Factorial Difference-in-Differences*

Yiqing Xu (Stanford) Anqi Zhao (Duke) Peng Ding (UC Berkeley)

(Forthcoming, *Journal of the American Statistical Association*)

Abstract

We formulate *factorial difference-in-differences* (FDID), a research design that extends canonical difference-in-differences (DID) to settings in which an event affects all units. In many panel data applications, researchers exploit cross-sectional variation in a baseline factor alongside temporal variation in the event, but the corresponding estimand is often implicit and the justification for applying the DID estimator remains unclear. We frame FDID as a factorial design with two factors, the baseline factor G and the exposure level Z, and define effect modification and causal moderation as the associative and causal effects of G on the effect of Z, respectively. Under standard DID assumptions of no anticipation and parallel trends, the DID estimator identifies effect modification but not causal moderation. Identifying the latter requires an additional *factorial parallel trends* assumption, that is, mean independence between G and potential outcome trends. We extend the framework to conditionally valid assumptions and regression-based implementations, and further to repeated cross-sectional data and continuous G. We demonstrate the framework with an empirical application on the role of social capital in famine relief in China.

Keywords: difference-in-differences, factorial design, panel data, parallel trends

*Yiqing Xu and Anqi Zhao contribute equally to this paper. Yiqing Xu, Assistant Professor, Department of Political Science, Stanford University. Email: yiqingxu@stanford.edu. Anqi Zhao, Assistant Professor, Fuqua School of Business, Duke University. Email: anqi.zhao@duke.edu. Peng Ding, Associate Professor, Department of Statistics, University of California, Berkeley. Email: pengdingpku@berkeley.edu.

1 Introduction

Social science research often relies on panel data to establish causality. One common approach involves exploiting cross-sectional variation in a baseline factor G and temporal variation in exposure to a common event affecting all units, and applying the difference-in-differences (DID) estimator in a panel setting. As our running example, Cao et al. (2022) examine how social capital (G), measured by the density of genealogy books, mitigated the mortality surge during China's Great Famine from 1958 to 1961 (the event), using a county-year panel. The authors interpret the coefficient of the interaction term between G and an indicator of the famine years from a two-way fixed effects (TWFE) regression as the causal effect of social capital on famine relief, and describe their approach as a DID method. Section 2 gives details on this study and five additional examples that employ a similar approach.

Although the coefficient from a TWFE regression is numerically equal to a DID estimate and this approach is often referred to as a DID method in the empirical literature, it differs from the canonical DID popularized by Card and Krueger (1994) because it lacks a clean control group unexposed to the event. The corresponding causal interpretation of the DID estimator in such settings is absent from the methodological literature, and empirical applications using this approach lack clear definitions of target causal estimands.

This paper aims to bring conceptual clarity to this empirical approach, which we term *factorial difference-in-differences* (FDID). We define FDID as a research design, or an identification strategy, that employs the DID estimator to recover interpretable quantities of interest using observations before and after a one-time event that affects all units, provided clearly stated identifying assumptions hold. Here, we highlight the distinction between the DID estimator and the canonical DID and FDID research designs. Applied to panel data, the DID estimator, denoted by $\hat{\tau}_{\text{DID}}$, calculates the difference in before-after differences with respect to an event between two groups, denoted by $G_i = 1$ and $G_i = 0$. In contrast, a research design encompasses not only the estimator but also the identifying assumptions and identification results. To simplify the presentation, we will refer to the FDID and canonical DID research designs as "FDID" and "canonical DID," respectively, when no confusion is likely to arise.

We present our main theoretical results for FDID in the two-group, two-period case. The key innovation is to augment the potential outcomes framework to include both the baseline factor G and the exposure indicator Z, motivating the term "factorial" in FDID. This formulation clarifies what the probability limit of $\hat{\tau}_{\text{DID}}$, denoted by τ_{DID}, identifies under different identification assumptions, and how these results relate to canonical DID.

In canonical DID, τ_{DID} identifies the average treatment effect on the treated (ATT) under the no anticipation and parallel trends assumptions (Angrist and Pischke, 2009). In FDID, the focus shifts to two other estimands: effect modification and causal moderation (VanderWeele, 2009; Bansak, 2020). Effect modification captures how the effect of Z varies across groups defined by G, but does not represent G's causal effect. In our running example, this corresponds to the statement that the mortality increase caused by the famine is smaller in counties with higher levels of social capital. Causal moderation, by contrast, has a direct causal interpretation as G's effect on the impact of Z, or symmetrically, Z's effect on the impact of G. In our example, causal moderation corresponds to two equivalent statements: (i) social capital reduced the famine's negative impact on mortality, as emphasized in the original paper, or (ii) social capital had a stronger effect on mortality during the famine than it would have otherwise. Our key result is that under the no anticipation and canonical parallel trends assumptions, τ_{DID} identifies effect modification, whereas recovering causal moderation requires additional assumptions. One such condition is the *factorial parallel trends* assumption, that is, mean independence between G and potential outcomes trends.

Intuitively, it holds in the absence of any unobserved confounder correlated with both G and the outcome trends. For example, to identify the causal moderation of social capital on the famine's effect on mortality, one must rule out any unobserved factor—such as income or governance quality—that is correlated with both social capital and changes in mortality between famine and non-famine years.

Moreover, we show that canonical DID can be reframed as a special case of FDID under an additional exclusion restriction requiring that exposure to the event has no effect on the outcome for units with $G_i = 0$. Under this assumption, G's effect modification simplifies to the average effect of Z on units with $G_i = 1$, analogous to the ATT in canonical DID. In our running example, this assumption implies that the famine had no impact on localities with low (or high) levels of social capital, an implausible claim in this setting. Alternatively, if researchers assume that G has no causal impact on the outcome in the absence of the event, then causal moderation reduces to G's average conditional effect given exposure to the event. With our example, this assumption implies that social capital would have no effect on mortality had the famine not occurred, which is possible but not suggested by the original paper.

We extend the framework to settings where canonical and factorial parallel trends hold only conditional on additional time-invariant covariates. We formalize identification results for the conditional DID estimator and clarify the assumptions needed to justify regression-based analysis. We also provide theoretical results for applications with repeated cross-sectional data and with a continuous baseline factor G.

Our contributions are twofold. First, we establish the causal interpretation of a widely used empirical approach in the social sciences, clarifying the identifying assumptions required to recover causal estimands of interest. This framework advances the discussion of causal panel analysis with the DID estimator and TWFE models—for recent reviews, see Roth et al.

(2023), Chiu et al. (2026), and Arkhangelsky and Imbens (2024). Second, we contribute to the literature on factorial designs (VanderWeele, 2009; Bansak, 2020; Han and Rubin, 2021; Pashley and Bind, 2023; Yu and Ding, 2026) by, to our knowledge, being the first to extend factorial designs to observational panel settings and to analyze the role of parallel trends assumptions in this context.

The rest of the paper is organized as follows. Section 2 gives six FDID examples appearing in the empirical literature. Section 3 formalizes FDID under the two-group, two-period panel case. Section 4 states identification results for FDID and reconciles FDID with canonical DID. Sections 5–6 discuss extensions to conditionally valid assumptions, repeated cross-sectional data, and continuous G. Section 7 illustrates our theory with the running example. Section 8 concludes. The Supplementary Materials provide technical details, and the replication files are available on GitHub at https://github.com/xuyiqing/fdid_paper.

2 FDID Examples

In this section, we present six empirical examples from economics, political science, and finance that align with the FDID research design. In each case, researchers obtain key estimates using a TWFE regression and describe the approach as a DID method. However, their intended estimands often differ.

Squicciarini (2020) examines whether Catholicism, proxied by the share of refractory clergy in 1791, hindered economic growth during the Second Industrial Revolution in France. One main analysis relies on a longitudinal dataset of French departments from 1866 to 1911. The baseline factor G is the 1791 share of refractory clergy, and the event is the Second Industrial Revolution in the late 19th century (post-1870), to which all departments were presumably exposed. It is not entirely clear about which estimand the study aims to identify. The author argues that the results pointed to a causal interpretation of the relationship between religiosity and economic development *during* the Second Industrial Revolution, which corresponds most closely to G's average conditional effect. The study also describes itself as examining "the differential diffusion of technical education and industrial development" (p. 3455), which can be interpreted as targeting the effect modification of G.

Fouka (2019) studies "the effect of taste-based discrimination on the assimilation decisions of immigrant minorities" in the United States (abstract). The repeated cross-sectional data include all men born in the US from 1880 to 1930 to a German-born father, organized by state and birth year. The baseline factor G is state-level measures of anti-Germanism, such as support for Woodrow Wilson in the 1916 presidential Election. The event is World War I, which started in 1917. The outcome is a foreign name index. Although not stated explicitly, the intended estimand is likely either the causal moderation or G's average conditional effect post World War I.

Charnysh (2022) argues that during crises, states allocate fewer resources to less "legible" groups—those from which they cannot gather reliable information or effectively collect taxes. Using district-level panel data with yearly observations from Imperial Russia, the author studies the 1891–1892 Russian famine. The baseline factor G is the district-level share of Muslims, and the event is the famine. The findings show that districts with larger Muslim populations experienced higher mortality rates during the famine. The author avoids explicit causal claims, so the analysis can be interpreted as targeting the effect modification of G.

Chen et al. (2025) argue during a major state-building episode in ancient China, the number of aristocrats from prefectures recruited into the imperial bureaucracy increased more in localities with strong military presence. The study uses prefecture-level panel data from the Northern Wei Dynasty (384–534 CE). The baseline factor G is whether a prefecture had fourth-century military strongholds, and the event is a state-building reform initiated by Empress Dowager Feng (477–490 CE). The authors describe their empirical strategy as a "canonical DD strategy," which "relies on the parallel-trend assumption to adopt a causal interpretation." Given the use of explicit causal language, the intended estimand is likely either the causal moderation or the average conditional effect of G.

Chen et al. (2023) examine how the ability to use corporate bonds as collateral (pledgeability) affects their prices, leveraging a policy change in Chinese bond markets. The study uses panel data on daily bond prices. The baseline factor G is bond ratings, and the event is a policy change on December 8, 2014, when Chinese policymakers prohibited bonds rated below AAA from being used as collateral. The intended estimand is the ATT, the policy effect on AA and AA+ bonds. This setting reduces to canonical DID because the exclusion restriction—i.e., no policy effect on AAA and AA− bonds—is plausible, given that AA− bonds were already ineligible for pledging before the policy change.

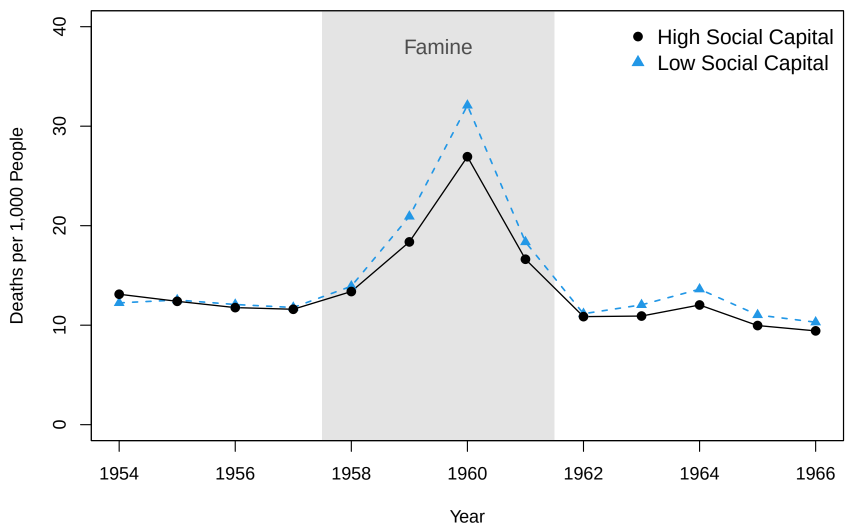

Figure 1: Average mortality rate across counties with high and low social capital. This figure resembles Figure 5(a) in Cao et al. (2022), although we use a balanced panel, which represents a subset of the data used by the authors.

Finally, Cao et al. (2022) is our running example. Recall from Section 1 that the baseline factor G is social capital, and the event is China's Great Famine from 1958 to 1961. The authors intend to estimate the causal moderation of social capital on the famine's impact on mortality. In Section 7, we reanalyze this application using a balanced panel of 921 counties spanning the years 1954 to 1966. Figure 1 displays the average mortality rates during this period for two types of counties in the sample: those with high social capital and those with low social capital. While the average mortality rate rose sharply during the famine years in both groups of counties, the increase was noticeably higher in counties with low social capital compared with those with high social capital.

3 FDID in Two-Group, Two-Period Panel Setting

We present our main theoretical results in the two-group, two-period panel setting. This section introduces the notation, data structure, and DID estimator, and then defines the potential outcomes that form the basis for the estimands.

3.1 Observed data and FDID setting

Assume a standard two-group, two-period panel setting with a one-time event and a study sample of n units, indexed by $i = 1, \ldots, n$. For each unit i, we observe a binary baseline factor $G_i \in \{0, 1\}$, an outcome measured at two time points—before and after the event—denoted by $Y_{i,\text{pre}} \in \mathbb{R}$ and $Y_{i,\text{post}} \in \mathbb{R}$, and an exposure indicator $Z_i \in \{0, 1\}$, where $Z_i = 1$ if unit i is exposed when the event occurs and $Z_i = 0$ otherwise. A defining feature of FDID is that all units are exposed to the event, so $Z_i = 1$ for all i, as formalized in Assumption 1 and Definition 1 below.

Assumption 1 (Universal exposure)**.** $Z_i = 1$ for all $i = 1, \ldots, n$.

Definition 1 (FDID setting)**.** The two-group, two-period (2×2) panel setting for FDID consists of the 2×2 panel data $\{(G_i, Y_{i,\text{pre}}, Y_{i,\text{post}}, Z_i) : G_i \in \{0, 1\}\}_{i=1}^{n}$ and Assumption 1.

In the FDID setting, the exposure indicator Z_i equals one for all units and thus appears redundant. However, it is essential for defining the potential values of $Y_{i,\text{pre}}$ and $Y_{i,\text{post}}$ that would have been observed in the absence of the event. These potential outcomes provide the basis for defining the causal estimands and stating the identification assumptions in FDID. A similar use of Z_i appears in Holland and Rubin (1986) to clarify Lord's paradox in settings identical to FDID. The departure from canonical DID arises from the absence of a one-to-one mapping between G and Z.

3.2 DID estimator

Let $\Delta Y_i = Y_{i,\text{post}} - Y_{i,\text{pre}}$ denote the before-after difference in the outcomes of unit i. The DID estimator is the difference in the average ΔY_i between the two groups defined by the baseline factor G:

$$\hat{\tau}_{\text{DID}} = n_1^{-1} \sum_{i:G_i=1} \Delta Y_i - n_0^{-1} \sum_{i:G_i=0} \Delta Y_i, \tag{1}$$

where n_g is the number of units with $G_i = g$. We assume throughout that units are drawn from a common population distribution. Define

$$\tau_{\text{DID}} = \mathbb{E}[\Delta Y_i \mid G_i = 1] - \mathbb{E}[\Delta Y_i \mid G_i = 0] \tag{2}$$

as the probability limit of $\hat{\tau}_{\text{DID}}$ as $n \to \infty$, commonly referred to as the DID estimand. Our goal is to clarify the causal interpretation of τ_{DID} in the FDID setting under various identifying assumptions.

Remark 1. From (1)–(2), the DID estimator and estimand, $(\hat{\tau}_{\text{DID}}, \tau_{\text{DID}})$, depend on the panel outcomes $(Y_{i,\text{pre}}, Y_{i,\text{post}})$ only through their difference ΔY_i, and equal the difference-in-means estimator and estimand based on the cross-sectional data $(G_i, \Delta Y_i)_{i=1}^n$. This equivalence between DID and cross-sectional analyses underlies the key identifying assumptions that justify the causal interpretation of τ_{DID}, as discussed in Section 4.

3.3 Potential outcomes

We now define the potential outcomes under FDID. Unlike the classic DID framework, we define them with respect to both the baseline factor G_i and the exposure level Z_i. For $t \in \{\text{pre}, \text{post}\}$, let $Y_{it}(g, z)$ denote the potential value of Y_{it} if G were set at g and Z were set at z for unit i. Let $1_{\{\cdot\}}$ be the indicator function. The observed outcome satisfies $Y_{it} = \sum_{g,z=0,1} 1_{\{(G_i,Z_i)=(g,z)\}} Y_{it}(g, z) = Y_{it}(G_i, Z_i)$, which reduces to $Y_{it} = Y_{it}(G_i, 1)$ under Assumption 1. Thus, the four potential outcomes with $z = 0$, $\{Y_{it}(g, 0) : t \in \{\text{pre}, \text{post}\},\ g \in \{0, 1\}\}$, are unobservable for all units in the FDID setting. Figure 2 illustrates this setup.

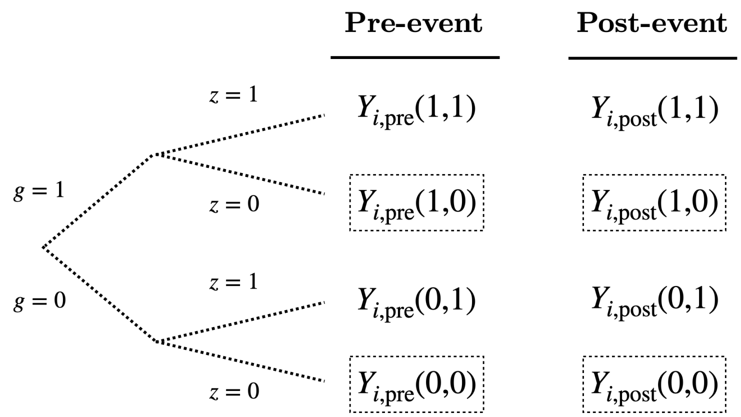

Figure 2: Potential outcomes under FDID. The four potential outcomes with $z = 0$ in dotted boxes are unobservable for any unit in the FDID setting.

3.4 Estimands

We now formalize four estimands that τ_{DID} can identify in the FDID setting under different assumptions. Following the literature on causal inference with factorial experiments (Dasgupta et al., 2015; Bansak, 2020; Zhao and Ding, 2021), we begin by defining three unit-level effects. Let $\tau_{i,Z|G=g} = Y_{i,\text{post}}(g, 1) - Y_{i,\text{post}}(g, 0)$ denote the effect of exposure on unit i if the baseline factor G were at level g. Let $\tau_{i,G|Z=z} = Y_{i,\text{post}}(1, z) - Y_{i,\text{post}}(0, z)$ denote the effect of the baseline factor G on unit i if its exposure level were at z. Define

$$\tau_{i,\text{cm}} = \tau_{i,Z|G=1} - \tau_{i,Z|G=0}$$

as the *causal moderation* of G on the effect of exposure for unit i, capturing how the effect of exposure differs between the two levels of G. Note that

$$\tau_{i,\text{cm}} = Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) - Y_{i,\text{post}}(0,1) + Y_{i,\text{post}}(0,0) = \tau_{i,G|Z=1} - \tau_{i,G|Z=0}.$$

Thus, $\tau_{i,\text{cm}}$ is symmetric in G and Z, and can also be interpreted as the causal moderation of Z on the effect of G for unit i. In the literature (see, e.g., VanderWeele, 2009), $\tau_{i,\text{cm}}$ is also referred to as the *interaction* between G and Z.

Definition 2 below formalizes effect modification and causal moderation as comparisons of the exposure effect, $\tau_{i,Z|G=g}$, between the two levels of G, extending VanderWeele (2009) and Bansak (2020) to the panel setting. To highlight the distinction between these concepts, define $\tau_{i,Z|G}$ as the value of $\tau_{i,Z|G=g}$ when the baseline factor G_i is at its observed level, i.e.,

$$\tau_{i,Z|G} = \begin{cases} \tau_{i,Z|G=1} & \text{if } G_i = 1; \\ \tau_{i,Z|G=0} & \text{if } G_i = 0. \end{cases}$$

Definition 2 (Effect modification and causal moderation)**.** Define

$$\tau_{\text{em}} = \mathbb{E}[\tau_{i,Z|G} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G} \mid G_i = 0] = \mathbb{E}[\tau_{i,Z|G=1} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G=0} \mid G_i = 0]$$

as the *effect modification* of the baseline factor G on the effect of exposure;

$$\tau_{\text{cm}} = \mathbb{E}[\tau_{i,\text{cm}}] = \mathbb{E}[\tau_{i,Z|G=1} - \tau_{i,Z|G=0}]$$

as the *causal moderation* of the baseline factor G on the effect of exposure.

Both τ_{em} and τ_{cm} capture heterogeneity in the effect of exposure, $\tau_{i,Z|G=g}$, across the two levels of the baseline factor G, but they emphasize different comparisons. On the one hand, τ_{cm} compares $\tau_{i,Z|G=1}$ and $\tau_{i,Z|G=0}$ across all units and is conventionally regarded as a causal quantity (Holland and Rubin, 1986; Frangakis and Rubin, 2002; Imbens and Rubin, 2015). It addresses the causal question: Would the effect of exposure change if one intervened on G? In our running example, this corresponds to whether the impact of exposure to China's Great Famine would have differed if a locality had randomly developed social capital, perhaps due to the migration of a large kinship clan. By contrast, τ_{em} compares the average of $\tau_{i,Z|G=1}$ among units with $G_i = 1$ to the average of $\tau_{i,Z|G=0}$ among units with $G_i = 0$, each evaluated at the observed value of G_i. Accordingly, τ_{em} describes differences in exposure effects between the two groups of units defined by G, but does not address the same causal question as τ_{cm}. See VanderWeele (2009) for further discussion in the cross-sectional setting.

Remark 2. In some applications, G is an inherent characteristic of the unit, such as geography of a locality or race of a person, which cannot be meaningfully manipulated. In these cases, the augmented potential outcomes $Y_{i,t}(g, z)$ are not well defined at the counterfactual level of G, and thus τ_{cm} is not a coherent causal estimand. Instead, τ_{em} is a more relevant target. In addition, the causal moderation τ_{cm} is the average of $\tau_{i,\text{cm}}$ across all units, analogous to the average treatment effect. One may also consider $\mathbb{E}[\tau_{i,\text{cm}} \mid G_i = g]$, the average causal moderation among units with $G_i = g$, analogous to the ATT.

Definition 3 below formalizes two conditional causal effects of G.

Definition 3 (Conditional causal effects of G)**.** Define

$$\tau_{\text{att}} = \mathbb{E}[\tau_{i,Z|G} \mid G_i = 1] = \mathbb{E}[\tau_{i,Z|G=1} \mid G_i = 1]$$

as the average causal effect of exposure on units with $G_i = 1$, and

$$\tau_{G|Z=1} = \mathbb{E}[\tau_{i,G|Z=1}]$$

as the average causal effect of the baseline factor G conditional on exposure.

From Definition 3, τ_{att} is analogous to the ATT in canonical DID, with units satisfying $G_i = 1$ viewed as the treated group. The subscript "att" reflects this connection. The estimand $\tau_{G|Z=1}$ is the average causal effect of G conditional on exposure to the event. Figure 3 summarizes the relationships among these estimands, serving as a roadmap for our identification results.

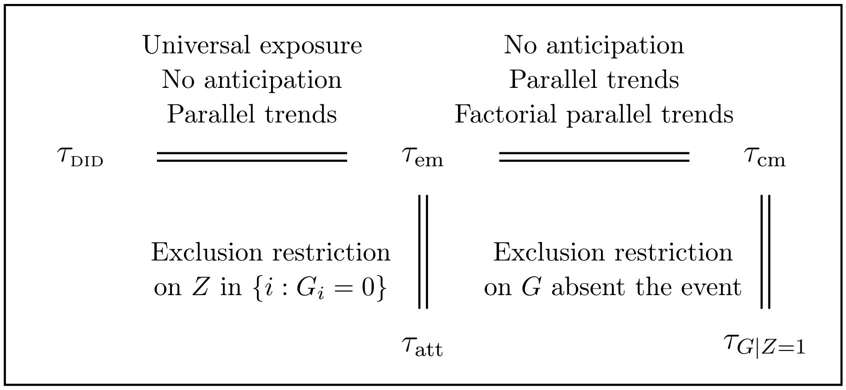

Figure 3: Roadmap for key identification results

4 Identification

We now present the identification results. To preview, τ_{DID} identifies τ_{em} under the canonical DID assumptions; identifying τ_{cm}, τ_{att}, and $\tau_{G|Z=1}$ requires additional assumptions.

4.1 Identification under canonical DID assumptions

The observed pre-event outcome $Y_{i,\text{pre}}$ and its potential values $\{Y_{i,\text{pre}}(g,z) : g, z = 0, 1\}$ all occur before the event. A common, often implicit, assumption in the DID literature is that future events do not influence past potential outcomes, known as the *no anticipation* assumption. We state this assumption explicitly in Assumption 2 below. It may be violated if units anticipate the event and adjust their behavior in advance.

Assumption 2 (No anticipation)**.** $Y_{i,\text{pre}}(g, 0) = Y_{i,\text{pre}}(g, 1)$ for $i = 1, \ldots, n$ and $g = 0, 1$.

Recall that $\Delta Y_i = Y_{i,\text{post}} - Y_{i,\text{pre}}$ is the before-after difference in outcome for unit i. Let $\Delta Y_i(g, z) = Y_{i,\text{post}}(g, z) - Y_{i,\text{pre}}(g, z)$ denote the potential value of ΔY_i if G were set at g and Z were set at z for unit i. Define $\Delta Y_i(G_i, 0) = Y_{i,\text{post}}(G_i, 0) - Y_{i,\text{pre}}(G_i, 0)$ as the potential before-after change for unit i under its observed baseline factor G but assuming no exposure. Assumption 3 below restates the canonical parallel trends assumption using the augmented potential outcomes.

Assumption 3 (Canonical parallel trends)**.** $\mathbb{E}[\Delta Y_i(G_i, 0) \mid G_i = 1] = \mathbb{E}[\Delta Y_i(G_i, 0) \mid G_i = 0]$.

Assumption 3 states that, in the absence of the event, the average change in outcome over time, $\Delta Y_i(G_i, 0)$, would be the same across the two groups defined by the baseline factor G. Together, Assumptions 2–3 form the canonical identifying assumptions in the DID literature, under which τ_{DID} identifies the ATT in the canonical DID setting. Proposition 1 below extends this classic result to the FDID setting and shows that, under these assumptions, τ_{DID} identifies τ_{em}. Figure 4 illustrates this identification result.

Proposition 1. If Assumptions 1–3 hold, then $\tau_{\text{DID}} = \tau_{\text{em}}$.

Definition 1 and Proposition 1 underscore two key differences between FDID and canonical DID. First, FDID assumes that all units are exposed to the event, whereas canonical DID relies on a clean, unexposed control group. Second, under the no anticipation and canonical parallel trends assumptions, the DID estimator identifies the ATT, a causal quantity, in canonical DID, but identifies τ_{em}, a descriptive quantity, in FDID. Despite these differences, we show below that the canonical DID research design can be reframed as a special case of FDID under an additional exclusion restriction assumption that the event has no effect on a group of units defined by G.

Assumption 4 (Exclusion restriction)**.** $\mathbb{E}[Y_{i,\text{post}}(0, 1) \mid G_i = 0] = \mathbb{E}[Y_{i,\text{post}}(0, 0) \mid G_i = 0]$.

Assumption 4 ensures that the average post-event outcome of units with $G_i = 0$ is unaffected by exposure. Conceptually, these units are exposed but unaffected, thus resembling the clean control group in canonical DID. Together, Assumption 1 (universal exposure) and Assumption 4 (exclusion restriction) reproduce a canonical DID setting under FDID, with units satisfying $G_i = 1$ and $G_i = 0$ serving as the treated and control groups, respectively. This justifies interpreting τ_{att} as the ATT analog in FDID. Moreover, Assumption 4 implies

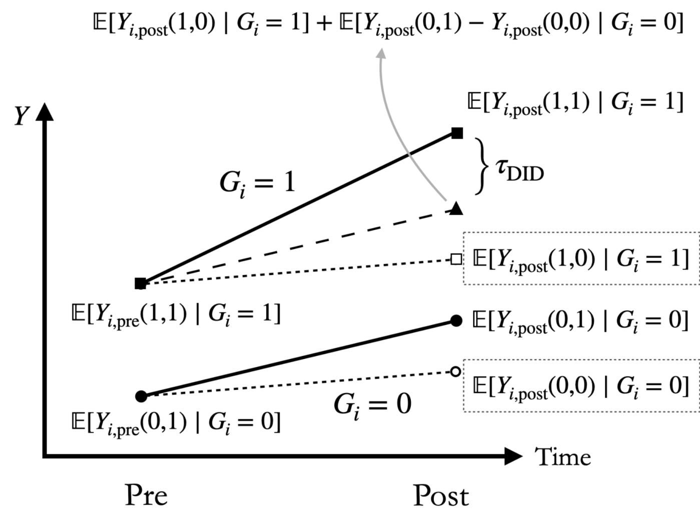

Figure 4: Identification result under FDID in Proposition 1

Note: Squares (■ and □) represent the group with $G_i = 1$, and circles (● and ○) represent the group with $G_i = 0$. By Assumption 2 (no anticipation), in the pre-event period, ■ $= \mathbb{E}[Y_{i,\text{pre}}(1,0) \mid G_i = 1] = \mathbb{E}[Y_{i,\text{pre}} \mid G_i = 1]$ and ● $= \mathbb{E}[Y_{i,\text{pre}}(0,0) \mid G_i = 0] = \mathbb{E}[Y_{i,\text{pre}} \mid G_i = 0]$, so ■ and ● represent the pre- and post-event observed outcomes of the two groups. The long dashed line is drawn parallel to the solid line for $G_i = 0$ to represent the definition of the DID estimand, so in the post-period the distance between ■ and ▲ equals τ_{DID}. This gives the first expression of ▲:

$$▲ = ■ - \tau_{\text{DID}} = \mathbb{E}[Y_{i,\text{post}}(1,1) \mid G_i = 1] - \tau_{\text{DID}}.$$

Under Assumption 3, the two short dashed lines are also parallel. These parallel lines ensure that in the post period, the distance between ▲ and □ equals the distance between ● and ○. This gives the second expression of ▲:

$$▲ = □ + (● - ○) = \mathbb{E}[Y_{i,\text{post}}(1,0) \mid G_i = 1] + \mathbb{E}[Y_{i,\text{post}}(0,1) - Y_{i,\text{post}}(0,0) \mid G_i = 0].$$

Equating the two expressions for ▲ yields

$$\tau_{\text{DID}} = \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 1] - \mathbb{E}[Y_{i,\text{post}}(0,1) - Y_{i,\text{post}}(0,0) \mid G_i = 0] = \tau_{\text{em}}.$$

$\mathbb{E}[\tau_{i,Z|G=0} \mid G_i = 0] = 0$, so that $\tau_{\text{em}} = \mathbb{E}[\tau_{i,Z|G=1} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G=0} \mid G_i = 0] = \tau_{\text{att}}$. This provides a causal interpretation of τ_{em}, as formalized in Proposition 2 below.

Proposition 2. (i) If Assumption 4 holds, then $\tau_{\text{em}} = \tau_{\text{att}}$. (ii) If Assumptions 1–4 hold, then $\tau_{\text{DID}} = \tau_{\text{att}}$.

Recall that τ_{att} is the ATT analog under FDID. Proposition 2 reframes the classic identification result for the ATT in canonical DID within the FDID framework. Definition 4 builds on this result and characterizes canonical DID as a special case of FDID. In Figure 4, this corresponds to the post-period gaps between the solid triangle and hollow square, and between the solid circle and hollow circle, being closed.

Definition 4 (Reframed canonical DID research design)**.** The canonical DID setting is equivalent to the FDID setting in Definition 1 combined with Assumption 4, if the groups defined by G, $\{i : G_i = 1\}$ and $\{i : G_i = 0\}$, are viewed as the treatment and control groups, respectively. Under this setting, effect modification τ_{em} reduces to the ATT analog τ_{att}, and is identified by τ_{DID} under Assumptions 2–3.

4.2 Identifying causal moderation and G's conditional effect

Propositions 1–2 establish the identification of τ_{em} and τ_{att} under FDID. In many applied studies, however, the primary quantity of interest is the causal moderation τ_{cm} or the causal effect of G given exposure, $\tau_{G|Z=1}$; c.f. Section 2. We discuss their identification below.

Recall that $\Delta Y_i(g, z) = Y_{i,\text{post}}(g, z) - Y_{i,\text{pre}}(g, z)$ denotes the potential before-after change in outcome for unit i. Assumption 5 below introduces a *factorial parallel trends* assumption, which requires mean independence between G_i and $\Delta Y_i(g, z)$.

Assumption 5 (Factorial parallel trends)**.** $\mathbb{E}[\Delta Y_i(g, z) \mid G_i = 1] = \mathbb{E}[\Delta Y_i(g, z) \mid G_i = 0]$ for $g, z = 0, 1$.

We call Assumption 5 the factorial parallel trends assumption because, like canonical parallel trends (Assumption 3), it requires equal average changes in potential outcomes across groups. However, now $\Delta Y_i(g, z)$ varies both G and Z, implying mean independence between G and all four potential outcome changes, $\Delta Y_i(g, z)$. By contrast, Assumption 3 requires only mean independence between G_i and $\Delta Y_i(G_i, 0)$, holding G fixed at its observed value. A sufficient condition for Assumption 5 is

$$G_i \perp\!\!\!\perp \{\Delta Y_i(g, z) : g, z = 0, 1\}, \tag{3}$$

which states that G is independent of the before-after changes in all four potential outcomes. As noted in Remark 1, DID analysis with $(G_i, Y_{i,\text{pre}}, Y_{i,\text{post}})$ in FDID corresponds to cross-sectional analysis with $(G_i, \Delta Y_i)$. Analogously, condition (3) resembles the standard random assignment assumption for $(G_i, \Delta Y_i, Z_i)$ in cross-sectional settings. Condition (3) reflects the belief that differencing the potential outcomes removes confounding with respect to G. As noted earlier, some applied researchers do not recognize that Assumption 5 or condition (3) is required to interpret DID estimates as causal moderation. Others, while not explicitly stating it, appear to have an intuitive understanding of this assumption and view it as more plausible than full random assignment $G_i \perp\!\!\!\perp Y_i(g, z)$.[1] Nevertheless, the assumption, like unconfoundedness, is strong and untestable.

Proposition 3 below formalizes the conditions under which τ_{cm} and τ_{em} to coincide, thereby ensuring that τ_{DID} identifies τ_{cm}.

Proposition 3. (i) Under Assumptions 2–3 and 5, $\tau_{\text{em}} = \tau_{\text{cm}}$. (ii) Under Assumptions 1–3 and 5, $\tau_{\text{DID}} = \tau_{\text{em}} = \tau_{\text{cm}}$.

Remark 3. The condition required for $\tau_{\text{em}} = \tau_{\text{cm}}$ is not unique. For example, mean independence between G_i and $\{\tau_{i,Z|G=g} : g = 0, 1\}$ also implies $\tau_{\text{cm}} = \tau_{\text{em}}$. We focus on Assumption 5 because it is the one most often invoked, implicitly or explicitly, in empirical applications of FDID as researchers frequently argue that G_i and $\Delta Y_i(g, z)$ are mean independent after conditioning on additional baseline covariates (e.g., Fouka, 2019; Cao et al., 2022), a setting we extend to in Section 5.

Remark 4. Assumption 5 (factorial parallel trends), together with Assumption 2 (no

[1] For example, Fouka (2019) writes: "I control for the potential time-varying effect of the share of the German population in the state, which is plausibly correlated with both (lower) support for Wilson and assimilation," and "state-level anti-Germanism is potentially endogenous to pre-existing trends in German assimilation" (p. 419), clearly acknowledging potential correlation between G_i and $\Delta Y_i(g, z)$.

anticipation), ensures that $\tau_{\text{att}} = \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 0] = \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0)]$, which is the conditional average causal effect of exposure to the event with $g = 1$. However, Assumption 5 does not imply that $\tau_{\text{att}} = \mathbb{E}[Y_{i,\text{post}}(G_i,1) - Y_{i,\text{post}}(G_i,0)]$, the average treatment effect of exposure for all units. See Section A2.1 in the Supplementary Materials for more discussion.

Assumption 6 below introduces an alternative exclusion restriction, which requires that in the absence of the event, G would not have affected the average post-period outcome. It implies $\tau_{\text{cm}} = \tau_{G|Z=1}$ so that τ_{DID} also identifies $\tau_{G|Z=1}$, as formalized in Proposition 4.

Assumption 6 (Exclusion restriction absent the event)**.** $\mathbb{E}[Y_{i,\text{post}}(1,0)] = \mathbb{E}[Y_{i,\text{post}}(0,0)]$.

Proposition 4. (i) If Assumption 6 holds, then $\tau_{\text{cm}} = \tau_{G|Z=1}$. (ii) If Assumptions 1–3 and 5–6 hold, then $\tau_{\text{DID}} = \tau_{\text{cm}} = \tau_{G|Z=1}$.

Propositions 1–4 complete the roadmap in Figure 3. Definition 5 below formalizes the *FDID research design* as the combination of the FDID setting in Definition 1 with these identification results.

Definition 5. The *FDID research design* consists of (i) the FDID setting in Definition 1, including Assumption 1 (universal exposure); (ii) the identification results in Propositions 1–4, under which τ_{DID} identifies (a) τ_{em} given Assumptions 1–3, (b) τ_{att} given Assumptions 1–4, (c) τ_{cm} given Assumptions 1–3 and 5, and (d) $\tau_{G|Z=1}$ given Assumptions 1–3 and 5–6.

5 Extension to Conditionally Valid Assumptions

In many applications, the canonical and factorial parallel trends assumptions in Assumptions 3 and 5 are plausible only after conditioning on baseline covariates X_i in addition to G_i. In this section, we present the corresponding identification and estimation results and connect them to regression methods commonly used in applied research. The unconditional setting is a special case where $X_i = \emptyset$.

The main takeaways are twofold. First, all results in Section 4 extend to the conditional setting. Under suitable assumptions, the conditional DID estimand identifies the conditional effect modification and causal moderation, and averaging over covariates yields the marginal effects. Second, coherent with Remark 1, standard cross-sectional methods based on unconfoundedness (Rosenbaum and Rubin, 1983), such as outcome regression and inverse propensity score weighting, carry over to the FDID setting when $(G_i, \Delta Y_i, X)_{i=1}^n$ are treated as the data. These approaches resemble the covariate-adjustment methods used in canonical DID under the conditional parallel trends assumption (e.g., Roth et al., 2023). We focus here on stratification and outcome regression using linear and TWFE specifications, and provide details on inverse propensity score weighting in Section A2.6 in the Supplementary Materials.

5.1 Identification

Let X_i denote the vector of covariates beyond G_i, taking values in $\mathcal{X} \subseteq \mathbb{R}^p$. Assumption 7 states the overlap condition that ensures the conditional expectation $\mathbb{E}[\,\cdot \mid G_i = g, X_i]$ is well defined for $g = 0, 1$. We maintain this assumption throughout the rest of the paper.

Assumption 7 (Overlap)**.** For all $x \in \mathcal{X}$, $\mathbb{P}(G_i = 1 \mid X_i = x) \in (0, 1)$.

Under Assumption 7, define

$$\begin{aligned} \tau_{\text{DID}}(x) &= \mathbb{E}[\Delta Y_i \mid G_i = 1, X_i = x] - \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i = x], \\ \tau_{\text{em}}(x) &= \mathbb{E}[\tau_{i,Z|G=1} \mid G_i = 1, X_i = x] - \mathbb{E}[\tau_{i,Z|G=0} \mid G_i = 0, X_i = x], \\ \tau_{\text{cm}}(x) &= \mathbb{E}[\tau_{i,\text{cm}} \mid X_i = x] = \mathbb{E}[\tau_{i,Z|G=1} - \tau_{i,Z|G=0} \mid X_i = x] \end{aligned} \tag{4}$$

as the *conditional* DID estimand, effect modification, and causal moderation, respectively, generalizing $(\tau_{\text{DID}}, \tau_{\text{em}}, \tau_{\text{cm}})$ in (2) and Definition 2. Define their marginal counterparts as

$$\tau_{\text{DID-x}} = \mathbb{E}[\tau_{\text{DID}}(X_i)], \quad \tau_{\text{em-x}} = \mathbb{E}[\tau_{\text{em}}(X_i)], \quad \tau_{\text{cm}} = \mathbb{E}[\tau_{\text{cm}}(X_i)], \tag{5}$$

where the expectations are taken over the marginal distribution of X_i. Since $\tau_{\text{DID}}(X_i)$ and $\tau_{\text{em}}(X_i)$ compare ΔY_i and $\tau_{i,Z|G=g}$, respectively, between the two levels of G_i conditional on X_i, their marginal averages $(\tau_{\text{DID-x}}, \tau_{\text{em-x}})$ in (5) generally differ from $(\tau_{\text{DID}}, \tau_{\text{em}})$ unless G_i and X_i are independent. Assumptions 8–9 below build on Assumption 7, and extend the canonical and factorial parallel trends assumptions to the conditional setting.

Assumption 8 (Conditional canonical parallel trends)**.** For all $x \in \mathcal{X}$, $\mathbb{E}[\Delta Y_i(G_i, 0) \mid G_i = 1, X_i = x] = \mathbb{E}[\Delta Y_i(G_i, 0) \mid G_i = 0, X_i = x]$.

Assumption 9 (Conditional factorial parallel trends)**.** For all $x \in \mathcal{X}$, $\mathbb{E}[\Delta Y_i(g, z) \mid G_i = 1, X_i = x] = \mathbb{E}[\Delta Y_i(g, z) \mid G_i = 0, X_i = x]$ for $g, z = 0, 1$.

Echoing the discussion below Assumption 5, a sufficient condition for Assumption 9 is $G_i \perp\!\!\!\perp \{\Delta Y_i(g, z) : g, z = 0, 1\} \mid X_i$, which states that G_i is as-if randomly assigned with respect to changes in all four potential outcomes, conditional on covariates X_i. This condition is analogous to the unconfoundedness assumption for the cross-sectional data $(\Delta Y_i, G_i, Z_i, X_i)$. Like unconfoundedness, this assumption is inherently untestable; researchers can only assess its plausibility through auxiliary checks, such as placebo or sensitivity analyses (Imbens and Xu, 2025).

Corollary 1 extends Propositions 1 and 3, and establishes the identification of $\tau_{\text{em-x}}$ and τ_{cm}.

Corollary 1. (i) If Assumptions 1–2 and 7–8 hold, then $\tau_{\text{DID}}(x) = \tau_{\text{em}}(x)$ and $\tau_{\text{DID-x}} = \tau_{\text{em-x}}$. (ii) If Assumptions 1–2 and 7–9 hold, then $\tau_{\text{DID}}(x) = \tau_{\text{em}}(x) = \tau_{\text{cm}}(x)$ and $\tau_{\text{DID-x}} = \tau_{\text{em-x}} = \tau_{\text{cm}}$.

Corollary 1(i) follows from Proposition 1 and ensures that $\tau_{\text{DID-x}}$ identifies $\tau_{\text{em-x}}$ under Assumption 8 (conditional canonical parallel trends). Corollary 1(ii) follows from Proposition 3 and ensures that $\tau_{\text{DID-x}}$ identifies $\tau_{\text{em-x}} = \tau_{\text{cm}}$ under Assumptions 8–9 (conditional canonical and factorial parallel trends). Moreover, Corollary 1 shows that $\tau_{\text{DID}}(x)$ allows us to identify the expected values of $\tau_{\text{em}}(X_i)$ and $\tau_{\text{cm}}(X_i)$ under any distribution of X_i, not just the marginal one. In particular, we can recover group-specific expectations, $\mathbb{E}[\tau_{\text{em}}(X_i) \mid G_i = g]$ and $\mathbb{E}[\tau_{\text{cm}}(X_i) \mid G_i = g]$, by averaging $\tau_{\text{DID}}(X_i)$ over the conditional distribution of X_i given $G_i = g$, paralleling the ATT and group causal moderation $\mathbb{E}[\tau_{i,\text{cm}} \mid G_i = g]$ in Remark 2.

5.2 Estimation

To apply Corollary 1, we need to estimate $\tau_{\text{DID}}(x)$. From (4), it suffices to estimate $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$. Stratification and outcome regression are two approaches, suited to categorical and continuous X_i, respectively.

For categorical X_i with K levels indexed by $k = 1, \dots, K$, we can estimate $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = k]$ by the sample average of ΔY_i among units with $(G_i, X_i) = (g, k)$, denoted by $\widehat{\Delta Y}(g, k)$. The resulting estimator of $\tau_{\text{DID}}(k)$ is $\hat{\tau}_{\text{DID}}(k) = \widehat{\Delta Y}(1, k) - \widehat{\Delta Y}(0, k)$, as the *stratum-specific DID estimator* based on units with $X_i = k$. The marginal estimand $\tau_{\text{DID-x}}$ can be estimated following (5) as $\hat{\tau}_{\text{DID-x}} = n^{-1} \sum_{i=1}^{n} \hat{\tau}_{\text{DID}}(X_i) = \sum_{k=1}^{K} \pi_k \hat{\tau}_{\text{DID}}(k)$, where π_k is the sample proportion of units with $X_i = k$. This approach also applies when X_i can be meaningfully discretized.

For continuous X_i, we can estimate $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$ using regression, denoted by $\widehat{\Delta Y}(g, x)$, and obtain $\tau_{\text{DID}}(x)$ and $\tau_{\text{DID-x}}$ from (4)–(5) as

$$\hat{\tau}_{\text{DID}}(x) = \widehat{\Delta Y}(1, x) - \widehat{\Delta Y}(0, x), \quad \hat{\tau}_{\text{DID-x}} = n^{-1} \sum_{i=1}^{n} \hat{\tau}_{\text{DID}}(X_i). \tag{6}$$

Note that

$$\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x] = \mathbb{E}[Y_{i,\text{post}} \mid G_i = g, X_i = x] - \mathbb{E}[Y_{i,\text{pre}} \mid G_i = g, X_i = x],$$

where the two sides correspond to two common regression-based approaches to DID analysis. The left-hand side corresponds to cross-sectional linear regression of ΔY_i on (G_i, X_i), which directly estimates $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$. The right-hand side corresponds to TWFE regression of Y_{it} on (G_i, X_i) with unit and time fixed effects using *long-format* data, where each row corresponds to a unit-time observation. We discuss both approaches below.

5.2.1 Linear regression of ΔY_i

Definition 6 below presents two common specifications for estimating $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$ using cross-sectional regression of ΔY_i.

Definition 6. (i) Let OLS_* be the ordinary least squares (OLS) fit of the model

$$\Delta Y_i = \beta_1 + \beta_G G_i + \beta_X^\top X_i + \beta_{GX}^\top G_i X_i + \epsilon_i,$$

and denote the estimated coefficients by $(\hat{\beta}_{1,*}, \hat{\beta}_{G,*}, \hat{\beta}_{X,*}, \hat{\beta}_{GX,*})$.

(ii) Let OLS_+ be the OLS fit of the model

$$\Delta Y_i = \beta_1 + \beta_G G_i + \beta_X^\top X_i + \epsilon_i,$$

and denote the estimated coefficients by $(\hat{\beta}_{1,+}, \hat{\beta}_{G,+}, \hat{\beta}_{X,+})$.

OLS_+ is a restricted version of OLS_* excluding the interactions between G_i and X_i. Define

$$\widehat{\Delta Y_*}(g, x) = \hat{\beta}_{1,*} + \hat{\beta}_{G,*} g + \hat{\beta}_{X,*}^\top x + \hat{\beta}_{GX,*}^\top g x, \quad \widehat{\Delta Y_+}(g, x) = \hat{\beta}_{1,+} + \hat{\beta}_{G,+} g + \hat{\beta}_{X,+}^\top x$$

as the estimators of $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$ under OLS_* and OLS_+, respectively. Following (6), the DID estimators based on OLS_* and OLS_+ are:

$$\begin{aligned} \text{OLS}_*: \quad & \hat{\tau}_{\text{DID},*}(x) = \widehat{\Delta Y_*}(1, x) - \widehat{\Delta Y_*}(0, x) = \hat{\beta}_{G,*} + \hat{\beta}_{GX,*}^\top x, \\ & \hat{\tau}_{\text{DID-X},*} = n^{-1} \sum_{i=1}^{n} \hat{\tau}_{\text{DID},*}(X_i) = \hat{\beta}_{G,*} + \hat{\beta}_{GX,*}^\top \bar{X}, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \text{OLS}_+: \quad & \hat{\tau}_{\text{DID},+}(x) = \widehat{\Delta Y_+}(1, x) - \widehat{\Delta Y_+}(0, x) = \hat{\beta}_{G,+}, \\ & \hat{\tau}_{\text{DID-X},+} = n^{-1} \sum_{i=1}^{n} \hat{\tau}_{\text{DID},+}(X_i) = \hat{\beta}_{G,+}, \end{aligned} \tag{8}$$

respectively, where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ is the sample mean of X_i. Proposition 5 below establishes the consistency of these estimators for $\tau_{\text{DID}}(x)$ and $\tau_{\text{DID-X}}$ when the corresponding models in Definition 6 are correctly specified.

Proposition 5. (i) If $\mathbb{E}[\Delta Y_i \mid G_i, X_i] = \beta_1 + \beta_G G_i + \beta_X^\top X_i + \beta_{GX}^\top G_i X_i$ for some constant $(\beta_1, \beta_G, \beta_X, \beta_{GX})$, then (a) $\tau_{\text{DID}}(x) = \beta_G + \beta_{GX}^\top x$; $\tau_{\text{DID-X}} = \beta_G + \beta_{GX}^\top \mathbb{E}[X_i]$ and (b) $\hat{\tau}_{\text{DID},*}(x)$

and $\hat{\tau}_{\text{DID-x},*}$ in (7) based on OLS$_*$ are consistent for $\tau_{\text{DID}}(x)$ and $\tau_{\text{DID-x}}$.

(ii) If $\mathbb{E}[\Delta Y_i \mid G_i, X_i] = \beta_1 + \beta_G G_i + \beta_X^\top X_i$ for some constant $(\beta_1, \beta_G, \beta_X)$, then (a) $\tau_{\text{DID}}(x) = \tau_{\text{DID-x}} = \beta_G$ and (b) $\hat{\tau}_{\text{DID},+}(x) = \hat{\tau}_{\text{DID-x},+} = \hat{\beta}_{G,+}$ in (8) based on OLS$_+$ are consistent for $\tau_{\text{DID}}(x) = \tau_{\text{DID-x}}$.

Proposition 5 justifies the use of OLS$_*$ and OLS$_+$ for estimating $\tau_{\text{DID}}(x)$ and $\tau_{\text{DID-x}}$ under linearity assumptions on $\mathbb{E}[\Delta Y_i \mid G_i, X_i]$. The identification of $\{\tau_{\text{em}}(x), \tau_{\text{em-x}}\}$ under Assumption 8, and of $\{\tau_{\text{cm}}(x), \tau_{\text{cm}}\}$ under Assumptions 8–9, then follows from Corollary 1. Together, Proposition 5 and Corollary 1 justify the use of OLS$_+$ and OLS$_*$ for DID analysis of FDID when linearity holds. In particular, Proposition 5(i) implies that if $\mathbb{E}[X_i] = 0$, then $\beta_G = \tau_{\text{DID-x}}$, so the coefficient of G_i from OLS$_*$, $\hat{\beta}_{G,*}$, has a direct causal interpretation as a consistent estimator of $\tau_{\text{em-x}}$ and τ_{cm} under the corresponding assumptions. The sample version of this condition can be ensured by centering the covariates so that $\bar{X} = 0$, as in Hirano and Imbens (2001) and Lin (2013). To account for the uncertainty in $\hat{\beta}_{G,*}$, one can implement a unit-level cluster bootstrap procedure (Bertrand et al., 2004). A subtlety is that the bootstrap samples must be generated using the original X_i's, which are then recentered within each bootstrap replication. Using pre-centered covariates for resampling yields invalid inference because it ignores the sampling variability in $\bar{X}$.

By contrast, Proposition 5(ii) shows that the more parsimonious OLS$_+$ may be inconsistent for $\{\tau_{\text{DID}}(x), \tau_{\text{DID-x}}\}$ if the effect of G_i on ΔY_i varies with X_i. However, when $\mathbb{E}[\Delta Y_i \mid G_i, X_i]$ is truly linear in (G_i, X_i), then $\tau_{\text{DID}}(x) = \tau_{\text{DID-x}}$, and the coefficient $\hat{\beta}_{G,+}$ from OLS$_+$ is consistent for their common value without requiring covariate centering.

5.2.2 Two-way fixed-effects regression of Y_{it}

Definition 7 presents two common TWFE specifications for estimating $\mathbb{E}[Y_{it} \mid G_i = g, X_i = x]$, where $t = \text{pre}, \text{post}$. Let $1_{\{t=\text{post}\}}$ denote an indicator for the post-period.

Definition 7. (i) Let TWFE_* be the OLS fit of the TWFE model

$$Y_{it} = b_G G_i \cdot 1_{\{t=\text{post}\}} + b_X X_i \cdot 1_{\{t=\text{post}\}} + b_{GX} G_i X_i \cdot 1_{\{t=\text{post}\}} + \alpha_i + \xi_t + \epsilon_{it},$$

where α_i and ξ_t are unit and time fixed effects, respectively, and ϵ_{it} is an idiosyncratic error.

(ii) Let TWFE_+ be the OLS fit of the TWFE model

$$Y_{it} = b_G G_i \cdot 1_{\{t=\text{post}\}} + b_X X_i \cdot 1_{\{t=\text{post}\}} + \alpha_i + \xi_t + \epsilon_{it}.$$

TWFE_+ is a restricted version of TWFE_* that excludes the interactions among G_i, X_i, and $1_{\{t=\text{post}\}}$. Standard OLS theory ensures that

(i) the estimated coefficients of $(G_i \cdot 1_{\{t=\text{post}\}}, X_i \cdot 1_{\{t=\text{post}\}}, G_i X_i \cdot 1_{\{t=\text{post}\}})$ from TWFE_* equal those of $(G_i, X_i, G_i X_i)$ from OLS_*;

(ii) the estimated coefficients of $(G_i \cdot 1_{\{t=\text{post}\}}, X_i \cdot 1_{\{t=\text{post}\}})$ from TWFE_+ equal those of (G_i, X_i) from OLS_+.

Thus, causal interpretation of TWFE_* and TWFE_+ follows directly from Proposition 5. As in OLS_*, when using TWFE_* for FDID analysis, it is important to center the covariates so that the coefficient on $G_i \cdot 1_{\{t=\text{post}\}}$ has a standalone causal interpretation.

6 Other Extensions

In this section, we generalize our main theoretical results to accommodate repeated cross-sectional data and to allow for a general (such as discrete or continuous) baseline factor G.

6.1 Repeated Cross-Sectional Data

So far, our discussion has focused on analyses based on panel data. In many applications, however, researchers only have access to repeated cross-sectional data, where a new set of units is sampled at each time point. We extend our framework to this setting, and unify the panel and repeated cross-sections as two sampling schemes for the same population.

6.1.1 Unification from a sampling perspective

Previously, i indexed observed units drawn from a common population. With slight abuse of notation, we now also let i denote a generic random unit from this population, which may or may not belong to the study sample. Our theory ensures that the conditional DID estimand $\tau_{\text{DID}}(x)$ identifies $\tau_{\text{em}}(x)$ and $\tau_{\text{cm}}(x)$ under suitable assumptions. Panel and repeated cross-sectional data then correspond to two sampling schemes: in panel data, the same units are re-observed over time, while in repeated cross-sectional data, new units are sampled at each time. Recall from (4) that

$$\tau_{\text{DID}}(x) = \mathbb{E}[\Delta Y_i \mid G_i = 1, X_i = x] - \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i = x],$$

where $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x] = \mathbb{E}[Y_{i,\text{post}} \mid G_i = g, X_i = x] - \mathbb{E}[Y_{i,\text{pre}} \mid G_i = g, X_i = x]$. With panel data, ΔY_i is observed directly, so both $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$ and $\mathbb{E}[Y_{it} \mid G_i = g, X_i = x]$ can be estimated. With repeated cross-sectional data, ΔY_i is not observed at the unit level, so estimation proceeds via $\mathbb{E}[Y_{it} \mid G_i = g, X_i = x]$. In either case, stratification and TWFE regression, as discussed in Section 5.2, are feasible, suitable for discrete X_i and continuous X_i, respectively. However, outcome regression based on the OLS specifications using ΔY_i is only applicable in the panel setting.

6.1.2 Nested sampling with subunits and analysis

A distinctive feature of many FDID applications (e.g., Fouka, 2019) using repeated cross-sectional data is nested sampling: the baseline factor G is defined at the regional level, while outcomes are measured at the individual level. For clarity, we refer to the higher-level units as regions and the lower-level (sub)units as individuals. In practice, "individual" may also denote other sub-regional entities such as firms, schools, households, or smaller administrative units. Three sampling designs are common in multi-period studies under nested sampling:

(i) *Panel–Panel*: Both regions and individuals are sampled once and followed over time, yielding panel data at both levels.

(ii) *Panel–Repeated Cross-section*: Regions are sampled once and followed over time, while a new set of sub-regional individuals is sampled within each region at each time point. This yields panel data at the regional level and repeated cross-sections at the individual level.

(iii) *Repeated Cross-section–Repeated Cross-section*: A new set of regions, and hence individuals, is sampled at each time point, yielding repeated cross-sections at both levels.

When individual-level data are available, researchers can conduct analysis to estimate $\tau_{\text{DID}}(x)$ either at the individual level (using individual-level data) or at the regional level (by aggregating individual level data to the regional level). When individual-level data are unavailable, methods for region-level analyses also apply.

(i) *Panel–Panel*: For region-level analysis, estimation reduces to the panel-data methods discussed in Section 5.2, treating regions as units. For individual-level analysis, the same methods apply with the unit index i replaced by the subunit index ij and $G_{ij} = G_i$. Details are provided in Section A2.4 in the Supplementary Materials.

(ii) *Panel–Repeated Cross-section*: For region-level analysis, the data remain panel, so estimation again reduces to methods discussed in Section 5.2, treating regions as units. For individual-level analysis, stratification and TWFE regression remain valid, with standard errors clustered at the regional level. A practical complication is that covariates may lack common support across time, which we leave for future research.

(iii) *Repeated Cross-section–Repeated Cross-section*: For both region-level and individual-level analyses, stratification and TWFE regression remain applicable, with standard errors clustered at the regional level. The same covariate overlap issue arises here and remains open for future work.

In summary, FDID extends naturally to repeated cross-sectional data, with panel and repeated cross-sections viewed as alternative sampling schemes from the same population. In FDID with nested data structures, across the three common nested sampling schemes, stratification and TWFE regression remain applicable for both individual- and region-level analyses, with standard errors clustered by region.

6.2 General Baseline Factor G

The discussion so far has assumed a binary baseline factor G. When G takes values in a general set $\mathcal{G} \subseteq \mathbb{R}$, all results from the binary case continue to hold for comparisons between any two levels $g, g' \in \mathcal{G}$.

Renew $Y_{it}(g, z)$, $\Delta Y_i(g, z)$, and $\tau_{i,Z|G=g} = Y_{i,\text{post}}(g, 1) - Y_{i,\text{post}}(g, 0)$ for general $g \in \mathcal{G}$. For $g, g' \in \mathcal{G}$, define $\tau_{i,\text{cm},g\to g'} = \tau_{i,Z|G=g'} - \tau_{i,Z|G=g}$ as the causal moderation of G when its level changes from g to g', generalizing $\tau_{i,\text{cm}}$. Define

$$\begin{aligned}
\tau_{\text{DID},g\to g'}(x) &= \mathbb{E}[\Delta Y_i \mid G_i = g', X_i = x] - \mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x],\\
\tau_{\text{em},g\to g'}(x) &= \mathbb{E}[\tau_{i,Z|G=g'} \mid G_i = g', X_i = x] - \mathbb{E}[\tau_{i,Z|G=g} \mid G_i = g, X_i = x],\\
\tau_{\text{cm},g\to g'}(x) &= \mathbb{E}[\tau_{i,\text{cm},g\to g'} \mid X_i = x] = \mathbb{E}[\tau_{i,Z|G=g'} - \tau_{i,Z|G=g} \mid X_i = x]
\end{aligned}$$

as the conditional DID estimand, effect modification, and causal moderation when G changes from g to g', extending $\{\tau_{\text{DID}}(x), \tau_{\text{em}}(x), \tau_{\text{cm}}(x)\}$ in (4). The marginal counterparts are $\tau_{\text{DID-x},g\to g'} = \mathbb{E}[\tau_{\text{DID},g\to g'}(X_i)]$, $\tau_{\text{em-x},g\to g'} = \mathbb{E}[\tau_{\text{em},g\to g'}(X_i)]$, and $\tau_{\text{cm},g\to g'} = \mathbb{E}[\tau_{\text{cm},g\to g'}(X_i)] = \mathbb{E}[\tau_{i,\text{cm},g\to g'}]$, extending $\tau_{\text{cm}}, \tau_{\text{DID-x}}, \tau_{\text{em-x}}$ in Definitions 2 and (5).

Identification. All identification results in Section 5.1 extend to general G. Parallel to Corollary 1, the conditional DID estimand $\tau_{\text{DID},g\to g'}(x)$ (i) identifies $\tau_{\text{em},g\to g'}(x)$ under the generalized conditional canonical parallel trends assumptions, and (ii) identifies

$\tau_{\text{cm},g\to g'}(x) = \tau_{\text{em},g\to g'}(x)$ under the generalized conditional canonical and factorial parallel trends assumptions. We provide the details in Section A2.5 in the Supplementary Materials.

Regression estimation. Renew (OLS_*, OLS_+) and (TWFE_*, TWFE_+) in Definitions 6–7 with general G_i. The numerical equivalence between OLS and TWFE continues to hold for general G_i. Parallel to Proposition 5, OLS_* and OLS_+ are consistent for estimating $\tau_{\text{DID},g\to g'}(x)$ and $\tau_{\text{DID-X},g\to g'}$ when the linear models are correctly specified.

With continuous G, we can also study incremental effects in the sense of Rothenhäusler and Yu (2019). For brevity, we relegate the details to Section A2.5 of the Supplementary Materials.

7 Empirical Application

We now reanalyze the data from our running example (Cao et al., 2022) to illustrate our theory. Each observation represents one county in the sample. The outcome of interest is the county-level annual mortality rate, measured as the number of deaths per thousand people. Following Cao et al. (2022), the baseline factor takes two forms: (i) a binary indicator for high social capital, equal to one if the per capita number of genealogy books is at or above the sample median and zero otherwise; and (ii) the logarithm of the per capita number of genealogy books plus one, a continuous measure. The Great Famine began in late 1958 and, according to most historians, ended in 1961. The set of pre-famine covariates, measured at the county level, include per capita grain production, ratio of non-farming land, urbanization ratio, distance from Beijing, distance from the provincial capital, share of ethnic minorities, suitability for rice cultivation, average years of education, and log population size.

We apply three estimators, DID, OLS_* (with interactions), and OLS_+ (without interactions), with the latter two incorporate covariates. As discussed earlier, in the FDID panel setting,

OLS_* and OLS_+ are numerically equivalent to TWFE_* and TWFE_+, respectively. We use 1957, one year before the famine, as the reference year for the pre-period, with $Y_{i,\text{pre}}$ representing the mortality rate in county i in 1957. For the post-period, we define three time windows: (i) the famine years from 1958 to 1961, (ii) the pre-famine years from 1954 to 1957, and (iii) the post-famine years from 1962 to 1966. The second time window, 1954–1957, precedes the famine and is used to construct a placebo test for the (conditional) canonical parallel trends assumption, similar to a pretrend test in canonical DID (Angrist and Pischke, 2009). Within each time window, we calculate the average mortality rate for each county, which serves as $Y_{i,\text{post}}$.

Cao et al. (2022) aim to estimate "the effect of social capital on famine relief," which we interpret as the causal moderation of social capital on the effect of exposure to famine. By Proposition 1, if Assumptions 1–3 (universal exposure, no anticipation, and canonical parallel trends) hold, then $\hat{\tau}_{\text{DID}}$ recovers effect modification, τ_{em}. If Assumption 5 (factorial parallel trends) also holds, then $\hat{\tau}_{\text{DID}}$ recovers causal moderation, τ_{cm}. Similarly, with covariates, Corollary 1 and Proposition 5 imply that under the conditional canonical parallel trends (Assumption 8), $\hat{\beta}_{G,*}$ and $\hat{\beta}_{G,+}$ recover effect modification, $\tau_{\text{em-x}}$, under their respective model specifications. If, in addition, the conditional factorial parallel trends (Assumption 9) holds, then they identify causal moderation, τ_{cm}.

Table 1 presents the results, where Panels A and B report estimates using binary and continuous measures of G, respectively. In each panel, Column (1) reports DID estimates without covariates, while columns (2) and (3) report estimates from OLS_* and OLS_+, respectively, with covariates included. Each row corresponds to a different comparison period relative to the reference year 1957. In Panel A, the first row shows that high social capital (relative to low social capital) are associated with a reduction the famine-induced increase in mortality by more than 2.9 deaths per 1,000 people per year, or nearly 12 fewer

Table 1: Estimating Causal Moderation of Social Capital on the Impact of Famine

Panel A: Binary G	DID ($\hat{\tau}_{\text{DID}}$)	OLS$_*$ ($\hat{\beta}_{G,*}$)	OLS$_+$ ($\hat{\beta}_{G,+}$)
Famine Years (1958–1961)	-2.32	-2.92	-2.79
	$[-3.75, -0.83]$	$[-4.29, -1.39]$	$[-4.12, -1.31]$
Pre-Famine Years (1954–1956)	0.32	0.35	0.33
	$[-0.10, 0.72]$	$[-0.04, 0.78]$	$[-0.06, 0.75]$
After Famine Ends (1962–1966)	-0.81	-0.51	-0.49
	$[-1.20, -0.40]$	$[-0.90, -0.11]$	$[-0.86, -0.11]$
Panel B: Continuous G	DID ($\hat{\tau}_{\text{DID}}$)	OLS$_*$ ($\hat{\beta}_{G,*}$)	OLS$_+$ ($\hat{\beta}_{G,+}$)
Famine Years (1958–1961)	-5.85	-5.14	-10.11
	$[-7.74, -4.01]$	$[-8.96, -0.57]$	$[-13.06, -6.96]$
Pre-Famine Years (1954–1956)	1.02	-0.51	0.69
	$[-0.04, 1.99]$	$[-2.03, 0.87]$	$[-0.41, 1.72]$
After Famine Ends (1962–1966)	-1.82	-1.35	-1.84
	$[-2.73, -0.99]$	$[-2.43, -0.26]$	$[-2.67, -0.97]$

Note: In Panel A, G is measured as a binary indicator of social capital, while in Panel B, G is measured as a continuous variable. The first column reports DID estimates without covariates. The second and third columns report regression estimates ($\hat{\beta}_{G,*}, \hat{\beta}_{G,+}$) from OLS$_*$ and OLS$_+$, respectively, which incorporate covariates. They equal estimates of $b_{G,*}$ and $b_{G,+}$ from TWFE$_*$ and TWFE$_+$, respectively. The pre-period reference year is 1957. Brackets report 95% bootstrap confidence intervals using the percentile method.

deaths per 1,000 people over the four-year period. The three estimators produce similar results and closely match those reported in the original paper using TWFE models. The estimates using the average mortality rate from 1954–1956 as $Y_{i,\text{post}}$ (second row) are close to zero, which support the canonical parallel trends assumption. The estimates using the average mortality rate from 1962–1966 as $Y_{i,\text{post}}$ (third row) are negative—though much smaller in magnitude than during the famine years—and statistically significant at the 5% level, suggesting a small but lasting effect of the famine on mortality rates after it ended.

To gain a better understanding on how the estimates evolves over time, we estimate it separately for each year from 1954 to 1966, excluding 1957 (the pre-period), defining $Y_{i,\text{post}}$ as

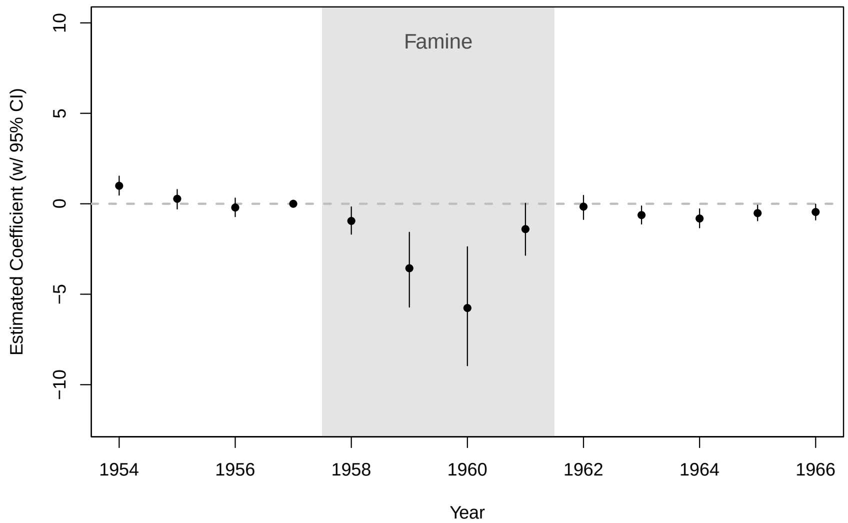

Figure 5: Estimated causal moderation over time. Each estimate is obtained from fitting OLS_* using data in the year of interest and 1957. G is a binary indicator of social capital.

the mortality rate for each year. Figure 5 shows the results. Coherent with findings in Panel A of Table 1, the estimates are close to zero in the pre-famine years, negative during the famine (with particularly large negative estimates in 1959 and 1960), and remain negative but with much smaller magnitudes after the famine ended. The pre-1957 estimates are close to zero, providing suggestive evidence on the no-anticipation and (conditional) canonical parallel trends assumptions, under which the first-row estimates are effect modification. Interpreting the results as causal moderation requires the stronger (conditional) factorial parallel trends assumption, which rules out any potential confounder correlated with both social capital and the dynamics of mortality. For example, if income or governance quality are correlated with both social capital and trends in mortality, this assumption would be violated. In Section A3 in the Supplementary Materials, we present sensitivity analysis showing that a confounder with correlations to social capital and mortality changes comparable to observed covariates would have only a small impact on the estimated moderation effect. Nevertheless, we emphasize that this causal interpretation rests on a strong and inherently untestable assumption.

The patterns in Panel B mirror those in Panel A: social capital is negatively correlated with famine-induced mortality increases, but not in years leading up to the famine. For example, with $\hat{\beta}_{G,*} = 5.15$, if linearity and constant effect assumptions hold, then a 10% increase in the per capita number of genealogy books corresponds to a reduction of 0.5 deaths per 1,000 people per year, or about 2 fewer deaths per 1,000 people over the four-year period. Overall, the evidence supports the original authors' claim that social capital is associated with a mitigated impact of the famine. Whether this association can be interpreted as causal, however, depends on how plausible the (conditional) factorial parallel trends assumption is.

8 Recommendations for Applied Researchers

Based on our results, we offer several recommendations for applied researchers. First, it is important to distinguish between an estimator and a research design. An estimator is merely an algorithm applied to observed data, while a research design specifies not only the estimator but also identifying assumptions that connect it to meaningful estimands. Researchers should clearly communicate each element of the research design, especially the target estimands (Lundberg et al., 2021). In particular, the use of the DID estimator should not be conflated with the canonical DID research design.

Second, researchers should be cautious when interpreting DID estimates as causal in FDID. Unless the exclusion restriction that exposure Z has no effect on one of the two groups is plausible, the DID estimator under the no anticipation and canonical parallel trends assumptions identifies only effect modification of G, not the causal effects of Z or G. To recover causal moderation by G, we propose a factorial parallel trends assumption, which requires mean independence between G_i and $\Delta Y_i(g, z)$ given X_i. Intuitively, it eliminates any unobserved confounders that are correlated with both G and the outcome dynamics. Like unconfoundedness in the cross-sectional setting, this is a strong and untestable assumption.

To identify G's conditional effect given exposure, a quantity many studies aim to estimate, an additional exclusion restriction is required, stating that G has no effect on the outcome when $Z = 0$. Table 2 summarizes these results and illustrates the interpretation of each estimand using the running example.

Table 2: Estimands, Interpretations, and Identifying Assumptions under FDID

Estimand	**Interpretation in the running example**	**Key identifying assumptions needed**
τ_{em}, effect modification	How did the average effect of exposure to famine on mortality differ between counties with high and low levels of social capital?	No anticipation, parallel trends
τ_{att}, the average effect of exposure for group $G = 1$	What was the average effect of exposure to famine on mortality in counties with high social capital?	No anticipation, parallel trends, exclusion restriction on Z in group $G = 1$
τ_{cm}, causal moderation	To what extent did high (vs. low) social capital mitigate the impact of exposure to famine on mortality?	No anticipation, parallel trends, factorial parallel trends
$\tau_{G\|Z=1}$, G's conditional effect given exposure	What was the average causal effect of social capital on mortality during the famine years?	No anticipation, parallel trends, factorial parallel trends, exclusion restriction on G absent the event

Third, for estimation, in the absence of covariates, a TWFE regression including the interaction between G and $1_{\{t=\text{post}\}}$ is numerically equivalent to the DID estimator. With additional covariates X_i, standard cross-sectional methods based on unconfoundedness apply to $(\Delta Y_i, G_i, X_i)$, where ΔY_i is the before-after difference of the outcome. Outcome regression, such as linear regression of ΔY_i on $(1, G_i, X_i, G_i X_i)$, is commonly used when X_i is continuous and is numerically equivalent to a TWFE regression with the interaction $G_i X_i \cdot 1_{\{t=\text{post}\}}$. In both approaches, it is important to center X_i to ensure that key regression coefficients admit a standalone causal interpretation. We discuss alternative estimation strategies in Section A2.6 of the Supplementary Materials and provide software support for their implementation through the `fdid` package in `R`.

Conflict of Interest Statement

The authors claim no conflicts of interest.

Acknowledgment

We thank Anran Liu and Rivka Lipkovitz for excellent research assistance. Peng Ding acknowledges support from the U.S. National Science Foundation (grants # 1945136 and # 2514234). We are grateful to Justin Grimmer, Erin Hartman, Jens Hainmueller, Laura Hatfield, Dennis Shen, Eric Tchetgen Tchetgen, Ye Wang, and seminar participants at Berkeley, MIT, Princeton, UW–Madison, UCSD, UCLA, Yale, PolMeth 2024, ACIC 2025, and the Stanford Online Causal Inference Seminar, as well as three anonymous reviewers and the Editor Hongtu Zhu, for their valuable comments. We also thank Jairui Cao and Chuanchuan Zhang for sharing the data used in Cao et al. (2022).

References

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press.

Arkhangelsky, D. and G. Imbens (2024). Causal models for longitudinal and panel data: A survey. *The Econometrics Journal 27*(3), C1–C61.

Bansak, K. (2020). Estimating causal moderation effects with randomized treatments and non-randomized moderators. *Journal of the Royal Statistical Society Series A: Statistics in Society 184*, 65–86.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics 119*, 249–275.

Cao, J., Y. Xu, and C. Zhang (2022, June). Clans and calamity: How social capital saved lives during China's Great Famine. *Journal of Development Economics 157*, 102865.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review 84*, 772–793.

Charnysh, V. (2022). Explaining out-group bias in weak states: Religion and legibility in the 1891/1892 Russian famine. *World Politics 74*, 205–248.

Chen, H., Z. Chen, Z. He, J. Liu, and R. Xie (2023). Pledgeability and asset prices: Evidence from the chinese corporate bond markets. *The Journal of Finance 78*, 2563–2620.

Chen, J., E. H. Wang, and X. Zhang (2025). From powerholders to stakeholders: State-building with elite compensation in early medieval China. *American Journal of Political Science 69*(2), 607–623.

Chiu, A., X. Lan, Z. Liu, and Y. Xu (2026). Causal panel analysis under parallel trends: lessons from a large reanalysis study. *American Political Science Review 120*(1), 245–266.

Cinelli, C. and C. Hazlett (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 39–67.

Dasgupta, T., N. Pillai, and D. B. Rubin (2015). Causal inference from 2^K factorial designs by using potential outcomes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 77*, 727–753.

Fouka, V. (2019). How do immigrants respond to discrimination? The case of Germans in the US during World War I. *The American Political Science Review 113*, 405–422.

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics 58*, 21–29.

Han, S. and D. B. Rubin (2021). Contrast-specific propensity scores. *Biostatistics & Epidemiology 5*, 1–8.

Hirano, K. and G. W. Imbens (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology 2*, 259–278.

Holland, P. W. and D. B. Rubin (1986). Research designs and causal inferences: On Lord's paradox. In *Survey Research Designs: Towards a Better Understanding of Their Costs and Benefits: Prepared under the Auspices of the Working Group on the Comparative Evaluation of Longitudinal Surveys Social Science Research Council*, pp. 7–37. Springer.

Imbens, G. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Imbens, G. W. and Y. Xu (2025). Comparing experimental and nonexperimental methods: What lessons have we learned four decades after lalonde (1986)? *Journal of Economic Perspectives 39*(4), 173–201.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics 7*, 295 – 318.

Lundberg, I., R. Johnson, and B. M. Stewart (2021). What is your estimand? defining the target quantity connects statistical evidence to theory. *American Sociological Review 86*(3), 532–565.

Pashley, N. E. and M.-A. C. Bind (2023). Causal inference for multiple treatments using fractional factorial designs. *Canadian Journal of Statistics 51*, 444–468.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

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

Rothenhäusler, D. and B. Yu (2019). Incremental causal effects. *arXiv preprint arXiv:1907.13258*.

Squicciarini, M. P. (2020). Devotion and development: Religiosity, education, and economic progress in Nineteenth-Century France. *The American Economic Review 110*(11), 3454–3491.

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology 20*, 863–871.

Yu, R. and P. Ding (2026). Balancing weights for causal inference in observational factorial studies. *Journal of the American Statistical Association* (in press).

Zhao, A. and P. Ding (2021). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika 109*, 799–815.

Factorial Difference-in-Differences
Supplementary Materials

A1 Proofs

We occasionally use the following abbreviations to simplify the presentation:

Assumption 1 (Universal exposure)	UE
Assumption 2 (No anticipation)	NA
Assumption 3 (Canonical parallel trends)	PT
Assumption 5 (Factorial parallel trends)	FPT

A1.1 A Lemma

Lemma A1 below gives an equivalent form of τ_{em} under Assumptions 2 (no anticipation) and 3 (parallel trends) that we will use in multiple proofs.

Lemma A1. Under Assumptions 2–3, we have

$$\tau_{\text{em}} = \mathbb{E}[\Delta Y_i(1,1) \mid G_i = 1] - \mathbb{E}[\Delta Y_i(0,1) \mid G_i = 0].$$

Proof of Lemma A1. Assumption 3 (canonical parallel trends) ensures

$$\mathbb{E}[\Delta Y_i(1,0) \mid G_i = 1] = \mathbb{E}[\Delta Y_i(0,0) \mid G_i = 0]. \tag{A1}$$

Therefore, we have

$$\begin{aligned}
\tau_{\text{em}} &= \mathbb{E}[\tau_{i,Z|G_i=1} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G_i=0} \mid G_i = 0] \\
&\overset{\text{(A1)}}{=} \mathbb{E}[\tau_{i,Z|G_i=1} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G_i=0} \mid G_i = 0] \\
&\quad + \mathbb{E}[\Delta Y_i(1,0) \mid G_i = 1] - \mathbb{E}[\Delta Y_i(0,0) \mid G_i = 0] \\
&= \mathbb{E}\Big[\tau_{i,Z|G_i=1} + \Delta Y_i(1,0) \mid G_i = 1\Big] - \mathbb{E}\Big[\tau_{i,Z|G_i=0} + \Delta Y_i(0,0) \mid G_i = 0\Big] \\
&= A_1 - A_0,
\end{aligned} \tag{A2}$$

where

$$A_1 = \mathbb{E}\Big[\tau_{i,Z|G_i=1} + \Delta Y_i(1,0) \mid G_i = 1\Big], \quad A_0 = \mathbb{E}\Big[\tau_{i,Z|G_i=0} + \Delta Y_i(0,0) \mid G_i = 0\Big]. \tag{A3}$$

Note that

$$\begin{aligned}
\tau_{i,Z|G_i=1} + \Delta Y_i(1,0) &= Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) + Y_{i,\text{post}}(1,0) - Y_{i,\text{pre}}(1,0) \\
&= Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,0)
\end{aligned} \tag{A4}$$

in the definition of A_1 in (A3). Under Assumption 2 (no anticipation), plugging (A4) in the expression of A_1 in (A3) ensures

$$\begin{aligned} A_1 &\overset{\text{(A3)+(A4)}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,0) \mid G_i = 1] \\ &\overset{\text{Assm. 2 (NA)}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1) \mid G_i = 1] \\ &= \mathbb{E}[\Delta Y_i(1,1) \mid G_i = 1], \\ A_0 &\overset{\text{by symmetry}}{=} \mathbb{E}[\Delta Y_i(0,1) \mid G_i = 0]. \end{aligned} \tag{A5}$$

Plugging (A5) in (A2) completes the proof. □

A1.2 Proofs of the results in Section 4

Proof of Proposition 1. Assumption 1 ensures

$$\Delta Y_i = \begin{cases} \Delta Y_i(1,1) & \text{if } G_i = 1 \\ \Delta Y_i(0,1) & \text{if } G_i = 0 \end{cases}$$

such that

$$\begin{aligned} \tau_{\text{DID}} &= \mathbb{E}[\Delta Y_i \mid G_i = 1] - \mathbb{E}[\Delta Y_i \mid G_i = 0] \\ &= \mathbb{E}[\Delta Y_i(1,1) \mid G_i = 1] - \mathbb{E}[\Delta Y_i(0,1) \mid G_i = 0]. \end{aligned}$$

In addition, Lemma A1 ensures

$$\tau_{\text{em}} = \mathbb{E}[\Delta Y_i(1,1) \mid G_i = 1] - \mathbb{E}[\Delta Y_i(0,1) \mid G_i = 0]$$

under Assumptions 2 (no anticipation) and 3 (canonical parallel trends). These two results together ensure $\tau_{\text{DID}} = \tau_{\text{em}}$ under Assumptions 1–3. □

Proof of Proposition 3. We verify below $\tau_{\text{em}} = \tau_{\text{cm}}$ under Assumptions 2–3 and 5. Their identification by τ_{DID} then follows from Proposition 1.

First, Assumptions 3–5 (canonical and factorial parallel trends) together ensure

$$\begin{aligned} & \mathbb{E}[Y_{i,\text{post}}(1,0) - Y_{i,\text{pre}}(1,0)] - \mathbb{E}[Y_{i,\text{post}}(0,0) - Y_{i,\text{pre}}(0,0)] \\ = \;& \mathbb{E}[\Delta Y_i(1,0)] - \mathbb{E}[\Delta Y_i(0,0)] \\ \overset{\text{Assm. 5 (FPT)}}{=} \;& \mathbb{E}\left[\Delta Y_i(1,0) \mid G_i = 1\right] - \mathbb{E}\left[\Delta Y_i(0,0) \mid G_i = 0\right] \\ \overset{\text{Assm. 3 (PT)}}{=} \;& 0. \end{aligned} \tag{A6}$$

Equation (A6) implies

$$\mathbb{E}[Y_{i,\text{post}}(1,0) - Y_{i,\text{post}}(0,0)] = \mathbb{E}[Y_{i,\text{pre}}(1,0) - Y_{i,\text{pre}}(0,0)] \tag{A7}$$

and, together with Lemma A1, ensures

$$\begin{aligned}
\tau_{\text{em}} &\overset{\text{Lemma A1}}{=} \mathbb{E}[\Delta Y_i(1,1) \mid G_i = 1] - \mathbb{E}[\Delta Y_i(0,1) \mid G_i = 0] \\
&\overset{\text{Assm. 5 (FPT)}}{=} \mathbb{E}[\Delta Y_i(1,1)] - \mathbb{E}[\Delta Y_i(0,1)] \\
&= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1)] - \mathbb{E}[Y_{i,\text{post}}(0,1) - Y_{i,\text{pre}}(0,1)] \\
&= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(0,1)] - \mathbb{E}[Y_{i,\text{pre}}(1,1) - Y_{i,\text{pre}}(0,1)] \\
&\overset{\text{Assm. 2 (NA)}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(0,1)] - \mathbb{E}[Y_{i,\text{pre}}(1,0) - Y_{i,\text{pre}}(0,0)] \\
&\overset{\text{(A7)}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(0,1)] - \mathbb{E}[Y_{i,\text{post}}(1,0) - Y_{i,\text{post}}(0,0)] \\
&= \tau_{\text{cm}}.
\end{aligned}$$

□

A2 Additional Theoretical Results

A2.1 ATE Analogs under Assumption 5

In the paper, we define τ_{att}, the ATT analog under FDID, as

$$\begin{aligned}\tau_{\text{att}} &= \mathbb{E}[\tau_{i,Z|G} \mid G_i = 1] \\ &= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 1].\end{aligned}$$

Lemma A2 shows its relationship with the conditional average causal effect under the no anticipation and factorial parallel trends assumptions.

Lemma A2. Under Assumptions 2 and 5, we have

$$\begin{aligned}\tau_{\text{att}} &= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 0] \\ &= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0)].\end{aligned}$$

Proof. We have

$$\begin{aligned}\tau_{\text{att}} &\overset{\text{Def.}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 1] \\ &= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1) \mid G_i = 1] + \mathbb{E}[Y_{i,\text{pre}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 1] \\ &\overset{\text{NA}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1) \mid G_i = 1] + \mathbb{E}[Y_{i,\text{pre}}(1,0) - Y_{i,\text{post}}(1,0) \mid G_i = 1] \\ &\overset{\text{FPT}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1) \mid G_i = 0] + \mathbb{E}[Y_{i,\text{pre}}(1,0) - Y_{i,\text{post}}(1,0) \mid G_i = 0] \\ &\overset{\text{NA}}{=} \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{pre}}(1,1) \mid G_i = 0] + \mathbb{E}[Y_{i,\text{pre}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 0] \\ &= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 0],\end{aligned}$$

which is the first identity. Therefore, we also have

$$\tau_{\text{att}} = \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0)],$$

which is the second identity. □

However, in the paper, we did not state that Assumption 5 implies ATT = ATE to avoid confusion because some readers might interpret $\mathbb{E}[Y_{i,\text{post}}(G_i,1) - Y_{i,\text{post}}(G_i,0)]$ (the average treatment effect of exposure for all units, regardless of their realized value of G_i) as the ATE. But

$$\tau_{\text{att}} \neq \mathbb{E}[Y_{i,\text{post}}(G_i,1) - Y_{i,\text{post}}(G_i,0)],$$

because

$$\begin{aligned}\mathbb{E}[Y_{i,\text{post}}(G_i,1)-Y_{i,\text{post}}(G_i,0)] &= \mathbb{E}[Y_{i,\text{post}}(0,1)-Y_{i,\text{post}}(0,0)\mid G_i=0]\cdot\mathbb{P}(G_i=0)\\ &\quad +\mathbb{E}[Y_{i,\text{post}}(1,1)-Y_{i,\text{post}}(1,0)\mid G_i=1]\cdot\mathbb{P}(G_i=1).\end{aligned}$$

We can construct a simple contraction, for example, by assuming Assumption 4 (the exclusion restriction) holds. Under the exclusion restriction, $\mathbb{E}[Y_{i,\text{post}}(0,1)-Y_{i,\text{post}}(0,0)\mid G_i=0]=0$, then

$$\begin{aligned}\mathbb{E}[Y_{i,\text{post}}(G_i,1)-Y_{i,\text{post}}(G_i,0)] &= \mathbb{E}[Y_{i,\text{post}}(1,1)-Y_{i,\text{post}}(1,0)\mid G_i=1]\cdot\mathbb{P}(G_i=1)\\ &= \mathbb{E}[Y_{i,\text{post}}(1,1)-Y_{i,\text{post}}(1,0)]\cdot\mathbb{P}(G_i=1).\\ &= \tau_{\text{att}}\cdot\mathbb{P}(G_i=1)\neq\tau_{\text{att}}.\end{aligned}$$

Therefore, we avoid stating that, under Assumption 5, ATT = ATE (when the latter is not clearly defined).

A2.2 Details about $\tau_{\textbf{DID-X}}$ and $\tau_{\textbf{em-X}}$

We show in the following that generally,

(i) τ_{DID} is not a weighted average of $\tau_{\text{DID}}(X_i)$ over any distribution of X_i;

(ii) τ_{em} is not a weighted average of $\tau_{\text{em}}(X_i)$ over any distribution of X_i

unless X_i and G_i are independent.

First, it follows from $\mathbb{E}[\Delta Y_i\mid G_i=1,X_i]=\tau_{\text{DID}}(X_i)+\mathbb{E}[\Delta Y_i\mid G_i=0,X_i]$ and the law of iterated expectations that

$$\begin{aligned}\tau_{\text{DID}} &= \mathbb{E}[\Delta Y_i\mid G_i=1]-\mathbb{E}[\Delta Y_i\mid G_i=0]\\ &= \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i\mid G_i=1,X_i]\mid G_i=1\Big\}-\mathbb{E}\Big\{\mathbb{E}[\Delta Y_i\mid G_i=0,X_i]\mid G_i=0\Big\}\\ &= \mathbb{E}[\tau_{\text{DID}}(X_i)\mid G_i=1]\\ &\quad +\mathbb{E}\Big\{\mathbb{E}[\Delta Y_i\mid G_i=0,X_i]\mid G_i=1\Big\}-\mathbb{E}\Big\{\mathbb{E}[\Delta Y_i\mid G_i=0,X_i]\mid G_i=0\Big\},\end{aligned}\tag{A8}$$

where the outer expectation of the first two terms in (A8) are with respective to the conditional distribution of X_i given $G_i=1$ and the outer expectation of the third term in (A8) is with respective to the conditional distribution of X_i given $G_i = 0$. This implies Statement (i).

In addition, replacing ΔY_i by $\tau_{i,Z|G}$ in (A8) ensures

$$
\begin{aligned}
\tau_{\text{em}} &= \mathbb{E}[\tau_{i,Z|G} \mid G_i = 1] - \mathbb{E}[\tau_{i,Z|G} \mid G_i = 0] \\
&= \mathbb{E}\Big\{\mathbb{E}[\tau_{i,Z|G} \mid G_i = 1, X_i] \mid G_i = 1\Big\} - \mathbb{E}\Big\{\mathbb{E}[\tau_{i,Z|G} \mid G_i = 0, X_i] \mid G_i = 0\Big\} \\
&= \mathbb{E}[\tau_{\text{em}}(X_i) \mid G_i = 1] \\
&\quad + \mathbb{E}\Big\{\mathbb{E}[\tau_{i,Z|G} \mid G_i = 0, X_i] \mid G_i = 1\Big\} - \mathbb{E}\Big\{\mathbb{E}[\tau_{i,Z|G} \mid G_i = 0, X_i] \mid G_i = 0\Big\}.
\end{aligned}
$$

This implies Statement (ii).

A2.3 Reconciling FDID with Canonical DID

Figure A1 illustrates Proposition 2(ii) and reconciles FDID with canonical DID. When Assumption 4 holds, that is, when $\mathbb{E}[Y_{i,\text{post}}(0,1) \mid G_i = 0] = \mathbb{E}[Y_{i,\text{post}}(0,0) \mid G_i = 0]$, the FDID setting reduces to canonical DID.

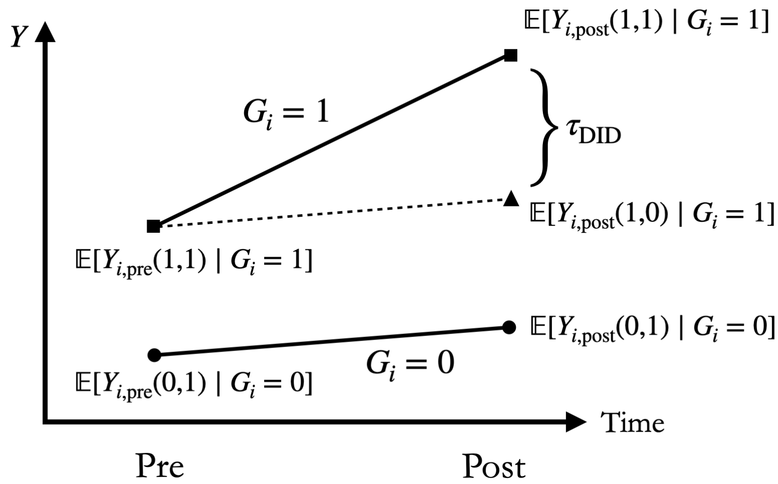

Figure A1: Reconcile FDID with Canonical DID.

Note. We use squares ■ to represent the group with $G_i = 1$, and circles ● to represent the group with $G_i = 0$. Assumption 2 (no anticipation) implies that in the pre-event period, ■ $= \mathbb{E}[Y_{i,\text{pre}}(1,0) \mid G_i = 1] = \mathbb{E}[Y_{i,\text{pre}} \mid G_i = 1]$ and ● $= \mathbb{E}[Y_{i,\text{pre}}(0,0) \mid G_i = 0] = \mathbb{E}[Y_{i,\text{pre}} \mid G_i = 0]$, so that ■ and ● represent the before and after observed outcomes of the two groups. The short dashed line parallels the solid line for $G_i = 0$ by construction of the DID estimator, so that in the post-period, the distance between ■ and ▲ equals τ_{DID}. Under Assumption 3, we have ▲ equals $\mathbb{E}[Y_{i,\text{post}}(1,0) \mid G_i = 1]$ such that the distance between the post-period ■ $= \mathbb{E}[Y_{i,\text{post}}(1,1) \mid G_i = 1]$ and ▲ also equals the ATT $= \mathbb{E}[Y_{i,\text{post}}(1,1) - Y_{i,\text{post}}(1,0) \mid G_i = 1]$. Under Assumption 4 (exclusion restriction), the post-period ● $= \mathbb{E}[Y_{i,\text{post}}(0,1) \mid G_i = 0]$ and ○ $= \mathbb{E}(Y_{i,\text{post}}(0,0) \mid G_i = 0]$ coincide, and only ● is shown. This ensures □ $= \mathbb{E}[Y_{i,\text{post}}(1,0) \mid G_i = 1]$ (not shown) merges into ▲, so Figure 4 reduces to Figure A1. This illustrates the correspondence between FDID and canonical DID under Assumption 4.

A2.4 Details for subunit-level analysis in Section 6.1

In this subsection, we discuss the identification and estimation of FDID using repeated cross-sectional (RCS) data from a nested sampling perspective. We consider three scenarios described in the main text. The preferred estimation strategies are summarized in Table A1.

Table A1: FDID with RCS Data: A Nested Sampling Perspctive

<table>
<tr><th></th><th>Unit-level analysis</th><th>Subunit-level analysis</th></tr>
<tr><td>Panel-panel</td><td rowspan="2">Stratification + OLS or TWFE (Section 5.2)</td><td>Stratification + OLS or TWFE</td></tr>
<tr><td>Panel-RCS</td><td rowspan="2">Stratification + TWFE</td></tr>
<tr><td>RCS-RCS</td><td>Stratification + TWFE</td></tr>
</table>

A2.4.1 DID estimand at the subunit level

Assume a *two-stage* sampling mechanism, also known as *grouped* or *cluster* sampling, where

(i) the first stage selects a simple random sample of units from a population of units;

(ii) the second stage selects a simple random sample of subunits within each selected unit.

Let i index a generic random unit from the population, and let ij index a generic random subunit within unit i. Let $G_i \in \{0, 1\}$ denote the baseline factor G of unit i. Let $Y_{ij,\text{pre}}, Y_{ij,\text{post}} \in \mathbb{R}$, $G_{ij} \in \{0, 1\}$, and $X_{ij} \in \mathbb{R}^p$ denote the before and after outcomes, baseline factor G, and covariates for subunits ij, where $G_{ij} = G_i$. The two-stage sampling mechanism ensures that $\{(Y_{ij,\text{pre}}, Y_{ij,\text{post}}), G_{ij}, X_{ij}\}$ are identically distributed across all subunits in the population. Let $\Delta Y_{ij} = Y_{ij,\text{post}} - Y_{ij,\text{pre}}$. The conditional DID estimand at the subunit level is

$$\tau'_{\text{DID}}(x) \quad = \quad \mathbb{E}[\Delta Y_{ij} \mid G_i = 1, X_i = x] - \mathbb{E}[\Delta Y_{ij} \mid G_i = 0, X_i = x],$$

where we use the prime superscript ($'$) to denote subunit-level quantities. Given

$$\mathbb{E}[\Delta Y_{ij} \mid G_i = g, X_{ij} = x] = \mathbb{E}[Y_{ij,\text{post}} \mid G_i = g, X_{ij} = x] - \mathbb{E}[Y_{ij,\text{pre}} \mid G_i = g, X_{ij} = x],$$

to estimate $\tau'_{\text{DID}}(x)$, it suffices to estimate either $\mathbb{E}[\Delta Y_{ij} \mid G_i = g, X_{ij} = x]$ or $\mathbb{E}[Y_{ijt} \mid G_i = g, X_{ij} = x]$. Depending on the definition of the unit-level outcome Y_{it} in terms of Y_{ijt}, $\tau'_{\text{DID}}(x)$ may or may not equal $\tau_{\text{DID}}(x)$.

A2.4.2 Regression estimation of $\tau'_{\text{DID}}(x)$

Let i_t index the i-th unit in the sample at time t, and let i_tj denote the j-th subunit in unit i_t. With slight abuse of notation, let G_{it} denote the level of the baseline factor G of unit i_t, and let (Y_{ijt}, X_{ijt}) denote the outcome and covariates of subunit i_tj, with $Y_{ijt} = Y_{i_tjt}$ under the population notation in Section A2.4.1. This notation covers the panel-panel, panel-RCS, and RCS-RCS schemes. Under the panel-panel scheme, we can suppress the subscript t in i_t, and use i and ij to index the units and subunits, respectively, with $\{(G_{it}, Y_{ijt}, X_{ijt}) : t = \text{pre}, \text{post}\}$ reduced to $(G_i, Y_{ij,\text{pre}}, Y_{ij,\text{post}}, X_{ij})$.

A2.4.2.1 TWFE for panel or repeated cross-section subunit-level data. Parallel to Definition 7, two common specifications for subunit-level TWFE regression based on $\{(G_{it}, Y_{ijt}, X_{ijt}) : t = \text{pre}, \text{post}\}$ are the OLS fits of the following TWFE models, with and without the three-way interactions among G_{it}, X_{ijt} and $1_{\{t=\text{post}\}}$:

$$\begin{aligned} \text{TWFE}'_* : \quad & Y_{ijt} = b_G G_{it} \cdot 1_{\{t=\text{post}\}} + b_X X_{ijt} \cdot 1_{\{t=\text{post}\}} + b_{GX} G_{it} X_{ijt} \cdot 1_{\{t=\text{post}\}} + \alpha_{it} + \xi_t + \epsilon_{ijt}, \\ \text{TWFE}'_+ : \quad & Y_{ijt} = b_G G_{it} \cdot 1_{\{t=\text{post}\}} + b_X X_{ijt} \cdot 1_{\{t=\text{post}\}} + \alpha_{it} + \xi_t + \epsilon_{ijt}, \end{aligned} \tag{A9}$$

where $(\alpha_{it}, i_t, \epsilon_{ijt})$ denote the fixed effect of unit i_t, the time fixed effect, and an idiosyncratic error of subunit-time pair (i_tj, t). With panel subunit-level data, (A9) simplifies to

$$\begin{aligned} \text{TWFE}'_* : \quad & Y_{ijt} = b_G G_i \cdot 1_{\{t=\text{post}\}} + b_X X_{ij} \cdot 1_{\{t=\text{post}\}} + b_{GX} G_i X_{ij} \cdot 1_{\{t=\text{post}\}} + \alpha_{it} + \xi_t + \epsilon_{ijt}, \\ \text{TWFE}'_+ : \quad & Y_{ijt} = b_G G_i \cdot 1_{\{t=\text{post}\}} + b_X X_{ij} \cdot 1_{\{t=\text{post}\}} + \alpha_{it} + \xi_t + \epsilon_{ijt}. \end{aligned}$$

A2.4.2.2 OLS for panel subunit-level data. With panel subunit-level data $(G_i, Y_{ij,\text{pre}}, Y_{ij,\text{post}}, X_{ij})$, parallel to Definition 6, two common specifications for subunit-level cross-sectional regression based on $(\Delta Y_{ij}, G_i, X_{ij})$ are the OLS fits of the following models, with and without interactions between G_i and X_{ij}:

$$\begin{aligned} \text{OLS}'_* : \quad & \Delta Y_{ij} = \beta_1 + \beta_G G_i + \beta_X^\top X_{ij} + \beta_{GX}^\top G_i X_{ij} + \epsilon_{ij}, \\ \text{OLS}'_+ : \quad & \Delta Y_{ij} = \beta_1 + \beta_G G_i + \beta_X^\top X_{ij} + \epsilon_{ij}. \end{aligned}$$

A2.4.2.3 Justification. All results in Section 5.2 at the unit level extend to the subunit level:

(i) OLS'_* and TWFE'_* are numerically equivalent, and are consistent for estimating τ'_{DID} if $\mathbb{E}[\Delta Y_{ij} \mid G_i = g, X_{ij} = x]$ is linear in $(G_i, X_{ij}, G_i X_{ij})$.

(ii) OLS'_+ and TWFE'_+ are numerically equivalent, and are consistent for estimating τ'_{DID} if $\mathbb{E}[\Delta Y_{ij} \mid G_i = g, X_{ij} = x]$ is linear in (G_i, X_{ij}).

In practice, subunit-level outcomes may depend on unit size. When this is the case, it is important to include unit size as a covariate in X_{ijt} to improve the plausibility of the linearity assumption.

A2.5 Details for extension to general baseline factor G

This subsection discusses results when G is extended to be a discrete or continuous variable.

A2.5.1 Identification

Assumption A1 and Corollary A1 below generalize Corollary 1 to general G, clarifying the conditions for $\tau_{\text{em-x},g\to g'}$ and $\tau_{\text{cm},g\to g'}$ by $\tau_{\text{DID-x},g\to g'}$.

Assumption A1. (i) *No anticipation:* $Y_{i,\text{pre}}(g,0) = Y_{i,\text{pre}}(g,1)$ for all $g \in \mathcal{G}$.

(ii) *Overlap:* For all $x \in \mathcal{X}$,

- For discrete G, $\mathbb{P}(G_i = g \mid X_i = x) \in (0,1)$ for all $g \in \mathcal{G}$;
- For continuous G, the conditional density satisfies $f_{G|X}(g \mid X_i = x) > 0$ for all $g \in \mathcal{G}$;

(iii) *Conditional canonical parallel trends:* $\mathbb{E}[\Delta Y_i(G_i, 0) \mid G_i = g, X_i]$ is constant across $g \in \mathcal{G}$.

(iv) *Conditional factorial parallel trends:* For $g \in \mathcal{G}$ and $z = 0, 1$, $\mathbb{E}[\Delta Y_i(g,z) \mid G_i = g', X_i]$ is constant across $g' \in \mathcal{G}$.

Corollary A1. (i) If Assumption 1 and Assumption A1(i)–(iii) hold, then $\tau_{\text{DID},g\to g'}(x) = \tau_{\text{em},g\to g'}(x)$ and $\tau_{\text{DID-x},g\to g'} = \tau_{\text{em-x},g\to g'}$.

(ii) If Assumption 1 and Assumption A1(i)–(iv) hold, then $\tau_{\text{DID},g\to g'}(x) = \tau_{\text{em},g\to g'}(x) = \tau_{\text{cm},g\to g'}(x)$ and $\tau_{\text{DID-x},g\to g'} = \tau_{\text{em-x},g\to g'} = \tau_{\text{cm},g\to g'}$.

A2.5.2 Regression estimation

Echoing Section 5.2, regression gives a convenient way to estimate $\tau_{\text{DID},g\to g'}(x)$ and $\tau_{\text{DID-x},g\to g'}$ for general G. We establish below the identification results based on regression coefficients.

Renew $(\hat\beta_{1,*}, \hat\beta_{G,*}, \hat\beta_{X,*}, \hat\beta_{GX,*})$ and $(\hat\beta_{1,+}, \hat\beta_{G,+}, \hat\beta_{X,+})$ as the coefficients from OLS$_*$ and OLS$_+$, respectively, with general G_i. Renew

$$\widehat{\Delta Y_*}(g,x) = \hat\beta_{1,*} + \hat\beta_{G,*}g + \hat\beta_{X,*}^\top x + \hat\beta_{GX,*}^\top gx, \quad \widehat{\Delta Y_+}(g,x) = \hat\beta_{1,+} + \hat\beta_{G,+}g + \hat\beta_{X,+}^\top x$$

for $(g,x) \in \mathcal{G} \times \mathcal{X}$ as the corresponding estimators of $\mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]$, and

$$\begin{aligned} \hat\tau_{\text{DID},g\to g',*}(x) &= \widehat{\Delta Y_*}(g',x) - \widehat{\Delta Y_*}(g,x) &&= (g'-g)(\hat\beta_{G,*} + \hat\beta_{GX,*}^\top x), \\ \hat\tau_{\text{DID-x},g\to g',*} &= n^{-1}\sum_{i=1}^n \hat\tau_{\text{DID},g\to g',*}(X_i) &&= (g'-g)(\hat\beta_{G,*} + \hat\beta_{GX,*}^\top \bar X), \end{aligned} \tag{A10}$$

$$\begin{aligned} \hat\tau_{\text{DID},g\to g',+}(x) &= \widehat{\Delta Y_+}(g',x) - \widehat{\Delta Y_+}(g,x) &&= (g'-g)\hat\beta_{G,+}, \\ \hat\tau_{\text{DID-x},g\to g',+} &= n^{-1}\sum_{i=1}^n \hat\tau_{\text{DID},g\to g',+}(X_i) &&= (g'-g)\hat\beta_{G,+} \end{aligned} \tag{A11}$$

as the corresponding DID estimators, generalizing $(\hat\tau_{\text{DID},*}(x), \hat\tau_{\text{DID-x},*})$ and $(\hat\tau_{\text{DID},+}(x), \hat\tau_{\text{DID-x},+})$. Proposition A1 below generalizes Proposition 5 to general G. Standard OLS theory ensures that the numerical equivalence between OLS$_+$ and TWFE$_+$, and that between OLS$_*$ and TWFE$_*$, also holds for general G.

Proposition A1. (i) If $\mathbb{E}[\Delta Y_i \mid G_i, X_i] = \beta_1 + \beta_G G_i + \beta_X^\top X_i + \beta_{GX}^\top G_i X_i$ for some constant $(\beta_1, \beta_G, \beta_X, \beta_{GX})$, then

$$\tau_{\text{DID},g\to g'}(x) = (g'-g)(\beta_G + \beta_{GX}^\top x), \qquad \tau_{\text{DID-x},g\to g'} = (g'-g)(\beta_G + \beta_{GX}^\top \mathbb{E}[X_i]),$$

for $g, g' \in \mathcal{G}$ and $x \in \mathcal{X}$, so that $\hat\tau_{\text{DID},g\to g',*}(x) = (g'-g)(\hat\beta_{G,*} + \hat\beta_{GX,*}^\top x)$ and $\hat\tau_{\text{DID-x},g\to g',*} = (g'-g)(\hat\beta_{G,*} + \hat\beta_{GX,*}^\top \bar X)$ in (A10) based on OLS$_*$ are consistent for $\tau_{\text{DID}}(x)$ and $\tau_{\text{DID-x}}$.

(ii) If $\mathbb{E}[\Delta Y_i \mid G_i, X_i] = \beta_1 + \beta_G G_i + \beta_X^\top X_i$ for some constant $(\beta_1, \beta_G, \beta_X)$, then

$$\tau_{\text{DID},g\to g'}(x) = \tau_{\text{DID-x},g\to g'} = (g'-g)\beta_G,$$

for $g, g' \in \mathcal{G}$ and $x \in \mathcal{X}$, so that $\hat\tau_{\text{DID},g\to g',+}(x) = \hat\tau_{\text{DID-x},g\to g',+} = (g'-g)\hat\beta_{G,+}$ in (A11) based on OLS$_+$ are consistent for $\tau_{\text{DID},g\to g'}(x) = \tau_{\text{DID-x},g\to g'}$.

Proof of Corollary A1(i). Under Assumption A1, we have

$$\begin{aligned}
\tau_{\text{DID},g\to g'}(x) &= \mathbb{E}[\Delta Y_i \mid G_i = g', X_i = x] - \mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x] \\
&\overset{\text{Assumption 1}}{=} \mathbb{E}[\Delta Y_i(g',1) \mid G_i = g', X_i = x] - \mathbb{E}[\Delta Y_i(g,1) \mid G_i = g, X_i = x] \\
&\overset{\text{Assumption A1(iii)}}{=} \mathbb{E}[\Delta Y_i(g',1) \mid G_i = g', X_i = x] - \mathbb{E}[\Delta Y_i(g',0) \mid G_i = g', X_i = x] \\
&\quad -\Big\{\mathbb{E}[\Delta Y_i(g,1) \mid G_i = g, X_i = x] - \mathbb{E}[\Delta Y_i(g,0) \mid G_i = g, X_i = x]\Big\} \\
&= \mathbb{E}[\Delta Y_i(g',1) - \Delta Y_i(g',0) \mid G_i = g', X_i = x] \\
&\quad -\mathbb{E}[\Delta Y_i(g,1) - \Delta Y_i(g,0) \mid G_i = g, X_i = x] \\
&\overset{\text{Assumption A1(i)}}{=} \mathbb{E}[\tau_{Z\mid G=g'} \mid G_i = g', X_i = x] - \mathbb{E}[\tau_{Z\mid G=g} \mid G_i = g, X_i = x] \\
&= \tau_{\text{em},g\to g'}(x).
\end{aligned}$$

□

Proof of Proposition A1. When $\mathbb{E}[\Delta Y_i \mid G_i, X_i] = \beta_1 + \beta_G G_i + \beta_X^\top X_i + \beta_{GX}^\top G_i X_i$, we have

$$\begin{aligned}
\tau_{\text{DID},g\to g'}(x) &= \mathbb{E}[\Delta Y_i \mid G_i = g', X_i = x] - \mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x] \\
&= (g' - g)(\beta_G + \beta_{GX}^\top x)
\end{aligned}$$

by definition, and properties of least squares ensure

$$\text{plim}\,(\hat{\beta}_{G,*}, \hat{\beta}_{X,*}, \hat{\beta}_{GX,*}) = (\beta_G, \beta_X, \beta_{GX}).$$

This ensures Proposition A1(i).

Proposition A1(ii) is a direct consequence of Proposition A1(i) with $\beta_{GX} = 0$. □

A2.5.3 Extension to the incremental effects

When $\mathcal{G}$ is an interval, define

$$\begin{aligned}
\delta_{\text{DID}}(g,x) &= \frac{\partial \mathbb{E}[\Delta Y_i \mid G_i = g, X_i = x]}{\partial g}, \\
\delta_{\text{em}}(g,x) &= \frac{\partial \mathbb{E}[\tau_{i,Z\mid G=g} \mid G_i = g, X_i = x]}{\partial g}, \\
\delta_{\text{cm}}(g,x) &= \frac{\partial \mathbb{E}[\tau_{i,Z\mid G=g} \mid X_i = x]}{\partial g}
\end{aligned}$$

as the "incremental effects" in the sense of Rothenhäusler and Yu (2019).

Parallel to Corollary 1, the DID estimand $\delta_{\text{DID}}(g,x)$ (i) identifies $\delta_{\text{em}}(g,x)$ under the generalized conditional canonical parallel trends assumptions, and (ii) identifies $\delta_{\text{em}}(g,x) = \delta_{\text{cm}}(g,x)$ under the generalized conditional canonical and factorial parallel trends assumptions.

Under the assumption in Proposition A1(i), we have $\delta_{\text{DID}}(g,x) = \beta_G + \beta_{GX}^\top x$, and under the assumption in Proposition A1(ii), we have $\delta_{\text{DID}}(g,x) = \beta_G$. In addition, (i) when OLS$_*$ is correctly specified, then $\delta_{\text{DID}}(g,x) = \beta_G + \beta_{GX} x$ is constant with each level of x, and the estimated coefficient of G_i is consistent for β_G; (ii) when OLS$_+$ is correctly specified, then $\delta_{\text{DID}}(g,x) = \beta_G$, and the estimated coefficient of G_i from OLS$_+$ is consistent for β_G.

A2.6 Identification by inverse propensity score weighting

Outcome regression and inverse propensity score weighting are two leading methods in causal inference from observational cross-sectional data. We extend below the discussion in Section 5 to inverse propensity score weighting, and establish the corresponding identification results in FDID.

Let $e(x) = \mathbb{P}(G_i = 1 \mid X_i = x)$ denote the propensity score of G_i given X_i (Rosenbaum and Rubin, 1983). Define

$$\tau_{\text{IPW}} = \mathbb{E}\left[\frac{G_i}{e(X_i)} \cdot \Delta Y_i - \frac{1-G_i}{1-e(X_i)} \cdot \Delta Y_i\right]$$

as the inverse-propensity-score-weighted estimand by treating $\{\Delta Y_i, G_i : i = 1, \dots, n\}$ as the cross-sectional input data. Define

$$\tau_{\text{IPW}}(x) = \mathbb{E}\left[\frac{G_i}{e(x)} \cdot \Delta Y_i - \frac{1-G_i}{1-e(x)} \cdot \Delta Y_i \mid X_i = x\right]$$

as the conditional version of τ_{IPW} with $\tau_{\text{IPW}} = \mathbb{E}[\tau_{\text{IPW}}(X_i)]$. Define

$$\begin{aligned}
\tau_{\text{IPW}|G=1} &= \mathbb{E}[\Delta Y_i \mid G_i = 1] - \mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1-G_i}{1-e(X_i)} \cdot \Delta Y_i\right], \\
\tau_{\text{IPW}|G=0} &= \mathbb{E}\left[\frac{1-e(X_i)}{1-e} \cdot \frac{G_i}{e(X_i)} \cdot \Delta Y_i\right] - \mathbb{E}[\Delta Y_i \mid G_i = 0],
\end{aligned}$$

where $e = \mathbb{P}(G_i = 1)$. Proposition A2 below ensures the numerical equivalence between $\tau_{\text{IPW}}(x)$ and $\tau_{\text{DID}}(x)$ and that between $\tau_{\text{IPW}|G=g}$ and $\mathbb{E}[\tau_{\text{DID}}(X_i) \mid G_i = g]$ for $g = 0, 1$.

Identification based on $\tau_{\text{IPW}}(x)$, τ_{IPW}, and $\tau_{\text{IPW}|G=g}$ ($g = 0, 1$) then follows from Corollary 1. Specifically, $\tau_{\text{IPW}|G=g}$ identifies $\mathbb{E}[\tau_{\text{em}}(X_i) \mid G_i = g]$ under Assumptions 1–2 and 7–8 (universal exposure, no anticipation, overlap, and conditional canonical parallel trends), and identifies $\mathbb{E}[\tau_{\text{cm}}(X_i) \mid G_i = g]$ if we further assume Assumption 9 (conditional factorial parallel trends).

Proposition A2. Under Assumption 7, we have

(i) $\tau_{\text{IPW}}(x) = \tau_{\text{DID}}(x)$ with $\tau_{\text{IPW}} = \tau_{\text{DID-X}}$.

(ii) $\tau_{\text{IPW}|G=g} = \mathbb{E}[\tau_{\text{DID}}(X_i) \mid G_i = g]$ for $g = 0, 1$.

Proof of Proposition A2. We verify below the results about $\tau_{\text{IPW}}(X_i)$ and $\tau_{\text{IPW}|G=g}$ ($g = 0, 1$), respectively.

Results about $\tau_{\text{IPW}}(X_i)$. The law of iterated expectations ensures

$$\begin{aligned}
\mathbb{E}\left[\frac{G_i}{e(X_i)} \cdot \Delta Y_i \mid X_i\right] &= \mathbb{E}\left[\frac{G_i}{e(X_i)} \cdot \Delta Y_i \mid G_i = 1, X_i\right] \cdot \mathbb{P}(G_i = 1 \mid X_i) \\
&= \mathbb{E}[\Delta Y_i \mid G_i = 1, X_i]. \qquad \text{(A12)}
\end{aligned}$$

By symmetry, we have

$$\mathbb{E}\left[\frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i \mid X_i\right] = \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]. \qquad \text{(A13)}$$

Plugging (A12)–(A13) in the definition of $\tau_{\text{IPW}}(X_i)$ ensures

$$\begin{aligned}
\tau_{\text{IPW}}(X_i) &= \mathbb{E}\left[\frac{G_i}{e(X_i)} \cdot \Delta Y_i - \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i \mid X_i\right] \\
&= \mathbb{E}\left[\frac{G_i}{e(X_i)} \cdot \Delta Y_i \mid X_i\right] - \mathbb{E}\left[\frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i \mid X_i\right] \\
&\overset{\text{(A12)+(A13)}}{=} \mathbb{E}[\Delta Y_i \mid G_i = 1, X_i] - \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \\
&= \tau_{\text{DID}}(X_i).
\end{aligned}$$

Results about $\tau_{\text{IPW}|G=g}$. We verify in the following

$$\tau_{\text{IPW}|G=1} = \mathbb{E}[\tau_{\text{DID}}(X_i) \mid G_i = 1], \qquad \text{(A14)}$$

where

$$\tau_{\text{IPW}|G_i=1} = \mathbb{E}[\Delta Y_i \mid G_i = 1] - \mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i\right]. \qquad \text{(A15)}$$

That $\tau_{\text{IPW}|G=0} = \mathbb{E}[\tau_{\text{DID}}(X_i) \mid G_i = 0]$ follows by symmetry.

First, the right-hand side of (A14) equals

$$\begin{aligned}
\mathbb{E}[\tau_{\text{DID}}(X_i) \mid G_i = 1] &= \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 1, X_i] - \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid G_i = 1\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 1, X_i] \mid G_i = 1\Big\} - \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid G_i = 1\Big\} \\
&= \mathbb{E}[\Delta Y_i \mid G_i = 1] - \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid G_i = 1\Big\}.
\end{aligned}$$

This, together the definition of $\tau_{\text{IPW}|G=1}$ in (A15), ensures (A14) equals

$$\mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i\right] = \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid G_i = 1\Big\}. \tag{A16}$$

We show (A16) in the following. First, the law of iterated expectations ensures

$$\begin{aligned}
\mathbb{E}\Big[(1 - G_i) \cdot \Delta Y_i \mid X_i\Big] &= \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \cdot \mathbb{P}(G_i = 0 \mid X_i) \\
&= \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \cdot \{1 - e(X_i)\}\,. \qquad \text{(A17)}
\end{aligned}$$

This ensures the conditional version of the left-hand side of (A16) equals

$$\begin{aligned}
\mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i \mid X_i\right] &= \frac{e(X_i)}{e} \cdot \frac{1}{1 - e(X_i)} \cdot \mathbb{E}\Big[(1 - G_i) \cdot \Delta Y_i \mid X_i\Big] \\
&\overset{\text{(A17)}}{=} \frac{e(X_i)}{e} \cdot \frac{1}{1 - e(X_i)} \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \cdot \{1 - e(X_i)\} \\
&= \frac{e(X_i)}{e} \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]. \qquad \text{(A18)}
\end{aligned}$$

In addition, $\mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]$ is a function of X_i so that

$$\mathbb{E}(G_i \mid X_i) \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] = \mathbb{E}\Big(G_i \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid X_i\Big). \tag{A19}$$

This, together with (A18) and the law of iterated expectations, ensures (A16) as follows:

$$\begin{aligned}
\mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i\right] &= \mathbb{E}\left\{\mathbb{E}\left[\frac{e(X_i)}{e} \cdot \frac{1 - G_i}{1 - e(X_i)} \cdot \Delta Y_i \mid X_i\right]\right\} \\
&\overset{\text{(A18)}}{=} \mathbb{E}\left\{\frac{e(X_i)}{e} \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]\right\} \\
&= e^{-1} \cdot \mathbb{E}\Big\{\mathbb{E}(G_i \mid X_i) \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]\Big\} \\
&\overset{\text{(A19)}}{=} e^{-1} \cdot \mathbb{E}\left\{\mathbb{E}\Big(G_i \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid X_i\Big)\right\} \\
&= e^{-1} \cdot \mathbb{E}\Big\{G_i \cdot \mathbb{E}[\Delta Y_i \mid G_i = 0, X_i]\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}[\Delta Y_i \mid G_i = 0, X_i] \mid G_i = 1\Big\}.
\end{aligned}$$

□

A3 Additional Information on the Application

Below we provide additional information on the empirical application.

- Table A2: Descriptive statistics
- Figure A2: Assessing Assumption 7 (overlap) using the estimated propensity score
- Figure A3: Sensitivity analysis for the relationship between G and ΔY

Table A2: Descriptive Statistics

Variable	N	Mean	Median	SD	Min	Max
High social capital	921	0.50	1.00	0.50	0.00	1.00
Per capita grain production	921	285.39	272.14	96.35	69.72	613.36
Ratio of non-farming land	921	0.22	0.21	0.09	0.01	0.48
Share of urban population	921	6.86	5.62	4.54	0.56	25.93
Distance from Beijing (km)	921	937.29	808.83	563.43	43.11	2149.29
Distance from provincial capital (km)	921	194.90	182.51	101.94	0.00	481.79
Suitable for rice cultivation	921	0.42	0.00	0.49	0.00	1.00
Share of ethnic minorities	921	0.17	0.00	0.37	0.00	1.00
Average years of education	921	2.30	2.26	0.63	0.72	4.37
Log population	921	12.49	12.57	0.72	9.59	13.95
Mortality rate (‰) in 1954	921	12.68	12.38	4.91	0.30	92.56
Mortality rate (‰) in 1955	921	12.47	11.88	4.35	1.36	46.94
Mortality rate (‰) in 1956	921	11.92	11.61	3.90	1.26	39.20
Mortality rate (‰) in 1957	921	11.69	11.41	2.94	1.36	22.64
Mortality rate (‰) in 1958	921	13.65	11.86	6.53	1.11	47.32
Mortality rate (‰) in 1959	921	19.65	14.38	16.53	1.15	183.47
Mortality rate (‰) in 1960	921	29.51	19.47	26.84	1.15	218.88
Mortality rate (‰) in 1961	921	17.49	14.00	11.30	1.16	86.49
Mortality rate (‰) in 1962	921	11.01	10.40	4.01	1.43	51.49
Mortality rate (‰) in 1963	921	11.48	10.99	3.73	1.24	63.00
Mortality rate (‰) in 1964	921	12.82	11.81	3.91	1.39	28.96
Mortality rate (‰) in 1965	921	10.50	9.99	2.91	1.30	21.45
Mortality rate (‰) in 1966	921	9.86	9.48	2.71	1.18	24.52

Note: High social capital is defined as a county having a number of genealogy books greater than or equal to the sample median.

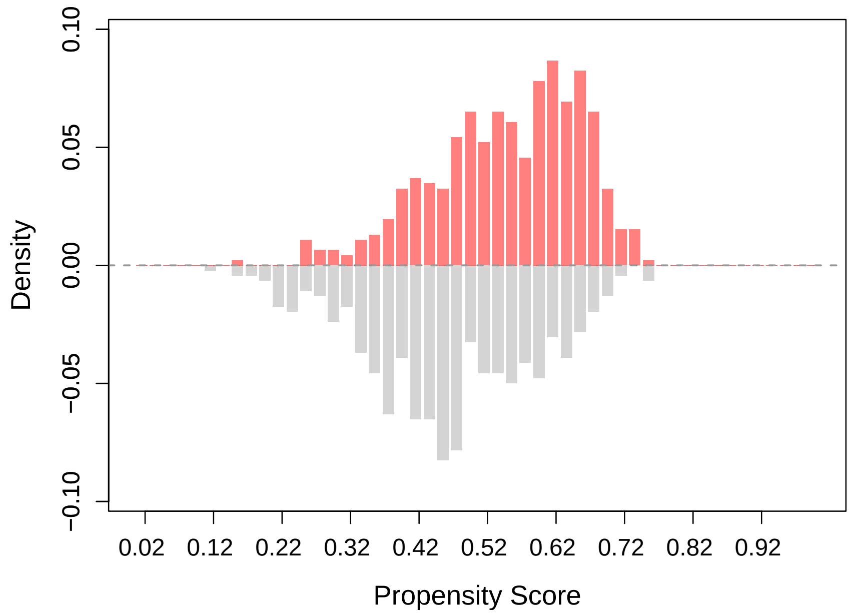

Figure A2: Accessing Assumption 7 (overlap) using the estimated propensity score. Histograms of propensity scores estimated using a generalized random forest, based on the same set of covariates as in the regression analysis in Table 1 columns (2)–(3), are shown in red for the group $\{i : G = 1\}$ and gray for the group $\{i : G = 0\}$.

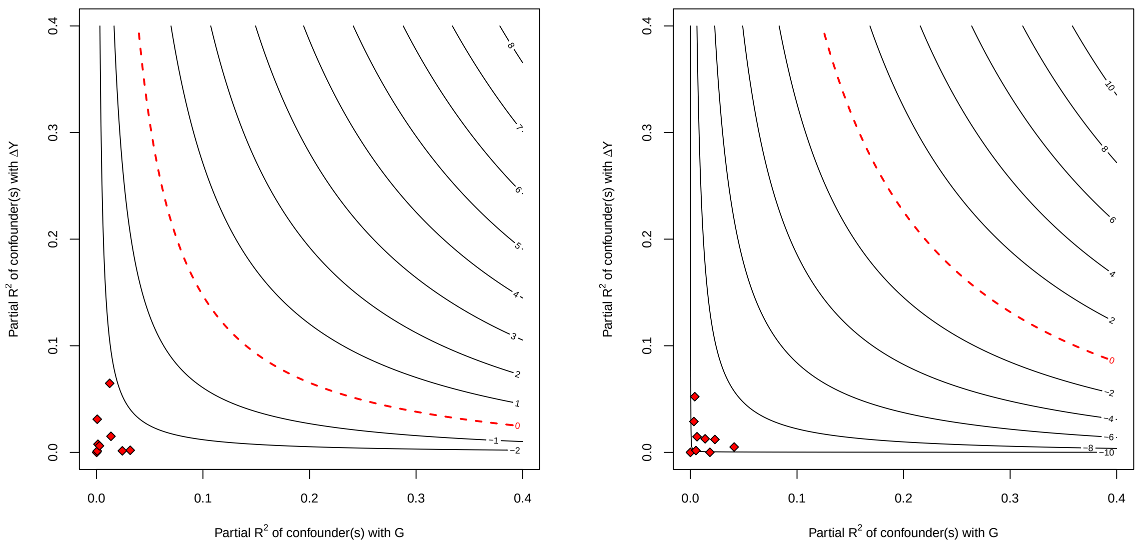

Figure A3: Sensitivity analysis based on Cinelli and Hazlett (2020). The left and right panels use binary and continuous measures of social capital G, respectively, with OLS_+ as the estimator. The outcome ΔY is defined as the average mortality from 1958–1961 (the famine year) minus the 1957 mortality rate (baseline). The results show that a confounder whose associations with G and ΔY are comparable to those of existing pre-event covariates would not be strong enough to eliminate the negative association between G and ΔY. The red diamonds in each plot illustrate the correlation of an existing covariate with both G and ΔY.