NBER WORKING PAPER SERIES

A FLEXIBLE, HETEROGENEOUS TREATMENT EFFECTS DIFFERENCE-IN-DIFFERENCES ESTIMATOR FOR REPEATED CROSS-SECTIONS

Partha Deb
Edward C. Norton
Jeffrey M. Wooldridge
Jeffrey E. Zabel

Working Paper 33026
http://www.nber.org/papers/w33026

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024, Revised April 2026

The authors thank conference participants at the New York Camp Econometrics XVIII, the American Society of Health Economists Annual Meeting, the 31st European Workshop on Econometrics and Health Economics, the 2024 Annual Health Econometrics Workshop, the Chinese Economists Society, the 2025 North American Winter Meeting of the Econometrics Society, and seminar participants at Oregon Health Sciences University and Universitat Pompeu Fabra. We also thank Anjelica Gangaram, Govert Bijwaard and Irene Botosaru, the Editor and three anonymous reviewers, for their valuable comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Flexible, Heterogeneous Treatment Effects Difference-in-Differences Estimator for Repeated Cross-Sections

Partha Deb, Edward C. Norton, Jeffrey M. Wooldridge, and Jeffrey E. Zabel

NBER Working Paper No. 33026
October 2024, Revised April 2026
JEL No. C01, C21, C50

ABSTRACT

This paper proposes a method to estimate treatment effects in difference-in-differences designs for repeated cross-section data in which the treatment start is staggered over time and treatment effects are heterogeneous by group, time, and observation-level covariates. We show that a linear-in-parameters regression specification with a sufficiently flexible functional form consisting of group-by-time treatment effects, two-way fixed effects, and interaction terms yields consistent estimates of heterogeneous treatment effects under general conditions. We also show that our method is identical to an imputation estimator. Under homoskedasticity assumptions the estimators are efficient, and aggregation of the treatment effects and inference are straightforward. We illustrate the use of this flexible linear model estimated by OLS with covariates (X) – FLEX – with an empirical example and provide comparisons to some benchmark estimators.

Partha Deb
Hunter College of the City University of New York
Department of Economics
and NBER
partha.deb@hunter.cuny.edu

Edward C. Norton
University of Michigan
School of Public Health
Department of Health Management and Policy,
and Department of Economics
and NBER
ecnorton@umich.edu

Jeffrey M. Wooldridge
Michigan State University
wooldri1@msu.edu

Jeffrey E. Zabel
Tufts University
jeff.zabel@tufts.edu

1 Introduction

The difference-in-differences (DID) study design is an important tool for causal inference in economics. In recent years, there has been an extraordinary number of new theoretical papers on how to obtain consistent estimates in regression-based implementations when the treatment is staggered over time across treated units and the treatment effects are heterogeneous across units and over time (e.g., Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). These approaches were motivated by the discovery that the so-called two-way fixed effects model that imposes a constant treatment effect across units and time does not generally recover an interesting causal effect; in particular, it does not identify a suitably defined weighted average of causal effects across units and time (e.g., de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021).

Consider a scenario where treatment is applied to groups of observed units and that collections of groups are treated for the first time in different time periods, which we define as a cohort. The above citations specify a model where the treatment effect varies by time and cohort. If there is treatment heterogeneity at the group level, these cohort-level treatment effects are a weighted average of the group-level treatment effects for the groups in each cohort and hence they mask this group-level heterogeneity. In this paper, we specify a flexible linear regression model where the treatment effects are allowed to vary by time and cohort or group. Given the simplicity of the underlying linear regression, aggregated treatment effects, such as the average treatment effect on the treated (ATET) (or aggregates at the cohort or exposure-time levels), can be calculated easily from the estimated coefficients. Inference on disaggregate and aggregate effects also follows naturally from estimating the

linear regression specification by OLS.

Much of the recent work on staggered interventions has focused on the case of panel data. Exceptions include Callaway and Sant'Anna (2021) and Borusyak et al. (2024) who consider repeated cross-sections. Callaway and Sant'Anna (2021) extend Abadie (2005) and Sant'Anna and Zhao (2020) from the two-period case, and propose regression-based and doubly robust estimators that combine regression with propensity score weighting. The Callaway and Sant'Anna (2021) approach is based on 2×2 comparisons of treated and control units, using the period just before the intervention as the reference period; as such, it is in the spirit of event study estimators of pre-treatment and treatment effects. For reasons that will become clear below, we refer to these as "lags-and-leads" approaches. Borusyak et al. (2024) is what we will characterize as a "lags-only" approach; effectively, all pre-treatment periods are aggregated to form a reference period for subsequent treated periods (rather than using only the time period just before the intervention).

Our approach extends the regression-based panel data method in Wooldridge (2025) to repeated cross-sections. Consequently, both lags-only and lags-and-leads approaches can be covered in the same framework, where the estimation method is ordinary least squares (OLS) on suitably defined treatment variables, group and time fixed effects, and various interactions with observation-level regressors. Additional advantages include ease of inference – because we are applying OLS to repeated cross-sections – and the ability to directly estimate moderating (heterogeneous) effects as a function of observed covariates, along with providing valid standard errors for these effects. Moreover, by studying coefficients on the group/cohort indicators and interactions with time dummies and covariates, one can learn about selection effects and the nature of any heterogeneous trends.

Our paper makes several contributions to the literature. We make two theoretical contributions that are presented in Section 2. First, we propose FLEX—flexible linear model estimated by OLS with covariates (X)—for heterogeneous treatment effects in a difference-in-differences study design with repeated cross-section data and staggered treatment starting times. We lay out the assumptions required for this pooled-regression specification to produce consistent average treatment effects on the treated. The linear-in-parameters specification, estimated using a single linear regression, delivers consistent parameter estimates, allows for a flexible functional form with respect to individual-level covariates, and provides access to the OLS toolbox for inference and specification testing.

Second, we prove that the FLEX treatment effect parameter estimates are consistent estimates of the group-time heterogeneous treatment effects that can be obtained in the repeated cross-section setting by an imputation method. To be precise, the FLEX estimates are identical to those from the Borusyak et al. (2024) imputation estimator when treatment heterogeneity is specified at the cohort-year level. This imputation estimator extends both Borusyak et al. (2024) and Wooldridge (2025) to the group-time level for repeated cross-section data. We show that applying OLS to the FLEX model using all of the data is equivalent to an imputation estimator for repeated cross-sections. Unlike imputation, which requires special coding of standard errors or computationally extensive bootstrapping, FLEX is a one-step OLS estimator with standard errors readily available.

In Section 3, we develop the regression specification, and describe the implementation, of FLEX. We demonstrate its flexibility in its ability to control for heterogeneous time trends (subsection 3.3). Given the simplicity of this underlying linear regression, in Section 3.1, we show that it is straightforward to aggregate the heterogeneous treatment effects to obtain

higher-level treatment effects such as the event study and the average treatment effect on the treated (ATET). Inference on disaggregate and aggregate effects also follow naturally from the linear regression specification. The nature of our approach also makes it easy to impose restrictions when preferred, or when the data structure demands it.

In this context, we develop a consistent language and terminology when referring to the different DID models in the staggered treatment setting. We distinguish between models that include lags-only versus lags-and-leads, models with static versus dynamic treatment effects, and models with homogeneous versus heterogeneous treatment effects where the latter means that the treatment effects vary by group or cohort. For example, the so-called two way fixed effects model is a static homogeneous model, the event study model is a dynamic homogeneous model, and the FLEX is a dynamic heterogeneous model. In Section 3.2, we discuss other approaches that have become popular in empirical research (Cengiz et al., 2019; Callaway and Sant'Anna, 2021).

Along with this consistent terminology, FLEX provides flexibility and transparency. We allow for covariates to be entered individually and possibly interacted with the treatment effects. We write down our model as one regression such that the number of parameters, the specification of the treatment effects, and how covariates are specified is clear. These components are not as obvious with other approaches, such as those mentioned in the preceding paragraph.

In addition to the theoretical results, we demonstrate the use of FLEX with an empirical example of how an immigration enforcement program named Secure Communities may have affected enrollment in a government nutrition program (SNAP). Effects of this policy on SNAP have been substantively examined by East et al. (2023) and Alsan and Yang (2024).

This policy-relevant example uses individual-level data from repeated cross-sections. The policy was implemented across counties in the US in a staggered way over several years. The policy may not have constant treatment effects at the group-time (county-year) level, so we allow for heterogeneous treatment effects. We compare the estimated average treatment effects on the treated (ATET) and their standard errors with those obtained using "benchmarks" that assume some degree of homogeneity and the popular method of Callaway and Sant'Anna (2021). The example demonstrates the features of our proposed FLEX estimator for repeated cross-sections with heterogeneous treatment effects, including being easy to implement, transparent about the number of parameters estimated, and allowing flexible controls for individual-level covariates. Moreover, FLEX produces legitimately smaller standard errors than other methods, suggesting at least some efficiency gains. We also show how the results from our regressions can be displayed in a variety of commonly used graphical forms.

2 Theory

2.1 The Population Setting, Definitions, and Assumptions

Here, we set out a framework that focuses on the analysis of repeated cross-sections where new units, i, belonging to one and only one group g, where $g = 1, 2, ..., G$, are (randomly) sampled from the population in every time period $t = 1, 2, ..., T$. Our framework also applies to panel data settings, most transparently when units i in groups g are followed over time t. In that case we are studying a stable population. When the data available are repeated cross-sections, we assume the existence of a stable population from which units are randomly sampled in each time period. In defining parameters and stating assumptions, we assume

that the populations are the same across t. But, in practice, there may be changes in the population across t, a problem that can, at least partly, be overcome by controlling for observable characteristics.

We now turn to the policy setting. We assume that the intervention occurs at the group level where groups are indexed by $g = 1, 2, ..., G$. Each group, among those that receive treatment in the period of observation $t = 1, 2, ..., T$, receives the intervention for the first time in a time period from $t = 2, ..., T$ implying that there are no "always treated" groups. One or more groups ($< G$) receive the intervention for the first time in a particular time period. Let $c(g)$ denote the first time period in which treatment is received by units in group g. We follow the existing literature by referring to each collection of groups associated with $c(g)$ as a cohort, and by identifying each cohort by the time at which the groups within the cohort first receive treatment, i.e., $c(g) \in 2, ..., T$. Therefore, by definition, different cohorts have different values of $c(g)$. Let $\bar{g}$ groups receive treatment by time period $c(\bar{g}) \leq T$. Without loss of generality, let groups $g = 1, 2, ..., \bar{g}$ index the treated groups ordered by their cohort membership so that $c(1) \leq c(2) \leq ... \leq c(\bar{g})$. The never-treated groups can be indexed by $g = \bar{g} + 1, ..., G$. In what follows, it is convenient to refer to never-treated groups as being "treated" at ∞, i.e., $c = \infty$. Note that new entrant groups are not needed for each and every period. Also, for each group g there is a unique cohort, $c(g)$. There can be multiple groups within the same treatment cohort.

Assumption 2.1 (SUTVA)**.** For each unit i in the population, the potential outcome depends only on unit i's assignment and not on the assignment of the other units. In other

words,

$$Y_{it}\left(g_1, ..., g_{i-1}, g_i, g_{i+1}, ...\right) = Y_{it}\left(g_i\right),\ t = 1, ..., T,$$

where g_j is the group assignment for unit j. $\square$

The Stable Unit Treatment Value Assumption (SUTVA) [Rubin (1977,1986); Lechner (2011)] ensures that we observe one potential outcome for each unit i corresponding to i's treatment assignment. It also rules out spillover effects. Because we are assuming the treatment is an absorbing state, given SUTVA we need only index the potential outcomes by a single treatment assignment. Also, in stating Assumption 2.1, we allow for the possibility of an infinite or finite population; later, we assume independent sampling for each time period.

Let $R_g \in R_1, ..., R_{\bar{g}}, R_{\bar{g}+1}, ..., R_G$ denote a binary indicator for group membership. Let the single indicator variable $\{R_g\}^c$ denote the subset of groups in cohort c. The ATETs commonly of interest in staggered DID settings are the mean differences in the potential outcomes using $Y_t\left(\infty\right)$ as the reference outcome in a treated period t, i.e., the counterfactual untreated counterparts:

$$\tau_{gt} = E\left[Y_t(g) - Y_t\left(\infty\right)|R_g = 1\right],\ t = c(g), ..., T;\ \ g = 1, ..., \bar{g}. \tag{2.1}$$

For each (eventually) treated group g, τ_{gt}, $t = c(g), ..., T$ are the ATETs in all time periods, including the first period in which treatment is received, $c(g)$, and all following ones through the end of the observation period, T.

Assumption 2.2 (No Bad Controls, NBC)**.** For a $1 \times K$ vector of covariates $\{\mathbf{X}_t(g)$: $t = 1, ..., T; g = 1, ..., G\}$, which may be time-varying, the covariates are the same across all

potential treatment assignments: $\mathbf{X}_t(g) = \mathbf{X}_t(\infty) = \mathbf{X}_t$ for all $g \in \{1, ..., G\}$ and $t = 1, ..., T$.

□

Assumption 2.2 implies that the covariates are not influenced by treatment status. Much of the literature on difference-in-differences in panel data contexts assumes the controls are dated prior to the first intervention date and do not change over time. This restriction helps ensure that one is not including 'bad controls' in the analysis—that is, elements in $\mathbf{X}_t$ that might be affected in the current or future periods by the intervention. Borusyak et al. (2024) allow for time-varying covariates, but they treat them as nonrandom (or fixed in repeated samples). Necessarily, this means the covariates cannot be changed in response to treatment assignment. It also implies the covariates are strictly exogenous with respect to shocks to Y_t, something we do not need to explicitly assume. Caetano et al. (2024) formalizes the use of time-varying covariates, including to a situation we do not consider, i.e., to allow time-varying covariates to be affected by the treatment. We do not index the $\mathbf{X}_t$ using potential outcomes notation [such as $\mathbf{X}_t(g)$], and so we are maintaining that the covariates do not change with the treatment assignment. (This is different from saying that the treatment assignment cannot depend on $\mathbf{X}_t$—which, of course, we allow.) Allowing $\mathbf{X}_t$ to have time variation means that we can include predictors of the outcome whose paths are not influenced by treatment. For example, the outcome may be affected by local weather conditions, which are time-varying but not influenced by the treatment. More commonly, in the repeated cross-section case, the controls will necessarily be time varying—due to sampling variation across time periods.

Difference-in-differences analyses requires two additional key assumptions. The first as-

sumption rules out anticipatory changes in the potential outcomes prior to the intervention occuring for each eventually treated group. The second, parallel trends assumption, is stated conditional on group indicators that are less coarse than the cohort indicators. Consequently, because cohorts typically include more than one group and because the groups within a cohort may be heterogeneous, by conditioning on group indicators the assumptions are more general. Note that it is always possible to treat cohorts as "groups" in our framework.

Assumption 2.3 (Conditional No Anticipation, CNA)**.** For groups $g \in 1, 2, ..., \bar{g}$ and $t \in \{1, ..., c(g) - 1\}$,

$$E\left[Y_t(g)|R_1, ..., R_G, \mathbf{X}_t\right] = E\left[Y_t\left(\infty\right)|R_1, ..., R_G, \mathbf{X}_t\right]. \qquad \square$$

This simply means that, in any time period before the intervention occurs for any of the eventually treated groups, the potential outcomes are the same as the potential outcomes in the never treated state. This assumption can be violated if units within groups that are eventually treated anticipate the intervention and change their behavior prior to the first period of intervention. This formulation is adapted from Wooldridge (2025) (also in Wooldridge, 2023) which is stated for the panel data case.

Let $P_t \in P_1, P_2, ..., P_T$ denote binary indicators for observations in time periods $t = 1, 2, ..., T$. Then, in the staggered intervention case without exit, the time-varying treatment indicator is

$$W_t = \{R_g\}^{c(g)} \cdot (P_{c(g)} + ... + P_T) + \{R_g\}^{c(g)+1} \cdot (P_{c(g)+1} + ... + P_T) + ... + \{R_g\}^{c(\bar{g})} \cdot (P_{c(\bar{g})} + ... + P_T).$$

The observed outcome in every period is

$$Y_t = \{R_g\}^{c(g)} \cdot Y_t\left(c(g)\right) + \{R_g\}^{c(g)+1} \cdot Y_t\left(c(g)+1\right) + ...\{R_g\}^{c(\bar{g})} \cdot Y_t\left(c(\bar{g})\right) + \{R_g\}^{\infty} \cdot Y_t\left(\infty\right)$$

We state the parallel trends assumption assuming linearity of the conditional expectations and conditional on covariates and groups:

Assumption 2.4 (Conditional Parallel Trends, CPT)**.** For $t = 1, 2, ..., T$,

$$E\left[Y_t\left(\infty\right)|R_1, ..., R_G, \mathbf{X}_t\right] = \sum_{g=1}^{G} \beta_g R_g + \sum_{g=1}^{G} \left(R_g \cdot \mathbf{X}_t\right) \gamma_g + \mathbf{X}_t \pi_t + \eta_t. \qquad \square \tag{2.2}$$

Note that Assumption 2.4 also relaxes the analogous parallel trends assumption in Callaway and Sant'Anna (2021) and other specifications where conditioning on cohort identity is replaced by conditioning on group identity, g. Note also that, although covariates are fully interacted above, the approach does not prohibit some covariates to enter additively. In this regard, the FLEX approach also extends the imputation method of Borusyak et al. (2024) by allowing for covariates to enter the model flexibly, giving the user more control over how the covariates are modeled.

Technically, we need not condition on the entire history of the covariates, $\{\mathbf{X}_t, t = 1, ..., T\}$ in Assumption 2.4, and so the covariates need not satisfy a strict exogeneity assumption (see Wooldridge, 2010, Chapter 10). Nevertheless, if we think the treatment assignment influences the covariates in the future, Assumption 2.4 would generally fail. For a recent discussion of 'bad controls' in the DID setting, see Wooldridge (2024). Also, we require a sufficient number of observations per stratum in order to get precise estimates of the β_g and γ_g in Assumption 2.4.

Even if the covariates do not change over time, the terms $\mathbf{X}_t \pi_t$ and $\left(R_g \cdot \mathbf{X}_t\right) \gamma_g$ allow relaxation of the usual parallel trends assumption. Their inclusion plays the same role as in Callaway and Sant'Anna (2021), who apply standard treatment effects estimators when covariates are available; see also Wooldridge (2025, 2023). The presence of $\left(R_g \cdot \mathbf{X}_t\right) \gamma_g$ al-

lows for substantial heterogeneity in how the average potential outcome changes with the groups (and therefore with the treatment groups). Also note that, in the case of time-varying covariates, the FLEX pooled-data regression approach allows covariates to enter only additively, as is typically specified in the canonical constant effect regression specification, fully interacted, as is needed to demonstrate equivalence to the imputation method of Borusyak et al. (2024), or any subset of the fully interacted specification, giving more control over how the covariates are modeled.

Under Assumptions 2.3 and 2.4, the parameters in 2.4 are identified using the untreated observations. In the subpopulation of untreated units at time t, we can write

$$E\left(Y_t|R_1,...,R_G,\mathbf{X}_t,W_t=0\right)=\sum_{g=1}^{G}\beta_g R_g+\sum_{g=1}^{G}\left(R_g\cdot\mathbf{X}_t\right)\gamma_g+\mathbf{X}_t\pi_t+\eta_t. \tag{2.3}$$

Equation (2.3) shows that all of the parameters are identified using the untreated observations, provided we have some units in every group.

The identification argument is easier to see in the simple 2×2 case, i.e., let $T=2$ and $G=2$. Let W denote the (only) treatment indicator and, to frame it like a typical 2×2 DID specification, let α denote the intercept for the value of the outcome for $G=1$ and $t=1$. We also remove group and time subscripts from all coefficients for simplicity. Then, equation (2.3) can be written as

$$E\left[Y_1\left(\infty\right)|W,\mathbf{X}_1\right]=\alpha+\beta W+\left(W\cdot\mathbf{X}_1\right)\gamma+\mathbf{X}_1\zeta \tag{2.4}$$

$$E\left[Y_2\left(\infty\right)|W,\mathbf{X}_2\right]=\alpha+\beta W+\left(W\cdot\mathbf{X}_2\right)\gamma+\eta_2+\mathbf{X}_2\pi+\mathbf{X}_2\zeta \tag{2.5}$$

Under CNA, the expectation in equation (2.4) is the same when we replace $Y_1\left(\infty\right)$ with $Y_1\left(2\right)$. Because $W=0$ implies $Y_1=Y_1\left(\infty\right)$ and $W=1$ implies $Y_1=Y_1\left(2\right)$,

$$E\left(Y_1|W,\mathbf{X}_1\right)=\alpha+\beta W+\left(W\cdot\mathbf{X}_1\right)\gamma+\mathbf{X}_1\zeta,$$

which shows that, provided there are some treated and control units and $\mathbf{X}_1$ does not have perfectly collinear elements, the parameters α, β, γ, and ζ are identified using the control and (eventually) treated units in $t = 1$. By equation (2.5),

$$E\left(Y_2|W=0,\mathbf{X}_2\right)=\alpha+\eta_2+\mathbf{X}_2\pi+\mathbf{X}_2\zeta=(\alpha+\eta_2)+\mathbf{X}_2(\pi+\zeta)$$

Again, ruling out perfect collinearity in $\mathbf{X}_2$, $(\alpha+\eta_2)$ and $(\pi+\zeta)$ are identified by the second period control units. Because α and ζ are identified, so are η_2 and π. Because we are conditioning on the control units in the second time period, which corresponds to G = 1, we do require variation in the covariates within group. Returning to equation (2.5), we have

$$E\left[Y_2\left(\infty\right)|W=1,\mathbf{X}_2\right]=(\alpha+\beta+\eta_2)+\mathbf{X}_2\left(\gamma+\pi+\zeta\right)$$

and so

$$E\left[Y_2\left(\infty\right)|W=1\right]=(\alpha+\beta+\eta_2)+E\left(\mathbf{X}_2|W=1\right)\left(\gamma+\pi+\zeta\right)$$

Identification of $E\left(\mathbf{X}_2|W=1\right)$ follows immediately because we observe the second-period covariates for the all units, including the treated units. Because the other parameters are identified by assumptions 2.3 and 2.4, $E\left[Y_2\left(\infty\right)|W=1\right]$ is identified. Then

$$\tau_2=E\left[Y_2\left(2\right)-Y_2\left(\infty\right)|W=1\right]=E\left(Y_2|W=1\right)-E\left[Y_2\left(\infty\right)|W=1\right]$$

is identified. The argument in the general staggered case is similar, with $E\left[Y_t\left(g\right)|W_g=1\right]=E\left(Y_t|W_g=1\right)$ always identified and $E\left[Y_t\left(\infty\right)|W_g=1\right]$ identified under assumptions 2.3 and 2.4.

2.2 Estimation by Imputation and Ordinary Least Squares

The identification argument in subsection 2.1 immediately suggests an imputation approach to estimation. We assume the availability of random samples from the population at each

time period t. The draws, indicated by i, are independent, but not generally identically distributed because the population distribution of both the outcome and control variables may change across t. (This same kind of heterogeneity is allowed in panel data settings. Even if the population is stable across time, we allow for changing distributions across t.) For unit i, $t(i)$ represents the time period. The observed data for unit i is $Y_{i,t(i)}$, $\mathbf{X}_{i(t)}$, and the group indicators, $R_{i,t(i),g}$, $g = 1, ..., G$. As discussed earlier, the treatment assignment is determined by the group. Typically, the group for a unit does not change over time—an individual lives in the same state, say, over the time periods in question—but even if that is true, the repeated cross-sections setting means that the draws of group indicators over time are from different samples of units.

Estimation of the parameters in equation (2.3) can proceed by estimating an OLS regression of the outcome on the untreated observations in the repeated cross-section dataset. Then, one mimics iterated expectations in the sample by using an out-of-sample prediction for $Y_t(\infty)$ (for the treated observations). Equivalently, the out-of-sample residuals are averaged over the appropriate group-time period pair to produce group-time specific ATET estimates. The following procedure extends Wooldridge (2025) to allow repeated cross-sections and time-varying covariates. It is also related to the imputation method in Borusyak et al. (2024), who mention applying imputation in the repeated cross-sections case. Here, we have derived an imputation estimator for repeated cross-sections under clearly stated assumptions that allow random control variables to appear in a flexible way.

Procedure 2.1. [Imputation Estimation]

1. Use the control observations to estimate the parameters

$$(\beta_1, ..., \beta_G, \eta_1, ..., \eta_T, \gamma_2, ..., \gamma_G, \pi_2, ..., \pi_T) \tag{2.6}$$

by OLS:

$$\begin{aligned} &Y_{i,t(i)} \text{ on } R_{i,t(i),1}, ..., R_{i,t(i),G},\ P_{2,t(i)}, ..., P_{T,t(i)},\ \mathbf{X}_{i,t(i)} \\ &R_{i,t(i),1} \cdot \mathbf{X}_{i,t(i)}, ..., R_{i,t(i),G} \cdot \mathbf{X}_{i,t(i)},\ P_{2,t(i)} \cdot \mathbf{X}_{i,t(i)}, ..., P_{T,t(i)} \cdot \mathbf{X}_{i,t(i)} \end{aligned} \tag{2.7}$$

2. For unit i, impute $Y_{i,t(i)}(\infty)$ as

$$\begin{aligned} \hat{Y}_{i,t(i)}(\infty) = &\sum_{g=1}^{G} \hat{\beta}_g R_{i,t(i),g} + \sum_{g=1}^{G} \left(R_{i,t(i),g} \cdot \mathbf{X}_{i,t(i)}\right) \hat{\gamma}_g \\ &+ \sum_{t=2}^{T} \hat{\eta}_t P_{t(i)} + \sum_{t=2}^{T} \left(P_{t(i)} \cdot \mathbf{X}_{i,t(i)}\right) \hat{\pi}_t \end{aligned} \tag{2.8}$$

3. For treatment group g, in period t, obtain

$$\begin{aligned} \hat{\tau}_{gt} &= N_{gt}^{-1} \sum_{i=1}^{N} R_{i,t(i),g} \cdot 1\left[t(i) = t\right] \cdot \left[Y_{i,t(i)}(g) - \hat{Y}_{i,t(i)}(\infty)\right] \\ &\equiv N_{gt}^{-1} \sum_{i=1}^{N} R_{i,t(i),g} \cdot 1\left[t(i) = t\right] \cdot \widehat{TE}_{i,t(i)} \\ &= \bar{Y}_{gt} - N_{gt}^{-1} \sum_{i=1}^{N} R_{i,t(i),g} \cdot 1\left[t(i) = t\right] \cdot \hat{Y}_{i,t(i)}(\infty) \end{aligned} \tag{2.9}$$

where

$$N_{gt}^{-1} = \sum_{i=1}^{N} R_{i,t(i),g} \cdot 1\left[t(i) = t\right]$$

is the number of units in treatment group g in time period t, $\widehat{TE}_{i,t(i)} = Y_{i,t(i)} - \hat{Y}_{i,t(i)}(\infty)$ is the unit-specific estimated treatment effect, and

$$\bar{Y}_{gt} = N_{gt}^{-1} \sum_{i=1}^{N} R_{i,t(i),g} \cdot 1\left[t(i) = t\right] \cdot Y_{i,t(i)} \tag{2.10}$$

is the average of the observed outcomes for units in treatment group g in period t. □

With a sufficient number of observations in each (g,t) cell, $\hat{\tau}_{gt}$, will have good statistical properties by the law of large numbers and central limit theorem. Nevertheless, the multi-step nature of the estimation makes inference challenging. A similar issue arises in the panel data setting in Borusyak et al. (2024), where unit-specific fixed effects are included in the first imputation step. Fortunately, there is an algebraic equivalence between imputation and a longer regression that uses all of the data. To describe the longer regression, let $\bar{\mathbf{X}}_{gt}$ be the average value of the covariates for treatment group g and time period t. Specifically, to the list of regressors in equation (2.7) we add treatment indicators and treatment indicators interacted with demeaned covariates:

$$R_{i,t(i),g} \cdot P_{i,t(i)},\ R_{i,t(i),g} \cdot P_{i,t(i)} \cdot \left(\mathbf{X}_{i,t(i)} - \bar{\mathbf{X}}_{gt(i)}\right),\ t = c(g), ..., T;\ g = 1, 2, ..., \bar{g} \tag{2.11}$$

If $R_{i,t(i),c} \cdot P_{i,t(i)} = 1$ then unit i is receiving treatment in time period t. The interactions $R_{i,t(i),c} \cdot P_{i,t(i)} \cdot \left(\mathbf{X}_{i,t(i)} - \bar{\mathbf{X}}_{gt(i)}\right)$ allow for heterogeneity in the treatment effects as a function of the observed covariates.

Proposition 2.1. [Equivalence of Imputation and OLS regression]

Using all of the data, consider the regression that includes all regressors in equations 2.7 and 2.11:

$Y_{i,t(i)}$ on $\mathbf{X}_{i,t(i)}$,

$$R_{i,t(i),1},\ ...,\ R_{i,t(i),G},\ R_{i,t(i),1} \cdot \mathbf{X}_{i,t(i)},\ ...,\ R_{i,t(i),G} \cdot \mathbf{X}_{i,t(i)},$$

$$P_{2,t(i)},\ ..., P_{T,t(i)},\ P_{2,t(i)} \cdot \mathbf{X}_{i,t(i)},\ ...,\ P_{T,t(i)} \cdot \mathbf{X}_{i,t(i)}$$

$$R_{i,t(i),c(1)} \cdot P_{c(1),t(i)},\ ...,\ R_{i,t(i),c(1)} \cdot P_{\bar{g},t(i)},\ ..., R_{i,t(i),c(\bar{g})} \cdot P_{c(\bar{g}),t(i)},$$

$$R_{i,t(i),c(1)} \cdot P_{c(1),t(i)} \cdot \left(\mathbf{X}_{i,t(i)} - \bar{\mathbf{X}}_{gt(i)}\right),\ ...,\ R_{i,t(i),c(\bar{g})} \cdot P_{c(\bar{g}),t(i)} \cdot \left(\mathbf{X}_{i,t(i)} - \bar{\mathbf{X}}_{gt(i)}\right)$$

Let the coefficients be $\left\{\tilde{\beta}_g : j = 1, ..., G\right\}$, $\{\tilde{\eta}_g : j = 1, ..., G\}$, $\{\tilde{\gamma}_t : t = 2, ..., T\}$, $\{\tilde{\pi}_r : t = 2, ..., T\}$, $\{\tilde{\tau}_{gt} : t = c(g), ..., T; g = 1, ..., \bar{g}\}$, and $\left\{\tilde{\delta}_{gt} : t = c(g), ..., T; g = 1, ..., \bar{g}\right\}$. Then

(i) For all g and t, $\tilde{\beta}_g = \hat{\beta}_g$, $\tilde{\eta}_g = \hat{\eta}_g$, $\tilde{\gamma}_t = \hat{\gamma}_t$, and $\tilde{\pi}_t = \hat{\pi}_t$.

(ii) For all $g \in \{1, 2, ..., \bar{g}\}$ and $t \in \{c(g), ..., T\}$,

$$\tilde{\tau}_{gt} = \hat{\tau}_{gt} \qquad \square$$

Appendix A provides a proof of this proposition. The equivalences in Proposition 2.1 are practically useful. Standard errors for the treatment effects $\tilde{\tau}_{gt} = \hat{\tau}_{gt}$ from the regression in equation 2.1 are readily available, and issues of clustering can be resolved in a standard pooled OLS setting. In addition to providing the $\hat{\tau}_{gt}$ and their standard errors, the $\tilde{\delta}_{gt}$ can be studied to determine if there are heterogeneous treatment effects.

The equivalence in Proposition 2.1 holds even if, in some time periods, we sample no units from a particular group. Naturally, we will not be able to estimate an ATE for that group/time pair. The OLS regression will simply drop all terms depending on the group-time interaction and estimate the remaining effects; nothing special is needed. As a special case of the regression in equation (2.1), one might have a single binary indicator, X_{it} (probably not time-varying), separating units into one of two groups. The scalar coefficients $\tilde{\delta}_{gt}$ would be the difference-in-difference-in-differences (DIDID) estimator for the group represented by $X_{it} = 1$. Again, inference is straightforward.

3 Regression Specifications

As the literature on methods for estimation of models for data with staggered entry into treatment has grown rapidly, the nomenclature used to describe various techniques has also

proliferated. In naming methods, we think it is a) important to distinguish between the estimands and the estimation methods, and b) to name the various estimators recently developed for estimating models in ways that indicate functional distinctions. In terms of the characteristics of the treatment effects parameters, some estimators assume homogeneous effects across groups and time. Others assume heterogeneity in one or the other dimension but not both. Yet others allow for heterogeneous effects across both cohorts and time but not across groups and time.

As for the parallel trends assumption, some estimators impose the parallel trends assumption for all periods prior to treatment, i.e., they assume that the regression specifications have only lagged treatment parameters. Other estimators impose the parallel trends assumption only for one baseline period prior to treatment (typically the period just preceding treatment) and the regression specifications include lag and lead treatment parameters with the lag parameters being the coefficients of interest for the ATET. Such models are often referred to as *event-study* models but that terminology corrupts definitions of event-study models from other areas of the econometrics literature.

We clarify our own use of language and notation by formally specifying the regression specifications we estimate using OLS to implement equation (2.1). In equation (3.1) below, we present a FLEX specification that explicitly displays all the regression parameters (and associated variables). The coefficients in the line denoted *lags* are the treatment effects at the group-time level in post-treatment periods. The coefficients in the line denoted *leads* are the pre-treatment differences between treated groups and the never-treated groups, except in one "baseline" period (chosen, without loss of generality, to be the period prior to treatment initiation, $c(g) - 1$). When the parameters in *leads* are estimated, we refer to

these as lags-and-leads specifications. This specification is commonly referred to as an *event study* in the recent literature, (e.g. Roth, 2024), but we prefer the term lags-and-leads to disambiguate from an older, distinct use of the term "event study" in the econometrics literature (e.g. MacKinlay, 1997). In lags-only specifications, all *lead* coefficients are set equal to zero. In other words, lags-only specifications assume that all pre-treatment differences between treated cohorts and the never-treated cohort are identically equal to zero. These specifications include group and year fixed effects. These specifications also allow covariates to enter additively and interacted with each lag, lead, and fixed effect indicators.

$$
\begin{aligned}
&E\left(Y_{i,t(i)}|\{R_{ig}\},\{P_{i,t(i)}\},\boldsymbol{X}_{i,t(i)}\right) = \\
&\sum_{g=1}^{\bar{g}}\sum_{t=c(g)}^{T}\tau_{gt}R_{ig}P_{i,t(i)} + \sum_{g=1}^{\bar{g}}\sum_{t=c(g)}^{T}R_{ig}P_{i,t(i)}\cdot(\boldsymbol{X}_{i,t(i)}-\bar{\boldsymbol{X}}_g)\boldsymbol{\kappa}_{gt} \\
&+\sum_{g=1}^{\bar{g}}\sum_{t=1}^{c(g)-2}\tau_{gt}R_{ig}P_{i,t(i)} + \sum_{g=1}^{\bar{g}}\sum_{t=1}^{c(g)-2}R_{ig}P_{i,t(i)}\cdot(\boldsymbol{X}_{i,t(i)}-\bar{\boldsymbol{X}}_g)\boldsymbol{\kappa}_{gt} \\
&+\sum_{g=1}^{G}\beta_g R_{ig} + \sum_{t=2}^{T}\eta_t P_{i,t(i)} + \sum_{g=1}^{G}R_{ig}\cdot\boldsymbol{X}_{i,t(i)}\boldsymbol{\gamma}_g + \sum_{t=1}^{T}P_{i,t(i)}\cdot\boldsymbol{X}_{i,t(i)}\boldsymbol{\pi}_t + \boldsymbol{X}_{i,t(i)}\boldsymbol{\zeta}
\end{aligned}
\tag{3.1}
$$

where $i = 1, ..., N$, $g = 1, ..., G$, $t = 1, ..., T$, and $c(g)$ denotes the first period in which treatment is received by observations in group g. The first row in equation (3.1) shows the lags terms that model treatment in the treated periods, the second row shows the leads terms that model "treatment" prior to the treated periods and the third row shows how all other regressors are specified.

When covariates are interacted with treatment indicators, they are specified as deviations from group means. Deviating the covariates from means when they are interacted only with the group and only with the time indicators is not necessary, but doing say generally

makes the coefficients on R_{ig} and $P_{i,t(i)}$ identical to ATETs. Note also that the specification in equation (3.1) shows that all covariates are interacted, i.e., it is a "fully interacted" specification. If desired, one may put restrictions on the coefficients so that some covariates appear only additively, which means the covariates that affect selection into treatment or determine heterogeneous trends can differ from those that provide heterogeneous treatment effects as a function of observables.

Formally, the group, g, is the level of aggregation of individual units of observation at which the treatment is applied. But, FLEX can also be specified using aggregates of the groups as g as long as each aggregate contains units from one cohort only. At the highest level of aggregation, a group may simply be a cohort. This feature allows researchers to consider tradeoffs between generality and the risks of parameter proliferation inherent in applications in which the number of groups is much bigger than the number of cohorts.

Generally, the lags-and-leads specification can be expected to be less efficient than the lags-only specification because the latter uses all implications of the parallel trends assumption – although it need not use this information in the most efficient way. One case where we can be sure of the superiority of lags-only is if we maintain independent sampling and conditional homoskedasticity (in addition to the assumptions used for identification). Then, lags-only is the best linear unbiased estimator because it is an OLS estimator using a correctly specified conditional mean under the Gauss-Markov assumptions, and lags-and-leads adds redundant regressors. With heteroskedasticity or clustered treatment assignment, we cannot state an efficiency theorem for lags-only, but neither estimator exploits heteroskedasticity or cluster correlation in estimation, meaning lags-only is likely to be more efficient.

In the panel data case, Harmon (2024) shows that using subgroup difference-in-differences

estimators are be more efficient than the **??** imputation estimator when the underlying idiosyncratic errors follow a random walk. We can use the algebraic equivalence results in Wooldridge (2025) to conclude that the lags-only estimator is efficient when idiosyncratic errors are serially uncorrelated and lags-and-leads is efficient when the errors follow a random walk. For other cases, it is not possible to rank the efficiency of lags-and-leads versus lags-only. The results in Harmon (2024) are not relevant for FLEX applied to repeated cross-sections because we cannot exploit serial correlation across time within a unit.

3.1 Aggregating the Treatment Effects

Rather than report a full set of treatment effect estimates for each treatment group and year, it is common to aggregate the effects to the cohort level, to the time level, to the exposure-time level (for staggered start), or most commonly to the aggregate level. One could do this by simple averaging. For example, the immediate effects, $\hat{\tau}_{gt}$, can be averaged over time periods $t = c(g)$, ..., T. The one-period dynamic effects, $\hat{\tau}_{t,t+1}$, can be averaged over $t = c(g)$, ..., $T-1$; and so on for each of the exposure lengths. To be precise, the aggregated average treatment effect on the treated (ATET) corresponding to equation (3.1) is given by:

$$\hat{\tau} = \frac{1}{N}\sum_{g=1}^{\bar{g}}\sum_{t=c(g)}^{T}\sum_{i=1}^{n_{gt}}\left[\hat{\tau}_{gt}\left(R_{ig}P_{i,t(i)}\right) + \left(R_{ig}P_{i,t(i)}\cdot\left(\boldsymbol{X}_{i,t(i)} - \bar{\boldsymbol{X}}_{g}\right)\right)\hat{\boldsymbol{\kappa}}_{gt}\right] \tag{3.2}$$

where $i = 1, ..., N$, $g = 1, ..., G$, $t = 1, ..., T$, and $c(g)$ denotes the first period in which treatment is received by observations in group g. Note that n_{gt} is the number of observations

in the g^{th} treated group in the t^{th} time period and $N = \sum_{g=1}^{\bar{g}} \sum_{t=c(g)}^{T} n_{gt}$. This simplifies to

$$\hat{\tau} = \frac{1}{N} \sum_{g=1}^{\bar{g}} \sum_{t=c(g)}^{T} [n_{gt} \hat{\tau}_{gt}] \tag{3.3}$$

because $\boldsymbol{X}_{i,t(i)} - \bar{\boldsymbol{X}}_g$) is de-meaned so that part of the equation sums to zero.

Note that $\hat{\tau}$ is simply the average of the treatment effects over all treated observations in all treated time periods. In specifications with no covariates, or covariates entered only additively, or if the covariates are de-meaned (using the means of covariates for the treated sample), then $\hat{\tau}$ is also a weighted average of the group-time effects $\hat{\tau}_{gt}$, where each weight is the sample size associated with each group and year, n_{gt}. Standard errors for the unweighted or weighted averages can be obtained via the delta method – as described, for example, in Wooldridge (2010).

3.2 Other Approaches

Other approaches have become popular in empirical research. For summaries, see de Chaisemartin and D'Haultfœuille (2023), Roth et al. (2023), Freedman et al. (2023), and Baker et al. (2025). We contrast FLEX with two such approaches, Callaway and Sant'Anna (2021) and Cengiz et al. (2019), noting that Borusyak et al. (2024) is equivalent to FLEX in fundamental ways.

As mentioned previously, the Callaway and Sant'Anna (2021) estimators fall into the lags-and-leads category. Without covariates and using the never treated units as controls, Callaway and Sant'Anna (2021) reduces to the repeated cross-sections version of Sun and Abraham (2021) (who provide results only for panel data, and only with covariates appearing additively). **??** discusses this equivalence in the panel data case. When using regression adjustment with covariates and the never treated units as controls, Callaway and Sant'Anna

(2021) produces ATETs identical to our fully flexible lags-and-leads estimator. Our pooled OLS approach has the advantage of delivering estimates of selection effects, heterogeneous trends, and moderating effects. Technically, identifying moderating effects from the FLEX regression adds the assumption that the treatment effects are linear in the covariates, X (but with heterogeneous coefficients across group and time). If X is a set of mutually exclusive and exhaustive dummies then the linearity assumption is for free. Generally, assuming linearity of the treatment effects in X seems like a modest assumption given that we have already assumed linearity of the conditional mean in the untreated state.

The doubly robust estimators proposed by Callaway and Sant'Anna (2021) have some resilience to the linear functional form of the conditional mean, but it is more restrictive than our FLEX approach along other dimensions, including the level at which to include fixed effects. In some cases, we may want to define fixed effects at, say, the county level even if the treatment effects vary only at, say, the cohort-by-state level. This flexibility allows us to combine attractive features from Callaway and Sant'Anna (2021) and Borusyak et al. (2024) and still obtain valid inference using standard software.

Another advantage of our approach is that we may want to impose restrictions on the treatment effects and the heterogeneous trends, or both. For example, in some cases we may not have enough data per cohort or group to provide reliable inference in the fully flexible regression, especially when there might be few treated units per cohort. In such cases we might impose constant effects by exposure time – as in a traditional event study analysis. In addition, we may not have enough observations to estimate fully heterogeneous effects with respect to the covariates, and these are easily dropped from the regression. Imposing restrictions on heterogeneity is essentially trivial in our pooled regression setting, but not

in the Callaway and Sant'Anna (2021) approach or the Borusyak et al. (2024) imputation approach. And, remember, we encompass both lags-only and lags-and-leads approaches in one setting.

A novel aspect of Callaway and Sant'Anna (2021) is that, with covariates, they allow estimation methods other than linear regression adjustment. One estimator with nice robustness properties combines propensity score weighting with linear regression adjustment (a version of IPWRA). Without the propensity score and without covariates, the Callaway and Sant'Anna (2021) approach is the same as the lags-and-leads specification of FLEX without covariates. When covariates are introduced, the Callaway and Sant'Anna (2021) approach with regression-adjustment is equivalent to a fully-saturated FLEX specification. Callaway and Sant'Anna (2021) can also estimate a doubly-robust estimator with propensity score weights, which is an extension that FLEX does not have. Note that the IPWRA estimator can have some resiliency to the assumed linear functional form.

Cengiz et al. (2019) propose a procedure that creates for each cohort a sample (stack) of treated cohort observations pooled with never-treated observations, then combines those samples, which resolves the issue of unwanted comparisons. Separate cohort, control (never treated), and year fixed effects are specified for each cohort sample. Exposure-time level treatment effects are then estimated using a linear regression commonly referred to as a *stacked regression*. Since the never treated group is included in each stack, the generated sample can be much larger than the original sample if the number of units in the never treated group is large. This contrasts with FLEX, Callaway and Sant'Anna (2021) and Borusyak et al. (2024) that use the original sample to obtain estimates of the treatment effects that do not include already treated units in the control group. The overall ATETs

are variance-weighted aggregates of the exposure-time level estimates (Dube et al., 2025).

3.3 Heterogeneous Time Trends

Like the estimates obtained from procedure 3.1, the event-study estimates require a kind of parallel trends assumption for consistency (Roth, 2022; Dette and Schumann, 2024). One approach to accounting for violations of parallel trends that occur even after including covariates is to assume relatively simple heterogeneous trends in the absence of the intervention. In particular, with at least two pre-intervention periods per treated group, one can include in equation (3.1) interactions between the group indicators, $R_{i,t(i),j}$ and a linear time trend, t. The coefficients on the trend terms can be used to test for pre-trends (Dette and Schumann, 2024). As shown in Wooldridge (2025) in the panel data case, such tests do not suffer from 'contamination bias', provided the covariates are included flexibly, as in equation (3.1). That is because the imputation and OLS approaches continue to be identical.

Including group-specific trends can be costly in terms of precision because their inclusion creates collinearity with the treatment indicators. Of course, using a pre-test to decide whether to drop these terms can be problematic—just as when using the event-study approach.

3.4 Practical Considerations

In this section we discuss several practical considerations when estimating difference-in-differences models for data with repeated cross-sections, heterogeneous treatment effects, and staggered timing. Estimating treatment effects in a difference-in-differences model is usually a two-step process. The first step is to estimate the treatment parameters of a model and the second step is to aggregate some of those treatment parameters into estimated average

treatment effects. For example, estimating the average treatment effect on the treated, or treatment effects by group or by time since the start of treatment.

The set of questions for the first step revolves around identification of treatment effects and model specification. How are the treatment effects identified? What level of heterogeneity is allowed, specifically, heterogeneity at the group or the cohort level? Should the model assume that parallel trends holds in the pre period or allow for heterogeneous effects across groups in each time period before treatment starts? To what extent should the model specification control for covariates?

The set of questions for the second step includes answering the research question, presenting the results, and describing the extent of the heterogeneity in the treatment effects. After estimating the regression model, how should the estimated treatment coefficients be combined to form an overall ATET? Should the treatment effects be aggregated to other levels, for example, to show an event-study graph?

3.4.1 Modeling Decisions

There are three main modeling decisions, which can be seen as whether to allow more or less flexibility in the estimation step. They can also be seen as variations on equation (3.1). The first decision is whether to allow heterogeneity at the group level or the cohort level, when there are multiple groups per cohort. In our empirical examples, there are multiple states (groups) that start the policy treatment in some years and so are in the same cohort. In general we think that it is restrictive to impose homogeneity across all states that happen to start their policy in the same year. Therefore, our model specifications are at the group level. However, if the number of groups is large (e.g., $3,000$ counties), then one needs to

think harder about the tradeoffs between model flexibility and possible over-fitting.

A second decision is whether to model lags-only or both lags-and-leads. One might think that the event-study approach, which allows for different effects for treatment observations in the pre-treatment periods that follow a trend that is not parallel to that for the control observations, is more flexible than the leads only model because it does not impose the conditional parallel trends assumptions in the pre-treatment periods. But if parallel trends is violated in the treated periods, then the OLS estimator applied to the lags-and-leads model can be worse than the OLS estimator applied to the lags-only model as it will have different biases for different violations of parallel trends.

The third is about how to include covariates. Covariates can be included separately as additional controls or interacted with treatment effects. Group fixed effects sweep away any time-invariant covariates. However, unlike covariates in a balanced panel, covariates in a repeated cross-section can change either because they are time-varying or because the random subsample has a different composition of covariates than the full sample. Therefore, one can control for covariates in a cross-section even when the covariates are inherently time-invariant. Endogenous covariates should not be included in the model. For example, controlling for experience when predicting income is not appropriate if the model is measuring the effect of a job training program on income and that job training program endogenously changes experience.

Interacting covariates with treatment effects allows for further treatment heterogeneity. This is a decision about whether to include the third and fourth rows of equation (3.1) or to assume that some of those coefficients are zero. The potential downside of this flexibility includes over-fitting and slower estimation speed. Although adding covariates when estimat-

ing group-level treatment effects can greatly increase estimation time, in some cases data compression can be used to increase the speed of estimation (Wong et al., 2021). Alternatively, one can lean on economic theory and knowledge of institutions to focus on just one or a few covariates to interact with the treatment effects instead of all covariates. In our experience, including interactions between treatment effects and covariates does not lead to larger standard errors.

3.4.2 Aggregation and Graphing

After estimating treatment effects at the group-time level, researchers have several options for how to present the results. One common approach is to aggregate to the time-since-treatment level to create an event study plot. This is useful when estimating a lags-and-leads model because one can plot the estimated difference between ever-treated groups and control groups in each pre period. This provides a visual test of the parallel trends and no anticipation assumptions. If one suspects or wants to test for heterogeneity by calendar time or by group, one could aggregate treatment effects to those levels and plot the results.

Researchers often want an overall ATET. This is easy to do in the FLEX approach. The ATET is the weighted average of treatment effects. In our `flexdid` Stata package, we allow two options for weights. The default option is that the weights equal the number of treated observations in each group-time set in the post periods. The alternative option is that the weights are proportional to the number of treated observations, and therefore vary by group but not over time.

3.4.3 Software

A Stata package `flexdid` that estimates the parameters of FLEX and allows users to conveniently produce a variety of post-estimation aggregate ATETs and tests of hypotheses about those ATETs is available from the Boston College Statistical Software Components (SSC) archive (Deb, 2025). To download the software and help files, type `ssc install flexdid` in Stata.

3.5 Transparency

One advantage of our FLEX approach is that it is transparent. The model specification is clear; there are no hidden estimated parameters. In equation (3.1), we write down our model as one regression such that the number of parameters, the specification of the treatment effects, and how covariates are specified in transparent.

The `flexdid` Stata package allows the user to report all estimated parameters, which makes it easy to see exactly how many treatment effects are estimated and how many additional covariates are part of the model specification, including all interactions. This transparency is also useful for comparing different possible FLEX model specifications and for comparing FLEX with other models developed in Callaway and Sant'Anna (2021), Sun and Abraham (2021), Cengiz et al. (2019) and de Chaisemartin and D'Haultfœuille (2023). This comparison, though, can be complicated as the transparency of these other estimators (e.g. the number of estimated parameters) is not always obvious.

3.5.1 Computational Time

Because FLEX uses a single least-squares regression, its computational time is fast. It does not require multiple regressions (e.g., the imputation method), numerous pairwise compar-

isons (e.g., Callaway and Sant'Anna (2021)), or replicating the data (e.g., stacked). The computational time does increase, in proportion to the number of parameters, in models with many covariates and interactions between covariates and treatment effects, but even then we have found the computational time to be reasonable.

4 Empirical Example

We show an empirical example using a data set that has features typical of repeated cross-section data with staggered difference-in-differences study design. We estimate the effect of the Secure Communities immigration enforcement program on Supplemental Nutrition Assistance Program (SNAP) enrollment. The idea is that when immigration enforcement increases in a community, some people will be more reluctant to sign up for a government program, even if they are U.S. citizens and eligible for the benefits (Alsan and Yang, 2024). The data are collected at the individual level. Treatment happens at the county level and the timing of the start of treatment is staggered over several years. We measure the outcome and covariates using 12 years of the American Community Survey, which is not a panel data set but instead a repeated cross-section. The example is typical of many difference-in-differences study designs and is appropriate to illustrate our theoretical results.

4.1 Prior Related Empirical Literature

There is a large and rapidly expanding empirical literature that uses difference-in-differences study design to assess the effects of policy changes. Here we briefly mention two papers that are most relevant to our empirical example (East et al., 2023; Alsan and Yang, 2024). Both papers study the effect of the Secure Communities immigration enforcement policy using

repeated cross-sectional data with staggered treatment timing.

East et al. (2023) measure the effects of the Secure Communities immigration enforcement on labor market outcomes. They exploited variation in the timing of exposure to Secure Communities over commuting zones to measure its effects on employment and hourly wages. They estimated a difference-in-differences regression with a homogeneous treatment effect, the traditional two-way fixed effects model, using data from the American Community Survey. East et al. (2023) limited their analysis to working-age populations who were most likely to be affected by Secure Communities—low-educated foreign born persons and U.S.-born individuals. They found that Secure Communities immigration enforcement decreased the employment of people who were likely to be undocumented immigrants, an expected finding. However, they also found that U.S.-born persons had slightly lower employment and hourly wages as a result of Secure Communities, an unexpected finding, perhaps due to a decline in labor demand.

Alsan and Yang (2024) explore a different policy question related to Secure Community immigration enforcement. They ask whether noncitizen deportations affect co-ethnic citizen participation in means-tested social insurance programs, even if those citizens are not personally at risk of deportation. Persons who decide not to enroll in those benefit programs if they are concerned that providing information to the government could result in deportation of relatives or persons in their network. This is one example of a chilling effect of deportation policies. Alsan and Yang examine at the effects of Secure Communities enforcement on two means-tested social insurance programs: Supplemental Security Income and the Supplemental Nutrition Assistance Program (SNAP). They use data from both the American Community Survey and the Panel Survey of Income Dynamics. Their treatment

is measured at the county-year level. In addition, instead of a difference-in-differences study design, they use a triple difference by comparing the effect of Secure Communities for Hispanics compared to black and white non-Hispanics. They allow for heterogeneous treatment effects by racial/ethnic group and by year since start of treatment, but not across counties. They found evidence for a chilling effect on Hispanic-headed households.

Our example borrows elements from both East et al. (2023) and Alsan and Yang (2024). We also use the American Community Survey to study the effect of Secure Communities immigration enforcement, which has a staggered implementation across US counties and data from a repeated cross-section. Like Alsan and Yang (2024) we examine the effect on SNAP enrollment and like East et al. (2023) we use a difference-in-differences study design.

4.2 Secure Communities Immigration Enforcement and ACS Data

We use repeated cross-section data from the American Community Survey (ACS) to show how to use heterogeneous effects difference-in-differences methods to estimate the effect of Secure Communities immigration enforcement on SNAP enrollment. The U.S. Census Bureau conducts the ACS monthly to collect information on individuals in communities, including information on employment, education, and enrollment in government benefit programs. The ACS publicly reports the location of respondents using Public Use Microdata Areas (PUMAs). PUMAs are non-overlapping, statistical geographic areas with at least 100,000 people. PUMA boundaries respect state boundaries but are often collections of counties. There are 2,487 PUMAs, somewhat fewer than the 3,144 counties in the US. When a PUMA consists of multiple counties, we assign the earliest data of implementation of the program in the constituent counties to the PUMA.

We use the data archive of Alsan and Yang (2022) to construct our analysis sample, including a crosswalk between counties (the geographic unit at which the treatment was assigned) and PUMAs (the geographic unit available in the ACS). Following Alsan and Yang (2024), we limit our sample to Hispanic citizens (including naturalized citizens) who did not move between states and who were either heads of households or their spouses. We also drop observations from Illinois, Massachusetts and New York because these states resisted the activation of Secure Communities. Unlike Alsan and Yang who restricted their sample to those with less than a high school education, we include those with a high school diploma (but not those with additional education beyond high school). We also exclude PUMAs that have fewer than 10 observations in any year 2005-2016. Few observations make it hard to estimate treatment effects at the group-time level, especially for models that include interactions with other regressors. For purposes of illustration, although not required for estimating FLEX, we further restrict our sample to individuals who live in 682 PUMAs observed in each sample year. Our sample includes observations for more than half a million people who live in 682 PUMAs across 32 states during the years 2005-2016. Therefore, this is an individual-level data set of repeated cross-sections of all 682 PUMAs for all 12 years.

Out of the 682 PUMAs in our sample, 259 began Secure Communities immigration enforcement during our time period while 423 did not and are the never-treated controls (see Table 1). During the years 2005-2008, no PUMAs in our data set started Secure Communities immigration enforcement. Beginning in 2009, at least 50 PUMAs in our data set started Secure Communities immigration enforcement each year through 2012. Figure 1 shows that the Secure Communities immigration enforcement began in the South and West Census regions. Treatment did not start in the Northeast region until 2012. See Table 1 for the

number of PUMAs that were treated in the years 2009-2012.

A cohort comprises all PUMAs that implemented the Secure Communities immigration enforcement program in a calendar year. Each of the four treated cohorts is observed for at least four periods before the start of treatment and for at least five years after the start of treatment. In addition, one cohort consists of never-treated PUMAs.

There are a total of 581,660 observations on individuals. The ACS has an average of 850 observations per PUMA per year. The majority of observations are in the West and the fewest are in the Midwest (see Table 2). Around 20 percent of the sample was enrolled in SNAP (see Table 3). Slightly more than half the sample is female, the average age is 50, the income to poverty ratio is around 240, and about 60 percent have either a high-school diploma or a GED.

4.3 Model Specifications

We estimate several alternative FLEX specifications based on the general model specification in equation (3.1). The lags-only models estimate treatment effects only in the periods after the start of the policies, i.e., these specifications use indicators for treatment lags-only. For the lags-only models we estimated a static, homogeneous effects (SHo) model to provide a benchmark, knowing that it is likely biased. Then we estimated several different FLEX regressions at either the cohort level or the group level, where the group is the unit of geography where enforcement happened (PUMA).

In lags-and-leads models, we allow all periods—except one reference baseline period, which we specify as the period preceding initiation of treatment—to have treatment indicators for each ever-treated group or cohort. There are cohort-by-time or group-by-time

coefficients in the pre-treatment periods as well as the post-treatment periods. In the lags-and-leads models, the treatment effects are compared to the year prior to the start of treatment; in the lags-only models, the treatment effects are compared to the average of all the years prior to the start of treatment. Note that one obtains the lags-only model by imposing the restriction that all the pre-treatment effects equal zero.

We also estimate treatment effects for the lags-and-leads models using two popular alternative techniques, each of which has a way to resolve the bias issues arising from staggered entry into treatment. One is the event study (labeled ES) model as described in Sun and Abraham (2021), that is homogeneous across groups. The other uses the method of Callaway and Sant'Anna (2021) which implements a flexible regression adjustment estimator (labeled CS).

There are additional choices about how to estimate the heterogeneous treatment effects. We can allow lots of heterogeneity or little. We can allow treatment effects at the cohort-year level or the group-year level. We can allow more heterogeneity by interacting other covariates with the cohort-year or group-year treatment effects. In our empirical analysis, other covariates include individual-level characteristics: female sex, age, poverty status of the family, and an indicator for whether the head of the family has a high school diploma. Because Secure Communities generally rolled out from the southern border northwards, there could be a strong regional geographic component to treatment effect heterogeneity. Therefore, in some model specifications we include region fixed effects as part of the set of other covariates interacted with the treatment effects.

4.4 Results

The outcome variable in our empirical analysis is whether the individual enrolled in SNAP. We report estimates of the overall ATETs from a number of lags-only specifications in Figure 2, where all standard errors are clustered at the PUMA level to reflect policy assignment at the PUMA level. Among specifications without covariates, the estimates of ATET are negative and statistically significant in the static, homogeneous specification (SHo) and in FLEX specifications with heterogeneity specified at the cohort and PUMA levels (the first three rows of Figure 2). The FLEX estimates are, however, substantially larger in relative terms. The estimates are that immigration enforcement lowered SNAP enrollment by about 1.2 percentage points. Compared to the sample enrollment rate of around 20 percent, these estimates imply a 6% decrease in the probability of enrolling in SNAP, among Hispanic citizens with a high school education or less. The comparable value for SHo is less than 1 percentage point.

When individual-level regressors (gender, age, income, and education) are added to the specification, the SHo estimate declines substantially (to -0.003) and is now not statistically significant at conventional levels. However, we know that this result may not represent any reasonable aggregation of cohort by time ATETs. The ATET estimate from the FLEX specification with groups defined as cohorts, and the covariates also interacted with the treatment, cohort, and year fixed effects, is twice as large and significant at the 10% level. When region indicators are added to the specification, the magnitude of ATET more than doubles again to about -0.019 (almost a 10% reduction) and is significant at all conventional levels. Given the substantial across-region heterogeneity in implementation of treatment, it's

not surprising that a specification that flexibly models that heterogeneity produces sizably different estimates. Finally, we add PUMA indicators additively to the specification (formally an extension to the specification in equation 3.1) which reduces the standard error of a very similar ATET estimate. Modeling the additional heterogeneity is substantively important.

We report estimates of the overall ATETs from several lags-and-leads specifications in Figure 3. ATET estimates from ES, CS and FLEX specifications with and without individual-level regressors are very similar. When region indicators are introduced, the ATET estimates increase substantially in magnitude, corresponding to a 10% change ((the last two rows of Figure 3).

The standard errors in the group-by-year specifications and in the cohort-by-year specifications with additive PUMA fixed effects are consistently smaller than those in the cohort-by-year specifications. These results can be seen in both lags-only and lags-and-leads specifications. It appears that the additional generality implied in the group-by-year heterogeneous specifications produces treatment estimates with greater precision because there is considerable within-cohort (by year) heterogeneity in outcomes. The CS estimates, which allow for general heterogeneity at the cohort-by-year level, are similar to those obtained using FLEX that also model heterogeneity at the cohort-by-year level. These estimates are also statistically significant but the standard errors are slightly larger than those obtained using FLEX.

We use our preferred FLEX specification with heterogeneous treatment coefficients at the cohort-by-year level with covariates (individual characteristics and region indicators), with cohort and year fixed effects and with PUMA indicators entered additively to the specification to calculate estimates of ATET at disaggregate levels of interest to researchers.

These are the model specifications at the bottom of Figures 2 and 3. We show the plots of ATETs by exposure year, from lags-only models in panel (a) and from lags-and-leads models in panel (b) of Figure 4. The latter is what is often referred to as the event study figure. An eyeball check shows that the parallel trends assumption is good for five pre-periods. The negative treatment effects by exposure year in the treatment periods are quite similar in magnitude across the lags-only specification and the lags-and-leads specification.

The ATET estimates in each calendar year in a selection of treated periods are shown in panels (c) and (d) of Figure 4. These are similar across the lags-only specification (c) and lags-and-leads specification (d). The ATET by cohorts are shown in panels (e) and (f). They show some heterogeneity by cohort, with the early cohorts generally having more negative effects and the later cohorts having either no or positive effects. This heterogeneity would be worth exploring further.

Our preferred specification allows for the cohort-year treatment effects to vary by region. Given the heterogeneity of treatment implementation across regions, it is of policy interest to visualize whether treatment effects were very different across regions. We display the event study plots from the lags-only specification for each census region in Figure 5, and from the lags-and-leads specification for each census region in Figure 6. What is interesting is that the only treatment effects that are significant are in the South region. The overall ATETs are greater than 0.03 in magnitude implying a 15% decline.

As stated in Section 3.5, one advantage of our FLEX approach is that it is transparent. To demonstrate this transparency for the Secure Communities empirical example, we created a table showing the number of parameters estimated across different model specifications and estimators, before any aggregation of the treatment effects (see Table 4). The numbers of

parameters are shown separately for the treatment effects, other regressors, and the total. The model specifications match those in Figures 2 and 3, with the lags-only models in the upper half (Panel A) and the lags-and-leads models in the lower half (Panel B) of Table 4.

There are several important patterns apparent in Table 4. Allowing for more treatment parameters is one way to increase the flexibility of the model specification. We include additional regressors as reported in the second column of Table 4. These include either 682 PUMA (groups) or four cohort fixed effects (2009-2012) and 11 year effects that are part of every specification in Table 4.

The static homogeneous treatment effects model (SHo) includes only a single treatment effect, by definition, as shown in the first row of Panel A. Note that there are 693 parameters from the group (682) and time fixed-effects (11).The first FLEX model specification has 26 cohort-year treatment effects, because the four cohorts have 8, 7, 6, and 5 years of treatment, and a total of 42 parameters. Because cohorts are collections of at least one group, there are at least as many treatment parameters in the group models as in the cohort models. The third row, with group-year treatment effects at the PUMA level, has 1,655 treatment effects and a total of 2,348 parameters.

The number of treatment effects for the lags-and-leads models, reported in Panel B of Table 4, is always higher than the corresponding number for lags-only models because of the additional effects estimated for the (not yet) treated groups in the periods before treatment starts. The Event Study model includes 14 dynamic treatment effects (the left-out treatment effect is the one just prior to the period when the treatment is applied). This allows one to plot the event study graph and test for parallel pre-trends and no anticipation effects. Now the FLEX model with group by year treatment effects and group fixed effects has 3,542

parameters (line 3 of Panel B).

Four individual-level regressors (X) are added in the last four specifications in both Panels A and B in Table 4: indicators for female sex, age, poverty status of the family and whether the head of the family has a high school diploma. In SHo and ES, these regressors are included additively only, leading to only four additional parameters (row 4 in Panel A and B). In the FLEX specifications with cohort as group, these four regressors are included fully interacted with the treatment effects and with the cohort and year fixed effects. Fully interacting the four covariates in the cohort-level FLEX model increases the number of parameters from 42 to 210 for the lags-only model and from 60 to 300 in the lags-and-leads model (line 5 in Panels A and B). We do not estimate models with covariates and PUMA as the group because that fully interacted specification has about 10,000 parameters.

Despite adding additional covariates, in our experience this additional treatment effect heterogeneity does not generally increase the standard errors of the estimated ATET. Also, the researcher has the option to only interact a small number of the other regressors with the treatment effects, based on what economic theory predicts or the evidence from prior related research supports.

5 Conclusions

Our paper makes several theoretical and practical contributions to the literature on difference-in-differences models with staggered treatments using repeated cross-section data. On the theoretical dimension, we prove that a linear regression with a sufficiently flexible functional form consisting of group-by-time treatment effects, two-way fixed effects, covariates and interaction terms yields consistent estimates of heterogeneous treatment effects. The estimates

are efficient under traditional assumptions and likely are comparable in efficiency to other methods with clustered assignment or heteroskedasticity. Aggregation of treatment effects and inference are straightforward. The results hold under standard assumptions of stable unit treatment value, no bad controls, conditional no-anticipation, and conditional parallel trends. We prove that FLEX with lags-only and appropriate interaction terms, estimated by ordinary least squares, returns numerically identical results as the imputation method by Borusyak et al. (2024).

The theoretical result about repeated cross-section data is of importance to many applied researchers, because data are often not balanced panel data. Our FLEX model extends other well-known DID models in other ways. FLEX with lags-and-leads extends Sun and Abraham (2021) and Callaway and Sant'Anna (2021) to repeated cross-sections and allows for other regressors to be included more flexibly. FLEX also allows heterogeneous treatment effects at the group level, instead of only at the cohort level. FLEX is also transparent. We write down our model as one regression such that the number of parameters in the model is easily determined. This is not as simple in other models (e.g., Cengiz et al., 2019; Callaway and Sant'Anna, 2021).

On the empirical side, we demonstrate our FLEX methods with one publicly available data set to answer a policy-relevant research question about the effect of Secure Communities immigration enforcement. The empirical example use individual-level cross-section data with staggered treatment at the PUMA level. Our FLEX method and the imputation method obtain the same numerical result. Our FLEX method generally has smaller standard errors than other popular estimators. In summary, FLEX has the advantage of being easy to implement, flexible, fast, and best among linear unbiased estimators.

References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

Alsan, M. and Yang, C. (2022). Replication data for: Fear and the Safety Net: Evidence from Secure Communities.

Alsan, M. and Yang, C. S. (2024). Fear and the Safety Net: Evidence from Secure Communities. *The Review of Economics and Statistics*, 106(6):1427–1441.

Baker, Andrew C. Callaway, B., Scott, C., Goodman-Bacon, A., and Sant'Anna, P. H. C. (2025). Difference-inDiferences Designs: A Practitioner's Guide. Technical report.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies*, 91(6):3253–3285.

Caetano, C., Callaway, B., Payne, S., and Rodrigues, H. S. (2024). Difference in Differences with Time-Varying Covariates. *Working paper*.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

de Chaisemartin, C. and D'Haultfœuille, X. (2023). Two-way fixed effects and difference-indifferences with heterogeneous treatment effects: a survey. *Econometrics Journal*, 26:C1–C30.

Deb, P. (2025). FLEXDID: Stata module for flexible estimation of difference-in-differences regression with staggered implementation. *Statistical Software Components*.

Dette, H. and Schumann, M. (2024). Testing for Equivalence of Pre-Trends in Difference-in-Differences Estimation. *Journal of Business & Economic Statistics*, 1–13.

Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2025). A Local Projections Approach to Difference-in-Differences. *Journal of Applied Econometrics*, 40(7):741–758.

East, C. N., Hines, A. L., Luck, P., Mansour, H., and Velásquez, A. (2023). The Labor Market Effects of Immigration Enforcement. *Journal of Labor Economics*, 41(4):957–996.

Freedman, S. M., Hollingsworth, A., Simon, K. I., Wing, C., and Yozwiak, M. (2023). Designing Difference in Difference Studies With Staggered Treatment Adoption: Key Concepts and Practical Guidelines. NBER Working Paper 31842. https://www.nber.org/papers/w31842.

Goodman-Bacon, A. (2021). Difference-in-differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Harmon, N. A. (2024). Difference-in-differences and efficient estimation of treatment effects. *University of Copenhagen Working Paper*.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, 35(1):13–39.

Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4(3):305–322.

Roth, J. (2024). Interpreting Event-Studies from Recent Difference-in-Differences Methods. *arXiv working paper*. https://doi.org/10.48550/arXiv.2401.12309.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2023). What's Trending in Difference-in-differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, 235(2):2218–2244.

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

Sun, L. and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2):175–199.

Wong, J., Forsell, E., Lewis, R., Mao, T., and Wardrop, M. (2021). You Only Compress Once: Optimal Data Compression for Estimating Linear Models. arXiv:2102.11297 [cs].

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Wooldridge, J. M. (2023). Simple Approaches to Nonlinear Difference-in-differences with Panel data. *The Econometrics Journal*, 26(3):C31–C66.

Wooldridge, J. M. (2025). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Empirical Economics*, 69(5):2545–2587.

Table 1: Public Use Microdata Area Participation in the Secure Communities Immigration Enforcement Program by Cohort and Census Region

	Northeast	Midwest	South	West	Total
Never treated	10	24	143	246	423
2009	0	0	49	20	69
2010	0	0	32	18	50
2011	0	9	20	24	53
2012	47	6	5	29	87
Total	57	39	249	337	682

Notes: The repeated cross-section data are from the American Community Survey (ACS) data for residents of 682 PUMAs in 32 states for 12 years from 2005–2016. A cohort comprises PUMAs that implemented the program in a particular year.

Table 2: Secure Communities Observations by Cohort and Region

	Northeast	Midwest	South	West	Total
Never treated	5680	8256	132281	235625	381842
2009	0	0	54124	14349	68473
2010	0	0	30950	15589	46539
2011	0	3336	15837	19629	38802
2012	27092	1429	1346	16137	46004
Total	32772	13021	234538	301329	581660

Notes: The repeated cross-section data are from the American Community Survey (ACS) data for residents of 682 PUMAs in 32 states for 12 years from 2005–2016. A cohort comprises PUMAs that implemented the program in a particular year.

Table 3: Sample Means by Secure Communities Status

	Treated	Never treated
Enrolled in SNAP	0.204	0.166
Secure Communities is activated	0.557	0.000
Female	0.544	0.550
Age	50.582	51.403
Income to Poverty ratio	235.350	247.876
High school diploma or GED	0.607	0.631
Observations	199818	381842

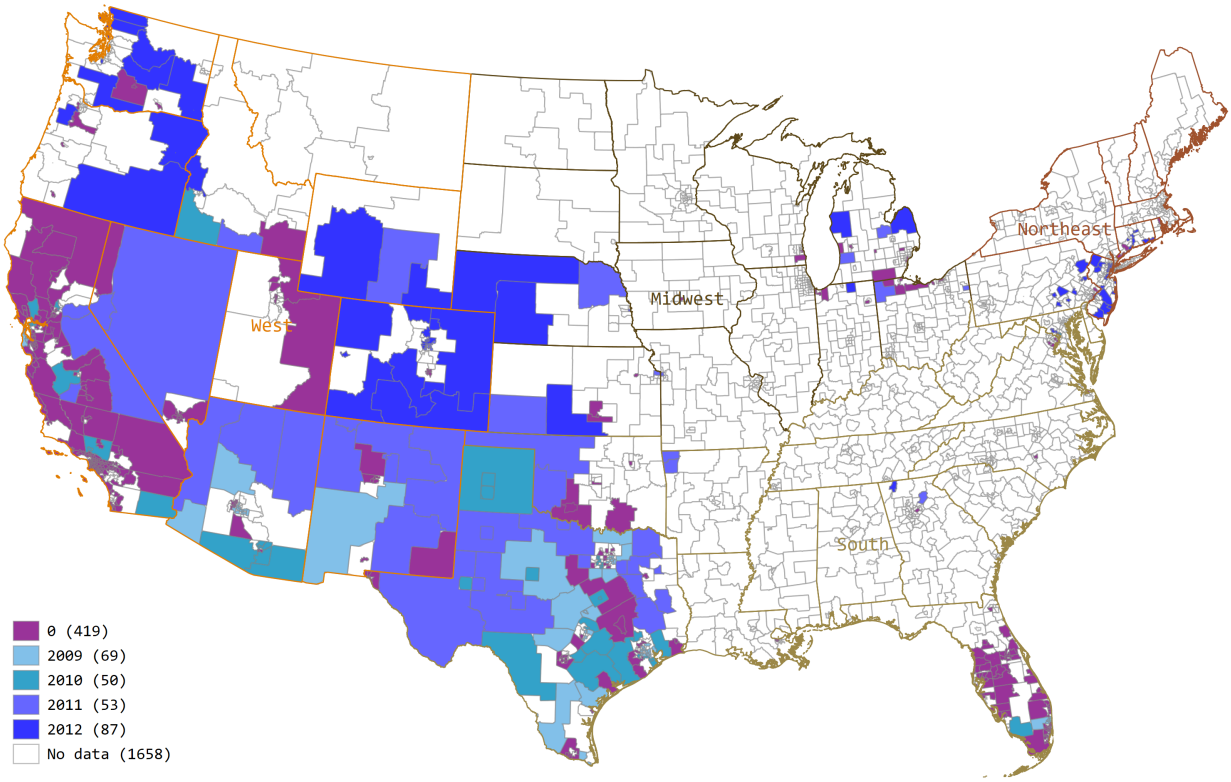
Notes: The repeated cross-section data are from the American Community Survey (ACS) data for residents of 682 PUMAs in 32 states for 12 years from 2005–2016.

Table 4: Numbers of Parameters Estimated in Alternative Specifications

Specification of			Number of Parameters		
Model	Treatment	Other regressors	Treatment	Other regressors	Total
Panel A: Lags only models					
SHo	constant	grp yr	1	693	694
FLEX	coh.yr	coh yr	26	16	42
FLEX	grp.yr	grp yr	1655	693	2348
SHo	constant	grp yr x	1	697	698
FLEX	coh.yr.x	coh.x yr.x	130	80	210
FLEX	coh.yr.(x+rg)	coh.(x+rg) yr.(x+rg)	234	61	295
FLEX	coh.yr.(x+rg)	coh.(x+rg) yr.(x+rg) grp	234	728	962
Panel B: Lags and leads models					
ES	event-time	grp yr	14	693	707
FLEX	coh.yr	coh yr	44	16	60
FLEX	grp.yr	grp yr	2849	693	3542
ES	event-time	grp yr x	14	697	711
FLEX	coh.yr.x	coh.x yr.x	220	80	300
FLEX	coh.yr.(x+rg)	coh.(x+rg) yr.(x+rg)	396	24	420
FLEX	coh.yr.(x+rg)	coh.(x+rg) yr.(x+rg) grp	396	691	1087

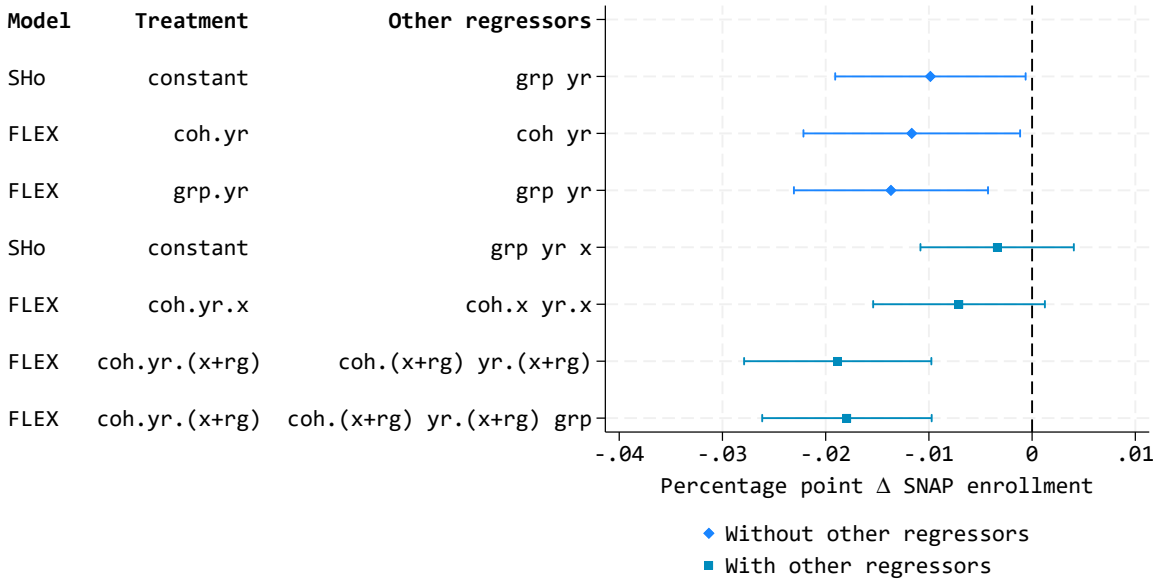
Notes: The repeated cross-section data are from the American Community Survey (ACS) data for residents of 682 PUMAs in 32 states for 12 years from 2005–2016. FLEX refers to the model developed in this paper. SHo refers to the static homogeneous effect two-way fixed effects estimator. ES refers to the heterogeneous over exposure-time but constant across groups two-way fixed effects estimator.
`grp` denotes PUMA (the group), `yr` denotes year, `coh` denotes cohort, `rg` denotes indicators for four regions of the country and `x` denotes the vector of individual-level regressors (an indicator for female sex, age, poverty status of the family and an indicator for whether the head of the family has a high school diploma). Dot (.) is used to denote interactions of indicator levels.

Figure 1: Map Showing the Year of Entry of PUMAs across the United States into the Secure Communities Immigration Enforcement Program



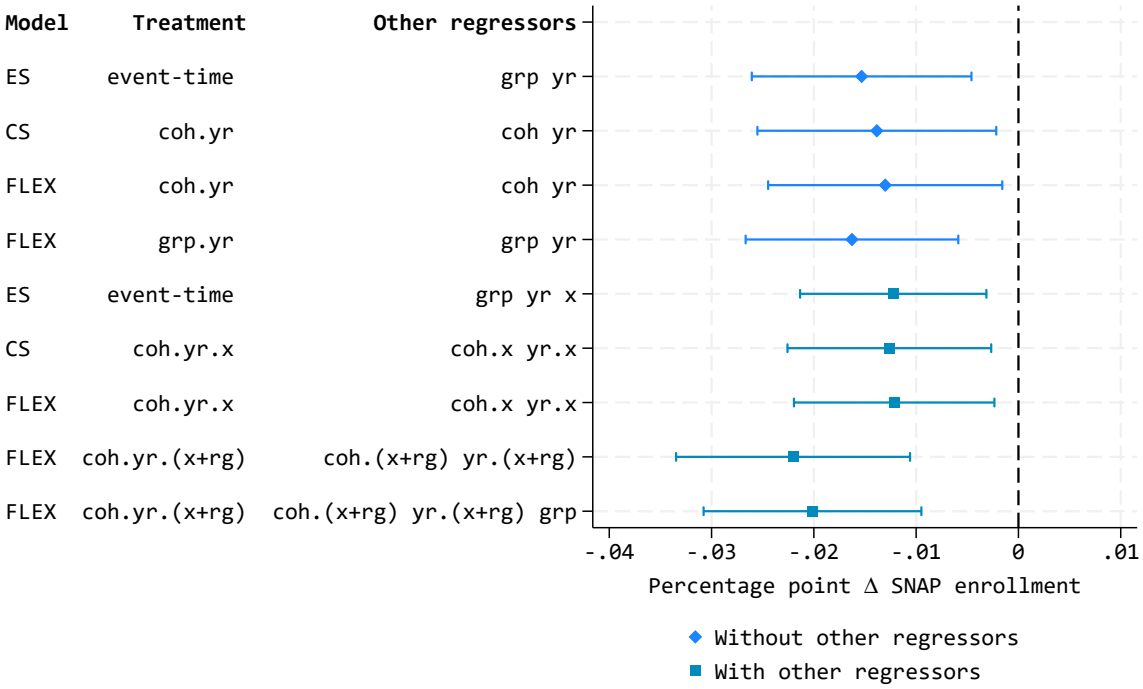
Notes: A number of Public Use Microdata Areas (PUMAs) have no data because they are not identified in the American Community Survey (ACS) for confidentiality reasons. We also exclude PUMAs that have fewer than 10 observations in any year 2005-2016. Our sample includes 682 PUMAs across 32 states during the years 2005-2016.

Figure 2: Overall ATET of the Secure Communities Immigration Enforcement Program on SNAP Enrollment in Lags-only Specifications



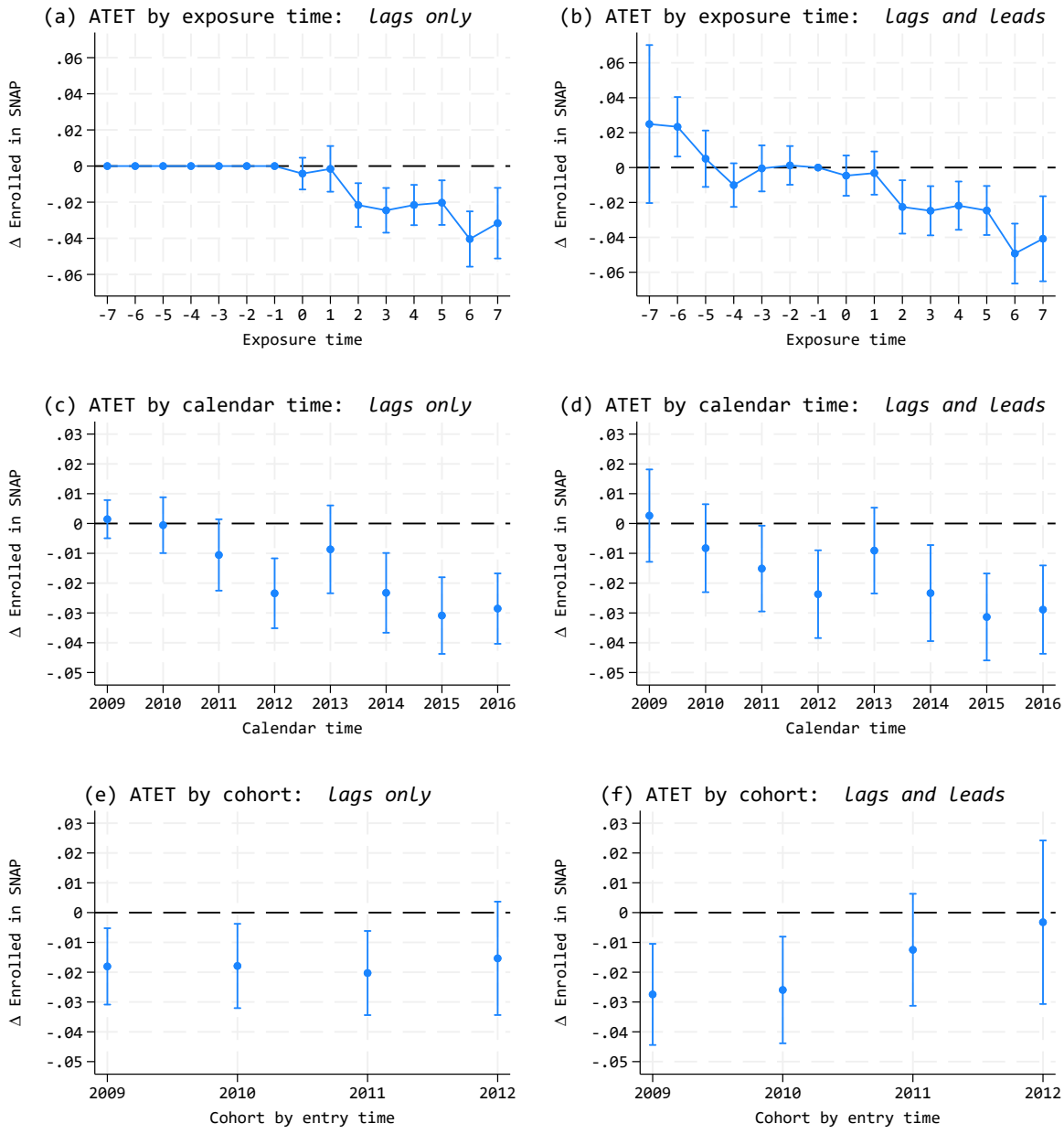
Notes: For models with heterogeneous effects, ATET is a weighted average of the estimand. Standard errors of ATET are based on cluster (at the PUMA level) robust standard errors of the coefficients in the estimand. FLEX refers to the model developed in this paper. SHo refers to the static homogeneous effect two-way fixed effects estimator.
`grp` denotes PUMA (the group), `yr` denotes year, `coh` denotes cohort, `rg` denotes indicators for the four regions of the country and `x` denotes the vector of individual-level regressors (an indicator for female sex, age, poverty status of the family and an indicator for whether the head of the family has a high school diploma). Dot (.) is used to denote interactions of indicator levels.

Figure 3: Overall ATET of the Secure Communities Immigration Enforcement Program on SNAP Enrollment in Lags-and-leads Specifications



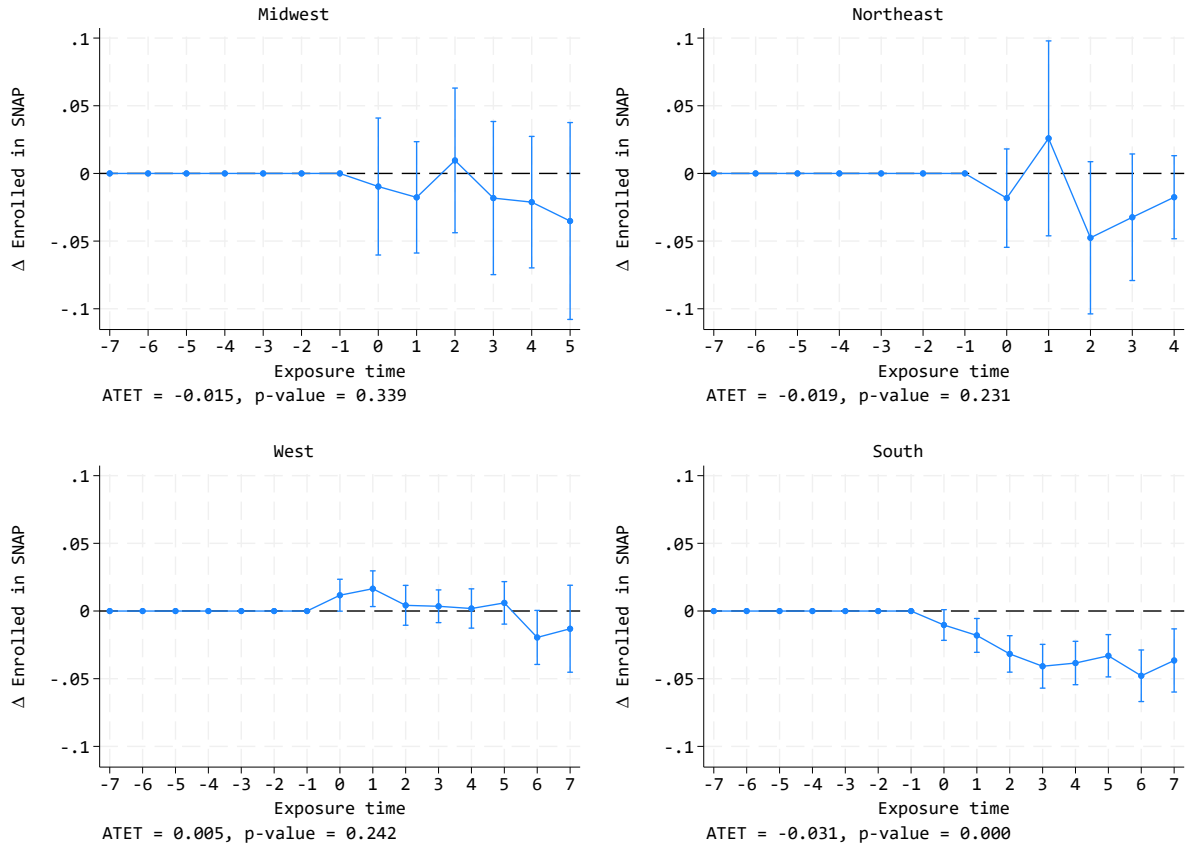
Notes: For models with heterogeneous effects, ATET is a weighted average of the estimand. Standard errors of ATET are based on cluster (at the PUMA level) robust standard errors of the coefficients in the estimand. FLEX refers to the model developed in this paper. ES refers to the heterogeneous over exposure-time but constant across groups two-way fixed effects estimator. CS refers to the Callaway and Sant'Anna (2021) regression-adjustment estimator.
`grp` denotes PUMA (the group), `yr` denotes year, `coh` denotes cohort, `rg` denotes indicators for the four regions of the country and `x` denotes the vector of individual-level regressors (an indicator for female sex, age, poverty status of the family and an indicator for whether the head of the family has a high school diploma). Dot (.) is used to denote interactions of indicator levels.

Figure 4: ATETs of the Secure Communities Immigration Enforcement Program on SNAP Enrollment by Exposure Time, Calendar Year and Cohort



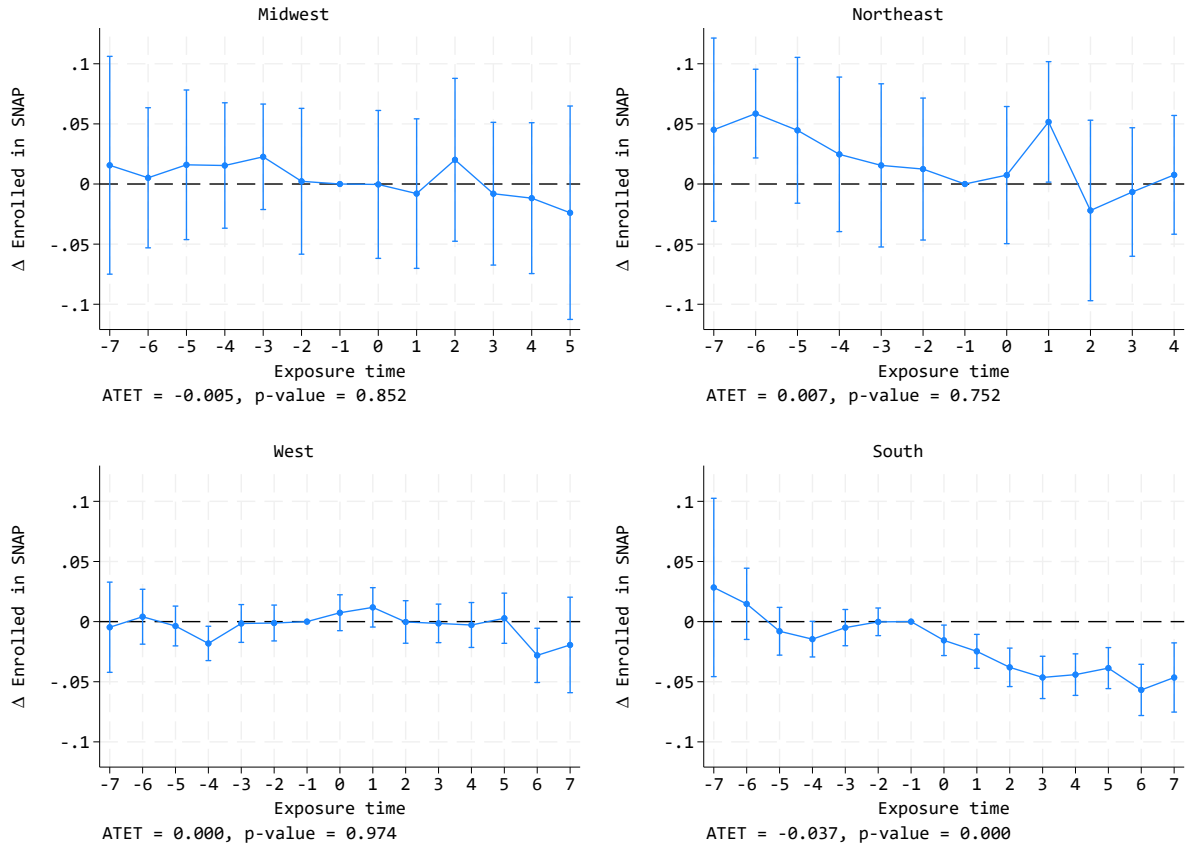
Notes: The FLEX specification estimates treatment effects at the cohort by year level fully interacted with individual socioeconomic characteristics and region-of-the-country indicators. The other regressors include cohort and year fixed effects, interactions of cohort and indicators with the other regressors and PUMA fixed effects entered additively. In the notation of Figure 3, this specification has treatment coefficients `coh.yr.(x+rg)` and other regressors `coh.x.reg yr.(x+rg) grp`.

Figure 5: ATET by Exposure Time from Lags-only Specifications for each Region



Notes: The lags-only FLEX specification estimates treatment effects at the cohort by year level fully interacted with individual socioeconomic characteristics and region-of-the-country indicators. The other regressors include cohort and year fixed effects, interactions of cohort and indicators with the other regressors and PUMA fixed effects entered additively. In the notation of Figure 3, this specification has treatment coefficients `coh.yr.(x+rg)` and other regressors `coh.x.reg yr.(x+rg) grp`.

Figure 6: ATET by Exposure Time from Lags-and-leads Specifications for each Region



Notes: The lags-and-leads FLEX specification estimates treatment effects at the cohort by year level fully interacted with individual socioeconomic characteristics and region-of-the-country indicators. The other regressors include cohort and year fixed effects, interactions of cohort and indicators with the other regressors and PUMA fixed effects entered additively. In the notation of Figure 3, this specification has treatment coefficients `coh.yr.(x+rg)` and other regressors `coh.x.reg yr.(x+rg) grp`.

Appendix A: Proofs

To prove Proposition 2.1, it is helpful to start with a lemma about general algebraic equivalences when using long and short regressions.

Lemma A.1. Consider a repeated cross-section data set $\left(Y_{i,t(i)}, \mathbf{G}_{i,t(i)}, \mathbf{H}_{i,t(i)}, W_{i,t(i)}\right)$ where $i = 1, ..., N$,$t = 1, ..., T$, and $\mathbf{G}_{i,t(i)}$ is $1 \times K$, $\mathbf{H}_{i,t(i)}$ is $1 \times L$, and $W_{i,t(i)}$ is a binary indicator. Here, $t\,(i)$ is the time period associated with observation i. Further, assume that $W_{i,t(i)}\mathbf{H}_{i,t(i)} = \mathbf{H}_{i,t(i)}$ for all i [so that $\left(1 - W_{i,t(i)}\right)\mathbf{H}_{i,t(i)} = \mathbf{0}$]. Let $\hat{\beta}$ and $\hat{\gamma}$ be the OLS estimates from the regression pooled across all observations:

$$Y_{i,t(i)} \text{ on } \mathbf{G}_{i,t(i)}, \mathbf{H}_{i,t(i)}, i = 1, ..., N. \tag{A.1}$$

Let $\tilde{\beta}$ be the estimates from the short POLS regression using the $W_{i,t(i)} = 0$ sample:

$$Y_{i,t(i)} \text{ on } \mathbf{G}_{i,t(i)} \text{ if } W_{i,t(i)} = 0 \tag{A.2}$$

Define the residuals from the short regression, for all i,

$$\tilde{V}_{i,t(i)} = Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\tilde{\beta} \tag{A.3}$$

Let $\tilde{\gamma}$ be the $L \times 1$ vector of coefficients from the regression

$$\tilde{V}_{i,t(i)} \text{ on } \mathbf{H}_{i,t(i)} \text{ if } W_{i,t(i)} = 1 \tag{A.4}$$

Suppose the OLS rank condition holds such that $\hat{\beta}$ and $\hat{\gamma}$ are unique and, for all i,

$$W_{i,t(i)}\mathbf{G}_{i,t(i)} = \mathbf{H}_{i,t(i)}\mathbf{A}$$

for an $L \times K$ matrix $\mathbf{A}$. Then

$$\tilde{\beta} = \hat{\beta} \text{ and } \tilde{\gamma} = \hat{\gamma}. \ \square$$

Proof: The POLS problem defining $\left(\hat{\beta}, \hat{\gamma}\right)$ is

$$\min_{\beta,\gamma} \sum_{i=1}^{N} \left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\beta - \mathbf{H}_{i,t(i)}\gamma\right)^2 \tag{A.5}$$

Because $\left(1 - W_{i,t(i)}\right)\mathbf{H}_{i,t(i)} = \mathbf{0}$, this problem is the same as

$$\min_{\beta,\gamma} \sum_{i=1}^{N} \left(1 - W_{i,t(i)}\right)\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\beta\right)^2 + \sum_{i=1}^{N} W_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\beta - \mathbf{H}_{i,t(i)}\gamma\right)^2$$

The first order conditions (FOCs) can be written as

$$\sum_{i=1}^{N} \left(1 - W_{i,t(i)}\right)\mathbf{G}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta}\right) + \sum_{i=1}^{N} W_{i,t(i)}\mathbf{G}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta} - \mathbf{H}_{i,t(i)}\hat{\gamma}\right) = \mathbf{0} \tag{A.6}$$

$$\sum_{i=1}^{N} W_{i,t(i)}\mathbf{H}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta} - \mathbf{H}_{i,t(i)}\hat{\gamma}\right) = \mathbf{0} \tag{A.7}$$

Now substitute $W_{i,t(i)}\mathbf{G}_{i,t(i)} = W_{i,t(i)}\mathbf{H}_{i,t(i)}\mathbf{A}$ and use simple algebra to write the FOCs as

$$\sum_{i=1}^{N} \left(1 - W_{i,t(i)}\right)\mathbf{G}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta}\right) + \mathbf{A}'\sum_{i=1}^{N} W_{i,t(i)}\mathbf{H}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta} - \mathbf{H}_{i,t(i)}\hat{\gamma}\right) = \mathbf{0} \tag{A.8}$$

$$\sum_{i=1}^{N} W_{i,t(i)}\mathbf{H}'_{i,t(i)}\left(Y_{i,t(i)} - \mathbf{G}_{i,t(i)}\hat{\beta} - \mathbf{H}_{i,t(i)}\hat{\gamma}\right) = \mathbf{0} \tag{A.9}$$

The second term in equation (A.8), being a linear combination of the left hand side of equation (A.9), is redundant. In particular, $\left(\hat{\beta}, \hat{\gamma}\right)$ solves the two sets of equations

$$\sum_{i=1}^{N}\left(1-W_{i,t(i)}\right)\mathbf{G}'_{i,t(i)}\left(Y_{i,t(i)}-\mathbf{G}_{i,t(i)}\hat{\beta}\right)= \qquad \mathbf{0} \tag{A.10}$$

$$\sum_{i=1}^{N}W_{i,t(i)}\mathbf{H}'_{i,t(i)}\left(Y_{i,t(i)}-\mathbf{G}_{i,t(i)}\hat{\beta}-\mathbf{H}_{i,t(i)}\hat{\gamma}\right)= \qquad \mathbf{0} \tag{A.11}$$

But equation A.10 just says that $\hat{\beta}$ is the OLS estimator of $Y_{i,t(i)}$ on $\mathbf{G}_{i,t(i)}$ using $W_{i,t(i)}=0$ – that is, $\hat{\beta}=\tilde{\beta}$. Inserting $\tilde{\beta}$ for $\hat{\beta}$ in equation (A.11) gives

$$\sum_{i=1}^{N}W_{i,t(i)}\mathbf{H}'_{i,t(i)}\left(\tilde{V}_{i,t(i)}-\mathbf{H}_{i,t(i)}\hat{\gamma}\right)=\mathbf{0};$$

but this simply says that $\hat{\gamma}$ can be obtained from equation (A.4), and so $\hat{\gamma}=\tilde{\gamma}$. □

Proof of Proposition 2.1

We can apply this result to prove Proposition 2.1 with $\hat{\gamma}=\hat{\tau}$ (the vector of ATETs) and $\hat{\beta}=\tilde{\beta}$ the estimates in the first-step of the imputation. Take

$$\begin{aligned}\mathbf{G}_{i,t(i)}=\big(&R_{i1},...,R_{iG},P_{2t(i)},...,P_{Tt(i)}\ ,\\ &R_{i1}\cdot\mathbf{X}_{i,t(i)},...,R_{iG}\cdot\mathbf{X}_{i,t(i)},P_{2t(i)}\cdot\mathbf{X}_{i,t(i)},...,P_{Tt(i)}\cdot\mathbf{X}_{i,t(i)}\big)\end{aligned} \tag{A.12}$$

$$\begin{aligned}\mathbf{H}_{i,t(i)}=\quad&\big(R_{i1}\cdot P_{c(1),t(i)},...,R_{i1}\cdot P_{T,t(i)},...,R_{i\bar{g}}\cdot P_{c(\bar{g}),t(i)},...,R_{i\bar{g}}\cdot P_{T,t(i)}\\ &R_{i1}\cdot P_{c(1),t(i)}\cdot\dot{\mathbf{X}}_{i,t(i),1,c(1)},...,R_{i1}\cdot P_{c(1),t(i)}\cdot\dot{\mathbf{X}}_{i,t(i),1,T},...,\\ &R_{i\bar{g}}\cdot P_{c(\bar{g}),t(i)}\cdot\dot{\mathbf{X}}_{i,t(i),\bar{g},c(\bar{g})},...,R_{i\bar{g}}\cdot P_{T,t(i)}\cdot\dot{\mathbf{X}}_{i,t(i),\bar{g},T}\big)\end{aligned} \tag{A.13}$$

Note that the interactions involving the demeaned controls have zero averages, by construction, when averaged over the appropriate group-time period pair. For group one, the group

dummies are interacted with the time period dummies for $c(1)$, ..., T. For the final treated group, $\bar{g}$, $R_{i\bar{g}}$ is interacted with the period dummies for $c(\bar{g})$, ..., T.

The condition $W_{i,t(i)}\mathbf{H}_{i,t(i)} = \mathbf{H}_{i,t(i)}$ holds because of the definition of $W_{i,t(i)}$ as a linear combination of the $R_{ig}P_{s,t(i)}$. In particular, $W_{i,t(i)}R_{ig}P_{s,t(i)} = R_{ig}P_{s,t(i)}$. Also, each element of $W_{i,t(i)}\mathbf{G}_{i,t(i)}$ is a fixed linear combination of $\mathbf{H}_{i,t(i)}$. For an untreated group g, $W_{i,t(i)}R_{ig} = 0$, for a treated group, $W_{i,t(i)}R_{ig} = R_{ig}P_{c(g),t(i)} + \cdots + R_{ig}P_{T,t(i)}$. Also, $W_{i,t(i)}P_{s,t(i)} = \sum_{c(g)\leq s} R_{ig}P_{s,t(i)}$. The terms involving the $\mathbf{X}_{i,t(i)}$ are handled similarly.

Given the definition of $\mathbf{G}_{i,t(i)}$, it is immediate that $\tilde{V}_{it}$ in Lemma A.1 is $\widetilde{TE}_{i,t(i)}$ given in equation (A.13) (with a slight change in notation). All that remains is to characterize the coefficients in the regression

$$\widetilde{TE}_{i,t(i)} \text{ on } \mathbf{H}_{i,t(i)} \text{ if } W_{i,t(i)} = 1 \tag{A.14}$$

for the choice of $\mathbf{H}_{i,t(i)}$ in equation (A.13). Because every term in $\mathbf{H}_{i,t(i)}$ is a treatment indicator or a treatment indicator multiplied by covariates, the $W_{i,t(i)} = 1$ restriction is redundant. (We cannot have $R_{ig} \cdot P_{s,t(i)} = 1$ and $W_{i,t(i)} = 0$.) Moreover, the treatment indicators are mutually exclusive, so we can separate the regression of $\widetilde{TE}_{i,t(i)}$ on $\mathbf{H}_{i,t(i)}$ into regressions on several orthogonal pairs:

$$\widetilde{TE}_{i,t(i)} \text{ on } R_{ig} \cdot P_{s,t(i)},\ R_{ig} \cdot P_{s,t(i)}\dot{\mathbf{X}}_{i,t(i),g,s} \tag{A.15}$$

Because the covariates are centered about $\bar{\mathbf{X}}_{gt(i)}$,

$$\sum_{i=1}^{N} R_{ig} \cdot P_{s,t(i)} \cdot \left(R_{ig} \cdot P_{s,t(i)} \cdot \dot{\mathbf{X}}_{i,t(i),g,s}\right) = \sum_{i=1}^{N} R_{ig} \cdot P_{s,t(i)} \cdot \dot{\mathbf{X}}_{i,t(i),g,s} = \mathbf{0}$$

– that is, $R_{ig} \cdot P_{s,t(i)}$ and $R_{ig} \cdot P_{s,t(i)} \cdot \dot{\mathbf{X}}_{i,t(i),g,s}$ are orthogonal in sample. It follows that the coefficient on $R_{ig} \cdot P_{s,t(i)}$ in equation (A.14) is the same as dropping $R_{ig} \cdot P_{s,t(i)}\dot{\mathbf{X}}_{i,t(i),g,s}$.

Because $R_{ig} \cdot P_{s,t(i)}$ is a dummy variable, the coefficient on it is simply the average of $\widetilde{TE}_{i,t(i)}$ over the $R_{ig} \cdot P_{s,t(i)} = 1$ subsample – precisely as in equation (2.9) that produces $\hat{\tau}_{gt}$.

The previous argument can be extended to the lags-and-leads estimator at the cost of even more complicated notation.