeneity is likely to be time-varying, we conduct four replications. We first follow Imai and Kim (2019) by replicating Rose (2004) and Tomz *et al.* (2007), who look at the relationship between country-level participation in the General Agreement on Tariffs and Trade (GATT) and international trade. We then replicate Leipziger (2024) who assesses the impact of democracy on ethnic inequality. Next, we replicate James (2015), who looks at the relationship between state-level resource-based government revenue and state-level fiscal measures. In all three replications, we show that the choice of estimator can have important consequences not just for the magnitude of the coefficients of interest, but also for their sign and statistical significance. In contrast, when replicating Djankov and Reynal-Querol (2010), we find the FE estimator to yield broadly robust results.

4.2.1 Imai and Kim (2019)

Table 1 replicates results in Imai and Kim (2019), who in turn replicate the analyses of the relationship between country-level GATT participation and bilateral trade in Rose (2004) and Tomz *et al.* (2007). The results in Table 1 use a sample of 175 countries over the period 1948-1994. Dyad-specific fixed effects (i.e., a unit is a country pair) are included in a standard "gravity" model specification; Rose (2004) and Tomz *et al.* (2007) also consider separate country-specific fixed effects. In Tomz *et al.* (2007), researchers justify the inclusion of country fixed effects following the theoretical model in Anderson and van Wincoop (2003), which "implies the presence of a 'multilateral resistance' term that can be approximated using country and time fixed effects."

Each column in Table 1 denotes a different estimator and each panel denotes a different specification. In Panels A and B the variable of interest is formal GATT membership, where Panel B includes year FEs. In Panels C and D, the variable of interest is informal GATT membership, where Panel D includes year FEs. In each panel, we show the estimated coefficient on GATT membership as well as the result of a Laporte and Windmeijer (2005) test of equality of coefficients between the FE and FD specifications. Figure C.1 in Appendix C plots the RFD and RTFD (with an intercept) estimates by year to examine temporal heterogeneity in the effects of GATT.[34] Tables C.1–C.4 in the Appendix display the full set of coefficient estimates.

The results are striking. First, FE and FD are statistically different at the $p < 0.01$ level in all four panels. While the difference in point estimates is small in Panel B, this is not the case in the remaining panels. Thus, the choice between FE and FD fundamentally alters the conclusions. Second, while the TFD estimates are always close to the FD estimate, the RFD and RTFD (with or without a constant for either) tend to be statistically insignificant, except for the RFD estimates in the case of informal GATT membership, in which case the estimates are closer to the FD estimate than the FE estimate. Third, the IFE1 estimates are quite similar to the FE estimates, as suggested by our simulation study. The IFE2 estimates are more similar to the FD estimates, although are statistically insignificant. Finally, the year-specific estimates in Figure C.1 show marginally less volatile and more likely to be non-negative estimates for informal GATT membership, particularly during 1960s and 1970s.

4.2.2 Leipziger (2024)

Tables 2 and 3 replicate results in Leipziger (2024), who looks at the relationship between democracy and ethnic inequality using country-level panel data. The objective is to test the hypothesis that democracy elevates ethnic minority groups through the ability to mobilize and hold the government accountable. In addition, it is hypothesized that this effect may be greater in countries where the pre-democratic regime was particularly unequal due to the greater incentives to achieve equality created by starting from a lower (relative) baseline.

We analyze the effects of a binary measure of democracy on two measures of ethnic inequality: a measure of group inequality in access to public services and inequality in educational attainment.[35] The first measure is available for an unbalanced panel of 175 countries over 121 years. The median country is in the sample for 84 years. The second measure is available for an unbalanced panel of 86 countries over 68 years. The median country is in the sample for 49 years. Despite the sample spanning many decades, the author relies heavily on country fixed effects to interpret the results causally. Leipziger (2024, p. 1341) states:

> "I regard country-fixed effects as crucial for addressing potential endogeneity issues. Countries characterized by high ethnic inequality are likely to be different from countries with less pronounced inequality on a range of confounding characteristics, such

[34] Note, estimates are missing for certain years in which there is no change in GATT status from the prior year. In addition, one should interpret the confidence bands carefully as they are likely to have coverage rates that are too small as discussed in Section 3.4.

[35] Leipziger (2024) examines a third outcome based on a measure of inequality constructed from nighttime lights data. However, this measure is only available for the years 1992, 2000, and 2012.

as colonization, state antiquity, and geographic location. By including country-fixed effects, the specifications control for such unobservable time-invariant factors."

While some of the country-level factors listed are (mostly) time invariant (e.g., even borders change over time), the effects of such characteristics on ethnic inequality are not likely to be constant over such a long time period. The unobserved heterogeneity, while time-varying, is likely to be highly persistent.

The results for inequality in public services are in Table 2. The results for inequality in educational attainment are in Table 3. Again, each column in the tables denotes a different estimator, and each panel denotes a different specification. Each panel also displays the result of the Laporte and Windmeijer (2005) test of equality of coefficients between the FE and FD specifications. Figures C.2 and C.3 in Appendix C plots the RFD and RTFD (with an intercept) estimates by year to examine temporal heterogeneity in the effects of democracy and the interaction between democracy and the pre-democratic level of inequality.[36]

The results highlight the importance of estimator choice. First, FE and FD are statistically different at the $p < 0.01$ level in three of four panels. In Panel A in both tables, the effect of democracy is reduced in magnitude by at least 80% (in absolute value), although it remains statistically significant at the $p < 0.01$ level in Table 2. In Panel B in both tables, the coefficients on the interaction term between democracy and the pre-democratic level of inequality become precisely estimated null effects. Second, in Panel A in both tables, the effect of democracy is a precisely estimated zero according to TFD and all rolling estimators. The IFE estimates continue to be similar to the FE estimates, particularly in Table 2, albeit with smaller magnitudes. Finally, in Table 2 the TFD and the rolling estimates are not consistent with the hypothesis that greater pre-democratic inequality leads to a larger reduction in inequality under democracy. In fact, the TFD and RFD (with constant) are now positive and statistically significant at at least the $p < 0.05$ level. In contrast, in Table 3 the rolling estimates are all negative and statistically significant at the $p < 0.01$ level. The point estimates, however, are at least 80% smaller than the FE estimate.

4.2.3 James (2015)

Tables 4 and 5 replicate results in James (2015), who looks at the relationship between state-level resource-based government revenue and taxation, spending, and savings at the state-level using annual data from 1958-2008. The goal is to test the theory that a benevolent government will reduce taxes and increase both spending and savings in response to an exogenous increase in resource-based government revenue. Despite the sample covering 51 years, James (2015, p. 243) claims that "time-invariant, state-specific characteristics such as average population density, political preferences, wealth, unemployment, culture, and institutional quality are captured by state fixed effects." Clearly, this is not the case. For example, Caughey and Warshaw (2016) shows that political preferences as measured by state policy liberalism varied tremendously over the past century, at least

[36]No additional controls are included in the original study.

for select states. Frank (2009) documents the intrastate temporal variation in income inequality. The unemployment rate in California varied from a low of 4.8% to a high of 11.1% between 1976 and 2008.[37]

The results in Table 4 use the full sample of all 50 states. The specifications in Table 5 omit Alaska but are otherwise identical. As before, each column in the tables denotes a different estimator, and each panel denotes a different dependent variable. In each panel, we show the estimated coefficient on resource-based government revenue as well as the result of the Laporte and Windmeijer (2005) test of equality of coefficients between the FE and FD specifications. Figures C.4 and C.5 in Appendix C plots the RFD and RTFD (with an intercept) estimates by year to examine temporal heterogeneity in the effects of resource-based revenue.[38]

Our analysis leads to a few salient findings. First, the FE and FD estimates in Table 4 are statistically different at the $p < 0.05$ level in three out of five cases. The estimates are even of opposite sign for education expenditures (Panel D). Moreover, while statistically different at only the $p < 0.11$ level, the FE and FD estimates are also of opposite sign for nonresource revenue (Panel A), only the FD estimate is statistically significant (at the $p < 0.01$ level), and the FD estimate is more than three times as large in absolute value. Second, the divergence between the FE and FD estimates is even more pronounced in Table 5 when Alaska is omitted. While the FE estimates are statistically significant at the $p < 0.05$ level for all outcomes, the FE and FD estimates continue to be statistically different at the $p < 0.05$ level in three out of five cases. Moreover, the FD estimates are statistically indistinguishable from zero in three out of five cases. Thus, the choice between FE and FD matters.

Third, the TFD and the various RFD estimators are much more similar to the FD estimates than the FE estimates, consistent with our simulation study. In Table 4, in particular, FE is a clear outlier among this group of estimators. Fourth, the IFE estimates are a mixed bag, but they tend to look closer to the FE estimates than the FD or RFD estimates, at least in the case of the IFE1 estimate. The IFE2 estimates often differ in magnitude from the IFE1 estimate. Again, this is consistent with our simulation results. Fifth, while the year-specific estimates shown in Figures C.4 and C.5 are noisy, there is no discernible pattern of temporal heterogeneity.

Finally, a striking pattern emerges when examining the instrumental variable (IV) estimates in James (2015). Omitting the details, the author worries about the endogeneity of resource revenues even with the inclusion of state FEs and thus combines IV with FE (IV-FE) in the specifications omitting Alaska. Interestingly, in four of five cases the FD estimates are much closer to the IV-FE estimates than the FE estimates. This is consistent with FD removing more relevant unobserved heterogeneity than FE (but not accounting for as much unobserved heterogeneity as IV-FE). While certainly not a general result, in this case FD, TFD, and RFD isolate much of the same (or similar) exogenous variation in resource-based government revenue as does the instrument. For example, for total expenditures (Panel C), the point estimates change from 0.43 (FE) to 0.19 (FD) to -0.03

[37] See https://fred.stlouisfed.org/series/CAUR.

[38] Note, no additional controls are included in the original study.

(IV-FE). For education expenditures (Panel D), the point estimates change from 0.15 (FE) to 0.02 (FD) to -0.03 (IV-FE). For public savings (Panel E), the point estimates change from 0.32 (FE) to 0.55 (FD) to 0.76 (IV-FE).

4.2.4 Djankov and Reynal-Querol (2010)

Table 6 replicates the results in Djankov and Reynal-Querol (2010), who look at the relationship between country-level poverty and the onset of civil war using data on roughly 200 countries for the period 1960-2000. The goal is to test the assertion that poverty reduction is a critical factor in reducing the incidence of civil war and violence. While prior evidence supports the assertion, the authors contend that failure to include country fixed effects biases these results.[39]

The results in Table 6 rely on data from 1960-2000 collapsed to five- or ten-year intervals. The outcomes are indicators for the onset of civil war at any point during the interval and the covariates are measured at the start of the interval. This leads to $T = 8$ and $T = 4$, respectively.[40] Figure C.6 in Appendix C plots the RFD and RTFD (with an intercept) estimates by period to examine temporal heterogeneity in the effects lagged income. Tables C.5–C.7 in Appendix C display the full set of coefficient estimates.

In contrast to the previous replications, here the original FE estimates are broadly similar to those of alternative estimators. Specifically, we never reject the null of equality between the FE and FD estimates below the $p < 0.09$ level and the only cases where the alternative estimators indicate a negative and statistically significant impact of income are IFE1 and IFE2 in Panel B(I) and IFE1 in Panel B(II). The remaining point estimates are a mix of positive and negative, with TFD being positive and marginally statistically significant in Panel B(I). Figure C.6 bears out these mixed results. Specifically, in Panel (A) the effects are negative and statistically significant early in the sample, but a mix of positive and negative later in the sample. In Panel (B), the estimates oscillate between roughly positive and negative across periods. Finally, in Panel (C) the estimate are positive during the middle of the sample period, but predominantly negative at the beginning and end of the sample period.

Djankov and Reynal-Querol (2010) also use a second data set spanning 1825-2000 and group the data in 25- and 50-year intervals. While we omit our replication of those results for brevity, our findings are again broadly consistent with the FE estimates. Specifically, while we now find the FD and FE estimates to be statistically different at the $p < 0.01$ level, all alternative estimates indicate either no statistically meaningful relationship between income and the onset of civil war, or a positive and statistically significant relationship, consistent with Miguel *et al.* (2004).

[39] The authors also discuss instrumenting for poverty, finding similar results to their FE estimates.

[40] The authors also consider 20-year intervals, but this implies $T = 2$ which precludes the need for alternative estimators.

5 Conclusion

When everything else is held constant, more data is better. But when is everything ever held constant? In this paper, we highlight the oft-overlooked fact that as T increases, estimators such as FE are of decreasing usefulness because the amount of time-invariant heterogeneity is (weakly) decreasing in T. We then compare the FE estimator with the FD estimator, which is often taught alongside the FE estimator but which has fallen out of favor relative to the FE estimator. Importantly, we show that FD is preferred to FE when the persistence in the time-varying unobserved heterogeneity is weakly greater than the persistence in the time-varying covariates of interest net of the unobserved heterogeneity.

We also present additional estimators that improve on FE under large T and unobserved heterogeneity that is relatively persistent, but not time invariant, and are simple to implement. While the proposed alternatives to FE are inefficient under the assumptions of the standard fixed effects model, they can be more robust to the types of misspecification discussed here as the amount of unobserved heterogeneity removed by these alternatives is (weakly) greater.[41]

On the basis of these findings, we offer several recommendations for applied researchers when using panel data with $N >> T$. Researchers should

1. Explicitly discuss (and justify) the assumptions made regarding what is captured by unit fixed effects.

2. At the very least, report estimates based on temporal differences (i.e., estimates from FD, RFD, TFD, and RTFD) when $T > 2$ and compare them with results from the fixed effects estimator.

3. Test for violations of strict exogeneity and/or equality between the fixed effects and first differences estimators.

4. Seek to assemble higher-frequency data (i.e., more sub-periods within each period t) rather than (or in addition to) longer (i.e., larger T) panels.[42]

Our analysis suggests several avenues for future research. First, a rolling FE estimator may be worth exploring where one rolls over three or more consecutive periods at a time. This is similar to existing approaches of allowing for multiple, common structural breaks and has efficiency gains relative to RFD if the slopes and intercepts are temporally locally homogeneous over more than two time periods. Second, similar to Qian and Su (2016), Li *et al.* (2016), Okui and Wang (2021), and Chan and Mátyás (2022), it may be advantageous to treat α_{it} as parameters to be estimated using machine learning techniques under a sparsity assumption. Third, the derivation of proper

[41] For time-varying unobserved heterogeneity in panel event-study designs, see Freyaldenhoven *et al.* (2019), who provide sufficient conditions for identification even in the presence of pre-trends in the outcome variable.

[42] This is consistent with McKenzie (2012), who advocates that researchers take multiple measurements of noisy outcomes characterized by a low degree of autocorrelation (e.g., agricultural yields, business profits, household expenditures).

confidence bands for the rolling estimates of the slope parameters is needed when one estimates the model period-by-period. Estimation by GMM offers a solution but can be computationally intensive with large T. Fourth, one should consider the bias-variance tradeoff involved with the use of time fixed effects as N increases. For example, as one adds more countries to country-by-time panel data, unobserved heterogeneity that is constant over the larger set of counties within a given time period will be (weakly) less. Finally, given the popularity of TWFE in difference-in-difference designs, as well as the now well-known issues created by staggered adoption, investigating the performance of rolling estimators in that context seems warranted (see, e.g., Roth *et al.*, 2023)—an effort we rather happily leave for future research.

References

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2022). When should you adjust standard errors for clustering?

Ahn, S. C., Hoon Lee, Y. and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, **101** (2), 219–255.

—, — and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, **174** (1), 1–14.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, **93** (1), 170–192.

Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, **31** (1), 163–191.

Antonelli, J. and Cefalu, M. (2020). Averaging causal estimators in high dimensions. *Journal of Causal Inference*, **8** (1), 92–107.

Aquaro, M. and Čížek, P. (2013). One-step robust estimation of fixed-effects panel data models. *Computational Statistics & Data Analysis*, **57** (1), 536–548.

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, **21** (1), 1–42.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, **77** (4), 1229–1279.

Balestra, P. and Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, **34** (3), 585–612.

Baltagi, B. H., Feng, Q. and Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, **191** (1), 176–195.

Bellemare, M. F., Masaki, T. and Pepinsky, T. B. (2017). Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics*, **79** (3), 949–963.

— and Millimet, D. L. (2025). Retrospectives: Yair Mundlak and the fixed effects estimator. *Journal of Economic Perspectives*, **39** (2), 261–274.

Boldea, O., Drepper, B. and Gan, Z. (2020). Change point estimation in panel data with time-varying individual effects. *Journal of Applied Econometrics*, **35** (6), 712–727.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, **83** (3), 1147–1184.

Bramati, M. C. and Croux, C. (2007). Robust estimators for the fixed effects panel data model. *The Econometrics Journal*, **10** (3), 521–540.

Cai, Z. and Juhl, T. (2023). The distribution of rolling regression estimators. *Journal of Econometrics*, **235**, 1447–1463.

Campello, M., Galvao, A. F. and Juhl, T. (2019). Testing for slope heterogeneity bias in panel data models. *Journal of Business & Economic Statistics*, **37** (4), 749–760.

Caughey, D. and Warshaw, C. (2016). The dynamics of state policy liberalism, 1936–2014. *American Journal of Political Science*, **60** (4), 899–913.

Chan, F. and Mátyás, L. (2022). Linear econometric models with machine learning. In F. Chan and L. Mátyás (eds.), *Econometrics with Machine Learning*, Springer Nature Switzerland AG, pp. 1–37.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25** (1), 1–37.

Djankov, S. and Reynal-Querol, M. (2010). Poverty and civil war: revisiting the evidence. *The Review of Economics and Statistics*, **92** (4), 1035–1041.

Frank, M. W. (2009). Inequality and growth in the United States: evidence from a new state-level panel of income inequality measures. *Economic Inquiry*, **47** (1), 55–68.

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

Gibbons, C. E., Serrato, J. C. S. and Urbancic, M. B. (2019). Broken or fixed effects? *Journal of Econometric Methods*, **8** (1), article 20170002.

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

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, **31** (1), 93–118.

Hahn, J., Hausman, J. and Kuersteiner, G. (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics*, **140** (2), 574–617.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, **75** (4), 1175–1189.

— (2009). Averaging estimators for regressions with a possible structural break. *Econometric Theory*, **25** (6), 1498—-1514.

Hill, T. D., Davis, A. P., Roos, J. M. and French, M. T. (2020). Limitations of fixed-effects models for panel data. *Sociological Perspectives*, **63** (3), 357–369.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *Test*, **16**, 1–22.

Imai, K. and Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*, **63** (2), 467–490.

— and — (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, **29** (3), 405–415.

Ishimaru, S. (2025). What do we get from two-way fixed effects regressions? implications from numerical equivalence. *arXiv preprint arXiv:2103.12374*.

James, A. (2015). Us state fiscal policy and natural resources. *American Economic Journal: Economic Policy*, **7** (3), 238–257.

Kaddoura, Y. and Westerlund, J. (2023). Estimation of panel data models with random interactive effects and multiple structural breaks when T is fixed. *Journal of Business & Economic Statistics*, **41** (3), 778–790.

Keane, M. and Neal, T. (2020). Climate change and U.S. agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, **11** (4), 1391–1429.

Laporte, A. and Windmeijer, F. (2005). Estimation of panel data models with binary indicators when treatment effects are not constant over time. *Economics Letters*, **88** (3), 389–396.

Leipziger, L. E. (2024). Does democracy reduce ethnic inequality? *American Journal of Political Science*, **68** (4), 1335–1352.

Lewis, D. J., Melcangi, D., Pilossoph, L. and Toner-Rodgers, A. (2023). Approximating grouped fixed effects estimation via fuzzy clustering regression. *Journal of Applied Econometrics*, **38** (7), 1077–1084.

LI, D., QIAN, J. and SU, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, **111** (516), 1804–1819.

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

LIU, R., SHANG, Z., ZHANG, Y. and ZHOU, Q. (2020). Identification and estimation in panel models with overspecified number of groups. *Journal of Econometrics*, **215** (2), 574–590.

LUMSDAINE, R. L., OKUI, R. and WANG, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, **233** (1), 45–65.

MCKENZIE, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, **99** (2), 210–221.

MCKINNISH, T. (2008). Panel data models and transitory fluctuations in the explanatory variable. In T. Fomby, R. Carter Hill, D. L. Millimet, J. A. Smith and E. J. Vytlacil (eds.), *Advances in Econometrics*, vol. 21, Emerald Group Publishing Limited, pp. 335–358.

MEHRABANI, A. (2023). Estimation and identification of latent group structures in panel data. *Journal of Econometrics*, **235** (2), 1464–1482.

MIGUEL, E., SATYANATH, S. and SERGENTI, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, **112** (4), 725–753.

MULLAHY, J. (2016). Estimation of multivariate probit models via bivariate probit. *The Stata Journal*, **16** (1), 37–51.

MUNDLAK, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, **43** (1), 44–56.

— (1978). On the pooling of time series and cross section data. *Econometrica*, **46** (1), 69–85.

MURAYAMA, K. and GFRÖRER, T. (forthcoming). Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for choosing a statistical model from a causal inference perspective. *Psychological Methods*.

NICKELL, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, **49** (6), 141 Inference7–1426.

OKUI, R. and WANG, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, **220** (2), 447–473.

PAPKE, L. E. and WOOLDRIDGE, J. M. (2023). A simple, robust test for choosing the level of fixed effects in linear panel data models. *Empirical Economics*, **64** (6), 2683–2701.

Pesaran, M. H. and Zhou, Q. (2018). To pool or not to pool: revisited. *Oxford Bulletin of Economics and Statistics*, **80** (2), 185–217.

Plümper, T. and Troeger, V. E. (2019). Not so harmless after all: The fixed-effects model. *Political Analysis*, **27** (1), 21–45.

Qian, J. and Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, **191** (1), 86–109.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review*, **94** (1), 98–114.

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

Sarafidis, V. and Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*, **77** (2), 274–296.

Spierdijk, L. (2023). Assessing the consistency of the fixed-effects estimator: a regression-based Wald test. *Empirical Economics*, **64** (4), 1599–1630.

Su, L. and Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, **29** (6), 1079–1135.

—, Shi, Z. and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, **84** (6), 2215–2264.

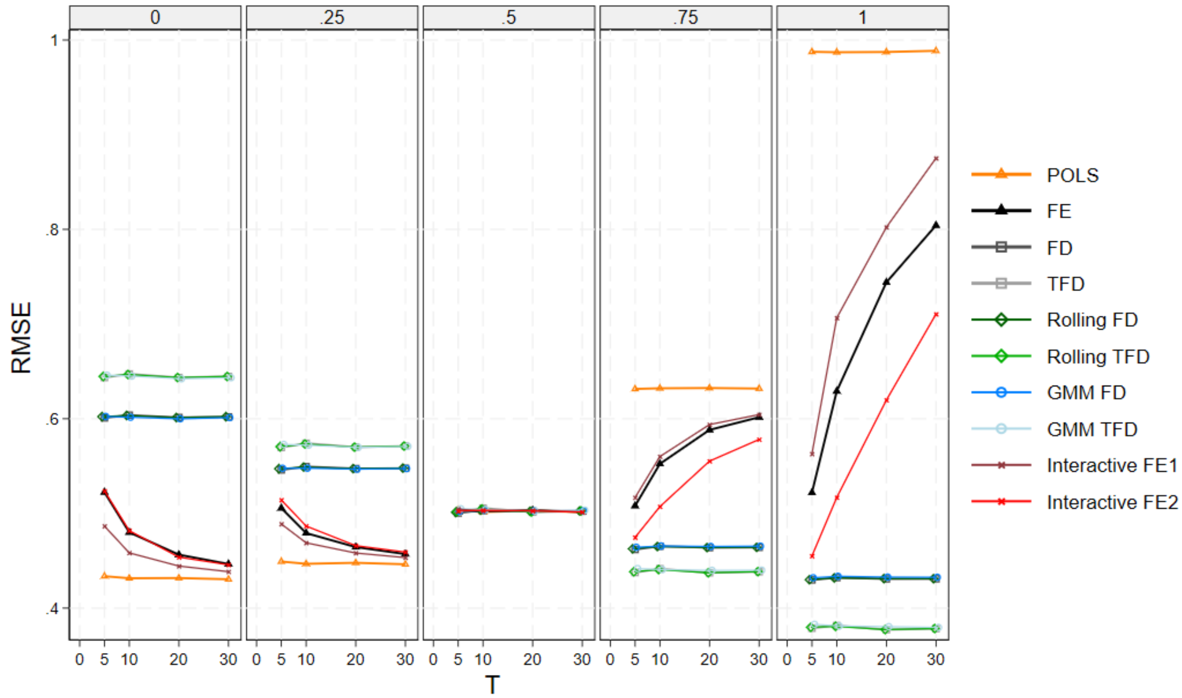
Sun, L. and Shapiro, J. M. (2022). A linear panel model with heterogeneous coefficients and variation in exposure. *Journal of Economic Perspectives*, **36** (4), 193–204.

Tomz, M., Goldstein, J. L. and Rivers, D. (2007). Do we really know that the WTO increases trade? comment. *American Economic Review*, **97** (5), 2005–2018.

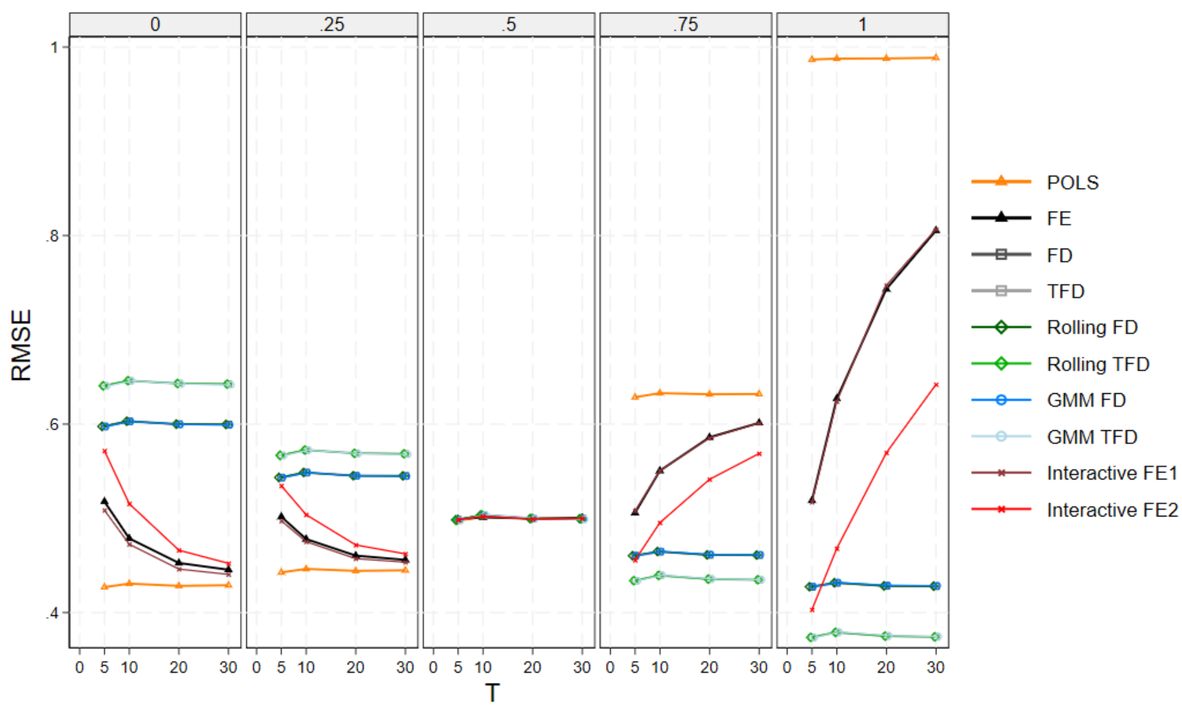
Wang, X., Kang, Y., Hyndman, R. J. and Li, F. (2023). Distributed arima models for ultra-long time series. *International Journal of Forecasting*, **39** (3), 1163–1184.

Wang, Y., Phillips, P. C. B. and Su, L. (2024). Panel data models with time-varying latent group structures. *Journal of Econometrics*, **240** (1), 105685.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press, 2nd edn.



(A) $N = 100$



(B) $N = 500$

FIGURE 1: Simulation Results: Root Mean Squared Error ($\sigma^2_{\varepsilon_\alpha} = 1$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.

Table 1: Replication: Imai and Kim (2019)

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)	IFE1 (8)	IFE2 (9)
Panel A. Formal GATT Membership — No Year FEs									
GATT (Formal)	-0.048**	0.051**	0.063*	0.006	0.021	0.007	0.008		
	(0.024)	(0.024)	(0.033)	(0.017)	(0.020)	(0.022)	(0.022)		
LW Test		p=0.000							
N	196,207	175,000	160,916						
Panel B. Formal GATT Membership — With Year FEs									
GATT (Formal)	0.036	0.037	0.061*	0.006	0.021	0.007	0.008	0.056**	0.023
	(0.024)	(0.024)	(0.033)	(0.017)	(0.020)	(0.022)	(0.022)	(0.024)	(0.021)
LW Test		p=0.000							
N	196,207	175,000	160,916					196,207	196,207
Panel C. Informal GATT Membership — No Year FEs									
GATT (Participate)	0.147***	0.066***	0.060*	0.041***	0.064***	0.021	0.025		
	(0.030)	(0.025)	(0.036)	(0.014)	(0.014)	(0.021)	(0.021)		
LWTest		p=0.000							
N	196,207	175,000	160,916						
Panel D. Informal GATT Membership — With Year FEs									
GATT (Participate)	0.227***	0.044*	0.054	0.041***	0.064***	0.021	0.025	0.262***	0.014
	(0.030)	(0.026)	(0.036)	(0.014)	(0.014)	(0.021)	(0.021)	(0.029)	(0.027)
LW Test		p=0.000							
N	196,207	175,000	160,916					196,207	196,207

Notes: Dependent variable: *Log(Bilateral Trade Volume)*. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. IFE1 and IFE2 are omitted in Panels A and C since IFE always includes time-varying factor(s). RFD estimates based on 46 rolling regressions. RTFD based on 45 rolling regressions. Formal membership includes only formal GATT members as in Rose (2004); informal includes nonmember participants as in Tomz *et al.* (2007). Other controls include log product real GDP, log product real GDP per capita, and indicators for Generalized System of Preferences, a regional free trade agreement, a currency union, and currently colonized. Full results are in Tables C.1–C.4 in the Appendix. * p <.10, ** p< .05, *** p<.01.

Table 2: Replication: Leipziger (2024) Public Services

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. No Interaction									
Democracy	-0.035***	-0.005***	0.001	-0.001	-0.000	0.000	-0.000	-0.031***	-0.031***
	(0.012)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.012)	(0.010)
GDP (per capita)	-0.050***	0.001	0.043***	-0.001	-0.000	-0.001	0.001	-0.063***	-0.045***
	(0.011)	(0.007)	(0.016)	(0.001)	(0.000)	(0.001)	(0.001)	(0.010)	(0.008)
LW Test		p=0.000							
N	15,191	15,004	14,818					15,191	15,191
Panel B. With Interactions									
Democracy	0.042*	0.001	-0.008	-0.001	-0.000	0.001	-0.000	0.019	0.006
	(0.025)	(0.005)	(0.006)	(0.003)	(0.002)	(0.001)	(0.000)	(0.018)	(0.009)
Predemocratic Inequality	0.727***	0.036***	-0.411***	-0.002	0.000	-0.000	-0.000	0.979***	0.966***
	(0.030)	(0.013)	(0.020)	(0.003)	(0.000)	(0.001)	(0.001)	(0.013)	(0.005)
Democracy ×	-0.200***	-0.013	0.026**	0.343***	0.005	-0.005	-0.006**	-0.153***	-0.097***
Predemocratic Inequality	(0.050)	(0.010)	(0.012)	(0.076)	(0.010)	(0.004)	(0.003)	(0.041)	(0.021)
GDP (per capita)	-0.045***	0.001	0.039***	-0.000	-0.000	-0.000	0.000**	-0.004	-0.002
	(0.006)	(0.007)	(0.014)	(0.001)	(0.000)	(0.002)	(0.000)	(0.003)	(0.002)
LW Test		p=0.000							
N	15,102	14,915	14,729					15,102	15,102

Notes: Dependent variable: Indicator for *Ethnic Inequality in Access to Basic Services*. N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 119 rolling regressions. RTFD based on 118 rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

Table 3: Replication: Leipziger (2024) Education

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. No Interaction									
Democracy	-0.005	-0.001	0.002	0.001	-0.001	-0.001	-0.000	-0.011**	-0.007
	(0.007)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.005)	(0.004)
GDP (per capita)	0.013	0.004	0.014	-0.007	-0.016**	-0.002	0.001	-0.012	-0.017*
	(0.015)	(0.012)	(0.022)	(0.007)	(0.006)	(0.003)	(0.001)	(0.008)	(0.009)
LW Test		p=0.273							
N	3,945	3,848	3,752					3,945	3,945
Panel B. With Interactions									
Democracy	0.012**	-0.001	0.005*	-0.004***	-0.005***	-0.001	-0.000	-0.003	-0.003
	(0.006)	(0.002)	(0.003)	(0.001)	(0.000)	(0.001)	(0.001)	(0.004)	(0.004)
Predemocratic Inequality	0.718***	-0.018	-0.436***	0.699***	0.910***	-0.142***	0.019***	0.863***	0.879***
	(0.034)	(0.047)	(0.049)	(0.056)	(0.027)	(0.038)	(0.005)	(0.024)	(0.021)
Democracy ×	-0.197***	0.001	-0.017	-0.039***	-0.038***	-0.004***	-0.005***	-0.108***	-0.109***
Predemocratic Inequality	(0.056)	(0.014)	(0.019)	(0.004)	(0.008)	(0.001)	(0.001)	(0.041)	(0.040)
GDP (per capita)	-0.011*	0.004	0.004	-0.002	-0.001	-0.002	-0.001	-0.006***	-0.005***
	(0.006)	(0.012)	(0.022)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
LW Test		p=0.000							
N	3,865	3,769	3,673					3,865	3,865

Notes: Dependent variable: *Ethnic Inequality in Educational Attainment.* N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 66 rolling regressions. RTFD based on 65 rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

Table 4: Replication: James (2015) Full Sample

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Dependent Variable: Nonresource Revenue/Private Income									
Resource Revenue	0.006	-0.020***	-0.014**	-0.027***	-0.025***	-0.019**	-0.025***	-0.028	-0.018
	(0.020)	(0.005)	(0.005)	(0.009)	(0.003)	(0.009)	(0.002)	(0.020)	(0.018)
LW Test		p= 0.106							
N	2,550	2,500	2,450					2,550	2,550
Panel B. Dependent Variable: Income Tax Revenue/Private Income									
Resource Revenue	0.018*	0.022***	0.013***	0.014*	0.001	0.004	-0.001	0.022***	0.072***
	(0.010)	(0.001)	(0.001)	(0.007)	(0.001)	(0.007)	(0.001)	(0.008)	(0.019)
LW Test		p=0.671							
N	2,550	2,500	2,450					2,550	2,550
Panel C. Dependent Variable: Total Expenditures/Private Income									
Resource Revenue	0.397***	-0.003	-0.020***	-0.008	-0.056***	-0.013*	-0.033***	0.328***	0.109**
	(0.006)	(0.003)	(0.002)	(0.012)	(0.010)	(0.008)	(0.001)	(0.021)	(0.048)
LW Test		p= 0.000							
N	2,550	2,500	2,450					2,550	2,550
Panel D. Dependent Variable: Education Expenditures/Private Income									
Resource Revenue	0.063***	-0.010***	-0.019***	-0.003	-0.034***	-0.015***	-0.018***	0.061***	0.075***
	(0.007)	(0.001)	(0.001)	(0.005)	(0.005)	(0.003)	(0.002)	(0.008)	(0.007)
LW Test		p=0.000							
N	2,550	2,500	2,450					2,550	2,550
Panel E. Dependent Variable: Public Savings/Private Income									
Resource Revenue	0.609***	0.983***	1.006***	1.001***	1.043***	0.996***	1.006***	0.669***	0.224***
	(0.021)	(0.007)	(0.005)	(0.013)	(0.006)	(0.009)	(0.001)	(0.038)	(0.068)
LW Test		p=0.000							
N	2,550	2,500	2,450					2,550	2,550

Notes: N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 50 rolling regressions. RTFD based on 49 rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

TABLE 5: Replication: James (2015) Omit Alaska

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)	IFE1 (8)	IFE2 (9)
Panel A. Dependent Variable: Nonresource Revenue/Private Income									
Resource Revenue	-0.249***	-0.255	-0.340	-0.124	-0.137	-0.212*	-0.229*	-0.316***	-0.224
	(0.061)	(0.169)	(0.327)	(0.094)	(0.105)	(0.114)	(0.114)	(0.061)	(0.219)
LW Test		p=0.969							
N	2,499	2,450	2,401					2,499	2,499
Panel B. Dependent Variable: Income Tax Revenue/Private Income									
Resource Revenue	-0.104**	-0.010	0.016	0.009	0.001	0.019	0.012	-0.067*	-0.076
	(0.039)	(0.030)	(0.040)	(0.016)	(0.014)	(0.015)	(0.011)	(0.036)	(0.056)
LW Test		p= 0.039							
N	2,499	2,450	2,401					2,499	2,499
Panel C. Dependent Variable: Total Expenditures/Private Income									
Resource Revenue	0.429***	0.191**	0.000	0.101	0.198**	-0.118*	-0.121	0.429***	0.563**
	(0.063)	(0.085)	(0.172)	(0.071)	(0.088)	(0.060)	(0.088)	(0.090)	(0.232)
LW Test		p= 0.004							
N	2,499	2,450	2,401					2,499	2,499
Panel D. Dependent Variable: Education Expenditures/Private Income									
Resource Revenue	0.147***	0.017	-0.042	-0.016	-0.023	-0.056	-0.078*	0.156***	0.136**
	(0.036)	(0.039)	(0.077)	(0.034)	(0.035)	(0.039)	(0.040)	(0.037)	(0.063)
LW Test		p=0.012							
N	2,499	2,450	2,401					2,499	2,499
Panel E. Dependent Variable: Public Savings/Private Income									
Resource Revenue	0.322***	0.553***	0.661***	0.761***	0.544***	0.931***	0.755***	0.214***	0.230***
	(0.038)	(0.139)	(0.197)	(0.112)	(0.115)	(0.112)	(0.132)	(0.060)	(0.076)
LW Test		p= 0.133							
N	2,499	2,450	2,401					2,499	2,499

Notes: N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 50 rolling regressions. RTFD based on 49 rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

Table 6: Replication: Djankov and Reynal-Querol (2010)

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Dependent Variable: Incidence of Civil Wars (25 Deaths)									
A(I). Five-Year Intervals									
$log(GDP_{t-1})$	-0.087*	-0.035	0.005	-0.027	0.011	0.042	0.041	0.002	0.007
	(0.046)	(0.060)	(0.120)	(0.071)	(0.063)	(0.111)	(0.106)	(0.032)	(0.041)
LW Test		p=0.528							
N	1,169	985	816					1,169	1,169
A(II). Ten-Year Intervals									
$log(GDP_{t-1})$	-0.059	-0.044	-0.003	-0.022	-0.021	0.069	0.080	0.028	0.088
	(0.057)	(0.064)	(0.153)	(0.063)	(0.051)	(0.136)	(0.151)	(0.054)	(0.130)
LW Test		p=0.481							
N	576	407	251					576	576
Panel B. Dependent Variable: Onset of War (ACD, 1,000+ Deaths)									
B(I). Five-Year Intervals									
$log(GDP_{t-1})$	0.030	0.059	0.180**	0.040	0.028	0.094	0.071	-0.018***	-0.018***
	(0.023)	(0.042)	(0.082)	(0.040)	(0.037)	(0.058)	(0.056)	(0.005)	(0.006)
LW Test		p=0.096							
N	1,169	985	816					1,169	1,169
B(II). Ten-Year Intervals									
$log(GDP_{t-1})$	0.082*	0.089	0.037	0.115	0.057	0.024	-0.021	-0.025**	-0.027
	(0.048)	(0.060)	(0.131)	(0.080)	(0.046)	(0.088)	(0.151)	(0.011)	(0.018)
LW Test		p=0.110							
N	576	407	251					576	576
Panel C. Dependent Variable: Incidence of Civil War (1,000 Deaths per Year)									
C(I). Five-Year Intervals									
$log(GDP_{t-1})$	-0.035	-0.040	-0.057	-0.004	0.012	-0.033	-0.035	-0.011	-0.015
	(0.035)	(0.041)	(0.070)	(0.026)	(0.024)	(0.037)	(0.037)	(0.010)	(0.011)
LW Test		p=0.919							
N	1,169	985	816					1,169	1,169
C(II). Ten-Year Intervals									
$log(GDP_{t-1})$	-0.013	0.029	0.098	0.018	0.009	0.095	0.091	0.016	-0.001
	(0.056)	(0.058)	(0.094)	(0.040)	(0.021)	(0.036)	(0.069)	(0.025)	(0.034)
LW Test		p=0.232							
N	576	407	251					576	576

Notes: Dependent variables are indicator variables. N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 7 (Panels A(I), B(I), and C(I)) and 3 (Panels A(II), B(II), and C(II)) rolling regressions. RTFD based on 6 (Panels A(I), B(I), and C(I)) and 2 (Panels A(II), B(II), and C(II)) rolling regressions. $log(POP_{t-1})$ and time fixed effects included in all models. Full results are in Tables C.5–C.7 in the Appendix. * p <0.10, ** p<0.05, *** p<0.01.

On the (Mis)Use of the Fixed Effects Estimator

Supplemental Appendix

Daniel L. Millimet & Marc F. Bellemare

November 2, 2025

A Proofs

Declaration: AI tools were used extensively in this Section (specifically, Google Gemini 2.5 Pro). The authors are responsible and accountable for the analysis, accuracy, interpretation, etc. of all content provided.

A.1 Setup

Assume the DGP is

$$\begin{aligned}
y_{it} &= \alpha_0 + x_{it}\beta + \alpha_{it} + \varepsilon_{it}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T} \\
x_{it} &= \lambda \alpha_{it} + z_{it} \\
\alpha_{it} &= \delta \alpha_{it-1} + \varepsilon_{\alpha,it} \\
z_{it} &= \rho z_{it-1} + \varepsilon_{z,it} \\
\varepsilon_{it} &\sim WN(0, \sigma_\varepsilon^2) \\
\varepsilon_{\alpha,it} &\sim WN(0, \sigma_{\varepsilon_\alpha}^2) \\
\varepsilon_{z,it} &\sim WN(0, \sigma_{\varepsilon_z}^2) \\
\texttt{Cov}(z_{it}, \varepsilon_{is}) = \texttt{Cov}(z_{it}, \varepsilon_{\alpha,is}) &= 0 \quad \forall\, t, s \\
\texttt{Cov}(\varepsilon_{\alpha,it}, \varepsilon_{is}) = \texttt{Cov}(\varepsilon_{\alpha,it}, \varepsilon_{z,is}) = \texttt{Cov}(\varepsilon_{it}, \varepsilon_{z,is}) &= 0 \quad \forall\, t, s
\end{aligned}$$

A.2 Requirements for x_{it} to be Stationary

For estimation of the model in Section A.1 to be meaningful, y and x should be stationary. As such, we need to determine the parameter restrictions required for this to be the case.

Using the lag operator and the known properties of AR(1) processes, we can write

$$\begin{aligned}
\alpha_{it} &= \delta \alpha_{it-1} + \varepsilon_{\alpha,it} = \frac{1}{1-\delta L} \varepsilon_{\alpha,it} \\
z_{it} &= \rho z_{it-1} + \varepsilon_{z,it} = \frac{1}{1-\rho L} \varepsilon_{z,it}
\end{aligned}$$

where L is the lag operator. Substituting into the equation for x_{it} yields

$$\begin{aligned}
x_{it} &= \lambda \left(\frac{1}{1-\delta L} \varepsilon_{\alpha,it} \right) + \left(\frac{1}{1-\rho L} \varepsilon_{z,it} \right) \\
&= \frac{\lambda(1-\rho L)\varepsilon_{at} + (1-\delta L)\varepsilon_{zt}}{(1-\delta L)(1-\rho L)}.
\end{aligned}$$

Multiplying both sides by the denominator and expanding terms yields

$$x_t - (\delta + \rho) x_{t-1} + \delta\rho x_{t-2} = \lambda \varepsilon_{at} - \lambda\rho\varepsilon_{at-1} + \varepsilon_{z,it} - \delta \varepsilon_{z,it-1}$$

which can be re-written as

$$x_t = (\delta + \rho) x_{t-1} - \delta\rho x_{t-2} + \eta_{it},$$

where $\eta_{it} := \lambda\varepsilon_{\alpha,it} - \lambda\rho\varepsilon_{\alpha,it-1} + \varepsilon_{z,it} - \delta\varepsilon_{z,it-1}$ is an MA(1) process. So, x_{it} is an ARMA(2,1) process.

For an ARMA(2,1) process to be stationary, the roots of its autoregressive characteristic equation must lie outside the unit circle. The characteristic equation is

$$1 - (\delta + \rho)L + \delta\rho L^2 = 0$$

Letting r be the roots, we obtain $r_1 = 1/\delta$ and $r_2 = 1/\rho$. Consequently, stationarity requires

$$\begin{aligned} |r_1| > 1 \quad &\Longrightarrow \quad |1/\delta| > 1 \Longrightarrow |\delta| < 1 \\ |r_2| > 1 \quad &\Longrightarrow \quad |1/\rho| > 1 \Longrightarrow |\rho| < 1 \end{aligned}$$

The parameter λ affects the moving average part of the process, but it does not affect the stationarity. Restrictions on λ are required for invertibility if one wishes to estimate the full set of parameters of the ARMA process.

A.3 Estimator `plim`'s

The `plim`'s for the estimators considered are

$$\begin{aligned}
\text{POLS}: \qquad \mathtt{plim}\,\widehat{\beta}_{ols} &= \frac{\mathtt{Cov}(y,x)}{\mathtt{Var}(x)} = \frac{\mathtt{Cov}(\alpha_0 + \beta x + \alpha + \varepsilon, x)}{\mathtt{Var}(x)} \\
&= \beta + \frac{\mathtt{Cov}(\alpha,x)}{\mathtt{Var}(x)} = \beta + \lambda\frac{\mathtt{Var}(\alpha)}{\mathtt{Var}(x)} \\
\text{FD}: \qquad \mathtt{plim}\,\widehat{\beta}_{fd} &= \frac{\mathtt{Cov}(\Delta y,\Delta x)}{\mathtt{Var}(\Delta x)} = \frac{\mathtt{Cov}(\beta\Delta x + \Delta\alpha + \Delta\varepsilon, \Delta x)}{\mathtt{Var}(\Delta x)} \\
&= \beta + \frac{\mathtt{Cov}(\Delta\alpha,\Delta x)}{\mathtt{Var}(\Delta x)} = \beta + \lambda\frac{\mathtt{Var}(\Delta\alpha)}{\mathtt{Var}(\Delta x)} \\
\text{TFD}: \qquad \mathtt{plim}\,\widehat{\beta}_{tfd} &= \frac{\mathtt{Cov}(\Delta^2 y,\Delta^2 x)}{\mathtt{Var}(\Delta^2 x)} = \frac{\mathtt{Cov}(\beta\Delta^2 x + \Delta^2\alpha + \Delta^2\varepsilon, \Delta^2 x)}{\mathtt{Var}(\Delta^2 x)} \\
&= \beta + \frac{\mathtt{Cov}(\Delta^2\alpha,\Delta^2 x)}{\mathtt{Var}(\Delta^2 x)} = \beta + \lambda\frac{\mathtt{Var}(\Delta^2\alpha)}{\mathtt{Var}(\Delta^2 x)} \\
\text{FE}: \qquad \mathtt{plim}\,\widehat{\beta}_{fe} &= \frac{\mathtt{Cov}(\ddot{y},\ddot{x})}{\mathtt{Var}(\ddot{x})} = \frac{\mathtt{Cov}(\beta\ddot{x} + \ddot{\alpha} + \ddot{\varepsilon}, \ddot{x})}{\mathtt{Var}(\ddot{x})} \\
&= \beta + \frac{\mathtt{Cov}(\ddot{\alpha},\ddot{x})}{\mathtt{Var}(\ddot{x})} = \beta + \lambda\frac{\mathtt{Var}(\ddot{\alpha})}{\mathtt{Var}(\ddot{x})}
\end{aligned}$$

Thus, the bias term for each estimator is $\mathtt{Var}(f(\alpha))/\mathtt{Var}(f(x))$ for differing functions $f(\cdot)$. The basic building blocks for the bias terms are the properties of α, z, and x, where $x = \lambda\alpha + z$ and α and z are independent AR(1) processes:

- Unconditional Variance:

$$\begin{aligned}
\texttt{Var}(z) &= \frac{\sigma^2_{\varepsilon_z}}{1-\rho^2} \\
\texttt{Var}(\alpha) &= \frac{\sigma^2_{\varepsilon_\alpha}}{1-\delta^2} \\
\texttt{Var}(x) &= \lambda^2\texttt{Var}(\alpha) + \texttt{Var}(z) = \lambda^2\left[\frac{\sigma^2_{\varepsilon_\alpha}}{1-\delta^2}\right] + \left[\frac{\sigma^2_{\varepsilon_z}}{1-\rho^2}\right]
\end{aligned}$$

- First Difference Variance:

$$\begin{aligned}
\texttt{Var}(\Delta z) &= \texttt{Var}(z_{it} - z_{it-1}) \\
&= \texttt{Var}(z_{it}) + \texttt{Var}(z_{it-1}) - 2\texttt{Cov}(z_{it}, z_{it-1}) \\
&= 2\texttt{Var}(z) - 2\rho\texttt{Var}(z) \\
&= 2(1-\rho)\texttt{Var}(z) \\
\texttt{Var}(\Delta\alpha) &= 2(1-\delta)\texttt{Var}(\alpha) \\
\texttt{Var}(\Delta x) &= \lambda^2\texttt{Var}(\Delta\alpha) + \texttt{Var}(\Delta z) = 2\left[\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)\right]
\end{aligned}$$

- Second Difference Variance:

$$\begin{aligned}
\texttt{Var}(\Delta^2 z) &= \texttt{Var}(z_{it} - 2z_{it-1} + z_{it-2}) \\
&= \texttt{Var}(z) + 4\texttt{Var}(z) + \texttt{Var}(z) - 4\texttt{Cov}(z_{it}, z_{it-1}) + 2\texttt{Cov}(z_{it}, z_{it-2}) - 4\texttt{Cov}(z_{it-1}, z_{it-2}) \\
&= 6\texttt{Var}(z) - 4\rho\texttt{Var}(z) + 2\rho^2\texttt{Var}(z) - 4\rho\texttt{Var}(z) \\
&= (6 - 8\rho + 2\rho^2)\texttt{Var}(z) \\
&= 2(1-\rho)(3-\rho)\texttt{Var}(z) \\
\texttt{Var}(\Delta^2\alpha) &= 2(1-\delta)(3-\delta)\texttt{Var}(\alpha) \\
\texttt{Var}(\Delta^2 x) &= \lambda^2\texttt{Var}(\Delta^2\alpha) + \texttt{Var}(\Delta^2 z) = 2\left[\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)\right]
\end{aligned}$$

since $\texttt{Cov}(z_{it}, z_{it-k}) = \rho^k\texttt{Var}(z)$ and similarly for α.

- Within Variance:

$$\begin{aligned}
\texttt{Var}(\ddot{z}) &\approx g(\rho, T)\texttt{Var}(z) \\
\texttt{Var}(\ddot{\alpha}) &\approx g(\delta, T)\texttt{Var}(\alpha) \\
\texttt{Var}(\ddot{x}) &= \lambda^2\texttt{Var}(\ddot{\alpha}) + \texttt{Var}(\ddot{z}) \approx \lambda^2 g(\delta, T)\texttt{Var}(\alpha) + g(\rho, T)\texttt{Var}(z)
\end{aligned}$$

where $g(\rho, T) = 1 - \frac{1}{T}\left[\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^T)}{T(1-\rho)^2}\right]$ and similarly for α.

Substitution yields

$$
\begin{aligned}
\text{POLS}: \quad & \texttt{plim}\ \widehat{\beta}_{ols} &&= \beta + \lambda \frac{\texttt{Var}(\alpha)}{\texttt{Var}(x)} \\
\text{FD}: \quad & \texttt{plim}\ \widehat{\beta}_{fd} &&= \beta + \lambda \frac{\texttt{Var}(\Delta\alpha)}{\texttt{Var}(\Delta x)} \\
& &&= \beta + \frac{\lambda(1-\delta)\texttt{Var}(\alpha)}{\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)} \\
\text{TFD}: \quad & \texttt{plim}\ \widehat{\beta}_{tfd} &&= \beta + \lambda \frac{\texttt{Var}(\Delta^2\alpha)}{\texttt{Var}(\Delta^2 x)} \\
& &&= \beta + \frac{\lambda(1-\delta)(3-\delta)\texttt{Var}(\alpha)}{\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)} \\
\text{FE}: \quad & \texttt{plim}\ \widehat{\beta}_{fe} &&= \beta + \lambda \frac{\texttt{Var}(\ddot{\alpha})}{\texttt{Var}(\ddot{x})} \\
& &&\approx \beta + \frac{\lambda g(\delta,T)\texttt{Var}(\alpha)}{\lambda^2 g(\delta,T)\texttt{Var}(\alpha) + g(\rho,T)\texttt{Var}(z)}
\end{aligned}
$$

A.4 Comparison of biases

As stated in Propositions 1 and 2, the absolute bias of each estimator can be ranked as follows:

$$
\begin{aligned}
\text{If } \delta = \rho: \quad & |\text{Bias}(\widehat{\beta}_{ols})| = |\text{Bias}(\widehat{\beta}_{fe})| = |\text{Bias}(\widehat{\beta}_{fd})| = |\text{Bias}(\widehat{\beta}_{tfd})| \\
\text{If } \delta < \rho: \quad & |\text{Bias}(\widehat{\beta}_{tfd})| > |\text{Bias}(\widehat{\beta}_{fd})| > |\text{Bias}(\widehat{\beta}_{fe})| \gtrsim |\text{Bias}(\widehat{\beta}_{ols})| \\
\text{If } \delta > \rho: \quad & |\text{Bias}(\widehat{\beta}_{ols})| \gtrsim |\text{Bias}(\widehat{\beta}_{fe})| > |\text{Bias}(\widehat{\beta}_{fd})| > |\text{Bias}(\widehat{\beta}_{tfd})|
\end{aligned}
$$

where

$$
\begin{aligned}
|\text{Bias}(\widehat{\beta}_{ols})| &= \left| \frac{\lambda\texttt{Var}(\alpha)}{\lambda^2\texttt{Var}(\alpha) + \texttt{Var}(z)} \right| \\
|\text{Bias}(\widehat{\beta}_{fd})| &= \left| \frac{\lambda(1-\delta)\texttt{Var}(\alpha)}{\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)} \right| \\
|\text{Bias}(\widehat{\beta}_{tfd})| &= \left| \frac{\lambda(1-\delta)(3-\delta)\texttt{Var}(\alpha)}{\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)} \right| \\
|\text{Bias}(\widehat{\beta}_{fe})| &= \left| \frac{\lambda g(\delta,T)\texttt{Var}(\alpha)}{\lambda^2 g(\delta,T)\texttt{Var}(\alpha) + g(\rho,T)\texttt{Var}(z)} \right|
\end{aligned}
$$

Proof: Let $Q := \frac{\mathtt{Var}(z)}{\lambda^2 \mathtt{Var}(\alpha)}$. Then

$$|\text{Bias}(\widehat{\beta}_{ols})| = \left| \frac{1}{\lambda(1+Q)} \right|$$

$$|\text{Bias}(\widehat{\beta}_{fd})| = \left| \frac{1}{\lambda \left[1 + \left(\frac{1-\rho}{1-\delta}\right) Q\right]} \right|$$

$$|\text{Bias}(\widehat{\beta}_{tfd})| = \left| \frac{1}{\lambda \left[1 + \frac{(1-\rho)(3-\rho)}{(1-\delta)(3-\delta)} Q\right]} \right|$$

$$|\text{Bias}(\widehat{\beta}_{fe})| = \left| \frac{1}{\lambda \left[1 + \frac{g(\rho,T)}{g(\delta,T)} Q\right]} \right|$$

Assuming $\lambda, \rho, \delta > 0$ and strictly positive variances, the biases are positive in all cases. Thus, we can drop the absolute value signs. The relative magnitudes of the biases depend on the scalars multiplying Q in the denominator (ignoring the common factor $1/\lambda$). Let the scalar be k_j for estimator $j \in \{POLS, FD, TFD, FE\}$. A larger k_j implies a smaller bias for estimator j.

Case 1: $\rho > \delta \geq 0$ If the AR(1) process for z is more persistent than for α:

- $k_{fd} = \frac{1-\rho}{1-\delta} < 1 = k_{ols} \implies \text{Bias}(\widehat{\beta}_{\mathbf{fd}}) > \text{Bias}(\widehat{\beta}_{\mathbf{ols}})$.
- Comparing k_{tfd} to k_{fd}: $\frac{k_{tfd}}{k_{fd}} = \frac{3-\rho}{3-\delta} < 1 \implies k_{tfd} < k_{fd} \implies \text{Bias}(\widehat{\beta}_{\mathbf{tfd}}) > \text{Bias}(\widehat{\beta}_{\mathbf{fd}})$.

This gives the ranking: $|\text{Bias}(\widehat{\beta}_{\mathbf{tfd}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{fd}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{ols}})|$.

Case 2: $\delta > \rho \geq 0$ If the AR(1) process for α is more persistent than for z:

- $k_{fd} = \frac{1-\rho}{1-\delta} > 1 = k_{ols} \implies \text{Bias}(\widehat{\beta}_{\mathbf{fd}}) < \text{Bias}(\widehat{\beta}_{\mathbf{ols}})$.
- Comparing k_{tfd} to k_{fd}: $\frac{k_{tfd}}{k_{fd}} = \frac{3-\rho}{3-\delta} > 1 \implies k_{tfd} > k_{fd} \implies \text{Bias}(\widehat{\beta}_{\mathbf{tfd}}) < \text{Bias}(\widehat{\beta}_{\mathbf{fd}})$.

This gives the ranking: $|\text{Bias}(\widehat{\beta}_{\mathbf{ols}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{fd}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{tfd}})|$.

Detailed Comparison of $|P|$ and $|E|$ for Large T As $T \to \infty$, $g(\delta, T) \to 1$ and $g(\rho, T) \to 1$. Consequently, $k_{fe} = \frac{g(\rho,T)}{g(\delta,T)} \to 1$, which implies that $\text{Bias}(\widehat{\beta}_{\mathbf{fe}}) \to \text{Bias}(\widehat{\beta}_{\mathbf{ols}})$. For a large but finite T, the comparison depends on whether k_{fe} is greater or less than $k_{ols} = 1$. This is determined by the relative magnitudes of $g(\rho, T)$ and $g(\delta, T)$. The function $g(\phi, T)$ is a decreasing function of the term $h(\phi, T) = \frac{1}{T}\left(\frac{1+\phi}{1-\phi} - \dots\right)$. For large T, the behavior of $h(\phi, T)$ is dominated by its first component, $m(\phi) = \frac{1+\phi}{1-\phi}$, which is a monotonically increasing function of ϕ for $\phi \in [0, 1)$.

- **Case 1:** $\rho > \delta > 0$: Since $m(\phi)$ is increasing in ϕ, we have $m(\rho) > m(\delta)$. This implies that, for large T, $h(\rho, T) > h(\delta, T)$. Because $g = 1 - h$, it follows that $g(\rho, T) < g(\delta, T)$. Therefore, $k_{fe} = \frac{g(\rho,T)}{g(\delta,T)} < 1 = k_{ols}$. This leads to the conclusion: $|\text{Bias}(\widehat{\beta}_{\mathbf{fe}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{ols}})|$.
- **Case 2** $\delta > \rho > 0$: By the same logic, $m(\delta) > m(\rho)$ implies $h(\delta, T) > h(\rho, T)$, and thus $g(\delta, T) < g(\rho, T)$. Therefore, $k_{fe} = \frac{g(\rho,T)}{g(\delta,T)} > 1 = k_{ols}$. This leads to the conclusion: $|\text{Bias}(\widehat{\beta}_{\mathbf{fe}})| < |\text{Bias}(\widehat{\beta}_{\mathbf{ols}})|$.

The within-mean transformation removes more variance from the more persistent series. When $\rho > \delta$, demeaning is more effective on z, reducing its relative variance. This increases the bias in FE compared to $POLS$. Conversely, when $\delta > \rho$, demeaning is more effective on α, reducing its relative variance and thus decreasing the bias of FE compared to $POLS$.

A.5 x Depends on Lagged α

The DGP is identical to before except now $x_{it} = \lambda\alpha_{it-1} + z_{it}$. The `plim`'s for the estimators considered are

$$
\begin{aligned}
\text{POLS}: \quad \texttt{plim}\,\widehat{\beta}_{ols} &= \frac{\texttt{Cov}(y_t, x_t)}{\texttt{Var}(x_t)} = \frac{\texttt{Cov}(\alpha_0 + \beta x_t + \alpha_t + \varepsilon_t, x_t)}{\texttt{Var}(x_t)} \\
&= \beta + \frac{\texttt{Cov}(\alpha_t, x_t)}{\texttt{Var}(x_t)} = \beta + \frac{\texttt{Cov}(\delta\alpha_{t-1}, \lambda\alpha_{t-1} + z_t)}{\texttt{Var}(x_t)} = \beta + \lambda\delta\frac{\texttt{Var}(\alpha)}{\texttt{Var}(x)} \\
\text{FD}: \quad \texttt{plim}\,\widehat{\beta}_{fd} &= \frac{\texttt{Cov}(\Delta y_t, \Delta x_t)}{\texttt{Var}(\Delta x_t)} = \frac{\texttt{Cov}(\beta\Delta x_t + \Delta\alpha_t + \Delta\varepsilon_t, \Delta x_t)}{\texttt{Var}(\Delta x_t)} \\
&= \beta + \frac{\texttt{Cov}(\Delta\alpha_t, \Delta x_t)}{\texttt{Var}(\Delta x_t)} = \beta + \frac{\texttt{Cov}(\Delta\alpha_t, \lambda\Delta\alpha_{t-1} + \Delta z_t)}{\texttt{Var}(\Delta x_t)} \\
&= \beta + \lambda\frac{\texttt{Cov}(\Delta\alpha_t, \Delta\alpha_{t-1})}{\texttt{Var}(\Delta x_t)} = \beta - \lambda(1-\delta)^2\frac{\texttt{Var}(\alpha)}{\texttt{Var}(\Delta x)} \\
\text{TFD}: \quad \texttt{plim}\,\widehat{\beta}_{tfd} &= \frac{\texttt{Cov}(\Delta^2 y_t, \Delta^2 x_t)}{\texttt{Var}(\Delta^2 x_t)} = \frac{\texttt{Cov}(\beta\Delta^2 x_t + \Delta^2\alpha_t + \Delta^2\varepsilon_t, \Delta^2 x_t)}{\texttt{Var}(\Delta^2 x_t)} \\
&= \beta + \frac{\texttt{Cov}(\Delta^2\alpha_t, \Delta^2 x_t)}{\texttt{Var}(\Delta^2 x_t)} = \beta + \frac{\texttt{Cov}(\Delta^2\alpha_t, \lambda\Delta^2\alpha_{t-1} + \Delta^2 z_t)}{\texttt{Var}(\Delta^2 x_t)} \\
&= \beta + \lambda\frac{\texttt{Cov}(\Delta^2\alpha_t, \Delta^2\alpha_{t-1})}{\texttt{Var}(\Delta^2 x_t)} = \beta + \lambda\left(\delta^3 - 4\delta^2 + 7\delta - 4\right)\frac{\texttt{Var}(\alpha)}{\texttt{Var}(\Delta^2 x)} \\
\text{FE}: \quad \texttt{plim}\,\widehat{\beta}_{fe} &= \frac{\texttt{Cov}(\ddot{y}_t, \ddot{x}_t)}{\texttt{Var}(\ddot{x}_t)} = \frac{\texttt{Cov}(\beta\ddot{x}_t + \ddot{\alpha}_t + \ddot{\varepsilon}_t, \ddot{x}_t)}{\texttt{Var}(\ddot{x}_t)} \\
&= \beta + \frac{\texttt{Cov}(\ddot{\alpha}_t, \ddot{x}_t)}{\texttt{Var}(\ddot{x}_t)} = \beta + \frac{\texttt{Cov}(\ddot{\alpha}_t, \lambda\ddot{\alpha}_{t-1} + \ddot{z}_t)}{\texttt{Var}(\ddot{x}_t)} \\
&= \beta + \frac{\texttt{Cov}(\ddot{\alpha}_t, \lambda\ddot{\alpha}_{t-1})}{\texttt{Var}(\ddot{x}_t)} \approx \beta + \lambda\delta\frac{\texttt{Var}(\alpha)}{\texttt{Var}(\ddot{x})}
\end{aligned}
$$

Thus, the bias term for each estimator is $f_0(\texttt{Var}(\alpha))/\texttt{Var}(f_1(x))$ for differing functions $f_0(\cdot), f_1(\cdot)$. The basic building blocks for the bias terms are the properties of α, z, and x, where $x = \lambda\alpha + z$ and α and z are independent AR(1) processes:

- Unconditional Variance:

$$
\begin{aligned}
\texttt{Var}(z) &= \frac{\sigma^2_{\varepsilon_z}}{1-\rho^2} \\
\texttt{Var}(\alpha) &= \frac{\sigma^2_{\varepsilon_\alpha}}{1-\delta^2} \\
\texttt{Var}(x) &= \lambda^2\texttt{Var}(\alpha) + \texttt{Var}(z) = \lambda^2\left[\frac{\sigma^2_{\varepsilon_\alpha}}{1-\delta^2}\right] + \left[\frac{\sigma^2_{\varepsilon_z}}{1-\rho^2}\right]
\end{aligned}
$$

- First Difference Variance:

$$\begin{aligned}
\texttt{Var}(\Delta z) &= \texttt{Var}(z_{it} - z_{it-1}) \\
&= \texttt{Var}(z_{it}) + \texttt{Var}(z_{it-1}) - 2\texttt{Cov}(z_{it}, z_{it-1}) \\
&= 2\texttt{Var}(z) - 2\rho\texttt{Var}(z) \\
&= 2(1-\rho)\texttt{Var}(z) \\
\texttt{Var}(\Delta\alpha) &= 2(1-\delta)\texttt{Var}(\alpha) \\
\texttt{Var}(\Delta x) &= \lambda^2\texttt{Var}(\Delta\alpha) + \texttt{Var}(\Delta z) = 2\left[\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)\right]
\end{aligned}$$

- Second Difference Variance:

$$\begin{aligned}
\texttt{Var}(\Delta^2 z) &= \texttt{Var}(z_{it} - 2z_{it-1} + z_{it-2}) \\
&= \texttt{Var}(z) + 4\texttt{Var}(z) + \texttt{Var}(z) - 4\texttt{Cov}(z_{it}, z_{it-1}) + 2\texttt{Cov}(z_{it}, z_{it-2}) - 4\texttt{Cov}(z_{it-1}, z_{it-2}) \\
&= 6\texttt{Var}(z) - 4\rho\texttt{Var}(z) + 2\rho^2\texttt{Var}(z) - 4\rho\texttt{Var}(z) \\
&= (6 - 8\rho + 2\rho^2)\texttt{Var}(z) \\
&= 2(1-\rho)(3-\rho)\texttt{Var}(z) \\
\texttt{Var}(\Delta^2\alpha) &= 2(1-\delta)(3-\delta)\texttt{Var}(\alpha) \\
\texttt{Var}(\Delta^2 x) &= \lambda^2\texttt{Var}(\Delta^2\alpha) + \texttt{Var}(\Delta^2 z) = 2\left[\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)\right]
\end{aligned}$$

  since $\texttt{Cov}(z_{it}, z_{it-k}) = \rho^k\texttt{Var}(z)$ and similarly for α.

- Within Variance:

$$\begin{aligned}
\texttt{Var}(\ddot{z}) &\approx g(\rho, T)\texttt{Var}(z) \\
\texttt{Var}(\ddot{\alpha}) &\approx g(\delta, T)\texttt{Var}(\alpha) \\
\texttt{Var}(\ddot{x}) &= \lambda^2\texttt{Var}(\ddot{\alpha}) + \texttt{Var}(\ddot{z}) \approx \lambda^2 g(\delta, T)\texttt{Var}(\alpha) + g(\rho, T)\texttt{Var}(z)
\end{aligned}$$

  where $g(\rho, T) = 1 - \frac{1}{T}\left[\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^T)}{T(1-\rho)^2}\right]$ and similarly for α.

Substitution yields

$$\begin{aligned}
\text{POLS}: \quad & \texttt{plim}\,\widehat{\beta}_{ols} &= \beta + \frac{\lambda\delta\texttt{Var}(\alpha)}{\lambda^2\texttt{Var}(\alpha) + \texttt{Var}(z)} \\
\text{FD}: \quad & \texttt{plim}\,\widehat{\beta}_{fd} &= \beta + \frac{-\lambda(1-\delta)^2\texttt{Var}(\alpha)}{2\left[\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)\right]} \\
\text{TFD}: \quad & \texttt{plim}\,\widehat{\beta}_{tfd} &= \beta + \frac{\lambda\left(\delta^3 - 4\delta^2 + 7\delta - 4\right)\texttt{Var}(\alpha)}{2\left[\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)\right]} \\
\text{FE}: \quad & \texttt{plim}\,\widehat{\beta}_{fe} &\approx \beta + \frac{\lambda\delta\texttt{Var}(\alpha)}{\lambda^2 g(\delta, T)\texttt{Var}(\alpha) + g(\rho, T)\texttt{Var}(z)}
\end{aligned}$$

A.6 Comparison of biases

Now, the absolute bias of each estimator is given by:

$$
\begin{aligned}
|\text{Bias}(\widehat{\beta}_{ols})| &= \left| \frac{\lambda\delta \texttt{Var}(\alpha)}{\lambda^2 \texttt{Var}(\alpha) + \texttt{Var}(z)} \right| \\
|\text{Bias}(\widehat{\beta}_{fd})| &= \left| \frac{-\lambda(1-\delta)^2 \texttt{Var}(\alpha)}{2\left[\lambda^2(1-\delta)\texttt{Var}(\alpha) + (1-\rho)\texttt{Var}(z)\right]} \right| \\
|\text{Bias}(\widehat{\beta}_{tfd})| &= \left| \frac{\lambda\left(\delta^3 - 4\delta^2 + 7\delta - 4\right) \texttt{Var}(\alpha)}{2\left[\lambda^2(1-\delta)(3-\delta)\texttt{Var}(\alpha) + (1-\rho)(3-\rho)\texttt{Var}(z)\right]} \right| \\
|\text{Bias}(\widehat{\beta}_{fe})| &= \left| \frac{\lambda\delta \texttt{Var}(\alpha)}{\lambda^2 g(\delta, T)\texttt{Var}(\alpha) + g(\rho, T)\texttt{Var}(z)} \right|
\end{aligned}
$$

Let $Q := \frac{\texttt{Var}(z)}{\lambda^2 \texttt{Var}(\alpha)}$. Then

$$
\begin{aligned}
|\text{Bias}(\widehat{\beta}_{ols})| &= \left| \frac{\delta}{\lambda(1+Q)} \right| \\
|\text{Bias}(\widehat{\beta}_{fd})| &= \left| \frac{-(1-\delta)}{2\lambda\left[1 + \frac{1-\rho}{1-\delta}Q\right]} \right| \\
|\text{Bias}(\widehat{\beta}_{tfd})| &= \left| \frac{\delta^3 - 4\delta^2 + 7\delta - 4}{2\lambda(1-\delta)(3-\delta)\left[1 + \frac{(1-\rho)(3-\rho)}{(1-\delta)(3-\delta)}Q\right]} \right| \\
|\text{Bias}(\widehat{\beta}_{fe})| &= \left| \frac{\delta}{\lambda(1+Q)} \right|
\end{aligned}
$$

where $\text{Bias}(\widehat{\beta}_{fe})$ assumes large T. Assuming $\lambda, \rho, \delta > 0$ and strictly positive variances, the biases are positive for $POLS$ and FE but negative for FD and TFD. In general, it is not possible to rank the magnitudes. When relaxing the assumption of large T, there is a small difference between $POLS$ and FE that is identical to Section A.4.

- **Case 1:** $\rho > \delta > 0$: $|\text{Bias}(\widehat{\beta}_{\mathbf{fe}})| > |\text{Bias}(\widehat{\beta}_{\mathbf{ols}})|$.
- **Case 2** $\delta > \rho > 0$: $|\text{Bias}(\widehat{\beta}_{\mathbf{fe}})| < |\text{Bias}(\widehat{\beta}_{\mathbf{ols}})|$.

B Monte Carlo Study

We simulate data from four experimental designs nested in the following setup:

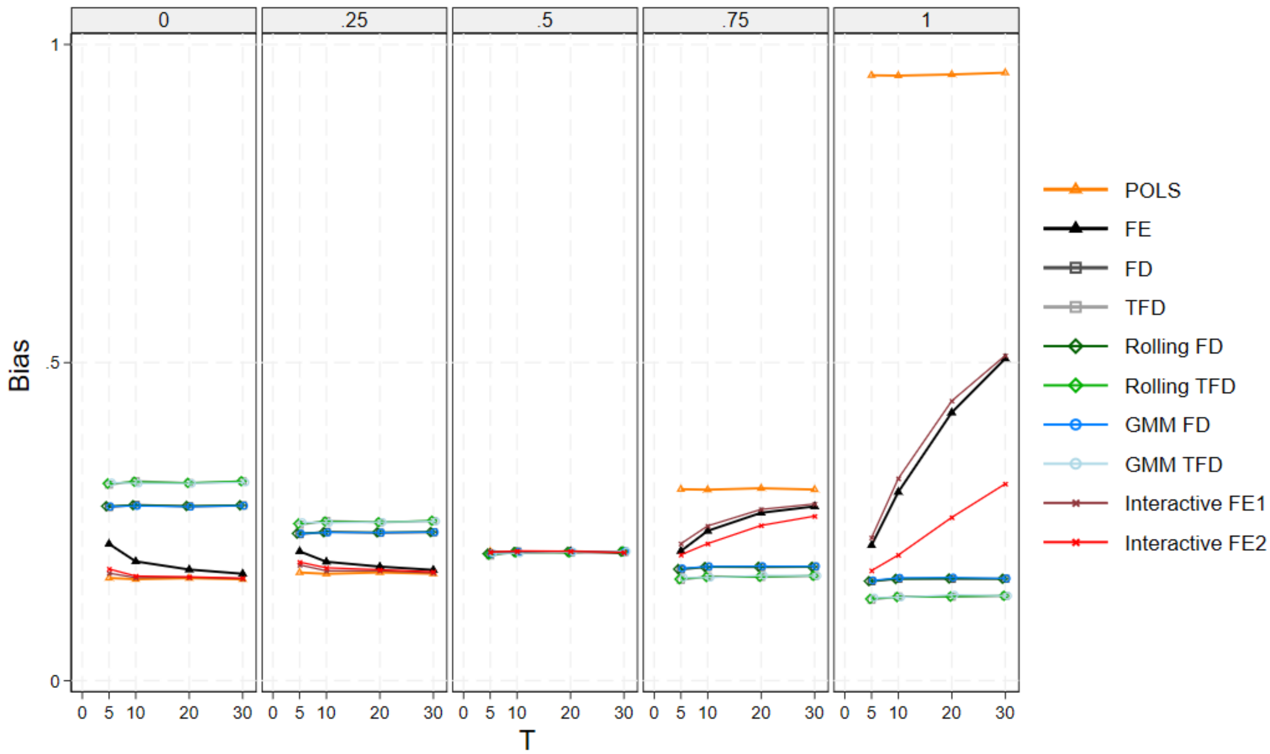
$$\begin{aligned}
y_{it} &= \beta x_{it} + \alpha_{it} + \varepsilon_{it} \\
x_{it} &= \lambda \alpha_{it} + z_{it} \\
\alpha_{it} &= \delta \alpha_{it-1} + \varepsilon_{\alpha,it} \\
z_{it} &= \rho z_{it-1} + \varepsilon_{z,it} \\
\varepsilon_{it} &\sim WN(0, \sigma_{\varepsilon}^2) \\
\varepsilon_{\alpha,it} &\sim WN(0, \sigma_{\varepsilon_\alpha}^2) \\
\varepsilon_{z,it} &\sim WN(0, \sigma_{\varepsilon_z}^2) \\
\mathtt{Cov}\left(z_{it}, \varepsilon_{is}\right) = \mathtt{Cov}\left(z_{it}, \varepsilon_{\alpha,is}\right) &= 0 \quad \forall\, t, s \\
\mathtt{Cov}\left(\varepsilon_{\alpha,it}, \varepsilon_{is}\right) = \mathtt{Cov}\left(\varepsilon_{\alpha,it}, \varepsilon_{z,is}\right) = \mathtt{Cov}\left(\varepsilon_{it}, \varepsilon_{z,is}\right) &= 0 \quad \forall\, t, s
\end{aligned}$$

Across all experiments, we set $\beta = 1$, $\lambda = 1$, $\rho = 0.5$, and $\sigma_{\varepsilon}^2 = \sigma_{\varepsilon_z}^2 = 1$, and conduct 200 simulations. We vary $N \in \{100, 500\}$, T, $T \in \{5, 10, 20, 30\}$, $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$, and $\sigma_{\varepsilon_\alpha}^2 \in \{0.25, 0.50, 0.75, 1\}$. Given the fixed ρ, varying δ controls the relative persistence of z and α. Varying δ and $\sigma_{\varepsilon_\alpha}^2$ affect the degree of time variation in α. Finally, α follows a random walk when $\delta = 1$. In this case, $\sigma_{\varepsilon_\alpha}^2$ alone affects the degree of time variation in α. We report the mean bias, mean absolute bias, and the root mean squared error (RMSE). We display the results graphically.

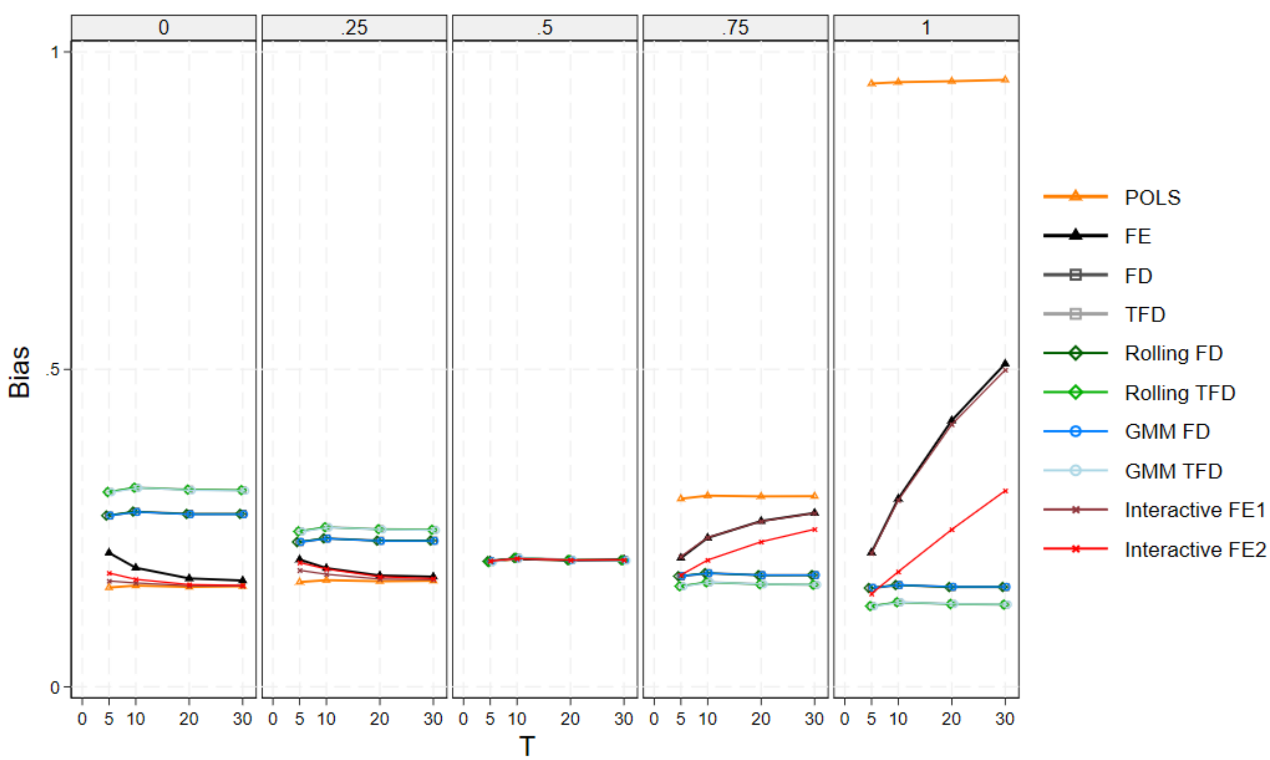
We also simulate two additional experimental designs that are identical to the previous designs except now

$$x_{it} = \lambda \alpha_{it-1} + z_{it}$$

with $\delta = \rho = 0.1$ and $\delta = \rho = 0.9$. We fix $\sigma_{\varepsilon_\alpha}^2 = 0.25$.



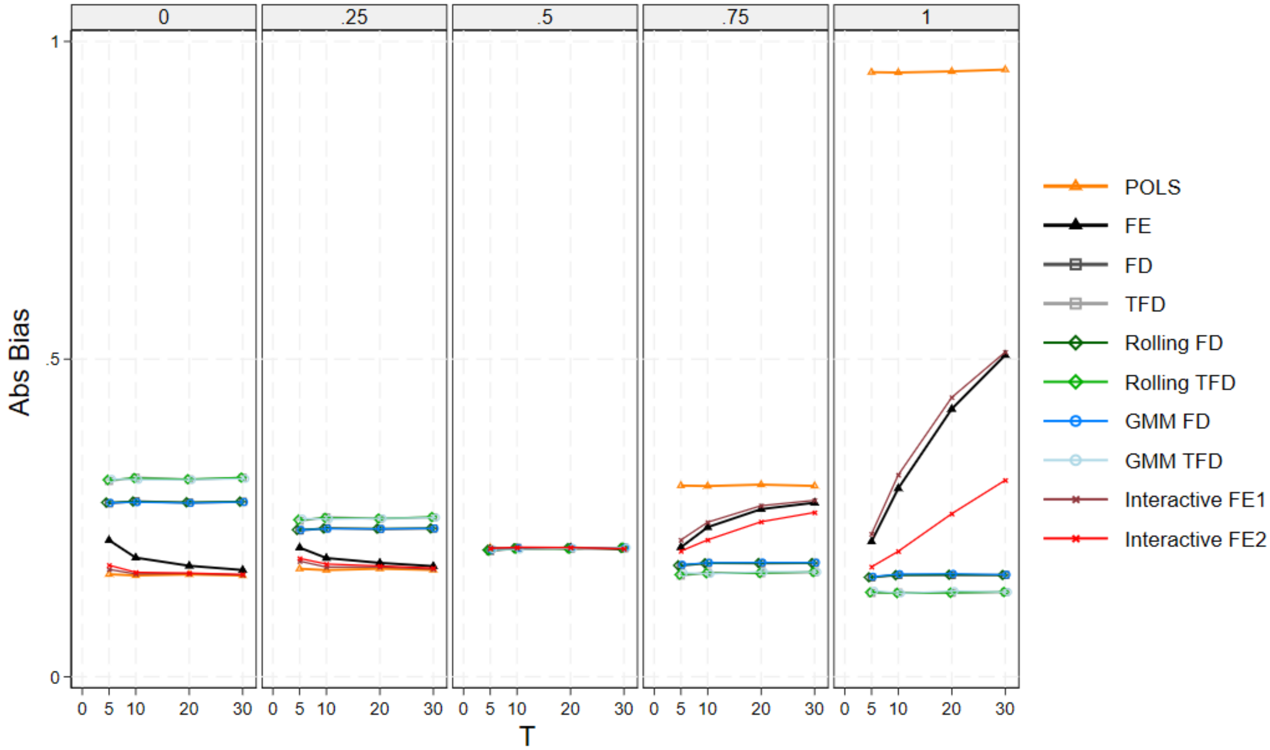
(A) $N = 100$



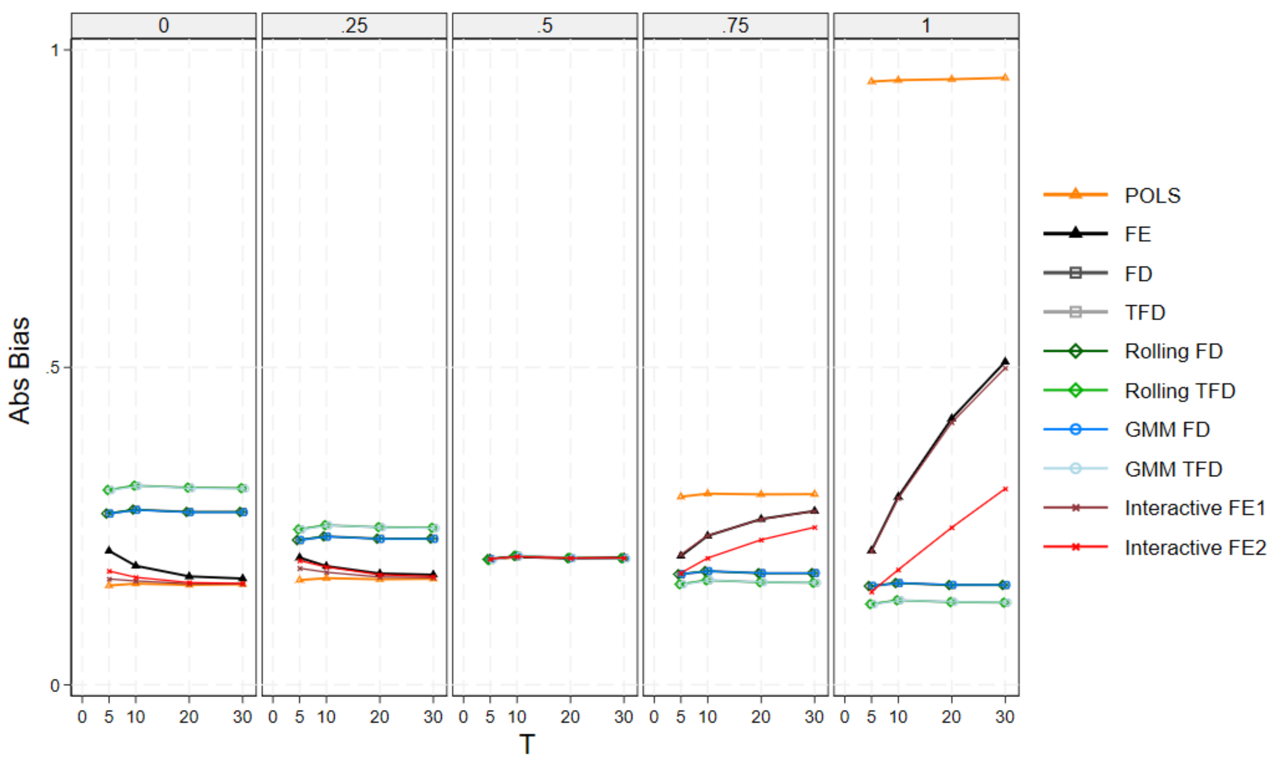
(B) $N = 500$

FIGURE B.1: Simulation Results: Bias ($\sigma^2_{\varepsilon_\alpha} = 0.25$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



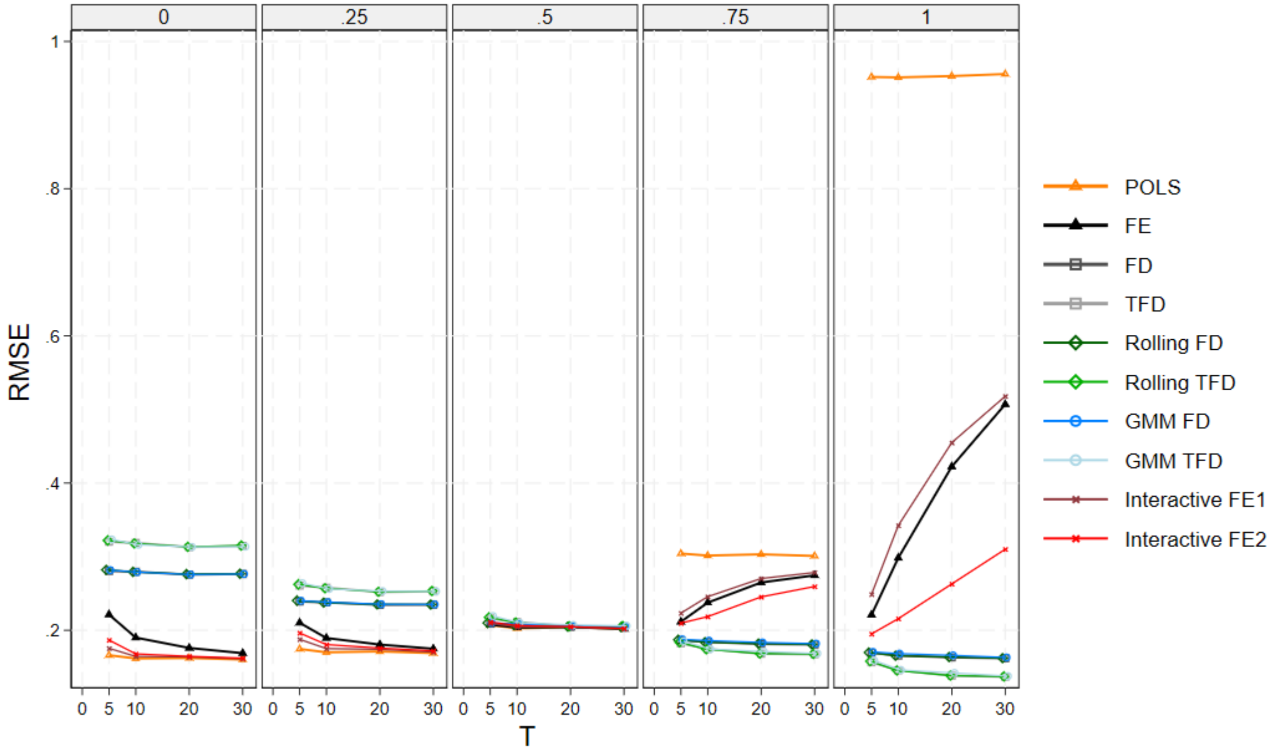
(A) $N = 100$



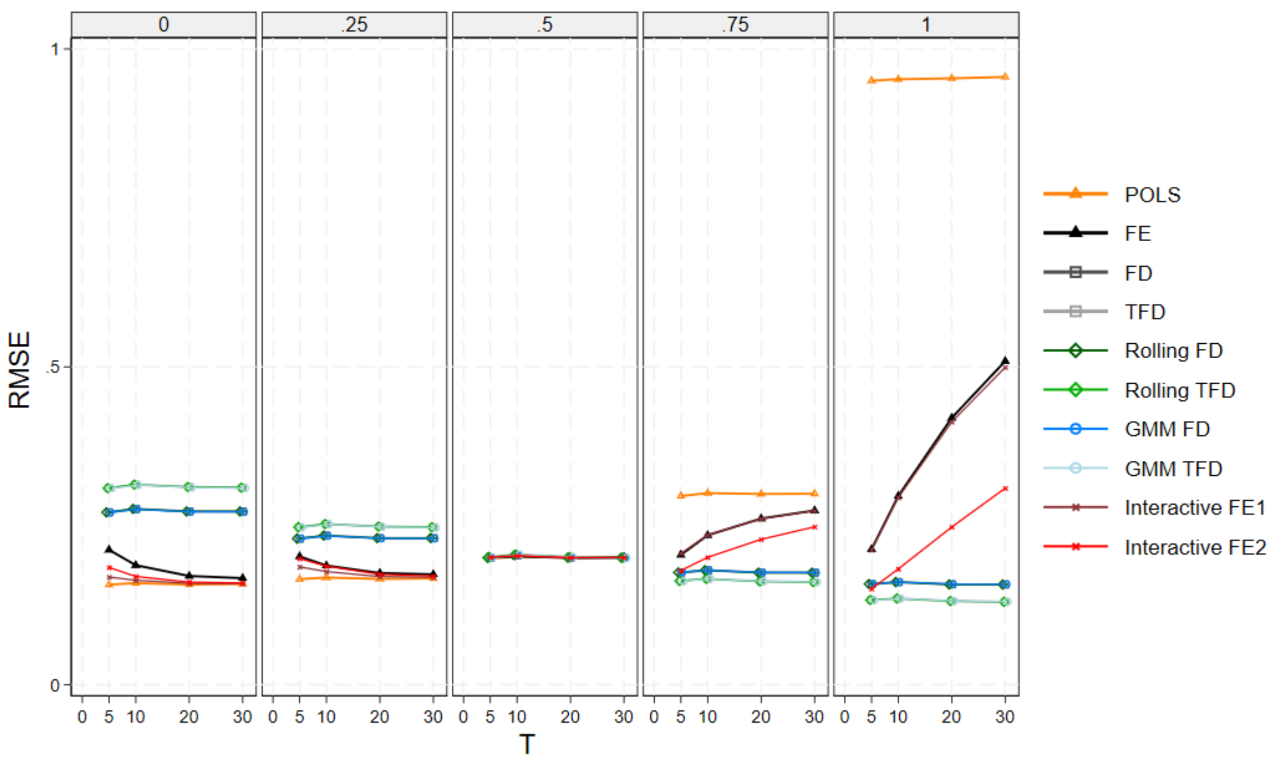
(B) $N = 500$

FIGURE B.2: Simulation Results: Absolute Bias ($\sigma^2_{\varepsilon_\alpha} = 0.25$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



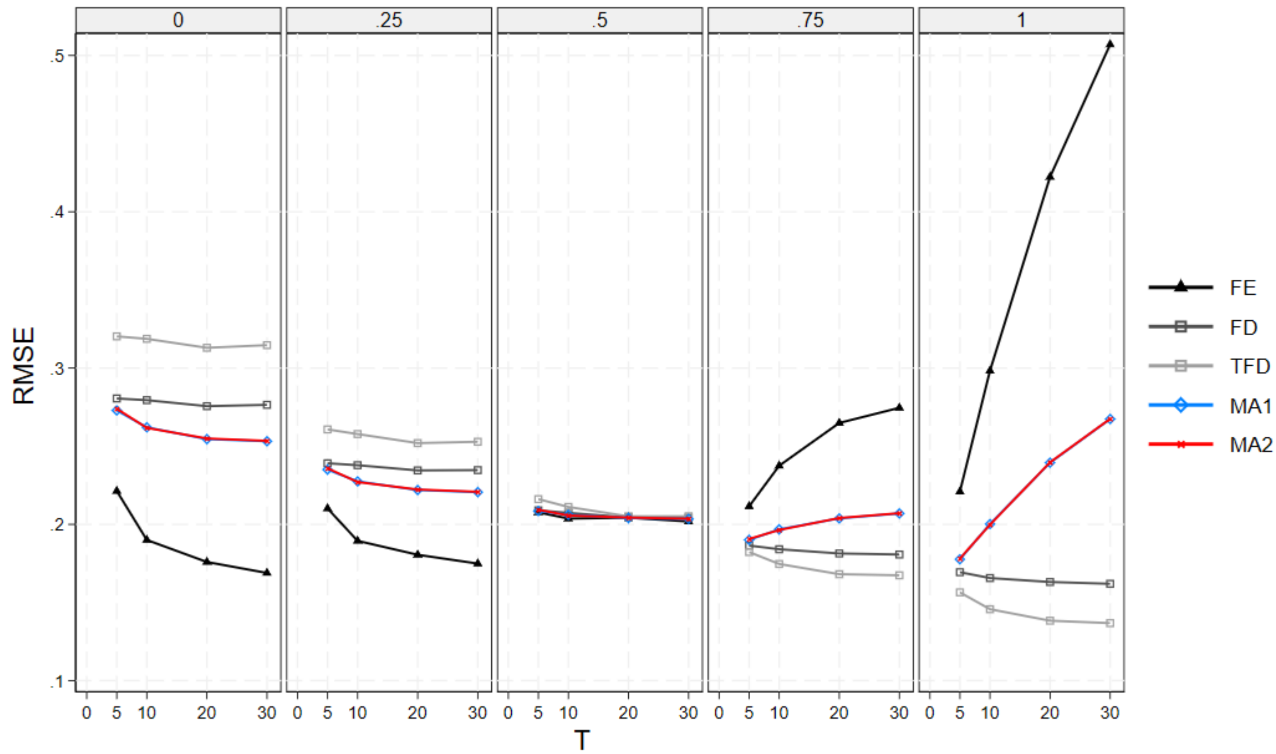
(A) $N = 100$



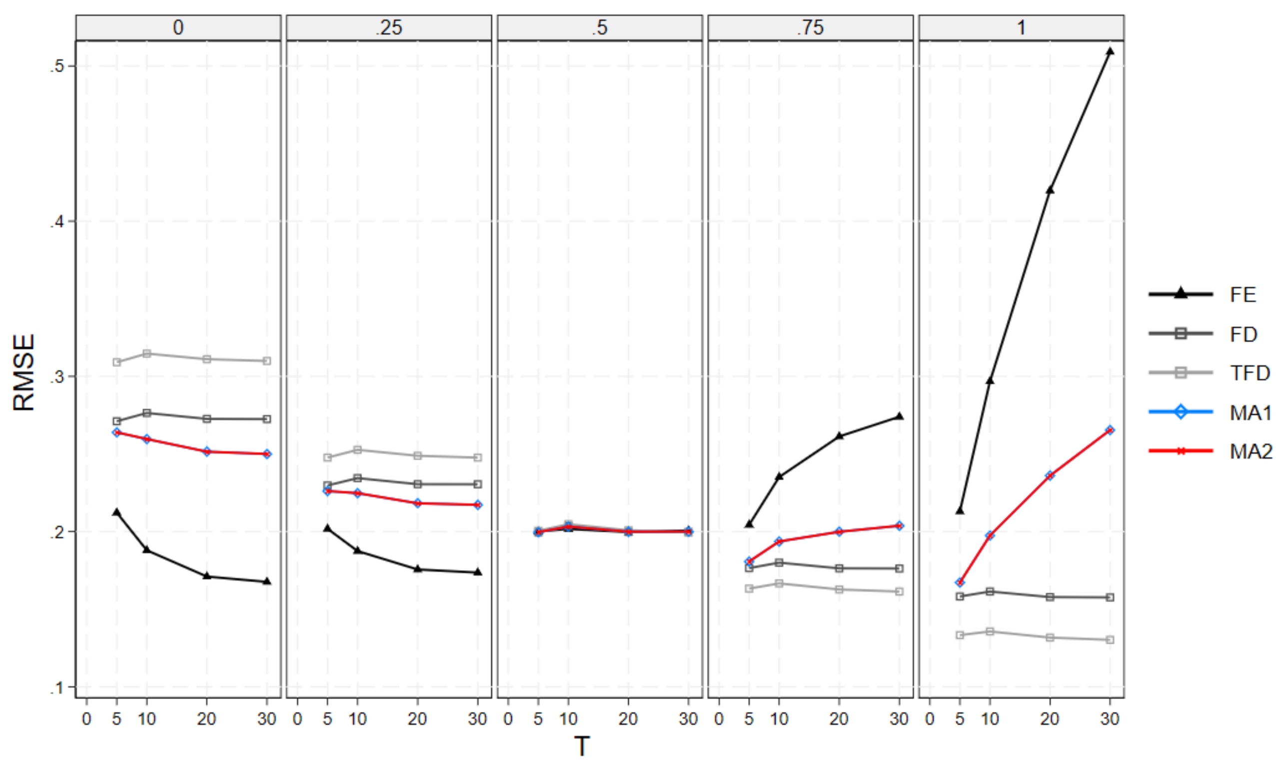
(B) $N = 500$

FIGURE B.3: Simulation Results: Root Mean Squared Error ($\sigma^2_{\varepsilon_\alpha} = 0.25$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



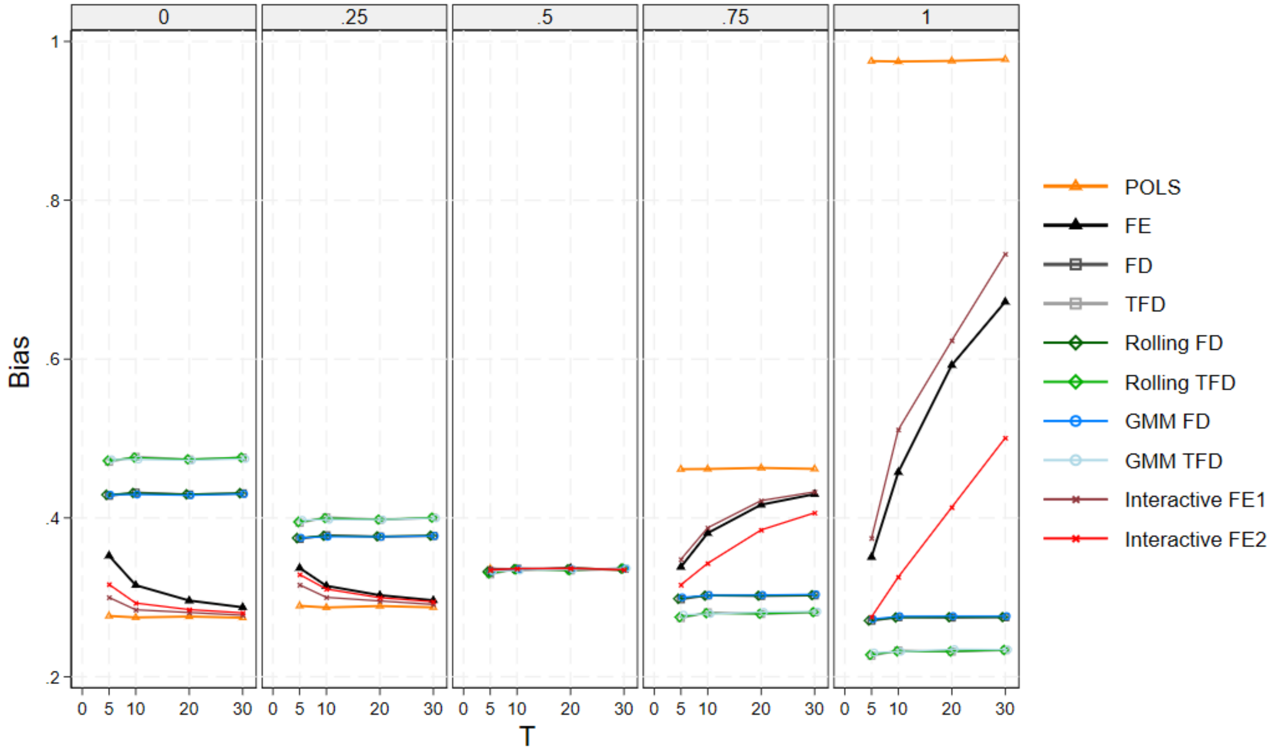
(A) $N = 100$



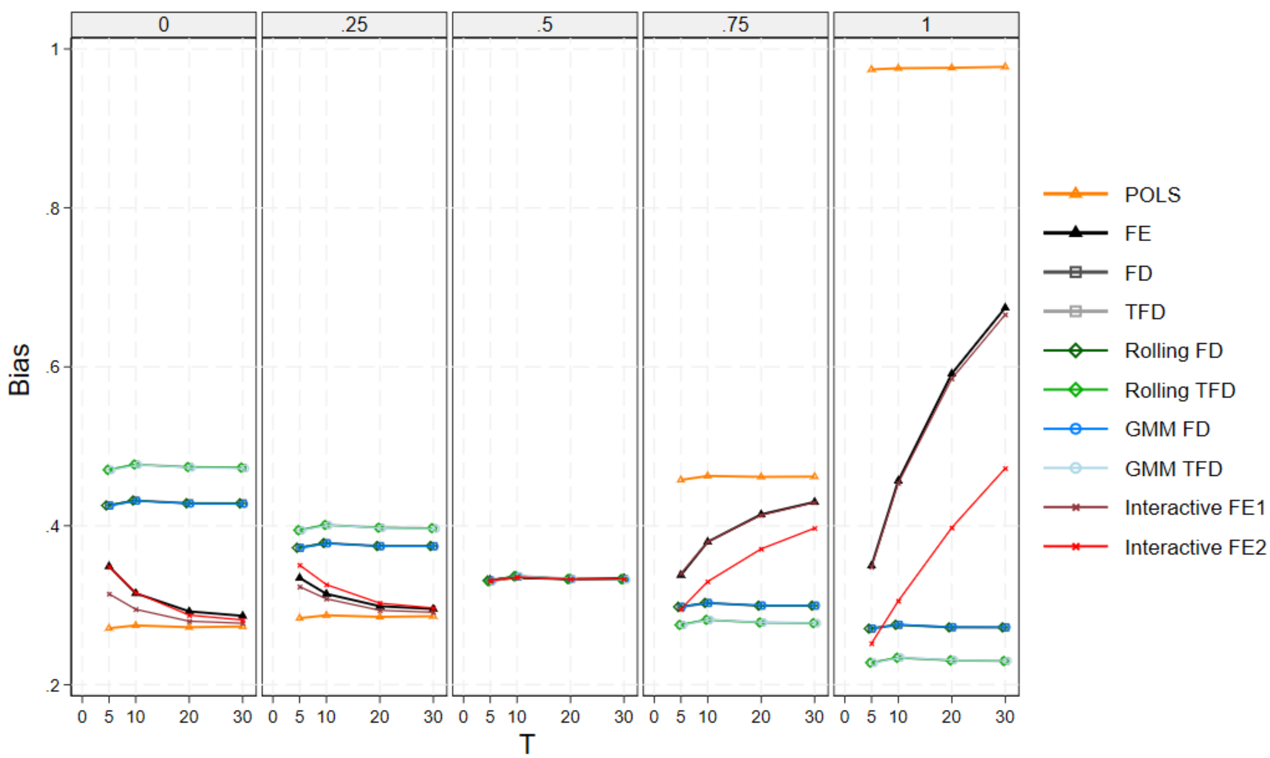
(B) $N = 500$

FIGURE B.4: Simulation Results: Model Averaging Estimators ($\sigma^2_{\varepsilon_\alpha} = 0.25$)

Notes: RMSE = Root Mean Squared Error. Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



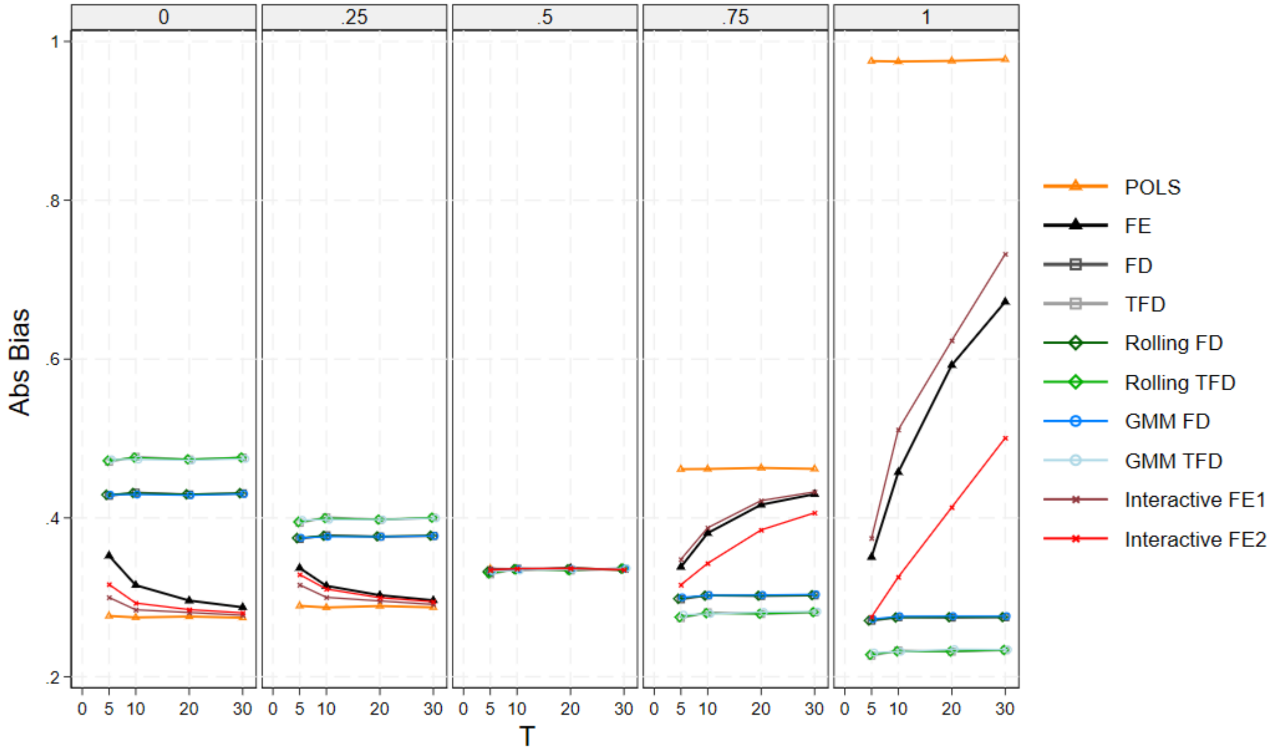
(A) $N = 100$



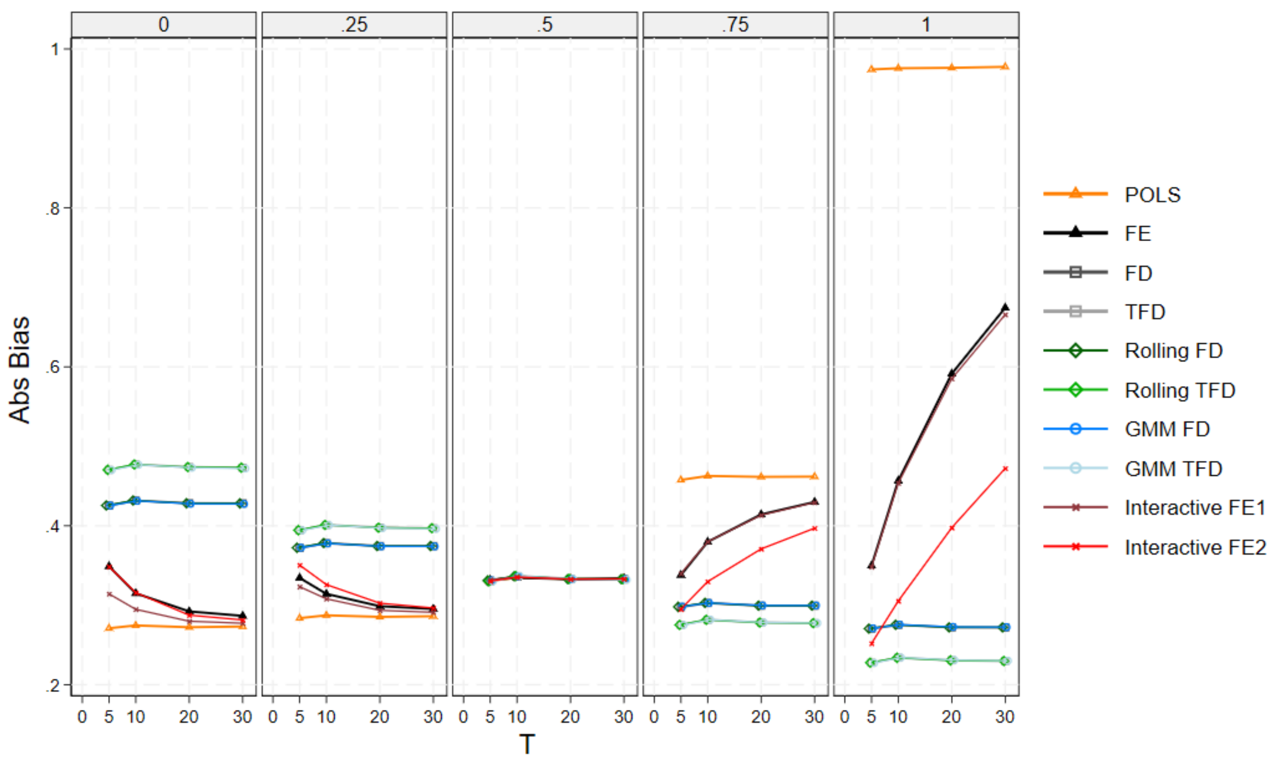
(B) $N = 500$

FIGURE B.5: Simulation Results: Bias ($\sigma^2_{\varepsilon_\alpha} = 0.50$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



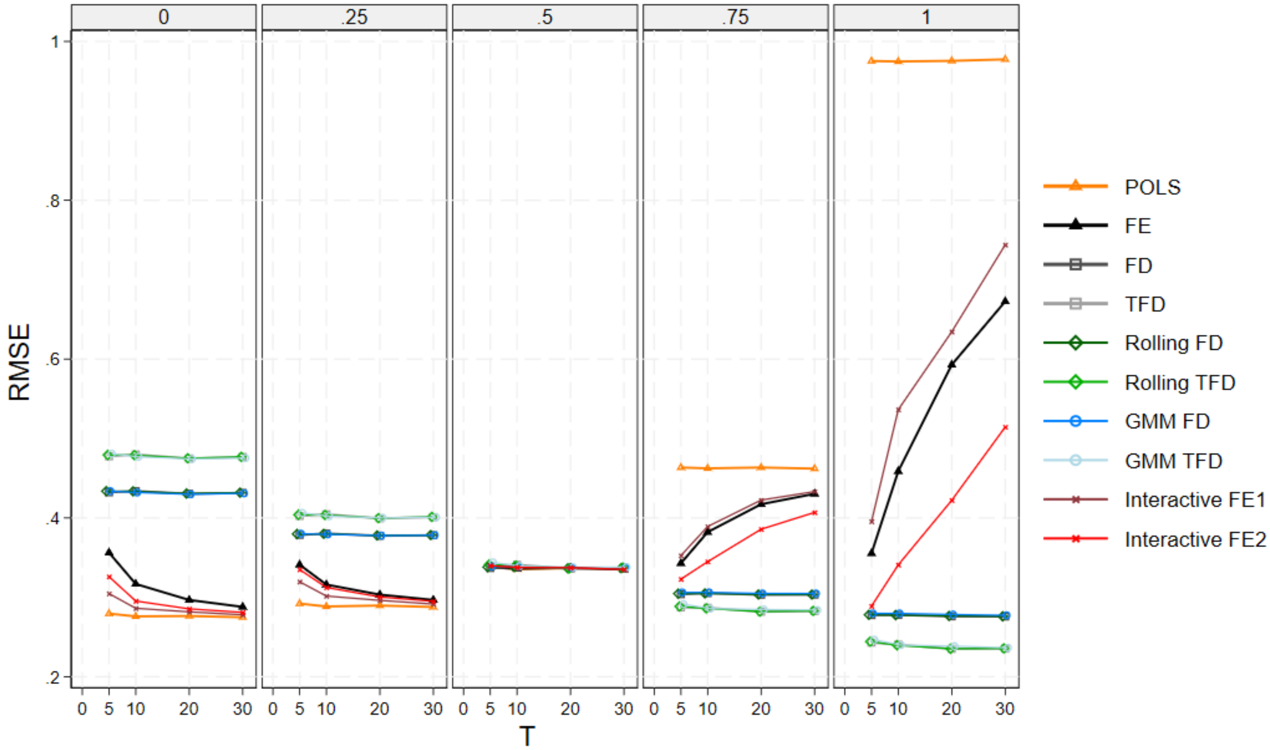
(A) $N = 100$



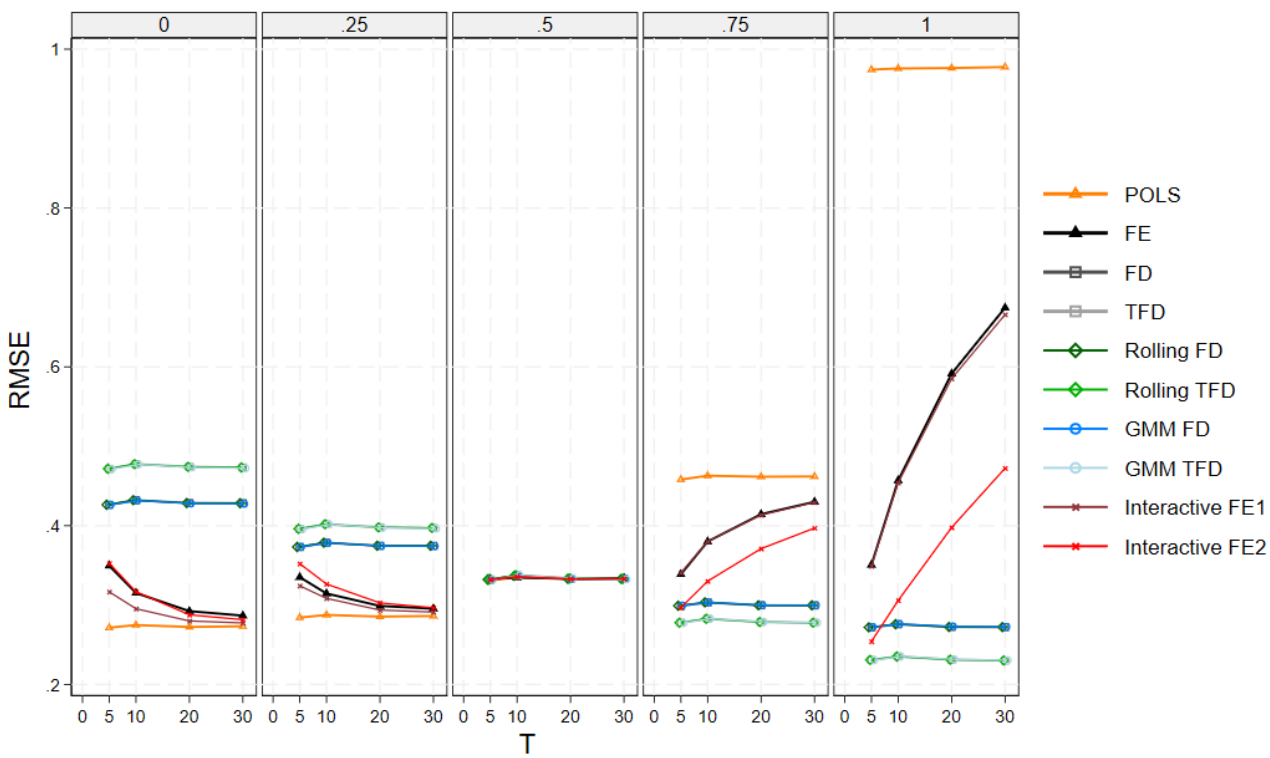
(B) $N = 500$

FIGURE B.6: Simulation Results: Absolute Bias ($\sigma^2_{\varepsilon_\alpha} = 0.50$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



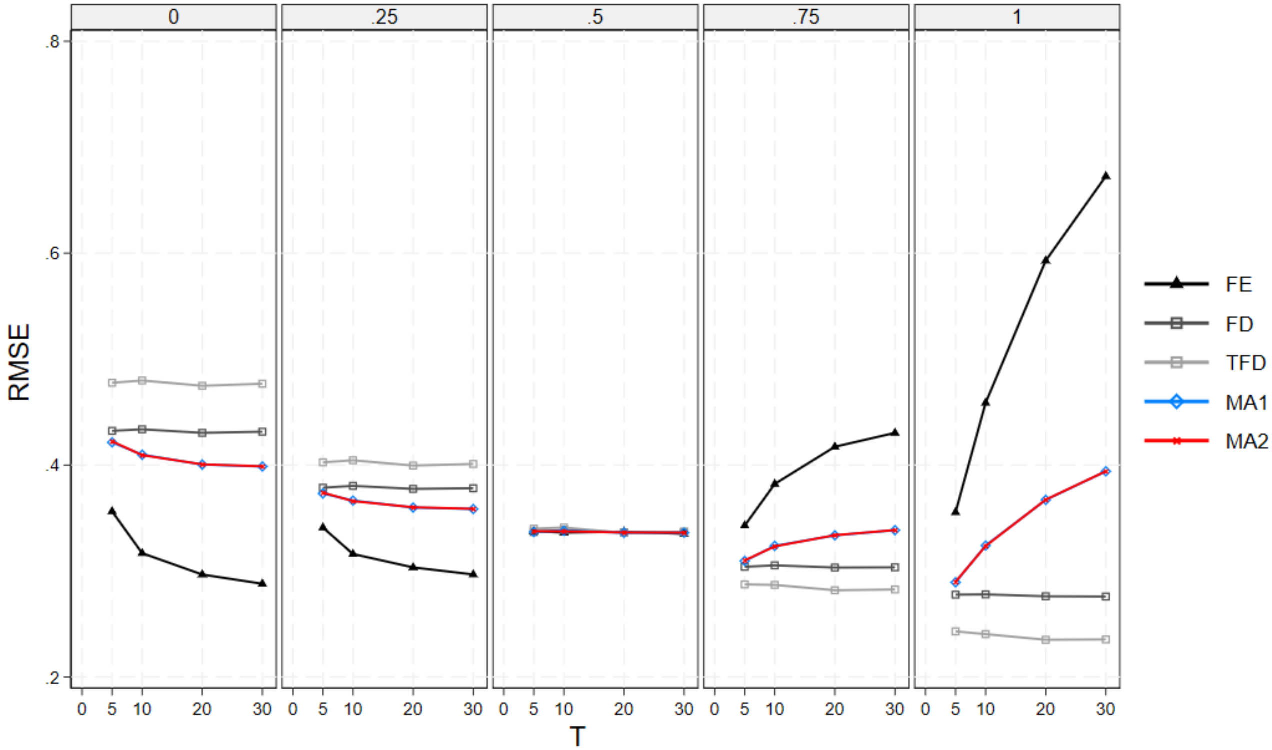
(A) $N = 100$



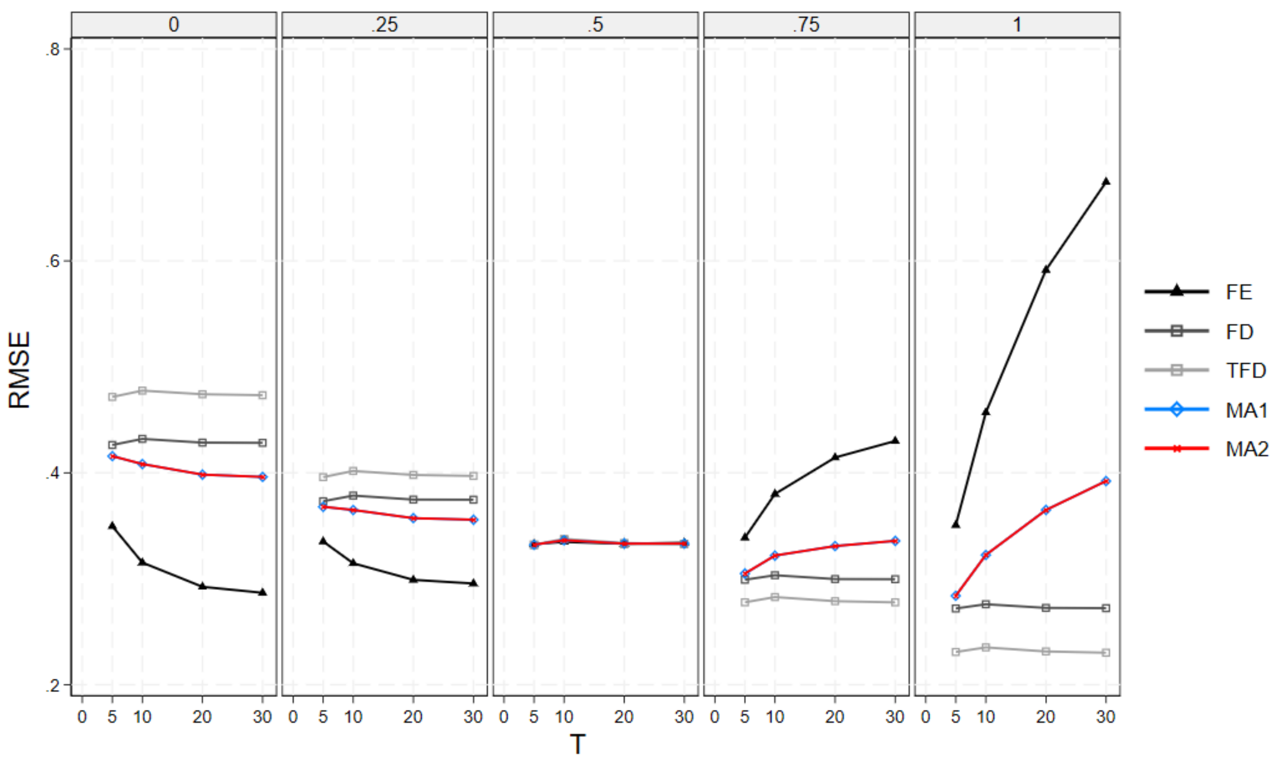
(B) $N = 500$

FIGURE B.7: Simulation Results: Root Mean Squared Error ($\sigma^2_{\varepsilon_\alpha} = 0.50$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



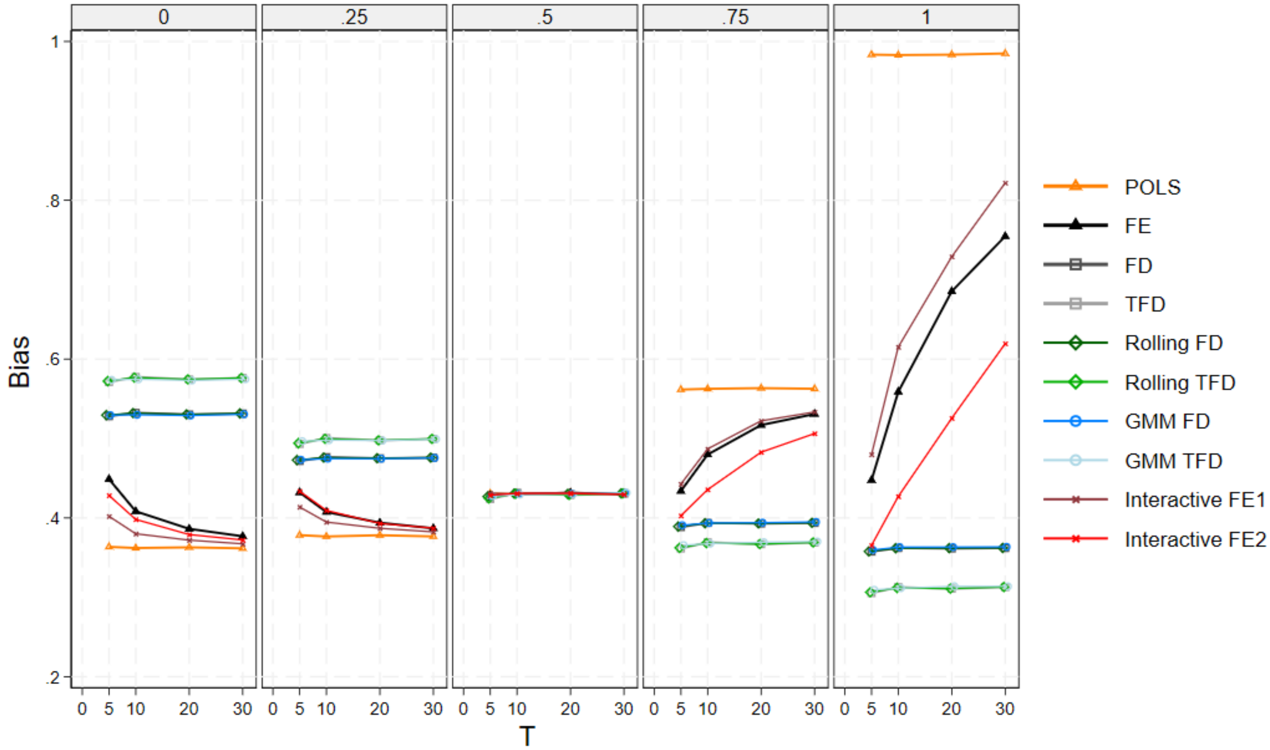
(A) $N = 100$



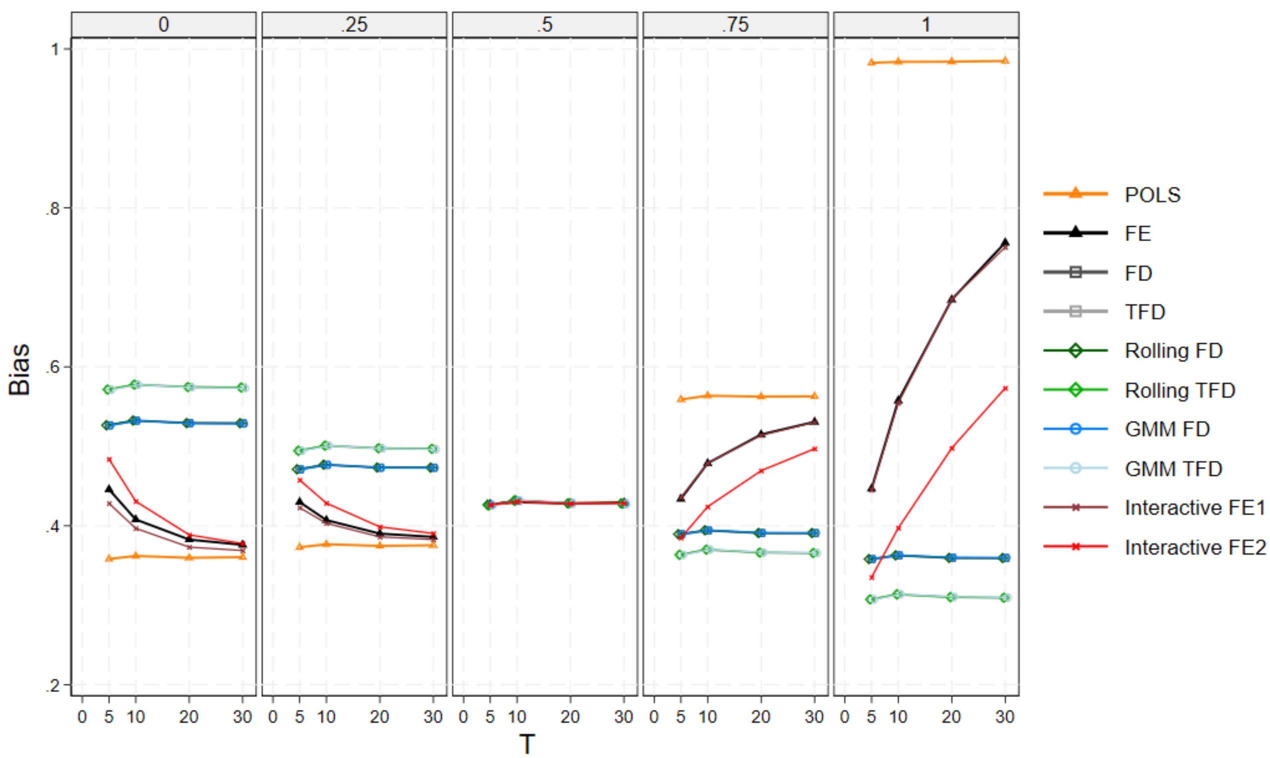
(B) $N = 500$

FIGURE B.8: Simulation Results: Model Averaging Estimators ($\sigma^2_{\varepsilon_\alpha} = 0.50$)

Notes: RMSE = Root Mean Squared Error. Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



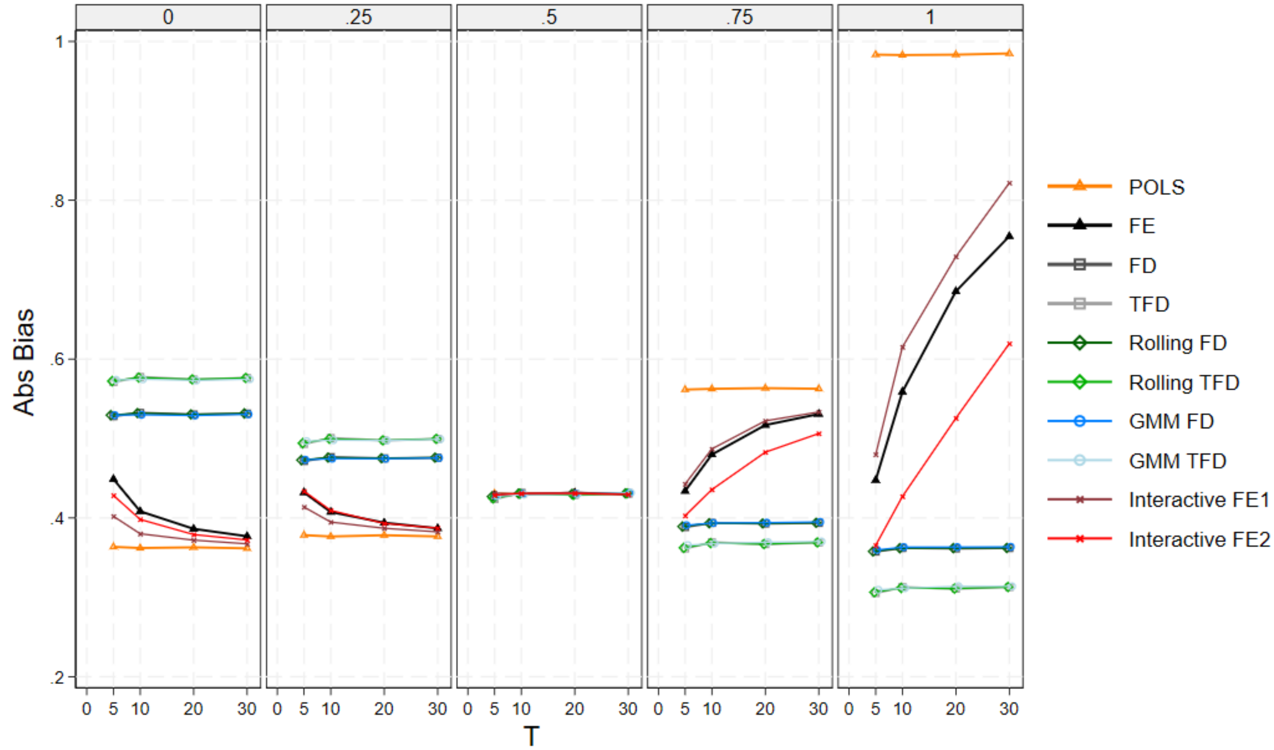
(A) $N = 100$



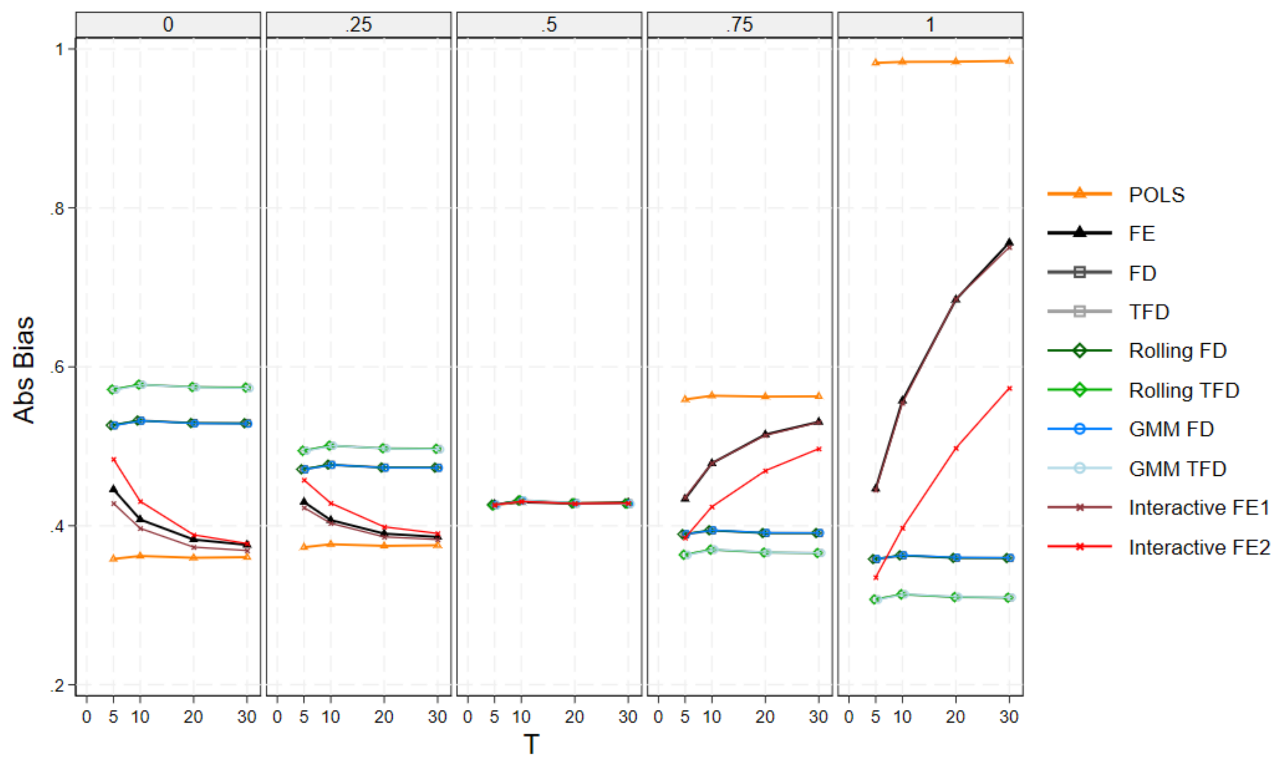
(B) $N = 500$

FIGURE B.9: Simulation Results: Bias ($\sigma^2_{\varepsilon_\alpha} = 0.75$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



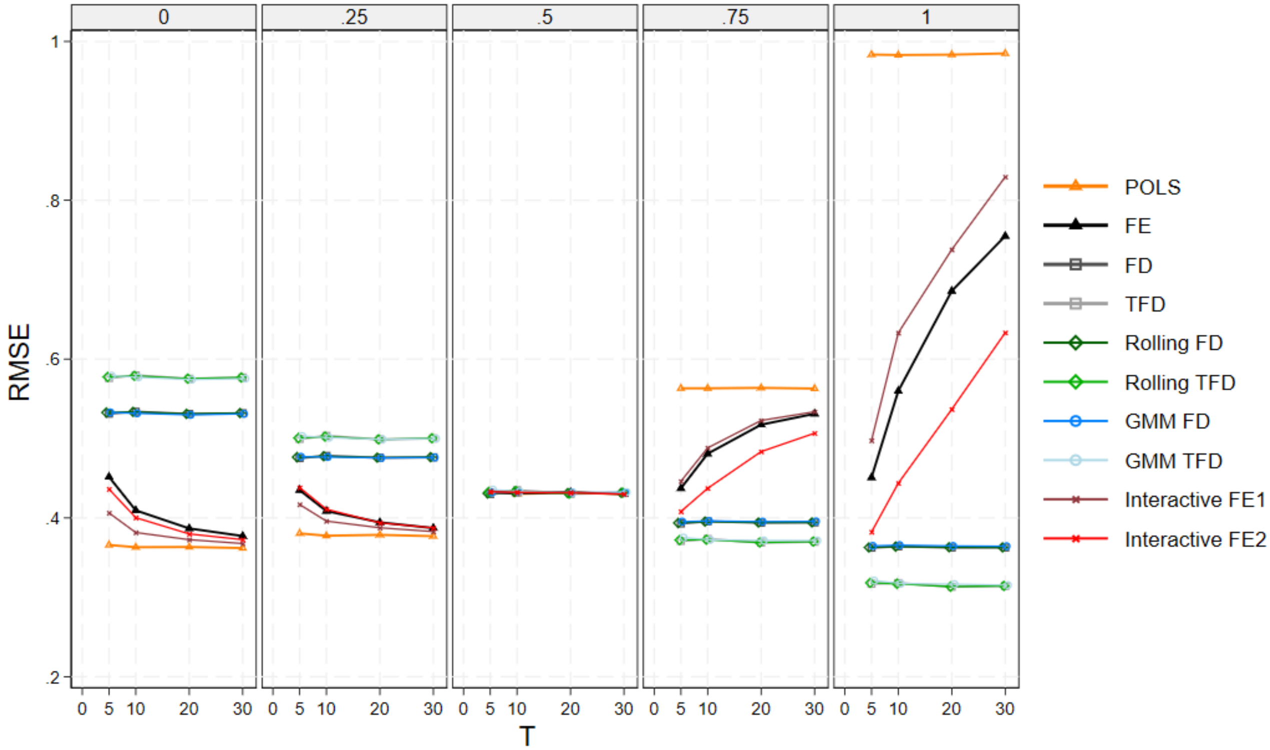
(A) $N = 100$



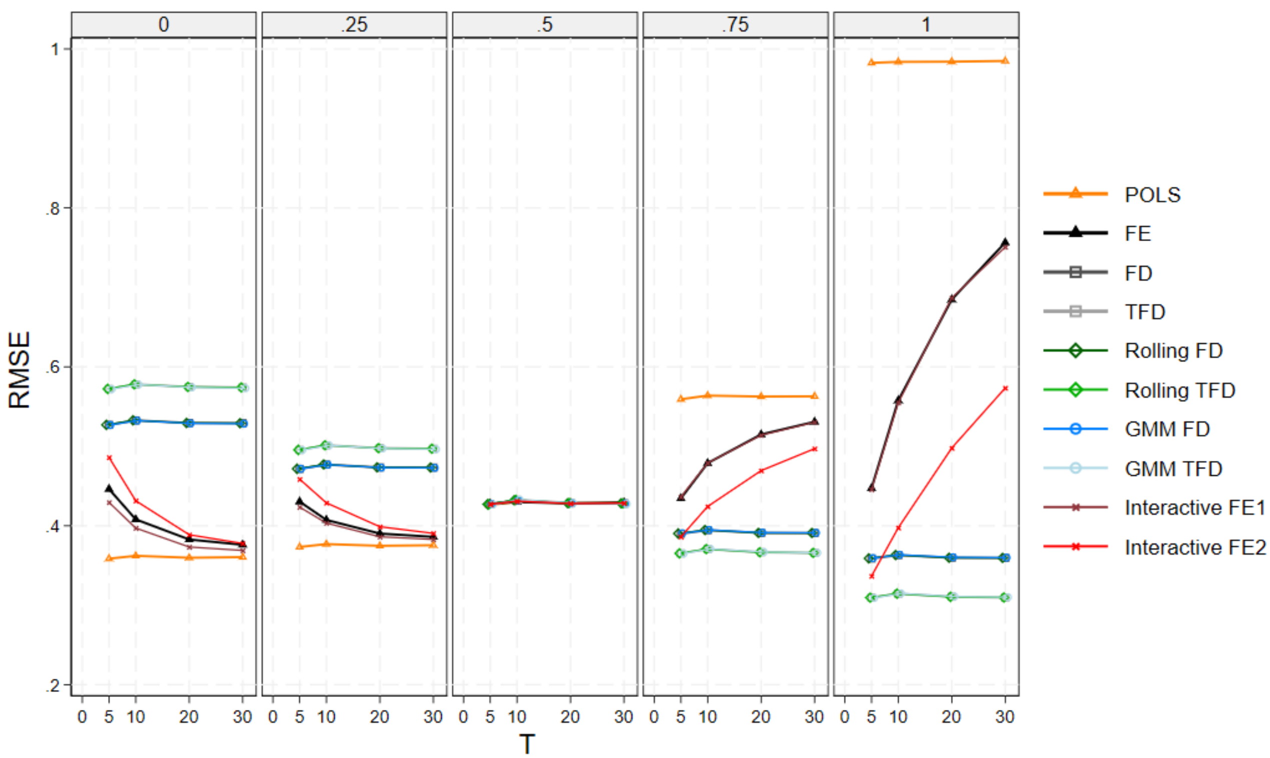
(B) $N = 500$

FIGURE B.10: Simulation Results: Absolute Bias ($\sigma^2_{\varepsilon_\alpha} = 0.75$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



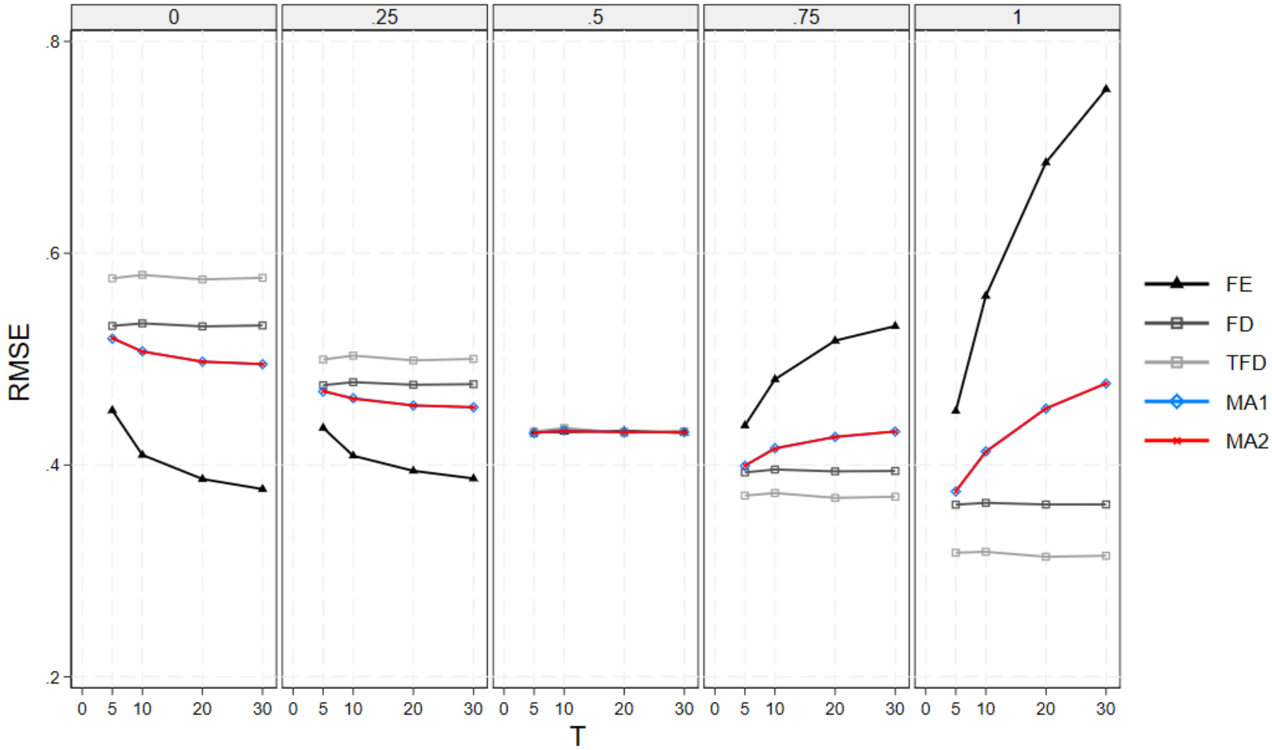
(A) $N = 100$



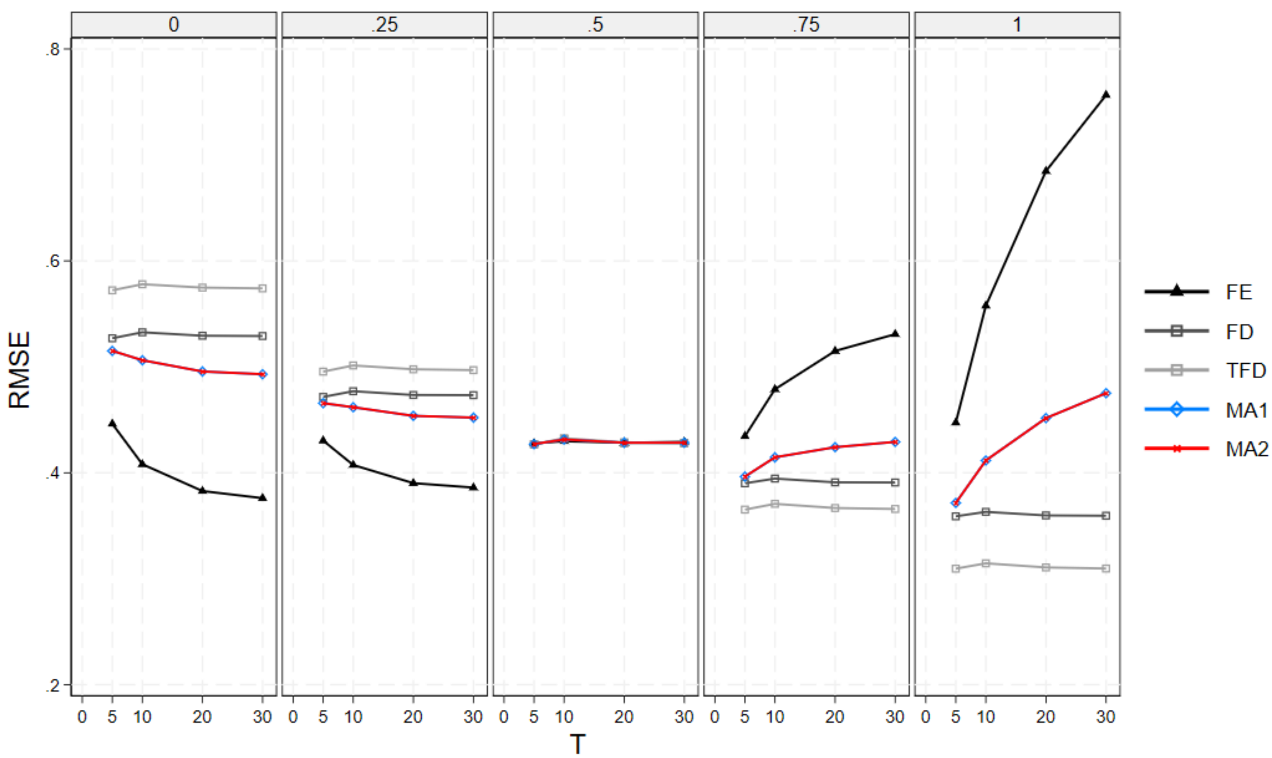
(B) $N = 500$

FIGURE B.11: Simulation Results: Root Mean Squared Error ($\sigma^2_{\varepsilon_\alpha} = 0.75$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



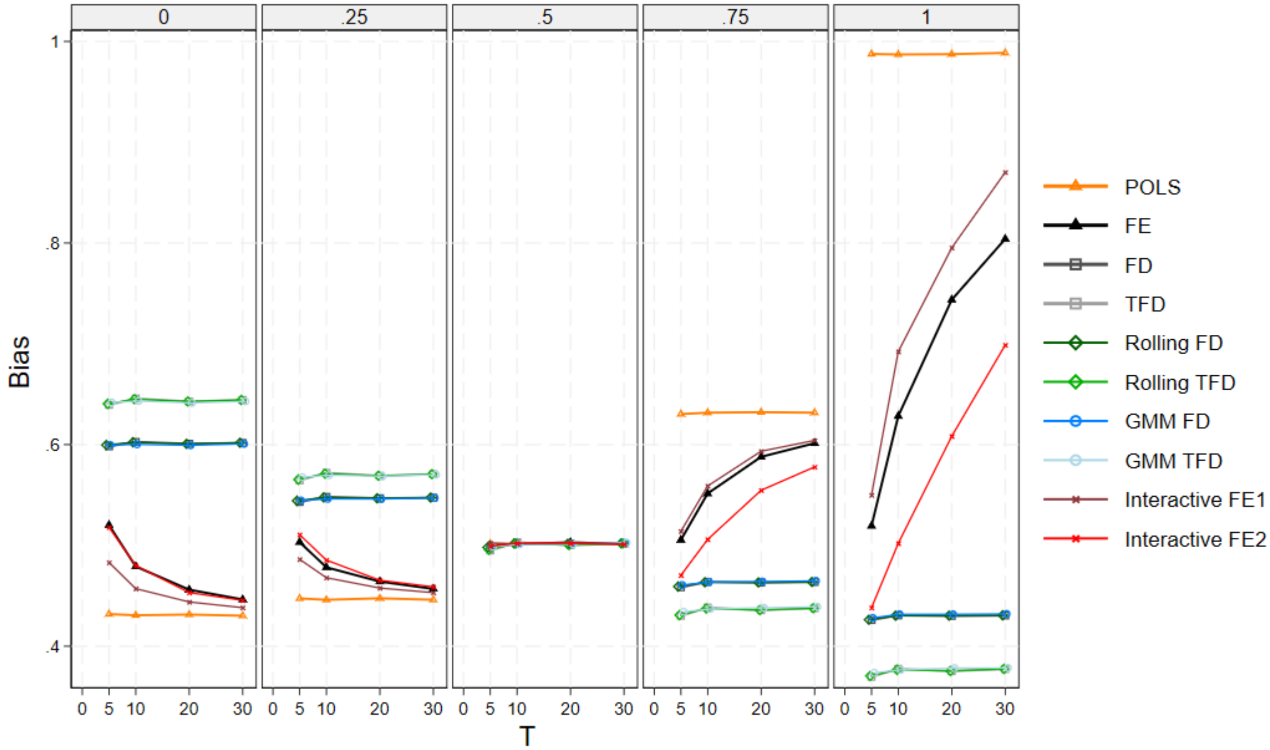
(A) $N = 100$



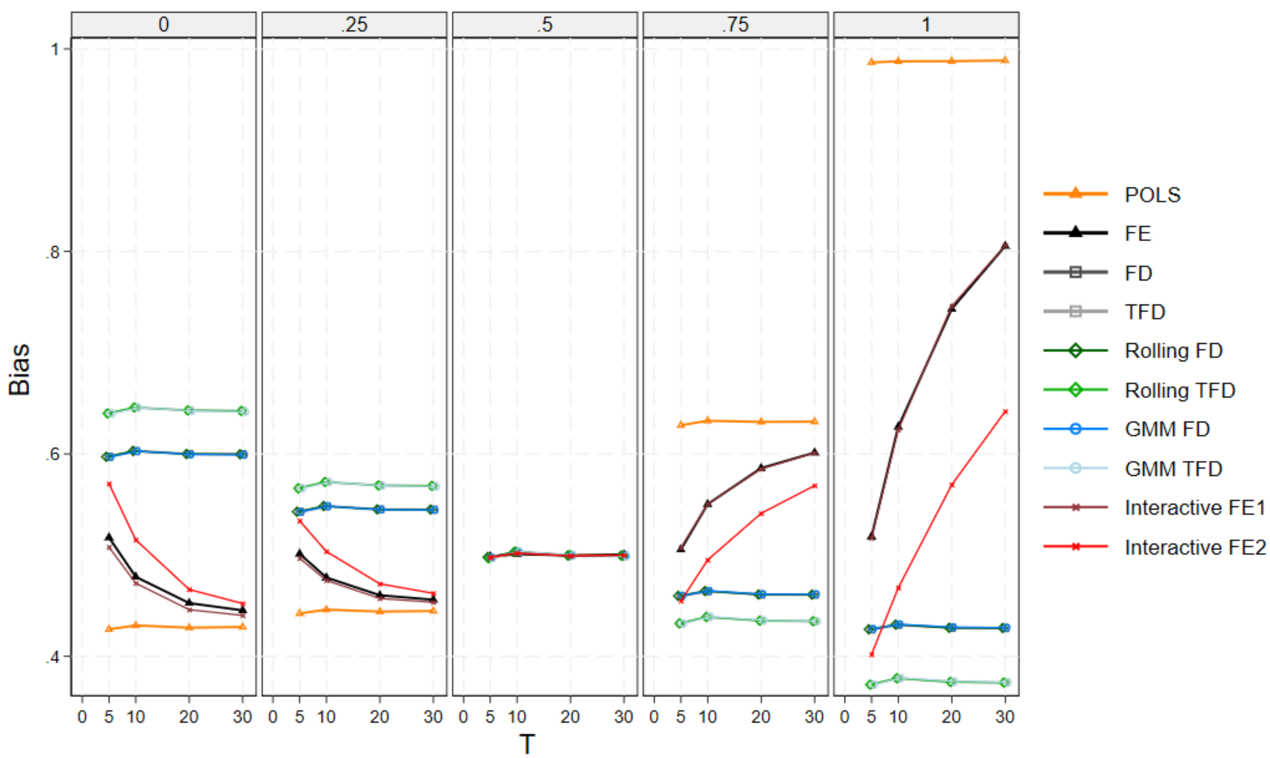
(B) $N = 500$

FIGURE B.12: Simulation Results: Model Averaging Estimators ($\sigma^2_{\varepsilon_\alpha} = 0.75$)

Notes: RMSE = Root Mean Squared Error. Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



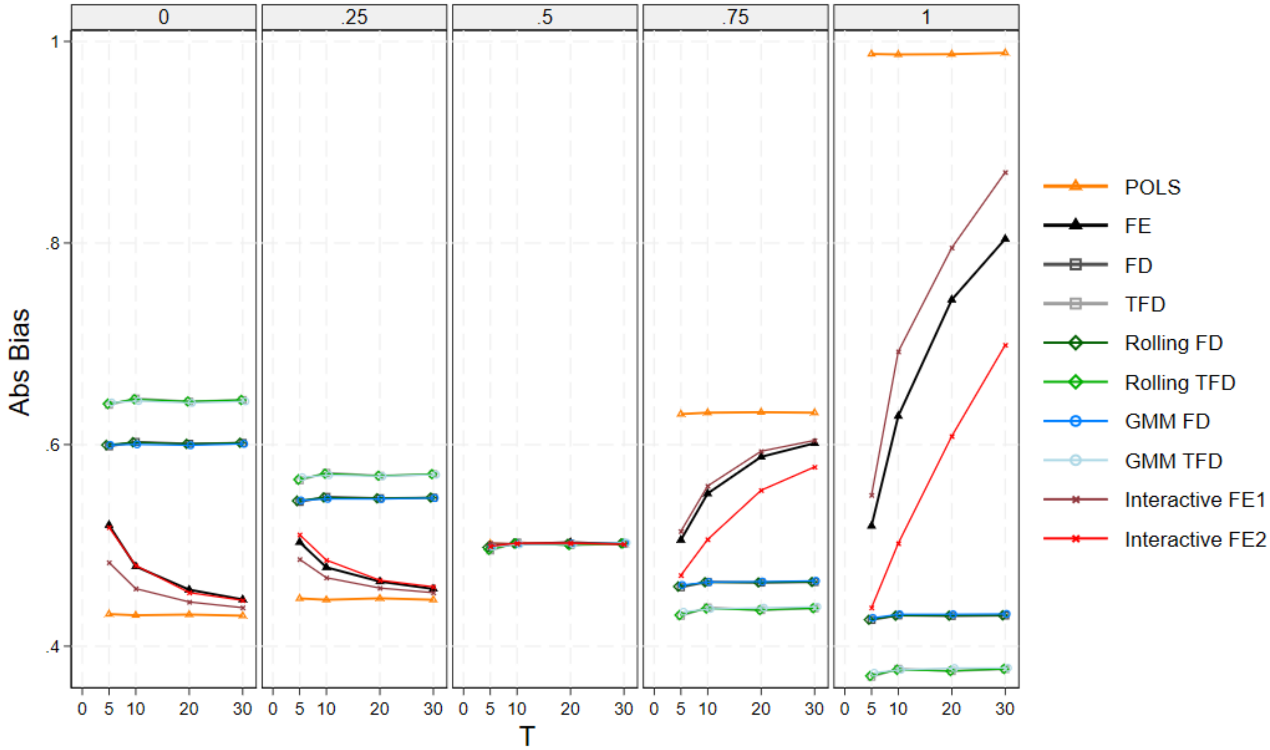
(A) $N = 100$



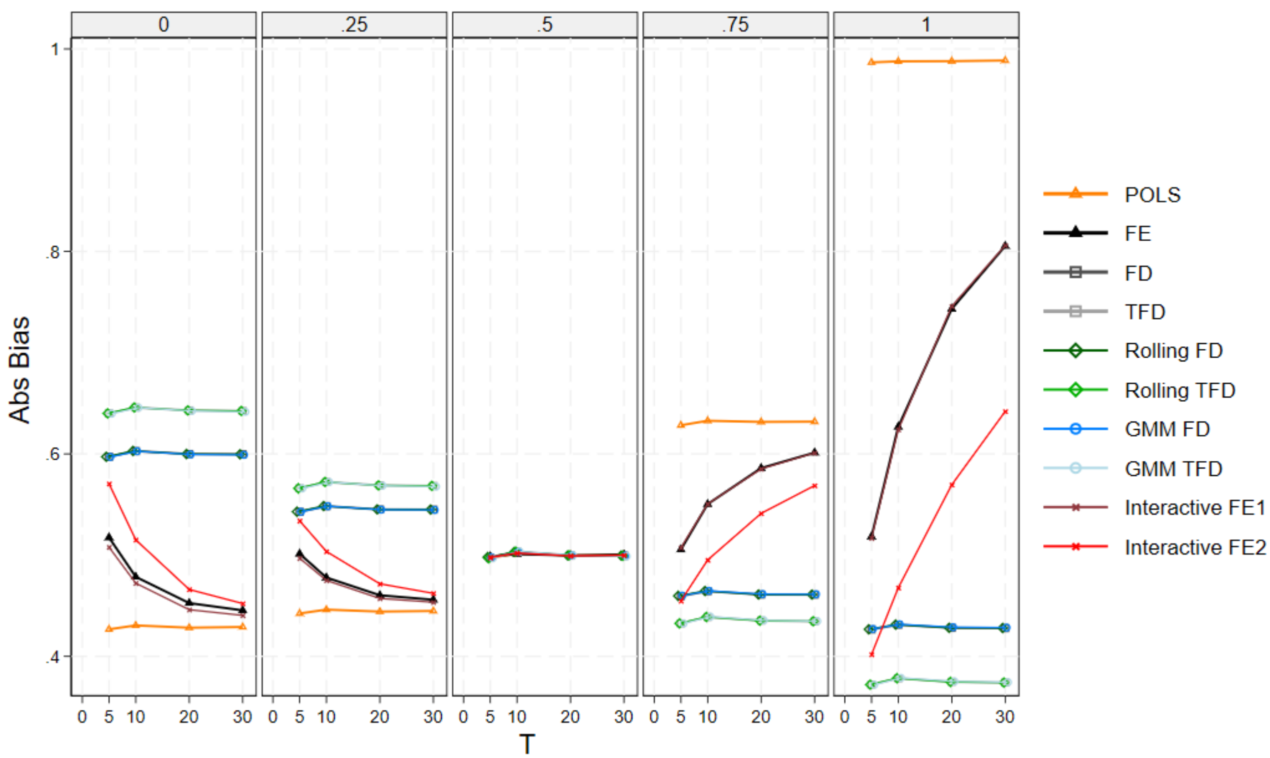
(B) $N = 500$

FIGURE B.13: Simulation Results: Bias ($\sigma^2_{\varepsilon_\alpha} = 1$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



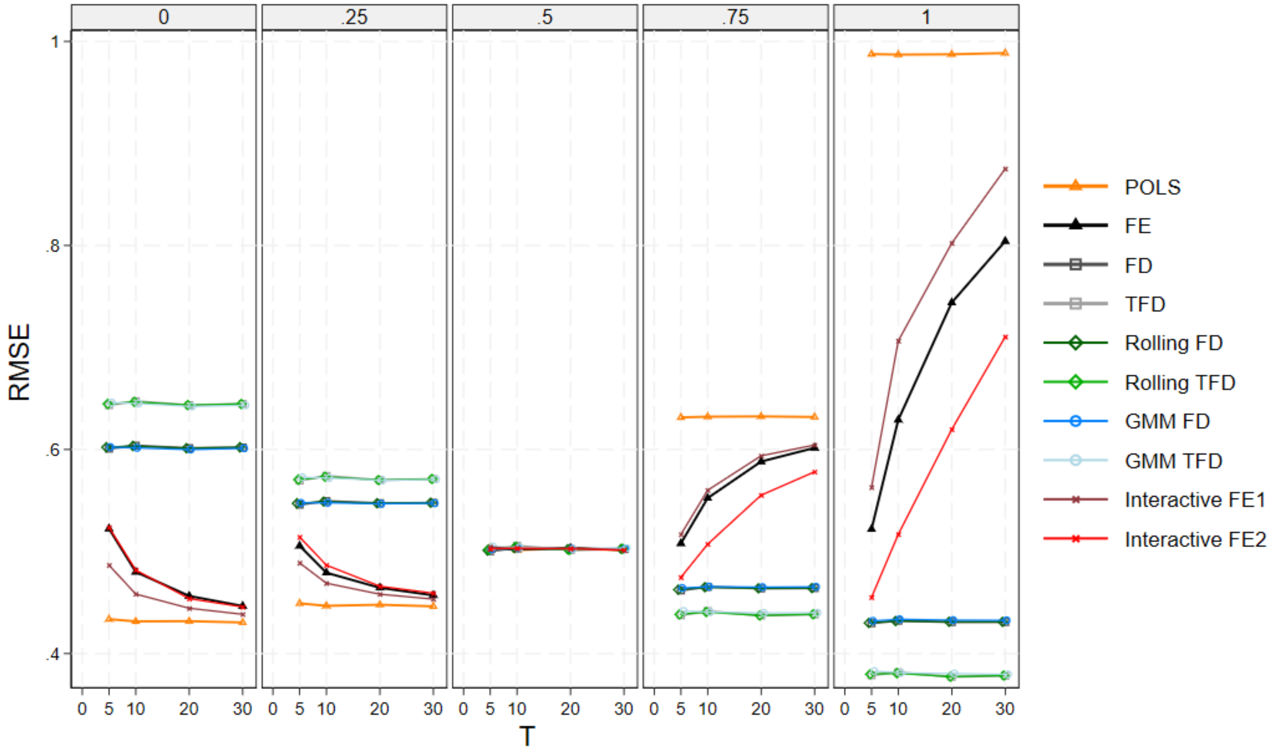
(A) $N = 100$



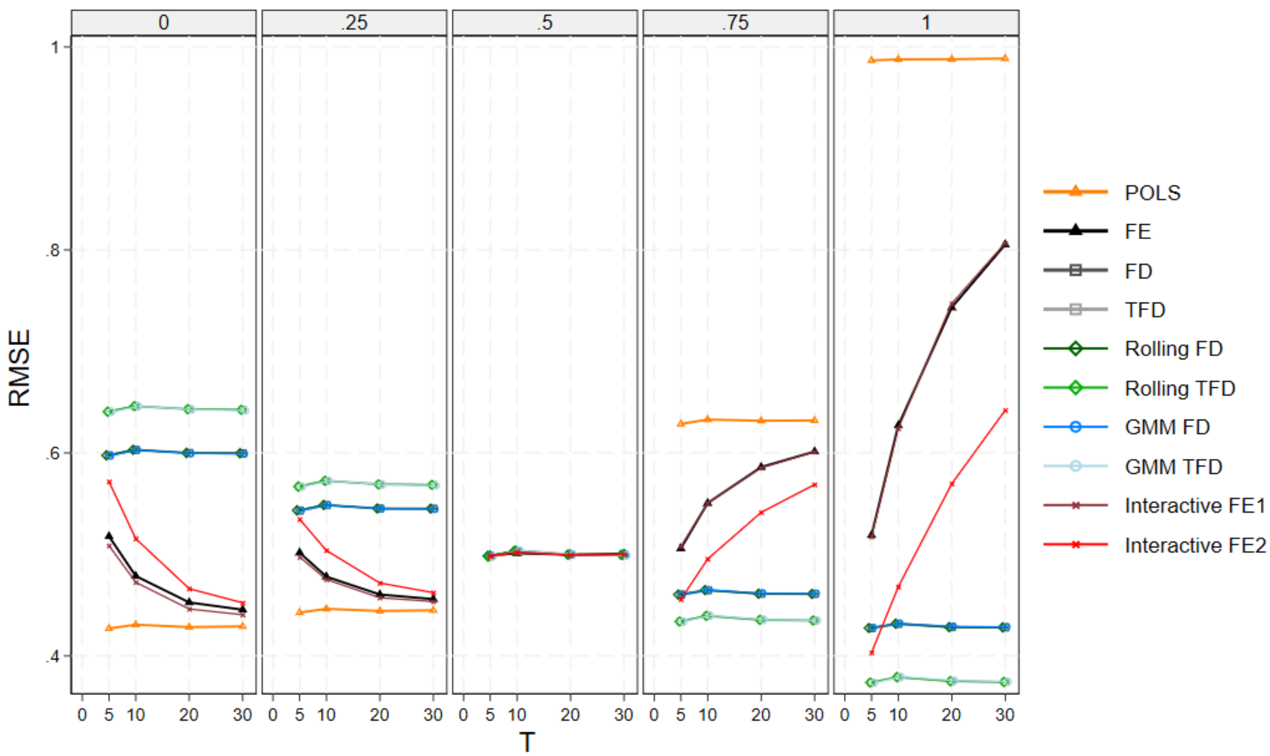
(B) $N = 500$

FIGURE B.14: Simulation Results: Absolute Bias ($\sigma^2_{\varepsilon_\alpha} = 1$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



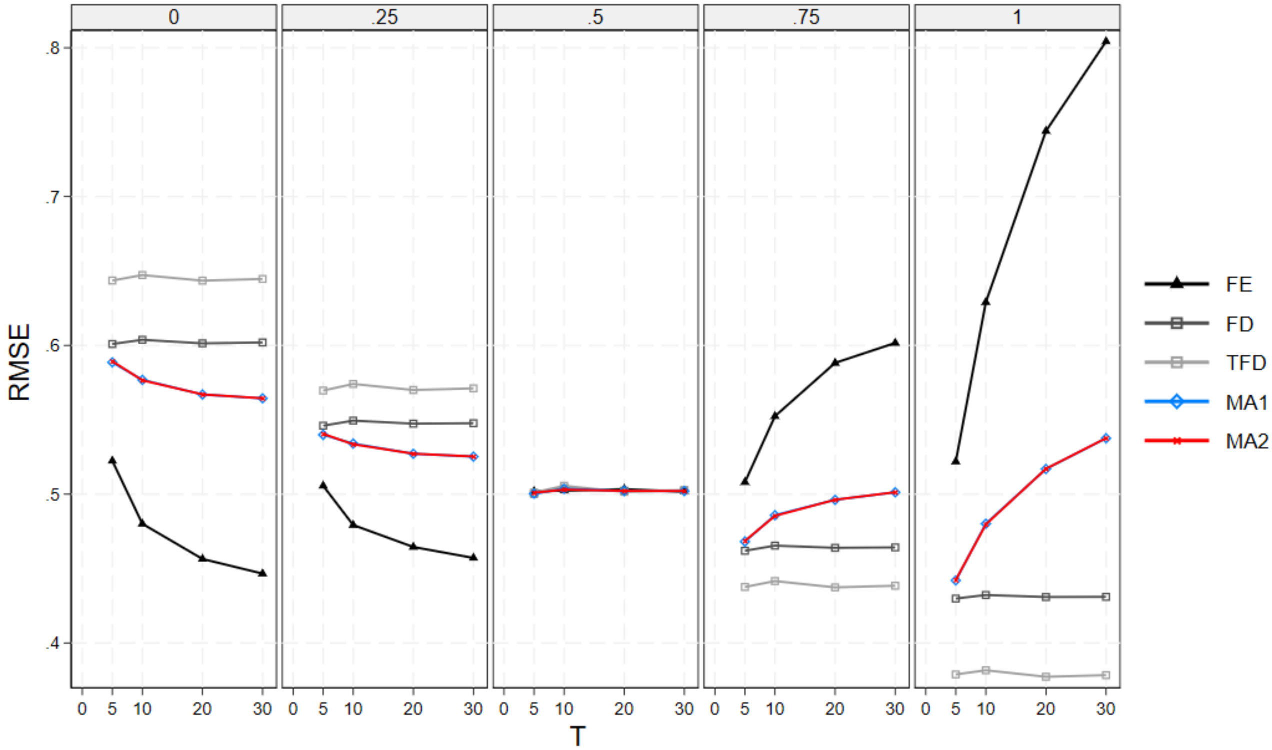
(A) $N = 100$



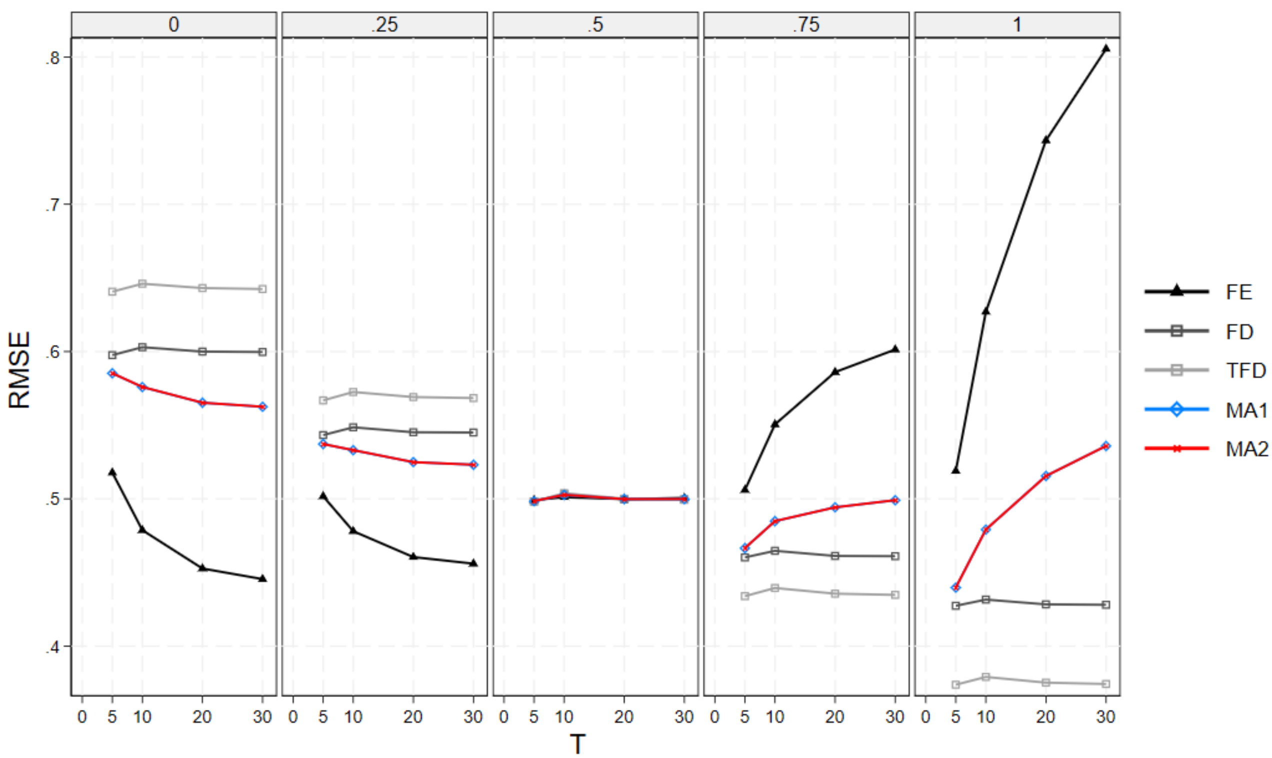
(B) $N = 500$

FIGURE B.15: Simulation Results: Root Mean Squared Error ($\sigma^2_{\varepsilon_\alpha} = 1$)

Notes: Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



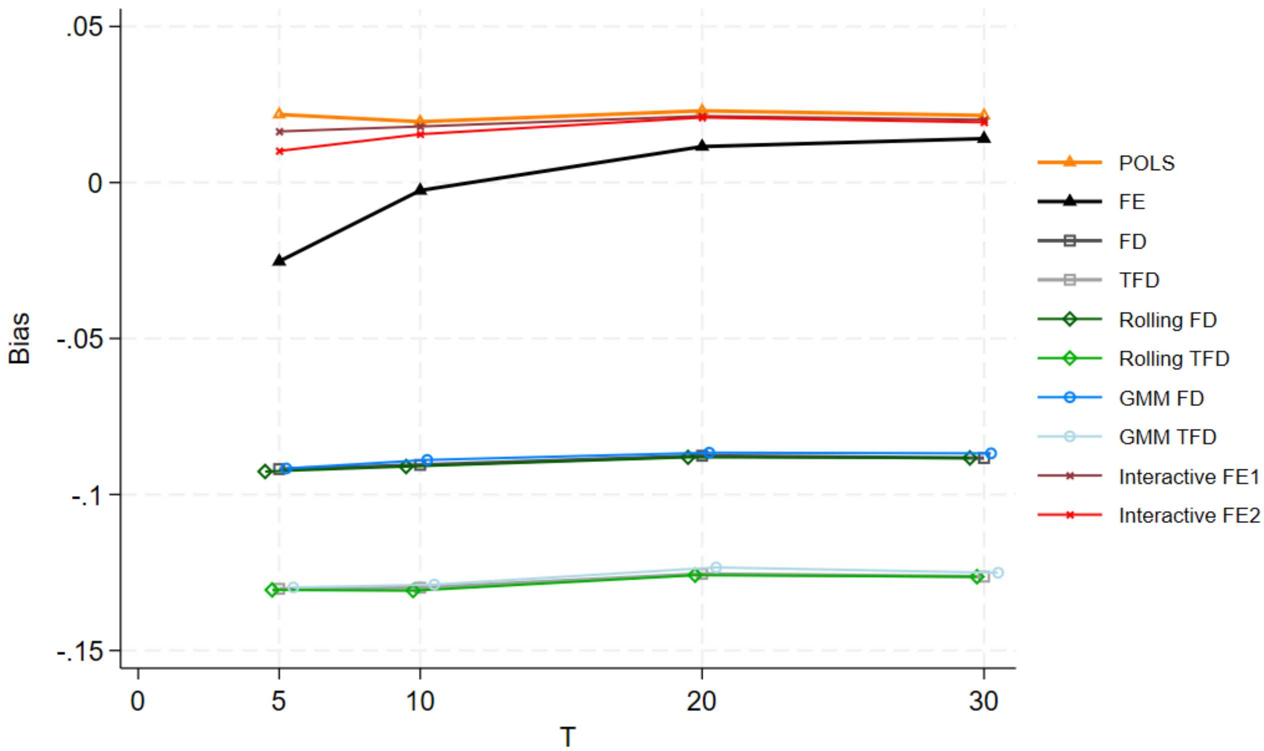
(A) $N = 100$



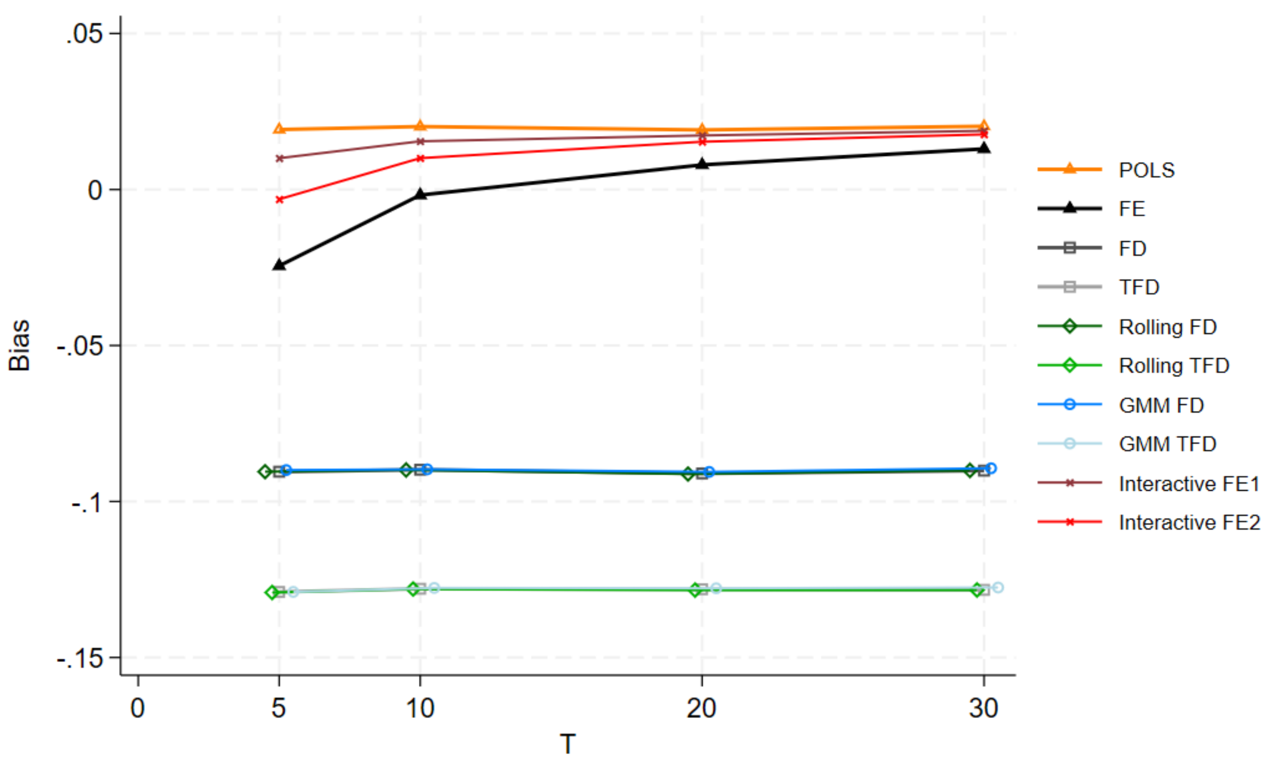
(B) $N = 500$

FIGURE B.16: Simulation Results: Model Averaging Estimators ($\sigma^2_{\varepsilon_\alpha} = 1$)

Notes: RMSE = Root Mean Squared Error. Column headings denote the value of $\delta \in \{0, 0.25, 0.50, 0.75, 1\}$. T = number of time periods per panel. See text for further details.



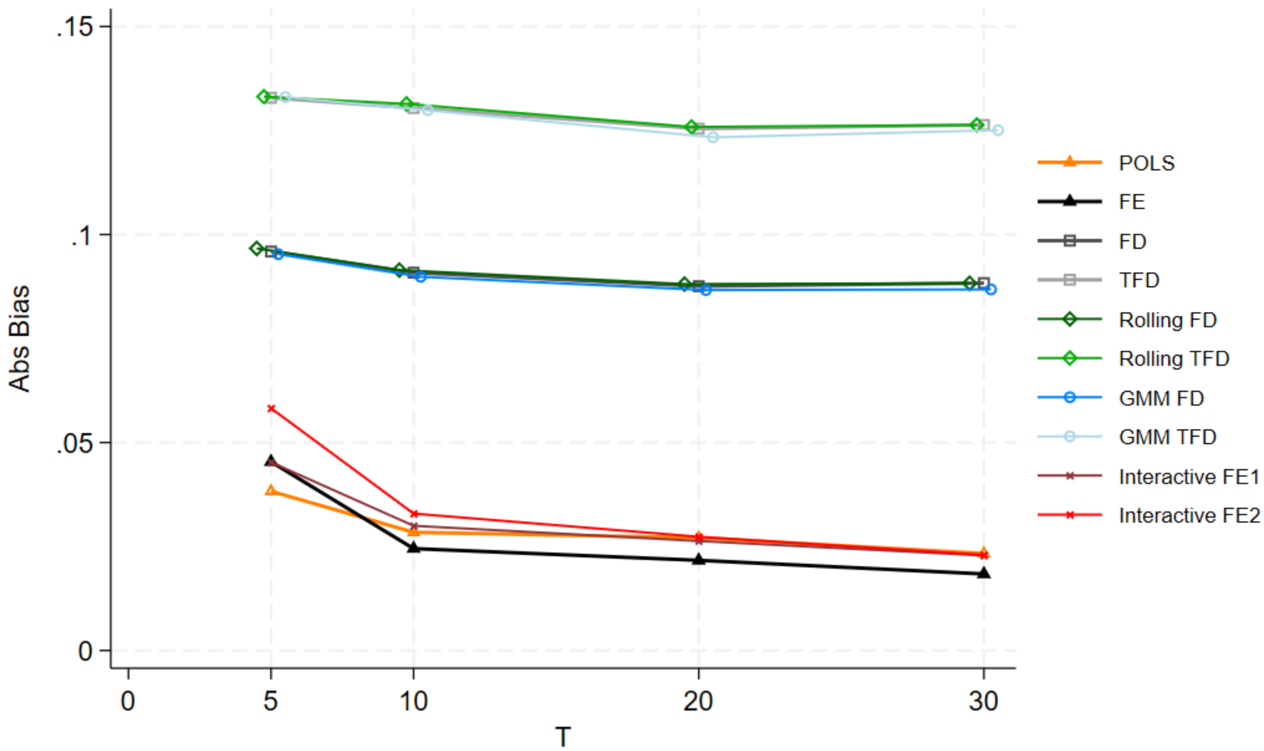
(A) $N = 100$



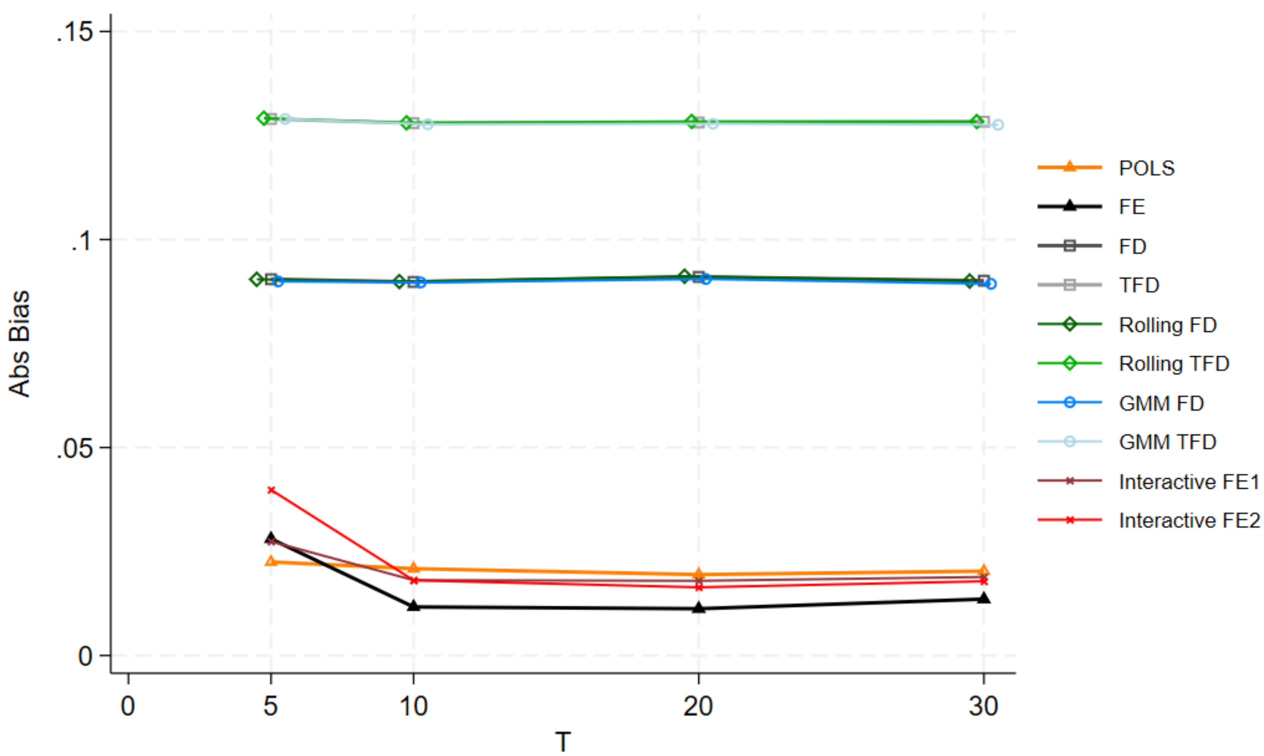
(B) $N = 500$

FIGURE B.17: Simulation Results: Bias ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.1$)

Notes: T = number of time periods per panel. See text for further details.



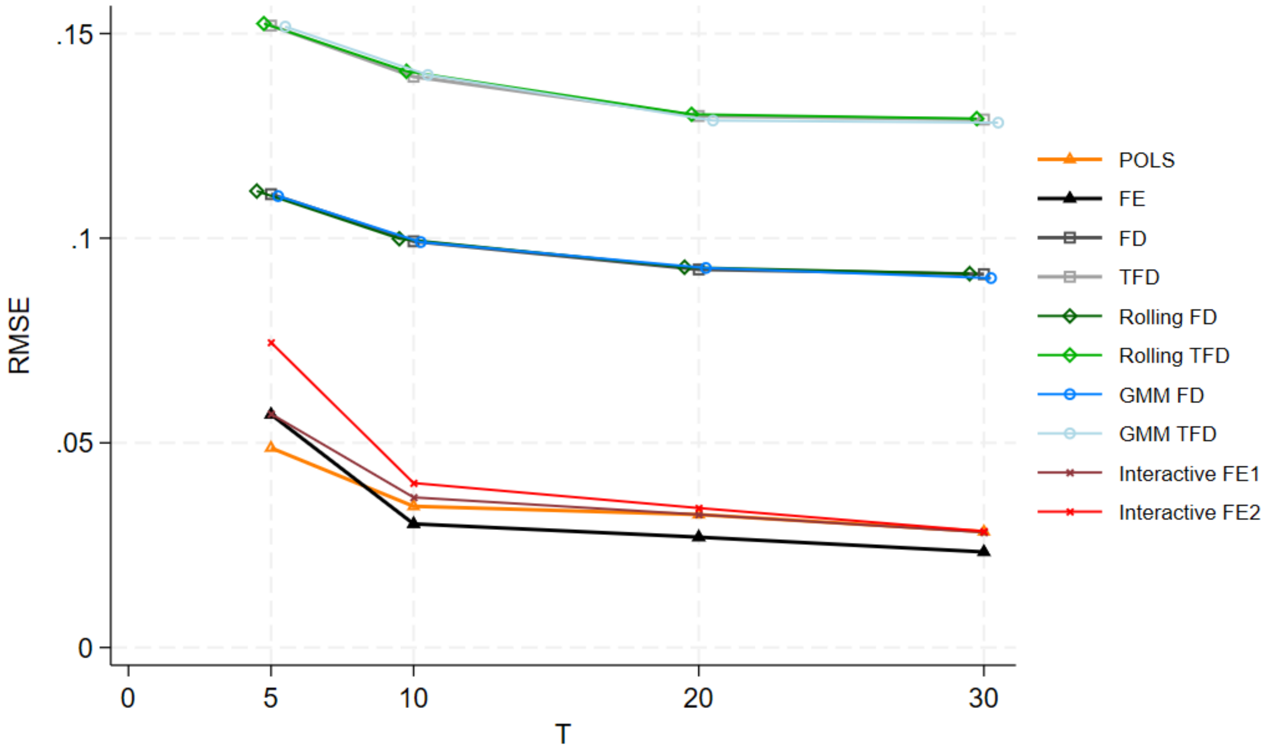
(A) $N = 100$



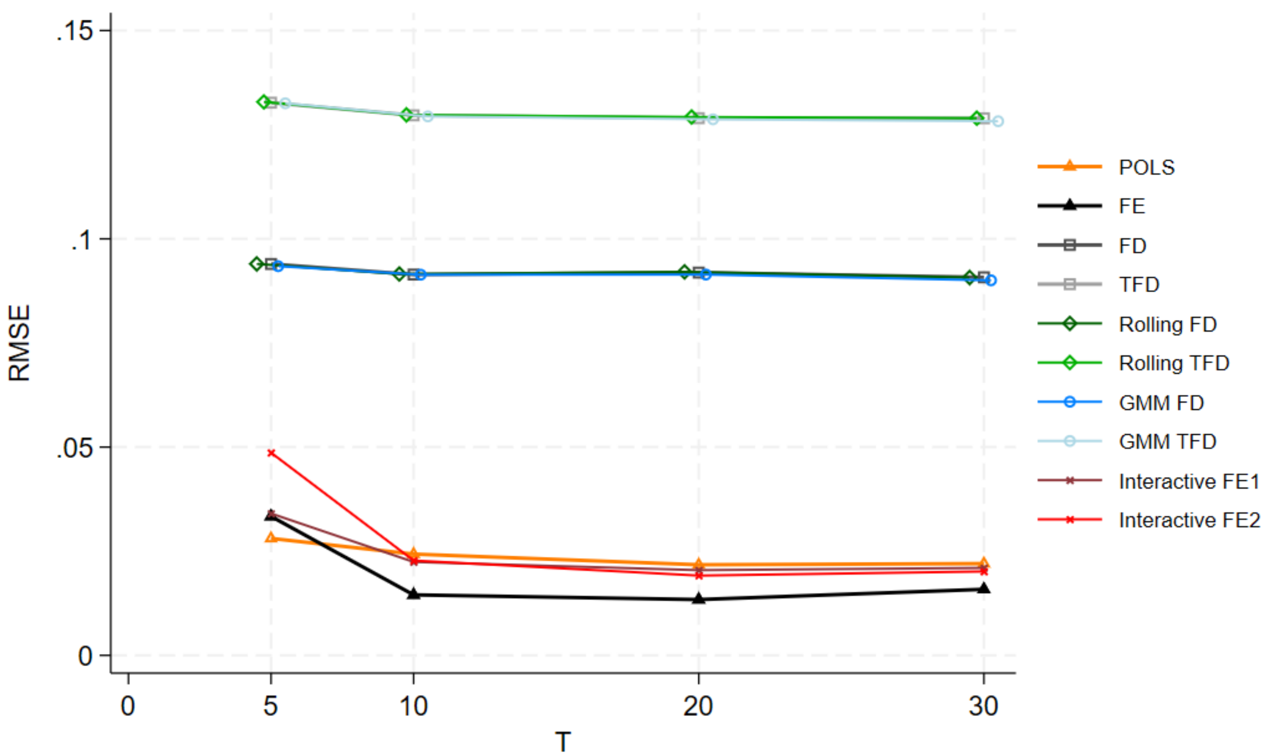
(B) $N = 500$

FIGURE B.18: Simulation Results: Absolute Bias ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.1$)

Notes: T = number of time periods per panel. See text for further details.



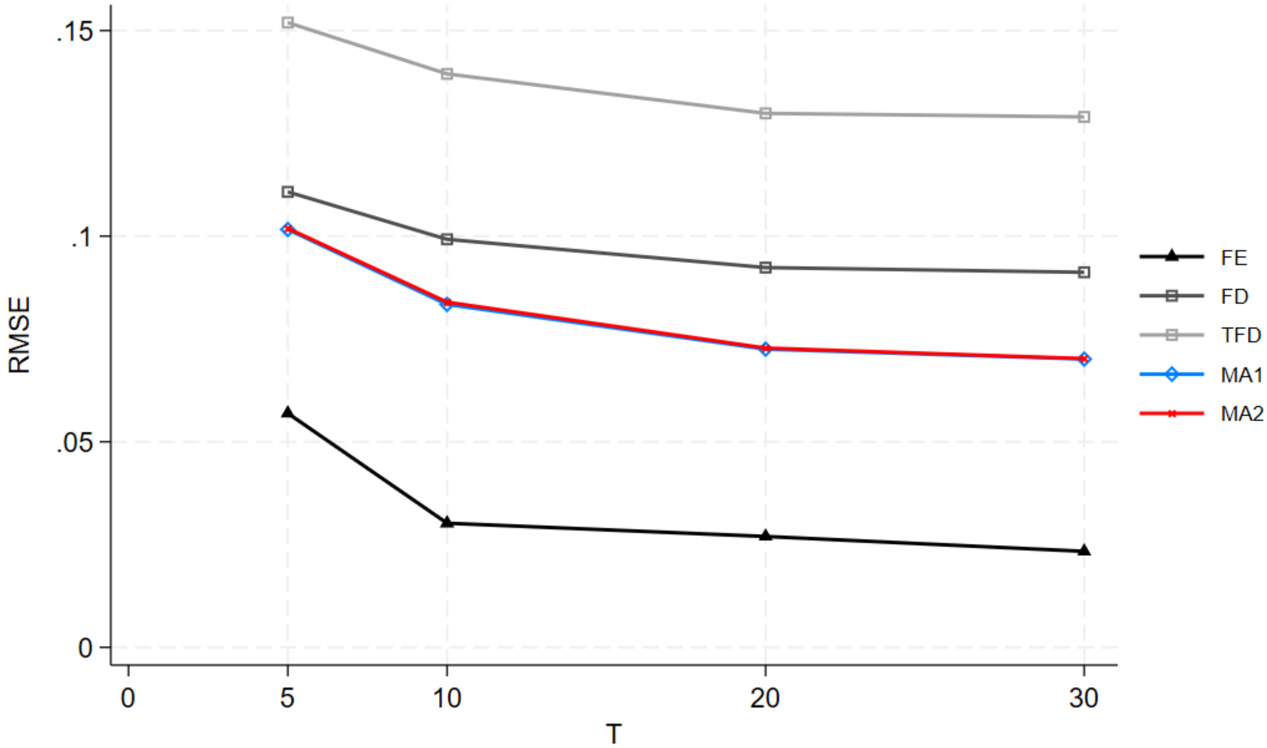
(A) $N = 100$



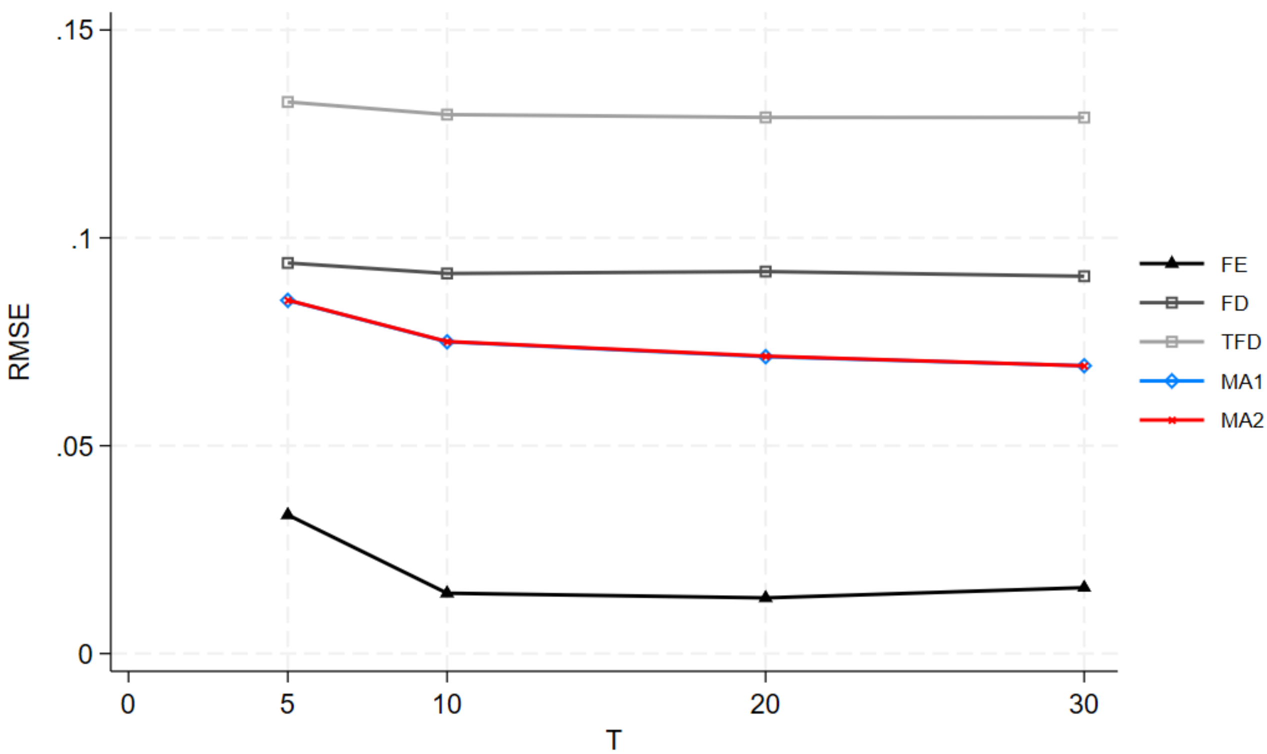
(B) $N = 500$

FIGURE B.19: Simulation Results: Root Mean Squared Error ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.1$)

Notes: T = number of time periods per panel. See text for further details.



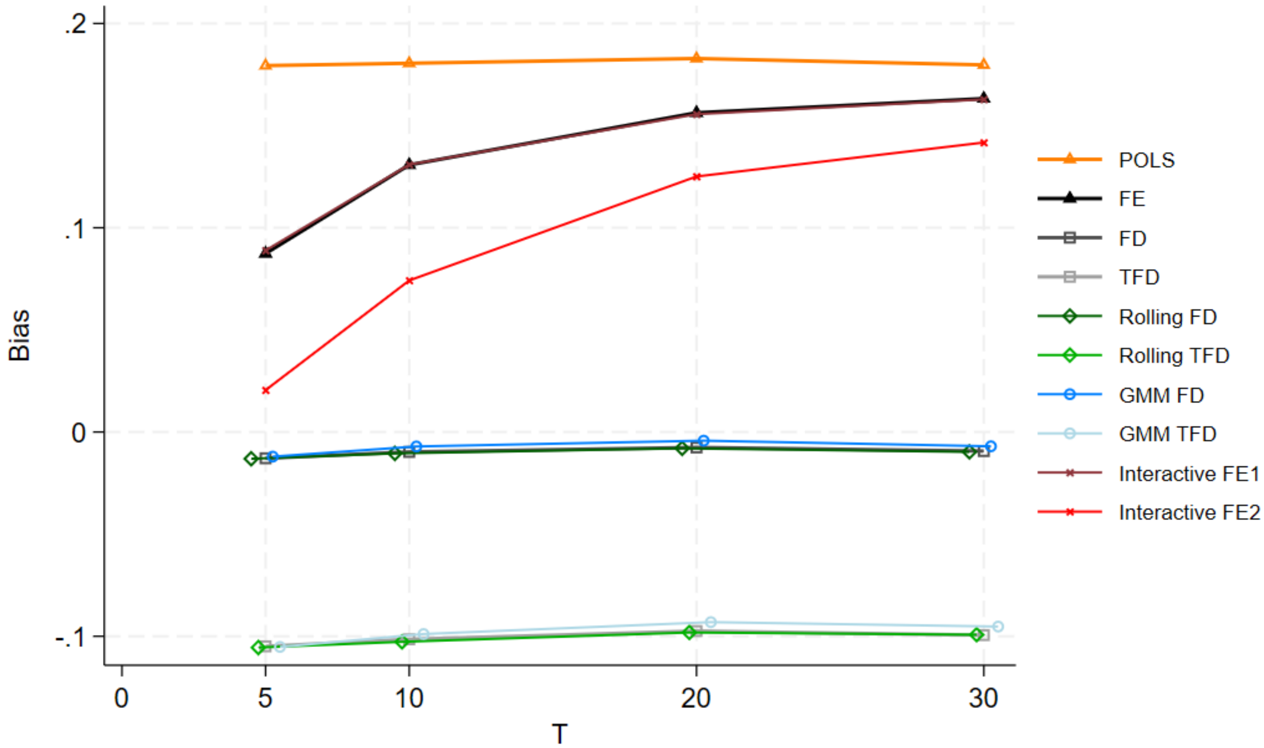
(A) $N = 100$



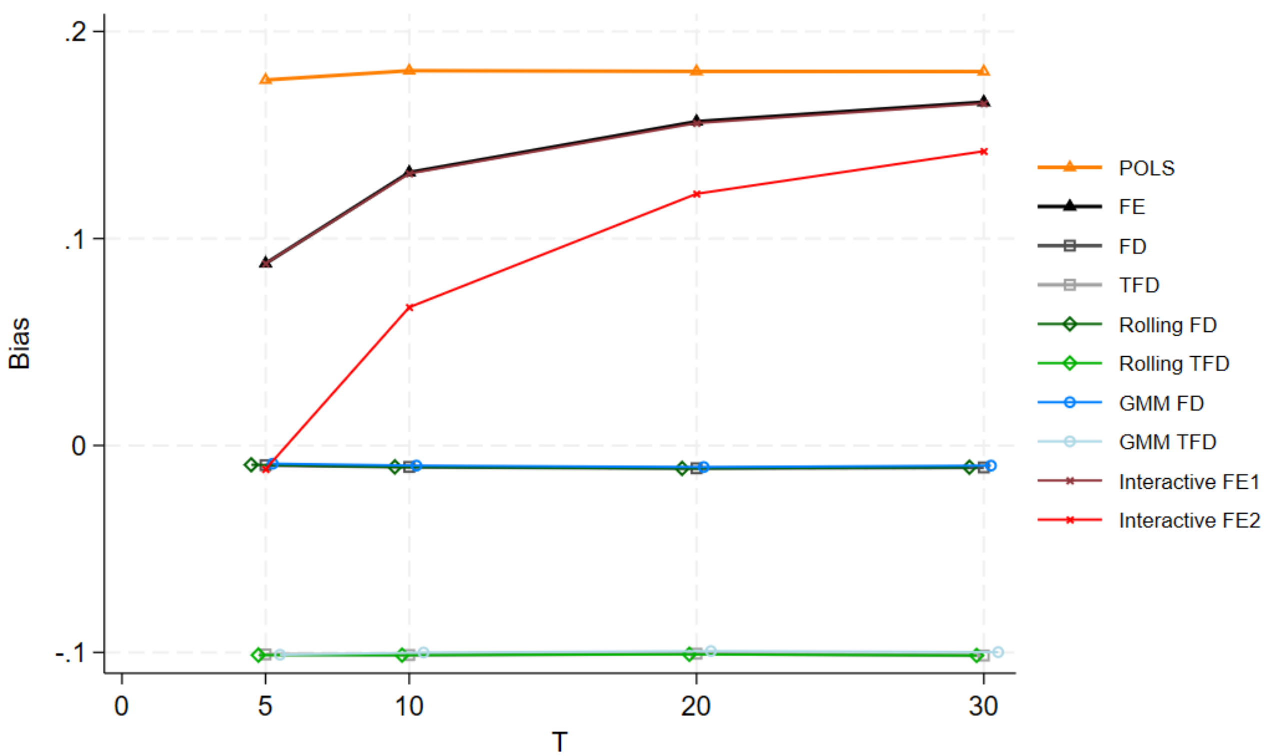
(B) $N = 500$

FIGURE B.20: Simulation Results: Model Averaging Estimators ($\delta = \rho = 0.1$)

Notes: RMSE = Root Mean Squared Error. $x_{it} = \lambda\alpha_{it-1} + z_{it}$. T = number of time periods per panel. See text for further details.



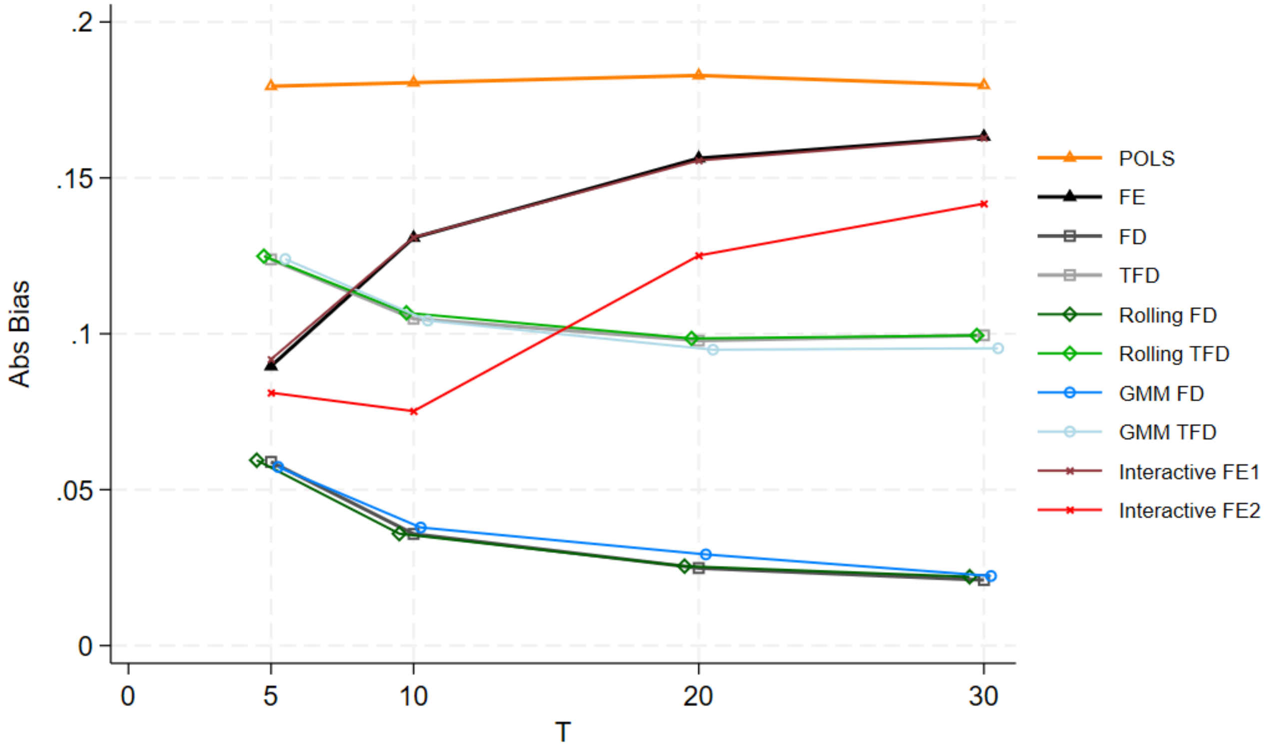
(A) $N = 100$



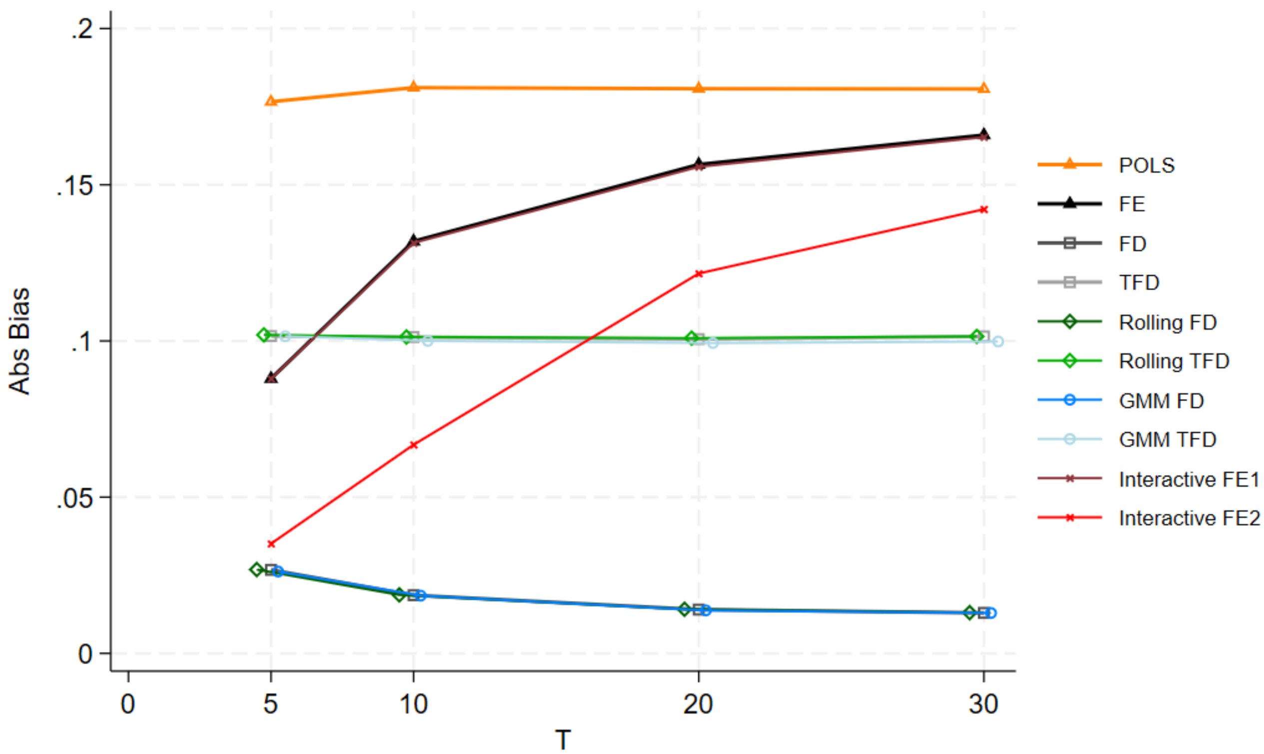
(B) $N = 500$

FIGURE B.21: Simulation Results: Bias ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.9$)

Notes: T = number of time periods per panel. See text for further details.



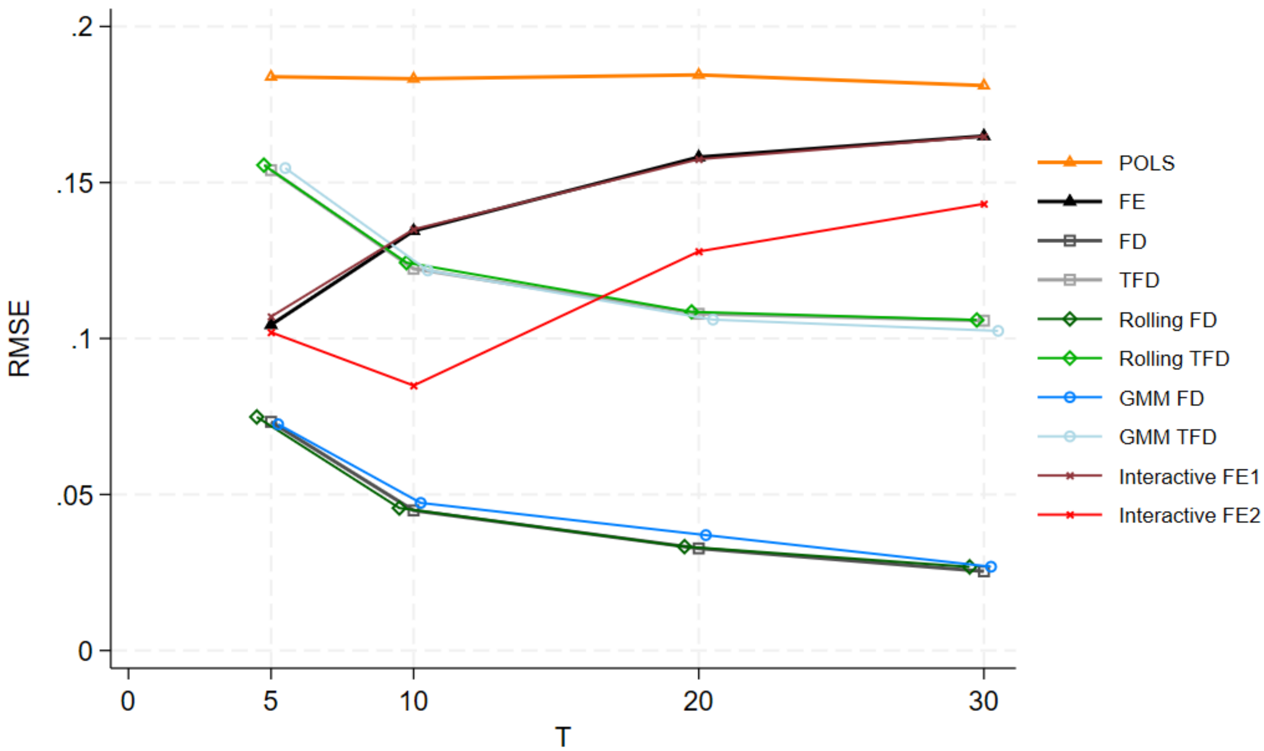
(A) $N = 100$



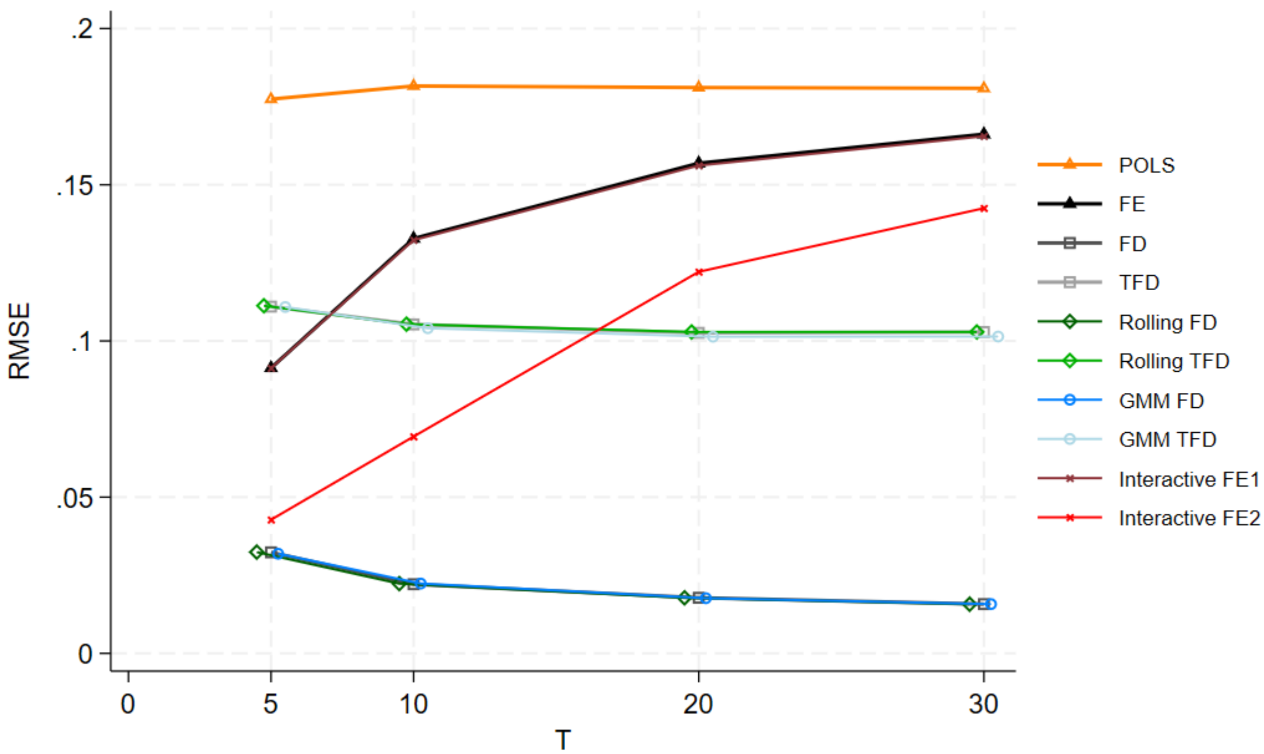
(B) $N = 500$

FIGURE B.22: Simulation Results: Absolute Bias ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.9$)

Notes: T = number of time periods per panel. See text for further details.



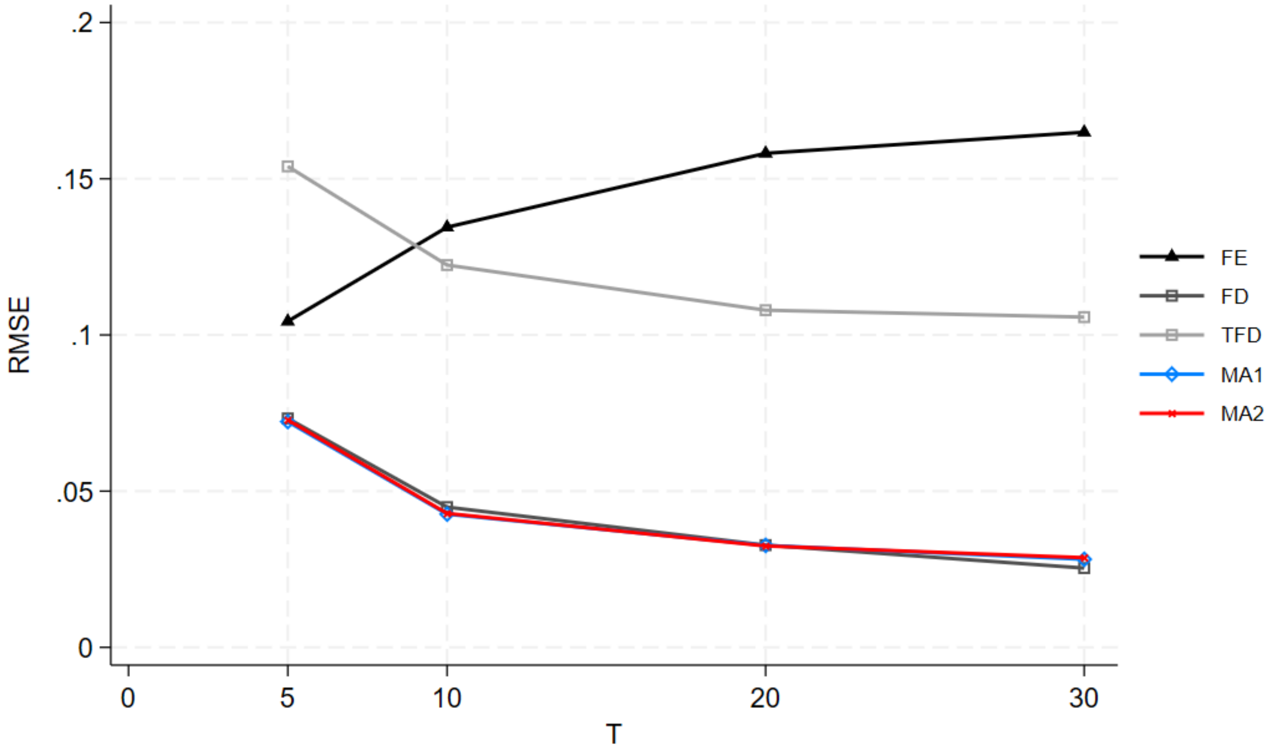
(A) $N = 100$



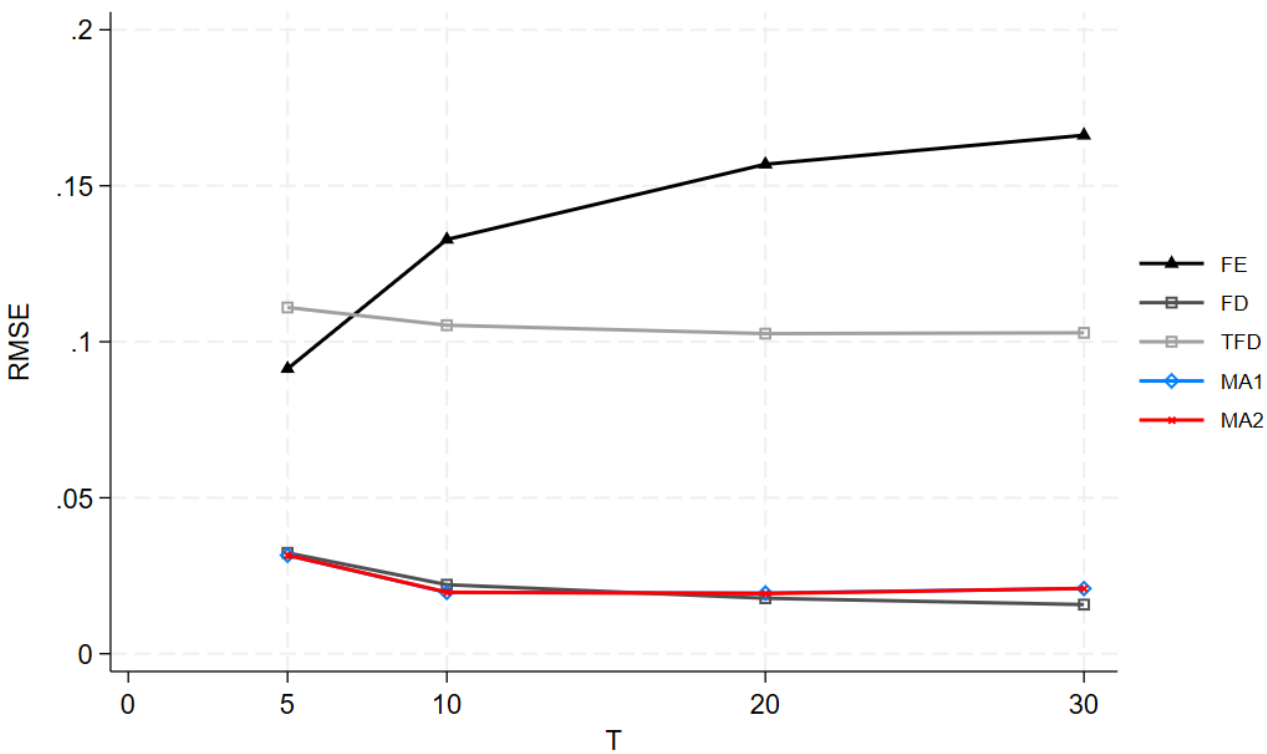
(B) $N = 500$

FIGURE B.23: Simulation Results: Root Mean Squared Error ($x_{it} = \lambda\alpha_{it-1} + z_{it}$, $\delta = \rho = 0.9$)

Notes: T = number of time periods per panel. See text for further details.



(A) $N = 100$



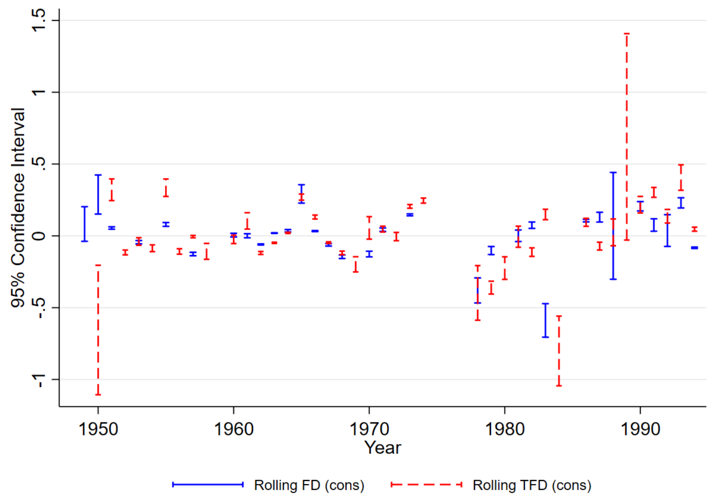
(B) $N = 500$

FIGURE B.24: Simulation Results: Model Averaging Estimators ($\delta = \rho = 0.9$)

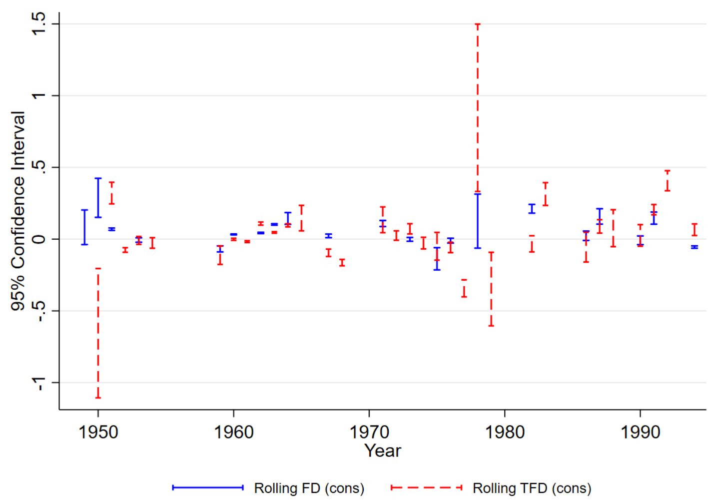
Notes: RMSE = Root Mean Squared Error. $x_{it} = \lambda\alpha_{it-1} + z_{it}$. T = number of time periods per panel. See text for further details.

C Additional Replication Results

C.1 Imai and Kim (2019)



(A) Formal Membership



(B) Informal Membership

Figure C.1: Effect of GATT Membership by Year

Note: Results are missing in years in which there is no change in GATT status from the previous year.

TABLE C.1: Replication: Formal GATT Membership – No Year FEs (Imai and Kim, 2019)

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)
GATT (Formal)	-0.048**	0.051**	0.063*	0.006	0.021	0.007	0.008
	(0.024)	(0.024)	(0.033)	(0.017)	(0.020)	(0.022)	(0.022)
GSP Dummy	0.114***	-0.002	-0.025	-0.002	0.002	-0.010	-0.021
	(0.024)	(0.016)	(0.021)	(0.014)	(0.016)	(0.019)	(0.021)
Log of Product of Real GDPs	-0.059*	0.376***	0.413**	0.227	0.318	0.121	0.002
	(0.030)	(0.088)	(0.162)	(0.170)	(0.214)	(0.108)	(0.143)
Log of Product of Real GDPs per capita	0.921***	0.012	0.016	0.137	0.240	0.118	0.223
	(0.044)	(0.096)	(0.169)	(0.189)	(0.240)	(0.140)	(0.181)
RTA Dummy	0.670***	0.027	-0.042	0.062**	0.068**	0.049*	-0.017
	(0.092)	(0.055)	(0.069)	(0.021)	(0.024)	(0.026)	(0.021)
Strict Currency Union	0.639***	0.250***	0.244***	0.132**	0.130**	0.130***	0.010
	(0.119)	(0.074)	(0.093)	(0.059)	(0.051)	(0.047)	(0.009)
Currently in Colonial Relationship	0.364**	0.004	-0.065	0.034	0.083	-0.123**	-0.146***
	(0.154)	(0.094)	(0.091)	(0.030)	(0.054)	(0.054)	(0.046)
LW Test		p=0.000					
N	196,207	175,000	160,916				

Notes: Dependent variable: *Log(Bilateral Trade Volume)*. Formal membership includes only formal GATT members as in Rose (2004); informal includes nonmember participants as in Tomz *et al.* (2007). Other controls include log product real GDP, log product real GDP per capita, and indicators for Generalized System of Preferences, a regional free trade agreement, a currency union, and currently colonized. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. IFE1 and IFE2 are omitted in Panels A and C since IFE always includes time-varying factor(s). RFD estimates based on 46 rolling regressions. RTFD based on 45 rolling regressions. * p <.10, ** p< .05, *** p<.01.

TABLE C.2: Replication: Formal GATT Membership – Year FEs (Imai and Kim, 2019)

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)	IFE1 (8)	IFE2 (9)
GATT (Formal)	0.036	0.037	0.061*	0.006	0.021	0.007	0.008	0.056**	0.023
	(0.024)	(0.024)	(0.033)	(0.017)	(0.020)	(0.022)	(0.022)	(0.024)	(0.021)
GSP Dummy	0.240***	-0.004	-0.023	-0.002	0.002	-0.010	-0.021	0.233***	0.331***
	(0.027)	(0.018)	(0.023)	(0.014)	(0.016)	(0.019)	(0.021)	(0.026)	(0.022)
Log of Product of Real GDPs	0.509***	0.412***	0.456***	0.227	0.318	0.121	0.002	0.411***	0.598***
	(0.052)	(0.088)	(0.163)	(0.170)	(0.214)	(0.108)	(0.143)	(0.019)	(0.019)
Log of Product of Real GDPs per capita	0.545***	-0.094	-0.100	0.137	0.240	0.118	0.223	0.362***	0.309***
	(0.053)	(0.095)	(0.171)	(0.189)	(0.240)	(0.140)	(0.181)	(0.028)	(0.029)
RTA Dummy	0.890***	0.043	-0.006	0.062**	0.068**	0.049*	-0.017	0.287***	0.532***
	(0.090)	(0.053)	(0.069)	(0.021)	(0.024)	(0.026)	(0.021)	(0.090)	(0.063)
Strict Currency Union	0.564***	0.246***	0.227**	0.132**	0.130**	0.130***	0.010	0.909***	0.498***
	(0.114)	(0.074)	(0.094)	(0.059)	(0.051)	(0.047)	(0.009)	(0.112)	(0.124)
Currently in Colonial Relationship	0.296*	-0.003	-0.079	0.034	0.083	-0.123**	-0.146***	0.845***	0.371***
	(0.151)	(0.093)	(0.089)	(0.030)	(0.054)	(0.054)	(0.046)	(0.168)	(0.116)
LW Test		p=0.000							
N	196,207	175,000	160,916					196,207	196,207

Notes: Dependent variable: *Log(Bilateral Trade Volume)*. Formal membership includes only formal GATT members as in Rose (2004); informal includes nonmember participants as in Tomz *et al.* (2007). Other controls include log product real GDP, log product real GDP per capita, and indicators for Generalized System of Preferences, a regional free trade agreement, a currency union, and currently colonized. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. IFE1 and IFE2 are omitted in Panels A and C since IFE always includes time-varying factor(s). RFD estimates based on 46 rolling regressions. RTFD based on 45 rolling regressions. * p <.10, ** p< .05, *** p<.01.

Table C.3: Replication: Informal GATT Membership – No Year FEs (Imai and Kim, 2019)

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)
GATT (Participate)	0.147***	0.066***	0.060*	0.041***	0.064***	0.021	0.025
	(0.030)	(0.025)	(0.036)	(0.014)	(0.014)	(0.021)	(0.021)
GSP Dummy	0.115***	-0.002	-0.026	-0.002	0.002	-0.011	-0.023
	(0.024)	(0.016)	(0.021)	(0.014)	(0.016)	(0.019)	(0.021)
Log of Product of Real GDPs	-0.080***	0.376***	0.411**	0.232	0.320	0.121	-0.006
	(0.030)	(0.088)	(0.162)	(0.170)	(0.213)	(0.108)	(0.140)
Log of Product of Real GDPs per capita	0.918***	0.012	0.018	0.129	0.236	0.127	0.238
	(0.043)	(0.096)	(0.169)	(0.189)	(0.239)	(0.142)	(0.179)
RTA Dummy	0.663***	0.026	-0.042	0.062**	0.068**	0.051*	-0.016
	(0.091)	(0.055)	(0.069)	(0.021)	(0.024)	(0.026)	(0.021)
Strict Currency Union	0.626***	0.248***	0.241***	0.078**	0.091***	0.076**	0.008
	(0.119)	(0.074)	(0.093)	(0.035)	(0.032)	(0.032)	(0.007)
Currently in Colonial Relationship	0.372**	-0.004	-0.071	0.019	0.042	-0.089**	-0.119**
	(0.156)	(0.094)	(0.091)	(0.028)	(0.047)	(0.042)	(0.047)
LW Test		p=0.000					
N	196,207	175,000	160,916				

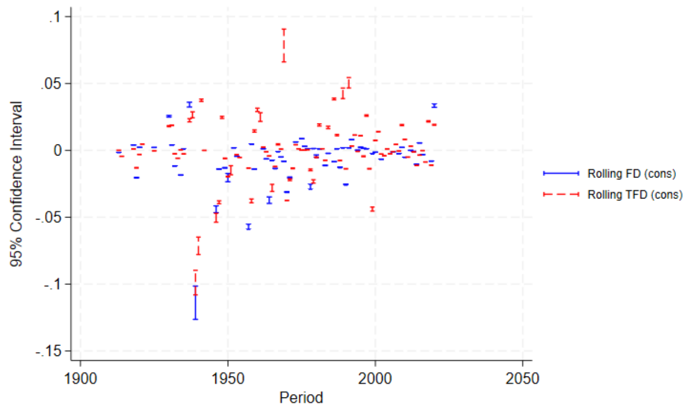
Notes: Dependent variable: *Log(Bilateral Trade Volume)*. Formal membership includes only formal GATT members as in Rose (2004); informal includes nonmember participants as in Tomz *et al.* (2007). Other controls include log product real GDP, log product real GDP per capita, and indicators for Generalized System of Preferences, a regional free trade agreement, a currency union, and currently colonized. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. IFE1 and IFE2 are omitted in Panels A and C since IFE always includes time-varying factor(s). RFD estimates based on 46 rolling regressions. RTFD based on 45 rolling regressions. * p <.10, ** p< .05, *** p<.01.

Table C.4: Replication: Informal GATT Membership – Year FEs (Imai and Kim, 2019)

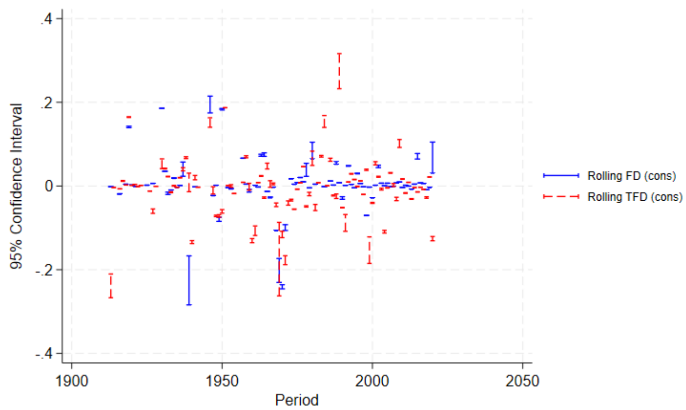
	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GATT (Participate)	0.227***	0.044*	0.054	0.041***	0.064***	0.021	0.025	0.262***	0.014
	(0.030)	(0.026)	(0.036)	(0.014)	(0.014)	(0.021)	(0.021)	(0.029)	(0.027)
GSP Dummy	0.250***	-0.004	-0.023	-0.002	0.002	-0.011	-0.023	0.246***	0.331***
	(0.027)	(0.018)	(0.023)	(0.014)	(0.016)	(0.019)	(0.021)	(0.026)	(0.022)
Log of Product of Real GDPs	0.525***	0.411***	0.454***	0.232	0.320	0.121	-0.006	0.420***	0.599***
	(0.052)	(0.088)	(0.163)	(0.170)	(0.213)	(0.108)	(0.140)	(0.019)	(0.019)
Log of Product of Real GDPs per capita	0.521***	-0.094	-0.098	0.129	0.236	0.127	0.238	0.345***	0.308***
	(0.053)	(0.095)	(0.171)	(0.189)	(0.239)	(0.142)	(0.179)	(0.028)	(0.029)
RTA Dummy	0.892***	0.043	-0.006	0.062**	0.068**	0.051*	-0.016	0.272***	0.532***
	(0.089)	(0.053)	(0.069)	(0.021)	(0.024)	(0.026)	(0.021)	(0.088)	(0.063)
Strict Currency Union	0.545***	0.245***	0.224**	0.078**	0.091***	0.076**	0.008	0.880***	0.498***
	(0.114)	(0.074)	(0.094)	(0.035)	(0.032)	(0.032)	(0.007)	(0.111)	(0.124)
Currently in Colonial Relationship	0.269*	-0.009	-0.086	0.019	0.042	-0.089**	-0.119**	0.806***	0.363***
	(0.152)	(0.093)	(0.088)	(0.028)	(0.047)	(0.042)	(0.047)	(0.166)	(0.116)
LW Test		p=0.000							
N	196,207	175,000	160,916					196,207	196,207

Notes: Dependent variable: *Log(Bilateral Trade Volume)*. Formal membership includes only formal GATT members as in Rose (2004); informal includes nonmember participants as in Tomz *et al.* (2007). Other controls include log product real GDP, log product real GDP per capita, and indicators for Generalized System of Preferences, a regional free trade agreement, a currency union, and currently colonized. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. IFE1 and IFE2 are omitted in Panels A and C since IFE always includes time-varying factor(s). RFD estimates based on 46 rolling regressions. RTFD based on 45 rolling regressions. * p <.10, ** p< .05, *** p<.01.

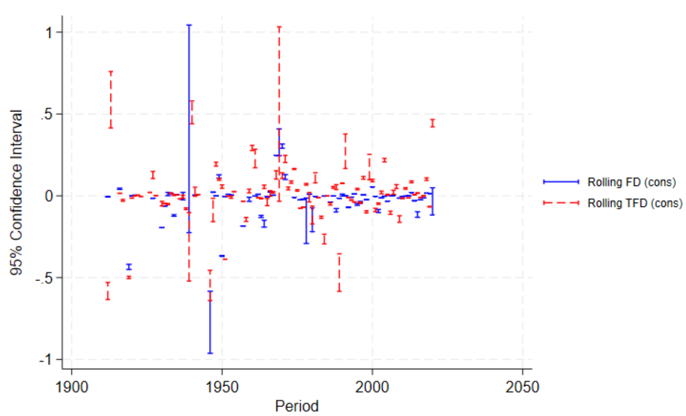
C.2 Leipziger (2024)



(A) Democracy (No Interaction)



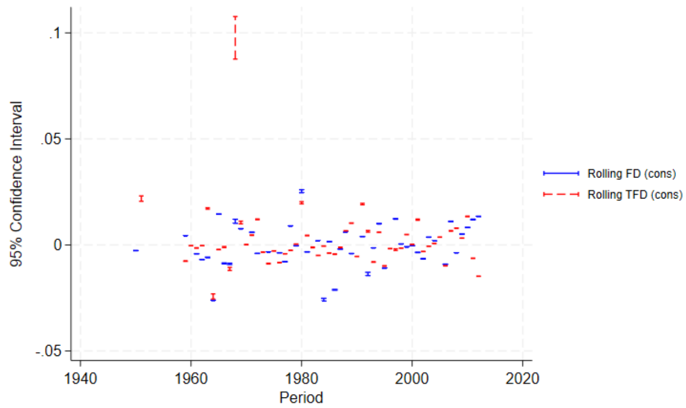
(B) Democracy (With Interaction)



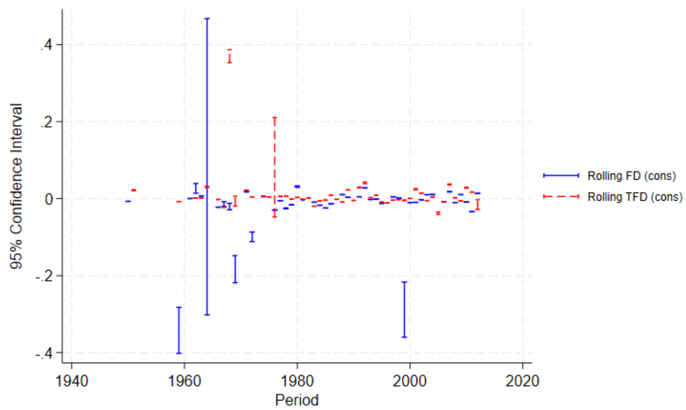
(C) Democracy × Predemocratic Inequality (With Interaction)

FIGURE C.2: Effect of Democracy on Public Service Inequality by Year

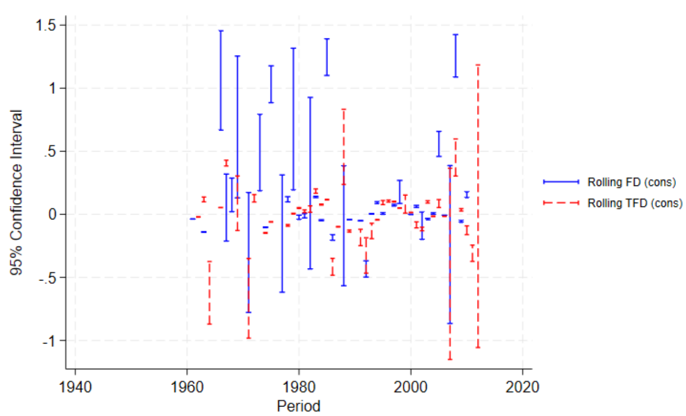
Note: Results are omitted in years in which the confidence intervals are extremely wide and distort the scaling of the figure.



(A) Democracy (No Interaction)



(B) Democracy (With Interaction)

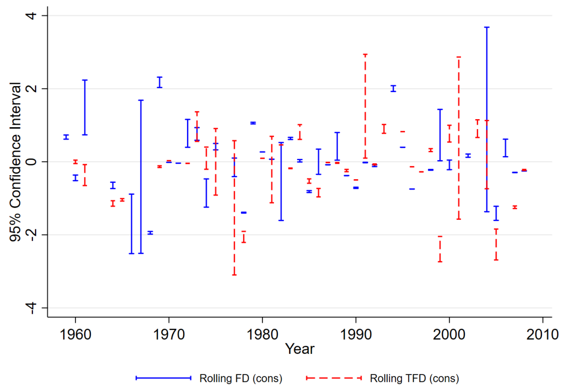


(C) Democracy × Predemocratic Inequality (With Interaction)

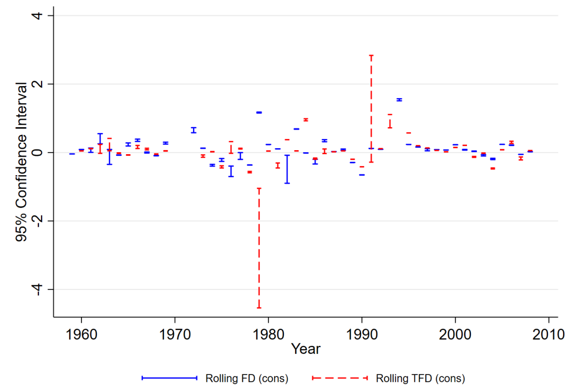
FIGURE C.3: Effect of Democracy on Education Inequality by Year

Note: Results are omitted in years in which the confidence intervals are extremely wide and distort the scaling of the figure.

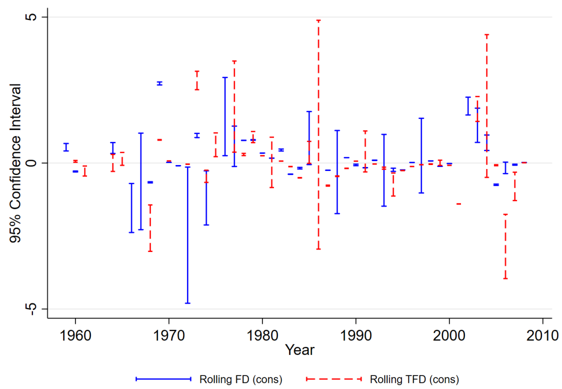
C.3 James (2015)



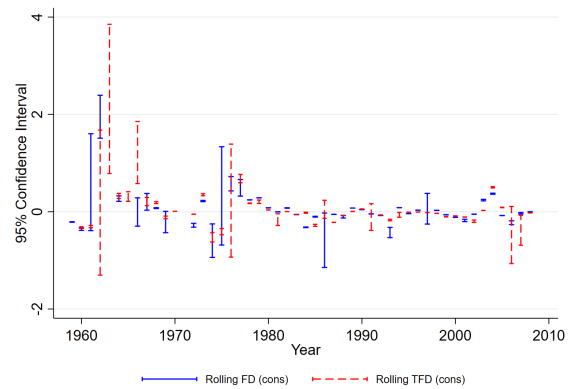
(A) Nonresource Revenue



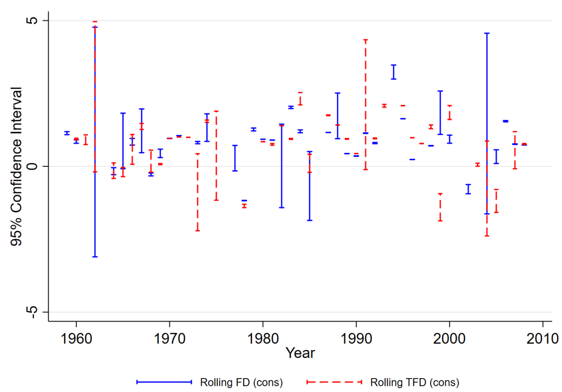
(B) Income Tax Revenue



(C) Total Expenditures



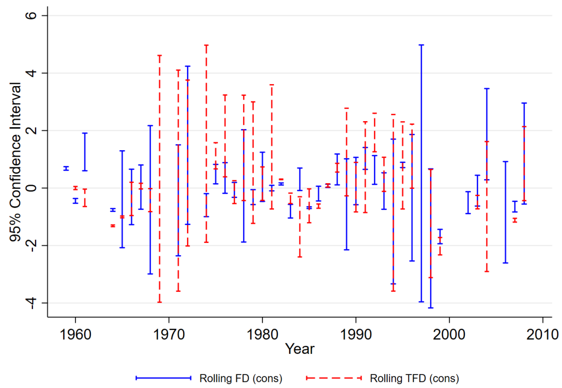
(D) Education Expenditures



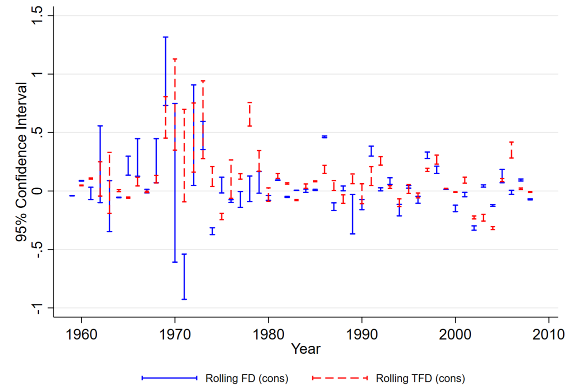
(E) Public Savings

FIGURE C.4: Effect of Resource Revenue on Fiscal Policy by Year: Full Sample

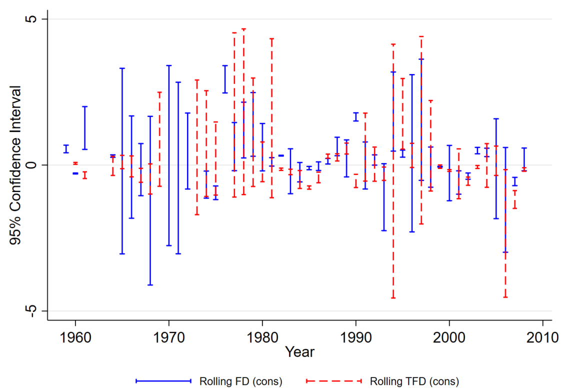
Note: Results are omitted in years in which the confidence intervals are extremely wide and distort the scaling of the figure.



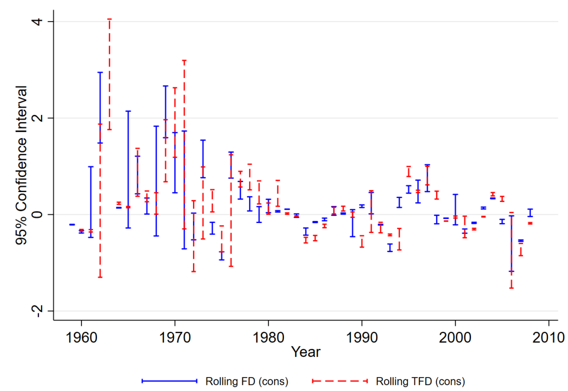
(A) Nonresource Revenue



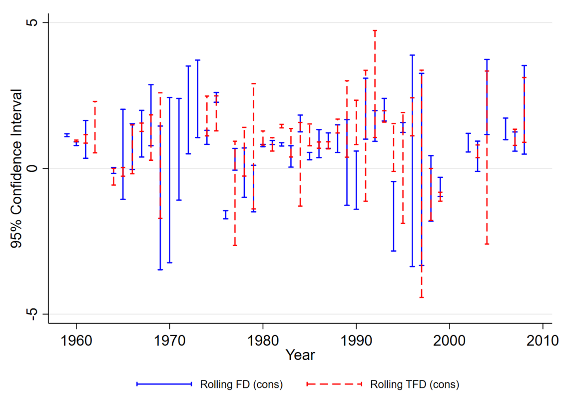
(B) Income Tax Revenue



(C) Total Expenditures



(D) Education Expenditures

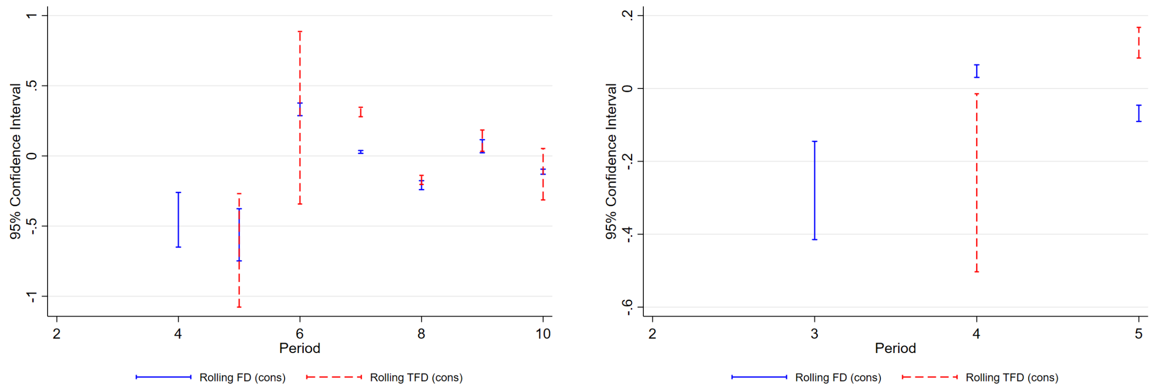


(E) Public Savings

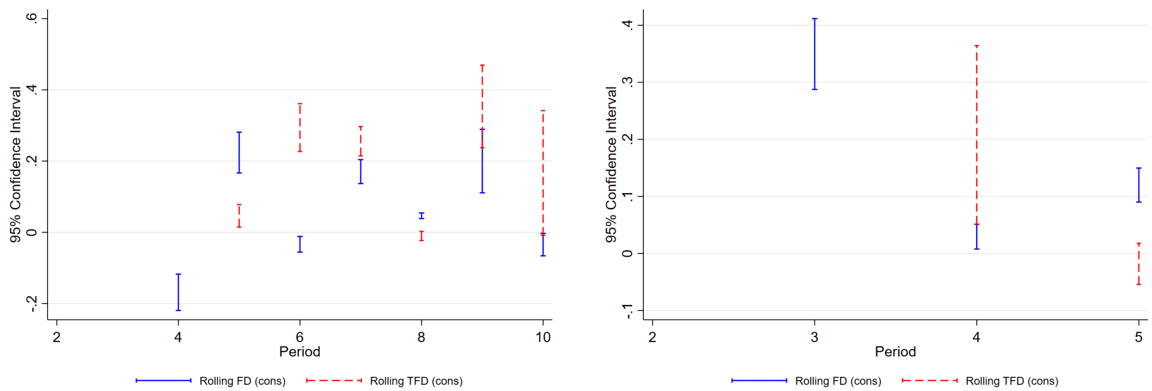
Figure C.5: Effect of Resource Revenue on Fiscal Policy by Year: Omit Alaska

Notes: Results are omitted in years in which the confidence intervals are extremely wide and distort the scaling of the figure.

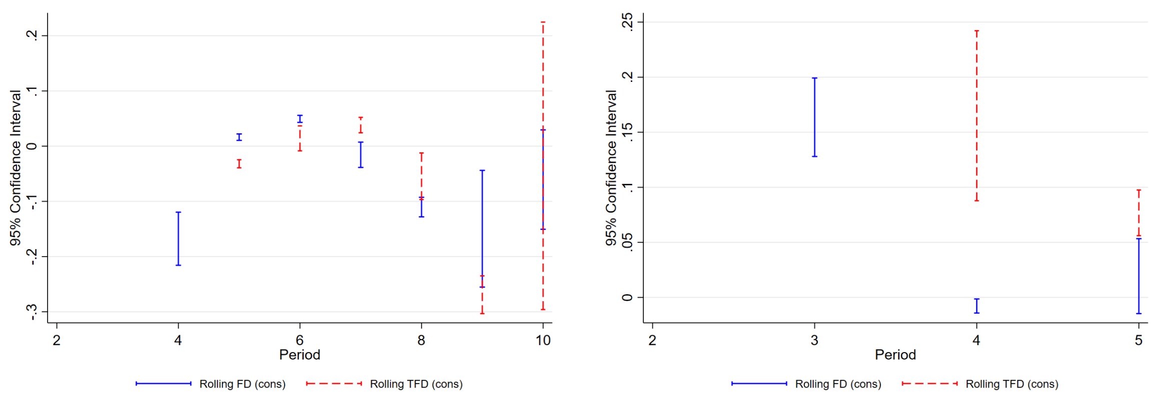
C.4 Djankov and Reynal-Querol (2010)



(A) Incidence of Civil Wars (25 Deaths)



(B) Onset of War (ACD, 1,000+ Deaths)



(C) Incidence of Civil Wars (1,000 Deaths Per Year)

FIGURE C.6: Effect of $log(GDP_{t-1})$ by Period

Notes: Data in five-year intervals in left column. Data in ten-year intervals in right column.

TABLE C.5: Replication: Incidence of Civil Wars (25 Deaths) (Djankov and Reynal-Querol, 2010)

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Five-Year Intervals									
$log(GDP_{t-1})$	-0.087*	-0.035	0.005	-0.027	0.011	0.042	0.041	0.002	0.007
	(0.046)	(0.060)	(0.120)	(0.071)	(0.063)	(0.111)	(0.106)	(0.032)	(0.041)
$log(POP_{t-1})$	0.016	-0.001	-0.579	-0.126	0.044	-0.745	-0.684	0.033	0.031
	(0.078)	(0.111)	(0.599)	(0.205)	(0.153)	(0.726)	(0.748)	(0.021)	(0.026)
LW Test		p=0.528							
N	1,169	985	816					1,169	1,169
Panel B. Ten-Year Intervals									
$log(GDP_{t-1})$	-0.059	-0.044	-0.003	-0.022	-0.021	0.069	0.080	0.028	0.088
	(0.057)	(0.064)	(0.153)	(0.063)	(0.051)	(0.136)	(0.151)	(0.054)	(0.130)
$log(POP_{t-1})$	-0.002	0.030	0.783	-0.002	0.114	0.580	0.591	0.066	0.076
	(0.097)	(0.102)	(0.548)	(0.017)	(0.078)	(0.919)	(0.924)	(0.056)	(0.088)
LW Test		p=0.481							
N	576	407	251					576	576

Notes: Dependent variable: Indicator for *Incidence of Civil Wars (25 Deaths)*. N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 7 (Panel A) and 3 (Panel B) rolling regressions. RTFD based on 6 (Panel A) and 2 (Panel B) rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

Table C.6: Replication: Onset of War (ACD, 1,000+ Deaths) (Djankov and Reynal-Querol, 2010)

	FE (1)	FD (2)	Twice FD (3)	RFD (cons) (4)	RFD (no cons) (5)	Twice RFD (cons) (6)	Twice RFD (no cons) (7)	IFE1 (8)	IFE2 (9)
Panel A. Five-Year Intervals									
$log(GDP_{t-1})$	0.030	0.059	0.180**	0.040	0.028	0.094	0.071	-0.018***	-0.018***
	(0.023)	(0.042)	(0.082)	(0.040)	(0.037)	(0.058)	(0.056)	(0.005)	(0.006)
$log(POP_{t-1})$	0.067**	0.225**	0.697*	0.159	0.039	0.548**	0.506*	0.013***	0.012***
	(0.032)	(0.104)	(0.389)	(0.116)	(0.111)	(0.207)	(0.219)	(0.003)	(0.004)
LW Test		p=0.096							
N	1,169	985	816					1,169	1,169
Panel B. Ten-Year Intervals									
$log(GDP_{t-1})$	0.082*	0.089	0.037	0.115	0.057	0.024	-0.021	-0.025**	-0.027
	(0.048)	(0.060)	(0.131)	(0.080)	(0.046)	(0.088)	(0.151)	(0.011)	(0.018)
$log(POP_{t-1})$	0.112*	0.195**	0.542	0.149	0.040	0.594**	0.689	0.019***	0.018
	(0.061)	(0.094)	(0.495)	(0.151)	(0.097)	(0.019)	(0.144)	(0.007)	(0.012)
LW Test		p=0.110							
N	576	407	251					576	576

Notes: Dependent variable: Indicator for *Onset of War (ACD, 1,000+ Deaths)*. N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 7 (Panel A) and 3 (Panel B) rolling regressions. RTFD based on 6 (Panel A) and 2 (Panel B) rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.

Table C.7: Replication: Incidence of Civil War (1,000 Deaths per Year) (Djankov and Reynal-Querol, 2010)

	FE	FD	Twice FD	RFD (cons)	RFD (no cons)	Twice RFD (cons)	Twice RFD (no cons)	IFE1	IFE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Five-Year Intervals									
$log(GDP_{t-1})$	-0.035	-0.040	-0.057	-0.004	0.012	-0.033	-0.035	-0.011	-0.015
	(0.035)	(0.041)	(0.070)	(0.026)	(0.024)	(0.037)	(0.037)	(0.010)	(0.011)
$log(POP_{t-1})$	0.036	0.025	0.033	0.040	0.101*	-0.098	-0.107	0.009	0.006
	(0.043)	(0.091)	(0.216)	(0.083)	(0.051)	(0.251)	(0.259)	(0.006)	(0.006)
LW Test		p=0.919							
N	1,169	985	816					1,169	1,169
Panel B. Ten-Year Intervals									
$log(GDP_{t-1})$	-0.013	0.029	0.098	0.018	0.009	0.095	0.091	0.016	-0.001
	(0.056)	(0.058)	(0.094)	(0.040)	(0.021)	(0.036)	(0.069)	(0.025)	(0.034)
$log(POP_{t-1})$	0.098	0.103*	-0.020	0.082	0.112	0.001	0.048	0.009	0.003
	(0.060)	(0.062)	(0.300)	(0.091)	(0.094)	(0.010)	(0.096)	(0.017)	(0.023)
LW Test		p=0.232							
N	576	407	251					576	576

Notes: Dependent variable: Indicator for *Incidence of Civil War (1,000 Deaths per Year)*. N = number of observations. LW = Laporte and Windmeijer (2005) test of equality of FE and FD. FE = fixed effects. FD = first-differences. RFD = rolling first differences. IFE1 = interactive fixed effects (1 factor). IFE2 = interactive fixed effects (2 factors). cons/no cons refers to the inclusion of a constant in the first-differenced or twice-differenced specifications. RFD estimates based on 7 (Panel A) and 3 (Panel B) rolling regressions. RTFD based on 6 (Panel A) and 2 (Panel B) rolling regressions. Time fixed effects included in all models. * p <0.10, ** p<0.05, *** p<0.01.