Matched Triple-Differences: A Framework for Covariate Adjustment

Yihong Liu (UCSB)

Job Market Talk

Nov 6th, 2026

Motivation

- Difference-in-Differences (DiD) is widely used for causal inference and its identification relies on the parallel trends assumption.
- In practice, researchers may worry that this assumption is too restrictive.
- This concern is particularly relevant when treatment requires two criteria: exposure and eligibility.
- In such a setting, researchers often use a **triple-differences design** (DDD) and incorporate **covariates** to make the identification assumption more credible.

Motivation

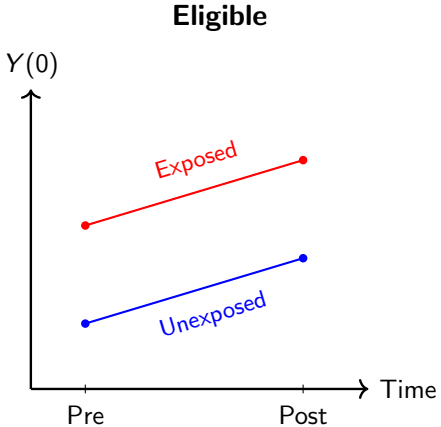
- In 1978, three states (exposed states) in the U.S. passed laws requiring employer-provided health insurance plans to cover pregnancy and childbirth.
  - Gruber (1994) studies the effect of mandated maternity coverage on workers' labor market outcomes.
- A worker needs to satisfy two criteria to be treated:
  - Exposure: Live in exposed states that introduced maternity coverage;
  - Eligibility: Married women of childbearing age.

Motivation

- A worker needs to satisfy two criteria to be treated:
  - Exposure: Live in exposed states that introduced maternity coverage;
  - Eligibility: Married women of childbearing age.

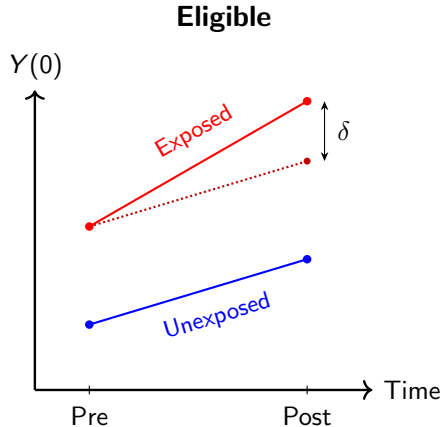
	Exposed	**Unexposed**
Eligible	**Exposed Eligible (Treated)**	Unexposed Eligible
Ineligible	Exposed Ineligible	Unexposed Ineligible

Motivation



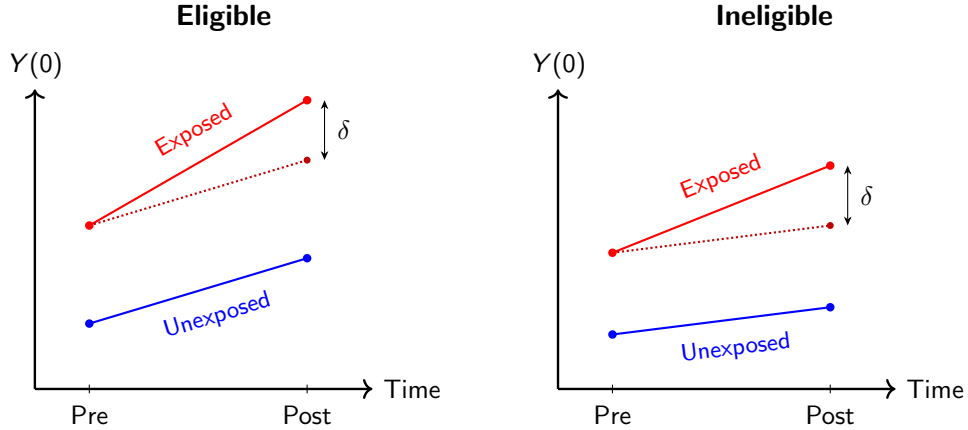
Motivation

- Violation of Parallel Trends: $DiD_{Eligible} = ATT + \delta$



Motivation

- Violation of Parallel Trends: $DiD_{Eligible} = ATT + \delta$
- **Triple-Differences**: $DDD = DiD_{Eligible} - DiD_{Ineligible} = ATT + \delta - \delta = ATT$



Motivation

- State-specific shocks may affect workers differently depending on their covariates.
  - Workers with different education levels may respond differently to the same shock.
  - Violation of parallel trends, $\delta(x)$, may therefore vary with covariates.
- Incorporating **covariates** in DDD makes the identification more credible.
- Common empirical practice: include covariate trends or covariate levels in a 3WFE regression.
  - 18 papers published in leading economics journals (AER, QJE, etc.) use at least one of these two specifications.

Research Question

- Do these two 3WFE specifications consistently estimate the ATT?
  - No, they are inconsistent in general.

Research Question

- Do these two 3WFE specifications consistently estimate the ATT?
  - No, they are inconsistent in general.
- What alternative procedure can consistently estimate the ATT?
  - Matching + Triple-differences.

Preview of Results I: Diagnosis of Empirical Practice

- Both 3WFE specifications inconsistently estimate the ATT in general.
- Both specifications use the wrong weights to reweight the four subgroups.
- Both specifications require additional assumptions to recover the ATT:
  - Strong restrictions on treatment effect heterogeneity, conditional outcome means, or propensity scores.

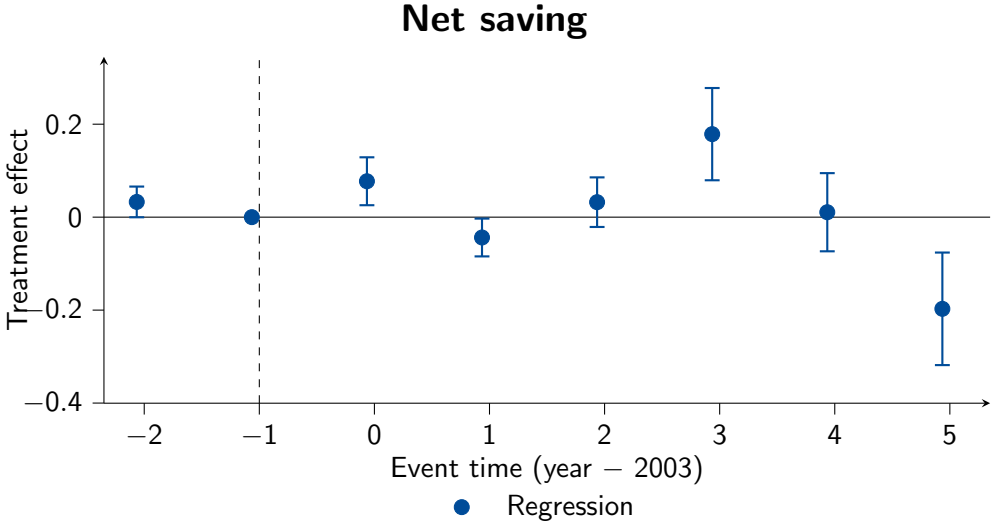
Preview of Results II: Matched Triple-Differences Framework

- **Estimation:** I propose a two-step procedure: matching + 3WFE.
- **Inference:** I establish the asymptotic normality + variance estimator.
- **Generality:** The asymptotic theory accommodates commonly used covariate matching methods, including nearest-neighbor and kernel matching.

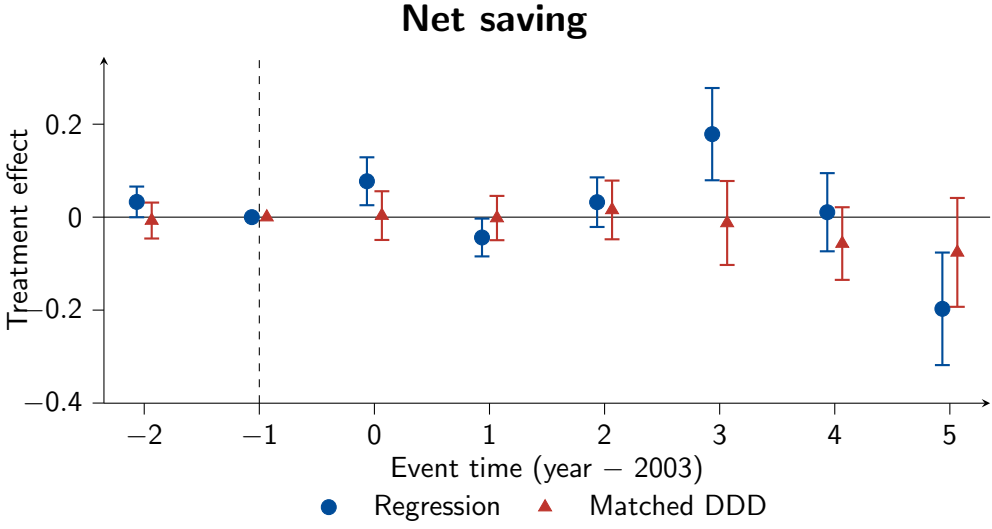
Empirical Application: Agricultural Insurance and Household Saving

- Cai (2016): a household-level agricultural insurance program implemented in one of 12 tobacco-producing counties in China.
- Treatment: A household needs to satisfy two criteria to be offered the insurance:
  - Exposure: Live in the county providing the agricultural insurance program.
  - Eligibility: Grow tobacco.
- Parameter of Interest: The effect of the insurance program on treated households' saving behavior.

Empirical Application: Agricultural Insurance and Household Saving



Empirical Application: Agricultural Insurance and Household Saving



Related Literature and Contribution

- **3WFE and Triple-Differences**
  - 3WFE without covariates: Olden and Moen (2022), Strezhnev (2023)
  - **Contribution:** I formally analyze the two commonly used 3WFE specifications with covariates and characterize their limitations.

Related Literature and Contribution

- **3WFE and Triple-Differences**
  - **Contribution:** I formally analyze the two commonly used 3WFE specifications with covariates and characterize their limitations.
- **Matching Methods**
  - Cross-sectional: Heckman, Ichimura, and Todd (1998), Abadie and Imbens (2006), Lin, Ding, and Han (2023)
  - DiD: Heckman, Ichimura, and Todd (1997), Imai, Kim, and Wang (2023), Liu and Vazquez-Bare (2026)
  - **Contribution:** I formalize the use of matching in triple-differences designs and develop asymptotic theory and inference for a class of matching methods.

Related Literature and Contribution

- **3WFE and Triple-Differences**
  - **Contribution:** I formally analyze the two commonly used 3WFE specifications with covariates and characterize their limitations.
- **Matching Methods**
  - **Contribution:** I formalize the use of matching in triple-differences designs and develop asymptotic theory and inference for a class of matching methods.
- **Causal Interpretation of Linear Regression Estimands**
  - Chattopadhyay and Zubizarreta (2023), Caetano and Callaway (2026), Słoczyński (2022, 2026), Blandhol et al. (2026)

Road Map

- **Setup and Identification**
- Limitations of 3WFE with Covariates
- Matched Triple-Differences Framework
- Empirical Application
- Conclusion

Setup

- Balanced panel data: units $i = 1, \ldots, n$, periods $t = 1, \ldots, T$
- Subgroup indicators:
  - $G \in \{0, 1\}$: exposure indicator; e.g., live in states introducing maternity coverage
  - $S \in \{0, 1\}$: eligibility indicator; e.g., a married woman of childbearing age
- I use a shorthand notation $\mathcal{I}_{gs}$ to denote subgroup $(G = g, S = s)$

Setup

- Treatment is implemented in period $2 \leq t^* \leq T$
  - Treated subgroup indicator: $D = G \times S$
  - Treatment indicator: $D_t = G \times S\mathbb{1}(t \geq t^*)$
  - Treatment is an absorbing state: Treated units cannot "become untreated"

Setup

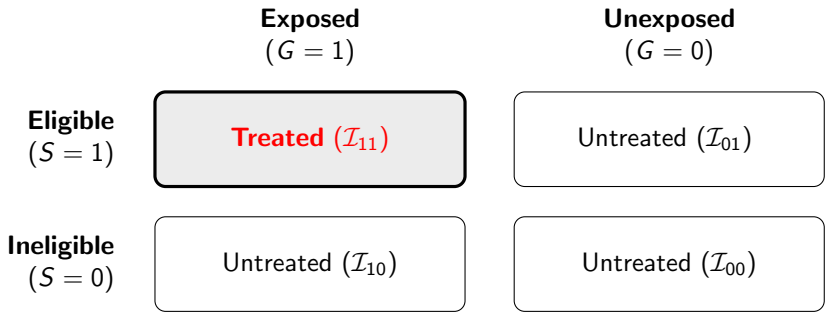
- Treatment is implemented in period $2 \leq t^* \leq T$
  - Treated subgroup indicator: $D = G \times S$
  - Treatment indicator: $D_t = G \times S\mathbb{1}(t \geq t^*)$
  - Treatment is an absorbing state: Treated units cannot "become untreated"
- Potential outcomes: treated $Y_t(1)$, untreated $Y_t(0)$
  - Observed outcome: $Y_t = Y_t(1)D_t + Y_t(0)(1 - D_t)$

Setup

- Treatment is implemented in period $2 \leq t^* \leq T$
  - Treated subgroup indicator: $D = G \times S$
  - Treatment indicator: $D_t = G \times S\mathbb{1}(t \geq t^*)$
  - Treatment is an absorbing state: Treated units cannot "become untreated"
- Potential outcomes: treated $Y_t(1)$, untreated $Y_t(0)$
  - Observed outcome: $Y_t = Y_t(1)D_t + Y_t(0)(1 - D_t)$
- Parameters of interest: ATTs in all post-treatment periods.

$$\tau_l = \mathbb{E}[Y_{t^*+l}(1) - Y_{t^*+l}(0) \mid G = 1,\ S = 1], \quad l = 0, 1, \ldots, T - t^*$$

Setup

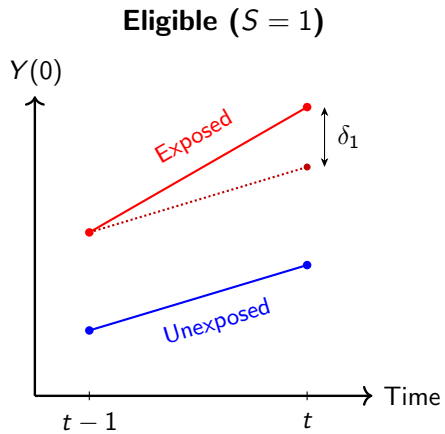


Parameters of interest: ATTs in all post-treatment periods.

$$\tau_l = \mathbb{E}[Y_{t^*+l}(1) - Y_{t^*+l}(0) \mid G = 1, S = 1]\,.$$

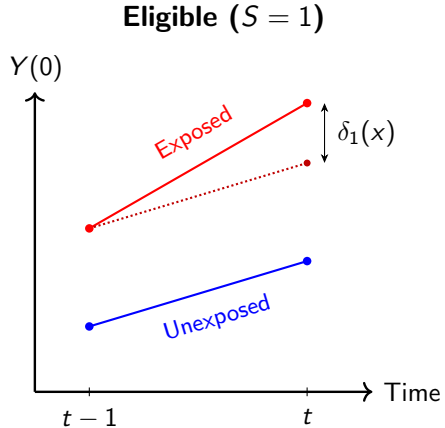
Identification Assumption: Conditional Parallel Gaps

- Allow violation of parallel trends: $\delta_1 \neq 0$.



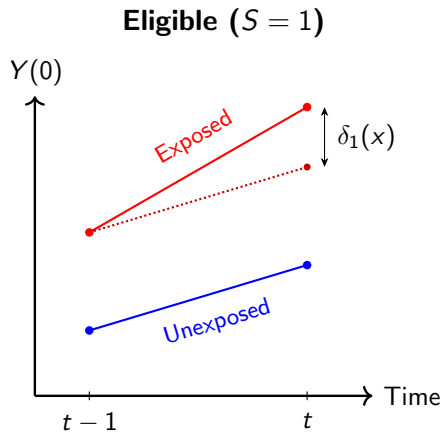
Identification Assumption: Conditional Parallel Gaps

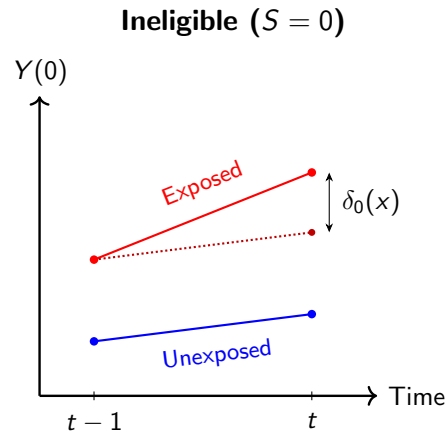
- Let X denote a q-dimensional vector of covariates.
- Allow the violation to depend on the covariate value x: $\delta_1(x)$.



Identification Assumption: Conditional Parallel Gaps

- Conditional parallel gaps: $\delta_1(x) = \delta_0(x)$. Formula

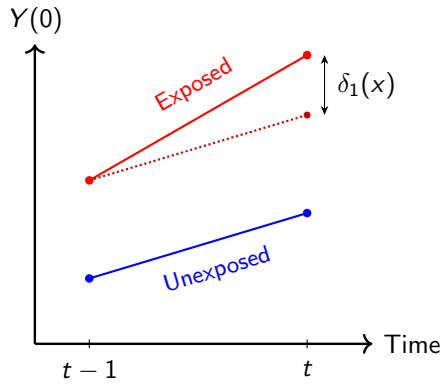




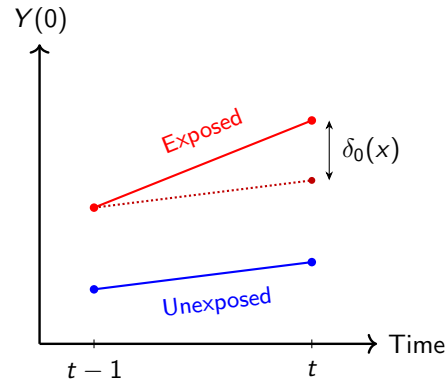
Identification Assumption: Conditional Parallel Gaps

- Conditional parallel gaps: $\delta_1(x) = \delta_0(x)$. Formula
- Conditional parallel trends: $\delta_1(x) = 0$.

Eligible ($S = 1$)



Ineligible ($S = 0$)



Identification

Theorem: Identification

Let $\Delta Y_l = Y_{t^*+l} - Y_{t^*-1}$. Then

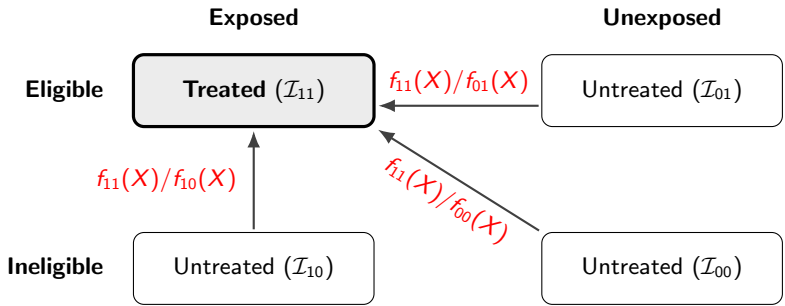
$$\tau_l = \mathbb{E}[\Delta Y_l \mid G = 1, S = 1] - \mathbb{E}\left[\frac{f_{11}(X)}{f_{01}(X)}\Delta Y_l \,\middle|\, G = 0, S = 1\right] - \mathbb{E}\left[\frac{f_{11}(X)}{f_{10}(X)}\Delta Y_l \,\middle|\, G = 1, S = 0\right] + \mathbb{E}\left[\frac{f_{11}(X)}{f_{00}(X)}\Delta Y_l \,\middle|\, G = 0, S = 0\right],$$

The treated subgroup is left unweighted, and each untreated subgroup (g, s) is reweighted by the density ratio $f_{11}(X)/f_{gs}(X)$.

Technical Conditions

Density Ratios and Covariate Balancing Condition

- The untreated subgroups are reweighted because the identification assumption are conditional on covariates.
- The density ratios reweight the covariate distribution of each untreated subgroup to match that of the treated subgroup.

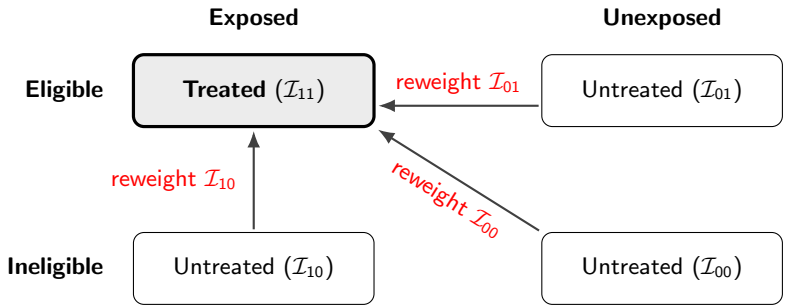


Density Ratios and Covariate Balancing Condition

Covariate Balancing Condition

An outcome weighted estimand equals the ATT if its weighting profile:

1. leaves the treated subgroup unweighted;
2. reweights each untreated subgroup so that its covariate distribution matches that of the treated subgroup.



Road Map

- Setup and Identification
- **Limitations of 3WFE with Covariates**
- Matched Triple-Differences Framework
- Empirical Application
- Conclusion

Empirical Practice: Two 3WFE Specifications

- Common empirical practice: includes the covariates in 3WFE regressions.
- **Covariate-Trend 3WFE Regression:**

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \beta_l^{CT} G_i S_i \mathbb{1}\{t - t^* = l\} + \sum_{r=1}^{T} X_i' \theta_r \mathbb{1}(t = r) + u_{it}.$$

- **Covariate-Level 3WFE Regression:**

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \beta_l^{CL} G_i S_i \mathbb{1}\{t - t^* = l\} + X_i' \theta + u_{it}.$$

Diagnosis of Empirical Practice

- Both OLS estimands, β_l^{CT} and β_l^{CL}, can be expressed as outcome weighted estimands.
- I diagnose these estimands in two steps:
  - Check whether their weighting profiles satisfy the covariate balancing condition.
  - Characterize the additional conditions under which each estimand equals the ATT.
- Main takeaway: **Neither** 3WFE regressions consistently estimate the ATT.

Covariate-Trend 3WFE

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \beta_l^{CT} G_i S_i \mathbb{1}\{t - t^* = l\} + \sum_{r=1}^{T} X_i' \theta_r \mathbb{1}(t = r) + u_{it}.$$

- **Corresponding regression estimand:**

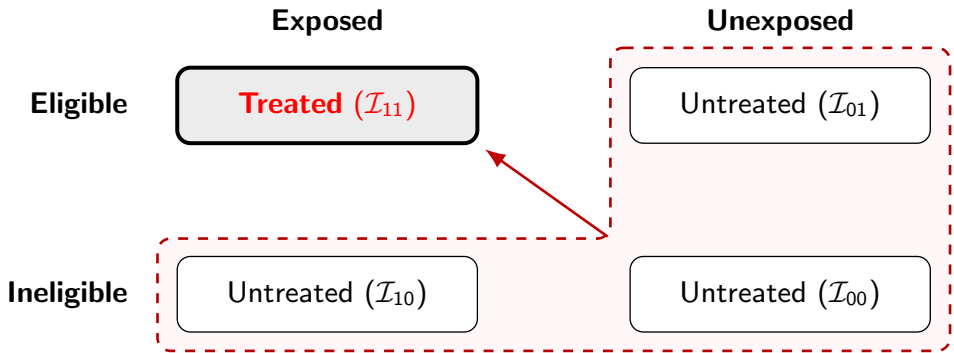
$$\beta_l^{CT} = \mathbb{E}\left[\frac{p_{11}\,(1 - L(D \mid G, S, X))}{\kappa}\,\Delta Y_l \;\middle|\; G = 1, S = 1\right] - \mathbb{E}\left[\frac{p_{01}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \;\middle|\; G = 0, S = 1\right]$$
$$- \mathbb{E}\left[\frac{p_{10}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \;\middle|\; G = 1, S = 0\right] + \mathbb{E}\left[\frac{-p_{00}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \;\middle|\; G = 0, S = 0\right],$$

  where $L(D \mid G, S, X)$ is the linear projection.

- All four subgroups are reweighted by **implied regression weights**.
- This weighting profile fails the covariate balancing condition in three ways.

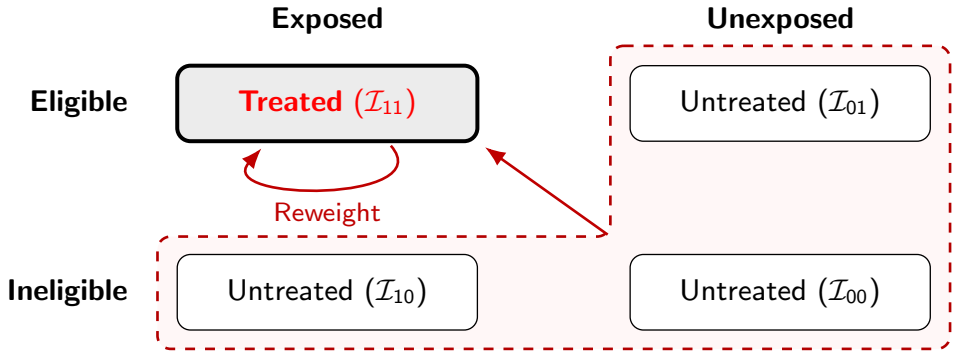
Covariate-Trend 3WFE: Pooling

- Targets covariate balance between the treated subgroup and the **pooled untreated group**, rather than each untreated subgroup separately.



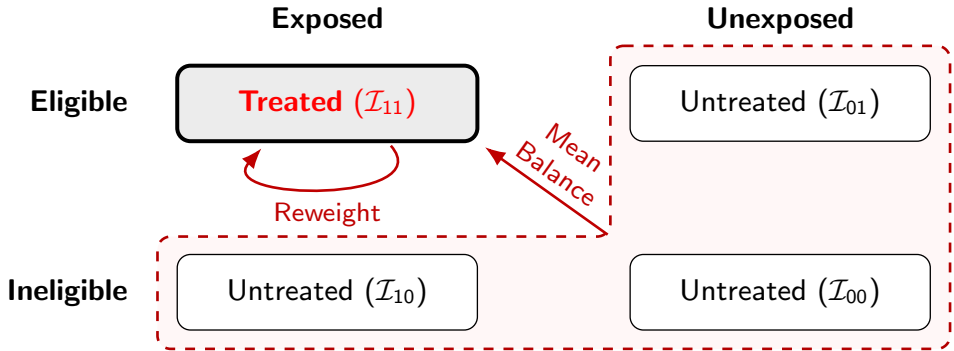
Covariate-Trend 3WFE: Treated Reweighting

- Reweights the **treated subgroup**.



Covariate-Trend 3WFE: Mean-only Balance

- Only guarantees balance of the **means**, rather than **full distributions**.



When does covariate-trend 3WFE recover the ATT?

Lemma: Sufficient Conditions for $\beta_I^{CT} = \tau_I$

1. **Treatment-effect homogeneity in X:** $\tau_I(X) = \tau_I$.
2. **Either** of the following conditions holds:
   - 2.1 **Linear conditional means with a common slope.**

$$\mathbb{E}[\Delta Y_I \mid G = g, S = s, X] = \gamma_{gs} + X'\gamma, \quad (g, s) \in \{(0,0), (0,1), (1,0)\}.$$

   - 2.2 **Inverse linear form for generalized propensity scores.**

$$P(G = g, S = s \mid X) = \frac{\left[\lambda_{gs} + (-1)^{g+s} X'\lambda\right]^{-1}}{\sum_{g=0}^{1}\sum_{s=0}^{1}\left[\lambda_{gs} + (-1)^{g+s} X'\lambda\right]^{-1}}.$$

Further Detail

Covariate-Level 3WFE

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \beta_l^{CL} G_i S_i \mathbb{1}(t - t^* = l) + X_i' \theta + u_{it}$$

- **Corresponding Regression Estimand:**

$$\begin{aligned} \beta_l^{CL} = \mathbb{E}[\Delta Y_l \mid G = 1, S = 1] - \mathbb{E}[\Delta Y_l \mid G = 0, S = 1] \\ - \mathbb{E}[\Delta Y_l \mid G = 1, S = 0] + \mathbb{E}[\Delta Y_l \mid G = 0, S = 0] \end{aligned}$$

- Fails the covariate balancing condition: all three untreated subgroups remain unweighted.
- $\widehat{\beta}_l^{CL}$ numerically equals the 3WFE estimate without covariates, even though X_i has regression coefficients.

When does covariate-level 3WFE recover the ATT?

Lemma: Necessary and sufficient condition for $\beta_l^{CL} = \tau_l$

The parallel gaps assumption holds **unconditionally**:

$$\begin{aligned}&\mathbb{E}[\Delta Y_l(0) \mid G = 1,\ S = 1] - \mathbb{E}[\Delta Y_l(0) \mid G = 0,\ S = 1] \\ &= \mathbb{E}[\Delta Y_l(0) \mid G = 1,\ S = 0] - \mathbb{E}[\Delta Y_l(0) \mid G = 0,\ S = 0]\end{aligned}$$

Under unconditional parallel gaps, there is no need to incorporate covariates into a triple-differences design.

Road Map

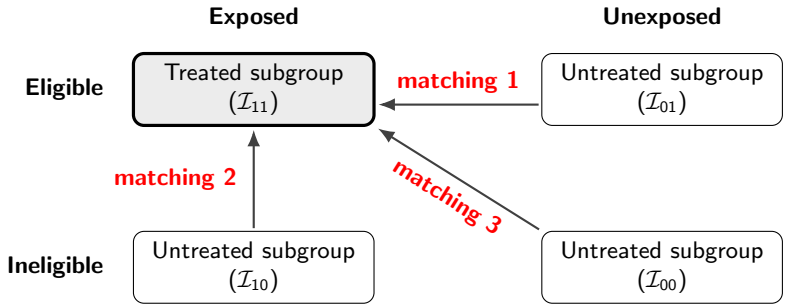
- Setup and Identification
- Limitations of 3WFE with Covariates
- **Matched Triple-Differences Framework**
- Empirical Application
- Conclusion

What's next?

- The covariate-trend and covariate-level 3WFE are inconsistent in general.
- I will propose an alternative matched triple-differences framework: consistent estimator + asymptotic theory + inference.

Estimation: Matching

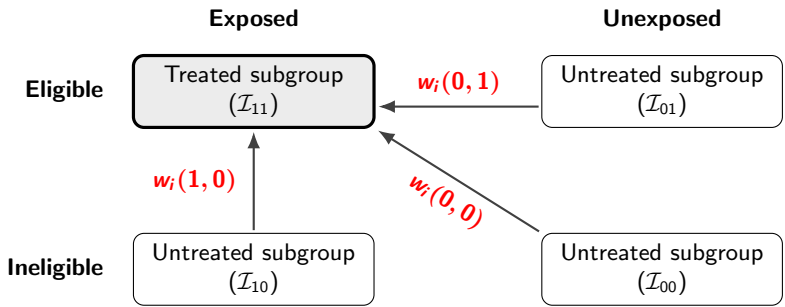
- **Three pairwise matchings**: In each, the treated group is matched with one of the three untreated subgroups.



Technical Detail

Estimation: Matching

- **Three pairwise matchings**: In each, the treated group is matched with one of the three untreated subgroups.
- **Matching weights**: number of times each untreated unit is used as a match.



Technical Detail

Estimation: Weighted 3WFE Regression

- Run a weighted 3WFE regression:

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \tau_l^{match} G_i S_i \mathbb{1}(t - t^* = l) + u_{it},$$

  with weights w_i
  - Each treated unit is unweighted: $w_i = 1$.
  - Each untreated unit is reweighted by its matching weights: $w_i = w_i(g, s)$.
- **Matched triple-differences estimator**: the coefficient of the triple interaction term $G_i S_i \mathbb{1}(t - t^* = l)$.

Matched Triple-Differences Estimator: $\widehat{\tau}_l^{match}$

- The estimator mirrors the structure of the identification result.

$$\widehat{\tau}_l^{match} = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{01}} w_i(0,1) \Delta Y_{il}$$

$$- \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{10}} w_i(1,0) \Delta Y_{il} + \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{00}} w_i(0,0) \Delta Y_{il}.$$

$$\tau_l = \mathbb{E}[\Delta Y_l \mid G{=}1, S{=}1] - \mathbb{E}\left[\frac{f_{11}(X)}{f_{01}(X)} \Delta Y_l \mid G{=}0, S{=}1\right]$$

$$- \mathbb{E}\left[\frac{f_{11}(X)}{f_{10}(X)} \Delta Y_l \mid G{=}1, S{=}0\right] + \mathbb{E}\left[\frac{f_{11}(X)}{f_{00}(X)} \Delta Y_l \mid G{=}0, S{=}0\right].$$

Matched Triple-Differences Estimator: $\widehat{\tau}_l^{match}$

- The estimator mirrors the structure of the identification result.

$$\widehat{\tau}_l^{match} = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{01}} w_i(0,1) \Delta Y_{il}$$
$$- \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{10}} w_i(1,0) \Delta Y_{il} + \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{00}} w_i(0,0) \Delta Y_{il}.$$

- Normalized matching weights approximate the density ratios.

$$\underbrace{\frac{n_{gs}}{n_{11}} w_i(g,s)}_{\text{Normalized matching weight}} \approx \underbrace{\frac{f_{11}(X_i)}{f_{gs}(X_i)}}_{\text{Density ratio}} \qquad (g,s) \neq (1,1).$$

Asymptotic Theory

Assumption: Regularity Conditions on Pairwise Matching

The pairwise matching procedure should satisfy:

(i) **Locality.** Untreated units with covariate values closer to those of the treated units receive larger weights.

(ii) **Bounded moments.** The matched untreated subgroup cannot be concentrated on a small number of heavily reused untreated units.

(iii) **Stability.** Changing the covariate value of a single observation cannot substantially change the matching results.

- Our framework includes nearest-neighbor matching and kernel matching.

Asymptotic Theory

Assumption: Regularity Conditions on Pairwise Matching

The pairwise matching procedure should satisfy:

(i) **Locality.**

$$\mathbb{E}\Big[\sum_{j\in\mathcal{I}_{gs}} w_{ij}\|X_j - X_i\| \;\Big|\; i \in \mathcal{I}_{11}\Big] = o(1).$$

(ii) **Bounded moments.** For some $\delta > 0$,

$$\mathbb{E}\Big[w_j^{4+\delta} \;\Big|\; j \in \mathcal{I}_{gs}\Big] = O(1).$$

(iii) **Stability.** Let $w_j^{(1)}$ denote the matching weight recomputed after replacing observation 1 with an independent copy.

$$\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_j - w_j^{(1)}\right|\right)^4\right] = O(1).$$

- Our framework includes nearest-neighbor matching and kernel matching.

Additional Regularity

Asymptotic Theory

Theorem 2: Consistency and Asymptotic Normality

$$\widehat{\tau}_I^{match} - \tau_I \xrightarrow{p} 0,$$

$$(V_{\tau(X)} + V_E)^{-1/2} \sqrt{n_{11}} \left(\widehat{\tau}_I^{match} - \tau_I - B \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Variance Formula

Asymptotic Theory

Theorem 2: Consistency and Asymptotic Normality

$$\widehat{\tau}_l^{match} - \tau_l \xrightarrow{p} 0,$$

$$(V_{\tau(X)} + V_E)^{-1/2} \sqrt{n_{11}} \left(\widehat{\tau}_l^{match} - \tau_l - B \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

- The convergence rate of the estimator depends on the treated subgroup size n_{11}.

Variance Formula

Asymptotic Theory

Theorem 2: Consistency and Asymptotic Normality

$$\widehat{\tau}_l^{match} - \tau_l \xrightarrow{p} 0,$$

$$(V_{\tau(X)} + V_E)^{-1/2}\sqrt{n_{11}}\left(\widehat{\tau}_l^{match} - \tau_l - B\right) \xrightarrow{d} \mathcal{N}(0,1).$$

- **Finite-sample bias:** $B = O_p(n_{01}^{-1/r} + n_{10}^{-1/r} + n_{00}^{-1/r})$.
  - Convergence rate depends on: (i) the untreated subgroup sizes n_{01}, n_{10}, n_{00}; (ii) the number of continuous covariates r.
  - Negligible at the $\sqrt{n_{11}}$ scale: (i) the untreated subgroups are sufficiently large; (ii) there is one continuous covariates ($r = 1$).

Variance Formula

Asymptotic Theory

Theorem 2: Consistency and Asymptotic Normality

$$\widehat{\tau}_I^{match} - \tau_I \xrightarrow{p} 0,$$

$$(V_{\tau(X)} + V_E)^{-1/2} \sqrt{n_{11}} \left(\widehat{\tau}_I^{match} - \tau_I - B \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

- **Bias correction:** B can be estimated using an off-the-shelf estimator, such as a series estimator.
- Under further conditions, $\widehat{\tau}_I^{match} - \widehat{B}$ is $\sqrt{n_{11}}$ consistent and has the same asymptotic variance as $\widehat{\tau}_I^{match} - B$.

Variance Formula

Variance Estimation

- $\widehat{\tau}_l^{match}$ is a weighted least squares estimator: use the cluster-robust standard error of the second-step regression
  - reg ..., vce(cluster id)
- The cluster-robust variance estimator ignores the matching step, leading to either over- and under-estimation of the true standard error.

Variance Estimation

Theorem 3: Consistency of the Variance Estimator

$$\frac{\widehat{V}_{\tau(X)} + \widehat{V}_E}{V_{\tau(X)} + V_E} \xrightarrow{p} 1.$$

- Generalize Abadie and Imbens' (2006) variance-estimation approach to a class of matching procedures.
- Tailored to the three pairwise matching procedures in the triple-differences design.

Variance Estimator Formula Simulation

Road Map

- Setup and Identification
- Limitations of 3WFE with Covariates
- Matched Triple-Differences Framework
- **Empirical Application**
- Conclusion

Empirical Application: Agricultural Insurance (Cai, 2016)

- Rural households in 12 main tobacco-producing counties in Jiangxi Province, China.
  - Agriculture is the main source of household income, with some households relying on tobacco and others on rice and white lotus.
  - Weather risks can reduce agricultural production and cause income losses.

Empirical Application: Agricultural Insurance (Cai, 2016)

- Rural households in 12 main tobacco-producing counties in Jiangxi Province, China.
  - Agriculture is the main source of household income, with some households relying on tobacco and others on rice and white lotus.
  - Weather risks can reduce agricultural production and cause income losses.
- Policy: In 2003, one county (exposed county) offered agricultural insurance to protect **tobacco-growing households** against these income losses.
  - A household receives compensation if an extreme weather event reduces tobacco yield by at least 30 percent.

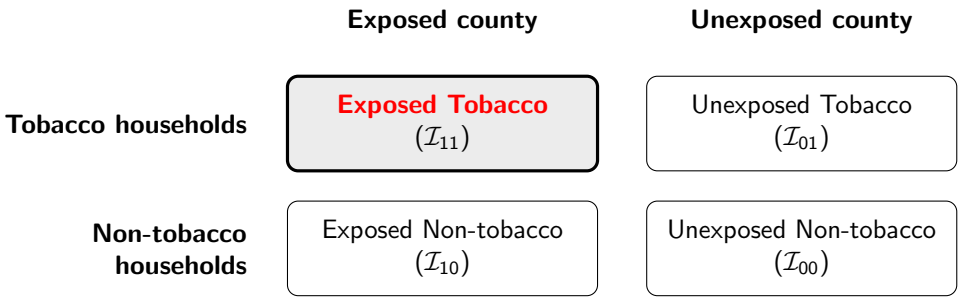
Empirical Application: Agricultural Insurance (Cai, 2016)

- Treatment: A household needs to satisfy two criteria to be offered the insurance:
  - Exposure: Live in the county providing the agricultural insurance program.
  - Eligibility: Grow tobacco.

Empirical Application: Agricultural Insurance (Cai, 2016)

- Treatment: A household needs to satisfy two criteria to be offered the insurance:
  - Exposure: Live in the county providing the agricultural insurance program.
  - Eligibility: Grow tobacco.
- Parameter of Interest: The effect of the insurance program on treated households' saving behavior.
- Dataset: A household-level panel from 2001-2008 covering 5,746 households across the 12 counties
  - Household borrowing, saving, and demographic characteristics

Four Types of Households



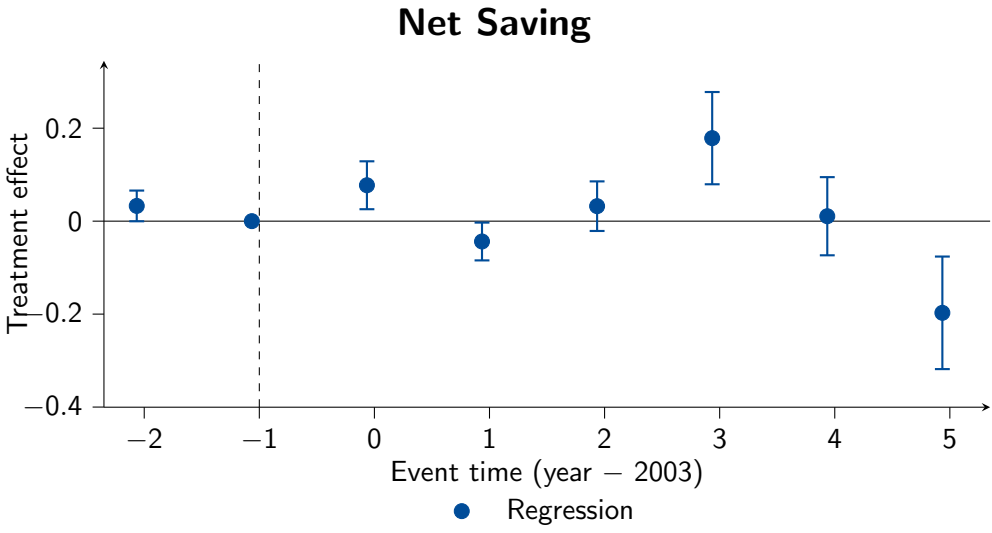
Empirical Application: Agricultural Insurance (Cai, 2016)

- Cai (2016) runs a covariate-level 3WFE regression (baseline year: 2002).

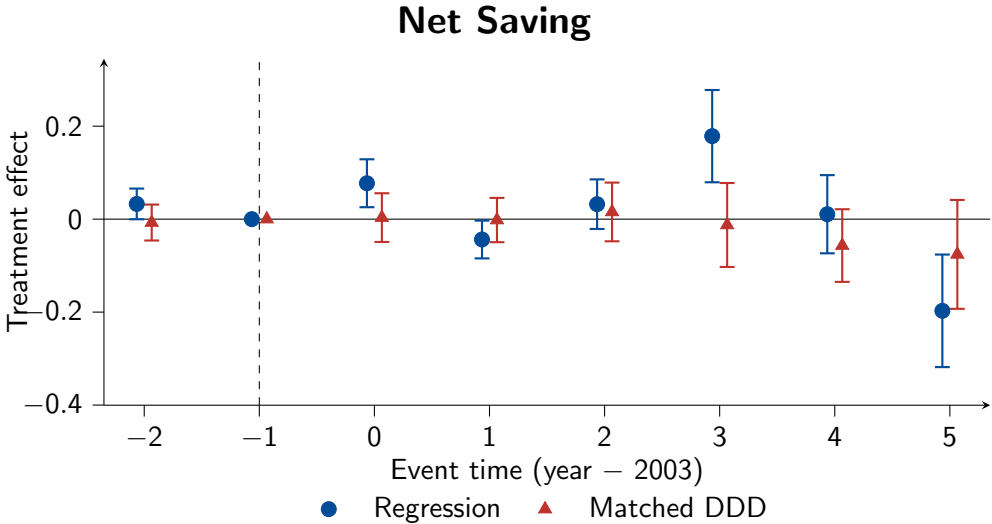
$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \beta_l \, G_i S_i \mathbb{1}(t - t^* = l) + X_i' \theta + u_{it}.$$

- Outcomes (Y_{it}): net saving and flexible-term saving ratio.
- Covariates (X_i): age and education level of the household head, household size in the baseline year.
- I report both event-study estimates from a covariate-level 3WFE regression and a matched-triple differences estimator.

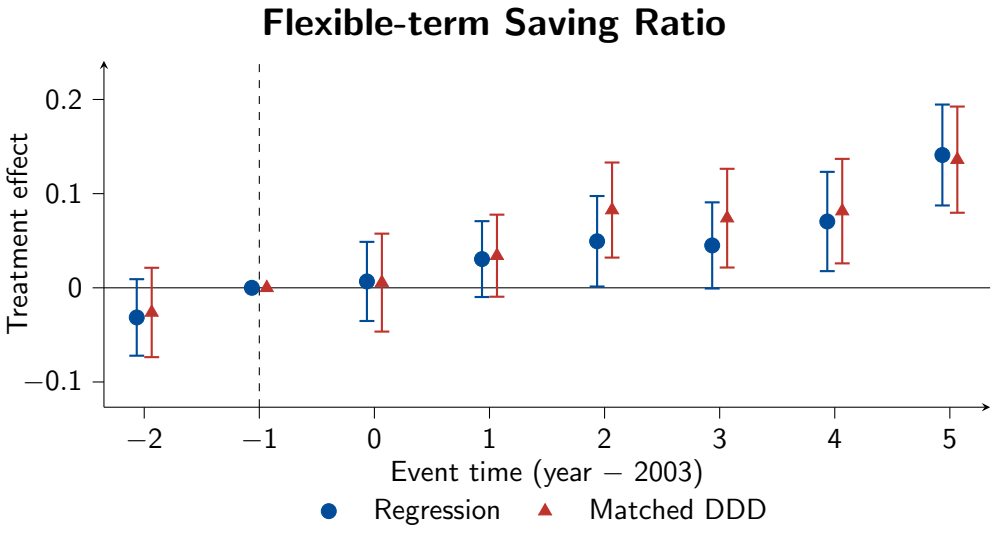
Effect on Net Saving



Effect on Net Saving



Effect on Flexible-term Saving Ratio



Road Map

- Setup and Identification
- Limitations of 3WFE with Covariates
- Matched Triple-Differences Framework
- Empirical Application
- **Conclusion**

Conclusion and Takeaways

- Incorporating covariates into a triple-differences design can make identification more credible, but estimation requires care.

Conclusion and Takeaways

- Incorporating covariates into a triple-differences design can make identification more credible, but estimation requires care.
- Commonly used covariate-trend and covariate-level 3WFE regressions are generally inconsistent for the ATT.
  - They do not satisfy the covariate balancing condition and require additional restrictive conditions for consistency.

Conclusion and Takeaways

- Incorporating covariates into a triple-differences design can make identification more credible, but estimation requires care.
- Commonly used covariate-trend and covariate-level 3WFE regressions are generally inconsistent for the ATT.
  - They do not satisfy the covariate balancing condition and require additional restrictive conditions for consistency.
- I propose a matched triple-differences framework.
  - I establish consistency and asymptotic normality for a class of matching estimators and provide a valid variance estimator for inference.

Conclusion and Takeaways

- Incorporating covariates into a triple-differences design can make identification more credible, but estimation requires care.
- Commonly used covariate-trend and covariate-level 3WFE regressions are generally inconsistent for the ATT.
  - They do not satisfy the covariate balancing condition and require additional restrictive conditions for consistency.
- I propose a matched triple-differences framework.
  - I establish consistency and asymptotic normality for a class of matching estimators and provide a valid variance estimator for inference.
- The empirical application shows that matching can yield different estimates and conclusions from 3WFE regression.

Conclusion and Takeaways

- Incorporating covariates into a triple-differences design can make identification more credible, but estimation requires care.
- Commonly used covariate-trend and covariate-level 3WFE regressions are generally inconsistent for the ATT.
  - They do not satisfy the covariate balancing condition and require additional restrictive conditions for consistency.
- I propose a matched triple-differences framework.
  - I establish consistency and asymptotic normality for a class of matching estimators and provide a valid variance estimator for inference.
- The empirical application shows that matching can yield different estimates and conclusions from 3WFE regression.
- Stata and R packages are in progress: `dddmatch`.

Thank you!

Conditional Parallel Gaps

Assumption: Conditional Parallel Gaps (CPG)

$$\underbrace{\mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid G = 1, S = 1, X] - \mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid G = 0, S = 1, X]}_{\delta_1(X)}$$

$$=$$

$$\underbrace{\mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid G = 1, S = 0, X] - \mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid G = 0, S = 0, X]}_{\delta_0(X)}$$

Back

Additional Identification Assumption

Assumption: No Anticipation

Potential outcomes do not depend on future treatments:

$$Y_t(d_1, \ldots, d_T) = Y_t(d_1, \ldots, d_t)$$

Assumption: Strong Overlap

There exists $\epsilon > 0$ such that

$$\Pr(G = g, S = s \mid X = x) > \epsilon$$

for every $(g, s) \in \{0, 1\}^2$ and almost every $x \in \mathcal{X}$, where $\mathcal{X}$ denotes the support of X.

Back

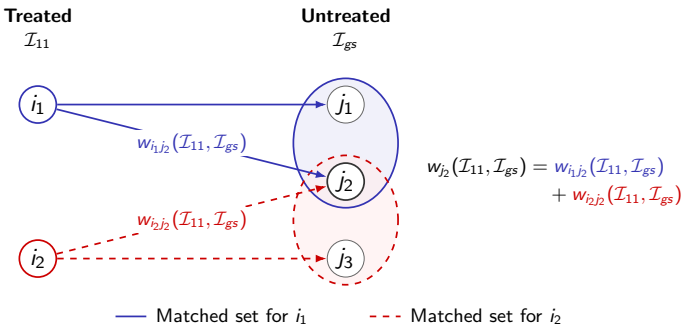
Enriching the Covariate-Trend 3WFE

- These problems are specific to the covariate-trend 3WFE regression, which is widely used in empirical work.
- Each issue can be (partly) addressed by enriching the 3WFE regression:
  - Pooling issue: interact the covariates with the subgroup indicators G_i and S_i.
  - Mean-only balance (linearity) issue: include higher-order covariate terms, such as X_i^2
- This enrichment is a step toward non-parametric regression adjustment.
- My proposed solution: matching is intuitively appealing, easy to implement, and fully non-parametric.

Back

Matching Weights

- Pairwise weight $w_{ij}(g, s)$: contribution of unit j to the matched set of unit i.
  - Usually a function of the covariate distance between i and j.
- **Matching weight**: $w_j(g, s) = \sum_{i \in \mathcal{I}_{11}} w_{ij}(g, s)$.



Back

Additional Regularity Conditions for Asymptotic Theory

Assumption: Random Sampling

$\{Y_{i1}, Y_{i2}, \ldots, Y_{iT}, G_i, S_i, X_i\}_{i=1}^n$ is a random sample from the population.

Assumption: Regularity Conditions

For all $(g, s) \in \{0,1\}^2$, $d \in \{0,1\}$, $t = 1, \ldots, T$, and event times l:

(i) **Conditional density.** The conditional density of X has compact convex support and is continuous, bounded, and bounded away from zero on its support.

(ii) **Conditional means.** $\mathbb{E}[Y_t(d) \mid G = g, S = s, X = x]$ is Lipschitz continuous in x.

(iii) **Conditional variances.** Lipschitz continuous in x, with uniform bounds

$$0 < \underline{\sigma}^2 \leq \text{Var}[\Delta Y_l(d) \mid G = g, S = s, X = x] \leq \overline{\sigma}^2 < \infty.$$

(iv) **Fourth moments.** $\mathbb{E}[Y_t(d)^4 \mid G = g, S = s, X = x]$ is uniformly bounded.

Back

Formula for the asymptotic variance

Asymptotic Variance

The asymptotic variance of the matched triple-differences estimator takes the form of:

$$V_{\tau(X)} = \mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right)^2 \middle| G_i = 1, S_i = 1\right],$$

$$V_E = \frac{1}{n_{11}}\left(\sum_{i \in \mathcal{I}_{11}} \sigma^2_{11,l}(X_i) + \sum_{(g,s)\neq(1,1)} \sum_{j \in \mathcal{I}_{gs}} w_j^2 \sigma^2_{gs,l}(X_j)\right).$$

Back

Formula for the asymptotic variance estimator

Asymptotic Variance Estimator

The asymptotic variance estimator of the matched triple-differences estimator takes the form of:

$$\widehat{V}_{\tau(X)} = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\Delta Y_{il} - \widehat{\Delta Y}_{il}^{01} - \widehat{\Delta Y}_{il}^{10} + \widehat{\Delta Y}_{il}^{00} - \widehat{\tau}_l \right)^2,$$

$$\widehat{V}_E = \frac{1}{n_{11}} \sum_{(g,s) \neq (1,1)} \sum_{j \in \mathcal{I}_{gs}} \left(w_j^2 - \widetilde{w}_j^2 \right) \widehat{\sigma}_{gs,l}^2(X_j).$$

Back

Simulation

- Four periods $t = 1, \ldots, 4$, four subgroups $(G_i, S_i) \in \{0,1\}^2$, $n = 5,000$
- Covariates $X_i = (X_{i1}, X_{i2})'$ are generated by:

$$X_{ij} = \xi_i \frac{|\zeta_{ij}|}{\|\zeta_i\|}, \quad \xi_i \sim \text{Uniform}(0,1), \quad \zeta_i \sim N(0, I_2), \quad \|X_i\| \sim \textit{Uniform}[0,1]$$

- Generalized Propensity score:

$$P(G_i = g, S_i = s|X_i) = \frac{exp(\alpha_{gs} + \gamma_{gs}(\|X_i\| - 0.5))}{\sum_{(g,s)} exp(\alpha_{gs} + \gamma_{gs}(\|X_i\| - 0.5))}$$

Simulation

- Small $(1,0)$ subgroup: $\alpha_{11} = 0, \alpha_{10} = -0.776, \alpha_{01} = 0.228, \alpha_{00} = 0.270$.
- Selection on the eligibility dimension: $\gamma_{11} = 0, \gamma_{10} = -2, \gamma_{01} = -0.5, \gamma_{00} = -3.0$
- Potential Outcomes

$$Y_{it}(0) = \alpha_i + 2G_i(\|X_i\| - 0.5)t + S_i(\|X_i\| - 0.5)t + v_{it}(0)$$

$$Y_{it}(1) = 1 + Y_{it}(0) + v_{it}(1) - v_{it}(0)$$

- Homogeneous treatment effect, $ATT = 1$.

Simulation

Table: Small $(1, 0)$ Subgroup, Homogeneous Treatment Effect

	Bias	MCSE	Avg. SE	Coverage
3WFE, spec. 1 $(X'\theta_t)$	**0.8118**	0.1102	0.1102	0.000
3WFE, spec. 2 $(X'\theta)$	**0.7087**	0.1253	0.1242	0.001
NN, $K = 1$	0.0134	0.1532	0.1496	0.951
bias-corrected	0.0011	0.1542	0.1496	0.947
NN, $K = n_{gs}^{1/3}$	0.0461	0.1307	0.1257	0.922
bias-corrected	−0.0003	0.1346	0.1257	0.942
Kernel	0.0206	0.1394	0.1353	0.942
bias-corrected	0.0020	0.1408	0.1353	0.941

Back