A Design-Based Matching Framework for Staggered Adoption with Time-Varying Confounding

Suehyun Kim [1] Dahai Jung [2] Kwonsang Lee [1]

[1]Seoul National University [2]Sungkyunkwan University

Motivating example: Causal impact of Netflix on IPTV viewing

- Motivated by the rapid expansion of OTT services, we address two empirical questions: **(i) whether Netflix subscription causally affects IPTV viewing behavior** and **(ii) whether the magnitude and dynamics of effects differ by subscription timing and duration.**
- We propose a novel framework for causal inference in staggered adoption settings that enables **design-based analysis with reduced model and assumption dependence** and **simultaneous inference of causal effects across different treatment groups and timings.**

Notation and causal framework

Notation

- We consider a panel of N units over discrete time periods $t = T_0, \dots, 0, 1, \dots, T$.
  - The period $t = 1$ marks the earliest possible treatment.
  - Time-varying confounders may be measured in the pre-treatment periods $t = T_0, \dots, 0$.
- For each unit $i = 1, \dots, N$, we observe $(Y_{i1}, \dots, Y_{iT}, Z_{i1}, \dots, Z_{iT}, \mathbf{X}_{iT_0}, \dots, \mathbf{X}_{iT})$.
  - Z_{it}: Binary treatment indicator at time t.
  - Y_{it} and $\mathbf{X}_{it}$: Outcome and covariates at time t.
- **Staggered treatment adoption**: We assume that once a unit receives treatment, it remains treated thereafter, i.e., for $t = T_0, \dots, 0$, $Z_t = 0$, and for $t = 2, \dots, T$,

$$Z_{t-1} = 1 \implies Z_t = 1.$$

  - Let $G_i \in \{1, 2, \dots, T, \infty\}$ denote the period in which unit i first receives treatment, with $G_i = \infty$ indicating that the unit is never treated.
  - Under staggered adoption, the treatment path is identified by G; we refer to $\{i : G_i = g\}$ as the **cohort** (or **group**) initiating treatment at time g. The **potential outcomes** $Y_{it}(g)$ are defined as functions of g.

Assumptions

(i) **No anticipation**: For all $g \in \{1, \dots, T, \infty\}$ and $t < g$, $Y_{it}(g) = Y_{it}(\infty)$.

(ii) **Time-specific unconfoundedness**: For each $g = 1, \dots, T$ and potential treatment adoption time $t \geq g$,

$$(Y_{it}(g), Y_{it}(\infty)) \perp Z_{ig} \mid (\mathbf{X}_{iT_0}, \dots \mathbf{X}_{i,g-1}).$$

(iii) Other assumptions: **SUTVA**, **positivity**. Note that we do not impose the *parallel trends assumption*, which is a relatively strong assumption required by DiD-based frameworks.

Causal estimand: Group-time average treatment effect $ATT(g, t)$

Following Callaway and Sant'Anna (2021), we target the **group-time average treatment effect** for $G = 1, \dots, T$ and $t = 1, \dots, T$:

$$ATT(g, t) := \mathbb{E}[Y_{it}(g) - Y_{it}(\infty) \mid G_i = g].$$

Under the assumptions above, $ATT(g, t)$ is identified as

$$ATT(g, t) = \mathbb{E}[\mathbb{E}[Y_{it} \mid G_i = g, \mathbf{X}_{i,g-1}] - \mathbb{E}[Y_{it} \mid G_i > t, \mathbf{X}_{i,g-1}]].$$

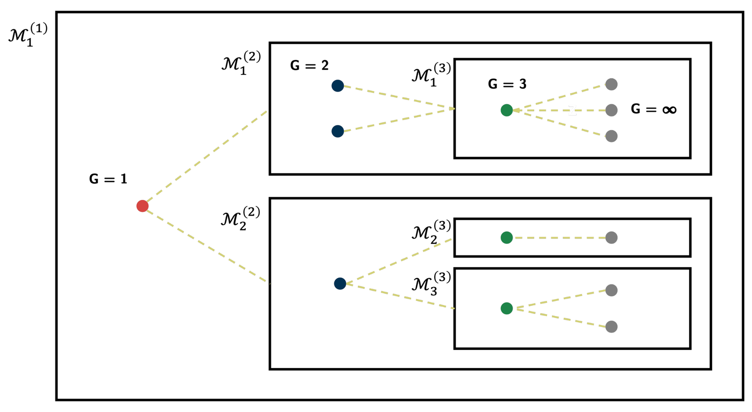
Proposed design

Nested design

- The identification of $ATT(g, t)$ requires conditioning on $\mathbf{X}_{T_0:(g-1)} := (\mathbf{X}_{T_0}, \dots, \mathbf{X}_{g-1})$, so that if $g < g'$, then

$$\mathbf{X}_{T_0:(g-1)} \subset \mathbf{X}_{T_0:(g'-1)}.$$

- This induces a natural **nested structure of time-varying covariates** that *evolves more finely over time and resembles an SRE* at a given cross-section.



Estimation and inference of $ATT(g, t)$ under the nested design

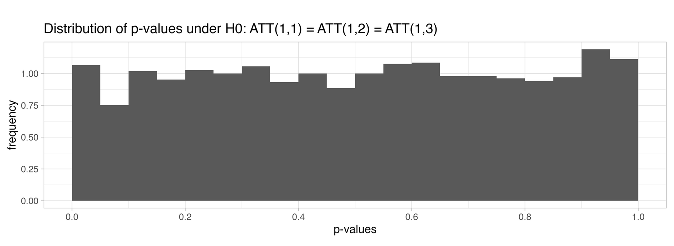
- **Estimation**: For a given (g, t), extract the stratified structure and compute

$$\widehat{ATT}(g, t) = \sum_{m=1}^{n_g} \omega_{m,g} \left(\bar{Y}^{(T)}_{m,g,t} - \bar{Y}^{(C)}_{m,g,t} \right).$$

- **Inference**: We adopt a **block-level bootstrap** by resampling the outermost strata with replacement.
  - This enables the **estimation of the covariance matrix of the $ATT(g, t)$s** and **hypothesis testing of the form $H_0 : R\tau = 0$** using a Wald-type statistic.

Simulation: 2,300 MC simulations with $B = 1{,}000$ bootstrap iterations each, $n = 100$ initial strata over three time points.

SE estimates	$ATT(1,1)$	$ATT(1,2)$	$ATT(1,3)$
MC	0.0086	0.0098	0.0133
Bootstrap means	0.0083	0.0097	0.0143



Matching algorithm

Reverse-Time Nested Matching (RTNM)

We propose a novel matching algorithm, *Reverse-Time Nested Matching*, that reconstructs the nested design from longitudinal observational datasets.

Step 1 (Initial step). Optimally match units from cohort $\{i : G_i = G\}$ to $\{i : G_i > G\}$ to form matched sets $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$.

Step 2 (Matching with pseudo-controls). Move to the previous cohort $G - 1$, and match units from cohort $\{i : G_i = G - 1\}$ to matched sets $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$ as follows:

(i) Compute the distance matrix between the treated cohort $\{i : G_i = G - 1\}$ and the not-yet-treated cohort $\{i : G_i > G - 1\}$, based on a prespecified metric d.

(ii) Based on the distance matrix from (i), compute the distance from each unit in the treated cohort $\{i : G_i = G - 1\}$ to the matched sets $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$.

(iii) Using this distance, optimally match units in $\{i : G_i = G - 1\}$ to $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$ to obtain matched sets $\mathcal{M}_1^{(G-1)}, \dots, \mathcal{M}_{n_{G-1}}^{(G-1)}$.

Step 3 (Iteration). Repeat Steps 1-2 to match $\{i : G_i = g\}$ with the previously-matched $\mathcal{M}_1^{(g+1)}, \dots, \mathcal{M}_{n_{g+1}}^{(g+1)}$ until the first cohort $\{i : G_i = 1\}$ is reached.

Data application

The Netflix-IPTV dataset

- Monthly panel data observed from March 2021 to November 2021.
- **Cohorts of interest**: June 2021 ($G = 1$) to September 2021 ($G = 4$).
  - March 2021 to May 2021: Data used for covariate adjustment.
  - October 2021 to November 2021: For comparisons with $t \geq g$.

G	1 (Jun)	2 (Jul)	3 (Aug)	4 (Sep)	∞
Count	237	360	302	838	7890

- **Treatment** (Z_t): Netflix subscription status at time t.
- **Outcomes** (Y_t): Total real-time / VoD viewing hours, VoD viewing status.
- **Covariates** ($\mathbf{X}_t$): Nine time-varying covariates, including total and genre-specific real-time and VoD viewing hours, and purchase history.

Causal impact of Netflix subscription on IPTV viewing behavior

(g,t)	1	2	3	4	5	6
1	-1.635 (5.881)	-0.145 (6.514)	-1.603 (5.971)	2.026 (5.927)	-9.733 (5.786)	-7.500 (5.759)
2	–	-3.561 (4.276)	-5.389 (5.086)	0.819 (5.161)	-1.149 (5.530)	-9.254 (4.973)
3	–	–	5.232 (5.087)	-0.851 (5.166)	-1.691 (5.083)	-0.590 (6.154)
4	–	–	–	1.613 (2.600)	-3.428 (3.334)	2.388 (3.510)

(a) Total real-time viewing hours

(g,t)	1	2	3	4	5	6
1	-1.281 (1.189)	**-4.453 (1.492)**	**-4.236 (1.456)**	**-3.443 (1.498)**	**-3.286 (1.492)**	**-3.207 (1.376)**
2	–	**-4.346 (0.985)**	**-4.000 (1.111)**	**-3.729 (1.049)**	**-4.823 (0.844)**	**-2.371 (0.948)**
3	–	–	**-2.900 (0.905)**	-1.782 (0.966)	-0.716 (1.239)	-1.183 (1.189)
4	–	–	–	**-1.641 (0.587)**	**-1.647 (0.625)**	**-2.259 (0.685)**

(b) Total VoD viewing hours

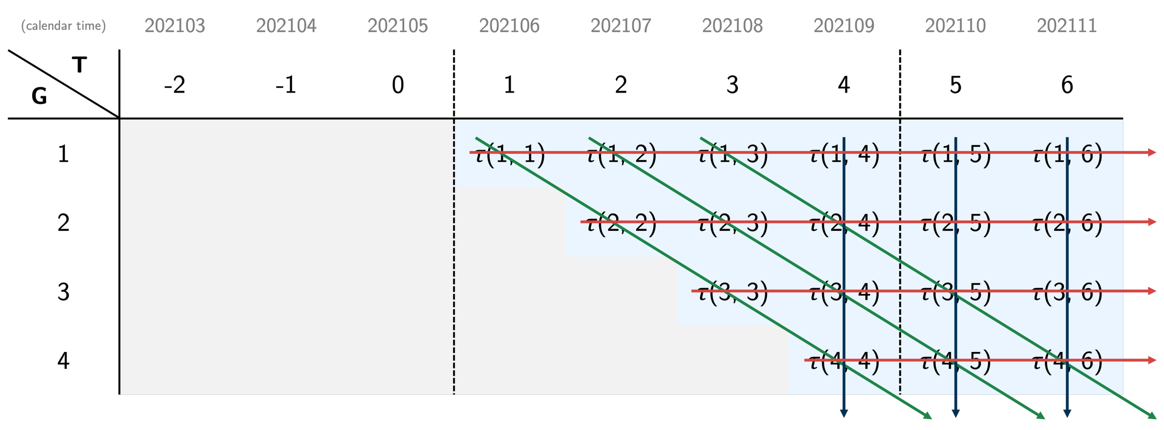
(g,t)	1	2	3	4	5	6
1	-0.012 (0.030)	-0.050 (0.032)	-0.036 (0.030)	-0.025 (0.029)	-0.030 (0.027)	**-0.078 (0.030)**
2	–	**-0.086 (0.032)**	**-0.080 (0.024)**	**-0.087 (0.028)**	**-0.088 (0.026)**	**-0.067 (0.025)**
3	–	–	-0.030 (0.024)	**-0.054 (0.027)**	-0.048 (0.027)	-0.001 (0.027)
4	–	–	–	0.006 (0.013)	**-0.034 (0.015)**	**-0.060 (0.015)**

(c) VoD viewing status

Table 1. Point estimates and standard errors of $ATT(g, t)$ for outcome variables. Bold values denote statistical significance at the $\alpha = 0.05$ level.

- Whereas Netflix subscription does not significantly affect total viewing hours per se, it has a significant negative impact on VoD viewing behavior.
- This implies that Netflix serves as an effective substitute for traditional VoD content.

Tests for homogeneity of $ATT(g, t)$



(i) **Fixed g (red)**: Temporal stability within cohort

(ii) **Fixed t (blue)**: Cross-cohort homogeneity at time t

(iii) **Fixed e (green)**: Constant e-lag effect

	Total real-time hours	Total VoD hours	VoD viewing status
$H_{0,g=1}$	0.3126	0.0718	0.5094
$H_{0,g=2}$	0.1636	0.0520	0.9296
$H_{0,g=3}$	0.6122	0.2960	0.1836
$H_{0,g=4}$	0.0734	0.5308	0.0006***
$H_{0,t=4}$	0.9780	0.1952	0.0056**
$H_{0,t=5}$	0.6842	0.0096**	0.3134
$H_{0,t=6}$	0.2032	0.7120	0.1386
$H_{0,e=0}$	0.5948	0.0608	0.0450*
$H_{0,e=1}$	0.8806	0.1048	0.4640
$H_{0,e=2}$	0.8938	0.1372	0.6278

Table 2. P-values from hypothesis tests assessing the homogeneity of $ATT(g, t)$ for the three outcome variables. The bootstrap-based tests are conducted with $B = 5{,}000$ iterations.

Discussion

- In this work, we introduce a novel design that identifies and estimates group-time average treatment effects, together with the RTNM algorithm for design-based analysis of observational panel data.
- We apply the proposed framework to assess the causal impact of Netflix subscription on IPTV viewing behavior.
- Future research directions include (i) extending the design to more complex treatment regimes and (ii) establishing the optimality properties of the matching algorithm.

References

[1] Brantly Callaway and Pedro H.C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. Themed Issue: Treatment Effect 1.

[2] Kosuke Imai, In Song Kim, and Erik H. Wang. Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 67(3):587–605, 2023.

[3] Yunfei Paul Li, Kathleen J. Propert, and Paul R. Rosenbaum. Balanced risk set matching. *Journal of the American Statistical Association*, 96(455):870–882, 2001.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00573615).

suehyunkim@snu.ac.kr