Dynamic Biases of Static Panel Data Estimators

Sylvia Klosin*†

January 21, 2026

Abstract

This paper identifies an important bias — termed dynamic bias — in fixed effects panel estimators that arises when dynamic feedback is ignored in the estimating equation. Dynamic feedback occurs if past outcomes impact current outcomes, a feature of many settings ranging from economic growth to labor markets. When estimating equations omit past outcomes, dynamic bias can lead to significantly inaccurate treatment effect estimates, even with randomly assigned treatments. I show that dynamic bias stems from the estimation of fixed effects, as their estimation generates confounding in the data. This dynamic bias in simulations is an order of magnitude larger than Nickell bias. To recover consistent treatment effects, I develop a flexible estimator that provides fixed-T bias correction. I apply this approach to study the impact of temperature shocks on GDP, a canonical example where economic theory points to an important feedback from past to future outcomes. Accounting for dynamic bias lowers the estimated effects of higher annual temperatures on GDP growth by 10% and GDP levels by 120%.

Keywords: treatment effects, fixed effects panel model, dynamic panel model, climate economics

JEL classification: C33, Q51

*Department of Agricultural and Resource Economics, University of California, Davis. Email: sklosin@ucdavis.edu.

†I thank Alberto Abadie, Isaiah Andrews, Karl Aspelund, Tamma Carleton, Victor Chernozhukov, Jonathan Colmer, Lindsey Currier, Ivan Fernandez-Val, Bulat Gafarov, Dalia Ghanem, Vitor Hadad, Chris Hansen, Peter Hull, Kelsey Jack, Katrina Jessoe, Sara Johns, Clair Lazar Reich, Pierre Merel, Douglas Miller, Anna Mikusheva, Ishan Nath, Whitney Newey, Benjamin Olken, Dev Patel, Deborah Plana, Jacquelyn Pless, Felix Pretis, Ashesh Rambachan, Jonathan Roth, Wolfram Schlenker, Naomi Shimberg, Rahul Singh, Liyang Sun, Jenny Wang; I am also very grateful to Andrew Goodman-Bacon for detailed feedback and help as a discussant of my paper at the 2026 ASSA meetings. I acknowledge generous support from the Jerry A. Hausman Graduate Dissertation Fellowship and the NSF GRFP. This draft is a work in progress and comments are welcome; all errors are my own.

1 Introduction

Applied researchers routinely use panel models with fixed effects to estimate causal effects. Nearly one in five empirical AER papers (2010–2012) used some form of fixed-effects panel regression [Chaisemartin and d'Haultfoeuille, 2020]. Crucially, these models are often deployed in settings where outcomes are dynamic - past outcome realizations affect current ones. A large body of theoretical and empirical work indicates that outcomes are dynamic across economic fields, including: aggregate output through capital accumulation [Solow, 1956, Olley and Pakes, 1992]; pollution and emissions via stock dynamics [Ang, 2006]; employment and human capital through persistence and state dependence [Blanchard and Summers, 1988, Cunha and Heckman, 2007] agricultural yields via soil-capital stocks [Griliches, 1963], and purchasing behavior through habit formation [Pollak, 1970].

Despite clear economic arguments for dynamics, many influential applied papers study these outcomes using what I call a *static* fixed-effects panel model: a panel model that does not include past outcomes ("lagged Y") as a control variable. Examples include papers that estimate the effects of temperature on GDP [e.g., Dell et al., 2012], crop yields [e.g., Annan and Schlenker, 2015, Burke et al., 2015], test scores [e.g., Cho, 2017, Graff Zivin et al., 2018, Garg et al., 2020], the effects of precipitation on employment [e.g., Jessoe et al., 2018], and policy changes on carbon emissions [e.g., Metcalf, 2019]. Across these settings, researchers typically treat the treatment variable—e.g., a temperature shock—as quasi-randomly assigned. Consequently, they conclude that omitting lagged outcomes does not bias their treatment-effect estimates, since no omitted-variable bias (OVB) arises in this case.[1]

This paper shows that this reasoning is incorrect. Although researchers are correct that there is no OVB, a different bias arises. The first contribution of the paper formalizes this problem and bias. I show that even when treatment is completely randomly assigned, estimating static fixed-effects panel models in settings with dynamic outcomes yields systematically biased treatment-effect estimates. I call this "dynamic bias." This paper characterizes this bias theoretically and empirically. The paper's second contribution is to demonstrate that accounting for dynamic bias

[1] OVB occurs when an omitted variable is correlated with both the treatment and the outcome. If the treatment is randomly assigned, past outcomes are uncorrelated with treatment and therefore do not generate OVB [Angrist and Pischke, 2009].

materially changes conclusions in canonical climate–economy applications and, in turn, to clarify the appropriate causal estimands by distinguishing the short-run (contemporaneous) effect from the long-run effect of treatment. Third, I develop a new estimator that I term Dynamic Bias Correction (DBC), which removes dynamic bias without the use of instrumental variables. I show that DBC is consistent and asymptotically normal.

First, I establish and quantify dynamic bias analytically. I derive closed-form expressions that characterize this bias. Dynamic bias arises in static models because fixed effect transformations - whether via unit dummies, demeaning, or differencing - induce correlation between the transformed treatment and the transformed error whenever past outcomes influence current outcomes. Thus, accounting for fixed effects creates confounding in the transformed equation. Consequently, because lagged outcomes are omitted from the model, treatment-effect estimates will be biased even under completely random treatment assignment. It is also worth noting that treatment effects are even more biased if treatment is related to past outcomes and therefore endogenous.[2]

One reason researchers often avoid including lagged outcomes is that doing so introduces another well-known problem - Nickell bias [Nickell, 1981]. Nickell bias arises when lagged outcomes are included in models with unit fixed effects. This creates a practical trade-off: omitting lags produces dynamic bias, while including them induces Nickell bias. I compare these two sources of bias analytically in the context of an AR(1) model where treatment is randomly assigned conditional on unit fixed effects. This analytical comparison shows that dynamic bias in the treatment-effect coefficient indeed decays more slowly than the Nickell bias in the treatment-effect coefficient. I further show in simulations across a range of data-generating processes that dynamic bias exceeds Nickell bias by more than an order of magnitude. This formal analysis underscores that the prevailing practice of omitting lagged outcomes to avoid Nickell bias actually exacerbates bias overall by inducing dynamic bias; the remedy has been worse than the original problem.

The second contribution is to show that dynamic bias matters empirically. I use the canonical Dell et al. [2012] dataset on temperature and GDP. Adding the past outcome to their static specification

[2] Marx et al. [2022] and Ghanem et al. [2022] show how many economic models lead to treatment selection that depends on past outcomes. For example, policies that target air pollution in particular areas are implemented because of past pollution rates [Chay and Greenstone, 2005]. As another example, regional deforestation protection in Brazil is based on past deforestation in the region [Harding et al., 2021, Assunção et al., 2023]. Despite endogenous treatment these papers do not control for past outcomes.

changes the estimated effect of temperature on GDP growth by 10 percent and on GDP levels by 120 percent.[3][4] Importantly, Dell et al. [2012] describe their parameter as the "contemporary causal effect of temperature on the development process," but do not formalize the estimand.

Applied work often leaves the estimand implicit when using fixed-effects panels in dynamic settings. I make this object explicit and show that two distinct targets are natural: (i) the short-run (contemporaneous) effect - the immediate impact of increasing today's treatment on the outcome, holding past and future treatments fixed—and (ii) the long-run effect—the eventual change in the outcome level under a permanent increase in treatment from today onward [Koyck, 1954, Nerlove, 1958].[5] I prove that the widely used static fixed-effects estimator recovers neither object because of dynamic bias, even under random treatment assignment. This paper clarifies what should be estimated in dynamic outcome environments, and demonstrates that prevailing practice can materially mislead inference about both short- and long-run effects.

My third contribution is that I develop a Dynamic Bias Correction (DBC) estimator that delivers consistent treatment-effect estimates, removing both dynamic bias and Nickell bias. My estimator can be used to obtain unbiased estimates of both the short and long-run treatment effects. DBC follows the analytical bias-correction tradition of Kiviet [1995]: given the model, it derives a closed-form bias term induced by fixed-effects demeaning and subtracts its estimate. Unlike existing analytic corrections, DBC accommodates endogenous treatment driven by past outcomes and allows heterogeneous treatment effects.[6] Allowing for endogenous treatment is important in many economic settings because selection into treatment is often a function of past outcomes. DBC neither requires instrumental variables nor depends on a long panel; it remains valid with few time periods. Alternative Nickell-bias corrections IV/GMM that instrument with lagged values of the regressors can face weak-instrument and variance-inflation issues. In contrast, DBC yields consistent

[3]The p-value for the GDP growth result is .06, so it is significant at the 10% level while the GDP level result is significant at the 5% level.

[4]Both GDP growth and levels are used as outcomes in the literature that studies the effect of temperature on economic outcomes [Newell et al., 2021, Nath et al., 2024].

[5]Short term treatment effect of treatment $D_{i,t}$ on outcome $Y_{i,t}$ is $\tau_0 = \mathbb{E}\left[\frac{\partial Y_{i,t}(D_{i,t})}{\partial D_{i,t}}\right]$. The long term effect is $\frac{\tau_0}{1-\rho_0}$ where ρ_0 is the autocorrelation in outcome. Note, this long-run treatment effect does not take into account adaptation [Mérel and Gammans, 2021]. Details are given in Appendix I.

[6]A large econometrics literature highlights how important it is to allow treatment heterogeneity in panel data [Sun and Abraham, 2020], Callaway and Sant'Anna [2021], Goodman-Bacon [2021], Chaisemartin and d'Haultfoeuille [2024].

estimates with OLS-like standard errors in simulations and in our application.

Related Work.— The introduction shows two main results. First, static fixed-effects estimators are biased in dynamic settings because of the dynamic bias introduced in this paper. Second, the DBC estimator removes this bias without using instruments. This section places the contribution within four strands of related work.

First, I discuss how the dynamic bias introduced here compares to previously studied biases of dynamic FE models, including Nickell bias [Griliches, 1967, Nickell, 1981, Kiviet, 1995]. I study the empirically common *static* specification that omits lagged outcomes and target the treatment effect as the parameter of interest. By contrast, the prior literature primarily analyzes dynamic specifications that include lagged outcomes and targets the lagged-outcome coefficient rather than a treatment effect. By focusing on the static model, my contribution is to characterize the resulting dynamic bias on the treatment coefficient—a phenomenon distinct from Nickell bias—and to show that, under empirically relevant conditions, this bias can be substantially larger than Nickell bias. To my knowledge, prior work has not provided a formal analysis of the treatment-effect coefficient when lagged outcomes are excluded from OLS with fixed effects. My treatment of the lag outcome as a nuisance parameter and quantifying the induced bias on the treatment effect is new.

I next situate the DBC correction within the fixed-T panel literature on bias correction for dynamic fixed-effects models. This literature offers two main approaches: IV and analytical corrections. In fixed-T panels, Nickell bias is addressed either by IV/GMM estimators [Holtz-Eakin et al., 1988, Arellano and Bond, 1991], which rely on deeper lags and are vulnerable to weak/many-instrument issues and instrument choice [Mikusheva and Sun, 2024], or by analytical bias corrections that subtract finite-sample bias directly [Kiviet, 1995, Juodis et al., 2015, Breitung et al., 2022, Basso et al., 2022]. Analytical bias corrections avoid instruments but require full specification of the outcome process: the correction is model-specific and must be re-derived for each data-generating process. My strategy corrects dynamic bias by including lagged outcomes and then applying a fixed-T analytical correction. Relative to existing analytical methods, my contribution is to accommodate endogenous treatments and heterogeneous treatment effects (via interactions), thereby removing both dynamic and Nickell bias. Because my object of interest is a treatment effect, I explicitly

model the treatment equation in order to compute the analytical correction. A separate line of work uses jackknife remedies [Hahn and Kuersteiner, 2002, Dhaene and Jochmans, 2015]. These are simple to implement but typically inflate standard errors and require longer panels, as they rely on large-T asymptotics [Fernández-Val and Weidner, 2016].

As for other econometric methods used in the literature, applied work often tries to proxy dynamics by using alternatives to controlling for lagging outcomes. Common alternatives include using (i) unit-specific time trends or (ii) factor-model methods e.g., synthetic controls, synthetic differences-in-differences [Abadie et al., 2010, Arkhangelsky et al., 2021], or (iii) transforming outcomes (differences/growth) to reduce persistence. However, none of these control *within-unit dynamics*. Unit trends address general time-specific patterns for each unit, but not autocorrelation - i.e., they cannot capture the effect of lagged outcomes on current outcomes.[7] Factor model based approaches rely on low-rank factor assumptions that do not accommodate unit-specific dynamics.[8] Finally, transformations such as differencing introduce an additional bias -"transformation bias" - which I define and analyze in the paper. Thus, to purge dynamic bias one must explicitly include lagged outcomes (and then correct the resulting Nickell bias); proxies alone do not suffice.[9]

Relatedly, Bonhomme [2025] surveys panel methods under feedback / sequential exogeneity, where past outcomes can affect future covariates or treatments. My focus is complementary: I show that even when treatment assignment is completely randomly assigned, static FE treatment-effect estimates are biased when outcomes are dynamic and lagged outcomes are omitted; the mechanism operates through fixed-effects estimation and the within transformation.[10]

The rest of the paper proceeds as follows. I provide empirical motivation and an overview of the biases discussed in this paper in Section 2. Section 3 presents theoretical results. Section 4 introduces the DBC estimator of treatment effects. Section 5 conducts a simulation study to illustrate the asymptotic properties of the estimators I propose. Section 6 provides an empirical

[7] This is also seen empirically in the applied example in Section 6.

[8] In the context of panel data, a low-rank factor model aims to capture cross-sectional correlations among different units (e.g., individuals, firms, or countries) by assuming that these correlations can be explained by a limited number of common factors. Importantly, such models focus primarily on capturing variation between units rather than time-series dynamics within each unit.

[9] For examples using trends and factor approaches, see Annan and Schlenker [2015], Damm et al. [2024]; on tests for outcome dynamics, see Chamberlain [1982].

[10] I use the term dynamic bias to refer to bias induced by dynamic outcome feedback when lagged outcomes are omitted; this differs from biases studied in the feedback/sequential-exogeneity literature.

example illustrating how correcting for dynamic bias impacts treatment effect estimation. Section 7 concludes.

2 Intuition: Mechanism and Magnitude

Before presenting the theoretical results of the paper, I provide a preview of the results to give intuition for dynamic bias in a simple empirical setting. I compare dynamic bias and Nickell bias both when treatment is randomly assigned and when it is endogenous (related to past outcomes). I then discuss how commonly used transformations of the outcome variable interact with these biases.

2.1 Randomly Assigned Treatment

Consider estimating the treatment effect of country temperature on country GDP using a panel of countries observed annually when temperature is taken to be randomly assigned conditional on country fixed effects. A widely used specification for treatment effect estimation is the static fixed-effects regression given in Equation (1).[11]

$$\text{Static Model:} \quad \text{GDP}_{i,t} = a_i + \tau_0 \, \text{Temp}_{i,t} + e_{i,t}. \tag{1}$$

A practical question is whether one should include lagged outcomes in this regression when estimating treatment effects?

To fix ideas, define a simple true model (Eq. (2)) with randomly assigned treatment that we take as ground truth. Suppose treatment evolves as in Dell et al. [2012]: conditional on country, annual temperature realizations are as good as random.[12] Additionally, the true model includes past GDP, as Solow [1956] explains that past GDP affects current GDP.[13]:

$$\text{True Model with Random Treatment:} \quad \text{GDP}_{i,t} = a_i + \tau_0 \, \text{Temp}_{i,t} + \rho_{10} \, \text{GDP}_{i,t-1} + \varepsilon_{i,t}. \tag{2}$$

[11] The framework accommodates time fixed effects and observed covariates; these are omitted here for expositional clarity.

[12] $\text{Temp}_{i,t} = c_i + u_{i,t}$ with a country effect c_i and an i.i.d. shock $u_{i,t}$.

[13] One way that past GDP affects current GDP through its impact on capital accumulation. Higher GDP in the past implies higher savings and investment, leading to a larger capital stock in the current period. Since the capital stock is an input in the production function, a larger capital stock results in higher current GDP.

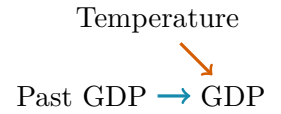

Figure 1: Causal Diagram for Dell et al. [2012]

The diagram in Figure 1 summarizes this environment: past GDP affects current GDP (dynamics), temperature affects current GDP (the treatment effect), and temperature is conditionally randomly assigned given country fixed-effects.

2.1.1 Models

Throughout, τ_0 denotes the short run causal effect in the *true* model (Eq. (2)) that we take as ground truth. In this section, I explain two models that could be used to estimate the treatment effect τ_0. Researchers often estimate these models using OLS. A necessary condition for OLS to be unbiased is that treatment in time period t, $\text{Temp}_{i,t}$, is uncorrelated with the model error $e_{i,t}$ in time period t.

The first model is the Dynamic Model in which the past outcome is included as a regressor.

$$\text{Dynamic Model:} \quad \text{GDP}_{i,t} = a_i + \tau_0 \,\text{Temp}_{i,t} + \rho_{10} \,\text{GDP}_{i,t-1} + \varepsilon_{i,t}. \tag{3}$$

Here in this model $\varepsilon_{i,t}$ simply is the structural error from the true model, so by construction we know that it is uncorrelated with treatment.

The second model is the Static Model in which the past outcome is not included as a regressor.

$$\text{Static Model:} \quad \text{GDP}_{i,t} = a_i + \tau_0 \,\text{Temp}_{i,t} + \underbrace{e_{i,t}}_{\rho_{10}\text{GDP}_{i,t-1}+\varepsilon_{i,t}}. \tag{4}$$

In the Static Model, researchers do not control for $\text{GDP}_{i,t-1}$. However, $\text{GDP}_{i,t-1}$ is part of the true model given Equation 2, therefore the $\text{GDP}_{i,t-1}$ term appears in the new model error along with the original error. Therefore, $e_{i,t} := \rho_{10}\text{GDP}_{i,t-1} + \varepsilon_{i,t}$. Still, $\text{Temp}_{i,t}$ remains uncorrelated with the error term $e_{i,t}$. This is because the temperature is randomly assigned, so it is not related to previous outcomes $\text{GDP}_{i,t-1}$ or model errors $\varepsilon_{i,t}$ that appear in $e_{i,t}$.

2.1.2 Cause of Bias

Researchers frequently adopt these models because they see $\text{Temp}_{i,t}$ as uncorrelated with the time-t model error. However, researchers still have to estimate fixed effects a_i by either using dummies, the within transformation, or first differences.[14] These transformations create new variables that are functions of data in multiple time periods. For example, consider the within transformation,

$$\widetilde{\text{Temp}}_{i,t} = \text{Temp}_{i,t} - \frac{1}{T}\sum_{s=1}^{T} \text{Temp}_{i,s}. \tag{5}$$

This new treatment variable $\widetilde{\text{Temp}}_{i,t}$ is now a function of $\text{Temp}_{i,s}$ in all time periods. Because of this, OLS requires *strict exogeneity* for unbiasedness: not only must contemporaneous treatment $\text{Temp}_{i,t}$ be uncorrelated with the period-t error, but the regressors at *all* dates must be uncorrelated with the error at every t. Since the new variable $\widetilde{\text{Temp}}$ is now a function of temperature in all periods, both models—the Static and Dynamic Model—lead to biased treatment effect estimates.

The fixed effect estimation turns past outcomes into generated confounders if they are correlated with *either* the outcome or the treatment.[15] This manifests in two ways:

1. In the Dynamic Model, treatment in all time periods is uncorrelated with model error because the past outcome, $\text{GDP}_{i,t}$, is controlled for. However, $\text{GDP}_{i,t}$ itself in other time periods is correlated with model error, leading to the well-studied **Nickell bias**. Then because the coefficient on $\text{GDP}_{i,t-1}$ is estimated with bias, this spills over and biases the treatment effect estimate. Nickell bias is discussed in detail in Section 4.1.

2. In the Static Model, treatment in other time periods is correlated with model error. Specifically, $\text{Temp}_{i,t-1}$ is correlated with $e_{i,t}$, which leads to biased treatment effect estimates. I term this **dynamic bias**, which is discussed in detail in Section 3.3.

Variation on dynamic bias: Before showing simulation results, I introduce a variation on the Static Model, which I call the Delta Model, often implemented in applied work that also leads to dynamic bias.

[14] Running a regression with within-transformed data leads to same coefficients as running a regression with unit dummies [Wooldridge, 2010]. Running first-differences also generates bias of a similar form.

[15] In classic cross sectional causal inference we think of variables as confounders if they are correlated with *both* the outcome and treatment.

$$\text{Delta Model:}^{16} \quad \underbrace{\Delta \text{GDP}_{i,t}}_{\text{GDP}_{i,t}-\text{GDP}_{i,t-1}} = a_i + \tau_0 \, \text{Temp}_{i,t} + \underbrace{\eta_{i,t}}_{(\rho_{10}-1)\,\text{GDP}_{i,t-1}+\varepsilon_{i,t}} . \tag{6}$$

Some researchers suspect that highly persistent outcomes (ρ_{10} close to 1) can cause problems with their analysis, so instead they study transformations of their outcomes, such as differences or growth. When taking the difference on the left-hand side, this imposes a ρ_{10} coefficient of 1, since $1 \times \text{GDP}_{i,t-1}$ is subtracted from the model. The difference between 1 and true ρ_{10} remains in the error of the model and therefore $\eta_{i,t} := (\rho_{10} - 1)\text{GDP}_{i,t-1} + \varepsilon_{i,t}$.[17] A type of dynamic bias occurs in this model because like in the Static Model, part of the outcome remains in the error term. It is the case that $\text{Temp}_{i,t-1}$ is correlated with $\eta_{i,t}$, discussed in detail in Appendix F.2, also leading to bias. I call this type of dynamic bias **transformation bias**.

2.1.3 Simulation

To complement the intuition above, I run a targeted Monte Carlo that varies two parameters — outcome persistence and panel length — and compare the resulting bias of the Dynamic, Static, and Delta estimators. All three specifications are biased, but the magnitude differs sharply.

Figure 2 reports a Monte Carlo based on the true model in Eq. (2) with randomly assigned treatment. I set the true effect to $\tau_0 = 0.5$ and simulate panels with $N = 1000$ units, varying the number of time periods T (x–axis). Each panel corresponds to a different persistence parameter, $\rho_{10} \in \{0.2, 0.5, 0.9\}$. For each design I estimate the three models by OLS and plot the resulting $\hat{\tau}$ (y–axis).

First, a practical lesson is that the bias decreases as there are more time periods. This means shorter panels exacerbate bias for all three approaches. This matters in applications that shorten panels - for example, when using long differences to differentiate temperature from climate change.[18] While longer differences can reduce serial correlation in outcomes (and thus some dynamic bias), they also reduce T, which tends to worsen bias overall.

[16] Since $\text{Temp}_{i,t}$ only impacts GDP_{it} and not $\text{GDP}_{i,t-1}$, the effect of $\text{Temp}_{i,t}$ on the transformed outcome, $\Delta\text{GDP}_{i,t}$, is the same the effect of $\text{Temp}_{i,t}$ on the untransformed outcome $\text{GDP}_{i,t}$.

[17] Note here that only the outcome ($GDP_{i,t}$) is being transformed through differences - the variables on the right-hand side are not being differenced - therefore this transformation is not equivalent to the first-differences transformation.

[18] See, e.g., Nordhaus [2006], Deryugina and Hsiang [2014], Burke et al. [2015].

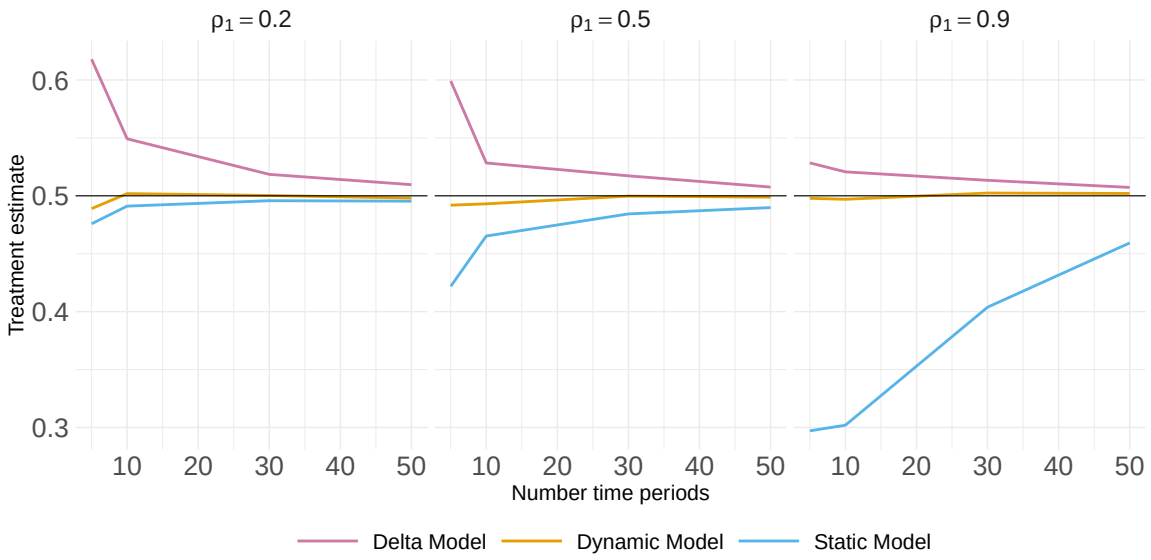

Figure 2: Bias of three different models.

Another key takeaway is the Dynamic Model (Eq. (3)) always delivers the smallest bias out of all the estimators - this is true for all ρ_{10} and all T. The intuition for this is that only the Dynamic Model explicitly controls for past outcome, and so only in this model does treatment remain strictly exogenous. Second, the ranking of Static vs. Delta depends on the value of ρ_{10}. The Delta Model effectively subtracts last period's outcome with a fixed coefficient of 1 (i.e., it "hard-codes" $Y_{i,t-1}$ with weight 1 because $\Delta Y_{i,t} = Y_{i,t} - 1 \times Y_{i,t-1}$). If $\rho_{10} \approx 1$ (right panel), this closely replicates the true omitted lag and therefore reduces the bias of the Delta Model. If ρ_{10} is small (left panel), imposing a unit coefficient on the lag through differencing in the Delta Model over-corrects and increases bias relative to the Static Model.

2.2 Endogenous Treatment

Treatment effect estimates are even more biased when treatment is not randomly assigned but instead related to past outcomes - that is, when treatment is endogenous. Many economic environments generate treatment selection on lagged outcomes which link treatment and past outcomes [Marx et al., 2022, Ghanem et al., 2022]. Examples include policy targeting based on prior conditions (air pollution [Chay and Greenstone, 2005]; Brazil deforestation [Harding et al., 2021, Assunção et al., 2023]), insurance enrollment shaped by prior utilization [Kowalski, 2023], drought relief tied to past outcomes [Tarquinio, 2022], and firm choices linked to past productivity [Olley

and Pakes, 1992].

With endogenous treatment, omitting $Y_{i,t-1}$ induces a standard omitted-variable bias that does not vanish as T grows. Thus, the Static and Delta Model specifications never converge to the causal effect. Our simulations (Appendix F.4) show that even modest endogeneity yields substantial bias.

Therefore, the remedy is to control for past outcomes and address the resulting Nickell bias. In empirically-calibrated simulations, the Nickell component is small relative to the omitted-lag bias from failing to control for $Y_{i,t-1}$.

3 Theoretical Results

Dynamic bias arises when there are dynamics in outcomes but researchers do not control for past outcomes when estimating treatment effects. Dynamic bias is explained and characterized in Section (3.3). Correcting dynamic bias requires controlling for past outcomes. Estimators that include past outcomes as controls still suffer from Nickell bias, which is discussed in Section (4.1). Section (4.2) provides the dynamic bias-corrected (DBC) estimator which eliminates Nickell bias.

3.1 Notation

I use $Y_{i,t}$ to denote the outcome, $Y_{i,t-h}$ to denote the h^{th} lag of the outcome, $D_{i,t}$ to denote treatment, and $W_{i,t}$ to denote other covariates. I use the superscript t to denote the vector of variables up to time period t, for example $D_i^t := (D_{i,1}, D_{i,2}, \cdots, D_{i,t})$. The concatenation of all possible regressors is given by $X_i^t := (D_i^t, W_i^t, Y_i^{t-1})$.

Note that the covariates $W_{i,t}$ can include time period dummies. That is for each time period t I can include a dummy Q_t with 1 in the t^{th} position. This enables the model to include time effects.

I use $\perp\!\!\!\perp$ to denote independence, and $\perp$ for uncorrelated. I use bar notation for averages over time, for example $\bar{Y}_{i,-1} := \frac{1}{T}\sum_{t=1}^{T} Y_{i,t-1}$ and $\bar{Y}_i := \frac{1}{T}\sum_{t=1}^{T} Y_{i,t}$. I use tilde to denote the within transformation, for example,

$$\tilde{Y}_{i,t-1} := Y_{i,t-1} - \bar{Y}_{i,-1}. \tag{7}$$

3.2 Setup

I begin by introducing the setting under the potential outcomes framework (Splawa-Neyman [1923], Rubin [1974]). I observe outcomes $Y_{i,t}$ for a sample of units $i = 1, \ldots, N$ for time periods $t = 1, \ldots, T$. The number of time periods T is fixed, but the cross-sectional dimension N grows with more observations. Therefore, the formal asymptotic results are for the large N setting where $N \to \infty$ and T is fixed. This is also sometimes known as the "short panel".

For now, let us assume that the treatment $D_{i,t}$ is continuous. For each value d of the treatment $D_{i,t}$, unit i has a corresponding potential outcome $Y_{i,t}(D_{i,t} = d)$. In the example in Dell et al. [2012], $D_{i,t}$ is temperature and $Y_{i,t}$ is country GDP. The authors are interested in estimating the "contemporary causal effect of temperature on the development process".[19] Formally, the causal object of interest is average effect of marginally increasing temperature in the current time period ($D_{i,t}$) on GDP, holding all else fixed. This is also referred to the contemporary average partial derivative (APD) and is written as

$$\tau_0 = \mathbb{E}\left[\frac{\partial Y_{i,t}(D_{i,t})}{\partial D_{i,t}}\right], \tag{8}$$

where the expectation is over all units and time periods.

I impose structure on the potential outcomes to allow estimation. In the rest of this section, I showcase that dynamic bias is a problem even in a simple (though common) stylized model. I use the formula for the bias in this example to provide intuition. However, when I provide a bias correction in Section 4 I introduce a more general model (Section 4.3.1), to which the DBC correction also applies.

Assumption 1. *(Stylized Example One: Random Treatment.) Let the true vector of parameters be given by $\theta_0 := (\rho_{10}, \tau_0)$. The errors $\varepsilon_{i,t}$ are i.i.d mean zero with variance $\sigma^2_{\varepsilon,i}$ and the errors $u_{i,t}$ are i.i.d mean zero with variance $\sigma^2_{u,i}$. The true underlying has the following structure:*

$$Y_{i,t}(D_{i,t}) = a_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}, \tag{9}$$

[19] One may also be interested in the estimation of the long term effect of treatment, as studied in Hausman [1985], as opposed to the contemporary. I focus on the contemporary effect in this paper as it is the most common effect of interest in a set of surveyed applied papers.

$$D_{i,t} = c_i + u_{i,t}. \tag{10}$$

This simple example is based off the setting of Dell et al. [2012]. I include $Y_{i,t-1}$ in the outcome model because economic theory tells us that past GDP ($Y_{i,t-1}$) impacts current GDP ($Y_{i,t}$) [Solow, 1956]. As is standard in panel data, I impose additive fixed effects a_i, often called individual fixed effects, which can be correlated with covariates. In the treatment model in Equation (10),treatment $D_{i,t}$ is also a function of individual fixed effects c_i.[20] This structure mimics Dell et al. [2012], who write that treatment is correlated with country fixed effects a_i. The presence of fixed effects in the true structural model generates the need to control for fixed effects in treatment effect estimation.

However, in the data we don't observe the fixed effects. Further, they cannot be estimated consistently in short panels, which is known as the incidental parameter problem [Neyman and Scott, 1948]. I bypass the estimation of fixed effects by using within-estimation, which leads to the same estimates of $\theta = (\rho_{10}, \tau_0)$ as would explicitly estimating fixed effects by including fixed effect dummies for each unit.

Assumption 2. *(Weak Exogeneity.)*

$$\mathbb{E}[\varepsilon_{i,t}|X_i^t, a_i] = 0, \quad \mathbb{E}[u_{i,t}|X_i^t, a_i] = 0, \quad X_i^t := (X_{i,1}, \ldots, X_{i,t-1}, X_{i,t}). \tag{11}$$

The weak exogeneity assumption says the that error has zero conditional expectation given current and past covariates. Because the structural outcome model allows for lagged outcomes to impact current outcomes, I cannot make the more familiar strict exogeneity assumption, which would require that the regressors are uncorrelated with the error term across all time periods, meaning that current, past, and future values of the regressors do not influence the current error term.

Assumption 3. *(Independence across i) The data vectors $Z_i = \{(Y_{i,t}, X_{i,t}')'\}_{t=1}^T$ are i.i.d. across units i.*

[20] It is important to note dynamic bias remains even if treatment is not a function of fixed effects and $D_{i,t} = u_{i,t}$.

3.3 Dynamic Bias

Given the model laid out above, I now introduce dynamic bias and show how it impacts treatment effect estimates. Dynamic bias arises when there is a relationship between past outcomes and current outcomes in the true model but past outcomes are not included in the estimation model. A crucial feature of dynamic bias is that it arises even when the treatment is random.

3.3.1 Causal Object

Given the simple stylized model presented in Assumption (1), in particular the homogeneous treatment assumption in Equation 9, the treatment object of interest is simply τ_0. This follows using our definition of the APD from Equation (8) along with potential outcome given in Equation 9.

$$\mathbb{E}\left[\frac{\partial Y_{i,t}(D_{i,t})}{\partial D_{i,t}}\right] = \mathbb{E}\left[\frac{\partial\big(a_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}\big)}{\partial D_{i,t}}\right] = \tau_0. \tag{12}$$

In the context of Dell et al. [2012], this represents the contemporaneous causal effect of temperature on the development process. This is the *short-run (contemporaneous) effect*; in a dynamic setting it differs from the long-run effect of a permanent change (equal to $\tau_0/(1-\rho_{10})$ in this model).

3.3.2 Estimation

In settings where treatment assignment is random, applied researchers often rely on a static model that does not account for past outcomes to estimate τ_0. I write the static model in Equation (13).

Model 1: Static Model

$$Y_{i,t} = a_i + \tau^D D_{i,t} + e_{i,t}, \tag{13}$$

Given the true model presented in Assumption 1, the error in the static model is given by $e_{i,t} := \rho_{10} \mathrm{Y}_{i,t-1} + \varepsilon_{i,t}$. Since applied researchers do not observe fixed effects a_i, they can not use OLS to estimate equation (13) directly. Instead they have to estimate the fixed effects with dummy variables, or by using the within or first difference transformation. I remove the fixed effects by using the within-transformation, which recall was defined in Equation (7).

Model 2: Within-Transformed Static Model

$$\tilde{Y}_{i,t} = \tau^D \tilde{D}_{i,t} + \tilde{e}_{i,t}. \tag{14}$$

OLS Estimator. To estimate the treatment effect, OLS is used to estimate Equation (14). I call the coefficient estimated this way $\hat{\tau}^D$. Unfortunately $\hat{\tau}^D$ is inconsistent—no matter how large the number of units grows, the estimator does not converge to the true parameter value.

Theorem 3.1. *(Dynamic Bias.) Under Assumptions* (1), (2), (3), *and that* $Var(D_{i,t}) > 0$, *running OLS to estimate the static model in Equation* (14) *leads to a biased treatment effect estimator* $\hat{\tau}^D$.

$$\operatorname*{plim}_{N\to\infty} \hat{\tau}^D = \tau_0 - \frac{\rho_{10}\tau_0}{T(T-1)}\left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)}\right]. \tag{15}$$

The proof of Theorem (3.1) is provided in Appendix A. This bias is what I term dynamic bias. This bias is of order $1/T$, and it diminishes as the number of time period increases. An analytical comparison between dynamic bias and Nickell bias is given in Appendix J. There I show under 1 dynamic bias in the treatment-effect coefficient decays more slowly than the Nickell bias in the treatment-effect coefficient.

3.3.3 Dynamic Bias Intuition

Dynamic bias arises due to the estimation of fixed effects, which are typically estimated by including dummies, applying within-transformation, or first-differencing the data. This estimation *generates* confounding if past outcomes are not controlled for. In classic cross-sectional causal inference, a covariate is a confounder only if it is correlated with *both* the treatment and the outcomes. This reasoning explains why many researchers opt for static panel regressions; in their contexts, treatment is random, so although past outcomes affect current outcomes, they are not correlated with the treatment and, therefore, are not classic confounders [Angrist and Pischke, 2009]. Due to fixed effects estimation, a past outcome is a generated confounder and generates bias if it is correlated with *either* the treatment or the outcome.

Estimation of models with fixed effects requires strict exogeneity for unbiased estimation. When

treatment is random, and the past outcome is controlled for in the model, then treatment is a strictly exogenous regressor. However, when the past outcome is not controlled for, the past outcome is part of the model error, and treatment is no longer strictly exogenous. This leads to biased treatment effects.

To see algebraic intuition for why strict exogeneity does not hold, consider the within-transformation approach for estimating fixed effects. After applying the within transformation, the newly transformed variables become functions of data from all time periods. Consequently, the errors $\tilde{e}_{i,t}$ are functions of errors $e_{i,t}$ across all time periods, causing the within-transformed error to be correlated with the treatment, which in turn leads to endogeneity.

1. $\widetilde{e}_{i,t}$ is a function of $e_{i,t+1}$.

   (Because of the definition of the within transformation: $\widetilde{e}_{i,t} = e_{i,t} - \frac{1}{T}\sum_{s=1}^{T} e_{i,s}$.)

2. $e_{i,t+1}$ is a function of $\mathrm{Y}_{i,t}$.

   (Because $e_{i,t+1} := \rho_{10}\mathrm{Y}_{i,t} + \varepsilon_{i,t+1}$ by definition of the Static Model error.)

3. $\mathrm{Y}_{i,t}$ is a function of $\mathrm{D}_{i,t}$

   (Because $\mathrm{Y}_{i,t} = a_i + \tau_0\mathrm{D}_{i,t} + \rho_{10}\mathrm{Y}_{i,t-1} + \varepsilon_{i,t}$ by definition of the true model in Equation (9).)

Therefore $\widetilde{e}_{i,t}$ is a function of $D_{i,t}$.

However, if past outcomes had been included in the model, the treatment would have been a strictly exogenous regressor. Dynamic bias occurs when the outcome lag is not considered, leading to potentially large biases in treatment effect estimates. This dynamic bias can be mitigated by including past outcomes in the regression model.

However, even when past outcomes are included, the well-known Nickell bias remains. In the next section, I will discuss Nickell bias and then demonstrate in Section 5 that, in simulations, Nickell bias is much smaller than dynamic bias. Regardless of the values of ρ_{10} or τ_0, I show that in simulation that dynamic bias is consistently larger than Nickell bias.

4 Debiased Estimator (DBC)

The DBC estimator works both when treatment is randomly assigned and when treatment is a function of past outcomes and is therefore endogenous.

To explain how the bias correction works, I assume the following simple data-generating processes for the remainder of the section. However, the results of the paper hold for a richer class of models.[21] Equation (17) shows that the treatment is a function of the past outcome. Recall biases get worse if, in addition to past outcome not being controlled for, treatment is related to past outcomes in the true model (not randomly assigned). This is because in addition to dynamic bias there is now omitted variable bias. Equation (16) shows that outcomes are a function of the past outcome and treatment. This simple example highlights the causal parameter of interest, inferential goal, and key assumptions of this approach.

Assumption 4. *(Stylized Example Two: Endogenous Treatment.) The true underlying data generating has the following structure:*

$$Y_{i,t}(D_{i,t}) = a_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}, \tag{16}$$

$$D_{i,t} = c_i + \rho_{20} Y_{i,t-1} + u_{i,t}, \tag{17}$$

where as before $\varepsilon_{i,t}$ and $u_{i,t}$ are i.i.d mean zero true errors with variance $\sigma^2_{i,\varepsilon}$ and $\sigma^2_{i,u}$.

4.1 Dynamic Model

Nickell bias occurs when a regression includes both estimated fixed effects and past outcome variables as regressors. Since fixed effects are not observed, they have to be estimated with dummies, the within-transformation, or first differences. These transformations impact all parts of the model, including the error terms. Taking the within transform as an example, the transformed error term becomes a function of error terms in all time periods. Therefore, if a regressor is a function of past error terms (like lagged outcome variables), it becomes correlated with the new within-transformed errors.

[21] The richer class of models are given in Section 4.3.1 below.

Model 3: Within-Transformed Dynamic Model

$$\tilde{Y}_{i,t} = \tau_0 \tilde{D}_{i,t} + \rho_{10} \tilde{Y}_{i,t-1} + \tilde{\varepsilon}_{i,t} \tag{18}$$

$$\tilde{D}_{i,t} = \rho_{20} \tilde{Y}_{i,t-1} + \tilde{u}_{i,t} \tag{19}$$

OLS Estimator. One could estimate treatment effects with a dynamic model by using OLS to estimate Equation (18). This OLS regression leads to estimates of $\hat{\tau}^{NB}$ and $\hat{\rho}_1^{NB}$ that have Nickell bias. Though less common in practice, researchers could also use OLS to estimate Equation (19) to get the estimate $\hat{\rho}_2^{NB}$. Our vector of estimated OLS parameters is $\hat{\theta}^{NB} := (\hat{\rho}_1^{NB}, \hat{\tau}^{NB}, \hat{\rho}_2^{NB})$. The estimated residuals for a given parameter vector $\hat{\theta}$ are given by $\tilde{\varepsilon}_{i,t}(\hat{\theta}) := \tilde{Y}_{i,t} - \hat{\tau}\tilde{D}_{i,t} - \hat{\rho}_1 \tilde{Y}_{i,t-1}$ and $\tilde{u}_{i,t}(\hat{\theta}) := \tilde{D}_{i,t} - \hat{\rho}_2 \tilde{Y}_{i,t-1}$. By definition we have that $\tilde{\varepsilon}_{i,t}(\theta_0) = \tilde{\varepsilon}_{i,t}$ and $\tilde{u}_{i,t}(\theta_0) = \tilde{u}_{i,t}$.

To see concretely how the Nickell bias arises, focus on the OLS estimation of Equation (18). I use the familiar formula for our OLS coefficients $\left(\hat{\theta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}\right)$ plugging in our stack of covariates, a $(N \cdot N_T) \times 2$ matrix $\mathbb{X}$ whose row i,t is $\mathbb{X}_{i,t} = [\tilde{Y}_{i,t-1} \quad \tilde{D}_{i,t}]'$, and the outcomes given in vector $\mathbb{Y}$ where the i,t-th element is $\tilde{Y}_{i,t}$.

$$\begin{bmatrix} \hat{\rho}_1^{NB} \\ \hat{\tau}^{NB} \end{bmatrix} = \left(\mathbb{X}'\mathbb{X}\right)^{-1}\left(\mathbb{X}'\mathbb{Y}\right) \tag{20a}$$

$$= \underbrace{\begin{bmatrix} \rho_{10} \\ \tau_0 \end{bmatrix}}_{20.1} + \underbrace{\left(\frac{1}{N}\mathbb{X}'\mathbb{X}\right)^{-1}}_{20.2} \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \begin{bmatrix} \tilde{Y}_{i,t-1} \\ \tilde{D}_{i,t} \end{bmatrix} \begin{bmatrix} \tilde{\varepsilon}_{i,t} \end{bmatrix}\right)}_{20.3} \tag{20b}$$

If the errors were strictly exogenous the expectation of term 20.3 would be zero, and the OLS estimator would be unbiased. This would imply that $\mathbb{E}[\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}] = 0$ and $\mathbb{E}[\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}] = 0$. However, given the model in Assumption (4) the expectation of term 20.3 is not zero. This is because $\mathbb{E}[\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}] \neq 0$ and $\mathbb{E}[\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}] \neq 0$. As an example, let us focus on why $\mathbb{E}[\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}] \neq 0$. It follows that $\tilde{Y}_{i,t-1}$ and $\tilde{\varepsilon}_{i,t}$ are correlated by observing that they are both correlated with $\varepsilon_{i,t-1}$. First, $\tilde{Y}_{i,t-1}$ is a function of $Y_{i,t-1}$, which itself is a function of $\varepsilon_{i,t-1}$. Second, $\tilde{\varepsilon}_{i,t} = \varepsilon_{i,t} - \bar{\varepsilon}_{i,t}$ and $\bar{\varepsilon}_{i,t} = \frac{1}{T}(e_{i,0} + \cdots \varepsilon_{i,t-1} + \cdots + \varepsilon_{i,T})$. Similar logic is used to understand why $\mathbb{E}[\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}] \neq 0$.

4.2 Bias Correction

4.2.1 OLS Intuition

Though my de-biased DBC estimator uses a GMM framework, I first build an understanding for how the bias correction works using OLS. Focusing on the OLS estimates in Equation (20) , I calculate an expression for the asymptotic OLS estimator bias (term 20.2 + 20.3). The bias correction is created by calculating an estimate of this bias term, and the de-biased estimator by re-centering the GMM moment condition.

Rewriting Equation (20), I get an expression of the asymptotic bias of the OLS estimator.

$$\begin{aligned} \text{Bias} &= \operatorname*{plim}_{N\to\infty} \begin{bmatrix} \hat{\rho}_1^{NB} \\ \hat{\tau}^{NB} \end{bmatrix} - \begin{bmatrix} \rho_{10} \\ \tau_0 \end{bmatrix} \\ &= \operatorname*{plim}_{N\to\infty} \Bigg[\underbrace{\left(\frac{1}{N} \mathbb{X}'\mathbb{X} \right)^{-1}}_{20.2} \Bigg] \operatorname*{plim}_{N\to\infty} \Bigg[\underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} \tilde{Y}_{i,t-1} \\ \tilde{D}_{i,t} \end{bmatrix} \begin{bmatrix} \tilde{\varepsilon}_{i,t} \end{bmatrix} \right)}_{20.3} \Bigg] \end{aligned} \tag{21}$$

The bias correction comes from solving for [20.3] and subtracting the estimated bias out. Therefore the bias-corrected estimate $\hat{\theta}^{DBC} = \hat{\theta}^{NB} - \hat{\text{Bias}}$. Since I am interested in $\theta_0 = (\rho_{10}, \tau_0, \rho_{20})$, not just ρ_{10} and τ_0, I need to estimate both Equations (18) and (19). In order to estimate both equation simultaneously, I cast the problem as a GMM system.

4.2.2 GMM Bias Correction

The original moment conditions, which contain bias, for estimating the parameters in Equations (18) and (19) are given in Equation (22). I index the moment conditions by iT to make it clear that they are functions of the data and that the number of time periods T is fixed.

$$\mathbf{m}_{iT}(\theta) = \begin{bmatrix} m_{\rho_1,iT}(\theta) \\ m_{\tau,iT}(\theta) \\ m_{\rho_2,iT}(\theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}(\theta) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta) \end{bmatrix} \tag{22}$$

Because of Nickell bias, the expectation of these moment equations are not zero at the true param-

eter θ_0. For each moment in $\mathbf{m}_{iT}(\theta)$, however, I can create a new de-biased moment by subtracting the mean of each original moment equation at θ_0. By construction, this new moment that is mean zero at the true parameters. This term that I subtract out is labeled $\mathbf{b}_{iT}(\theta_0)$ and is written

$$\mathbf{b}_{iT}(\theta_0) = \begin{bmatrix} b_{\rho_1,iT}(\theta_0) \\ b_{\tau,iT}(\theta_0) \\ b_{\rho_2,iT}(\theta_0) \end{bmatrix} = \mathbb{E}\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta_0) \end{bmatrix}. \tag{23}$$

The analytic bias correction comes from solving for $\mathbf{b}_{iT}(\theta_0)$.

Bias Correction Formula Intuition. A detailed calculation $\mathbf{b}_{iT}(\theta_0)$ is given in Appendix B.1. However, here, I give intuition for how the calculation works. As an example, consider $b_{\rho_1,iT}(\theta) = \mathbb{E}[\frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0)]$. To calculate this expectation, we need to derive how $\tilde{Y}_{i,t-1}$ relates to $\tilde{\varepsilon}_{i,t}$. To do this, it is helpful to unpack $\tilde{Y}_{i,t-1}$ by looking just at $Y_{i,t-1}$ and how it depends on $\varepsilon_{i,t}$ terms in all time periods. I use the outcome model in Assumption 4 to write $Y_{i,t-1}$ as a sum of past outcomes and errors by iteratively plugging in the definition of past outcomes and treatment.[22]:

$$\begin{aligned} Y_{i,t-1} = a_i \sum_{j=0}^{t-2}(\rho_{10}+\tau_0\rho_{20})^j + (\rho_{10}+\tau_0\rho_{20})^{t-2}Y_{i,0} + \tau_0\sum_{j=0}^{t-2}(\rho_1+\tau\rho_2)^j(c_i+u_{i,t-j-1}) \\ + \sum_{j=0}^{t-2}(\rho_{10}+\tau_0\rho_{20})^j\varepsilon_{i,t-j-1} \end{aligned} \tag{24}$$

Given equation (24), it is clear how $Y_{i,t-1}$ relates to $\varepsilon_{i,t}$ in all time periods. The dependence comes from the $\sum_{j=0}^{t-2}(\rho_1+\tau\rho_2)^j\varepsilon_{i,t-j-1}$ term. Using the properties of geometric sums and accounting for the fact that I am calculating the expectation of the within-transformed $\tilde{Y}_{i,t-1}$ and $\tilde{\varepsilon}_{i,t}$, I calculate the following bias correction terms.

Lemma 4.1. *(Bias Correction Terms.) Under Assumptions 2, 3, 4, 7 the expression for $b(\theta_0)$ is the following.*

$$b_{\rho_1,iT}(\theta_0) = \frac{-\sigma^2_{\varepsilon,i}}{T}\left(\frac{T-1}{1-\phi(\theta_0)} - \frac{\phi(\theta_0)-\phi(\theta_0)^T}{(1-\phi(\theta_0))^2}\right), \tag{25}$$

[22]This formula uses the fact that treatment is varying over time. If it is the case that treatment is "absorbing" (once a unit starts treatment they stay treated), a different formula applies, given in Appendix H.

$$b_{\tau,iT}(\theta_0) = \rho_{20} \times b_{\rho_1}(\theta_0), \tag{26}$$

$$b_{\rho_2,iT}(\theta_0) = \tau_0 \times \frac{-\sigma_{u,i}^2}{T^2}\left(\frac{T-1}{1-\phi(\theta_0)} - \frac{\phi(\theta_0)-\phi(\theta_0)^T}{(1-\phi(\theta_0))^2}\right). \tag{27}$$

Here $\phi(\theta) = (\rho_{10} + \tau_0\rho_{20})$. Proofs are given in Appendix B.1. The true variance parameters $\sigma_{\varepsilon,i}^2$ and $\sigma_{u,i}^2$ are not observed and have to be estimated. The estimated bias correction terms are the following.

Estimated Bias Correction Terms.

$$\hat{\sigma}_{\varepsilon,iT}^2(\theta) = \frac{1}{T-1}\sum_{t=1}^{T}\tilde{\varepsilon}_{i,t}(\theta)^2 \tag{28}$$

$$\hat{\sigma}_{u,iT}^2(\theta) = \frac{1}{T-1}\sum_{t=1}^{T}\tilde{u}_{i,t}(\theta)^2 \tag{29}$$

$$\hat{b}_{\rho_1,iT}(\theta) = \frac{-\hat{\sigma}_{\varepsilon,iT}(\theta)^2}{T}\left(\frac{T-1}{1-\phi(\theta)} - \frac{\phi(\theta)-\phi(\theta)^T}{(1-\phi(\theta))^2}\right) \tag{30}$$

$$\hat{b}_{\tau,iT}(\theta) = \rho_2 \times \hat{b}_{\rho_1,iT}(\theta). \tag{31}$$

$$\hat{b}_{\rho_2,iT}(\theta) = \tau \times \frac{-\hat{\sigma}_{u,iT}(\theta)^2}{T^2}\left(\frac{T-1}{1-\phi(\theta)} - \frac{\phi(\theta)-\phi(\theta)^T}{(1-\phi(\theta))^2}\right) \tag{32}$$

Here $\mathbb{E}[\hat{\sigma}_{\varepsilon,iT}^2(\theta_0)] = \sigma_{\varepsilon,iT}^2$ and $\mathbb{E}[\hat{\sigma}_{u,iT}^2(\theta_0)] = \sigma_{u,iT}^2$.

4.3 Bias-Corrected GMM Estimator

The bias-corrected moment equations are the original OLS moment equations minus the bias correction term:

$$\mathbf{m}_{iT}^{DBC}(\theta) := \mathbf{m}_{iT}(\theta) - \hat{\mathbf{b}}_{iT}(\theta). \tag{33}$$

Given the definition of the model, only at θ_0 do the residuals in the original moment equation equal the true noise variables in the bias correction term. Consequently, the moment conditions for the de-biased moment equations are satisfied:

$$\mathbb{E}\Big[\mathbf{m}_{iT}^{DBC}(\theta_0)\Big] = \mathbb{E}\begin{bmatrix} m_{\rho_1,iT}^{DBC}(\theta_0) \\ m_{\tau,iT}^{DBC}(\theta_0) \\ m_{\rho_2,iT}^{DBC}(\theta_0) \end{bmatrix} = \mathbb{E}\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0) - \hat{b}_{\rho_1,iT}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}(\theta_0) - \hat{b}_{\tau,iT}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta_0) - \hat{b}_{\rho_2,iT}(\theta_0) \end{bmatrix} = 0 \tag{34}$$

Bias-Corrected Estimator. The DBC estimator $\hat{\theta}_{DBC}$ is obtained by solving the GMM objective function under sample moment conditions:

$$\hat{\theta}^{DBC} = \underset{\theta}{\operatorname{argmin}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_{iT}^{DBC}\right)' \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_{iT}^{DBC}\right). \tag{35}$$

I now derive asymptotic properties. Since the proposed estimator is based on GMM, asymptotic normality is established following standard results from the GMM framework.

Assumption 5. *(Stationarity.) For some small $\delta_s > 0$*

$$|\rho_1 + \tau\rho_2| \leqslant 1 - \delta_s \tag{36}$$

This assumption ensures that the process is stationary, meaning that the statistical properties (such as the mean and variance) of the outcome are stable over time.

Assumption 6. *The parameter space Θ is compact, θ_0 is the unique solution to Equation 34, and θ_0 satisfies $E[\nabla_\theta \mathbf{m}^{DBC}(\theta_0)]$ having full column rank. Furthermore, θ_0 is in the interior of Θ.*

Assumption 7. *The noise $\varepsilon_{i,t}$ and $u_{i,t}$ are independent across i and t with $\mathbb{E}[\varepsilon_{i,t}] = 0$ and $\mathbb{E}[u_{i,t}] = 0$, and $\mathbb{E}[\varepsilon_{i,t}^2] = \sigma_{\varepsilon,i}^2 < C$ fand $\mathbb{E}[u_{i,t}^2] = \sigma_{u,i}^2 < C$. Also*

$$\begin{aligned} \max_i \mathbb{E}\big[\|\varepsilon_{i,t}\|^{4+\delta_\varepsilon}\big] < \infty \quad \forall t \text{ for some } \delta_\varepsilon > 0 \\ \max_i \mathbb{E}\big[\|u_{i,t}\|^{4+\delta_u}\big] < \infty \quad \forall t \text{ for some } \delta_u > 0. \end{aligned} \tag{37}$$

Errors are uncorrelated: $E[\varepsilon_{i,t}u_{i,t}] = 0$ for all i and t and $E[\varepsilon_{i,t}\varepsilon_{is}] = 0$ and $E[u_{i,t}u_{is}] = 0$ for $t \neq s$

and $E[\varepsilon_{i,t}u_{is}] = 0$. *Finally, the initial values* Y_{i0} *and* D_{i0} *satisfy* $\mathbb{E}[Y_{i0}^2] < \infty$ *and* $\mathbb{E}[D_{i0}^2] < \infty$ *for all* i.

Theorem 4.2. *Under Assumptions* (2) (3) (4), (5), (6), (7) *the limiting distribution as* T *remains fixed and as* $N \to \infty$ *of the estimator presented in Equation* (35) *is given by:*

$$\sqrt{N}(\hat{\theta}^{DBC} - \theta_0) \xrightarrow{d} \mathcal{N}(0, G^{-1}\Omega G'^{-1}), \tag{38}$$

with

$$\Omega = \operatorname*{plim}_{N\to\infty} \big[\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_{iT}^{DBC}(\theta_0)\mathbf{m}_{iT}'^{DBC}(\theta_0)\big], \tag{39}$$

and

$$G = \operatorname*{plim}_{N\to\infty} [\frac{1}{N}\sum_{i=1}^{N} \nabla_{\theta_0}\mathbf{m}_{iT}^{DBC}(\theta_0)] \tag{40}$$

The complete expressions for Ω, G and the proof are given in Appendix (C).

The DBC procedure delivers bias-corrected, consistent, and asymptotically normal estimates of all parameters, in particular (τ_0, ρ_{10}). Consequently, it provides estimators for both the short-run (contemporaneous) treatment effect $\hat{\tau}^{\text{DBC}}$, and the long-run effect of a permanent change. The long-run can be estimated by $\hat{\tau}^{\text{DBC}}/(1 - \hat{\rho}_1^{\text{DBC}})$, and inference follows from applying the delta method.

4.3.1 More Flexible Model

The bias correction formula for the more following more general model is given in Appendix B.2.

Assumption 8.

$$Y_{i,t} = a_i + \sum_{c=1}^{N_c} \tau_c(D_{i,t} \cdot W_{i,t}^c) + \beta_1 X_{1,i,t} + \rho_1 Y_{i,t-1} + \varepsilon_{i,t} \tag{41}$$

$$D_{i,t} = c_i + \rho_2 Y_{i,t-1} + \beta_2 X_{2,i,t} + u_{i,t} \tag{42}$$

Allowing the past outcomes $Y_{i,t-1}$ through $Y_{i,t-h}$ to impact current potential outcomes follow from Breitung et al. [2022] and general VAR structure follows from Juodis et al. [2015]. In an in-progress

extension to this paper, I have a machine learning estimator that still imposes additive fixed effects and errors, but allows for more flexible modeling of the regressors, such as the interaction terms in the model above. A discussion of this extension is given in Appendix G. Note that a general nonparametric form with a non-separable fixed effect is not possible, especially with fixed T.[23]

5 Simulation Study

In this section, I present a simulation study to compare the dynamic and Nickell bias along several dimensions. I provide additional details for the simulation that I introduced in Section 2 as well as additional results comparing my DBC estimator to other Nickell bias correction procedures.

5.1 Dynamic Bias Versus Nickell Bias

5.1.1 Simulation Design with Random Treatment

I start with a Monte Carlo simulation using the structural model presented in Assumption 1, which I reproduce here:

$$Y_{i,t} = a_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}. \tag{43}$$

$$D_{i,t} = a_i + u_{i,t}.^{24} \tag{44}$$

In Equation (43), the outcome is a function of the treatment and the past outcome. For the treatment model (Equation 44), I make treatment a function of the fixed effect, in line with the common practice in applied research of adding fixed effects to account for correlation between treatment and a time-invariant individual-level term.[25] This model is the simplest setting, where $D_{i,t}$ is not impacted by past $Y_{i,t-1}$. The errors $u_{i,t}$ and $\varepsilon_{i,t}$ are i.i.d random noise.

The individual fixed effects are drawn from a normal distribution $a_i \sim N(0, 5)$. The treatment is set to $\tau_0 = .5$. I vary the values of ρ_{10}. [26]

[23] The current methods that control for time varying unobserved heterogeneity rely on large T asymptotics so they can estimate latent factor structures, which is not the setting of this paper [Moon and Weidner, 2017].

[24] Here I keep the fixed effect term the same across the treatment and outcome model for simplicity, but exactly matching Assumption 1 by modify treatment to $D_{i,t} = c_i + u_{i,t}$ would produce similar results.

[25] It is important to note dynamic bias remains even if treatment is not a function of fixed effects and $D_{i,t} = u_{i,t}$.

[26] Though I show simulations for this setting, Nickell bias is smaller than dynamic bias in all DGPs I have studied.

Simulation estimators: I compare the OLS estimates of three different models.

1. Within-Transformed Static Model: From Equation (14). This model does not include past outcome as a control, and therefore has no lag.

$$\tilde{Y}_{i,t} = \tau^D \tilde{D}_{i,t} + \tilde{e}_{i,t}, \tag{45}$$

2. Within-Transformed Dynamic Model: From Equation (18). This model does include past outcome as a control, and therefore has a lag.

$$\tilde{Y}_{i,t} = \tau_0 \tilde{D}_{i,t} + \rho_{10} \tilde{Y}_{i,t-1} + \tilde{\varepsilon}_{i,t}, \tag{46}$$

3. Within-Transformed Delta Model:

$$\Delta\tilde{Y}_{i,t} = \tau_0 \tilde{D}_{i,t} + \tilde{\eta}_{i,t}, \tag{47}$$

I repeat the same data generating process for a range of different panel lengths. For each number of time periods N_T : 3 - 50, I generate datasets and plot the OLS estimate of the Static, Dynamic, and Delta Models.

Figure 3 (replicated from Section 2) plots the results. On the x-axis we have the number of time periods in the generated panel data set, and then the y-axis marks the treatment estimate. The true treatment effect value is marked with a black horizontal line.

The plots show that even when treatment is randomly assigned, dynamic bias and transformation bias are larger than Nickell bias. The Nickell bias is small because treatment effect coefficient is biased only because the coefficient on the past outcome is estimated with bias, which then leads to bias in the other parameters in the model.

Additional simulation results for a wide variety of parameters are given in AppendixD.1

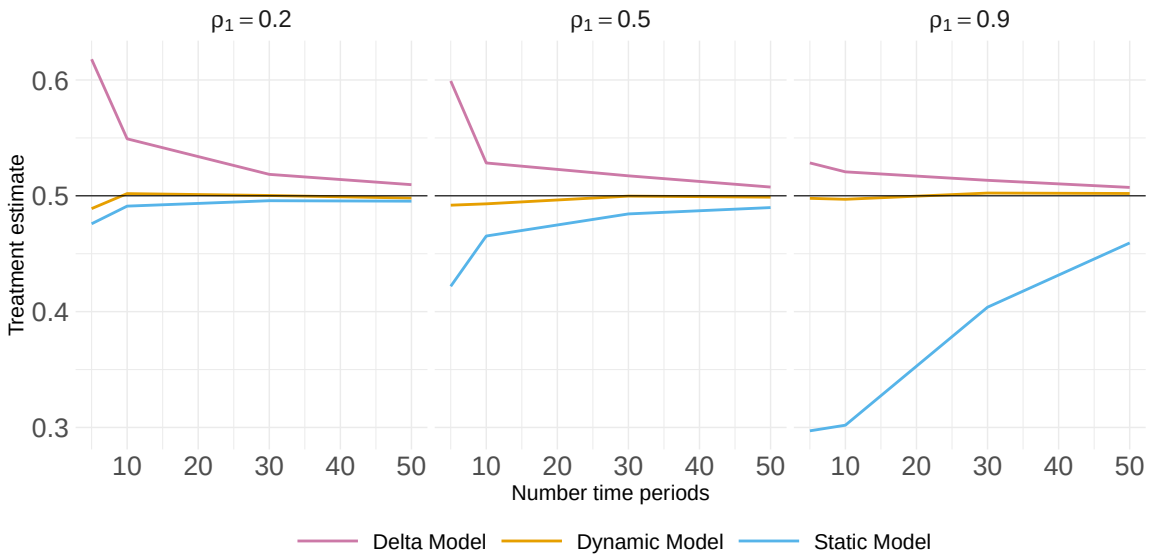

Figure 3: Bias of three different models.

5.1.2 Simulation Design with Endogenous Treatment

Now let us consider the case when treatment is not randomly assigned. I modify the DGP to follow the structural model outlined in Assumption 4:

$$Y_{i,t} = a_i + \rho_{01} Y_{i,t-1} + \tau_0 D_{i,t} + \varepsilon_{i,t}, \tag{48}$$

$$D_{i,t} = a_i + \rho_{02} Y_{i,t-1} + u_{i,t}. \tag{49}$$

I set $\rho_{20} = .1$, and generate results analogously to those in the previous section, but now using the endogenous treatment simulation design.

Figure 4 present the results. The bias gets larger the larger the absolute value of ρ_{20}. Recall that when treatment is endogenous, the bias for the Static and Delta model does not disappear as the number of time periods increases because omitted variable bias arises when past outcomes are not explicitly included in the model. Note in Figure 4, for the case of $\rho_{10} = .9$, the bias under the Static Model is so large it is omitted from the plot.

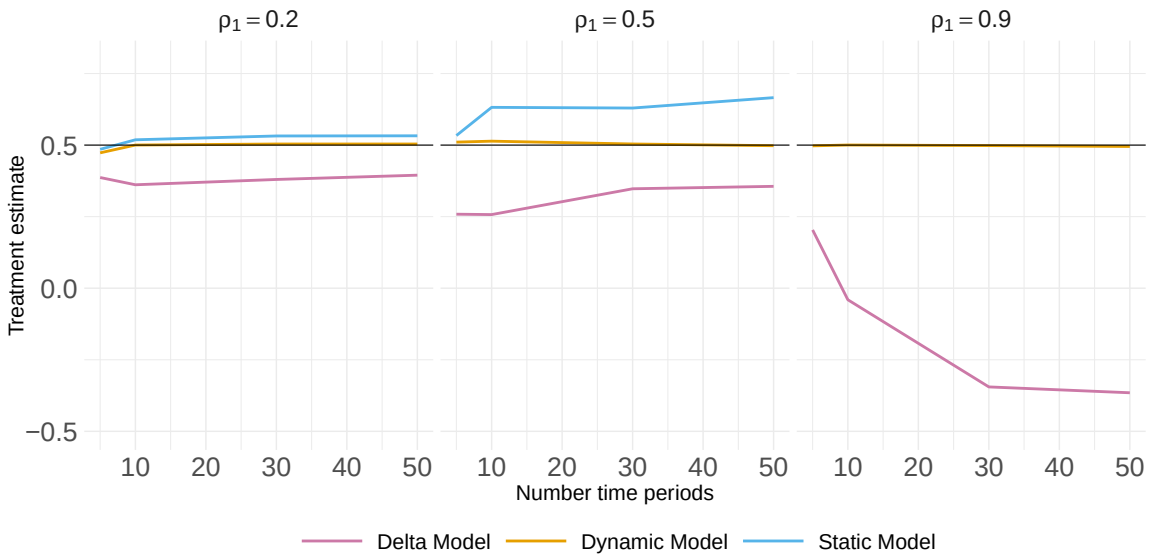

Figure 4: Bias of three different models.

5.2 Bias Correction Simulation

In these simulations I showcase the performance of my bias-corrected estimator. I also show how my estimator performs in comparison to the Arellano Bond estimator [Arellano and Bond, 1991].

I create a simulation data generating process with an endogenous treatment following Equations (48) and (49). Again, the individual fixed effects are drawn from a standard normal distribution $a_i \sim N(0, 5)$. The initial values $D_{i0} \sim N(a_i, 1)$. The vector of the true parameters is given by $\theta_0 = (\rho_{10}, \tau_0, \rho_{20}) = (.2, .5, .3)$.

Simulation Estimators:

1. Within-Transformed Dynamic Model from Equation (18). This model does include past outcome as a control, and therefore has a lag.

$$\tilde{Y}_{i,t} = \tau_0 \tilde{D}_{i,t} + \rho_{10} \tilde{Y}_{i,t-1} + \tilde{\varepsilon}_{i,t}, \tag{50}$$

2. My bias-corrected GMM estimator, presented in the previous section in Equation (35)

3. Arellano Bond estimator. Implemented using the R package `plm`[27].

[27] Identifying equations $E(\Delta Y_{i,t} - \Delta X_{i,t}\beta_0)X_i^{t-2} = 0, t = 2, \ldots, T.$. Our instruments are $Y_{i,t-2}, Y_{i,t-3}, Y_{i,t-4}, Y_{i,t-5}$, $D_{i,t-2}, D_{i,t-3}, D_{i,t-4}, D_{i,t-5}$

I run 1000 Monte Carlo simulations with the number of units $N = 1000$ and the number of time periods $N_T = 5$.

	DBC $\hat{\tau}$	OLS $\hat{\tau}$	AB $\hat{\tau}$
Mean	0.501	0.469	0.487
SD	0.015	0.014	0.311
95% cov	0.950	0.440	0.960

Table 1: Average estimates of the treatment effects τ_0.

	DBC $\hat{\rho}_1$	OLS $\hat{\rho}_1$	AB $\hat{\rho}_1$
Mean	0.198	-0.026	0.187
SD	0.020	0.014	0.099
95% cov	0.960	0.000	0.970

Table 2: Average estimates of parameter on past outcome in outome equation ρ_{10}.

Table 1 compares three different estimation strategies for treatment effects. The table compares my bias-corrected (DBC) estimator and ordinary least squares (OLS) estimates and Arellano Bond (AB) estimates. The DBC estimates are closest to the true treatment parameter $\tau_0 = .5$. The table presents the mean and standard deviation (sd) for each estimate, showing that DBC estimates yield proper 95% coverage, while OLS does not. The Arellano Bond (AB) estimator uses several lagged variables as instruments. The data was generated with random i.i.d errors, and so by construction the instruments are valid. The results demonstrate that AB models also achieve proper 95% coverage. I find that proper coverage is achieved regardless of the instruments used. However, the instrument choice does have a substantial effect on the standard deviations of the estimates. Using fewer instruments leads to an even larger standard deviation for the AB estimates. The large standard deviations explain why in practice for a particular dataset (as opposed to our Monte Carlo setting here where we average over 1000 datasets) the choice of instruments can lead to very different treatment effect estimates. In the table 1 present the AB results in which I selected the instruments that led to the smallest standard deviation for treatment (τ), using all possible instruments. Yet, even in this case, AB estimation still leads to standard deviations that are larger in comparison to DBC.

Table 2 compares the same three different estimation strategies for $\rho_{10} = .2$. While OLS led to biased estimates of treatment effects, the bias of the ρ_{10} is much larger and even changes sign. Both my DBC method and AB are able to achieve proper coverage of the true ρ_{10}, but the standard

deviation of method is again substantially smaller for my estimator.

6 Empirical Example

This paper is highly relevant for any applied research where past outcomes influence current outcomes. As discussed in the introduction, this includes studies focusing on variables such as agricultural yields, human capital, labor market outcomes, and migration flows.

To illustrate the application of my method, I focus on the relationship between temperature (as the treatment) and GDP (as the outcome). This relationship has been extensively explored in prior literature, and for this analysis, I utilize data from the seminal study by Dell et al. [2012].

6.1 GDP and Temperature

Dell et al. [2012] explores the relationship between temperature ($D_{i,t}$) and GDP ($Y_{i,t}$), and examines how rising temperatures can impact economic growth. The authors' results suggest that warmer temperatures can lead to reduced agricultural yields, decreased labor productivity, and increased health risk, all of which can hinder economic outcomes. The main result of their paper focuses on how higher temperatures affect poorer countries.

I use this empirical example to highlight two main points. First, I use the example to show how dynamic bias, transformation bias, and Nickell bias compare in a real world setting. This is discussed in detail in Section 6.1.1. Second, I highlight how my bias correction performs in practice and compare it to Arellano Bond. This is discussed in detail in Section 6.1.2.

6.1.1 Comparing Bias

In Table 3, I present estimates from regressions of GDP growth and levels on temperature using data from Dell et al. [2012]. The main results of Dell et al. [2012] focus on the impact of temperature on the GDP of poor countries, so my sample is restricted to poor countries. I do this for simplicity so that this analysis can focus on one treatment effect estimate. I exactly replicate the original analysis that uses all countries Dell et al. [2012] in Appendix F. The replication there is consistent with the results presented here.

In Table 3, I present the subset of results focused on poor countries. In the first two columns, I run baseline regressions where I regress GDP growth on temperature in the first column and then the second column shows how these estimates change once lagged GDP growth is included. In the third column and fourth columns I include all the original controls used in Dell et al. [2012], which are 371 region and time controls . The third column of my Table 3 follows the same specification given in the second column of Table 2 in Dell et al. [2012] - the outcome is GDP growth, and the past outcome is not controlled for. The treatment effect estimate without controlling for the lagged outcome is -1.421. Then in Column 4 I include lagged GDP growth, and the treatment effect estimate changes to -1.279. This is a 10% difference in treatment effects that is significant at the $p < .1$ level.[28]

Table 3

	Outcome: GDP Growth				*Outcome: GDP Level*	
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	−1.139***	−1.052***	−1.421***	−1.279***	174.202***	−19.330**
	(0.244)	(0.247)	(0.397)	(0.401)	(50.315)	(8.018)
Outcome Lag		0.136		0.109***		1.025***
		(0.083)		(0.021)		(0.003)
Controls	No	No	Yes	Yes	Yes	Yes
Observations	2,452	2,389	2,452	2,389	2,754	2,691
R^2	0.106	0.126	0.205	0.218	0.788	0.995
Adjusted R^2	0.082	0.102	0.114	0.127	0.761	0.994

*p<0.1; **p<0.05; ***p<0.01

These changes in coefficient occur despite the fact that I include the large number of time trend controls used in the original analysis. This highlights the point discussed in the introduction: time trend controls do not control for dynamics. To see this analytically, I run additional specifications to see how treatment effect estimates are impacted by the controls. In Table 3 Column 1 and 2 I run the same specifications in Column 3 and 4 respectively, just without controls. Adding

[28] To test whether the treatment estimate in Column 3 ($\hat{\tau}_3$) is statistically significantly smaller than the treatment estimate in Column 4, ($\hat{\tau}_4$) I conduct Wald test and the resulting p-value is .06. The Wald test requires accounting for the correlation between $\hat{\tau}_3$ and $\hat{\tau}_4$. These two estimates are highly correlated, which is why even though the estimates have a large variance, the null hypothesis is still rejected at the $p < .1$ level.

controls increases the estimated coefficient, whereas controlling for lags decreases the estimate – demonstrating that controls do not correct for dynamic relationships in outcomes. Adding the lagged outcome changes the treatment effect estimate by the same percentage regardless of whether or not controls are included.

It is important to note that GDP growth is calculated by transforming GDP levels. Looking at GDP growth rather then GDP levels may somewhat reduce dynamic bias, but it introduces transformation bias, as discussed in Section 2 and Appendix F.2. The magnitudes of both biases depend on ρ_{10} and the number of time periods. In this particular setting, the true ρ_{10} is very close to 1 and the number of time periods is 30, so the transformation bias is not as large as the dynamic bias.

To understand the impact of temperature on the original untransformed variable, I use GDP levels as the outcome in Table 3 Column 5 and Column 6. This is important because as discussed in Nath et al. [2024], the economic literature is divided over whether temperature affects GDP levels or growth rates. Therefore, it is also important to study GDP levels. Levels of outcomes are often studied in economics papers. [29] The impact of including past outcomes in the regression models is even greater when looking at GDP levels. Table 3 Column 5 shows that without controlling for past outcome the treatment estimate is unexpectedly positive (174.202), and only becomes negative (-19.330) as expected when controlling for past GDP level in Column 6.

6.1.2 Comparing Bias Correction to Arellano Bond

I benchmark the DBC estimator against AB using the Dell et al. [2012] panel data (T = 30), a setting where Nickell bias is expected to be modest due to the large number of time periods. I estimate the models on a balanced sample.[30]

Tables 4 and 5 report estimates of the treatment effect τ_0 and the lag coefficient ρ_{10} respectively. The OLS-with-lag column is the uncorrected OLS estimator and therefore contains Nickell bias. The DBC column reports the corresponding analytical bias-corrected estimates. As anticipated,

[29] Many economics papers study the levels of outcome variables without controlling for past outcomes. Somanathan et al. [2021] study the impact of temperature on economic productivity without controlling for levels, Annan and Schlenker [2015] study the temperature effects of crop yield levels.

[30] Hence small numerical differences from Table 3's unbalanced panel. The extension to unbalanced panels can be implemented but is left for future work.

DBC leaves $\hat{\tau}_0$ (Table 4), nearly unchanged and raises $\hat{\rho}_1$ (Table 5), consistent with Nickell bias loading on the lag coefficient.

I also estimate the coefficients using AB under three different instrument sets (columns "AB 1" - "AB 3" in Table 4 and Table 5). The key takeaway is that AB requires choosing instruments, and that choice materially affects the estimates. For τ_0, the sign flips from negative to positive and the magnitude changes by an order of magnitude (from -0.460 to 0.060); the same occurs for ρ_1, the estimate ranges from -0.016 to 0.101.

All standard errors are computed via a panel bootstrap clustered at the country level. DBC delivers smaller standard errors than AB, reflecting the variance inflation that accompanies many/weak-instrument designs; instrument choice materially shifts AB dispersion, amplifying the practical instability of its point estimates.

	DBC $\hat{\tau}$	OLS with Lag $\hat{\tau}$	AB 1 $\hat{\tau}$	AB 2 $\hat{\tau}$	AB 3 $\hat{\tau}$
Mean	-1.145	-1.143	0.060	-0.283	-0.460
SE	0.164	0.166	0.345	0.332	0.415

Table 4: Estimates of τ_0.

	DBC $\hat{\rho}_1$	OLS with Lag $\hat{\rho}_1$	AB 1 $\hat{\rho}_1$	AB 2 $\hat{\rho}_1$	AB 3 $\hat{\rho}_1$
Mean	0.200	0.163	0.094	0.101	-0.016
SE	0.094	0.090	0.100	0.108	0.141

Table 5: Estimates of ρ_{10}.

7 Conclusion

Dynamic outcomes are pervasive across applied economics—learning, health, production, and climate—yet a large share of empirical work still estimates static fixed-effects models that omit lagged outcomes.[31] This paper identifies and formalizes the resulting dynamic bias: a systematic bias of treatment-effect estimates that arises even under as-good-as-randomly assigned treatment. I show that omitting lagged outcomes breaks strict exogeneity and makes the transformed treatment correlated with the transformed error. In simulations, dynamic bias is larger by an order of magnitude than the previously studied Nickell bias that appears when lags are included. This dynamic bias

[31] Examples include Annan and Schlenker [2015], Burke et al. [2015], Cho [2017], Jessoe et al. [2018], Drabo and Mbaye [2015], Mahajan and Yang [2020], Missirian and Schlenker [2017], Graff Zivin et al. [2018] Garg et al. [2020].

implies that commonly used static models recover neither the short or long-run treatment effects. By bringing dynamic bias to light and quantifying its magnitude, the paper reframes the default use of static fixed effects models in applied economic work.

I provide a Dynamic Bias Correction (DBC) estimator that removes bias and works even in short panels. DBC builds an analytic correction for the bias and accommodates treatment processes that depend on past outcomes, and allows for heterogeneous effects via interactions. Relative to IV/GMM approaches, DBC avoids weak-instrument problems and delivers OLS-like precision in simulations.

An application to the Dell et al. [2012] temperature–GDP data illustrates the stakes. Accounting for dynamic bias reduces the estimated temperature effect on GDP growth by about 10% and on GDP levels by 120%. These results persist with even once rich time and region controls are included, underscoring that trends, factors, or mechanical transformations (levels-to-growth) are not substitutes for modeling within-unit dynamics. In fact, I show transformations can introduce their own transformation bias.

For researchers, the practical guidance is threefold. Because outcomes in economics are often dynamic, researchers should (i) state the causal estimand (short-run vs. long-run), (ii) include lagged outcomes to restore strict exogeneity of treatment, and (iii) correct the induced bias. DBC offers a transparent way to implement this in standard short panels without instruments.

Future work will focus on extending this estimator to more complex models by incorporating machine learning and adapting it to spatial data settings, where additional geographic dynamics play an important role.

References

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

B. W. Ang. Monitoring changes in economy-wide energy efficiency: From energy–gdp ratio to composite efficiency index. *Energy Policy*, 34(5):574–582, 2006. ISSN 0301-4215. doi: 10.1016/j.enpol.2005.11.011. URL https://doi.org/10.1016/j.enpol.2005.11.011.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.

Francis Annan and Wolfram Schlenker. Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review*, 105(5):262–266, 2015.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297, 1991.

Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021.

Juliano Assunção, Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues. Optimal environmental targeting in the amazon rainforest. *The Review of Economic Studies*, 90(4): 1608–1641, 2023.

Robert J Barro and Xavier Sala-i Martin. Convergence. *Journal of political Economy*, 100(2): 223–251, 1992.

Gaetano Basso, Douglas L. Miller, and Jessamyn Schaller. Dynamic treatment effects for empirical microeconomists: Local projections and quasi-experimental research designs. Unpublished working paper, 2022.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, Christian Hansen, and Kengo Kato. High-dimensional econometrics and regularized gmm. *arXiv preprint arXiv:1806.01888*, 2018.

Olivier J Blanchard and Lawrence H Summers. Beyond the natural rate hypothesis. *The American Economic Review*, 78(2):182–187, 1988.

Stéphane Bonhomme. Back to feedback: Dynamics and heterogeneity in panel data. Working paper / survey, 2025. URL https://sites.google.com/site/stephanebonhommeresearch/. Survey manuscript (Sargan Lecture, Royal Economic Society, July 2025).

Jörg Breitung, Sebastian Kripfganz, and Kazuhiko Hayakawa. Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics*, 24:116–132, 2022.

Marshall Burke, Solomon M Hsiang, and Edward Miguel. Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239, 2015.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

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

Clément Chaisemartin and Xavier d'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, pages 1–45, 2024.

Gary Chamberlain. Multivariate regression models for panel data. *Journal of econometrics*, 18(1): 5–46, 1982.

Kenneth Y Chay and Michael Greenstone. Does air quality matter? evidence from the housing market. *Journal of political Economy*, 113(2):376–424, 2005.

Hyunkuk Cho. The effects of summer heat on academic achievement: A cohort analysis. *Journal of Environmental Economics and Management*, 83:185–196, 2017.

Flavio Cunha and James Heckman. The technology of skill formation. *American economic review*, 97(2):31–47, 2007.

Yannic Damm, Elías Cisneros, and Jan Börner. Beyond deforestation reductions: Public disclosure, land-use change and commodity sourcing. *World Development*, 175:106481, 2024.

Melissa Dell, Benjamin F Jones, and Benjamin A Olken. Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66–95, 2012.

Tatyana Deryugina and Solomon M Hsiang. Does the environment still matter? daily temperature and income in the united states. Technical report, National Bureau of Economic Research, 2014.

Geert Dhaene and Koen Jochmans. Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030, 2015.

Alassane Drabo and Linguère Mously Mbaye. Natural disasters, migration and education: an empirical analysis in developing countries. *Environment and Development Economics*, 20(6): 767–796, 2015.

Iván Fernández-Val and Martin Weidner. Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312, 2016.

Teevrat Garg, Maulik Jagnani, and Vis Taraz. Temperature and human capital in india. *Journal of the Association of Environmental and Resource Economists*, 7(6):1113–1150, 2020.

Dalia Ghanem, Pedro HC Sant'Anna, and Kaspar Wüthrich. Selection and parallel trends. *arXiv preprint arXiv:2203.09001*, 2022.

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

Joshua Graff Zivin, Solomon M Hsiang, and Matthew Neidell. Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists*, 5(1):77–105, 2018.

Zvi Griliches. The sources of measured productivity growth: United states agriculture, 1940-60. *Journal of political economy*, 71(4):331–346, 1963.

Zvi Griliches. Distributed lags: A survey. *Econometrica: journal of the Econometric Society*, pages 16–49, 1967.

Jinyong Hahn and Guido Kuersteiner. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica*, 70(4):1639–1657, 2002.

Torfinn Harding, Julika Herzberg, and Karlygash Kuralbayeva. Commodity prices and robust environmental regulation: Evidence from deforestation in brazil. *Journal of Environmental Economics and Management*, 108:102452, 2021.

Jerry A Hausman. Taxes and labor supply. In *Handbook of public economics*, volume 1, pages 213–263. Elsevier, 1985.

Douglas Holtz-Eakin, Whitney Newey, and Harvey S Rosen. Estimating vector autoregressions with panel data. *Econometrica: Journal of the econometric society*, pages 1371–1395, 1988.

Katrina Jessoe, Dale T Manning, and J Edward Taylor. Climate change and labour allocation in rural mexico: Evidence from annual fluctuations in weather. *The Economic Journal*, 128(608): 230–261, 2018.

Arturas Juodis et al. Iterative bias correction procedures revisited: A small scale monte carlo study. Technical report, Universiteit van Amsterdam, Dept. of Econometrics, 2015.

Jan F Kiviet. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics*, 68(1):53–78, 1995.

Amanda E Kowalski. Reconciling seemingly contradictory results from the oregon health insurance experiment and the massachusetts health reform. *Review of Economics and Statistics*, 105(3): 646–664, 2023.

Leendert Marinus Koyck. *Distributed Lags and Investment Analysis*, volume 4 of *Contributions to Economic Analysis*. North-Holland Publishing Company, Amsterdam, 1954. Classic introduction to the geometric distributed-lag (Koyck) model.

Parag Mahajan and Dean Yang. Taken by storm: Hurricanes, migrant networks, and us immigration. *American Economic Journal: Applied Economics*, 12(2):250–277, 2020.

Philip Marx, Elie Tamer, and Xun Tang. Parallel trends and dynamic choices. *arXiv preprint arXiv:2207.06564*, 2022.

Pierre Mérel and Matthew Gammans. Climate econometrics: Can the panel approach account for long-run adaptation? *American Journal of Agricultural Economics*, 103(4):1207–1238, August 2021. doi: 10.1111/ajae.12200. URL https://doi.org/10.1111/ajae.12200.

Gilbert E. Metcalf. The impact of removing tax preferences for u.s. oil and gas production: Measuring tax subsidies by an equivalent price impact approach. *American Economic Journal: Economic Policy*, 11(4):360–390, 2019. doi: 10.1257/pol.20170179. URL https://www.jstor.org/stable/26798825.

Anna Mikusheva and Liyang Sun. Weak identification with many instruments. *The Econometrics Journal*, page utae007, 2024.

Anouch Missirian and Wolfram Schlenker. Asylum applications respond to temperature fluctuations. *Science*, 358(6370):1610–1614, 2017.

Hyungsik Roger Moon and Martin Weidner. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195, 2017.

Ishan B Nath, Valerie A Ramey, and Peter J Klenow. How much will global warming cool global growth? Technical report, National Bureau of Economic Research, 2024.

Marc Nerlove. *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*. Number 141 in Agriculture Handbook. U.S. Department of Agriculture, Agricultural Marketing Service, Washington, D.C., June 1958. Discusses Koyck's reduction approach around pp. 12–13.

Richard G Newell, Brian C Prest, and Steven E Sexton. The gdp-temperature relationship: implications for climate change damages. *Journal of Environmental Economics and Management*, 108:102445, 2021.

Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: journal of the Econometric Society*, pages 1–32, 1948.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, pages 1417–1426, 1981.

William D Nordhaus. An optimal transition path for controlling greenhouse gases. *Science*, 258 (5086):1315–1319, 1992.

William D Nordhaus. Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, 103(10):3510–3517, 2006.

Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry, 1992.

Robert A. Pollak. Habit formation and dynamic demand functions. *Journal of Political Economy*, 78(4, Part 1):745–763, 1970. doi: 10.1086/259669.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Aaron D. Smith and J. Edward Taylor. *Essentials of Applied Econometrics*. University of California Press, Oakland, CA, 1 edition, November 2016. ISBN 9780520288331. URL https://www.ucpress.edu/books/essentials-of-applied-econometrics/paper.

Robert M Solow. A contribution to the theory of economic growth. *The quarterly journal of economics*, 70(1):65–94, 1956.

Eswaran Somanathan, Rohini Somanathan, Anant Sudarshan, and Meenu Tewari. The impact of temperature on productivity and labor supply: Evidence from indian manufacturing. *Journal of Political Economy*, 129(6):1797–1827, 2021.

Jerzy Splawa-Neyman. Proba uzasadnienia zastosowafi rachunku prawdopodobiefistwa do doswiadczen polowych. *Roczniki Nauk Rolniczych*, 10:1–51, 1923.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event st udies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.

Lisa Tarquinio. *The politics of drought relief: Evidence from Southern India*. SSRN, 2022.

Nicholas Wheeler. Calculus of functionals. Classical Field Theory (Class Notes), Chapter 5. URL https://www.reed.edu/physics/faculty/wheeler/documents/Classical%20Field%20Theory/Class%20Notes/Field%20Theory%20Chapter%205.pdf. Lecture notes (PDF).

Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

A Dynamic Bias

A.1 Proof Theorem 3.1

Under Assumptions (1), (2), (3) and that $Var(D_{i,t}) > 0$ running OLS to estimate the static model in Equation (14) leads to a biased treatment effect estimator $\hat{\tau}^D$. Given that Equation (14) is a univariate panel regression, I have a simple formula for the OLS solution [Wooldridge, 2010].

$$\begin{aligned}\hat{\tau}^D &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Cov(\tilde{D}_{i,t}, Y_{i,t})}{\sum_{i=1}^{N}\sum_{t=1}^{T} Var(\tilde{D}_{i,t})} \\ &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Cov(\tilde{D}_{i,t}, \rho_{10} Y_{i,t-1} + \tau_0 D_{i,t} + \varepsilon_{i,t})}{\sum_{i=1}^{N}\sum_{t=1}^{T} Var(\tilde{D}_{i,t})}\end{aligned} \tag{51}$$

A.1.1 Numerator

$$Cov(\tilde{D}_{i,t}, \rho_{10} Y_{i,t-1} + \tau_0 D_{i,t} + \varepsilon_{i,t}) = \underbrace{Cov(\tilde{D}_{i,t}, \rho_{10} Y_{i,t-1})}_{N.1} + \underbrace{Cov(\tilde{D}_{i,t}, \tau_0 D_{i,t})}_{N.2} + \underbrace{Cov(\tilde{D}_{i,t}, \varepsilon_{i,t})}_{N.3} \tag{52}$$

$$\begin{aligned}N.1 = Cov(\tilde{D}_{i,t}, \rho_{10} Y_{i,t-1}) &= \rho_{10} Cov(D_{i,t} - \frac{1}{T}\sum_{s=1}^{T} D_{i,s}, Y_{i,t-1}) \\ &= -\frac{\rho_{10}}{T} Cov(\sum_{s=1}^{t-1} D_{i,s}, Y_{i,t-1}) \\ &= -\frac{\rho_{10}}{T} Cov(\sum_{s=1}^{t-1} D_{i,s}, \tau_0 \sum_{j=0}^{t-2} (\rho_{10})^j (D_{i,t-j-1})) \\ &= -\frac{\rho_{10}\tau_0}{T} \sum_{s=1}^{t-1} Cov(D_{i,s}, \sum_{j=0}^{t-2} (\rho_{10})^j (D_{i,t-j-1})) \\ &= -\frac{\rho_{10}\tau_0}{T} \sum_{s=1}^{t-1} Cov(D_{i,s}, (\rho_{10})^{t-s-1} (D_{i,s})) \\ &= -\frac{\rho_{10}\tau_0 Var(D_{i,t})}{T} \sum_{s=1}^{t-1} (\rho_{10})^{t-s-1}\end{aligned} \tag{53}$$

This follows by plugging in for $Y_{i,t-1}$, which follows from 80. Therefore:

$$\sum_{t=1}^{T} Cov(\tilde{D}_{i,t}, \rho_{10} Y_{i,t-1}) = -\frac{\rho_{10}\tau_0 Var(D_{i,t})}{T}\left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)}\right] \tag{54}$$

The next term in the numerator is:

$$\begin{aligned} N.2 = Cov(\tilde{D}_{i,t}, \tau_0 D_{i,t}) &= \tau_0 Cov(D_{i,t} - \frac{1}{T}\sum_{t=1}^{T} D_{i,t}, D_{i,t}) \\ &= \tau_0 Cov(D_{i,t}, D_{i,t}) - \tau_0 Cov(\frac{1}{T}\sum_{t=1}^{T} D_{i,t}, D_{i,t}) \\ &= \tau_0\left[1 - \frac{1}{T}\right] Var(D_{i,t}) \end{aligned} \tag{55}$$

The final term in the numerator is:

$$N.3 = Cov(\tilde{D}_{i,t}, \varepsilon_{i,t}) = 0 \tag{56}$$

Putting these parts together, and including the outside sums, our full numerator is.

$$\begin{aligned} \sum_{i=1}^{N}\sum_{t=1}^{T} Cov(\tilde{D}_{i,t}, Y_{i,t}) &= \sum_{i=1}^{N}\left(T\tau_0\left[1 - \frac{1}{T}\right] Var(D_{i,t}) - \frac{\rho_{10}\tau_0 Var(D_{i,t})}{T}\left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)}\right]\right) \\ &= \sum_{i=1}^{N}\left(T\tau_0\left[1 - \frac{1}{T}\right]\sigma_i^2 - \frac{\rho_{10}\tau_0\sigma_i^2}{T}\left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)}\right]\right) \end{aligned} \tag{57}$$

A.1.2 Denominator

$$\text{Var}\left(D_{it} - \bar{D}_i\right) = \text{Var}(D_{it}) + \text{Var}(\bar{D}_i) - 2 \cdot \text{Cov}(D_{it}, \bar{D}_i) \tag{58}$$

where:

$$\begin{aligned} \text{Var}(D_{it}) &= \sigma_i^2 \\ \text{Var}(\bar{D}_i) &= \frac{\sigma_i^2}{T} \\ \text{Cov}(D_{it}, \bar{D}_i) &= \frac{\sigma_i^2}{T} \end{aligned} \tag{59}$$

Substituting these into the variance formula:

$$\text{Var}\left(D_{it} - \bar{D}_i\right) = \sigma_i^2 + \frac{\sigma_i^2}{T} - 2 \cdot \frac{\sigma_i^2}{T} = \sigma_i^2 \left(1 - \frac{1}{T}\right) \tag{60}$$

Hence the denominator is

$$\sum_{i=1}^{N} \sum_{t=1}^{T} Var(\tilde{D}_{i,t}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i^2 \left(1 - \frac{1}{T}\right) \tag{61}$$

A.1.3 Combining Numerator and Denominator

$$\begin{aligned}
\hat{\tau}^D &= \frac{\sum_{i=1}^{N} \left(T\tau_0 \left[1 - \frac{1}{T}\right] \sigma_i^2 - \frac{\rho_{10}\tau_0\sigma_i^2}{T} \left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)} \right] \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i^2 \left(1 - \frac{1}{T}\right)} \\
&= \tau_0 - \frac{\sum_{i=1}^{N} \left(\frac{\rho_{10}\tau_0\sigma_i^2}{T} \left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)} \right] \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_i^2 \left(1 - \frac{1}{T}\right)} \\
&= \tau_0 - \frac{\frac{\rho_{10}\tau_0}{T} \left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)} \right]}{\sum_{t=1}^{T} \left(1 - \frac{1}{T}\right)} \\
&= \tau_0 - \frac{\frac{\rho_{10}\tau_0}{T} \left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)} \right]}{(T-1)} \\
&= \tau_0 - \frac{\rho_{10}\tau_0}{T(T-1)} \left[\frac{T}{1-\rho_{10}} - \frac{1-\rho_{10}^T}{(1-\rho_{10}^2)} \right]
\end{aligned} \tag{62}$$

B Bias Correction Formulas

B.1 Proof for Lemma 4.1

Under Assumptions 2, 3, 4 the following bias correction term is calculated.

$$\mathbf{b}(\theta_0) = \left[\begin{pmatrix} b_{\rho_1}(\theta_0) \\ b_{\tau}(\theta_0) \\ b_{\rho_2}(\theta_0) \end{pmatrix} \right] = \mathbb{E} \left[\begin{pmatrix} \frac{1}{T} \sum_{t=1}^{T} \tilde{Y}_{i,t-1} \tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T} \sum_{t=1}^{T} \tilde{D}_{i,t} \tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T} \sum_{t=1}^{T} \tilde{Y}_{i,t-1} \tilde{u}_{i,t}(\theta_0) \end{pmatrix} \right] \tag{63}$$

There are three expectations that have to be solved in order to calculate $\mathbf{b}(\theta_0)$. I have to solve for 1) $b_{\rho_1}(\theta_0)$ 2) $b_{\tau}(\theta_0)$ 3) $b_{\rho_2}(\theta_0)$. I first solve for $b_{\rho_1}(\theta_0) = \mathbb{E}_{\theta_0}[\frac{1}{T} \sum_{t=1}^{T} \tilde{Y}_{i,t-1} \tilde{\varepsilon}_{i,t}(\theta_0)]$.

B.1.1 Bias Term 1

Equations (64) to (70) below I outline the general argument for how to derive $b_{\rho_1}(\theta_0)$. Then after (70) I provide additional detail for each step.

$$b_{\rho_1}(\theta_0) = \mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0)\right] \tag{64}$$

$$= -\mathbb{E}_{\theta_0}\left[\bar{Y}_{i,-1}\bar{\varepsilon}_i(\theta_0)\right] \tag{65}$$

$$= -\mathbb{E}_{\theta_0}\left[\frac{1}{T}\left\{\sum_{l=0}^{T-2}\left(\sum_{j=0}^{l}\phi_0^j\right)\varepsilon_{i,T-1-l}(\theta_0)\right\}\bar{\varepsilon}_i(\theta_0)\right] \tag{66}$$

$$= -\frac{1}{T}\sum_{l=0}^{T-2}\left(\sum_{j=0}^{l}\phi_0^j\right)\mathbb{E}_{\theta_0}[\varepsilon_{i,T-1-l}(\theta_0)\bar{\varepsilon}_i(\theta_0)] \tag{67}$$

$$= -\frac{1}{T^2}\sum_{l=0}^{T-2}\left(\sum_{j=0}^{l}\phi_0^j\right)\sigma^2_{\varepsilon,iT}(\theta_0) \tag{68}$$

$$= -\frac{1}{T^2}\sum_{l=0}^{T-2}\left(\frac{1-\phi^{l+1}}{1-\phi}\right)\sigma^2_{\varepsilon,iT}(\theta_0) \tag{69}$$

$$= -\frac{\sigma^2_{\varepsilon,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \tag{70}$$

Now I explain the steps in greater detail.

Steps for Line (114)

$$\mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0)\right] = \mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}(Y_{i,t-1}-\bar{Y}_{i,-1})(\varepsilon_{i,t}(\theta_0)-\bar{\varepsilon}_i(\theta_0))\right] \tag{71}$$

$$= \mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}(Y_{i,t-1}-\bar{Y}_{i,-1})(\varepsilon_{i,t}(\theta_0))\right] \tag{72}$$

$$= -\mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\bar{Y}_{i,-1}\varepsilon_{i,t}(\theta_0)\right] \tag{73}$$

$$= -\mathbb{E}_{\theta_0}\left[\bar{Y}_{i,-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}(\theta_0)\right] \tag{74}$$

$$= -\mathbb{E}_{\theta_0}\left[\bar{Y}_{i,-1}\bar{\varepsilon}_i(\theta_0)\right] \tag{75}$$

Because of weak indepdence I have that the covariance of $Y_{i,t-1}$ and $\varepsilon_{i,t}$ is zero, and the mean of $\varepsilon_{i,t}$ is zero. Therefore $\mathbb{E}_{\theta_0}[\frac{1}{T}\sum_{t=1}^{T} Y_{i,t-1}\varepsilon_{i,t}] = 0$.

Steps for Line (66) Here I derive an expression for $\bar{Y}_{i,-1}$ as a sum of initial and past values.

I start by first deriving an expression for $Y_{i,t}$ as a sum of initial and past values.

$$D_{i,1} = c_i + \rho_2 Y_{i,0} + u_{i,1} \tag{76}$$

$$\begin{aligned} Y_{i,1} &= a_i + \rho_1 Y_{i,0} + \tau D_{i,1} + \varepsilon_{i1} \\ &= a_i + \rho_1 Y_{i,0} + \tau[c_i + \rho_2 Y_{i,0} + u_{i1}] + \varepsilon_{i1} \\ &= a_i + (\rho_1 + \tau\rho_2) Y_{i,0} + \tau(c_i + u_{i1}) + \varepsilon_{i1} \end{aligned} \tag{77}$$

$$\begin{aligned} Y_{i,2} &= a_i + (\rho_1 + \tau\rho_2) Y_{i,1} + \tau(c_i + u_{i1}) + \varepsilon_{i2} \\ &= a_i + (\rho_1 + \tau\rho_2)\Big[a_i + (\rho_1 + \tau\rho_2) Y_{i,0} + \tau(c_i + u_{i0}) + \varepsilon_{i1}\Big] + \tau(c_i + u_{i1}) + \varepsilon_{i2} \end{aligned} \tag{78}$$

Notice the recursive nature in the terms, use this pattern to generalize $Y_{i,t}$:

$$Y_{i,t} = a_i + (\rho_1 + \tau\rho_2) Y_{i,t-1} + \tau(c_i + u_{i,t}) + \varepsilon_{i,t} \tag{79}$$

Generalizing this pattern for any t:

$$\begin{aligned} Y_{i,t} = a_i \sum_{j=0}^{t-1} (\rho_1 + \tau\rho_2)^j + (\rho_1 + \tau\rho_2)^t Y_{i,0} + \tau \sum_{j=0}^{t-1} (\rho_1 + \tau\rho_2)^j (c_i + u_{i,t-j}) \\ + \sum_{j=0}^{t-1} (\rho_1 + \tau\rho_2)^j \varepsilon_{i,t-j} \end{aligned} \tag{80}$$

Recall $\phi_0 = (\rho_{10} + \tau_0\rho_{20})$. The expression for $\bar{Y}_{i,-1}$ follows:

$$\bar{Y}_{i,-1} = \frac{1}{T}\Bigg[\Bigg(\sum_{l=0}^{T-2} (T - l - 1)\phi_0^l\Bigg) a_i + \Bigg(\sum_{l=0}^{T-1} \phi_0^l\Bigg) Y_{i,0} + \sum_{l=0}^{T-2} \Bigg(\sum_{j=0}^{l} \phi_0^j\Bigg) \tau_0 (c_i + u_{i,T-1-l}) \tag{81}$$

$$+ \sum_{l=0}^{T-2} \Bigg(\sum_{j=0}^{l} \phi_0^j\Bigg) \varepsilon_{i,T-1-l}\Bigg] \tag{82}$$

Steps for Line (68) Follows by the definition of variance. I use that the noise is homoscedastic over time, but heteroscedastic across individuals.

Steps for Line (69) Follows by rules of genometrics sums.

B.1.2 Bias Term 2

The second bias term follows almost an identical argument to what was outlined in Section B.1.1 for Term 1.

$$b_\tau(\theta_0) = \mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}(\theta_0)\right] \tag{83}$$

$$= -\rho_2\frac{\sigma^2_{\varepsilon,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \tag{84}$$

It follows from the fact that.

$$\bar{D}_i = \rho_2\bar{Y}_{i,-} + \bar{e}_i \tag{85}$$

Since $\bar{Y}_{i,-}$ is solved above, it can be plugged in. The same exogeneity conditions hold.

B.1.3 Bias Term 3

The third bias term follows almost an identical argument to what was outlined in Section B.1.1 for Term 1. Again, since $\bar{Y}_{i,-}$ is solved above, it can be plugged in. The same exogeneity conditions hold.

$$b_{\rho_2}(\theta_0) = \mathbb{E}_{\theta_0}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta_0)\right] \tag{86}$$

$$= -\tau_0\frac{\sigma^2_{u,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \tag{87}$$

B.2 Interaction Model

Thie bias correction for the more general model is proven here.

$$\mathbf{b}(\theta_0) = \begin{bmatrix} b_{\rho_1}(\theta_0) \\ b_{\tau_c}(\theta_0) \\ b_{\rho_2}(\theta_0) \\ b_{\beta_1}(\theta_0) \\ b_{\beta_2}(\theta_0) \end{bmatrix} = \mathbb{E} \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} (\tilde{D}_{i,t}W^c_{i,t})\tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_{1,i,t}\tilde{\varepsilon}_{i,t}(\theta_0) \\ \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_{2,i,t}\tilde{u}_{i,t}(\theta_0) \end{bmatrix} \tag{88}$$

The outcome model has interactions with strictly exogenous covariates. Let there be N_c different covariates $W^1_{i,t}, W^2_{i,t}, \ldots, W^{N_c}_{i,t}$ that interact with treatment. The strictly exogenous regressors that do not interact with treatment are denoted by $X_{i,t}$.

The more general model that allows for interactions is given by:

$$Y_{i,t} = a_i + \sum_{c=1}^{N_c} \tau_c (D_{i,t} \cdot W^c_{i,t}) + \beta_1 X_{1,i,t} + \rho_1 Y_{i,t-1} + \varepsilon_{i,t} \tag{89}$$

$$D_{i,t} = c_i + \rho_2 Y_{i,t-1} + \beta_2 X_{2,i,t} + u_{i,t} \tag{90}$$

Substitute $D_{i,t}$ into the outcome equation:

$$\begin{aligned} Y_{i,t} = a_i + \sum_{c=1}^{N_c} \tau_c \left((c_i + \rho_2 Y_{i,t-1} + \beta_2 X_{2,i,t} + u_{i,t}) \cdot W^c_{i,t} \right) \\ + \beta_1 X_{1,i,t} + \rho_1 Y_{i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{91}$$

Distribute τ_c and group terms by $Y_{i,t-1}$:

$$\begin{aligned} Y_{i,t} = a_i &+ \sum_{c=1}^{N_c} \tau_c c_i W_{i,t}^c + \sum_{c=1}^{N_c} \tau_c \rho_2 Y_{i,t-1} W_{i,t}^c \\ &+ \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t} W_{i,t}^c + \sum_{c=1}^{N_c} \tau_c u_{i,t} W_{i,t}^c \\ &+ \beta_1 X_{1,i,t} + \rho_1 Y_{i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{92}$$

Grouping by $Y_{i,t-1}$: Let:

$$\gamma_t = \rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 W_{i,t}^c \tag{93}$$

$$\alpha_t = \sum_{c=1}^{N_c} \tau_c u_{i,t} W_{i,t}^c \tag{94}$$

Then:

$$\begin{aligned} Y_{i,t} = a_i &+ \sum_{c=1}^{N_c} \tau_c c_i W_{i,t}^c + \gamma_t Y_{i,t-1} \\ &+ \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t} W_{i,t}^c + \alpha_t + \beta_1 X_{1,i,t} + \varepsilon_{i,t} \end{aligned} \tag{95}$$

Iterate the equation backwards: Substitute $Y_{i,t-1}$ in terms of $Y_{i,t-2}$:

$$\begin{aligned} Y_{i,t-1} = a_i &+ \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-1}^c + \gamma_{t-1} Y_{i,t-2} \\ &+ \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-1} W_{i,t-1}^c + \alpha_{t-1} + \beta_1 X_{1,i,t-1} + \varepsilon_{i,t-1} \end{aligned} \tag{96}$$

So:

$$
\begin{aligned}
Y_{i,t} = a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t}^c + \gamma_t \Bigg(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-1}^c + \gamma_{t-1} Y_{i,t-2} \\
+ \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-1} W_{i,t-1}^c + \alpha_{t-1} + \beta_1 X_{1,i,t-1} + \varepsilon_{i,t-1} \Bigg) \\
+ \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t} W_{i,t}^c + \alpha_t + \beta_1 X_{1,i,t} + \varepsilon_{i,t}
\end{aligned} \tag{97}
$$

Continue substituting backward iteratively until reaching $Y_{i,0}$:

$$
\begin{aligned}
Y_{i,t} = \left(\prod_{j=0}^{t-1} \gamma_{t-j} \right) Y_{i0} \\
+ \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j} \right) \left(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c \right. \\
\left. + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c + \alpha_{t-k} + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \right)
\end{aligned} \tag{98}
$$

Summarize the final expression for $Y_{i,t}$:

$$
\begin{aligned}
Y_{i,t} = \left(\prod_{j=0}^{t-1} \gamma_{t-j} \right) Y_{i0} \\
+ \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j} \right) \left(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c \right. \\
\left. + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c + \sum_{c=1}^{N_c} \tau_c u_{i,t-k} W_{i,t-k}^c + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \right)
\end{aligned} \tag{99}
$$

Expression plugging in for variables:

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} Y_{i,t} = \frac{1}{T} \sum_{t=1}^{T} \Bigg(\left(\prod_{j=0}^{t-1} \gamma_{t-j} \right) Y_{i0} \\
+ \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j} \right) \left(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c \right. \\
\left. + \sum_{c=1}^{N_c} \tau_c u_{i,t-k} W_{i,t-k}^c + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \right) \Bigg)
\end{aligned} \tag{100}
$$

Expression for $D \times W$ Average: Similarly, for $D_{i,t}W_{i,t}^c$:

$$D_{i,t}W_{i,t}^c = \left(c_i + \rho_2 Y_{i,t-1} + \beta_2 X_{2,i,t} + u_{i,t}\right) W_{i,t}^c \tag{101}$$

Substitute $Y_{i,t-1}$:

$$\begin{aligned} D_{i,t}W_{i,t}^c = \Bigg(& c_i + \rho_2 \Bigg(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-1}^c + \gamma_{t-1} Y_{i,t-2} \\ & + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-1} W_{i,t-1}^c + \alpha_{t-1} + \beta_1 X_{1,i,t-1} + \varepsilon_{i,t-1} \Bigg) \\ & + \beta_2 X_{2,i,t} + u_{i,t} \Bigg) W_{i,t}^c \end{aligned} \tag{102}$$

Iterating backward:

$$\begin{aligned} D_{i,t}W_{i,t}^c = \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j-1} \right) \Bigg(& c_i W_{i,t-k}^c + \rho_2 \Bigg(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c \\ & + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c + \alpha_{t-k} + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \Bigg) \Bigg) \end{aligned} \tag{103}$$

Summarizing the result:

$$\begin{aligned} D_{i,t}W_{i,t}^c = \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j-1} \right) \Bigg(& c_i W_{i,t-k}^c + \rho_2 \Bigg(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c \\ & + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c + \alpha_{t-k} + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \Bigg) \Bigg) \end{aligned} \tag{104}$$

Therefore the average can be written as:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} D_{i,t}W_{i,t}^c = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=0}^{t-1} \left(\prod_{j=0}^{k-1} \gamma_{t-j-1} \right) \Bigg(& c_i W_{i,t-k}^c + \rho_2 \Bigg(a_i + \sum_{c=1}^{N_c} \tau_c c_i W_{i,t-k}^c \\ & + \sum_{c=1}^{N_c} \tau_c \beta_2 X_{2,i,t-k} W_{i,t-k}^c + \alpha_{t-k} + \beta_1 X_{1,i,t-k} + \varepsilon_{i,t-k} \Bigg) \Bigg) \end{aligned} \tag{105}$$

Therefore it follow by the same argument outlined in Appendix B.1. Here $\phi = \rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 \mu_c$.

Here μ_c is the mean of covariate W^c

$$\mathbf{b}(\theta_0) = \begin{bmatrix} b_{\rho_1}(\theta_0) \\ b_{\tau_c}(\theta_0) \\ b_{\rho_2}(\theta_0) \\ b_{\beta_1}(\theta_0) \\ b_{\beta_2}(\theta_0) \end{bmatrix} = \begin{bmatrix} -\frac{\sigma^2_{\varepsilon,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \\ -\rho_2\mu_c\frac{\sigma^2_{\varepsilon,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \\ -(\sum_{c=1}^{N_c}\tau_c\mu_c)\frac{\sigma^2_{u,iT}}{T}\left(\frac{T-1}{1-\phi} - \frac{\phi-\phi^T}{(1-\phi)^2}\right) \\ 0 \\ 0 \end{bmatrix} \tag{106}$$

B.3 Algebra interaction term

Note:

$$\begin{aligned} \bar{Y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} Y_{i,t} = \frac{1}{T}\sum_{t=1}^{T}\left(\prod_{j=0}^{t-1}\gamma_{t-j}\right)Y_{i0} \\ + \frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\left(\prod_{j=0}^{k-1}\gamma_{t-j}\right)\left(a_i + \sum_{c=1}^{N_c}\tau_c c_i W^c_{i,t-k} + \sum_{c=1}^{N_c}\tau_c u_{i,t-k}W^c_{i,t-k} + \beta X_{i,t-k} + \varepsilon_{i,t-k}\right) \end{aligned} \tag{107}$$

We want to find $\mathbb{E}_{\theta_0}\Big[\bar{Y}_{i,-1}\bar{\varepsilon}_i(\theta_0)\Big]$.

The part of $\bar{Y}_{i,-1}$ we care about is $\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\left(\prod_{j=0}^{k-1}\gamma_{t-j}\right)\varepsilon_{i,t-k}$

$$\begin{aligned} &\mathbb{E}_{\theta_0}\Big[\left\{\bar{Y}_{i,-1}\right\}\bar{\varepsilon}_i(\theta_0)\Big] \\ &= \mathbb{E}_{\theta_0}\Big[\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\left(\prod_{j=0}^{k-1}\gamma_{t-j}\right)\varepsilon_{i,t-k}\right\}\bar{\varepsilon}_i(\theta_0)\Big] \end{aligned} \tag{108}$$

Let us consider a simple example of the expression above.

$$\begin{aligned} \mathbb{E}\Big[\sum_{t=1}^{3}\left\{A_tB_t\right\}C\Big] &= \mathbb{E}\Big[\left\{A_1B_1 + A_2B_2 + A_3B_3\right\}C\Big] \\ &= \mathbb{E}\Big[\left\{A_1B_1C + A_2B_2C + A_3B_3C\right\}\Big] \\ &= \mathbb{E}\Big[A_1B_1C\Big] + \mathbb{E}\Big[A_2B_3C\Big] + \mathbb{E}\Big[A_3B_3C\Big] \end{aligned} \tag{109}$$

Consider $\mathbb{E}\Big[A_1 B_1 C\Big]$. In the first line below I use the fact that A is orthogonal to B and C.

$$\mathbb{E}\Big[A_1 B_1 C\Big] = \mathbb{E}\Big[A_1\Big] \times \mathbb{E}\Big[B_1 C\Big] \tag{110}$$

In our case

$$\mathbb{E}[A_t] \sim \mathbb{E}[\left(\prod_{j=0}^{k-1} \gamma_{t-j}\right)] \tag{111}$$

Below the first line uses the definition of γ_{t-j}. The second line uses the fact that the covariates are strictly exogenous variables that are iid across time and unit, so they are independent. The third line uses the fact that the expectation of a covaraiate is the same across time.

$$\begin{aligned}
\mathbb{E}\Big[\prod_{j=0}^{k-1} \gamma_{t-j}\Big] &= \mathbb{E}\Big[\prod_{j=0}^{k-1} \left\{\rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 W_{i,t-j}^c\right\}\Big] \\
&= \prod_{j=0}^{k-1} \left\{\mathbb{E}\Big[\rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 W_{i,t-j}^c\Big]\right\} \\
&= \prod_{j=0}^{k-1} \left\{\rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 E[W_{i,t-j}^c]\right\} \\
&= \prod_{j=0}^{k-1} \left\{\rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 \mu_c\right\} \\
&= \left\{\rho_1 + \sum_{c=1}^{N_c} \tau_c \rho_2 \mu_c\right\}^k \\
&= \phi^k
\end{aligned} \tag{112}$$

Therefore we have

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\Big[\left\{\bar{Y}_{i,-1}\right\} \bar{\varepsilon}_i(\theta_0)\Big] \\
&= \frac{1}{T} \sum_{t=1}^{T} \sum_{k=0}^{t-1} \left(\phi^k\right) \mathbb{E}_{\theta_0}[\varepsilon_{i,t-k} \bar{\varepsilon}_i(\theta_0)]
\end{aligned} \tag{113}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\Big[\left\{\bar{Y}_{i,-1}\right\} \bar{\varepsilon}_i(\theta_0)\Big] \\
&= \frac{1}{T} \sum_{l=0}^{T-2} \sum_{j=0}^{l} \left(\phi^j\right) \mathbb{E}_{\theta_0}[\varepsilon_{i,T-1-l} \bar{\varepsilon}_i(\theta_0)]
\end{aligned} \tag{114}$$

C Normality of bias-corrected Estimator

Proof for Theorem 4.2.

C.1 Gradient

The gradient of the bias-corrected estimator is the following.

$$\nabla_\theta \mathbf{m}_{NT}^{DBC}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta \mathbf{m}_{iT}^{DBC}(\theta) \tag{115}$$

Where the individual moment gradient is given by:

$$\nabla_\theta \mathbf{m}_{iT}^{DBC}(\theta) = \begin{pmatrix} \nabla_{\rho_1} m_{\rho_1,iT}^{DBC}(\theta) & \nabla_\tau m_{\rho_1,iT}^{DBC}(\theta) & \nabla_{\rho_2} m_{\rho_1,iT}^{DBC}(\theta) \\ \nabla_{\rho_1} m_{\tau,iT}^{DBC}(\theta) & \nabla_\tau m_{\tau,iT}^{DBC}(\theta) & \nabla_{\rho_2} m_{\tau,iT}^{DBC}(\theta) \\ \nabla_{\rho_1} m_{\rho_2,iT}^{DBC}(\theta) & \nabla_\tau m_{\rho_2,iT}^{DBC}(\theta) & \nabla_{\rho_2} m_{\rho_2,iT}^{DBC}(\theta) \end{pmatrix} \tag{116}$$

has the elements

$$\nabla_{\rho_1} m_{\rho_1,iT}^{DBC}(\theta) = -\frac{1}{T}\sum_{t=1}^{T} (\tilde{Y}_{i,t-1})\tilde{Y}_{i,t-1} - \nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \Phi(\theta)\nabla_{\rho_1}\hat{\sigma}^2_{\varepsilon,iT}(\theta), \tag{117}$$

$$\nabla_{\tau} m_{\rho_1,iT}^{DBC}(\theta) = -\frac{1}{T}\sum_{t=1}^{T} (\tilde{Y}_{i,t-1})D_{i,t} - \rho_2\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \Phi(\theta)\nabla_{\tau}\hat{\sigma}^2_{\varepsilon,iT}(\theta), \tag{118}$$

$$\begin{aligned} \nabla_{\rho_2} m_{\rho_1,iT}^{DBC}(\theta) &= -\tau\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \Phi(\theta)\nabla_{\rho_2}\hat{\sigma}^2_{\varepsilon,iT}(\theta) \\ &= -\tau\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta), \end{aligned} \tag{119}$$

$$\nabla_{\rho_1} m_{\tau,iT}^{DBC}(\theta) = -\frac{1}{T}\sum_{t=1}^{T} (\tilde{D}_{i,t-1})Y_{i,t-1} - \rho_2\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \rho_2\Phi(\theta)\nabla_{\rho_1}\hat{\sigma}^2_{\varepsilon,iT}(\theta), \tag{120}$$

$$\nabla_{\tau} m_{\tau,iT}^{DBC}(\theta) = -\frac{1}{T}\sum_{t=1}^{T} (\tilde{D}_{i,t-1})D_{i,t} - \rho_2^2\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \rho_2\Phi(\theta)\nabla_{\tau}\hat{\sigma}^2_{\varepsilon,iT}(\theta) \tag{121}$$

$$\begin{aligned}\nabla_{\rho_2} m^{DBC}_{\tau,iT}(\theta) &= -\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \rho_2\tau\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \rho_2\Phi(\theta)\nabla_{\rho_2}\hat{\sigma}^2_{\varepsilon,iT}(\theta) \\ &= -\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \tau\nabla\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta),\end{aligned} \tag{122}$$

$$\begin{aligned}\nabla_{\rho_1} m^{DBC}_{\rho_2,iT}(\theta) &= -\tau\nabla\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta) - \tau\Phi(\theta)\nabla_{\rho_1}\hat{\sigma}^2_{u,iT}(\theta) \\ &= -\tau\nabla\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta)\end{aligned} \tag{123}$$

$$\begin{aligned}\nabla_{\tau} m^{DBC}_{\rho_2,iT}(\theta) = \nabla_{\rho_2} m^{DBC}_{\tau,iT}(\theta) &= -\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta) - \rho_2\tau\nabla\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta) - \tau\Phi(\theta)\nabla_{\tau}\hat{\sigma}^2_{u,iT}(\theta) \\ &= -\Phi(\theta)\hat{\sigma}^2_{\varepsilon,iT}(\theta) - \tau\nabla\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta),\end{aligned} \tag{124}$$

$$\nabla_{\rho_2} m_{\rho_2,iT}(\theta) = -\frac{1}{T}\sum_{t=1}^{T}(\tilde{Y}_{i,t-1})Y_{i,t-1} - \tau^2\nabla\Phi(\theta)\hat{\sigma}^2_{u,iT}(\theta) - \tau\Phi(\theta)\nabla_{\rho_2}\hat{\sigma}^2_{u,iT}(\theta)(\theta), \tag{125}$$

where

$$\nabla\Phi(\theta) = -T^{-2}\sum_{l=1}^{T-2}\sum_{j=1}^{l} l(\rho_1 + \tau\rho_2)^{l-1} \tag{126}$$

$$\Phi(\theta) = -T^{-2}\sum_{l=1}^{T-2}\sum_{j=1}^{l}(\rho_1 + \tau\rho_2)^{l} \tag{127}$$

$$\begin{aligned}\nabla_{\rho_1}\hat{\sigma}_{\varepsilon,iT}(\theta) &= \frac{2}{T-1}\sum_{t=1}^{T}(\tilde{Y}_{i,t-1})\varepsilon_{i,t}(\theta) \\ \mathbb{E}[\nabla_{\rho_1}\hat{\sigma}_{\varepsilon,iT}(\theta)] &= 2\Phi(\theta)\sigma^2_{\varepsilon,iT}T/(T-1)\end{aligned} \tag{128}$$

$$\begin{aligned}\nabla_{\tau}\hat{\sigma}_{\varepsilon,iT}(\theta) &= \frac{2}{T-1}\sum_{t=1}^{T}(\tilde{D}_{i,t-1})\varepsilon_{i,t}(\theta) \\ \mathbb{E}[\nabla_{\tau}\hat{\sigma}_{\varepsilon,iT}(\theta)] &= 2\rho_2\Phi(\theta)\sigma^2_{\varepsilon,iT}T/(T-1)\end{aligned} \tag{129}$$

$$\nabla_{\rho_2}\hat{\sigma}_{\varepsilon,iT}(\theta) = 0, \tag{130}$$

$$\nabla_{\rho_1}\hat{\sigma}_{u,iT}(\theta) = 0, \tag{131}$$

$$\nabla_{\tau}\hat{\sigma}_{u,iT}(\theta) = 0, \tag{132}$$

$$\nabla_{\rho_2}\hat{\sigma}_{u,iT}(\theta) = \frac{2}{T-1}\sum_{t=1}^{T}(\tilde{Y}_{i,t-1})u_{i,t}(\theta) \quad = 2\tau\Phi(\theta)\sigma^2_{u,iT}T/(T-1) \tag{133}$$

C.2 GMM Normality

C.2.1 Consistency

Given the assumptions of Theorem 4.2, it follows by Theorem 14.1 from Wooldridge [2010] that the GMM estimator is consistent.

Theorem 14.1 from Wooldridge [2010] requires that a) $\Theta \subset \mathbb{R}^3$ is compact, which is satisfied by assumption b) For each $\theta \in \Theta$, $\mathbf{m}^{DBC}(\cdot, \theta)$ is Borel measurable on $\mathcal{Z}$, which is satisfied by the moments being a polynomial in data and having compact parameters c) for each $\mathbf{z} \in \mathcal{Z}$, $\mathbf{m}^{DBC}(\mathbf{z}, \cdot)$ is continuous on Θ. This follows from the fact moments are a polynomial in the random variables and $1/(1-\phi) < 1/\delta_s$ d) $|\mathbf{m}_j^{DBC}(\mathbf{z}, \theta)| \leqslant b(\mathbf{z})$ for all $\theta \in \Theta$ and $j = 1, 2, 3$ where $b(\cdot)$ is a nonnegative function on $\mathcal{Z}$ such that $E[b(\mathbf{z})] < \infty$, this follows from the assumption on the bounded moments of the data and that the moments on only quadratic in the random variables e) The GMM weighting matrix $\hat{\Xi} \xrightarrow{p} \Xi_0$ a positive definite weighting matrix, in my case the weighting matrix is just the identity matrix so this is satisfied, f) θ_0 is the unique solution to the problem, which the global identification condition that is satisfied by assumption.

Therefore a random vector $\hat{\theta}$ exists that solves Equation (35) and $\hat{\theta} \xrightarrow{p} \theta_0$.

C.2.2 Normality

Given the assumptions of Theorem 4.2, it follows by Theorem 14.2 from Wooldridge [2010] that the GMM estimator is asymptotically normal.

Theorem 14.2 from Wooldridge [2010] requires a) θ_0 is in the interior of Θ, which is satisfied by

assumption b) $\mathbf{m}^{DBC}(\mathbf{z},\cdot)$ is continuously differentiable on the interior of Θ for all $\mathbf{z} \in \mathcal{Z}$, this holds because the moments are ratios of polynomials with denominators bounded away from zero by the stationarity assumption c) each element of $\mathbf{m}^{DBC}(\mathbf{z},\theta_0)$ has finite second moment, which follows from assumption bounding the moments of the data d) each element of $\nabla\mathbf{m}^{DBC}(\mathbf{z},\theta)$ is bounded in absolute value by a function $b(\mathbf{z})$, where $E[b(\mathbf{z})] < \infty$ because the derivative of the polynomial is also bounded e) $E[\nabla\mathbf{m}^{DBC}(\mathbf{z},\theta)]$ is full rank follows by assumption.

Therefore, it follows that the limiting distribution of $\hat{\theta}$ follows the equation given in Theorem 4.2.

$$\sqrt{N}(\hat{\theta}^{DBC} - \theta_0) \xrightarrow{d} N(0, [\Sigma_T + B_T(\theta_0)]^{-1} S_T(\theta_0) [\Sigma_T + B_T(\theta_0)]^{-1} \tag{134}$$

With $S_T(\theta_0) = \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_{iT}^{DBC} \mathbf{m}_{iT}^{DBC}$,

$$\Sigma_T = \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} \tilde{Y}_{i,t-1}^2 & \tilde{Y}_{i,t-1}\tilde{D}_{i,t} & 0 \\ \tilde{D}_{i,t}\tilde{Y}_{i,t-1} & \tilde{D}_{i,t}^2 & 0 \\ 0 & 0 & \tilde{Y}_{i,t-1}^2 \end{bmatrix} \tag{135}$$

$$B_T(\theta) = \begin{bmatrix} \nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1} & \rho_2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & -\tau\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 \\ \rho_2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & \rho_2^2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & -\Phi(\theta)\sigma_{\varepsilon,iT}^2 - \tau\nabla\Phi(\theta)\sigma_{\varepsilon,iT}^2 \\ -\tau\nabla\Phi(\theta)\sigma_{u,iT}^2 & -\Phi(\theta)\sigma_{u,iT}^2 - \tau\nabla\Phi(\theta)\sigma_{u,iT}^2 & \tau^2(\nabla\Phi(\theta)\sigma_{u,i}^2 - \frac{2\Phi(\theta)^2\sigma_{u,i}^2 T}{T-1}) \end{bmatrix} \tag{136}$$

C.3 Local Identification

Let $\theta_0 \in \Theta \subset \mathbb{R}^p$ be the true parameter value. Suppose we have a set of moment conditions $\mathbb{E}[m^{DBC}(\theta)] = 0$, where $m : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$ is a vector-valued function. The parameter θ_0 is locally identified if the following conditions hold:

1. **Continuity and Differentiability**:
   - The moment condition function $m^{DBC}(\theta)$ is continuously differentiable with respect to θ in a neighborhood of θ_0.

2. **Rank Condition**:

- The Jacobian matrix $J_m^{DBC}(\theta)$ of the expected moment condition function with respect to θ, $J_m^{DBC}(\theta) = \frac{\partial}{\partial\theta}\mathbb{E}[m^{DBC}(\theta)]\big|_{\theta=\theta_0}$, has full column rank p (where p is the number of parameters).

Recall that the bias-corrected moments are the orginal OLS moment conditions minus the bias correction.

$$\begin{aligned}\mathbf{m}^{DBC}(\theta) := \mathbf{m}(\theta) - \mathbf{b}(\theta) &= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t} - b_{\rho_1,iT}(\theta) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t} - b_{\tau,iT}(\theta) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{u}_{i,t} - b_{\rho_2,iT}(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}(\theta) - \frac{-\hat{\sigma}_{\varepsilon,iT}(\theta)^2}{T}\left(\frac{T-1}{1-\phi(\theta)} - \frac{\phi(\theta)-\phi(\theta)^T}{(1-\phi(\theta))^2}\right) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}\tilde{\varepsilon}_{i,t}(\theta) - \rho_2\frac{-\hat{\sigma}_{\varepsilon,iT}(\theta)^2}{T}\left(\frac{T-1}{1-\phi(\theta)} - \frac{\phi(\theta)-\phi(\theta)^T}{(1-\phi(\theta))^2}\right) \\ \frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{u}_{i,t}(\theta) - \tau\frac{-\hat{\sigma}_{u,iT}(\theta)^2}{T^2}\left(\frac{T-1}{1-\phi(\theta)} - \frac{\phi(\theta)-\phi(\theta)^T}{(1-\phi(\theta))^2}\right) \end{bmatrix}\end{aligned} \tag{137}$$

In this section I provide conditions under which the Rank Condition holds. For the original OLS moment $\mathbf{m}(\theta)$ the corresponding Jacobian is $J_m(\theta) := \nabla_\theta \mathbf{m}(\theta)$. These are the original OLS moments, and therefore $\mathbb{E}[J_m(\theta)]$ is full rank when the columns of regressors are linearly independent. For the bias correction $J_b(\theta) := \nabla_\theta \mathbf{b}(\theta)$. Due to linearity I have $\mathbb{E}[J_m^{DBC}(\theta)] = \mathbb{E}[J_m(\theta)] - \mathbb{E}[J_b(\theta)]$. Therefore as long as columns of the 1) regressors are linearly independent, and 2) subtracting out $\mathbb{E}[J_b(\theta)]$ does not ruin the full rank of $\mathbb{E}[J_m(\theta)]$, then $\mathbb{E}[J_m^{DBC}(\theta)]$ has full column rank. $\mathbb{E}[J_b(\theta)]$ will not break the full rank condition as long as it is "small enough", formalized below.

Proof. In order to argue that $A - B$ is full rank, we need to argue that there does not exist an x such that $(A - B)x = 0$.

This is true if

$$x'(A - B)x > c\|x\|^2 \tag{138}$$

which is the same as saying

$$x'(A' - B')x > c\|x\|^2 \tag{139}$$

which is the same as saying

$$x'(A - \frac{B + B'}{2})x \geqslant c\|x\|^2 \tag{140}$$

this will hold if

$$\lambda_{\max}(\frac{B + B'}{2}) \leqslant \lambda_{\min}(A) \tag{141}$$

□

Therefore we have full rank if

$$\lambda_{\max}(\frac{\mathbb{E}[J_b(\theta)] + \mathbb{E}[J_b(\theta)]'}{2}) \leqslant \lambda_{\min}(\mathbb{E}[J_m^{DBC}(\theta)]) \tag{142}$$

$$J_m(\theta) = \begin{bmatrix} -\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}^2 & -\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}\tilde{D}_{i,t} & 0 \\ -\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}\tilde{Y}_{i,t-1} & -\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{i,t}^2 & 0 \\ 0 & 0 & -\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{i,t-1}^2 \end{bmatrix} \tag{143}$$

$$\mathbb{E}[J_b(\theta)] = \begin{bmatrix} \nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1} & \rho_2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & -\tau\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 \\ \rho_2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & \rho_2^2(\nabla\Phi(\theta)\sigma_{\varepsilon,i}^2 - \frac{2\Phi(\theta)^2\sigma_{\varepsilon,i}^2 T}{T-1}) & -\Phi(\theta)\sigma_{\varepsilon,iT}^2 - \tau\nabla\Phi(\theta)\sigma_{\varepsilon,iT}^2 \\ -\tau\nabla\Phi(\theta)\sigma_{u,iT}^2 & -\Phi(\theta)\sigma_{u,iT}^2 - \tau\nabla\Phi(\theta)\sigma_{u,iT}^2 & \tau^2(\nabla\Phi(\theta)\sigma_{u,i}^2 - \frac{2\Phi(\theta)^2\sigma_{u,i}^2 T}{T-1}) \end{bmatrix} \tag{144}$$

D Additional Simulations

D.1 Additional Comparison of Dynamic vs Nickell bias

Here I report additional simulation evidence extending the comparison between dynamic bias and Nickell bias from Section 5.1. Figure 5 summarizes the resulting treatment-effect estimates across a broad class of data-generating processes.

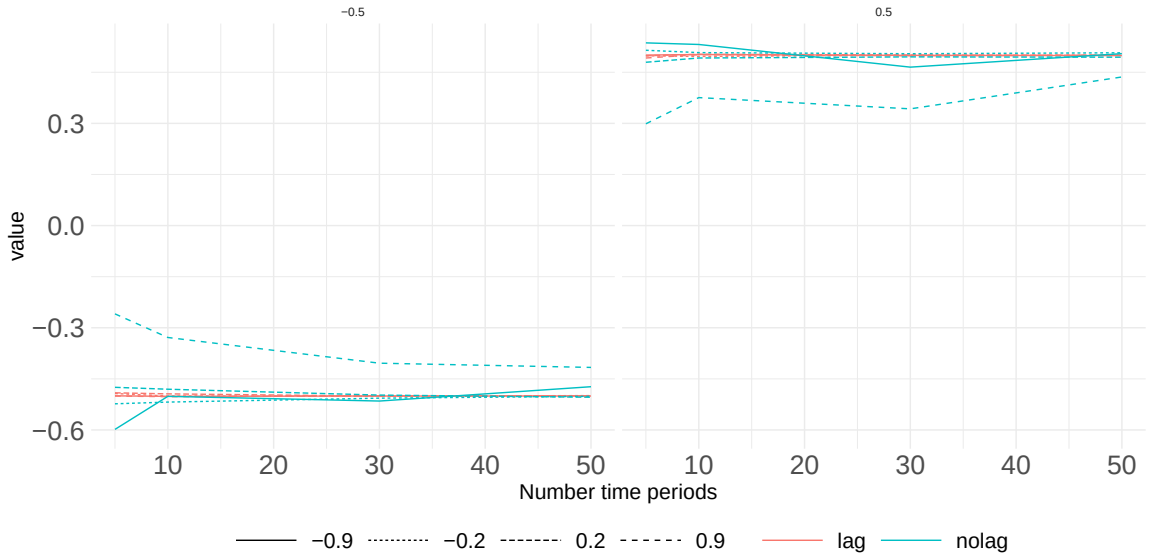

Figure 5: Line type represents different values of ρ_{10}. Line color represents whether the model was estimates with an outcome lag or not. The left column contains results when the true $\tau_0 = -.5$ and the right column contains results when $\tau_0 = .5$.

E Comparing Dynamic vs Nickell bias Analytically

This section covers a proof sketch comparing the two biases.

First I will calculate the Nickell bias for the OLS coefficent on treatment under Assumption (1). I will use FWL.

F Additional Empirical Results

F.1 Replicating Dell et al. [2012]

I replicate column 2 of Table 2 in Dell et al. [2012] and obtain the same results. My results are given in Table 6 column 3. In this Table 6, I also show in column 4 how the treatment effect changes once you control for lagged GDP Growth. The change the coefficient in Temperature x Poor Country changes 9% and is significantly different at the $p < .1$ level. I tested the difference in coefficients by using the clustered bootstrap.

Table 6

	Outcome variable: GDP Level		*Outcome variable:* GDP Growth	
	(1)	(2)	(3)	(4)
Temperature	118.568	82.071**	0.261	0.261
	(100.725)	(32.139)	(0.257)	(0.254)
Temperature x Poor Country	192.550	−88.364**	−1.655***	−1.511***
	(162.740)	(51.961)	(0.415)	(0.413)
Outcome Lag		0.922***		0.192***
		(0.005)		(0.015)
Controls	Yes	Yes	Yes	Yes
Observations	4,654	4,629	4,924	4,795
R^2	0.941	0.994	0.223	0.251
Adjusted R^2	0.935	0.993	0.150	0.179
Note:				*p<0.1; **p<0.05; ***p<0.01

F.2 Using growth vs levels in outcomes

Empirically, it is often observed that countries with lower initial GDP levels tend to have higher growth rates, which would suggest that GDP levels across countries should converge over time [Barro and Sala-i Martin, 1992]. By focusing on growth rates rather than levels, you might incorrectly suggest that these countries are diverging when, in fact, they could be converging in terms of absolute GDP levels. Using GDP growth as the dependent variable in regressions can lead to mis-

leading interpretations about the relative economic performance of countries, potentially suggesting divergence when, in reality, countries might be converging in terms of GDP levels.

Still, in practice, it is common to study GDP growth. I formalize the econometric problems that arise. Consider following simple model for country GDP. This comes from Solow [1956].

$$Y_{i,t} = a_i + \rho_0 Y_{i,t-1} + \tau_0 D_{i,t} + \varepsilon_{i,t} \tag{145}$$

Where we say treatment $D_{i,t}$ is temperature, and it is randomly assigned conditional on fixed effect. The error terms in both models are random idd shocks and $\varepsilon_{i,t}$ and $u_{i,t}$.

$$D_{i,t} = a_i + u_{i,t} \tag{146}$$

Let's say we want to run the regression using $\Delta Y_{i,t}$ as the outcome.

$$\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1} = \rho_0(Y_{i,t-1} - Y_{i,t-2}) + \tau_0(D_{i,t} - D_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \tag{147}$$

What is the causal object of interest? The impact of current temperature on growth.

$$\begin{aligned} \text{APD} &= \frac{\partial \Delta Y_{i,t}}{\partial D_{i,t}} \\ &= \frac{\partial \rho_0(Y_{i,t-1} - Y_{i,t-2}) + \tau_0(D_{i,t} - D_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1})}{\partial D_{i,t}} \\ &= \tau_0 \end{aligned} \tag{148}$$

The treatment effect is τ_0.

Consider the ordinary least squares (OLS) estimator of this effect.

$$\Delta Y_{i,t} = c_i + \beta D_{i,t} + e_{i,t} \tag{149}$$

Does $\hat{\beta}$ provide an unbiased estimate for τ_0? Recall that running OLS with fixed effects dummies is the same as running the within-transformed regression.

$$\Delta\tilde{Y}_{i,t} = \beta\tilde{D}_{i,t} + \tilde{e}_{i,t} \tag{150}$$

$$\hat{\beta} = \frac{Cov(\tilde{D}_{i,t}, \Delta\tilde{Y}_{i,t})}{Var(\tilde{D}_{i,t})} \tag{151}$$

$$\begin{aligned}
\frac{Cov(\tilde{D}_{i,t}, \Delta\tilde{Y}_{i,t})}{Var(\tilde{D}_{i,t})} &= \frac{Cov(\tilde{D}_{i,t}, \rho_0(\tilde{Y}_{i,t-1} - \tilde{Y}_{i,t-2}) + \tau_0(\tilde{D}_{i,t} - \tilde{D}_{i,t-1}) + (\tilde{\varepsilon}_{i,t} - \tilde{\varepsilon}_{i,t-1}))}{Var(\tilde{D}_{i,t})} \\
&= \frac{Cov(\tilde{D}_{i,t}, \rho_0(\tilde{Y}_{i,t-1} - \tilde{Y}_{i,t-2}) + \tau_0(\tilde{D}_{i,t} - \tilde{D}_{i,t-1}))}{Var(\tilde{D}_{i,t})} \\
&= \tau_0 + \underbrace{\frac{Cov(\tilde{D}_{i,t}, \rho_0(\tilde{Y}_{i,t-1} - \tilde{Y}_{i,t-2}) - \tau_0(\tilde{D}_{i,t-1}))}{Var(\tilde{D}_{i,t})}}_{\text{bias}}
\end{aligned} \tag{152}$$

This bias is therefore a form of dynamic bias. I refer to it as transformation bias, since it is induced by transforming the outcome prior to estimating the fixed-effects specification.

F.3 Replicating Annan and Schlenker [2015]

In this subsection, I replicate the findings of Annan and Schlenker [2015], which examine the impact of federal crop insurance subsidies on agricultural outcomes. The results are given in Table 7. The treatment variable of interest in this paper is "frac:ddayHot". This variable is the interaction of the insured fraction with exposure to hot days. In their original analysis, they used a static panel model that did not account for past outcomes. However, controlling for past outcomes is critical in settings where dynamic relationships are likely to exist, as previous outcomes can have a direct influence on current results.

In my replication, I modify their approach by introducing past outcomes into the model. The results show that the estimated treatment effects double when past outcomes are properly accounted for. This increase demonstrates the importance of controlling for dynamics in panel data settings. Failure to do so can lead to biased estimates.

Table 7

	Dependent variable:	
	yield_log	
	(1)	(2)
frac	0.0001	0.020
	(0.035)	(0.036)
ddayMod	0.430***	0.420***
	(0.017)	(0.017)
ddayHot	−0.619***	−0.633***
	(0.010)	(0.011)
prec	1.498***	1.560***
	(0.067)	(0.069)
prec2	−1.016***	−1.044***
	(0.049)	(0.051)
lag_Y_log		0.063***
		(0.004)
frac:ddayMod	0.008	0.004
	(0.013)	(0.013)
frac:ddayHot	0.046***	0.070***
	(0.015)	(0.015)
frac:prec	−0.153	−0.188*
	(0.106)	(0.108)
frac:prec2	0.142*	0.154*
	(0.079)	(0.080)
Observations	47,343	45,488
R^2	0.736	0.738
Adjusted R^2	0.725	0.727
Residual Std. Error	0.176 (df = 45447)	0.175 (df = 43638)
Note:		*p<0.1; **p<0.05; ***p<0.01

F.4 Endogenous Treatment

I repeat the same simulation exercise as in Section 2, but update the true model. The true model is now given in Equation (153). In this model I make treatment, $\text{Temp}_{i,t}$, a function of the past outcome, $\text{GDP}_{i,t-1}$. Therefore our model has a new parameter ρ_{20} which controls how much past GDP impacts temperature.[32]

$$\begin{aligned} \text{True Model with Endogenous Treatment: } \text{GDP}_{i,t} &= a_i + \tau_0 \text{Temp}_{i,t} + \rho_{10} \text{GDP}_{i,t-1} + \varepsilon_{i,t}, \\ \text{Temp}_{i,t} &= a_i + \rho_{20} \text{GDP}_{i,t-1} + u_{i,t}. \end{aligned} \tag{153}$$

For a simulation, I add just a little bit of endogeneity and set $\rho_{20} = .1$. The bias only gets larger the larger the absolute value of ρ_{20} is. Note in the plot, for the case of $\rho_{10} = .9$, the dynamic bias was so large it was omitted from the plot as it was much larger than all other biases. The plot of the bias is given in Figure 6.

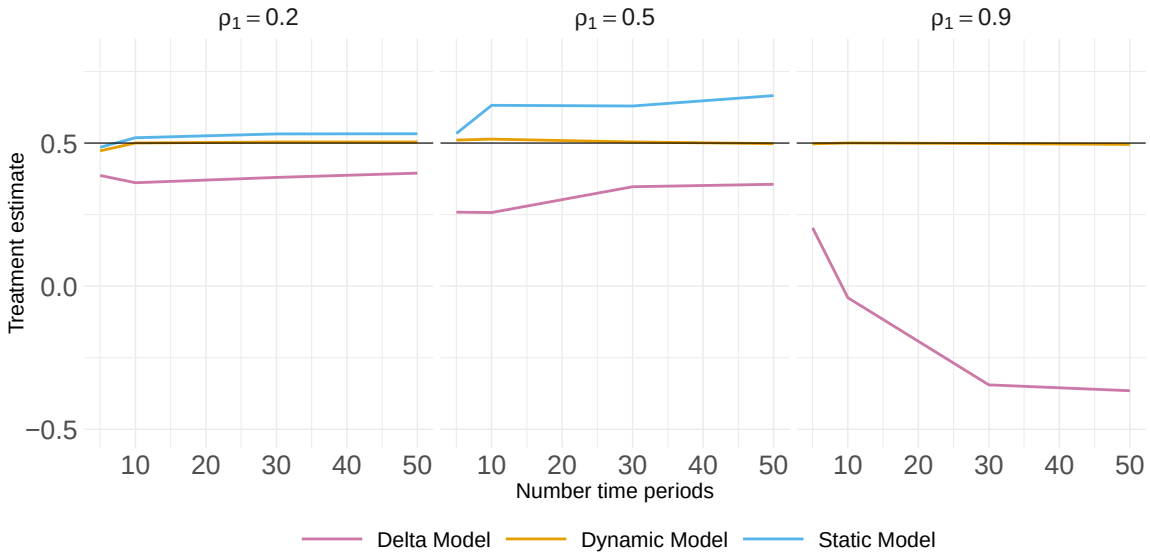

Figure 6: Bias of three different models.

[32] Most of the environmental literature does not think that past GDP impacts temperature, and I use setting mostly to illustrate my point on endogenous treatment. However, some papers discuss how economic growth, reflected in GDP, often correlates with increased industrial activity, energy consumption, and transportation. Historically, this has led to higher emissions of greenhouse gases (GHGs) such as carbon dioxide (CO2), which contribute to global warming [Nordhaus, 1992].

G Double Machine Learning Extension

This appendix outlines an extension of the DBC framework that accommodates high-dimensional controls and flexible heterogeneity while maintaining additive fixed effects. The key challenge is that standard machine-learning first-stage methods (e.g., Lasso) are not directly applicable in dynamic fixed-effects settings because Nickell bias contaminates first-stage estimation.

I therefore propose estimating nuisance components using Nickell-bias-corrected moment conditions with Regularized GMM, and then applying a DML-style orthogonal second stage for treatment effects. This appendix provides the setup and an implementable algorithmic outline; a full theoretical treatment is deferred.

G.1 Problem setup

Data

$$W_{i,t} = (Y_{i,t},\, Y_{i,t-1},\, D_{i,t},\, X_{i,t}), \qquad W_i = (W_{i,1},\, \ldots,\, W_{i,T}), \qquad \tilde{Y}_{i,t} = Y_{i,t} - \frac{1}{T}\sum_{s=1}^{T} Y_{i,s}.$$

Model setup

$$Y_{i,t} = a_i + \theta_{0,1} Y_{i,t-1} + \theta_{0,2} D_{i,t} + \theta_{0,3} X_{it} + \varepsilon_{i,t} \tag{154}$$

$$D_{i,t} = c_i + \theta_{0,4} Y_{i,t-1} + u_{i,t} \tag{155}$$

Let $\theta_{0,3}$ represent a high-dimensional component, while the endogenous parameters $\theta_{0,1}, \theta_{0,2}, \theta_{0,4}$ are low-dimensional.

The true parameter vector is $\theta_0 := (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \theta_{0,4})$.

The parameter of interest, $\theta_{0,2}$, corresponds to the treatment effect. The first equation represents the conditional expectation function (CEF), and the second equation represents the propensity function.

To implement Double Machine Learning (DML), I follow a two-stage process. In the first stage, I estimate the CEF and the propensity functions, which involves estimating the full parameter vector θ_0. This step is critical, as it reduces the high-dimensional problem to manageable components. I

rely on machine learning algorithms for flexible, consistent estimation of these nuisance parameters.

In the second stage, I use the first-stage estimates to construct a double-robust moment function. This moment is designed to be consistent, even if some of the first-stage estimates are slightly misspecified. The double-robust nature of DML ensures that the treatment effect $\theta_{0,2}$ can still be consistently estimated, provided that either the CEF or propensity function is estimated accurately.

Nickell bias prevents the consistent estimation of first stage using Lasso. To obtain consistent estimates, I use Nickell-bias-corrected moments with Regularized GMM (RGMM) instead of traditional Lasso. I outline the proof for the rates of RGMM in my setting to determine the rates of the first-stage function estimates.

Moment Conditions Note these are moments for one unit (average over time).

$$g_1(W_i, \theta) = \frac{1}{T} \sum_{i=1}^{T} \tilde{Y}_{i,t-1} e(\theta) - e(\theta)^2 \cdot C(\phi) \tag{156}$$

$$g_2(W_i, \theta) = \frac{1}{T} \sum_{i=1}^{T} \tilde{D}_{i,t} e(\theta) - e(\theta)^2 \cdot \theta_{0,4} \cdot C(\phi) \tag{157}$$

$$g_3(W_i, \theta) = \frac{1}{T} \sum_{i=1}^{T} \tilde{X}_{i,t} e(\theta) \tag{158}$$

$$g_4(W_i, \theta) = \frac{1}{T} \sum_{i=1}^{T} \tilde{Y}_{i,t-1} u(\theta) - u(\theta)^2 \cdot \theta_{0,2} \cdot C(\phi) \tag{159}$$

$$g(W_i, \theta) = [g_1(W_i, \theta), g_2(W_i, \theta), g_3(W_i, \theta), g_4(W_i, \theta)] \tag{160}$$

Where

$$e(\theta) = (\tilde{Y}_{i,t} - \theta_1 \tilde{Y}_{i,t-1} - \theta_2 \tilde{D}_{i,t} - \theta_3 \tilde{X}_{i,t}) \tag{161}$$

$$u(\theta) = (\tilde{D}_{i,t} - \theta_4 \tilde{Y}_{i,t-1}) \tag{162}$$

$$C(\phi) = \frac{1}{(1-\phi)T}(1 - \frac{1-\phi^T}{T(1-\phi)}) \tag{163}$$

$$\phi(\theta) = \theta_1 + \theta_2 \cdot \theta_4 \tag{164}$$

Averages Define the empirical moment average and population expectations as.

$$\hat{g}(\theta) = \frac{1}{N}\sum_{i=1}^{N} g(W_i, \theta), \qquad g(\theta) = \mathbb{E}[g(W_i, \theta)]\,. \tag{165}$$

G.2 RGMM

$$\min_{\theta=\Theta} \|\theta\|_1 : \|\hat{g}(\theta)\|_\infty \leqslant \lambda \tag{166}$$

Where λ is our regularization parameter. I only regularize the high dimensional parameter θ_3.

G.3 Rates for RGMM problem

I require sufficiently fast rates for my first-stage estimation. In this stage, I aim to estimate θ_0, as this allows me to construct consistent estimates of the conditional expectation function (CEF) and the propensity function.

My goal is to achieve a rate comparable to that of classical Lasso, but in the context of my Regularized GMM (RGMM) problem. Belloni et al. [2018] provide conditions and corresponding rates for RGMM when the moments possess an index structure. However, my moments do not naturally follow this structure.

To address this, I condition on a low-dimensional set of parameters, specifically θ_1, θ_2, and θ_4. After this conditioning, the conditional moment exhibits the required index structure. This allows me to apply the results from Belloni et al. [2018] to determine the rate for my high-dimensional parameter, θ_3, as a function of the low-dimensional parameters.

G.4 Steps

1. Create a grid with K points over the parameters θ_1, θ_2, and θ_4, which are the endogenous and low-dimensional parameters. For now, assume this grid is fixed (i.e., not increasing in size).

2. Select a point on the grid and fix the values of the parameters at this point. Denote this point as $(\theta_1^k, \theta_2^k, \theta_4^k)$.

3. Construct a new moment equation using these fixed values. This modified moment equation will still be a function of θ_3. For example, consider the moment equation for θ_1.

$$g_1(W_i, \theta_3) = \frac{1}{T}\sum_{i=1}^{T} \tilde{Y}_{i,t-1} e(\theta) - e(\theta)^2 \cdot \frac{1}{(1-\phi)T}(1 - \frac{1-\phi^T}{T(1-\phi)}) \tag{167}$$

   Where

$$e(\theta) = (\tilde{Y}_{i,t} - \theta_1^k \tilde{Y}_{i,t-1} - \theta_2^k \tilde{D}_{i,t} - \theta_3 \tilde{X}_{i,t}) \tag{168}$$

$$\phi = \theta_1^k + \theta_2^k \cdot \theta_4^k \tag{169}$$

4. Once θ_1^k, θ_2^k, and θ_4^k are fixed, the function ϕ becomes fixed and is treated as a constant, denoted C_ϕ^k.

$$g_1(W_i, \theta_3) = \frac{1}{T}\sum_{i=1}^{T} \tilde{Y}_{i,t-1} e(\theta) - e(\theta)^2 \cdot C_\phi^k \tag{170}$$

5. This allows me to express the moment equation in index form.

$$g_1(W_i, \theta_3) = \tilde{m}(W_i, Z_{u(j)}(W)' v_{u(j)}) \tag{171}$$

$$\tilde{m}(W_{it}, Z_{u(j)}(W_{it})' v_{u(j)}) = \frac{1}{T}\sum_{i=1}^{T} \tilde{Y}_{i,t-1} Z_{u(j)}(W_{it})' v_{u(j)} - (Z_{u(j)}(W_{it})' v_{u(j)})^2 \cdot C_\phi^k \tag{172}$$

$$Z_{u(j)}(W_{it})' v_{u(j)} = \tilde{Y}_{i,t} - \theta_1^k \tilde{Y}_{i,t-1} - \theta_2^k \tilde{D}_{i,t} - \theta_3 \tilde{X}_{i,t} \tag{173}$$

$$Z_{u(j)}(W_{it})' = (\tilde{Y}_{i,t}, \tilde{Y}_{i,t-1}, \tilde{D}_{i,t}, \tilde{X}_{i,t}) \tag{174}$$

$$v_{u(j)} = (-1, \theta_1^k, \theta_2^k, \theta_3) \tag{175}$$

6. Now that the moment equation is in index form, I can apply the bounds on empirical error for non-linear RGMM from Belloni et al. [2018] to obtain L_2 rates.

7. Now, apply this bound to the problem at hand. Define $\hat{\theta}_3(\theta_1^k, \theta_2^k, \theta_4^k)$ as the parameter estimate obtained from the RGMM, using the moment equation with fixed values $(\theta_1^k, \theta_2^k, \theta_4^k)$ from the grid, as specified in Equation (167).

   I introduce the following notation: let $\hat{f}_3(\theta_1^k, \theta_2^k, \theta_4^k)$ represent a function that maps the points $(\theta_1^k, \theta_2^k, \theta_4^k)$ on the grid to the corresponding RGMM solution $\hat{\theta}_3(\theta_1^k, \theta_2^k, \theta_4^k)$. Let $f_{0,3}(\theta_1^k, \theta_2^k, \theta_4^k)$ denote the oracle version of this function. Note that this is distinct from $\theta_{0,3}$. Theorem 2.1 provides the bound for every point on the grid.

$$\|\hat{f}_3(\theta_1^k, \theta_2^k, \theta_4^k) - f_{0,3}(\theta_1^k, \theta_2^k, \theta_4^k)\|_2 \leqslant \frac{2(\tilde{\ell}_n + \ell_n) s^{1/2}}{\mu_n \sqrt{n}} \forall k \tag{176}$$

8. This holds for every point on the grid, so I take the maximum to establish the bounds.

$$\max_{k=1,2,\ldots,K} \|\hat{f}_3(\theta_1^k, \theta_2^k, \theta_4^k) - f_{0,3}(\theta_1^k, \theta_2^k, \theta_4^k)\|_2 \leqslant \max_{k=1,2,\ldots,K} \frac{2(\tilde{\ell}_n + \ell_n) s^{1/q}}{\mu_n \sqrt{n}} \tag{177}$$

   From this maximum bound, I can calculate an L_2 rate for estimating the function $\hat{f}_3$. For now, assume that $\hat{f}_3$ is estimated at a rate of $O_p(n^{-\alpha})$.

$$\|\hat{f}_3 - f_{0,3}\|_2 = O_p(n^{-\alpha}) \tag{178}$$

9. Now, I focus on the convergence of $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_4$, which can be solved through a low-dimensional GMM problem. Our moment conditions include the estimated $\hat{f}_3$ as an input. I introduce the following new moments.

$$g(W_i, \theta_1, \theta_2, \theta_4, \hat{f}_3) = \frac{1}{T}\sum_{i=1}^{T} \tilde{Y}_{i,t-1} e(\theta) - e(\theta)^2 \cdot \frac{1}{(1-\phi)T}(1 - \frac{1-\phi^T}{T(1-\phi)}) \tag{179}$$

Where

$$e(\theta) = (\tilde{Y}_{i,t} - \theta_1 \tilde{Y}_{i,t-1} - \theta_2 \tilde{D}_{i,t} - \hat{f}_3(\theta_1, \theta_2, \theta_4)\tilde{X}_{i,t}) \tag{180}$$

$$\phi = \theta_1 + \theta_2 \cdot \theta_4 \tag{181}$$

Let

$$\hat{g}_n(\theta, f) = \frac{1}{N}\sum_{i=1}^{n} g(W_i, \theta, f) \tag{182}$$

$$g_0(\theta, f) = \mathbb{E} g(W_i, \theta, f) \tag{183}$$

The goal is to show:

$$\|\hat{g}_n(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0)\| = O_p(n^{-1/2} + n^{-\alpha}) \tag{184}$$

One possible argument for this case:

(a) I break the term on left hand side of Equation (184) into two parts.

$$\begin{aligned} &= \hat{g}_n(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0) \\ &= \hat{g}_n(\hat{\theta}, \hat{f}) \pm g_0(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0) \\ &= \underbrace{\hat{g}_n(\hat{\theta}, \hat{f}) - g_0(\hat{\theta}, \hat{f})}_{\text{CF}} - \underbrace{g_0(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0)}_{\text{Taylor}} \end{aligned} \tag{185}$$

The CF term can be controlled with either cross fitting or empirical process theory.

The "Taylor" term can be controlled by taking taylor expansions around θ_0 and f_0.

(b) Taylor Term:

$$
\begin{aligned}
&= \underbrace{g_0(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0)}_{\text{Taylor}} \\
&= g_0(\hat{\theta}, \hat{f}) \pm g_0(\hat{\theta}, f_0) - g_0(\theta_0, f_0) \\
&= \underbrace{g_0(\hat{\theta}, \hat{f}) - g_0(\hat{\theta}, f_0)}_{Term1} + \underbrace{g_0(\hat{\theta}, f_0) - g_0(\theta_0, f_0)}_{Term2}
\end{aligned} \tag{186}
$$

I perform a Taylor expansion of the first part of Term 1 around f_0, using a functional analog of a Taylor expansion [Wheeler]. Let $D_f(\hat{\theta}, f_0)$ represent the Jacobian matrix of the moment function g_0 with respect to the function f.

$$
\underbrace{g_0(\hat{\theta}, \hat{f}) - g_0(\hat{\theta}, f_0)}_{Term1} \approx g_0(\hat{\theta}, f_0) + D_f g(\hat{\theta}, f_0)(f - f_0) - g_0(\hat{\theta}, f_0) \tag{187}
$$

Let $D_\theta \hat{g}(\theta_0, f_0)$ denote the Jacobian matrix of the moment function $g_0(\theta_0, f_0)$ with respect to the parameter vector θ.

$$
g_0(\hat{\theta}, f_0) \approx g_0(\theta_0, f_0) + D_\theta g_0(\theta_0, f_0)(\hat{\theta} - \theta_0) \tag{188}
$$

$$
\underbrace{g_0(\hat{\theta}, f_0) - g_0(\theta_0, f_0)}_{Term2} \approx g_0(\theta_0, f_0) + D_\theta g_0(\theta_0, f_0)(\hat{\theta} - \theta_0) - g_0(\theta_0, f_0) \tag{189}
$$

I rewrite Equation (186) plugging in the expansions below.

$$
\underbrace{g_0(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0)}_{\text{Taylor}} \approx D_f g(\hat{\theta}, f_0)(f - f_0) + D_\theta g_0(\theta_0, f_0)(\hat{\theta} - \theta_0) \tag{190}
$$

I determine the rate of the LHS by substituting the rates of the two terms on the RHS. We have $|\hat{f} - f_0|_2 = O_p(n^{-\alpha})$. Since $g_0(\hat{\theta}, f0)$ is smooth and $Df\hat{g}(\hat{\theta}, f0)$ is a bounded linear operator, it follows that $|Df g_0(\hat{\theta}, f_0)(\hat{f} - f_0)|_2 = O_p(n^{-\alpha})$.

Conclude with Equation (191).

$$\|\hat{g}_n(\hat{\theta}, \hat{f}) - g_0(\theta_0, f_0)\| = O_p(n^{-1/2} + n^{-\alpha}) \tag{191}$$

10. Conclude with a convergence rate for the parameter estimates.

    I would want to show:

$$\|(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) - (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \theta_{0,4})\|_2 = O_p(n^{-\alpha}) \tag{192}$$

    where

$$\hat{\theta}_3 = \hat{f}_3(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_4) \tag{193}$$

    Given the smoothness of the moment conditions and the invertibility of the Jacobian, the goal is to apply the Delta method to transfer the rate of convergence from the sample moments to the parameter estimates.

H Absorbing Treatment

In this appendix I consider an empirically relevant setting with an *absorbing* treatment path, where each unit adopts once at time T_{start} and remains treated thereafter. Specifically, the regressor takes the form $D_i \cdot \mathbf{1}\{t > T_{\text{start}}\}$. I then redo the bias calculation for this absorbing regressor by explicitly computing its within-transformed mean and the induced correlation between the within-transformed regressor and the within-transformed error, which yields a closed-form bias term $b_4(\theta_0)$. The key implication is that, because treatment switches on only once, the relevant correlation is driven only by the error around $T_{\text{start}} - 1$, so the resulting Nickell-type bias is much smaller in the absorbing-adoption case.

Consider the following data generating process:

$$Y_{i,t} = a_i + \tau(D_i \times 1_{t>T_{start}}) + \varepsilon_{i,t} \tag{194}$$

$$D_i = a_i + \rho_2 Y_{i,T_{start}-1} + u_i \tag{195}$$

I label the bias in this model $b_4(\theta_0)$. I follow similar steps as above. To do this I need to calculate the formula for $D_i \times \overline{1_{t>T_{start}}}$.

$$\begin{aligned} D_i \times \overline{1_{t>T_{start}}} &= \frac{1}{T}\sum_{t=1}^{T}(a_i + \rho_2 Y_{i,T_{start}-1} + u_i)(1_{t>T_{start}}) \\ &= \frac{1}{T}\sum_{t=T_{start}}^{T}(a_i + \rho_2 Y_{i,T_{start}-1} + u_i) \\ &= \frac{T - T_{start}}{T}(a_i + \rho_2 Y_{i,T_{start}-1} + u_i) \\ &= \frac{T - T_{start}}{T}(a_i + \rho_2(a_i + \varepsilon_{T_{start}-1}) + u_i) \end{aligned} \tag{196}$$

$$b_4(\theta_0) = \operatorname*{plim}_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} D_i \times \widetilde{1_{t>T_{start}}}\tilde{\varepsilon}_{i,t}\right) \tag{197}$$

$$= -\operatorname*{plim}_{N\to\infty}\left(\frac{T}{N}\sum_{i=1}^{N} D_i \times \overline{1_{t>T_{start}}}\bar{\varepsilon}_i\right) \tag{198}$$

$$= -\operatorname*{plim}_{N\to\infty}\left(\frac{T}{N}\sum_{i=1}^{N} \frac{T - T_{start}}{T}(a_i + \rho_2(a_i + \varepsilon_{T_{start}-1}) + u_i)\bar{\varepsilon}_i\right) \tag{199}$$

$$= -\frac{1}{T}(T - T_{start} - 1)\rho_2\sigma_{\varepsilon}^2 \tag{200}$$

This formula is very intuitive. It is only error in the time period $T_{start}-1$ that is causing correlation in the errors. So the Nickell bias is much smaller in this case.

This is particularly interesting because past outcomes are typically viewed as time-varying covariates that are not absorbed by the individual fixed effect. However, in the case where the treatment is absorbed, the past outcome used for selection becomes fixed. As a result, the history remains constant over time and is controlled for by the fixed effect. Although there is still bias due to

violation of strict exogenity, it is not the standard OVB (omitted variable bias) type.

I Long-Run Effect: Definition, Intuition, and Why the Static FE Model Does Not Estimate It

I.1 What is long-run effect

Consider the dynamic outcome model I work with in the paper:

$$Y_{i,t} = \alpha_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}. \tag{201}$$

Two causal targets are natural:

- **Short-run (contemporaneous) effect.**

  The immediate effect of a one-unit increase in today's treatment holding the rest of the path fixed.

$$\tau_0 = \mathbb{E}\left[\frac{\partial Y_{i,t}(D_{i,t})}{\partial D_{i,t}}\right]. \tag{202}$$

- **Long-run (steady-state) effect.** The eventual change in the level of Y when treatment is permanently raised by one unit from period t onwards, i.e., the sum of the impulse responses that today's treatment sets in motion through outcome dynamics.

$$\tau^{LR} = \tau_0\left(1 + \rho_{10} + \rho_{10}^2 + \cdots\right) = \frac{\tau_0}{1-\rho_{10}} \quad \text{for} \quad |\rho_{10}| < 1. \tag{203}$$

As the paper shows running the Static Model does not recover the short-run or the long-run effect. Some researchers think that the Static Model approximates the long-run effect due to the following logic.

If you start from the dynamic-outcome model, iterating backwards m times the defintion of lag-Y gives a distributed representation in past D's.

$$Y_{i,t} = \alpha_i \sum_{j=0}^{m-1} \rho_{10}^j + \tau_0 \sum_{j=0}^{m-1} \rho_{10}^j D_{i,t-j} + \sum_{j=0}^{m-1} \rho_{10}^j \varepsilon_{i,t-j} + \rho_{10}^m Y_{i,t-1}. \tag{204}$$

Then the idea is if you take $m \to \infty$ then $\sum_{j=0}^{\infty} \rho_{10}^j = \frac{1}{1-\rho_{10}}$ and $\rho_{10}^m Y_{i,t-1} \to 0$, and so terms that look like the long-treatment treatment arise.

J Analytical Comparison of Nickell and Dynamic Bias

Consider the dynamic outcome model under Assumption (1).

$$Y_{i,t} = a_i + \tau_0 D_{i,t} + \rho_{10} Y_{i,t-1} + \varepsilon_{i,t}, \tag{205}$$

Consider the dynamic estimating equation corresponding to this data-generating process. The resulting coefficient estimates are subject to Nickell bias, and we therefore use the superscript NB. In this section, we apply the within transformation only to the right-hand-side regressors; by the Frisch–Waugh–Lovell theorem, this is equivalent to within-transforming all variables.

$$Y_{i,t} = \tau^{NB} \tilde{D}_{i,t} + \rho_1^{NL} \tilde{Y}_{i,t-1} + e_{i,t}, \tag{206}$$

The two-varriable OLS forumla give us the following expression for $\hat{\tau}^{NB}$ [Smith and Taylor, 2016].

$$\hat{\tau}^{NB} = \frac{\operatorname{Var}\left(\tilde{Y}_{i,t-1}\right) \operatorname{Cov}\left(\tilde{D}_{it}, Y_{it}\right) - \operatorname{Cov}\left(\tilde{D}_{it}, \tilde{Y}_{i,t-1}\right) \operatorname{Cov}\left(\tilde{Y}_{i,t-1}, Y_{it}\right)}{\operatorname{Var}\left(\tilde{D}_{it}\right) \operatorname{Var}\left(\tilde{Y}_{i,t-1}\right) - \operatorname{Cov}\left(\tilde{D}_{it}, \tilde{Y}_{i,t-1}\right)^2} \tag{207}$$

Dividing Equation (207) through by $\operatorname{Var}\left(\tilde{D}_{it}\right) \operatorname{Var}\left(\tilde{Y}_{i,t-1}\right)$ yeilds

$$\begin{aligned}\hat{\tau}^{NB} &= \frac{\left[\frac{\text{Cov}\left(\tilde{D}_{it},Y_{it}\right)}{\text{Var}\left(\tilde{D}_{it}\right)} - \frac{\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)}{\text{Var}\left(\tilde{D}_{it}\right)} \cdot \frac{\text{Cov}\left(\tilde{Y}_{i,t-1},Y_{it}\right)}{\text{Var}\left(\tilde{Y}_{i,t-1}\right)}\right]}{1 - \frac{\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)}{\text{Var}\left(\tilde{D}_{it}\right)\text{Var}\left(\tilde{Y}_{i,t-1}\right)}} \\ &= \frac{\left[\hat{\tau}^{D} - \frac{\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)}{\text{Var}\left(\tilde{D}_{it}\right)} \cdot \frac{\text{Cov}\left(\tilde{Y}_{i,t-1},Y_{it}\right)}{\text{Var}\left(\tilde{Y}_{i,t-1}\right)}\right]}{1 - \frac{\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)\text{Cov}\left(\tilde{D}_{it},\tilde{Y}_{i,t-1}\right)}{\text{Var}\left(\tilde{D}_{it}\right)\text{Var}\left(\tilde{Y}_{i,t-1}\right)}}\end{aligned} \tag{208}$$

Which gives us a clear expression comparing $\hat{\tau}^{NB}$ to $\hat{\tau}^{D}$. To make the comparison between $\hat{\tau}^{NB}$ and $\hat{\tau}^{D}$ more transparent, we re-express $\hat{\tau}^{NB}$ in terms of a small set of terms. Specifically, define:

$$V_D := \text{Var}\left(\tilde{D}_{it}\right), \qquad V_Y := \text{Var}\left(\tilde{Y}_{i,t-1}\right), \qquad C := \text{Cov}\left(\tilde{D}_{it},\, \tilde{Y}_{i,t-1}\right), \qquad \eta := \text{Cov}\left(\tilde{Y}_{i,t-1},\, \tilde{\varepsilon}_{it}\right).$$

The last term η can be thought of as a measure of "Nickell correlation". Positive AR(1) autocorrelation implies that $\eta \leqslant 0$.

Using this notation, we can rewrite the estimator from the static model $\hat{\tau}^{D}$ as follows. The equality between the first and second lines uses the assumption that the treatment is independent of ε.

$$\begin{aligned}\hat{\tau}^{D} &= \frac{\text{Cov}(\tilde{D}_{it},\tilde{Y}_{it})}{V_D} \\ &= \tau_0 + \rho_{10}\frac{C}{V_D}.\end{aligned} \tag{209}$$

Using this notation, we can express $\hat{\tau}^{NB}$ as

$$\hat{\tau}^{NB} = \frac{\hat{\tau}^{D} - \left(\frac{C}{V_D}\right)\left(\frac{\text{Cov}\left(\tilde{Y}_{i,t-1},\tilde{Y}_{it}\right)}{V_Y}\right)}{1 - \frac{C^2}{V_D V_Y}}. \tag{210}$$

Now expand the product in the numerator leading to:

$$\left(\frac{C}{V_D}\right)\left(\rho_{10} + \tau_0\frac{C}{V_Y} + \frac{\eta}{V_Y}\right) = \rho_{10}\frac{C}{V_D} + \tau_0\frac{C^2}{V_D V_Y} + \eta\frac{C}{V_D V_Y}.$$

Now also plugging in the definition of τ^D, the full numerator is:

$$\begin{aligned}
&= \tau_0 + \rho_{10}\frac{C}{V_D} - \left(\rho_{10}\frac{C}{V_D} + \tau_0\frac{C^2}{V_D V_Y} + \eta\frac{C}{V_D V_Y}\right) \\
&= \tau_0 - \tau_0\frac{C^2}{V_D V_Y} - \eta\frac{C}{V_D V_Y} \\
&= \tau_0\left(1 - \frac{C^2}{V_D V_Y}\right) - \eta\frac{C}{V_D V_Y}.
\end{aligned} \tag{211}$$

Therefore

$$\begin{aligned}
\hat{\tau}^{NB} &= \frac{\tau_0\left(1 - \frac{C^2}{V_D V_Y}\right) - \eta\frac{C}{V_D V_Y}}{1 - \frac{C^2}{V_D V_Y}} \\
&= \tau_0 - \frac{\eta C}{V_D V_Y - C^2}.
\end{aligned} \tag{212}$$

Now we can more easliy compare the two.

$$\begin{aligned}
\tau^{NB} - \tau^D &= \left(\tau_0 - \frac{\eta C}{V_D V_Y - C^2}\right) - \left(\tau_0 + \rho_{10}\frac{C}{V_D}\right). \\
&= -\rho_{10}\frac{C}{V_D} - \frac{\eta C}{V_D V_Y - C^2}.
\end{aligned} \tag{213}$$

In the case that autocrrelation and treamnet coefficents are possitive we have that Nickell and dynamic bias are both negative. Therefore dynamic bias is larger than Nickell bias when

$$\begin{aligned}
0 &< \tau^{NB} - \tau^D \\
&= -\rho_{10}\frac{C}{V_D} - \frac{\eta C}{V_D V_Y - C^2}.
\end{aligned} \tag{214}$$

Therefore a condition that ensures the dynamic bias is larger than Nickell bias is that:

$$\rho_{10}\left(V_Y - \frac{C^2}{V_D}\right) > (-\eta). \tag{215}$$

Looking at the right hand side, we see dynamic bias dominates when the lagged outcome still has variation after removing what D explains. This holds in the case of case of randomly assigned treatment. This is because the covariance between treatment and lagged Y, which is the C term,

is small because the covariance only arises from demeaning. That is even though treatment shocks are i.i.d., the within transformation creates a small mechanical correlation that is $O(1/T)$.

J.1 Rates of Decay: Dynamic Bias vs. Nickell Bias

Under Assumption (1), we next characterize the large-T order of the bias in the treatment-effect coefficient under both specifications: the static fixed-effects regression (which exhibits dynamic bias) and the dynamic fixed-effects regression (which exhibits Nickell bias). This yields rate statements for each bias term, which we then compare.

Recall from above we have that

$$\hat{\tau}^D - \tau_0 = \rho_{10} \frac{C}{V_D} \qquad \text{and} \qquad \hat{\tau}^{NB} - \tau_0 = -\frac{\eta C}{V_D V_Y - C^2}. \tag{216}$$

We next characterize the large-T orders of the terms appearing in these expressions. In particular, Lemma 4.1 implies that $\eta = O(T^{-1})$. Moreover, under random treatment assignment, we also have $C = O(T^{-1})$.

From the derivation of the Nickell correlation term,

$$b_{\rho_1}(\theta_0) = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{\varepsilon}_{i,t}\right], \tag{217}$$

it follows that,

$$\eta = -\frac{\sigma^2_{\varepsilon,i}}{T^2}\left(\frac{T-1}{1-\rho_{1,0}} - \frac{\rho_{1,0} - \rho^T_{1,0}}{(1-\rho_{1,0})^2}\right). \tag{218}$$

Under Assumption (1) , $\tilde{D}_{i,t} = \tilde{u}_{i,t}$ (because demeaning removes c_i). So

$$C = \mathrm{Cov}(\tilde{D}_{i,t}, \tilde{Y}_{i,t-1}) = \mathrm{Cov}(\tilde{u}_{i,t}, \tilde{Y}_{i,t-1}). \tag{219}$$

From Lemma 4.1's formula for

$$b_{\rho_2}(\theta_0) = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_{i,t-1}\tilde{u}_{i,t}\right], \tag{220}$$

and using that random treatment means $\rho_{2,0} = 0 \Rightarrow \phi(\theta_0) = \rho_{1,0}$, you get

$$C = -\frac{\tau_0 \sigma_{u,i}^2}{T^2} \left(\frac{T-1}{1-\rho_{1,0}} - \frac{\rho_{1,0} - \rho_{1,0}^T}{(1-\rho_{1,0})^2} \right). \tag{221}$$

Therefore, $\eta = O(T^{-1})$ and $C = O(T^{-1})$. We now use these rate results to analyze the bias in the treatment-effect objects of interest.

From Equation (209) we have that under Assumption (1) that for dynamic bias,

$$\hat{\tau}^D - \tau_0 = \rho_{10} \frac{C}{V_D}. \tag{222}$$

And $V_D = \text{Var}\left(\tilde{D}_{it}\right)$ is order 1. Therefore for dynamic bias we have that:

$$\hat{\tau}^D - \tau_0 = O(1/T). \tag{223}$$

We now carry out an analogous analysis for the treatment-effect coefficient in the dynamic fixed-effects regression, which is subject to Nickell bias.

$$\hat{\tau}^{NB} - \tau_0 = -\frac{\eta C}{V_D V_Y - C^2}. \tag{224}$$

We have that $\eta = O(1/T)$ and $C = O(1/T)$, therefore the product is $O(1/T^2)$. The denominator $V_D V_Y - C^2$ is order 1, so it does not change the rate. Therefore:

$$\hat{\tau}^{NB} - \tau_0 = O(1/T^2). \tag{225}$$

Therefore under Assumption (1), the dynamic bias in the treatment-effect coefficient decays more slowly than the Nickell leakage bias in the treatment-effect coefficient. This does not contradict the classic Nickell result: the Nickell bias in the lag coefficient ρ_1 is $O(1/T)$. However, the Nickell bias in the treatment-effect coefficient inherits an additional attenuation factor equal to the within correlation between $\tilde{D}_{it}$ and $\tilde{Y}_{i,t-1}$. Under Assumption (1), this correlation is itself $O(1/T)$, so the Nickell *leakage* into τ is $O(1/T^2)$.