

Estimation of Average Treatment Effects Using Panel Data when Treatment Effect Heterogeneity Depends on Unobserved Fixed Effects*

Shosei Sakaguchi [†]

December 20, 2019

Summary

This paper proposes a new panel data approach to identify and estimate the time-varying average treatment effect (ATE). The approach allows for treatment effect heterogeneity that depends on unobserved fixed effects. In the presence of this type of heterogeneity, existing panel data approaches identify the ATE for limited subpopulations only. In contrast, the proposed approach identifies and estimates the ATE for the entire population. The approach relies on the linear fixed effects specification of potential outcome equations and uses exogenous variables that are correlated with the fixed effects. I apply the approach to study the impact of a mother's smoking during pregnancy on her child's birth weight.

Keywords: Potential outcome, Program evaluation, Time-varying treatment, Treatment effect heterogeneity, Unobserved heterogeneity.

JEL classification: C21, C23.

*I would like to thank the co-editor Edward Vytlačil and three anonymous referees for their constructive suggestions and comments. This paper is based on a chapter of my dissertation at Kyoto University. I am grateful to Yoshihiko Nishiyama, Ken Yamada, Ryo Okui, and Keisuke Hirano for their helpful comments and guidance. I would also like to thank Naoya Sueishi, Susumu Imai, and participants in the seminars at Kyoto University and Hitotsubashi University, and those in the sessions at the 2017 Asian Meeting of the Econometric Society in Hong Kong and the IAAE 2017 Annual Conference in Sapporo for their comments. An earlier version of this paper was awarded the 2017 IAAE Student Prize at the IAAE 2017 Annual Conference. I acknowledge financial support from JSPS KAKENHI Grant Number JP16J01170.

[†]Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: s.sakaguchi@ucl.ac.uk.

1 Introduction

Panel data is popularly used in program evaluation studies to control for unobserved fixed effects. There exist many panel data methods that enable one to do so by relying on functional form restrictions or restrictions on the distribution of unobserved variables. However, when treatment effects are heterogeneous depending on unobserved fixed effects, existing panel data methods cannot identify and consistently estimate the average treatment effect (ATE) for the entire population. This paper proposes a new panel data approach that can identify and estimate the ATE for the entire population, even when treatment effect heterogeneity depends on unobserved fixed effects.

In program evaluation studies, treatment effects are often heterogeneous among individuals with different characteristics, and this heterogeneity often depends on unobserved fixed effects. For example, in job training evaluations, the effect of training on wages may be heterogeneous among individuals with different levels of unobserved ability. When treatment effect heterogeneity depends on unobserved fixed effects, existing panel data methods cannot identify and consistently estimate the ATE for the entire population. The frequently used fixed effects (FE) estimator, based on a conditional independence assumption, is biased in this case.¹ The difference-in-differences (DID) method, based on a common trend assumption, can identify and estimate the ATE for the treated (ATT), but not for the entire population. In many cases, however, estimating the ATE is often more desirable in terms of the external validity of estimation results, which has been capturing increasing attention in empirical studies (e.g., Athey and Imbens, 2017). For example, if we consider expanding an existing policy intervention from the present treated subpopulation to the entire population, the ATE should be the parameter of interest.

This paper proposes a new panel data approach that can identify and estimate the ATE for the entire population, while allowing for treatment effect heterogeneity that depends on unobserved fixed effects. The identification relies on the linear fixed effects specification of the treated and untreated potential outcome equations, wherein a scalar fixed effect enters additively the two equations with different coefficients. This specification allows treatment effect heterogeneity to depend on unobserved fixed effects, albeit it imposes some parametric restrictions. The approach also allows the ATE to be time-varying and can identify and estimate the ATE in each time period. Note, however, that it does not allow for the dynamic

¹In the context of group fixed effects, Gibbons et al. (2018) presented a number of empirical examples where the FE estimator is biased because treatment effect heterogeneity depends on group fixed effects.

structures in treatment effects, unlike Callaway and Sant’Anna (2018).

Under the specified model, we identify the ATE by predicting the unobserved potential outcome for each treated and untreated individual using the following three steps. First, we identify the coefficients of the observed covariates in the regression functions for the treated and untreated potential outcomes by applying the within-transformation to each regression function to difference out the fixed effects. Next, by looking at the residuals from the first step, we identify the ratio of the fixed effects terms between the two regression functions. To do that, we use changes over time in the treatment and control groups, which are observed in the subpopulation of individuals called movers,² and exogenous variables that are correlated with the fixed effects conditional on the treatment assignment. As discussed later, finding such exogenous variables is usually not difficult. Finally, we identify the ATE by combining observed data with the identified parameters to predict unobserved potential outcomes for each treated and untreated individual. Building on the identification result, the ATE can be estimated by a generalized method of moments (GMM) estimator.

The limitation of this approach is that it relies on the scalar and additive fixed-effects specification for the regression functions of the potential outcomes. Under nonseparable panel data models, Chernozhukov et al. (2013) considered identification and estimation of the average and quantile treatment effects based on the time homogeneity condition of the period-specific disturbance terms. With the same type of time homogeneity condition, Jun et al. (2016) derived the sharp identified bounds of the potential outcomes distributions. Sakaguchi (2016) proposed an approach to identify and estimate the ATE for the entire population as an extension of the DID approach. That approach exploits uniquely structured panel data, wherein the treatment exposure expands from no individuals to all individuals across time periods. Unlike those works, although the approach proposed in this paper relies on the scalar and additive fixed-effects specification, it does not require the time homogeneity condition or a specific data structure.

This paper also relates to the literature on the correlated random coefficient (CRC) panel data model. The CRC panel data model captures treatment effect heterogeneity that depends on unobserved fixed effects, and the average partial effect (APE) of the individual-specific slope in the model generally corresponds to the ATE.³ One common approach used to identify and estimate the APE in the model is a generalized within-group approach (e.g., Chamberlain, 1992), which identifies and estimates the APE for a subpopulation

²Movers are individuals who experience both treatment and no-treatment across observed time periods.

³This is illustrated in Wooldridge (2010, p. 968).

of movers. In contrast, in this paper we assume scalar fixed effects; instead, we identify and estimate the whole ATE. Other works studying the CRC panel data model include Wooldridge (2005), Arellano and Bonhomme (2012), and Graham and Powell (2012).

As an empirical application, I apply the approach to study the impact of a mother’s smoking during pregnancy on her child’s birth weight using the matched panel data constructed by Abrevaya (2006). By applying the proposed approach, we can allow the effect of mothers’ smoking to be heterogeneous depending on the mother’s unobserved fixed effects, and we can estimate the average smoking effect on the entire population of mothers. In the application, I find that a mother’s smoking has a negative average effect at each of three birth times, and the effect worsens with additional birth time.

The remainder of this paper is structured as follows. Section 2 describes the setting, specification of potential outcomes, and assumptions required for identification. Section 3 describes the identification approach and estimation method. Section 4 presents Monte Carlo simulation results to show the finite sample behavior of the estimator. Section 5 describes the empirical application. Section 6 concludes this paper with some remarks. An online Appendix contains all the proofs and additional simulation results.

2 Setup and Model

We suppose that $\{Y_{it}, D_{it}, X_{it}, Z_{it}\}$ is observed for N individuals ($i = 1, 2, \dots, N$) across T time periods ($t = 1, \dots, T$), where $T \geq 2$. We assume N is large, while T is small. Y_{it} is an observed outcome. $D_{it} \in \{0, 1\}$ is a binary treatment indicator, such that $D_{it} = 1$ if individual i is treated in period t and $D_{it} = 0$ otherwise. We suppose that the treatment assignment is time-varying. X_{it} is a $K \times 1$ vector of time-varying observed covariates that may include time dummies (or a deterministic time trend), time-varying observed confounding variables, and interactions among them. Z_{it} is an $L \times 1$ vector of exogenous variables that are correlated with unobserved fixed effects conditional on treatment assignment. Z_{it} may consist of some or all variables included in X_{it} , as well as external variables. No restrictions are required on the dimension of Z_{it} . We will discuss assumptions on Z_{it} and its use later. Finally, let $Y_{it}(1)$ and $Y_{it}(0)$ be the potential outcomes for individual i in period t with and without treatment, respectively. Then, Y_{it} is expressed as $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$.

Throughout this paper, we focus on the ATE for the entire population in each period $t \in \{1, \dots, T\}$: $\tau_t^{ate} \equiv E[Y_{it}(1) - Y_{it}(0)]$. In many cases, the ATE for the entire population, rather than for a limited subpopulation, is more interesting. We shall allow the ATE to

be time-varying, and consider identifying and estimating the ATE in each period. Note, however, that the time-varying ATE considered here is different from the time-varying treatment effect parameters considered by Lechner (2009) or Callaway and Sant’Anna (2018), in that it is defined only relative to the treatment assignment of the corresponding period, and does not capture the dynamics of the treatment effect considered by them.

Let C_i be a scalar fixed effect. Throughout this paper we suppose that the potential outcomes $Y_{it}(1)$ and $Y_{it}(0)$ are expressed as the following linear forms:

$$Y_{it}(1) = \alpha^1 + X'_{it}\beta^1 + \gamma^1 C_i + u_{it}^1, \quad (1)$$

$$Y_{it}(0) = \alpha^0 + X'_{it}\beta^0 + C_i + u_{it}^0, \quad (2)$$

for $t = 1, \dots, T$, where u_{it}^1 and u_{it}^0 are defined as $u_{it}^1 = Y_{it}(1) - E[Y_{it}(1) \mid X_{it}, C_i]$ and $u_{it}^0 = Y_{it}(0) - E[Y_{it}(0) \mid X_{it}, C_i]$, respectively. The equations are separable linear in X_{it} , C_i , and u_{it}^j . For scale normalization, we set the coefficient of C_i in equation (2) to one. We also restrict $\gamma^1 \neq 0$, which is required for identification.

In the model specification above, the coefficients of C_i and X_{it} are allowed to be different between (1) and (2). Further, the constant terms α^1 and α^0 are separated from the fixed effect terms, in contrast to the standard linear fixed effects panel data model. Under this specification, τ_t^{ate} in each period $t \in \{1, \dots, T\}$ is expressed as follows:

$$\tau_t^{ate} = (\alpha^1 - \alpha^0) + E[X_{it}]'(\beta^1 - \beta^0) + (\gamma^1 - 1)E[C_i],$$

where the expectation is not taken over time, and hence the time variation in X_{it} is not averaged out. As seen in this equation, the ATE consists of three parts: a constant part, and averages of two heterogeneous parts that depend on observed covariates and unobserved fixed effects, respectively. Notably, $\gamma^1 \neq 1$ allows the treatment effect heterogeneity to depend on unobserved fixed effects. $\beta^1 \neq \beta^0$ allows treatment effect heterogeneity to depend on observed covariates. In addition, the time variation of X_{it} causes time variation in the ATE. The separability of the constant terms from the fixed effects terms allows the treatment effects to consist of not only heterogeneous parts, but also a homogeneous part. Although we assume that X_{it} enters linearly in (1) and (2), a parametric nonlinear function of X_{it} could also be used as far as it is completely separable from the other terms.

In the remainder of this section, I describe four assumptions that are required for identification. First, throughout this paper, we suppose that all the defined random variables

are independent and identically distributed (i.i.d.) across individuals:

Assumption 2.1. $\{Y_{it}(0), Y_{it}(1), D_{it}, X_{it}, Z_{it}\}_{t=1}^T, C_i\}$ are i.i.d. across i .

This assumption imposes no restrictions on the distribution of the data across time periods unlike, for example, Chernozhukov et al. (2013) and Jun et al. (2016).

Next, to formalize the idea of confounding due to the presence of observed covariates and unobserved fixed effects, we impose the following assumption.

Assumption 2.2. $E[Y_{it}(j) \mid D_{i1}, \dots, D_{iT}, X_{i1}, \dots, X_{iT}, C_i] = E[Y_{it}(j) \mid X_{it}, C_i]$ for all $j = 0, 1$ and $t = 1, \dots, T$.

This assumption is the same as the fundamental assumption for the FE estimation. The assumption requires that, in each time period, the potential outcomes are mean independent of the treatment assignments across all periods conditional on the observed covariates in the given period and unobserved fixed effects. This assumption also requires that the potential outcomes in each time period do not depend on the past or future observed covariates conditional on the unobserved fixed effects. This implies strict exogeneity of X_{it} conditional on C_i .

We also impose the following assumption on the treatment assignment.

Assumption 2.3. $P(D_{it} = j, 0 < \sum_{t=1}^T D_{it} < T) > 0$ for all $j = 0, 1$ and $t = 1, \dots, T$.

This assumption requires that the treated and untreated groups in each time period include some individuals who have experienced both treatment and no-treatment at least once across the observed time periods (i.e., individuals with $\{0 < \sum_{t=1}^T D_{it} < T\}$). We call such individuals “*movers*,” borrowing the terminology introduced by Chamberlain (1982). Assumption 2.3 guarantees the existence of movers in each of the treated and untreated groups in each time period. Note that this assumption only requires some individuals, not all to be movers.

The last assumption requires that, in the subpopulation of movers, the variables in Z_{it} are strictly exogenous and correlated with the unobserved fixed effects conditional on treatment assignment. Z_{it} satisfies the following assumption:

Assumption 2.4. For any $j = 0, 1$, $k = 0, 1$, $s = 1, \dots, T$, and $t = 1, \dots, T$;

- (i) $E \left[Z_{it} u_{is}^k \mid D_{it} = j, D_{is} = k, 0 < \sum_{t=1}^T D_{it} < T \right] = 0,$
(ii) $\text{rank} \left(\sum_{t=1}^T E \left[(1, Z'_{it})'(1, C_i) \mid D_{it} = j, 0 < \sum_{t=1}^T D_{it} < T \right] \right) = 2.$

For movers, Assumption 2.4 (i) requires that Z_{it} is orthogonal to u_{is}^k conditional on treatment assignments in any pair of periods s and t ; Assumption 2.4 (ii) requires that Z_{it} is correlated with the fixed effects conditional on treatment assignment in period t .

In practice, finding Z_{it} that satisfies Assumption 2.4 is usually not very difficult. Some or all variables included in X_{it} are likely to satisfy Assumption 2.4, and may be included in Z_{it} . Assumption 2.2 implies that variables in X_{it} satisfy condition (i) in Assumption 2.4.⁴ Further, because both X_{it} and C_i should affect treatment assignment, they are likely to be correlated with each other after conditioning on the treatment assignment, even if they are not unconditionally correlated with each other.⁵ One exceptional case where $Z_{it} = X_{it}$ does not work is when X_{it} is not unconditionally correlated with C_i and, further, does not affect the treatment assignment. In this case, X_{it} is not correlated with C_i even conditional on D_{it} (does not satisfy condition (ii)).

We also often have some variables that are unconditionally correlated with C_i . For example, if C_i is individual's unobserved ability, years of education should be correlated with C_i and, hence, may be included in Z_{it} . Finally, although Z_{it} has the subscript t , time-invariant exogenous variables can also be included in Z_{it} for any $t = 1, \dots, T$.

Remark 2.1 (Relation to the CRC Panel Data Model) The observation rule, $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$, transforms the potential outcome model (1) and (2) into the following CRC panel data model:

$$\begin{aligned} Y_{it} &= D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0) \\ &= X'_{it}\beta^0 + \tau(C_i)D_{it} + D_{it}X'_{it}(\beta^1 - \beta^0) + \tilde{C}_i + u_{it}, \end{aligned}$$

where $\tau(C_i) = (\alpha^1 - \alpha^0) + (\gamma^1 - 1)C_i$, $\tilde{C}_i = \alpha^0 + C_i$, and $u_{it} = D_{it}u_{it}^1 + (1 - D_{it})u_{it}^0$. $\tau(C_i)$ is the correlated random coefficient in this model. Because the ATE is expressed

⁴This is because Assumption 2.2 is equivalent to $E[u_{it}^j \mid D_{i1}, \dots, D_{iT}, X_{i1}, \dots, X_{iT}, C_i] = 0$ for all $j = 0, 1$ and $t = 1, \dots, T$, which is a sufficient condition for condition (i) in Assumption 2.4 when $Z_{it} = X_{it}$.

⁵To illustrate this statement with an example, suppose that D_{it} denotes job training participation, C_i is individual ability, and X_{it} is individual age. Suppose further that individuals with low age and low ability tend to participate in the training. In this case, after conditioning on the training participation (D_{it}), age (X_{it}) becomes correlated with the ability (C_i), because individuals who participate in the training are likely to be young and have low ability and vice versa.

as $\tau_t^{ate} = E[\tau(C_i)] + E[X_{it}'(\beta^1 - \beta^0)]$, identifying the Average Partial Effect (APE) of D_{it} , $E[\tau(C_i)]$, is necessary to identify the ATE. However, the commonly used generalized within-group approach can identify the APE only for the subpopulation of movers; hence, this approach can identify the ATE only for movers. Under the specifications (1) and (2), the approach proposed in the following section can identify and estimate the ATE for the entire population.

Remark 2.2 (Relation to the Linear Fixed Effects Panel Data Model): The potential outcome model (1) and (2) nests the standard linear fixed effects panel data model. When $\gamma^1 = 1$, the observed outcome is expressed as the following:

$$Y_{it} = X_{it}'\beta^0 + (\alpha^1 - \alpha^0)D_{it} + D_{it}X_{it}'(\beta^1 - \beta^0) + \tilde{C}_i + u_{it}. \quad (3)$$

This corresponds to the standard linear fixed effects panel data model. The ATE is $\tau_t^{ate} = (\alpha^1 - \alpha^0) + E[X_{it}'(\beta^1 - \beta^0)]$ and does not depend on C_i . In this case, we can identify the ATE by either the usual fixed effects approach or the proposed approach.

3 Identification and Estimation

This section first describes the identification approach. Subsequently, in Section 3.2, the GMM estimation of the identified parameters is proposed.

3.1 Identification

This section describes the identification of τ_t^{ate} , based on the potential outcomes equations (1) and (2) and the assumptions described in the previous section and below. The identification argument proceeds in three steps. In the first step, we identify β^1 and β^0 . Then, we identify the ratio of fixed effects terms between the two regression functions, that is γ^1 . Finally, we identify τ_t^{ate} by combining the observed data with the identified parameters. The following describes the steps of the identification procedure.

First step

In the first step, we identify β^1 and β^0 based on the idea of the within-transformation method used in the FE approach. Let $\bar{A}_i^j = (1/\sum_{t=1}^T 1\{D_{it} = j\})\sum_{t=1}^T 1\{D_{it} = j\} \cdot A_{it}$ and $\check{A}_{it}^j = A_{it} - \bar{A}_i^j$ for any variable A_{it} and $j = 0, 1$, where $1\{\cdot\}$ is the indicator function.

\bar{A}_i^1 (\bar{A}_i^0) is the mean of A_{it} for individual i across time periods when he or she is treated (not treated). For each $j = 0, 1$ and individual i with $\sum_{t=1}^T 1\{D_{it} = j\} = 0$ (i.e., the individual who is always treated or never treated across the observed time periods), we adopt the convention that $\bar{A}_{it}^j = 0$. This transformation is a kind of within-transformation across treated and untreated periods.

We suppose here that the following assumption holds.

Assumption 3.1. For each $j = 0, 1$, $\text{rank}\left(E[\sum_{t=1}^T 1\{D_{it} = j\} \cdot \ddot{X}_{it}^j \ddot{X}_{it}^{j'}]\right) = K$.

This assumption requires that $E[\sum_{t=1}^T 1\{D_{it} = j\} \cdot \ddot{X}_{it}^j \ddot{X}_{it}^{j'}]$ is full rank for $j = 0, 1$. It implicitly ensures the presence of individuals who are treated in at least two time periods and individuals who are not treated in at least two time periods.⁶

For each of the treated and untreated individuals in period t (the individuals with $D_{it} = 1$ and $D_{it} = 0$, respectively), we have the following transformed model:

$$\ddot{Y}_{it}^j = \ddot{X}_{it}^{j'} \beta^j + \ddot{u}_{it}^j \quad \text{for } j = 0, 1 \text{ and } t = 1, \dots, T,$$

where the fixed effects term has been differenced out by the within-transformation. Then, from the above transformed model and under Assumptions 2.1, 2.2, and 3.1, we can identify β^j ($j = 0, 1$) as follows:

$$\beta^j = E \left[\sum_{t=1}^T 1\{D_{it} = j\} \cdot \ddot{X}_{it}^j \ddot{X}_{it}^{j'} \right]^{-1} E \left[\sum_{t=1}^T 1\{D_{it} = j\} \cdot \ddot{X}_{it}^j \ddot{Y}_{it}^j \right]. \quad (4)$$

This identification is similar to the identification in the standard linear fixed effects panel data model by the within-transformation. Note that this identification is based on the orthogonality condition: $E \left[\sum_{t=1}^T \ddot{X}_{it}^j \ddot{u}_{it}^{j'} \mid D_{i1}, \dots, D_{iT} \right] = 0$, which follows from Assumption 2.2.

The following proposition formalizes the identification result in this step.

Proposition 3.1. Under Assumptions 2.1, 2.2, and 3.1, β^j ($j = 0, 1$) is identified as (4).

A proof is provided in the online Appendix.

⁶The reason is as follows. By the definition of \ddot{X}_{it}^1 , $1\{D_{it} = 1\} \cdot \ddot{X}_{it}^1 = 0$ and $1\{D_{it} = 0\} \cdot \ddot{X}_{it}^0 = 0$ hold in any period t for individuals who have experienced treatment less than twice and individuals who have experienced no-treatment less than twice, respectively. Therefore, the rank condition in Assumption 3.1 does not hold if no individuals have experienced treatment or no-treatment at least twice.

Second step

We next identify the ratio of the scalar fixed effects terms, γ^1 , and the transformed constant terms, $\tilde{\alpha}^1$ and $\tilde{\alpha}^0$, defined below. The identification in this step is based on the subpopulation of movers (i.e., individuals with $\{0 < \sum_{t=1}^T D_{it} < T\}$), whose existence is secured under Assumption 2.3. Once β^1 and β^0 are identified in the previous step, we have the following equations for movers:

$$\eta_i^1(\beta^1) \equiv \bar{Y}_i^1 - \bar{X}_i^{1'}\beta^1 = \alpha^1 + \gamma^1 C_i + \bar{u}_i^1 \quad \text{and} \quad \eta_i^0(\beta^0) \equiv \bar{Y}_i^0 - \bar{X}_i^{0'}\beta^0 = \alpha^0 + C_i + \bar{u}_i^0.$$

Here, $\eta_i^1(\beta^1)$ ($\eta_i^0(\beta^0)$) is equal to the sum of the constant term α^1 (α^0), the fixed effects term $\gamma^1 C_i$ (C_i), and the disturbance term \bar{u}_i^1 (\bar{u}_i^0). Both equations above are nontrivial for movers only. For non-movers, one of them degenerates to zero.⁷ This is why we focus on the subpopulation of movers in this step.

Then, in each period t , we divide the subpopulation of movers into treated and untreated groups: $\{D_{it} = 1, \sum_{t=1}^T D_{it} \neq T\}$ and $\{D_{it} = 0, \sum_{t=1}^T D_{it} \neq 0\}$, respectively.⁸ For movers in the treated group $\{D_{it} = 1, \sum_{t=1}^T D_{it} \neq T\}$, we can re-express the regression function for $Y_{it}(1)$ as follows:

$$Y_{it}(1) = \tilde{\alpha}^1 + X_{it}'\beta^1 + \gamma^1 \eta_i^0(\beta^0) + \tilde{u}_{it}^1, \quad (5)$$

where $\tilde{\alpha}^1 \equiv \alpha^1 - \gamma^1 \alpha^0$ and $\tilde{u}_{it}^1 \equiv u_{it}^1 - \gamma^1 \bar{u}_i^0$. This equation is derived by replacing C_i with $\eta_i^0(\beta^0)$ in equation (1). Similarly, for movers in the untreated group $\{D_{it} = 0, \sum_{t=1}^T D_{it} \neq 0\}$, by replacing C_i with $\eta_i^1(\beta^1)$ in equation (2), we have the following regression function for $Y_{it}(0)$:

$$Y_{it}(0) = \tilde{\alpha}^0 + X_{it}'\beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) + \tilde{u}_{it}^0, \quad (6)$$

where $\tilde{\alpha}^0 \equiv \alpha^0 - \frac{1}{\gamma^1} \alpha^1$ and $\tilde{u}_{it}^0 \equiv u_{it}^0 - \frac{1}{\gamma^1} \bar{u}_i^1$. Note that if $\gamma^1 = 0$, $1/\gamma^1$ in (6) does not take a finite value. This is the reason we restrict $\gamma^1 \neq 0$.

We now consider identifying $\tilde{\alpha}^1$, $\tilde{\alpha}^0$, and γ^1 in the transformed regression functions (5) and (6), but an identification problem arises here. In (5) and (6), $\eta_i^0(\beta^0)$ and $\eta_i^1(\beta^1)$ are

⁷The reason is as follows. For individual i who is never treated, $\eta_i^1(\beta^1)$ degenerates to zero because $\bar{Y}_i^1 = 0$ and $\bar{X}_i^1 = 0$ by definition. Similarly, for individual i who is always treated, $\eta_i^0(\beta^0)$ degenerates to zero.

⁸Note that $\{D_{it} = 1, 0 \leq \sum_{t=1}^T D_{it} \leq T\} = \{D_{it} = 1, \sum_{t=1}^T D_{it} \neq T\}$ and $\{D_{it} = 0, 0 \leq \sum_{t=1}^T D_{it} \leq T\} = \{D_{it} = 0, \sum_{t=1}^T D_{it} \neq 0\}$.

endogenous because they include disturbance terms \bar{u}_{i0} and \bar{u}_{i1} and, hence, are correlated with \tilde{u}_{it}^1 and \tilde{u}_{it}^0 , respectively. To deal with this endogeneity problem, we use the exogenous variables in Z_{it} that satisfy the conditions in Assumption 2.4. Under this assumption, we can use Z_{it} as a vector of instrumental variables for $\eta_i^0(\beta^0)$ and $\eta_i^1(\beta^1)$, because, in the supposed subgroups, Z_{it} is exogenous to \tilde{u}_{it}^0 and \tilde{u}_{it}^1 under Assumption 2.4 (i) and is correlated with C_i under Assumption 2.4 (ii).⁹

From the above discussion, given that β^1 and β^0 are already identified, the parameters $\tilde{\alpha}^1$, $\tilde{\alpha}^0$, and γ^1 are identified by the following moment conditions:

$$\sum_{t=1}^T E \left[(1, Z'_{it})' (Y_{it}(1) - \tilde{\alpha}^1 - X'_{it}\beta^1 - \gamma^1 \eta_i^0(\beta^0)) \mid D_{it} = 1, \sum_{t=1}^T D_{it} \neq T \right] = 0, \quad (7)$$

$$\sum_{t=1}^T E \left[(1, Z'_{it})' \left(Y_{it}(0) - \tilde{\alpha}^0 - X'_{it}\beta^0 - \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \mid D_{it} = 0, \sum_{t=1}^T D_{it} \neq 0 \right] = 0. \quad (8)$$

The first components of the moment conditions (7) and (8) follow from the exogeneity condition: $\sum_{t=1}^T E [u_{it}^j \mid D_{i1}, \dots, D_{iT}] = 0$, which holds under Assumption 2.2. The others follow from the orthogonal condition in Assumption 2.4 (i). In these moment conditions, Z_{it} works as a vector of instruments for $\eta_i^0(\beta^0)$ and $\eta_i^1(\beta^1)$.¹⁰ $\tilde{\alpha}^1$ and $\tilde{\alpha}^0$ are identified by the moment conditions (7) and (8), respectively, and γ^1 is identified by both of them.

The following formalizes the identification result in this step.

Proposition 3.2. Suppose β^0 and β^1 are already identified. Then, under Assumptions 2.1–2.4, γ^1 and $\tilde{\alpha}^j$ ($j = 0, 1$) are identified by the moment conditions (7) and (8).

A proof is provided in the online Appendix.

Third step

We can finally identify $E[Y_{it}(1)]$ and $E[Y_{it}(0)]$, as follows:

$$E[Y_{it}(1)] = E \left[D_{it} Y_{it} + (1 - D_{it}) (\tilde{\alpha}^1 + X'_{it}\beta^1 + \gamma^1 \eta_i^0(\beta^0)) \right], \quad (9)$$

$$E[Y_{it}(0)] = E \left[(1 - D_{it}) Y_{it} + D_{it} \left(\tilde{\alpha}^0 + X'_{it}\beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \right], \quad (10)$$

⁹ \bar{u}_{i1} in $\eta_i^1(\beta^1)$ and \bar{u}_{i0} in $\eta_i^0(\beta^0)$ can be viewed as measurement errors for $\alpha^1 + \gamma^1 C_i$ and $\alpha^0 + C_i$, respectively. In this view, the variables in Z_{it} work as instruments to deal with the measurement errors.

¹⁰We can use Z_{it} that consist of some or all variables in X_{it} , even though X_{it} are already used in the first step of the identification, because we use different subpopulation, movers, in this step.

where the equations hold under Assumption 2.2. In equation (9), we predict $Y_{it}(1)$ for individuals who are not treated in period t by combining X_{it} and $\eta_i^0(\beta^0)$ with the identified parameters. This is based on the idea of the regression adjustment method (see, for example, Wooldridge 2010, Section 21.3.2). In equation (10), applying the same method to individuals treated in period t , we predict their unobserved potential outcomes $Y_{it}(0)$.

Consequently, under Assumption 2.2, we can identify τ_t^{ate} , as follows:

$$\begin{aligned}\tau_t^{ate} &= E[Y_{it}(1)] - E[Y_{it}(0)] \\ &= E \left[D_{it} \left\{ Y_{it} - \left(\tilde{\alpha}^0 + X'_{it}\beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \right\} + (1 - D_{it}) \left\{ (\tilde{\alpha}^1 + X'_{it}\beta^1 + \gamma^1 \eta_i^0(\beta^0)) - Y_{it} \right\} \right].\end{aligned}\tag{11}$$

The following proposition formalizes the identification result in this step.

Proposition 3.3. Suppose β^j and $\tilde{\alpha}^j$ for $j = 0, 1$, and γ^1 are already identified. Then, under Assumptions 2.1 and 2.2, τ_t^{ate} for $t = 1, \dots, T$ is identified as (11).

A proof is provided in the online Appendix.

The following theorem summarizes the complete identification result.

Theorem 3.1. Let $\theta' \equiv (\beta^{1'}, \beta^{0'}, \tilde{\alpha}^0, \tilde{\alpha}^1, \gamma^1, \tau_1^{ate}, \dots, \tau_T^{ate})$. Suppose the potential outcome model (1) and (2). Then, under Assumptions 2.1–2.4 and 3.1, θ is identified.

A proof is omitted since the result obviously follows from Propositions 3.1–3.3.

The following are some remarks on the identification result.

Remark 3.1. (Potential Outcome Model without Constant Terms) When the potential outcome equations (1) and (2) do not include constant terms, α^1 and α^0 , respectively, the identification procedure does not require the use of Z_{it} . To explain this, suppose here that the potential outcome equations are as follows:

$$\begin{aligned}Y_{it}(1) &= X'_{it}\beta^1 + \gamma^1 C_i + u_{it}^1, \\ Y_{it}(0) &= X'_{it}\beta^0 + C_i + u_{it}^0,\end{aligned}$$

where u_{it}^1 and u_{it}^0 are defined as in Section 2. These equations do not have the constant terms. Under this model and the maintained assumptions, β^j ($j = 0, 1$) can be identified

as (4), as in the first step of the identification procedure.

For the second step, we do not need to identify the transformed constant terms $\tilde{\alpha}^j$ ($j = 0, 1$) in this case, because the transformed equations (5) and (6) in this case do not contain these terms. Thus, we only need to identify the ratio of the fixed effects γ^1 in this step. In the absence of the constant terms, it follows for movers that $\eta_i^1(\beta^1) = \gamma^1 C_i + \bar{u}_i^1$ and $\eta_i^0(\beta^0) = C_i + \bar{u}_i^0$. Then, γ^1 can be identified, without using Z_{it} , as follows:

$$\frac{E \left[\eta_i^1(\beta^1) \mid 0 < \sum_{t=1}^T D_{it} < T \right]}{E \left[\eta_i^0(\beta^0) \mid 0 < \sum_{t=1}^T D_{it} < T \right]} = \frac{\gamma^1 E \left[C_i \mid 0 < \sum_{t=1}^T D_{it} < T \right]}{E \left[C_i \mid 0 < \sum_{t=1}^T D_{it} < T \right]} = \gamma^1,$$

where the first equality follows from Assumption 2.2. The ATE can be identified in the same way as in the third step.

Remark 3.2. (Time-invariant Treatment Effect): When X_{it} does not induce time-variation in the ATE, the ATE becomes time-invariant. This happens when $E[X_{it}]$ is time-invariant or β^1 and β^0 are equivalent. Instead of (11), the time-invariant ATE, denoted by τ^{ate} , can be identified as

$$\tau^{ate} = \frac{1}{T} \sum_{t=1}^T E \left[D_{it} \left\{ Y_{it} - \left(\tilde{\alpha}^0 + X'_{it} \beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \right\} + (1 - D_{it}) \left\{ (\tilde{\alpha}^1 + X'_{it} \beta^1 + \gamma^1 \eta_i^0(\beta^0)) - Y_{it} \right\} \right].$$

In this case, the numbers of moments and parameters to be estimated are reduced, which could decrease the variances of the estimators.

3.2 Estimation

Building on the identification procedure, the vector of parameters θ can be estimated by GMM. The GMM estimator of θ is constructed from the vector of moment functions as follows:

$$\begin{aligned} g_1(W_i, \theta) &= \sum_{t=1}^T D_{it} \cdot \ddot{X}_{it}^1 \left(\ddot{Y}_{it}^1 - \ddot{X}_{it}^{1'} \beta^1 \right), \\ g_2(W_i, \theta) &= \sum_{t=1}^T (1 - D_{it}) \cdot \ddot{X}_{it}^0 \left(\ddot{Y}_{it}^0 - \ddot{X}_{it}^{0'} \beta^0 \right), \\ g_3(W_i, \theta) &= \sum_{t=1}^T 1 \left\{ D_{it} = 1, \sum_{t=1}^T D_{it} \neq T \right\} \cdot (1, Z'_{it})' \left(Y_{it} - \tilde{\alpha}^1 - X'_{it} \beta^1 - \gamma^1 \eta_i^0(\beta^0) \right), \end{aligned}$$

$$\begin{aligned}
g_4(W_i, \theta) &= \sum_{t=1}^T 1 \left\{ D_{it} = 0, \sum_{t=1}^T D_{it} \neq 0 \right\} \cdot (1, Z'_{it})' \left(Y_{it} - \tilde{\alpha}^0 - X'_{it}\beta^0 - \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right), \\
g_5(W_i, \theta) &= D_{i1} \left\{ Y_{i1} - \left(\tilde{\alpha}^0 + X'_{i1}\beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \right\} \\
&\quad - (1 - D_{i1}) \left\{ (\tilde{\alpha}^1 + X'_{i1}\beta^1 + \gamma^1 \eta_i^0(\beta^0)) - Y_{i1} \right\} - \tau_1^{ate}, \\
&\quad \vdots \\
g_{4+T}(W_i, \theta) &= D_{iT} \left\{ Y_{iT} - \left(\tilde{\alpha}^0 + X'_{iT}\beta^0 + \frac{1}{\gamma^1} \eta_i^1(\beta^1) \right) \right\} \\
&\quad - (1 - D_{iT}) \left\{ (\tilde{\alpha}^1 + X'_{iT}\beta^1 + \gamma^1 \eta_i^0(\beta^0)) - Y_{iT} \right\} - \tau_T^{ate},
\end{aligned}$$

where $W_i = (\{Y_{it}, D_{it}, X'_{it}, Z'_{it}\}_{t=1}^T)'$. $g_1(W_i, \theta)$ and $g_2(W_i, \theta)$ are the moment functions to estimate β^1 and β^0 , respectively, which are derived from the first step of the identification procedure. $g_3(W_i, \theta)$ and $g_4(W_i, \theta)$ are the moment functions to estimate $\tilde{\alpha}^1, \tilde{\alpha}^0$, and γ^1 , which are analogues of moment conditions (7) and (8), respectively. For each $t = 1, \dots, T$, $g_{4+t}(W_i, \theta)$ is the moment function to estimate τ_t^{ate} by the prediction, which is derived from equation (11). The above moment functions can be easily modified for the cases considered in Remarks 3.1 and 3.2.

In the GMM framework, we can also test whether $\gamma^1 = 1$ or not. In our specification, this is equivalent to a test for whether the treatment effect heterogeneity depends on unobserved fixed effects. We could apply GMM-based tests for the null hypothesis: $\gamma^1 = 1$. If the null hypothesis is correct, we can use the usual fixed effects estimator to estimate the ATE.

4 Monte Carlo Simulation

In this section, we conduct a Monte Carlo experiment to study the finite sample behavior of the proposed estimator. Some additional Monte Carlo experiments are conducted in the online Appendix. We here consider two kinds of data generating processes (DGPs), DGP1 and DGP2, which consist of different outcome models and a same treatment assignment model with $T = 2$. The outcome models are as follows: For $t = 1, 2$,

$$Y_{it} = \begin{cases} 2 + X_{it,1} - 0.5X_{it,2} + C_i + u_{it}^0 & \text{if } D_{it} = 0 \\ 1 + 2X_{it,1} + 0.5X_{it,2} + \gamma_1 C_i + u_{it}^1 & \text{otherwise} \end{cases},$$

where $\gamma^1 = 1$ in DGP1 and $\gamma^1 = 3$ in DGP2. The treatment assignment model is $D_{it} = 1 \{-2 - X_{it,1} + X_{it,2} + C_i + u_{it}^D \geq 0\}$ for $t = 1, 2$. In both DGPs, $(X_{i1,1}, X_{i2,1}, C_i)'$ are drawn from a multivariate normal distribution with mean $(1, 1.5, 1)'$, standard deviation 1, and pairwise covariances 0.3; $(X_{i1,2}, X_{i2,2})'$ are drawn from a multivariate normal distribution with mean $(1, 1.5)'$, standard deviation 1, and pairwise covariance 0.3. Note here that $X_{it,1}$ is correlated with C_i while $X_{it,2}$ is not. The disturbance terms are drawn from the following multivariate normal distributions:

$$\begin{pmatrix} u_{i1}^0 \\ u_{i1}^1 \\ u_{i2}^0 \\ u_{i2}^1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.3 & 0.2 \\ & 1 & 0.2 & 0.3 \\ & & 1 & 0.5 \\ & & & 1 \end{pmatrix} \right), \begin{pmatrix} u_{i1}^D \\ u_{i2}^D \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ & 1 \end{pmatrix} \right)$$

In addition, we set an exogenous variable W_{it} as $W_{it} = C_i + u_{it}^W$, where $(u_{i1}^W, u_{i2}^W)'$ is drawn from a zero mean multivariate normal distribution with standard deviation 1 and pairwise covariance 0.3. This exogenous variable satisfies the conditions in Assumption 2.4.

Two remarks should be made on the DGPs. First, the difference between DGP1 and DGP2 is in the coefficients for C_i in the outcome equations. While DGP1 has the same coefficients for C_i in the treated and untreated outcome equations, DGP2 has different coefficients between them. Therefore, treatment effect heterogeneity does not depend on C_i in DGP1, but it does in DGP2. Second, each of the three variables W_{it} , $X_{it,1}$, and $X_{it,2}$ can be used as an exogenous variable in Z_{it} , whose role is described in Section 3, because they satisfy the conditions in Assumption 2.4. W_{it} and $X_{it,1}$ are unconditionally correlated with C_i , whereas $X_{it,2}$ is not. But $X_{it,2}$ is correlated with C_i after conditioning on D_{it} .

In this simulation, we compare the efficient two-step GMM estimator derived from the proposed identification procedure with the OLS estimator and the FE estimator in DGP1 and DGP2. The parameters of interest are the ATEs in periods 1 and 2. In DGP1, the true values of the ATEs in periods 1 and 2 are 1 and 2, respectively; in DGP 2, the true values of the ATEs in periods 1 and 2 are 3 and 4, respectively. For the proposed estimator, we use five different sets of Z_{it} : $\{W_{it}\}$, $\{X_{it,1}\}$, $\{X_{it,2}\}$, $\{X_{it,1}, X_{it,2}\}$, and $\{W_{it}, X_{it,1}, X_{it,2}\}$. The fourth set uses all the exogenous variables that are included in the potential outcome model; the last set uses all the exogenous variables. In both DGPs, the ratios of movers, always-treated individuals, and never-treated individuals are about 31.7%, 13.5%, and 54.8%, respectively.

Table 1 reports the results of 1,000 simulations with sample sizes $N = 200, 500, \text{ and } 800$. Panels I and II in Table 1 report the simulation results for DGP1 and DGP2, respectively. Several findings are worth noting. First, the OLS estimator is severely biased in both DGP1 and DGP2 due to the presence of unobserved fixed effects. Second, the FE estimator is not biased in DGP1, whereas it is biased in DGP2. This is because the treatment effect heterogeneity depends on unobserved fixed effects in DGP2. Third, in both DGP1 and DGP2, the proposed estimator has little bias, regardless of the type of Z_{it} . Fourth, even in DGP1, where the FE estimator is consistent, the performance of the proposed estimator is compatible with that of the FE estimator in terms of RMSEs. Fifth, comparing the different sets of Z_{it} , the proposed estimator has smaller RMSEs when Z_{it} consists of $X_{it,1}$ and/or $X_{it,2}$ than when Z_{it} consists of only W_{it} in both DGPs. In the online Appendix, additional simulation results are provided, where we evaluate the finite sample behavior in three cases: (i) fraction of movers is small, (ii) γ^1 is close to zero, and (iii) T is large.

5 Empirical Application

In this section, I apply the proposed approach to study the impact of a mother’s smoking during pregnancy on her child’s birth weight and its heterogeneity. There are many works that study the average smoking effect and its heterogeneity under various identifying and functional form assumptions (e.g., Abrevaya, 2006; Veiga and Wilder, 2008; Walker et al., 2009; Abrevaya et al., 2015; Lee et al., 2017). Walker et al. (2009) examine the difference in the average smoking effect between teen mothers and adult mothers. Abrevaya et al. (2015) and Lee et al. (2017) examine the heterogeneity of the average smoking effect in terms of the mother’s age in their application.

The analysis uses the matched panel data set constructed by Abrevaya (2006) from the U.S. Natality Data Sets for 1990–1998. Because the original data set does not have unique identifiers for mothers, he carefully matched mothers to children mainly based on pairs of the child’s state of birth and the mother’s state of birth, which have a small number of observations, and then constructed the matched panel data set. I select “matched panel #3,” because it is the most conservatively constructed. The same data set is also analyzed by Arellano and Bonhomme (2012) using the CRC panel data model, and by Jun et al. (2016) using a nonseparable panel data model. Because the numbers of births in the original data set are different among mothers, I focus on mothers who bore three children during the

observed period. The final sample contains 12,360 mothers. Among them, 1,349 mothers were smoking during their first pregnancy, 1,371 mothers were smoking during their second pregnancy, and 1,437 mothers were smoking during their third pregnancy. The ratio of movers in the sample is 11.7%.

Using this sample, I estimate the average effect of a mother’s smoking during pregnancy on her child’s birth weight at each birth time using the potential outcome model (1) and (2). In the model, D_{it} is an indicator for a mother’s smoking, where $D_{it} = 1$ indicates that i th mother was smoking during pregnancy for her t th birth and $D_{it} = 0$ indicates otherwise. The potential outcomes $Y_{it}(1)$ and $Y_{it}(0)$ are the child’s birth weight for the i th mother’s t th birth if she was smoking and not smoking during the pregnancy, respectively. X_{it} includes the age of the mother at the time of the birth, the gender of the child, dummy variables for birth times, dummy variables indicating the existence of prenatal visits, and the value of the “Kessner” index for the quality of prenatal care (for details, see Abrevaya, 2006). Unobserved fixed effects C_i are supposed to represent a mother’s health-related lifestyle factor (Jun et al., 2016, p.307). Under the potential outcome model (1) and (2), we allow the heterogeneity of the smoking effect to depend on the observed and unobserved characteristics. Note, however, that our model still cannot deal with some identification issues pointed out by Abrevaya (2006, Section 5) (e.g., feedback effect from a prior birth outcome and correlated changes in maternal behavior).

To apply the proposed estimator, we want exogenous variables Z_{it} that satisfy the conditions in Assumption 2.4. I use three sets of exogenous variables for Z_{it} : The age of the mother at her t th birth, the years of education of the mother, and both. The age is already included in X_{it} , while the education is not. Recall that we can use variables included in X_{it} for Z_{it} . Both variables are probably unconditionally correlated with unobserved fixed effects, because a woman’s decision about the timing of birth is likely to depend on her unobserved lifestyle factors and the lifestyle factors should be associated with her age and education, respectively. Even if they are not unconditionally correlated with the fixed effects, they should be correlated after conditioning on the smoking indicator. We also suppose that they are uncorrelated with error terms because we control for a sufficient set of the observables and unobserved fixed effects.

Table 2 presents the estimation results for the coefficient parameters in the transformed potential outcome equations (5) and (6), as well as the differences between them for equations (6) and (5) (i.e., $\beta^1 - \beta^0$), which indicate the heterogeneity of the smoking effect

depending on the observed covariates. Here, I applied the efficient two-step GMM estimator using the age and years of education as a set of Z_{it} . The result does not vary so much when using different Z_{it} . The effect of smoking is significantly higher for the second and third birth times and for mothers with no prenatal visits, and changes quadratically with respect to age. All of the variables, except for the no prenatal visit, first visit in the third trimester, second child, and third child, have the same signs of the difference estimates as those in Abrevaya’s (2006, Table VII) results. Note that Abrevaya (2006) does not consider the interaction effect of smoking with each birth time. The significant estimate of γ^1 indicates the presence of the heterogeneity of the smoking effect that depends on unobserved fixed effects.

Table 3 presents the estimation results for the average smoking effect using the proposed estimator and the FE estimator. For the FE estimator, I use a fixed-effects regression model including the smoking indicator, observed covariates X_{it} , and interactions between the smoking indicator and X_{it} . The proposed estimators are the two-step GMM estimator using the three sets of Z_{it} . The first three rows of Table 3 show the estimation results of the time-varying average effect using model (1) and (2) for the proposed estimator and using model (3) for the FE estimator. The last row shows the estimation results of the time-invariant average effect, where we impose the restriction $\beta^1 = \beta^0$ on model (1) and (2) and model (3). The results of both estimators show that a mother’s smoking has a negative average effect on her child’s birth weight at each birth time, and the effect is time-varying. Although the results of the proposed estimator are somewhat variable depending on the choice of Z_{it} , they show that the average smoking effect worsens with additional birth time. This might be because mothers who smoke during later pregnancy may be heavier smokers and/or because the negative effects of smoking accumulate over time.¹¹ The proposed estimator provides lower estimates at each birth time than that of the FE estimate. We might suspect that the FE estimates are upward biased for the average effects because it ignores the heterogeneity of the smoking effect that depends on the mother’s unobserved characteristics. The results of the proposed estimator also show relatively more negative effects of a mother’s smoking compared to the results of the previous works.

¹¹Note that the framework in this paper cannot explicitly capture the accumulated effect of smoking, unlike the dynamic treatment framework considered, for example, by Lechner (2009).

6 Conclusion

This paper proposed a new panel data approach to identify and estimate the time-varying ATE. The approach can identify and consistently estimate the ATE even when treatment effect heterogeneity depends on unobserved fixed effects. In such a case, existing panel data approaches identify the ATE for limited subpopulations only. In contrast, the proposed approach can identify and estimate the ATE for the entire population. The approach depends on the linear fixed effects specification to the potential outcome equations, wherein scalar fixed effects additively enter the two potential outcome equations with different coefficients, and uses exogenous variables that are correlated with fixed effects conditional on treatment assignment. I recommend that empirical researchers who use panel data apply this approach when treatment effect heterogeneity seems to depend on unobserved individual characteristics, and their interest is in the ATE for the entire population.

References

- Abrevaya, J. (2006), “Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach,” *Journal of Applied Econometrics*, 21, 489–519.
- Abrevaya, J., Hsu, Y-C., and Lieli, R. P. (2015), “Estimating Conditional Average Treatment Effects,” *Journal of Business & Economic Statistics*, 33, 485–505.
- Arellano, M., and Bonhomme, S. (2012), “Identifying Distributional Characteristics in Random Coefficients Panel Data Models,” *Review of Economic Studies*, 79, 987–1020.
- Athey, S., and Imbens, G. W. (2017), “The State of Applied Econometrics: Causality and Policy Evaluation,” *Journal of Economic Perspectives*, 31, 3–32.
- Chamberlain, G. (1982), “Multivariate Regression Models for Panel Data,” *Journal of Econometrics*, 18, 5–46.
- (1992), “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60, 567–596.
- Callaway, B., and Sant’Anna, P. (2018). “Difference-in-Differences With Multiple Time Periods,” Working Paper.
- Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013), “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 81, 535–580.

- Gibbons, C. E., Suárez Serrato, J. C., and Urbancic, M. B. (2018), “Broken or Fixed Effects?,” *Journal of Econometric Methods*, 8.
- Graham, B. S., and Powell, J. L. (2012), “Identification and Estimation of Average Partial Effects in “Irregular” Correlated Random Coefficient Panel Data Models,” *Econometrica*, 80, 2105–2152.
- Jun, S. J., Lee, Y., and Shin, Y. (2016), “Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach,” *Journal of Business & Economic Statistics*, 34, 302–311.
- Lechner, M. (2009), “Sequential Causal Models for the Evaluation of Labor Market Programs,” *Journal of Business & Economic Statistics*, 27, 71–83.
- Lee, S., Okui, R., and Whang, Y.-J. (2017), “Doubly robust uniform confidence band for the conditional average treatment effect function,” *Journal of Applied Econometrics*, 32, 1–19.
- Sakaguchi, S. (2016), “Estimation of Time-varying Average Treatment Effects Using Panel Data when Unobserved Fixed Effects Affect Potential Outcomes Differently,” *Economics Letters*, 146, 82–84.
- Veiga, P. V., and Wilder, R. P. (2008), “Maternal Smoking During Pregnancy and Birthweight: A Propensity Score Matching Approach,” *Maternal and Child Health Journal*, 12, 194–203.
- Walker, M. B., Tekin, E., and Wallace, S. (2009), “Teen Smoking and Birth Outcomes,” *Southern Economic Journal*, 75, 892–907.
- Wooldridge, J. M. (2005), “Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models,” *Review of Economics and Statistics*, 87, 385–390.
- (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, Cambridge, MA: MIT Press.

Table 1: Monte Carlo Simulation Results

		N=200			N=500			N=800		
True Value		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
Panel I: Simulation results for DGP1										
ATE for 1st period										
OLS	1	1.147	0.257	1.175	1.173	0.166	1.184	1.162	0.130	1.170
FE	1	-0.015	0.258	0.258	0.008	0.167	0.167	-0.004	0.134	0.134
Proposed Estimator										
(1)	1	0.025	0.195	0.196	0.034	0.130	0.134	0.255	0.107	0.110
(2)	1	0.023	0.190	0.192	0.030	0.126	0.129	0.021	0.099	0.101
(3)	1	0.038	0.177	0.181	0.042	0.117	0.124	0.037	0.092	0.099
(4)	1	0.034	0.180	0.183	0.037	0.118	0.123	0.028	0.093	0.097
(5)	1	0.036	0.184	0.187	0.035	0.123	0.128	0.025	0.098	0.101
ATE for 2nd period										
OLS	2	1.157	0.255	1.184	1.169	0.160	1.180	1.162	0.131	1.169
FE	2	-0.008	0.265	0.265	0.004	0.169	0.169	-0.002	0.132	0.132
Proposed Estimator										
(1)	2	0.041	0.203	0.207	0.044	0.132	0.140	0.041	0.111	0.118
(2)	2	0.027	0.191	0.193	0.032	0.131	0.134	0.025	0.100	0.103
(3)	2	0.040	0.198	0.201	0.048	0.122	0.131	0.045	0.097	0.107
(4)	2	0.031	0.194	0.196	0.037	0.125	0.130	0.029	0.098	0.102
(5)	2	0.031	0.196	0.198	0.036	0.129	0.134	0.028	0.102	0.106
Panel II: Simulation results for DGP2										
ATE for 1st period										
OLS	3	2.833	0.463	2.881	2.881	0.300	2.897	2.882	0.229	2.891
FE	3	0.921	0.378	0.996	0.953	0.247	0.984	0.954	0.186	0.972
Proposed Estimator										
(1)	3	0.092	0.310	0.323	0.076	0.203	0.217	0.061	0.169	0.180
(2)	3	0.081	0.287	0.298	0.077	0.192	0.207	0.062	0.155	0.167
(3)	3	0.085	0.285	0.297	0.077	0.182	0.197	0.069	0.146	0.161
(4)	3	0.071	0.295	0.304	0.072	0.197	0.210	0.065	0.153	0.166
(5)	3	0.074	0.317	0.326	0.068	0.215	0.225	0.061	0.174	0.184
ATE for 2nd period										
OLS	4	2.863	0.471	2.901	2.882	0.290	2.897	2.881	0.221	2.890
FE	4	0.938	0.387	1.015	0.949	0.232	0.977	0.953	0.189	0.972
Proposed Estimator										
(1)	4	0.106	0.324	0.341	0.082	0.214	0.229	0.071	0.176	0.190
(2)	4	0.085	0.320	0.331	0.076	0.212	0.225	0.066	0.171	0.183
(3)	4	0.092	0.299	0.313	0.083	0.195	0.211	0.075	0.159	0.176
(4)	4	0.079	0.326	0.335	0.069	0.212	0.223	0.069	0.167	0.181
(5)	4	0.078	0.331	0.339	0.070	0.223	0.234	0.064	0.181	0.192

Note: True Value is the true value of the ATE. Bias, SD, and RMSE are the mean bias, standard deviation, and the root mean squared error of the estimates across the simulations, respectively. For the Proposed Estimator, the rows (1), (2), (3), (4), and (5) report the results of the proposed estimator when Z_{it} is $\{W_{it}\}$, $\{X_{it,1}\}$, $\{X_{it,2}\}$, $\{X_{it,1}, X_{it,2}\}$, and $\{W_{it}, X_{it,1}, X_{it,2}\}$, respectively.

Table 2: Estimates of parameters in equations (5) and (6)

Variables	Estimates		
	Equation (5)	Equation (6)	Differences ($\beta^1 - \beta^0$)
Male	134.39 (14.82)	141.27 (4.64)	-6.88 (15.50)
Age	231.90 (2.18)	222.78 (1.68)	9.13 (2.78)
Age ²	-4.00 (0.07)	-3.55 (0.06)	-0.45 (0.09)
Kessner index = 2	-45.49 (27.62)	-61.99 (9.80)	16.50 (29.22)
Kessner index = 3	-169.53 (41.72)	-162.72 (24.45)	-6.81 (48.67)
No prenatal visit	-147.71 (61.03)	23.11 (47.86)	-170.82 (78.04)
First prenatal visit in 2nd trimester	43.61 (28.92)	79.93 (11.84)	-36.32 (31.40)
First prenatal visit in 3rd trimester	200.39 (52.00)	163.00 (30.18)	37.39 (60.21)
Second child	-6.54 (15.73)	29.76 (6.06)	-36.30 (16.89)
Third child	-72.97 (17.16)	27.05 (10.28)	-100.02 (20.08)
$\tilde{\alpha}^1$	-28.96 (16.92)		
$\tilde{\alpha}^0$		18.69 (17.77)	
γ^1		0.30 (0.05)	

Note: Robust standard errors are presented in the parentheses.

Table 3: Estimates of average effect of mother's smoking during pregnancy on her child's birth weight

	FE Estimator	Proposed Estimator		
		(1)	(2)	(3)
Average effect (first birth)	-121.47 (15.56)	-154.84 (17.18)	-219.68 (38.63)	-147.40 (16.27)
Average effect (second birth)	-76.21 (17.11)	-216.96 (16.73)	-300.42 (49.56)	-214.05 (15.83)
Average effect (third birth)	-128.17 (15.93)	-324.88 (15.92)	-403.12 (47.31)	-319.94 (15.42)
Average effect (time-invariant)	-244.23 (11.12)	-216.14 (22.68)	-217.87 (18.22)	-189.36 (11.77)

Note: Robust standard errors are presented in the parentheses. For the Proposed Estimator, the columns (1), (2), and (3) report the results of the proposed estimator using the age, the education, and both of them for Z_{it} , respectively.