

# SIMPLE TREATMENT EFFECT ESTIMATION IN FIXED- $T$ PANELS WITH INTERACTIVE FIXED EFFECTS\*

Nicholas Brown  
Queen's University

Kyle Butts  
University of Arkansas

Joakim Westerlund<sup>†</sup>  
Lund University  
and  
Deakin University

June 15, 2024

## Abstract

The present paper proposes a simple approach to treatment effects estimation in linear panel data models that is valid when the number of time periods is small and the parallel trends condition is violated due to the presence of interactive fixed effects. The procedure allows the covariates to be affected by treatment and enables separation of the part of the estimated treatment effect that is due to the covariates from the part that is not. The asymptotic properties of the procedure are established and their accuracy in small samples is investigated using Monte Carlo simulations. The procedure is illustrated using as an example the effect of increased trade competition on firm markups in China.

**JEL Classification:** C31, C33, C38.

**Keywords:** Treatments effects estimation; difference-in-differences; interactive fixed effects; common correlated effects.

---

\*The authors would like to thank Co-Editor Ivan Canay, an Associate Editor, and two anonymous referees for many valuable comments and suggestions. Westerlund would also like to thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship.

<sup>†</sup>Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

# 1 Introduction

Consider a panel data set comprised of  $N$  cross-sectional units and  $T$  time periods. Treatment effects studies based on this type of data typically estimate linear models with additive unit and time fixed effects. Such fixed effects are consistent with the so-called “parallel trends” assumption, which demands that the average outcomes for treated and untreated cross-sectional units must follow the same path over time in absence of treatment. Because this assumption is implausible in many settings, much effort has been spent trying to control for violations through the inclusion of covariates. A common approach estimates either time-varying slopes on time-invariant/‘baseline’ covariates or unit-varying slopes on covariates that only vary over time (see, for example, [Callaway and Karami, 2024](#), and [Kim and Oka, 2014](#)). The idea is that deviations from parallel trends can be modeled as a product of two terms; one captures the trend itself, while the other measures the extent to which the impact of this trend is equal, or parallel, across units. The problem is that many drivers of non-parallel trending are unobserved and lack good proxies.<sup>1</sup> The approach is also wasteful, in the sense that economists often have access to panel data on covariates that vary over both units and time, and that are likely more informative than covariates that vary only over one of the two dimensions. When these covariates are affected by treatment, including the covariates in the model will yield biased treatment effect estimates. Hence, while there are covariates available that could be used to control for non-parallel trends, they are rarely included since this is likely to do more damage than good (see [Caetano and Callaway, 2023](#), for a recent discussion).

The present paper is motivated by the observations made in the previous paragraph. The purpose is to develop a new approach to treatment effects estimation that allows for a broad range of non-parallel trends and treatment-affected covariates. As a starting point, we consider an “interactive” fixed effects specification, in which the unit fixed effects, also called “factor loadings”, interact multiplicatively with the time effects, also called “common fac-

---

<sup>1</sup>For example, in studies with time-varying-only covariates, deterministic linear and quadratic trends are often the only candidates even though there is plenty of evidence to suggest that they are not enough (see, for example, [Kim and Oka, 2014](#)).

tors”.<sup>2</sup> This specification contains the conventional additive two-way fixed effect model as a special case. It captures the idea that deviations from parallel trends can be modelled as macroeconomic shocks, or common factors, that impact each unit differently through specific multipliers, or loadings. The main advantage when compared to the usual approach of including covariates that are either unit- or time-invariant is that both factors and loadings can be treated as unknown.

The present paper is not the first to allow for interactive fixed effects in the treatment effects context; however, it is the first to do so while at the same time allowing for treatment-affected covariates. In fact, the new approach does not only allow for such treatment-affected covariates, but it enables researchers to separate the part of the treatment effect that is due to the covariates from the part that is not. As far as we are aware, there is currently no other treatment effects approach with this level of flexibility.

The new approach is not only flexible but is also very simple and user friendly. [Gobillon and Magnac \(2016\)](#), and [Xu \(2017\)](#) allow for interactive effects that are dealt with using versions of the principal components (PC)-based approach of [Bai \(2009\)](#). However, the PC approach is based on solving a non-convex optimization problem, which means that it is not only computationally costly but it can also be difficult to get to converge. Even if it does converge, it may not be to the global optimum (see [Moon and Weidner, 2019](#)). There are also generalized method of moments (GMM)-based approaches, such as those of [Callaway and Karami \(2024\)](#) and [Brown and Butts \(2023\)](#), but they require valid instruments that might not be available in practice. Both PC and GMM require that the number of unobserved factors can be accurately estimated, which is known to be difficult (see, for example, [Moon and Weidner, 2015](#), and [Breitung and Hansen, 2021](#)). Even if the estimator is accurate, the resulting PC and GMM estimators still suffer from “post-selection bias” (see [Leeb and Pötscher, 2005](#)). The new procedure uses only ordinary least squares (OLS) operations and there is no need for accurate estimation of the number of factors.

Another attractive feature of the new approach is that  $T$  does not have to be large. Initially

---

<sup>2</sup>When we refer to “interactive fixed effects”, we mean that both factors and factor loadings are treated as non-random parameters.

proposed as dimension reduction techniques, interactive fixed effects approaches tend to be “data-hungry”. A very common requirement is that both  $T$  and  $N$  are both large (see [Chan and Kwok, 2022](#), [Gobillon and Magnac, 2016](#), [Xu, 2017](#), and [Arkhangelsky et al., 2021](#)), which is a problem because in treatment effects studies  $T$  is often small (see [Bertrand et al., 2004](#), for a survey). The new procedure accounts for this smallness of  $T$  and is as a result widely applicable.

The attractiveness of the new approach is achieved in part by our novel use of the common correlated effects (CCE) approach of [Pesaran \(2006\)](#), which has a closed form, allows  $T$  to be fixed, and does not require estimation of the number of factors (see [Westerlund et al., 2019](#), for a discussion).<sup>3</sup> The “common-correlated effects” assumption requires the researcher to have access to a set of covariates that are impacted by the same set of common factors as the outcome variable. Using the information from these covariates (including the outcome variable) allows us to estimate the confounding shocks and isolate the causal effect of interest. The main requirement is that the number of factors cannot be larger than the number of observables that loads on those factors, which is a testable restriction.

The main object of interest is the average treatment effect on the treated (ATT), which is the average difference between the actual and counterfactual post-treatment outcomes of treated cross-sectional units. This average cannot be directly calculated since the counterfactual outcome is unobserved. We therefore use the CCE approach to estimate it. The proposed treatment effect CCE estimator, dubbed “TECCE”, is computed in three simple steps. We begin by estimating the common factors using cross-sectional averages of all the observables from the untreated sample. We then estimate the factor loadings by regressing the outcome variable for each cross-sectional unit on the first-step factor estimates. In the third and final step, we use the estimated interactive fixed effects, which is simply the product of the first- and second-step factor and loading estimates, to estimate the counterfactual outcomes. The estimated ATT is the average difference between the observed treated and estimated counterfactual outcomes.

---

<sup>3</sup>[Eberhardt \(2022\)](#) recently applied CCE in the treatment effects context to estimate the impact of democratisation on long-run growth. This other estimator is, however, materially different from ours, and with unknown finite sample and asymptotic properties.

The new estimator is shown to be consistent and asymptotically normal under general conditions provided only that the number of treated and untreated units is large, a result that is verified in finite samples by means of Monte Carlo simulations. This result is noteworthy because the TECCE estimation procedure described in the previous paragraph does not require explicit accounting of the covariates in steps two and three. In spite of this, consistency and asymptotic normality hold regardless of whether the covariates are affected by treatment. Researchers interested in studying the extent to which the covariates contributes to the ATT can do so by repeating steps two and three of the estimation procedure while conditioning on the observed covariates. This procedure gives an estimator of the part of the ATT that is not due to the covariates, which can be subtracted off the initial ATT estimator to produce an estimator of the part of the ATT that is due to the covariates.

As an empirical illustration, we consider the effect of China's accession into the World Trade Organization (WTO) in 2001 on the dispersion of industry-level markups. Our results suggest that the increased competition generated by the accession lead to reduced markup dispersion, and that almost half of this reduction was brought about by lower marginal cost dispersion.

The rest of the paper is structured as follows. Section 2 presents the model and defines the ATT, the estimation of which is the concern of Section 3. Sections 4, 5 and 6 contain the asymptotic, Monte Carlo and empirical studies, respectively. Section 7 concludes. All proofs are relegated to the online appendix.

## 2 The model

We are interested in estimating the ATT of a particular treatment on some outcome variable  $y_{i,t}$ , observable for  $i \in \{1, \dots, N\}$  cross-sectional units and  $t \in \{1, \dots, T\}$  time periods. We allow for the possibility that the  $N$  units can be divided into groups within which treatment timing is the same. There are  $G < \infty$  such groups indexed by  $g \in \{1, \dots, G\}$ . Treated units never leave their groups but remain exposed for all periods after entering treatment; that is, treatment is of the "staggered" type. Untreated units are members of group  $g = 0$ .

The timing of the treatment could be viewed as the outcome of a random process that is dealt with through suitable conditioning. In this paper we instead follow [Borusyak et al. \(2024\)](#), and view the treatment timing as fixed, which has the advantage that it is semi-parametric because no assumption on the distribution of the treatment timing needs to be made. It is particularly well suited for applications where treatment timing is not explicitly randomized, as in our empirical illustration of Section 6. Let us therefore denote by  $g_i \in \{0, \dots, G\}$  a variable stating the group membership of cross-sectional unit  $i$ , and by  $\mathcal{I}_g := \{i : g_i = g \in \{0, \dots, G\}\} \subset \{1, \dots, N\}$  the set of cross-sectional units that are members of group  $g$ , where  $a := b$  means that  $a$  is defined by  $b$ . The set of untreated units is therefore denoted  $\mathcal{I}_0$ , and it is convenient to let  $\mathcal{I}_0^c := \{1, \dots, N\} \setminus \mathcal{I}_0$  denote the set of treated units. The number of cross-sectional units within group  $g$  is given by the cardinality  $N_g := |\mathcal{I}_g|$ . Exact conditions on  $N_0, \dots, N_G$  will be specified later. For now, we just assume that these quantities are “large”.<sup>4</sup> We denote  $T_0$  as the first period before any unit is treated, so that the first treatment starts in period  $T_0 + 1$ . Unlike  $N_0, \dots, N_G$ , the number of pre-treatment periods  $T_0$  does not have to be large, provided that it is larger than the number of estimated factors. We describe this condition later. The start of the treatment of group  $g > 0$  is denoted  $t_g$ .

Denote by  $y_{i,t}(g)$  the “potential” outcome of cross-sectional unit  $i$  had it been member of group  $g$  in period  $t$ . The observed outcome for unit  $i$  at time  $t$  is given by  $y_{i,t} := y_{i,t}(g_i)$ . In this notation, the unit-specific treatment effect for a unit  $i$  that is member of treatment group  $g_i = g > 0$  in post-treatment periods  $t > T_0$  is given by

$$\Delta_{i,t} := y_{i,t} - y_{i,t}(0) \tag{1}$$

(see [Borusyak et al., 2024](#), for a similar definition). Because we do not observe  $y_{i,t}(0)$  for treated units in post-treatment periods,  $\Delta_{i,t}$  must be treated as unknown and estimated from the data. Of course, because  $T$  is fixed, accurate estimation of  $\Delta_{i,t}$  itself is impossible. The

---

<sup>4</sup>Hence, as usual, we assume that there are both treated and untreated units in the sample. The condition that  $N_0, \dots, N_G$  are all large is also very common, although it is not always articulated in the same way as here. Many studies assume the probability of treatment is strictly positive (see, for example, [Abadie, 2005](#), [Callaway and Sant’Anna, 2021](#), and [Sant’Anna and Zhao, 2020](#), for discussions), which in the current fixed treatment allocation context is tantamount to requiring that  $N_g/N \rightarrow \delta > 0$  as  $N_g, N \rightarrow \infty$ . [Chan and Kwok \(2022\)](#) estimate treatment effects for each cross-sectional unit in the treated group using only the time series variation. While this estimator do not require a large number of treated units, their pooled estimator does.

object of interest is therefore not  $\Delta_{i,t}$  but the ATT,  $\mathbb{E}(\Delta_{i,t})$ , which we now characterize.

**Assumption 1.**  $\Delta_{i,t} = \text{ATT}_{g,t} + v_{i,t}$ , where  $\text{ATT}_{g,t}$  is non-random and  $v_{i,t}$  is a mean zero random error that is independent of all other random elements of the model.

Assumption 1 implies that  $\mathbb{E}(\Delta_{i,t}) = \text{ATT}_{g,t}$ , which is the same “group-time” ATT considered by [Callaway and Sant’Anna \(2021\)](#) (see also [Callaway and Karami, 2024](#), and [Xu, 2017](#)).<sup>5</sup> Except for the constancy-within-groups condition, the ATT is unrestricted, which means that it is allowed to vary freely over both groups and time. One implication of this is that the effect of the treatment need not take place abruptly but can be gradual in nature. This is in contrast to existing studies that typically make quite strong assumptions about the variability of the ATT. In particular, it is common to assume that the ATT is time-invariant (see [Abadie, 2005](#), [Chan and Kwok, 2022](#), and [Gobillon and Magnac, 2016](#), to mention a few).<sup>6</sup>

**Assumption 2.**  $y_{i,t} = y_{i,t}(0)$  for  $t \leq T_0$ .

Assumption 2 is similar to Assumption 11 in [Callaway and Karami \(2024\)](#), and Assumption 3 in [Callaway and Sant’Anna \(2021\)](#). It requires that there is no treatment effect before any unit receives treatment. Once one group enters treatment, however, anticipation is not ruled out for other groups. This condition seems reasonable because before the first group enters treatment, the units might not know what to expect. Once treatment roll-out has commenced, however, it is possible that not-yet-treated units learn from the treated.

We typically do not observe  $y_{i,t}$  in isolation but together with covariates whose outcome may again depend on treatment status. In our empirical illustration, we estimate the effect of China’s accession into WTO on industry-level markup dispersion while controlling for the dispersion in marginal costs. According to theory, increased competition may affect markups directly through changes in market structure and indirectly through changes

---

<sup>5</sup>Most studies in the literature do not use random coefficient conditions like Assumption 1 (see [Chan and Kwok, 2022](#), and [Gobillon and Magnac, 2016](#), for exceptions), but state their conditions directly in terms of the (conditional) expectations of those random coefficients.

<sup>6</sup>[Chan and Kwok \(2022\)](#) assume that both  $N$  and  $T$  are large. This allows them to estimate unit-specific treatment effects, which is not possible when  $T$  is fixed.

in marginal costs. We therefore introduce the  $m \times 1$  vector  $\mathbf{x}_{i,t}(g)$ , whose realized value is given by  $\mathbf{x}_{i,t} := \mathbf{x}_{i,t}(g_i)$ .<sup>7</sup>

**Assumption 3.**

$$y_{i,t}(0) = \beta_i' \mathbf{x}_{i,t}(0) + \alpha_i' \mathbf{f}_t + \varepsilon_{i,t},$$

where  $\beta_i$  is a  $m \times 1$  vector of slope coefficients,  $\mathbf{f}_t$  is a  $r \times 1$  vector of unobservable common factors,  $\alpha_i$  is a  $r \times 1$  vector of factor loadings, and  $\varepsilon_{i,t}$  is a mean zero random error. Both  $\mathbf{f}_t$  and  $\alpha_i$  are assumed to be non-random.

Assumption 3 specifies  $y_{i,t}(0)$  as a linear function of  $\mathbf{x}_{i,t}(0)$ , which is a very common requirement in the empirical literature. There are some exceptions based on non-parametric approaches (see, for example, [Abadie, 2005](#), [Callaway and Sant'Anna, 2021](#), and [Sant'Anna and Zhao, 2020](#)). However, these suffer from the “curse of dimensionality” problem, the solution of which typically involves imposing additional structure, and still the small-sample properties can be unacceptably poor unless sample sizes are very large. Non-parametric approaches typically also place strong distributional assumptions directly on the observed data. A very common condition is that  $y_{i,t}$  and  $\mathbf{x}_{i,t}$  are independently and identically distributed (iid) over  $i$ , which is unrealistic.

The interactive fixed effects are given by  $\alpha_i' \mathbf{f}_t$ . The standard approach in the treatment effects literature is to include additive time and unit fixed effects (see, for example, [Caetano and Callaway, 2023](#), and [Callaway and Karami, 2024](#), for recent discussions). Such effects are nested within our interactive specification, as is clear from noting that  $\alpha_i' \mathbf{f}_t = \eta_i + \theta_t$  in the special case when  $\mathbf{f}_t = [1, \theta_t]'$  and  $\alpha_i = [\eta_i, 1]'$ . A major advantage of the interactive specification when compared to the additive one is that it accommodates violations of the parallel trends condition. Indeed, unless  $\mathbf{f}_t = \mathbf{f}$  for all  $t$ , trends will not be parallel unless  $\alpha_i \neq \alpha_j$  for all pairs  $i$  and  $j$ .<sup>8</sup> Hence, by leaving  $\mathbf{f}_t$  and  $\alpha_i$  unrestricted, we can accommodate very general

<sup>7</sup>Note how  $\mathbf{x}_{i,t} = \mathbf{x}_{i,t}(0)$  for  $t \leq T_0$  is a necessary condition for Assumption 2. We therefore require that  $\mathbf{x}_{i,t}$  satisfies the same anticipation condition as  $y_{i,t}$ .

<sup>8</sup>The parallel trends condition requires that  $\mathbb{E}[\Delta y_{i,t}(0) | \Delta \mathbf{x}_{i,t}(0)] = \mathbb{E}[\Delta y_{j,t}(0) | \Delta \mathbf{x}_{j,t}(0)]$  for all  $i \in \mathcal{I}_0$ ,  $j \in \mathcal{I}_0^c$  and  $t > T_0$ , which in terms of the model in Assumption 1 reads  $\beta_i' \Delta \mathbf{x}_{i,t}(0) + \alpha_i' \Delta \mathbf{f}_t = \beta_j' \Delta \mathbf{x}_{j,t}(0) + \alpha_j' \Delta \mathbf{f}_t$  (see, for example, [Callaway and Sant'Anna, 2021](#), or [Chan and Kwok, 2022](#)). A necessary condition for this last equality to hold is that  $\alpha_i = \alpha_j$  for  $i \in \mathcal{I}_0$  and  $j \in \mathcal{I}_0^c$ , which in turn requires  $\alpha_i = \alpha_j$  for all  $(i, j) \in \{1, \dots, N\}$ .



forms of non-parallel behaviour. In fact, the interactive effects specification considered here is general even when compared to other specifications of the same type, which typically assume that either  $\mathbf{f}_t$  or  $\alpha_i$ , or both, follow certain probability laws (see, for example, [Chan and Kwok, 2022](#), [Gobillon and Magnac, 2016](#), and [Xu, 2017](#)).<sup>9</sup>

#### Assumption 4.

$$\mathbf{x}_{i,t}(0) = \lambda_i' \mathbf{f}_t + \mathbf{v}_{i,t},$$

where  $\lambda_i$  is a  $r \times m$  matrix of non-random factor loadings and  $\mathbf{v}_{i,t}$  is a  $m \times 1$  vector of mean zero errors that are independent of  $\varepsilon_{i,t}$ .

Assumption 4 allow  $\mathbf{x}_{i,t}(0)$  to load on the same set of factors as  $y_{i,t}(0)$ , which means that it may be endogenous. Hence, in contrast to much of the previous literature, here treatment is not the only source of endogeneity (see [Caetano and Callaway, 2023](#), for a discussion). The common dependence on  $\mathbf{f}_t$  is consistent with the empirical observation that many variables are affected by the same common shocks (see, for example, [Westerlund et al., 2019](#)). The factors need not be the same, though, as loadings may be zero. There might therefore be factors that are unique to  $\mathbf{x}_{i,t}(0)$  and/or  $y_{i,t}(0)$ . The condition that  $\mathbf{x}_{i,t}(0)$  loads linearly on  $\mathbf{f}_t$  rules out specifications featuring, for example, dummy variables, powers or products of the covariates. However, additional covariates can be easily accommodated provided that they satisfy standard exogeneity conditions, as we explain in the online appendix.<sup>10</sup>

Most papers in the literature do not place any parametric assumptions on the covariates.<sup>11</sup> However, they assume instead that the covariates are exogenous (to treatment) and iid over the cross-section (see [Abadie, 2005](#), [Callaway and Sant'Anna, 2021](#), [Callaway and Karami, 2024](#), and [Caetano and Callaway, 2023](#), to mention a few), which is more restrictive than Assumption 4. In fact, the only studies that come close to ours in terms of the generality of

---

<sup>9</sup>An example of a common assumption is that  $\mathbf{f}_t$  is generated from a stationary stochastic process (see, for example, [Gobillon and Magnac, 2016](#), and [Xu, 2017](#)), which is likely restrictive in the present context because it rules out factors that are, for example, breaking or trending.

<sup>10</sup>If there are no covariates available, we define  $\beta_i' \mathbf{x}_{i,t}(0) := 0$  in Assumption 3. In this case, Assumption 4 is obviously not needed.

<sup>11</sup>One exception is given by [Caetano and Callaway \(2023\)](#), which is also one of the few studies in the literature to allow for treatment-affected covariates.

the allowable covariates are [Gobillon and Magnac \(2016\)](#), and [Xu \(2017\)](#). They allow the covariates to be arbitrarily correlated with the interactive fixed effects, which represents a more general consideration than Assumption 4. However, they require instead that the covariates are stationary and unaffected by treatment, which is not required here.<sup>12</sup>

[Caetano and Callaway \(2023\)](#) is one of the few papers in the literature to allow for covariates that are affected by treatment. They focus on the ATT; however, they also say that “it would be interesting to extend our arguments to additionally identifying direct and indirect effects of participating in the treatment” (page 6). We provide such an extension. In order to separate the part of the ATT that is due to changing values of the covariates from the part that is not, we define

$$\tau_{i,t} := \mathbf{x}_{i,t} - \mathbf{x}_{i,t}(0) \quad (2)$$

for a unit  $i$  that is member of group  $g_i = g > 0$  in post-treatment periods  $t > T_0$ . In view of Assumption 3, the covariates’ contribution to the unit-specific treatment effect on  $y_{i,t}$  is given by  $\beta'_i[\mathbf{x}_{i,t}(g) - \mathbf{x}_{i,t}(0)] = \beta'_i\tau_{i,t}$ . We also define the difference between these effects;

$$\eta_{i,t} := \Delta_{i,t} - \beta'_i\tau_{i,t}. \quad (3)$$

In the terminology of the mediation literature (see, for example, [Huber, 2014](#)),  $\eta_{i,t}$  is the “direct” effect of treatment and  $\beta'_i\tau_{i,t}$  is the “indirect” effect of treatment mediated through the covariates.

The random coefficient condition in Assumption 1 is enough if the purpose is to just estimate the ATT. If the purpose is to estimate also the direct and indirect ATTs, henceforth abbreviated “DATT” and “IATT”, respectively, more conditions of the same type are needed.

**Assumption 5.**  $\tau_{i,t} = \tau_{g,t} + \zeta_{i,t}$  and  $\beta_i = \beta + \nu_i$ , where  $\tau_{g,t}$  is non-random, and  $\zeta_{i,t}$  and  $\nu_i$  are mean zero random errors that are independent of each other as well as of all other random elements of the model.

---

<sup>12</sup>They also require that  $T$  is large and that the number of common factors,  $r$ , can be accurately estimated, which is again not a requirement in the present paper.

Assumption 5 implies that

$$\mathbb{E}(\beta'_i \tau_{i,t}) = \beta' \tau_{g,t} =: \text{IATT}_{g,t}. \quad (4)$$

Further use of Assumption 1 gives

$$\mathbb{E}(\eta_{i,t}) = \mathbb{E}(\Delta_{i,t}) - \mathbb{E}(\beta'_i \tau_{i,t}) = \text{ATT}_{g,t} - \text{IATT}_{g,t} =: \text{DATT}_{g,t} \quad (5)$$

or, equivalently,

$$\text{ATT}_{g,t} = \text{DATT}_{g,t} + \text{IATT}_{g,t}. \quad (6)$$

The condition that  $\beta_i$  is randomly distributed with a common mean relaxes the otherwise common homogeneous slope condition (see, for example, [Gobillon and Magnac, 2016](#), and [Xu, 2017](#)) and gives  $\beta$  a reasonable interpretation as the average effect, similar to the random slopes specification of the ATT. It allows for some heterogeneity across the cross-sectional units without the need for completely unrestricted estimation and has as a result become very popular in the panel data literature (see, for example, [Chudik et al., 2011](#), and [Pesaran, 2006](#)). If Assumption 5 is deemed too strong, one can still estimate the ATT itself, which does not place any conditions on  $\beta_i$ .

### 3 The TECCE estimator

#### 3.1 The ATT

We need an estimate of  $y_{i,t}(0)$  in order to estimate the ATT. Interestingly, because the objective here is to estimate only the ATT and not the other parameters of the model, the estimation of  $y_{i,t}(0)$  only requires accounting for the interactive fixed effects, as we now explain.

**Counterfactual estimation procedure:**

1. For all  $t$ , compute

$$\hat{\mathbf{f}}_t := \frac{1}{N_0} \sum_{i \in \mathcal{I}_0} \mathbf{z}_{i,t}, \quad (7)$$

where  $\mathbf{z}_{i,t} := [y_{i,t}, \mathbf{x}'_{i,t}]'$  is a  $(m+1) \times 1$  vector containing all the observables. The fact that  $\hat{\mathbf{f}}_t$  is computed with the untreated units only is crucial since in the present paper both  $y_{i,t}$  and  $\mathbf{x}_{i,t}$  can depend on the treatment, which may lead to inconsistency if we use the entire sample.

2. Estimate the following regression by OLS for all  $i$  and  $t \leq T_0$ :

$$y_{i,t} = \mathbf{a}'_i \hat{\mathbf{f}}_t + u_{i,t}, \quad (8)$$

where  $\mathbf{a}_i$  is a  $(m+1) \times 1$  vector of factor loadings and  $u_{i,t}$  is an error term. Define the  $T_0 \times 1$  vector  $\mathbf{y}_i := [y_{i,1}, \dots, y_{i,T_0}]'$  and the  $T_0 \times (m+1)$  matrix  $\hat{\mathbf{f}} := [\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_{T_0}]'$ . In this notation, the OLS estimator of  $\mathbf{a}_i$  is given by

$$\hat{\mathbf{a}}_i := (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}' \mathbf{y}_i, \quad (9)$$

which is computed for all  $i$ .

3. The sought counterfactual estimator is given by

$$\hat{y}_{i,t}(0) := \hat{\mathbf{a}}'_i \hat{\mathbf{f}}_t, \quad (10)$$

which is available for all treated observations  $i \in \mathcal{I}_0^c$  and  $t > T_0$ . Here  $\{\hat{\mathbf{a}}_i\}_{i \in \mathcal{I}_0^c}$  is from step 2, while  $\{\hat{\mathbf{f}}_t\}_{t > T_0}$  is from step 1.

With  $y_{i,t}$  known and  $y_{i,t}(0)$  estimated, the estimated treatment effect is given by

$$\hat{\Delta}_{i,t} := y_{i,t} - \hat{y}_{i,t}(0) \quad (11)$$

for  $i \in \mathcal{I}_0^c$  and  $t > T_0$ . The TECCE estimator of  $\text{ATT}_{g,t}$  for group  $g > 0$  at time  $t$  is obtained by averaging over all group members;

$$\widehat{\text{ATT}}_{g,t} := \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \hat{\Delta}_{i,t}. \quad (12)$$

A major point about the above estimation procedure is that there is no need to account for the covariates in steps 1 and 2. The intuition for why is as follows: the estimated unit-specific treatment effect can be written as

$$\hat{\Delta}_{i,t} = y_{i,t} - \hat{\mathbf{a}}'_i \hat{\mathbf{f}}_t = y_{i,t} - \mathbf{y}'_i \hat{\mathbf{f}} (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}_t, \quad (13)$$

which is a “defactored” version  $y_{i,t}$ . Asymptotically the defactoring eliminates the interactive fixed effects. This means that the source of the endogeneity of  $x_{i,t}$  is gone and therefore the covariates can be ignored without risking omitted variables bias. The defactoring is also the reason for why the procedure works in spite of the fact that  $\hat{a}_i$  is not consistent and in fact remains random even asymptotically because  $T$  is fixed. The defactoring leads to increased variance; however, the interactive fixed effects are gone and so the asymptotic validity of the procedure is unaffected. Hence, similarly to studies such as [Sant’Anna and Zhao \(2020\)](#), our main concern here is not efficiency, but robust estimation and inference. If there are covariates present that are either known not to have a factor structure or there is uncertainty over the process that generated them, the above procedure has to be modified as explained in online appendix.

As [Caetano and Callaway \(2023\)](#) point out, the validity of estimates of the ATT typically depend on whether or not the covariates are affected by treatment status. Most studies circumvent this problem by assuming that there are no covariates at all (see, for example, [Brown and Butts, 2023](#)), that any covariates are time-invariant (see, for example, [Callaway and Karami, 2024](#)), or, perhaps most commonly, that the covariates may depend on time but that they are unaffected by treatment (see [Chan and Kwok, 2022](#), [Gobillon and Magnac, 2016](#), and [Xu, 2017](#), to mention a few). The problem here is that if there are time-varying covariates present that depend on the treatment status, approaches that fail to properly control for these will be subject to omitted variables bias. However, the solution is not as simple as just including the relevant treatment-affected covariates into the model. For example, if we are estimating the effect of a certain policy aimed at reducing unemployment, we might want to control for the poverty rate (see [Caetano and Callaway, 2023](#), for more examples of this type). Because such policies might indirectly reduce poverty, the poverty rate covariate will absorb some of the treatment effect, typically referred to as “post-treatment bias”. Because of this, treatment-affected covariates are often considered to be “bad controls” (see [Angrist and Pischke, 2009](#)). There is therefore a dilemma; while including the covariates induces post-treatment bias, excluding them induces omitted variables bias (see [Aklin and Bayer, 2017](#)).

The defactoring implicit in TECCE not only eliminates the source of the omitted variables bias, but it also enables estimation of the ATT without including the treatment-affected covariates. It therefore resolves the dilemma.

Asymptotic standard errors of estimates of the ATT can be difficult to compute. Many studies therefore resort to bootstrap inference (see, for example, [Callaway and Karami, 2024](#), [Callaway and Sant’Anna, 2021](#), and [Xu, 2017](#)). As we show in the online appendix, bootstrap inference is possible also with TECCE. However, bootstrapping is computationally inconvenient, which means that it does not fit very well with the simple flavor of TECCE. Here we therefore instead follow [Brown and Butts \(2023\)](#), and [Chan and Kwok \(2022\)](#) and employ a version of the non-parametric variance estimator considered by [Pesaran \(2006\)](#):

$$\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) = \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,t} - \widehat{\text{ATT}}_{g,t})^2. \quad (14)$$

In addition to computational simplicity, non-parametric standard errors tend to perform well for CCE in small samples (see [Chudik et al., 2011](#), [Pesaran, 2006](#), and [Westerlund and Kaddoura, 2022](#)).

### 3.2 The DATT and IATT

As we demonstrated in Section 2, the ATT can be decomposed into the DATT and the IATT. We now show how to estimate these constituent parts. The idea behind our estimator of the DATT builds on the discussion of the previous subsection about bad controls. In particular, by making the counterfactual estimation procedure conditional on observed  $\mathbf{x}_{i,t}$ , we no longer estimate the ATT but instead estimate the DATT. In other words, including the bad controls induces post-treatment bias that is exactly equal to the DATT.<sup>13</sup>

Controlling for  $\mathbf{x}_{i,t}$  requires two changes to the counterfactual estimation procedure presented in Section 3.1. First, the appropriate step-2 regression model to be estimated is no

---

<sup>13</sup>The fact that controlling the covariates alters the object being estimated is important not only for the present paper, but also when considering the works of others. As alluded to earlier, many studies assume that the covariates are unaffected by treatment and use the observed covariates in their ATT estimations (see, for example, [Chan and Kwok, 2022](#), and [Xu, 2017](#)). Logic based on our findings suggests that if the unaffected covariates assumption is false, these estimators will only capture the DATT. The Monte Carlo and empirical studies of Sections 5 and 6 elaborate on this point.

longer given by (8) but by

$$y_{i,t} = \beta' \mathbf{x}_{i,t} + \mathbf{a}_i' \hat{\mathbf{f}}_t + u_{i,t}. \quad (15)$$

Define the  $T_0 \times m$  matrix  $\mathbf{x}_i := [\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,T_0}]'$  and the  $T_0 \times T_0$  matrix  $\mathbf{M}_\mathbf{A} := \mathbf{I}_{T_0} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ , where  $\mathbf{A}$  is any  $T_0$ -rowed matrix. The estimators of  $\beta$  and  $\mathbf{a}_i$  in the above model can now be written in the following way:

$$\hat{\beta} := \left( \sum_{i=1}^N \mathbf{x}_i' \mathbf{M}_{\hat{\mathbf{f}}} \mathbf{x}_i \right)^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{M}_{\hat{\mathbf{f}}} \mathbf{y}_i, \quad (16)$$

$$\hat{\mathbf{a}}_i := (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}' (\mathbf{y}_i - \mathbf{x}_i \hat{\beta}). \quad (17)$$

Second, the step-3 counterfactual estimator is now given by

$$\hat{y}_{i,t}(0) := \hat{\beta}' \mathbf{x}_{i,t} + \hat{\mathbf{a}}_i' \hat{\mathbf{f}}_t. \quad (18)$$

Given the above changes, the estimated DATT for group  $g$  at time  $t$  is entirely analogous to the estimated ATT;

$$\widehat{\text{DATT}}_{g,t} := \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \hat{\eta}_{i,t}, \quad (19)$$

where  $\hat{\eta}_{i,t} := y_{i,t} - \hat{y}_{i,t}(0)$ . The IATT can be estimated by simply subtracting off  $\widehat{\text{DATT}}_{g,t}$  from  $\widehat{\text{ATT}}_{g,t}$ ;

$$\widehat{\text{IATT}}_{g,t} := \widehat{\text{ATT}}_{g,t} - \widehat{\text{DATT}}_{g,t}. \quad (20)$$

The variances of the estimated DATT and IATT can be estimated non-parametrically in exactly the same way as the variance of the ATT.<sup>14</sup>

## 4 Asymptotic results

We can now study the asymptotic properties of the estimated ATT and its direct and indirect parts. Assumptions 1–5 characterize the model. In order to establish the required asymptotic results, more conditions are need.

---

<sup>14</sup>Let  $\hat{\tau}_{g,t} := N_g^{-1} \sum_{i \in \mathcal{I}_g} [\mathbf{x}_{i,t} - \hat{\mathbf{x}}_{i,t}(0)]$ , where  $\hat{\mathbf{x}}_{i,t}(0) := \hat{\lambda}_i' \hat{\mathbf{f}}_t$  with  $\hat{\lambda}_i := (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}' \mathbf{x}_i$ . Alternative, yet asymptotically equivalent, estimators of the DATT and the IATT can be obtained by replacing  $\tau_{g,t}$  and  $\beta$  in the definition of the IATT by  $\hat{\tau}_{g,t}$  and  $\hat{\beta}$ , respectively, and then estimating the DATT as the difference between  $\widehat{\text{ATT}}_{g,t}$  and the resulting estimated IATT.

**Assumption 6.**  $v_i$ ,  $v_{i,t}$ ,  $\zeta_{i,t}$ ,  $\varepsilon_{i,t}$  and  $\mathbf{v}_{i,t}$  are all independently distributed across  $i$  with finite fourth-order cumulants. Also,  $\lim_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\varepsilon_{i,t}^2) =: \sigma_{\varepsilon,g,t}^2$  and the  $m \times m$  matrix  $\lim_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\mathbf{v}_{i,t} \mathbf{v}_{i,t}') =: \Sigma_{\mathbf{v},g,t}$  exist and are positive definite.

**Assumption 7.** The  $r \times (m+1)$  matrix  $\Lambda_i := [\alpha_i + \lambda_i \beta_i, \lambda_i]$  is such that  $N_0^{-1} \sum_{i \in \mathcal{I}_0} \Lambda_i$  has full row rank  $r$  almost surely.

**Assumption 8.**  $T_0 > m + 1$ .

**Assumption 9.** The  $r \times r$  matrix  $\sum_{t=1}^{T_0} \mathbf{f}_t \mathbf{f}_t'$  is positive definite.

**Assumption 10.** The  $m \times m$  matrix  $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{M}_{\hat{\mathbf{f}}} \mathbf{x}_i$  exist and is positive definite.

Assumption 6 is not particularly restrictive in the sense that  $y_{i,t}$  and  $\mathbf{x}_{i,t}$  are still allowed to be strongly cross-sectionally correlated through  $\mathbf{f}_t$ .<sup>15</sup> While we do require that the covariance matrices of  $\varepsilon_{i,t}$  and  $\mathbf{v}_{i,t}$  are positive definite, we place no such restrictions on the covariance matrices of  $v_i$ ,  $v_{i,t}$  and  $\zeta_{i,t}$ . Parameter heterogeneity is therefore not a requirement.

Step 1 of the counterfactual estimation procedure uses the cross-sectional averages of the observables to estimate the factors. We therefore require that those averages are informative about the factors. Assumption 7 rules out situations in which there are factors present but their effect on  $\hat{\mathbf{f}}_t$  averages out. Another situation that is ruled out by this assumption is when  $m+1 \geq r$ , so that the number of factors is larger than the number of cross-sectional averages used to estimate them.<sup>16</sup> One way to relax this condition is if some of the factors are observed, because Assumption 7 only applies to the loadings of unobserved factors. Observed factors can be appended to  $\mathbf{f}_t$  in Assumptions 3 and 4, and to  $\hat{\mathbf{f}}_t$  in step 1 of the counterfactual estimation procedure. The rest is unaffected. In Section 6, we elaborate on this point and explain how to test Assumption 7.

Assumptions 8–10 are non-collinearity conditions. Assumptions 8 and 9 ensure that the  $(m+1) \times (m+1)$  matrix  $\hat{\mathbf{f}} \hat{\mathbf{f}}'$  appearing in the step-2 estimator of  $\mathbf{a}_i$  is positive definite both

<sup>15</sup>We illustrated this point using  $\mathbf{x}_{i,t}(0)$ . By Assumptions 4 and 6,  $\mathbb{E}[\mathbf{x}_{i,t}(0) \mathbf{x}_{j,t}(0)'] = \lambda_i' \mathbf{f}_t \mathbf{f}_t' \lambda_j$ , which is not zero unless the loadings are. Interactive fixed effects therefore provide a means to accommodate strong cross-sectional dependence (see Chudik et al., 2011, for a thorough discussion of the notions of weak and strong cross-sectional dependence).

<sup>16</sup>See Juodis et al. (2021) for a discussion and some results for the case when Assumption 7 fails.



asymptotically and in small samples.<sup>17</sup> Assumption 10 generalizes the usual “within assumption” in the unit-specific fixed effects only model, which rules out time-invariant covariates. Assumption 10 rules out more general “low-rank” covariates, such as deterministic constant and trend terms that only vary over time, as it is almost always done in models with interactive effects (see Moon and Weidner, 2015, for a discussion). The reason for this condition is that if there are such low rank covariates present, they will be captured by  $\hat{\mathbf{f}}_t$ , and therefore they cannot be included also in  $\mathbf{x}_i$ . Fortunately, there is an easy fix to this problem, which is to treat all low rank covariates as observed factors and to append them to  $\hat{\mathbf{f}}$ .

An important point about Assumptions 6–10 is that the time series properties of  $\mathbf{f}_t$ ,  $\varepsilon_{i,t}$  and  $\mathbf{v}_{i,t}$  are essentially unrestricted. Note in particular how, unlike most other treatment effects studies, stationarity and homoskedasticity are not needed.

We are now ready to state Theorem 1, which contains our first main result.

**Theorem 1.** *Suppose that Assumptions 1–4 and 6–9 are met. Then, uniformly in  $g \in \{1, \dots, G\}$  and  $t \in \{T_0 + 1, \dots, T\}$ , as  $N_g, N_0 \rightarrow \infty$  with  $N_g/N_0 \rightarrow \tau < \infty$ ,*

$$\frac{\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})}{\sqrt{\text{var}(\widehat{\text{ATT}}_{g,t})}} \rightarrow_d N(0, 1), \quad (21)$$

where  $\rightarrow_d$  signifies convergence in distribution and the definition of  $\text{var}(\widehat{\text{ATT}}_{g,t})$  is provided in the online appendix.

Theorem 1 does not require that Assumptions 5 and 10 hold, which is natural since the estimation of the ATT does not require explicit conditioning on  $\mathbf{x}_{i,t}$ . Asymptotic normality therefore holds without any restrictions on  $\beta_i$  and only minimal conditions on  $\mathbf{x}_{i,t}$ .

Theorem 1 implies that  $\widehat{\text{ATT}}_{g,t}$  is consistent for  $\text{ATT}_{g,t}$  and that the rate of convergence is  $N^{-1/2}$ . In the treatment effects literature it is common to plot estimates of the ATT over time and to interpret variations as being due to treatment effect dynamics. Theorem 1 says that  $\widehat{\text{ATT}}_{g,t}$  can be used for the same purpose. If one is not interested in dynamics but just want a

---

<sup>17</sup>Because our estimator of  $\mathbf{a}_i$  comes from a regression on pre-treatment observations, we need at least as many degrees of freedom as allowable factors. This assumption is similar in principle to requiring at least two time periods for a fixed effects regression.

summary measure of the evidence,  $\widehat{\text{ATT}}_{g,t}$  can be averaged over  $g$  or  $t$ .<sup>18</sup> Since  $G$  and  $T - T_0$  are both fixed, the consistency in Theorem 1 naturally carries over to such averages.

Inference based on Theorem 1 requires a consistent estimator of  $\text{var}(\widehat{\text{ATT}}_{g,t})$ . Unrestricted treatment effect heterogeneity is known to be problematic for variance estimation in the sense that exact asymptotic inference is not always possible (see, for example, [Borusyak et al., 2024](#), and [de Chaisemartin and D'Haultfoeulle, 2020](#)). Another problem is the presence of unobserved common factors and the additional estimation uncertainty that they bring. This problem does not necessarily interfere with asymptotic inference; however, it does require additional conditions to ensure that the factor estimation error is negligible. [Chan and Kwok \(2022\)](#) assume that  $N_0$ ,  $N$  and  $T$  are all large. They deal with the problem by requiring that  $T/N_0 \rightarrow 0$ . The appropriate condition in the present context is given by  $N_g/N_0 \rightarrow 0$ . In empirical work this means that while  $N_g$  should be large,  $N_0$  should be larger.

**Theorem 2.** *Suppose that the conditions of Theorem 1 are met. Then, uniformly in  $g \in \{1, \dots, G\}$  and  $t \in \{T_0 + 1, \dots, T\}$ , as  $N_g, N_0 \rightarrow \infty$  with  $N_g/N_0 \rightarrow 0$ ,*

$$\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) \rightarrow_p \text{var}(\widehat{\text{ATT}}_{g,t}), \quad (22)$$

where  $\rightarrow_p$  signifies convergence in probability.

Theorems 1 and 2 enable construction of test statistics and confidence intervals that are asymptotically valid for a particular constellation of  $g$  and  $t$ . Let us therefore consider the common problem of testing the null hypothesis of  $H_0 : \text{ATT}_{g,t} = 0$  using the following  $t$ -statistic:

$$T_{g,t} := \frac{\sqrt{N_g} \widehat{\text{ATT}}_{g,t}}{\sqrt{\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t})}}, \quad (23)$$

which by Theorems 1 and 2 is asymptotically  $N(0,1)$  under  $H_0$ . This test statistic is useful when  $G$  and  $T - T_0$  are small. As  $G$  and  $T - T_0$  increases, however, so does the number of tests that can be computed. A simple approach would be to compute one test for each

---

<sup>18</sup>See [Callaway and Sant'Anna \(2021\)](#) for a thorough discussion of various ways in which  $\text{ATT}_{g,t}$  can be meaningfully averaged.

$(g, t)$ -pair and to reject the null hypothesis of no treatment effect if at least one of the tests rejects at a particular significance level. However, this means ignoring the multiplicity of the testing problem, which is likely to result in too many rejections. Fortunately, testing multiple time periods and/or groups is almost as easy as testing a particular  $(g, t)$ -pair. Suppose as an example that we are interested in testing if  $H_0$  holds for all groups in a particular time period. This hypothesis can be tested using the following Wald-type test statistic:  $\sum_{g=1}^G T_{g,t}^2$ , which has a limiting  $\chi^2(G)$  distribution under the null. Alternatively, we may consider the empirical rejection frequency of  $T_{1,t}, \dots, T_{G,t}$ , as given by  $G^{-1} \sum_{g=1}^G 1(|T_{g,t}| > c_\alpha)$ , where  $1(A)$  is the indicator function for the event  $A$  and  $c_\alpha$  is the appropriate right-tail  $\alpha$ -level critical value from  $N(0, 1)$ . If  $H_0$  holds for all groups, the empirical rejection frequency should be close to  $2\alpha$  for  $G$  sufficiently large (see [Bai and Ng, 2006](#)).<sup>19</sup>

A major point about Theorems 1 and 2 is that they hold even if  $r$  is unknown, provided only that  $m + 1 \geq r$ , so that the number of factors is not under-specified. As we show in the proof, while  $\text{var}(\widehat{\text{ATT}}_{g,t})$  depends on whether  $m + 1 = r$  or  $m + 1 > r$ , this dependence is successfully mimicked in large samples by  $\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t})$ . Asymptotically valid inference is therefore possible for any  $r$  satisfying  $m + 1 \geq r$ .

The asymptotic approximations in Theorems 1 and 2 are not expected to be perfect in small samples. We propose a simple bootstrap algorithm to obtain bootstrapped  $p$ -values and confidence intervals, and prove their asymptotic validity in the appendix.

**Theorem 3.** *Suppose that Assumptions 1–10 are met. Then, the results reported in Theorems 1 and 2 for  $\widehat{\text{ATT}}_{g,t}$  apply also to  $\widehat{\text{DATT}}_{g,t}$  and  $\widehat{\text{IATT}}_{g,t}$ .*

Theorem 3 implies that the above discussion of Theorems 1 and 2 for  $\widehat{\text{ATT}}_{g,t}$  applies also to  $\widehat{\text{DATT}}_{g,t}$  and  $\widehat{\text{IATT}}_{g,t}$ . The variance estimates are formed the same way as the  $\widehat{\text{ATT}}_{g,t}$  and definitions are given in the online appendix.

---

<sup>19</sup>The Wald-type statistic and empirical rejection frequency are two examples of how a test statistic like  $T_{g,t}$  can be combined but there are many more (see, for example, [Bai and Ng, 2006](#), for a discussion in the context of testing interactive effects).

## 5 Monte Carlo simulations

In this section, we use Monte Carlo simulations as a means to study the small-sample properties of TECCE. Following the bulk of the previous literature, we assume that there is just one treated group, so that  $G = 1$  and  $g \in \{0, 1\}$ . Similarly to [Chan and Kwok \(2022\)](#), we set  $T_0 = T/2$  and  $\mathcal{I}_0 = \{1, \dots, N/2\}$ , so that  $t_1 = T/2 + 1$ ,  $N_0 = N/2$  and  $\mathcal{I}_0^c = \mathcal{I}_1 = \{N/2 + 1, \dots, N\}$ . The first half of the cross-sectional units are therefore untreated and for the second half treatment starts midway through the sample period.

The potential untreated outcome and covariates,  $y_{i,t}(0)$  and  $\mathbf{x}_{i,t}(0)$ , respectively, are generated according to Assumptions 1, 3 and 5 with  $r = m = \beta = 1$ . As in the heterogenous slope experiments of [Pesaran \(2006\)](#), we make  $\mathbf{v}_i$  a draw from  $N(0, \kappa^2)$  with  $\kappa = 0.2$ .

The factor loadings in the equations for  $y_{i,t}(0)$  and  $\mathbf{x}_{i,t}(0)$  are generated with  $\alpha_i = \theta 1(i \in \mathcal{I}_1) + e_i$  and  $\lambda_i = e_i$ , where  $\theta \in \{0, 1\}$  and  $e_i \sim N(1, 1)$ . The term  $\theta 1(i \in \mathcal{I}_1)$  introduces a break in the average loadings. If  $\theta = 0$ , then  $\mathbb{E}(\alpha_i) = 1$  for all  $i$ , whereas if  $\theta = 1$ , then  $\mathbb{E}(\alpha_i) = 1 + 1(i \in \mathcal{I}_1)$ , so that average loading for treated units is different from that of untreated units. The fact that  $e_i$  is present in both equations makes  $\alpha_i$  and  $\lambda_i$  correlated, which in turn means that  $\mathbf{x}_{i,t}(0)$  is endogenous. The loadings are drawn once and are then held fixed across replications. They are therefore non-random in the simulations. This is important because while breaks-in-loadings specifications are common (see, for example, [Chan and Kwok, 2022](#), [Gobillon and Magnac, 2016](#), and [Xu, 2017](#)), here the break does not determine whether or not trends are parallel, but just amplifies any non-parallel trending.

Three specifications of  $\mathbf{f}_t$  are considered. In the first,  $\mathbf{f}_t = 1$  is a constant, while the second,  $\mathbf{f}_t = t$  is a linear trend. In the third and final specification,  $\mathbf{f}_t$  follows an AR(1) process  $\mathbf{f}_t = (1 - \rho) + \rho \mathbf{f}_{t-1} + u_t$ , where  $\rho = 0.5$ ,  $\mathbf{f}_0 = 0$  and  $u_t \sim N(0, 1)$ . Just like the loadings, the factor is drawn once and is then kept fixed across replications. The interactive effects are therefore fixed, which in turn means that the only specification in which trends are parallel is when  $\mathbf{f}_t = 1$ . The errors  $\varepsilon_{i,t}$  and  $\mathbf{v}_{i,t}$  are generated as AR(1) processes with autoregressive coefficient 0.5, similarly to the third specification of  $\mathbf{f}_t$  but with zero mean.

The treated outcome and covariate are generated as

$$y_{i,t} = \eta_i + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t}, \quad (24)$$

$$\mathbf{x}_{i,t} = \boldsymbol{\tau}_i + \boldsymbol{\lambda}'_i \mathbf{f}_t + \mathbf{v}_{i,t} \quad (25)$$

for  $i \in \mathcal{I}_1$  and  $t \geq t_1$ . Here,  $\eta_i$  and  $\boldsymbol{\tau}_i$  are the direct effects of treatment to  $y_{i,t}$  and  $\mathbf{x}_{i,t}$ , which we assume for simplicity to be constant over time. Hence, in this section,  $\text{ATT}_{g,t} = \text{ATT}_g = \text{ATT}_1 = \mathbb{E}(\eta_i) + \beta' \mathbb{E}(\boldsymbol{\tau}_i) = \mathbb{E}(\eta_i) + \mathbb{E}(\boldsymbol{\tau}_i)$ . We consider two specifications of  $\boldsymbol{\tau}_i$ ,  $\boldsymbol{\tau}_i = 0$  and  $\boldsymbol{\tau}_i \sim N(1, \kappa^2)$ , and one specification of  $\eta_i$ ,  $\eta_i \sim N(1, \kappa^2)$ . This means that  $\text{ATT}_1 = 1$  if  $\boldsymbol{\tau}_i = 0$  and  $\text{ATT}_1 = 2$  if  $\boldsymbol{\tau}_i \sim N(1, \kappa^2)$ .

The TECCE estimator is implemented exactly as described in Section 3. We focus on the ATT itself. The results for the DATT and IATT were very similar and are available upon request. The TECCE results are compared to those obtained by regressing  $y_{i,t}$  onto a treatment dummy and  $\mathbf{x}_{i,t}$  using two-way fixed effects OLS, which represents the workhorse of the treatment effects literature. We also simulate the “PCDID–MG” estimator of [Chan and Kwok \(2022\)](#), and the “GSC” estimator of [Xu \(2017\)](#), which are two of the closest competitors to TECCE. Both estimators are implemented while treating the number of factors,  $r$ , as known. This is unlike TECCE, which in view of the fact that  $m + 1 = 2 > r = 1$  here is overestimating the number of factors. This is one difference that should be taken into account when considering the results as “overestimating” the number of factors should lean to efficiency losses for TECCE. Another difference is that while the OLS and PCDID estimators require that the ATT is time-invariant, the TECCE and GSC estimators do not. Hence, since in this section  $\text{ATT}_1$  is constant over time, the OLS and PCDID estimators make use of a true restriction. For each estimator, we report the average bias and the root mean squared error (RMSE) of the estimated ATT, which in case of GSC and TECCE is the estimated ATT for time period  $t_1 = T_0 + 1$ . The number of replications is set to 2,000.

INSERT TABLES [1](#) AND [2](#) ABOUT HERE

Tables [2](#) and [1](#) report the results for the cases when the loadings are subject to a structural break and when they are not, respectively. In both cases,  $\boldsymbol{\tau}_i = 0$  and therefore the covariate is

unaffected by treatment. Since trends are parallel when  $\mathbf{f}_t = 1$ , all four estimators considered are expected to work well in this case, which is just what we see in the tables. The ranking in terms of RMSE is also as expected with the PCDID–MG and OLS estimators that make use of the restriction that the ATT is time-invariant outperforming the GSC estimator, which in turn outperforms TECCE. The fact that the RMSE is generally lower for GSC than for TECCE is to be expected given the conjecture of [Bai \(2009\)](#) that his PC-based approach is asymptotically efficient under normal errors, and that overestimation of the number of factors is known to lead to finite sample inefficiency (see [Moon and Weidner, 2015](#)). The performance in terms of RMSE is reflected also in the bias results with the PCDID–MG and OLS estimators generally coming out on top, although the differences are very small. The overall best performance in terms of bias is obtained by using the TECCE, whereas in terms of RMSE, GSC is best.

The above picture is quite different when  $\mathbf{f}_t$  is no longer constant, so that trends are not parallel. The PCDID–MG and OLS estimators, that previously performed well, have large bias and RMSE, especially in the case when loadings are breaking. The fact that the OLS estimator does not work is just as expected, because with  $\mathbf{f}_t$  time varying two-way fixed effects are not enough to control for the endogeneity induced by the interactive fixed effects. The fact that PCDID–MG also does not work is consistent with the discussion in [Pesaran \(2006\)](#) where the point is made that PC based on OLS residuals, which is how PCDID is constructed, will in general be inconsistent, because it makes use of an inconsistent estimator of the model parameters.

INSERT TABLE 3 ABOUT HERE

Table 3 is the same as Tables 1, except that now  $\tau_i \sim N(1, \kappa^2)$ , so that the covariate is affected by treatment. As expected given the discussion of Section 3, except for TECCE, all estimators are seriously biased. [Chan and Kwok \(2022\)](#) point out that in PCDID covariates may be correlated with, yet not causally affected by, treatment. Of course, in practice one never knows whether an observed correlation is in fact causal or not. The results in Table 3 illustrate the risk involved.

All-in-all, we find that the TECCE estimator performs well and that it is essentially unbiased in the type of small- $T$  panels considered here. In fact, good performance in terms of bias seems remarkably robust. The estimator is not the most efficient, which is also not expected given its simplicity.

## 6 Empirical illustration

Since its membership in the WTO in 2001, China's role in the world economy has grown enormously. As a result, the pro-competitive effects of China's WTO accession have attracted considerable attention, so much so that there is by now a separate strand of literature devoted to them. The bulk of the evidence seems to suggest that both the level and dispersion of markups have gone down following the WTO entry, and that this development has had important welfare effects (see, for example, [Hsu et al., 2020](#)).

The standard approach to studying WTO membership on market characteristics is to exploit differences in tariffs across industries. For example, one can split industries into a treatment and a control group, where the former is relatively more exposed to the WTO accession. Given that pre-WTO tariffs varied greatly across industries, industries that had previously been protected with high tariffs experienced greater tariff reduction. They should therefore be relatively more exposed to the "treatment". The effect of the WTO accession is then estimated via a difference-in-differences-style OLS regression in which markup is regressed onto a dummy variable that takes on the value one for treated industries in post-WTO periods, control variables, and industry and time fixed effects.

While popular, the standard approach to WTO evaluation has (at least) two drawbacks. First, it requires the parallel trends assumption, which may not be realistic in this context. A commonly cited reason is that certain industries have more lobbying power for protection and are thus less responsive to changes in the macroeconomy. Tariffs may be granted to domestic special interest groups, the pressure of which may vary over time (see, for example, [Fan et al., 2018](#), [Deng et al., 2018](#), and [Xiang et al., 2017](#)). Differences in lobbying power may therefore cause the treatment and control groups to differ systematically over time, even if China had

not joined the WTO in 2001.<sup>20</sup> Because many sources of possible non-parallel trending are unknown and lack good proxies, it is common to control for industry-specific linear time trends (see, for example, [Liu and Qiu, 2016](#), and [Mao and Xu, 2019](#)). Deterministic trends can account for some non-parallel trending but not all. Moreover, results tend to be highly sensitive to the inclusion of such trends.<sup>21</sup>

The second main drawback of the standard approach is that it cannot handle covariates that are affected by treatment. This issue is important because the literature has identified many channels through which the WTO accession may affect markups (see [Mao and Xu, 2019](#), [Fan et al., 2018](#), [Deng et al., 2018](#), [Liu and Ma, 2021](#), and [Brandt et al., 2017](#), to mention a few). Two common examples are the price- and cost-change channels. Markup is defined as the ratio of price to marginal cost. Thus, markup changes can come from price changes, cost changes, or both. It is therefore common to include one of these variables as a covariate, and to estimate the effect of the WTO accession on it to understand the causal mechanism of treatment (see, for example, [Mao and Xu, 2019](#), [Fan et al., 2018](#), and [Lu and Yu, 2015](#)). But then we know from Section 3 that treatment-affected covariates require special treatment or else the estimated ATT will be misleading. Specifically, the inclusion of such covariates will absorb some effect of treatment. The following quotation, taken from [Fan et al. \(2018, page 116\)](#), suggests that researchers are aware of this problem: “If the marginal-cost channel indeed plays a role, then once the marginal costs are included as an explanatory variable, we would witness attenuation of the impact of input tariffs on markups.”

The present paper is not the first to point to these shortcomings, but it is the first to consider an econometric approach that is designed to deal with both. The TECCE approach allows for interactive fixed effects in which there may be unobserved differences between cross-sectional units that change over time as a result of common shocks. The parallel trend condition is therefore not required, which is a substantial advantage when compared to the standard fixed

---

<sup>20</sup>Similarly, policymakers may lower tariffs selectively in industries that are able to compete with relatively less expensive imports, for example, in industries experiencing a productivity boom (see [Brandt et al., 2017](#)).

<sup>21</sup>Some studies include common controls that are thought to be highly correlated with various kinds of protectionism, such as wage rates, employment, exports, and imports (see, for example, [Hsu et al., 2020](#)). Again the results tend to be very sensitive.



effects-based approach. Another advantage of the approach that we exploit in this section is that it not only allows for covariates that may be affected by treatment but that it makes it possible to assess the relative importance of the direct and indirect treatment channels. It should therefore be well suited for the problem at hand.

The data set that we use is taken from [Lu and Yu \(2015\)](#) (see also [Deng et al., 2018](#), who use the same data), and comprises 164 industries (three-digit Chinese industrial classification) observed over the 1998–2005 period. The smallness of  $T$  here, which is a feature of most data sets in the literature, means that it is important to use techniques that work even if  $T$  is not large. The Monte Carlo results reported in Section 5 suggest that the proposed TECCE approach should work well. Following [Lu and Yu \(2015\)](#), the outcome variable is markup dispersion, as measured by the markup Theil index (in logs). Industries are split in half into the treatment and control groups based on whether they faced tariffs above or below the sample median in 2001.

We focus on the [Lu and Yu \(2015\)](#) study in part because of their analysis of the price- and cost-change channels (see their Section E). The authors use the TFP Theil index as a proxy for marginal costs. They argue that its inclusion allows them to partially isolate the price-change channel. In order to assess the ATT of the WTO accession on costs, the authors run a second OLS regression with the TFP Theil index as dependent variable and markup dispersion as a covariate. The estimated ATTs are significant, which is taken as evidence that both channels are operational. The purpose of this illustration is to assess whether or not the treatment affects the controls.

The above discussion suggests that in terms of the notation of Section 2,  $y_{i,t}$  is the markup Theil index and  $x_{i,t}$  is the TFP Theil index. The estimated factors in  $\hat{\mathbf{f}}_t$  are made up of the cross-sectional averages of these variables. A constant is included as an observed factor (as explained in Section 4), which is tantamount to allowing for industry fixed effects. We therefore allow for one known and two unknown factors. In order to assess if this is enough, we apply the rank classifier of [De Vos et al. \(2024\)](#). The estimated number of factors and rank of the matrix of average factor loadings are equal to one and two, respectively, suggesting that

Assumption 7 is met.

INSERT FIGURE 1 ABOUT HERE

The estimated overall and indirect ATTs are reported in Figure 1. The estimates are reported for each year and averaged over all the post- and pre-treatment periods, as is customary in the literature. Both types are reported together with 95% confidence intervals. We first note that both the total and indirect ATT estimates are negative, suggesting that markup dispersion decreased more in industries that had relatively high tariffs in 2001. Given that industries with higher initial tariffs experienced greater tariff reduction, these results imply that the WTO accession reduced markup dispersion. We also note that all pre-treatment estimates are close to zero, which implies that our method adequately captures the non-parallel trending in markups before 2001.

While insignificant in 2002 and 2003, the year-specific total ATTs reported in Figure 1 (a) are significant in 2004 and 2005. The point estimate in 2003 is notably noisy. A possible reason for this is that the industry classification system changed in 2003, as noted by, for example, [Chen et al. \(2019\)](#), and [Lu and Yu \(2015\)](#). The estimated average ATT during the whole post-treatment period is about  $-0.1$  and significant, consistent with the results of [Chen et al. \(2019\)](#).

We also look at the estimated indirect ATT to understand the impact of productivity changes on markups. According to the results reported in Figure 1 (b), the estimated IATTs are negative and significant in the post-treatment period and insignificant in the pre-treatment period. [Lu and Yu \(2015\)](#) estimate the ATT on the TFP Theil index and find it to be significantly negative; however, their approach does not allow them to infer whether this negative response of the TFP Theil index has an effect on the markup Theil index. According to our results, the estimated IATTs are sizable, accounting for almost half of the total ATTs. This result is important in itself, but also for what it means for the results in [Lu and Yu \(2015\)](#), which are based on including the TFP Theil index as a covariate. In particular, we know from before that this type of conditioning will absorb the indirect effect. In this case, since both ATTs are estimated to be negative, and the magnitude of the indirect ATTs are about half of the overall ATTs, conditioning on the TFP Theil index will lead to an underestimation of the total ATTs

by about 50%.

## 7 Conclusion

In this paper, we propose a new estimator of the ATT, dubbed “TECCE”, that is applicable in fixed- $T$  panels when the parallel trends condition fails because of the presence of interactive fixed effects. Being CCE-based, TECCE relies on the presence of covariates that load on the same factors as the outcome variable. This assumption allows us to use the cross-sectional averages of the observables to estimate the untreated potential outcomes in post-treatment time periods. The covariates are allowed to depend on the treatment status so that TECCE makes it possible separate the direct ATT that is unrelated to the covariates from the indirect ATT that works through those covariates. The estimator is shown to be consistent and asymptotically normal, thereby enabling standard inference, provided only that the number of cross-sectional units,  $N$ , is large. This condition is a great advantage in practice because in the literature, many data sets involve only a few time periods  $T$ . We consider one such small- $T$  data set in our empirical illustration and estimate the effect of China’s accession into the WTO on the dispersion of industry-level markups.

## References

- ABADIE, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72, 1–19.
- AKLIN, M. AND P. BAYER (2017): "How can we estimate the effectiveness of institutions? Solving the post-treatment versus omitted variable bias dilemma," .
- ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.
- ARKHANGELSKY, D., S. ATHEY, D. A. HIRSHBERG, G. W. IMBENS, AND S. WAGER (2021): "Synthetic difference-in-differences," *American Economic Review*, 111, 4088–4118.
- BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.
- BAI, J. AND S. NG (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131, 507–537.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119, 249–275.
- BORUSYAK, K., X. JARAVEL, AND J. SPIESS (2024): "Revisiting event study designs: Robust and efficient estimation," *Review of Economic Studies*, forthcoming.
- BRANDT, L., J. VAN BIESEBROECK, L. WANG, AND Y. ZHANG (2017): "WTO accession and performance of Chinese manufacturing firms," *American Economic Review*, 107, 2784–2820.
- BREITUNG, J. AND P. HANSEN (2021): "Alternative estimation approaches for the factor augmented panel data model with small T," *Empirical Economics*, 60, 327–351.
- BROWN, N. AND K. BUTTS (2023): "Dynamic Treatment Effect Estimation with Interactive Fixed Effect Models and Short Panels," .
- CAETANO, C. AND B. CALLAWAY (2023): "Difference-in-Differences with Time-Varying Covariates in the Parallel Trends Assumption," ArXiv: 2202.02903.

- CALLAWAY, B. AND S. KARAMI (2024): "Treatment Effects in Interactive Fixed Effects Models with a Small Number of Time Periods," *Journal of Econometrics*, forthcoming.
- CALLAWAY, B. AND P. H. SANT'ANNA (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, S0304407620303948.
- CHAN, M. K. AND S. S. KWOK (2022): "The PCDID approach: difference-in-differences when trends are potentially unparallel and stochastic," *Journal of Business & Economic Statistics*, 40, 1216–1233.
- CHEN, W., X. CHEN, C.-T. HSIEH, AND Z. SONG (2019): "A forensic examination of China's national accounts," *Brookings Papers on Economic Activity*, 77–141.
- CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2011): "Weak and strong cross-section dependence and estimation of large panels," *The Econometrics Journal*, 14, C45–C90.
- DE CHAISEMARTIN, C. AND X. D'HAULTFŒUILLE (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–2996.
- DE VOS, I., G. EVERAERT, AND V. SARAFIDIS (2024): "A method to evaluate the rank condition for CCE estimators," *Econometric Reviews*, 43, 123–155.
- DENG, X., R. JING, AND Z. LIANG (2018): "Trade liberalisation and domestic brands: Evidence from China's accession to the WTO," *World Economy*, 43, 2237–2262.
- EBERHARDT, M. (2022): "Democracy, growth, heterogeneity, and robustness," *European Economic Review*, 147, 104173.
- FAN, H., X. GAO, Y. A. LI, AND T. A. LUONG (2018): "Trade liberalization and markups: Micro evidence from China," *Journal of Comparative Economics*, 46, 103–130.
- GOBILLON, L. AND T. MAGNAC (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551.

- HSU, W.-T., Y. LU, AND G. L. WU (2020): “Competition, markups, and gains from trade: A quantitative analysis of China between 1995 and 2004,” *Journal of International Economics*, 122, 103266.
- HUBER, M. (2014): “Identifying causal mechanisms (primarily) based on inverse probability weighting,” *Journal of Applied Econometrics*, 29, 920–943.
- JUODIS, A., H. KARABIYIK, AND J. WESTERLUND (2021): “On the Robustness of the Pooled CCE Estimator,” *Journal of Econometrics*, 220, 325–348.
- KIM, D. AND T. OKA (2014): “Divorce Law Reforms and Divorce Rates in the USA: An Interactive Fixed-Effects Approach,” *Journal of Applied Econometrics*, 29, 231–245.
- LEEB, H. AND B. M. PÖTSCHER (2005): “Model selection and inference: Facts and fiction,” *Econometric Theory*, 21, 21–59.
- LIU, Q. AND L. D. QIU (2016): “Intermediate input imports and innovations: Evidence from Chinese firms’ patent filings,” *Journal of International Economics*, 103, 166–6–183.
- LIU, Z. AND H. MA (2021): “Input trade liberalisation and markup distribution: Evidence from China,” *Economic Inquiry*, 59, 344–360.
- LU, Y. AND L. YU (2015): “Trade liberalization and markup dispersion: evidence from China’s WTO accession,” *American Economic Journal: Applied Economics*, 7, 221–253.
- MAO, Q. AND J. XU (2019): “Input trade liberalisation, institution and markup: Evidence from China’s accession to the WTO,” *World Economy*, 42, 3537–3568.
- MOON, H. R. AND M. WEIDNER (2015): “Linear regression for panel with unknown number of factors as interactive fixed effects,” *Econometrica*, 83, 1543–1579.
- (2019): “Nuclear Norm Regularized Estimation of Panel Regression Models,” .
- PESARAN, M. H. (2006): “Estimation and inference in large heterogeneous panels with a multifactor error structure,” *Econometrica*, 74, 967–1012.

- SANT'ANNA, P. H. AND J. ZHAO (2020): "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 219, 101–122.
- WESTERLUND, J. AND Y. KADDOURA (2022): "CCE in heterogenous fixed- $T$  panels," *Econometrics Journal*.
- WESTERLUND, J., Y. PETROVA, AND M. NORKUTE (2019): "CCE in fixed- $T$  panels," *Journal of Applied Econometrics*, 34, 746–761.
- XIANG, X., F. CHEN, C.-Y. HO, AND W. YUE (2017): "Heterogeneous effects of trade liberalisation on firm-level markups: Evidence from China," *World Economy*, 40, 1667–1686.
- XU, Y. (2017): "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76.

Table 1: Monte Carlo results without break in loadings and treatment-affected covariates.

$N$	$T$	Bias				RMSE			
		CCE	GSC	PC	OLS	CCE	GSC	PC	OLS
Common factor specification: Constant									
40	8	0.027	-0.027	-0.009	-0.010	0.653	0.328	0.315	0.328
80	8	0.011	-0.061	0.000	-0.001	0.621	0.286	0.226	0.231
160	8	-0.001	-0.027	-0.002	-0.002	0.306	0.156	0.185	0.165
320	8	0.010	-0.033	-0.003	0.002	0.233	0.118	0.134	0.117
40	16	0.031	0.005	0.003	-0.004	0.491	0.312	0.233	0.290
80	16	0.024	0.000	0.006	0.006	0.400	0.239	0.166	0.197
160	16	0.012	-0.013	-0.001	-0.002	0.279	0.159	0.125	0.140
320	16	0.009	-0.010	0.002	0.001	0.200	0.115	0.095	0.100
Common factor specification: AR(1) process									
40	8	0.036	0.024	0.553	-0.147	0.908	0.416	0.721	0.376
80	8	0.015	0.021	0.500	0.101	0.609	0.303	0.646	0.294
160	8	-0.003	-0.083	-0.357	0.080	0.385	0.247	0.448	0.226
320	8	0.005	-0.014	-0.787	0.097	0.206	0.105	0.842	0.178
40	16	0.014	-0.019	-0.098	0.017	0.550	0.309	0.377	0.291
80	16	0.009	0.011	-0.355	0.136	0.348	0.214	0.418	0.320
160	16	0.015	-0.042	0.039	0.092	0.350	0.179	0.317	0.197
320	16	0.006	-0.032	0.042	-0.059	0.212	0.121	0.481	0.129
Common factor specification: Linear trend									
40	8	0.008	-0.003	0.769	-0.316	0.785	0.381	0.917	0.646
80	8	0.019	0.020	0.589	0.277	0.744	0.330	0.695	0.480
160	8	-0.001	0.001	0.501	-0.166	0.370	0.184	0.557	0.325
320	8	0.014	0.003	0.334	-0.037	0.294	0.139	0.378	0.205
40	16	0.019	0.007	0.695	-0.308	0.551	0.362	0.826	0.842
80	16	0.020	0.008	0.621	0.018	0.453	0.268	0.696	0.554
160	16	0.005	-0.003	0.270	-0.007	0.310	0.177	0.353	0.398
320	16	0.005	0.004	0.566	0.005	0.229	0.128	0.608	0.279

Notes: “CCE”, “GSC”, “PC” and “OLS” refer to the proposed TECCE estimator, GSC estimator of Xu (2017), the PCDID–MG estimator of Chan and Kwok (2022), and the two-way fixed effects OLS estimator, respectively. “Bias” and “RMSE” refer to the average bias and the root mean squared error, respectively, across replications.



Table 2: Monte Carlo results with break in loadings but without treatment-affected covariates.

$N$	$T$	Bias				RMSE			
		CCE	GSC	PC	OLS	CCE	GSC	PC	OLS
Common factor specification: Constant									
40	8	0.034	-0.029	-0.009	-0.010	0.785	0.512	0.315	0.328
80	8	0.009	-0.111	0.000	-0.001	0.778	0.436	0.226	0.231
160	8	-0.001	-0.063	-0.002	-0.002	0.382	0.220	0.185	0.165
320	8	0.012	-0.068	-0.003	0.002	0.298	0.170	0.134	0.117
40	16	0.042	0.014	0.003	-0.004	0.579	0.429	0.233	0.290
80	16	0.037	-0.004	0.006	0.006	0.510	0.350	0.166	0.197
160	16	0.017	-0.025	-0.001	-0.002	0.369	0.223	0.125	0.140
320	16	0.014	-0.022	0.002	0.001	0.257	0.164	0.095	0.100
Common factor specification: AR(1) process									
40	8	0.056	0.112	1.163	0.610	1.129	0.705	1.367	0.701
80	8	0.020	0.053	1.255	1.454	0.766	0.475	1.490	1.480
160	8	-0.005	-0.576	-1.325	-1.427	0.517	1.317	1.480	1.442
320	8	0.008	-0.028	-2.312	-1.832	0.264	0.148	2.400	1.838
40	16	0.018	-0.042	-0.246	-0.425	0.708	0.489	0.738	0.514
80	16	0.011	0.034	-1.588	-2.297	0.436	0.446	1.650	2.314
160	16	0.020	-0.076	0.136	1.127	0.450	0.264	0.824	1.141
320	16	0.007	-0.062	0.095	-0.968	0.276	0.177	1.149	0.975
Common factor specification: Linear trend									
40	8	0.009	0.043	2.431	3.641	0.950	0.733	2.581	3.683
80	8	0.020	0.029	1.844	4.239	0.947	0.497	1.947	4.257
160	8	0.000	0.001	2.108	3.817	0.463	0.257	2.167	3.827
320	8	0.019	0.009	1.526	3.962	0.369	0.196	1.559	3.967
40	16	0.028	0.036	2.551	7.513	0.653	0.611	2.679	7.553
80	16	0.028	0.013	3.446	8.018	0.579	0.413	3.514	8.037
160	16	0.004	-0.004	2.325	7.993	0.409	0.262	2.384	8.003
320	16	0.008	0.007	2.593	8.005	0.294	0.190	2.652	8.010

Notes: “CCE”, “GSC”, “PC” and “OLS” refer to the proposed TECCE estimator, GSC estimator of Xu (2017), the PCDID–MG estimator of Chan and Kwok (2022), and the two-way fixed effects OLS estimator, respectively. “Bias” and “RMSE” refer to the average bias and the root mean squared error, respectively, across replications.

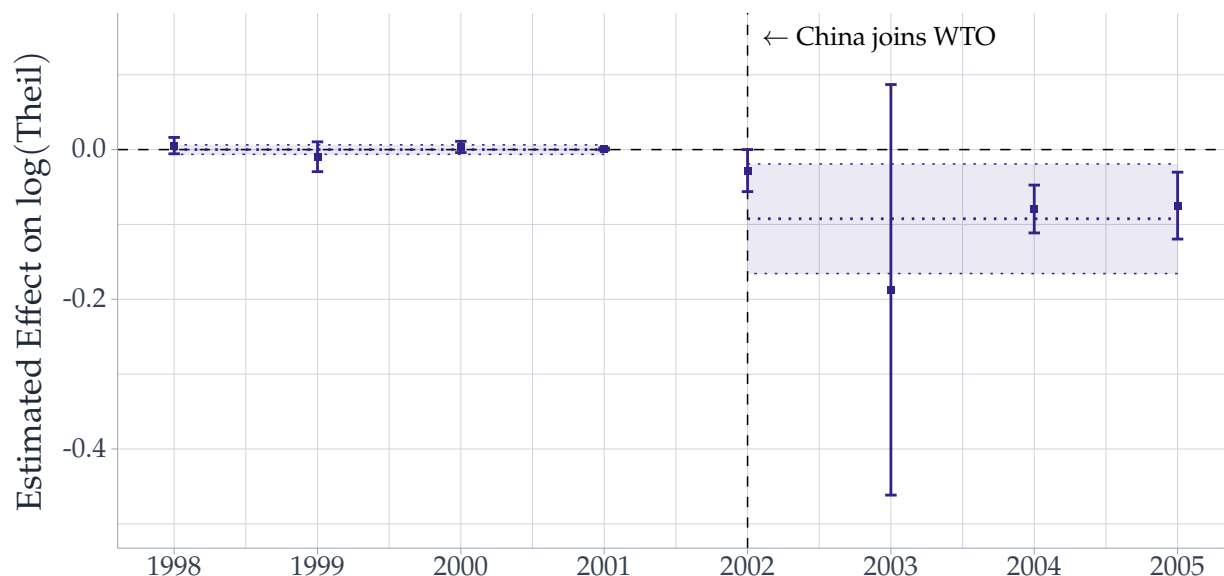
Table 3: Monte Carlo results without break in loadings but with treatment-affected covariates.

$N$	$T$	Bias				RMSE			
		CCE	GSC	PC	OLS	CCE	GSC	PC	OLS
Common factor specification: Constant									
40	8	0.028	-1.051	-1.011	-1.008	0.657	1.106	1.065	1.064
80	8	0.011	-1.068	-1.000	-0.997	0.625	1.108	1.028	1.027
160	8	-0.001	-1.032	-1.002	-1.000	0.307	1.045	1.016	1.020
320	8	0.010	-1.037	-0.998	-1.003	0.234	1.044	1.005	1.013
40	16	0.030	-1.004	-1.003	-0.996	0.494	1.056	1.046	1.026
80	16	0.025	-1.003	-0.994	-0.995	0.404	1.034	1.014	1.010
160	16	0.014	-1.013	-1.002	-1.002	0.282	1.026	1.012	1.010
320	16	0.008	-1.012	-0.999	-0.997	0.202	1.019	1.004	1.002
Common factor specification: AR(1) process									
40	8	0.036	-1.023	-1.322	-0.606	0.911	1.119	1.369	0.791
80	8	0.014	-0.990	-1.416	-1.032	0.610	1.038	1.445	1.122
160	8	-0.002	-1.279	-1.609	-1.988	0.385	1.377	1.623	2.004
320	8	0.006	-1.024	-1.534	-2.203	0.207	1.030	1.541	2.218
40	16	0.014	-1.053	-1.331	-1.374	0.553	1.114	1.363	1.426
80	16	0.009	-1.002	-1.603	-1.823	0.349	1.029	1.628	1.835
160	16	0.013	-1.047	-1.611	-1.567	0.353	1.062	1.621	1.596
320	16	0.006	-1.035	-1.609	-1.382	0.213	1.042	1.613	1.472
Common factor specification: Linear trend									
40	8	0.009	-1.098	-2.129	-0.834	0.788	1.180	2.201	1.015
80	8	0.020	-1.021	-1.545	-1.104	0.748	1.077	1.599	1.180
160	8	-0.001	-1.059	-2.019	-1.108	0.372	1.076	2.039	1.144
320	8	0.014	-1.056	-1.899	-1.373	0.295	1.065	1.910	1.388
40	16	0.017	-1.262	-2.246	-1.085	0.555	1.319	2.388	1.189
80	16	0.021	-1.129	-1.927	-1.043	0.456	1.163	2.005	1.098
160	16	0.006	-1.164	-1.966	-1.486	0.312	1.178	2.007	1.506
320	16	0.005	-1.152	-1.946	-1.192	0.230	1.159	1.967	1.217

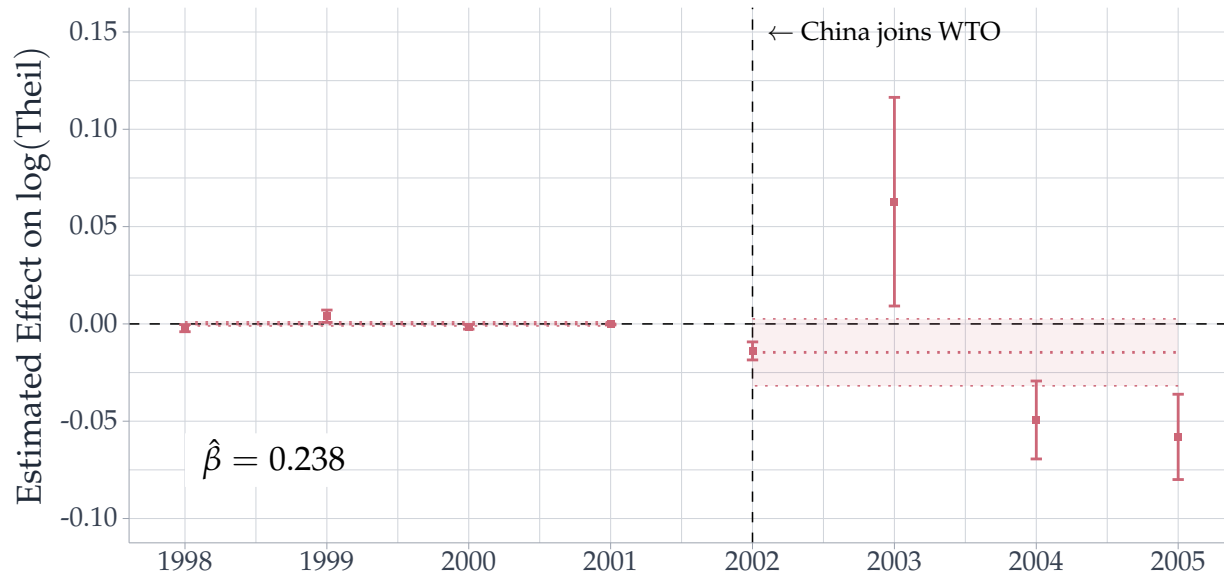
Notes: See Table 1 for an explanation.

Figure 1: Estimated ATTs of China's WTO accession in 2001 on the markup Theil index.

(a) Estimated total ATT



(b) Estimated indirect ATT via TFP dispersion



Notes: The figures present ATT estimates and 95% confidence intervals for the effect of China's WTO accession in 2001 on the dispersion of markups as measured by the markup Theil index. The treatment group comprise all industries that in 2001 had above-median tariff rates. Estimates are computed using the TECCE estimator with the TFP Theil index as a covariate. A constant is included as an observed factor. Figure (a) presents estimates of the total ATT and figure (b) presents the estimated indirect ATT operating through the TFP Theil index.  $\hat{\beta}$  in figure (b) refers to the estimated slope on the TFP Theil index in the markup Theil index regression.