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

EuroCIM 2026

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

[1]Department of Statistics, Seoul National University
[2]Department of Statistics, Sungkyunkwan University

15 April 2026

Overview

1. Motivation
2. Causal Framework and Design for Longitudinal Data
3. Matching Algorithm
4. Data Application: Netflix-IPTV Dataset
5. Summary

Overview

Motivation: Causal impact of Netflix on IPTV viewing

The outbreak of COVID-19 in the early 2020s led to a rapid shift in media consumption patterns.

IPTV (Internet Protocol TV)		OTT (Over-the-Top) Services
Telecom-provided live TV + VoD (video-on-demand) replay	$\longrightarrow$	Subscription-based streaming on personal devices

Examples of OTT platforms:

Motivation: Causal impact of Netflix on IPTV viewing

Motivated by the expansion of OTT services, we aim to address the following questions:

1. **Does Netflix subscription causally affect IPTV viewing behaviour?**
   - Would Netflix serve as a substitute for television viewing as a whole, or only for certain components (such as VoD services)?
2. **If so, does the causal effect vary by subscription timing and duration?**
   - The release of *Squid Game* in September 2021 led to a surge in new subscriptions. Do these shock-induced entrants respond differently compared to early adopters?

Our work

- We introduce a design-based framework for causal inference under staggered adoption which enables transparent and intuitive analysis for cohort- and time-specific effects.
- We propose a novel matching algorithm to implement the design in observational time series datasets.

Overview

Causal framework for longitudinal data

Notation

- We consider a panel of N units over discrete time periods $t = T_0, \ldots, 0, 1, \ldots, T$.
  - The period $t = 1$ indicates the earliest possible treatment.
  - Time-varying confounders may be measured in the pre-treatment periods $t = T_0, \ldots, 0$.
- For each unit $i = 1, \ldots, N$, we observe $(Y_{i1}, \ldots, Y_{iT}, Z_{i1}, \ldots, Z_{iT}, \mathbf{X}_{iT_0}, \ldots, \mathbf{X}_{iT})$.
  - Z_{it}: Binary treatment indicator at time t.
  - Y_{it} and $\mathbf{X}_{it}$: Outcome and covariates at time t.

Causal framework for longitudinal data

Staggered treatment adoption: We assume that once a unit receives treatment, it remains treated thereafter, i.e., for $t = 2, \ldots, T$,

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

- Let $G_i \in \{1, \ldots, T, \infty\}$ denote the period in which unit i first receives treatment, with $G_i = \infty$ indicating that the unit is never treated.
- 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 for a given time point t.

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

Following Callaway & Sant'Anna (2021), our causal estimand of interest is the **group-time average treatment effect** $ATT(g, t)$.

Definition (Group-time average treatment effect)

The group-time average treatment effect for cohort $\{i : G_i = g\}$ at time t is defined as

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

Such a causal estimand does *not* impose restrictions on heterogeneity of causal effects across different groups and treatment timings. In turn, several interesting questions arise:

(i) **Fixed cohort** g: *Does the causal effect evolve over time?*

(ii) **Fixed outcome timing** t: *Do early- and late-adopters differ?*

(iii) **Fixed event time** e: *Are the effects homogeneous with respect to the lagged period $e = t - g$?*

Identification of $ATT(g, t)$

In order to identify $ATT(g, t)$, we impose the following set of assumptions:

- **No anticipation**: For all $g \in \{1, \dots, T, \infty\}$ and $t < g$, we have $Y_{it}(g) = Y_{it}(\infty)$.
- **Sequential ignorability**: For each $g = 1, \dots, T$ and potential treatment adoption time $t \geq g$,

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

- Other assumptions: SUTVA, positivity.

Proposition (Identification of $ATT(g, t)$)

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

$$ATT(g, t) = \mathbb{E}[\mathbb{E}[Y_{it} \mid G_i = g, \mathbf{X}_{T_0:(g-1)}] - \mathbb{E}[Y_{it} \mid G_i > t, \mathbf{X}_{T_0:(g-1)}]].$$

Remark. Other identification strategies are also possible, such as using a DiD-based approach with a parallel trends assumption.

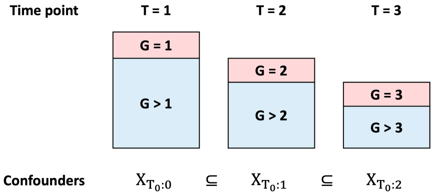
Nested stratification design

- Identification of $ATT(g,t)$ requires conditioning on $\mathbf{X}_{T_0:(g-1)} = (\mathbf{X}_{T_0}, \ldots, \mathbf{X}_{g-1})$, such that if $g < g'$, then
$$X_{T_0:(g-1)} \subseteq X_{T_0:(g'-1)}.$$

- If sequential randomization experiment were feasible, one would stratify units according to time-varying covariates at each time point.
- Under staggered adoption, this induces a **nested structure of time-varying covariates** that
  (i) evolves more finely over time within each strata of covariates,
  (ii) resembles a stratified randomization design at a given cross-section.

Nested stratification design

Example. One time-varying binary covariate with three time periods



Estimation and inference of $ATT(g, t)$

Estimation. Recall the identification formula

$$ATT(g,t) = \mathbb{E}[\mathbb{E}[Y_{it} \mid G_i = g, \mathbf{X}_{T_0:(g-1)}] - \mathbb{E}[Y_{it} \mid G_i > t, \mathbf{X}_{T_0:(g-1)}]].$$

For a given pair (g, t), we extract the relevant treatment and stratification structure:

- The subgroups $\{i : G_i = g\}$ and $\{i : G_i > t\}$ act as *treated* and *control* groups for estimating $\mathbb{E}[Y_{it}(g) \mid G_i = g]$ and $\mathbb{E}[Y_{it}(\infty) \mid G_i = g]$, respectively.
- After stratifying the observations with respect to covariates $\mathbf{X}_{T_0:(g-1)}$, treatment is as good as random within each stratum.
- This naturally yields a **stratification-based estimator**

$$\widehat{ATT}(g,t) = \sum_{m=1}^{n_g} \omega_{g,t,m} \widehat{\tau}(g,t)_m.$$

($n_g = \#$ of strata, $\widehat{\tau}(g,t)_m =$ difference-in-means in m-th stratum, $\omega_{g,t,m} =$ weight of stratum)

Simultaneous inference of multiple $ATT(g, t)$s

Inference.

- With the full nested stratification design at hand, marginal estimation and inference of each $ATT(g, t)$ is straightforward.
- Of particular interest is simultaneous inference for a set of $ATT(g, t)$s

$$ATT(\mathbf{g}, \mathbf{t}) := (ATT(g_1, t_1), \dots, ATT(g_k, t_k)),$$

  which requires the estimation of the covariance matrix Σ of $ATT(\mathbf{g}, \mathbf{t})$.
  - This can be achieved by implementing block-level bootstrapping, where we resample the outermost strata (blocks) with replacement.
- This enables hypothesis testing of the form $H_0 : R\tau = 0$, which can be used to assess the homogeneity of $ATT(g, t)$s across different dimensions.

Overview

1. Motivation
2. Causal Framework and Design for Longitudinal Data
3. **Matching Algorithm**
4. Data Application: Netflix-IPTV Dataset
5. Summary

Reverse-Time Nested Matching (RTNM) algorithm

We propose a novel matching algorithm, **Reverse-Time Nested Matching (RTNM)**, that reconstructs the nested design from observational time-series datasets.

The RTNM algorithm is

- Easy to implement using optimal full matching, thereby fully using the whole dataset.
- Maintains the main advantages of matching methods, such as reduced model dependence and transparent analysis.
- Enables the estimation of $ATT(g, t)$ by incorporating the nested structure of time-varying covariates.

Reverse-Time Nested Matching (RTNM) algorithm

Step 1 (Initial step). Starting from the latest time period of interest, optimally match units from $\{i : G_i = G\}$ to $\{i : G_i > G\}$ to form matched sets $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$ using optimal full matching.

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 $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$ ('pseudo-controls') 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.

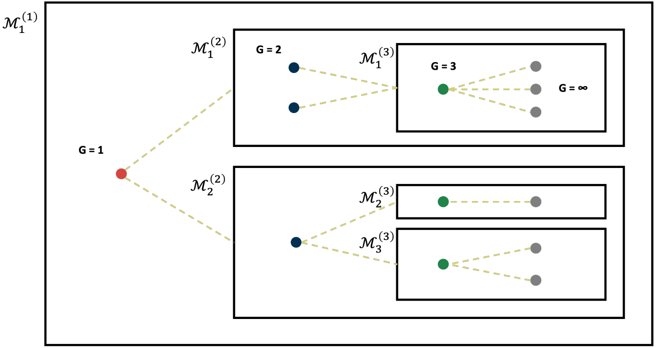
(ii) Based on the distance matrix from (i), compute the distance from each unit in the treated cohort $\{i : G_i = G-1\}$ to the matched sets $\mathcal{M}_1^{(G)}, \dots, \mathcal{M}_{n_G}^{(G)}$.

(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 $\{i : G_i = g-1\}$ with the previously-matched $\mathcal{M}_1^{(g)}, \dots, \mathcal{M}_{n_g}^{(g)}$ until the first cohort $\{i : G_i = 1\}$ is reached.

Reverse-Time Nested Matching (RTNM) algorithm

Example. RTNM with three time periods



Related work in time series matching

1. **Risk set matching** (Li et al., 2001)
   - At the time a unit is treated, it is paired with a not-yet-treated unit whose covariates are similar up to that treatment time point.
   - However, risk set matching *(i) does not account for treatment timing nor its duration*, and *(ii) uses each observation at most once, either as a treated or control unit*.
2. **PanelMatch** (Imai et al., 2023)
   - Each treated observation is matched to a control observation with an identical treatment history, then further matched with units with similar covariate values.
   - While PanelMatch incorporates more dynamic treatment regimes, it ignores treatment timing and imposes assumptions on lagged periods.

RTNM enables estimation of $ATT(g, t)$, allowing the assessment of heterogeneous effects across different dimensions by fully exploiting the data at each cross-section of (g, t).

Overview

1. Motivation
2. Causal Framework and Design for Longitudinal Data
3. Matching Algorithm
4. **Data Application: Netflix-IPTV Dataset**
5. Summary

Dataset description

- **Data source**: Monthly panel data from March 2021 to November 2021, provided by a Korean telecom company. Television viewing behaviour and Netflix subscription status are recorded at the household level.
- **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**: Netflix subscription status at time t.
- **Outcomes**: Television viewing behaviour.
- **Covariates**: Nine time-varying covariates, includeing total and genre-specific real-time and VoD viewing hours and purchase history.

Dataset description

We obtain the following structure for three outcome variables of interest:

(a) Total real-time viewing hours

(b) Total VoD viewing hours

(c) VoD viewing status (yes/no for each month)

	For covariate adjustment			Cohort (**G**) of interest				For comparisons $t > g$	
(calendar time)	202103	202104	202105	202106	202107	202108	202109	202110	202111
G \ **T**	-2	-1	0	1	2	3	4	5	6
1				$\tau(1, 1)$	$\tau(1, 2)$	$\tau(1, 3)$	$\tau(1, 4)$	$\tau(1, 5)$	$\tau(1, 6)$
2					$\tau(2, 2)$	$\tau(2, 3)$	$\tau(2, 4)$	$\tau(2, 5)$	$\tau(2, 6)$
3						$\tau(3, 3)$	$\tau(3, 4)$	$\tau(3, 5)$	$\tau(3, 6)$
4							$\tau(4, 4)$	$\tau(4, 5)$	$\tau(4, 6)$

Causal impact of Netflix on IPTV viewing behaviour

(g, t)	1	2	3	4	5	6
1	-1.635 (5.892)	-2.459 (6.931)	-2.694 (6.662)	-3.303 (7.575)	**-15.456 (7.836)**	**-15.597 (7.699)**
2	-	-3.561 (4.192)	-6.474 (6.055)	-0.809 (6.262)	-2.665 (6.460)	-9.000 (5.271)
3	-	-	5.232 (5.114)	-2.355 (5.975)	-3.313 (5.422)	-2.347 (6.403)
4	-	-	-	1.613 (2.690)	-2.834 (3.248)	3.213 (3.511)

(a) Total real-time viewing hours

(g, t)	1	2	3	4	5	6
1	-1.281 (1.181)	**-4.382 (1.519)**	**-5.124 (1.531)**	**-4.632 (1.513)**	**-5.118 (1.579)**	**-4.168 (1.334)**
2	-	**-4.346 (1.001)**	**-4.576 (1.185)**	**-4.628 (1.254)**	**-6.310 (1.180)**	**-3.117 (0.997)**
3	-	-	**-2.900 (0.923)**	-1.704 (1.044)	-1.770 (1.438)	-1.627 (1.214)
4	-	-	-	**-1.641 (0.576)**	**-1.864 (0.685)**	**-2.438 (0.671)**

(b) Total VoD viewing hours

(g, t)	1	2	3	4	5	6
1	-0.012 (0.030)	-0.064 (0.034)	-0.046 (0.032)	-0.043 (0.032)	**-0.083 (0.031)**	**-0.124 (0.032)**
2	-	**-0.086 (0.032)**	**-0.080 (0.025)**	**-0.102 (0.027)**	**-0.110 (0.027)**	**-0.085 (0.028)**
3	-	-	-0.030 (0.024)	-0.040 (0.028)	**-0.075 (0.029)**	-0.030 (0.029)
4	-	-	-	0.006 (0.013)	**-0.042 (0.016)**	**-0.067 (0.015)**

(c) VoD viewing status

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

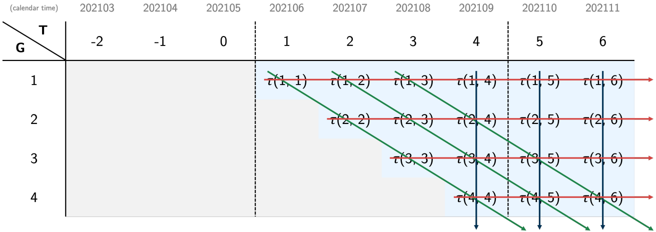
Test for homogeneity of $ATT(g,t)$s

We test homogeneity of $ATT(g,t)$ across three different dimensions:

(i) Fixed g (red): Temporal stability within cohort (e.g. $H_{0,g=1}: ATT(1,1) = \cdots = ATT(1,6)$)

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

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



Test for homogeneity of $ATT(g,t)$s

	Total real-time hours	Total VoD hours	VoD viewing status
$H_{0,g=1}$	0.5566	0.0382^{*}	0.0828
$H_{0,g=2}$	0.7420	0.0162^{*}	0.8448
$H_{0,g=3}$	0.5278	0.6006	0.3402
$H_{0,g=4}$	0.0972	0.4912	0.0002^{***}
$H_{0,t=4}$	0.8786	0.0264^{*}	0.0024^{**}
$H_{0,t=5}$	0.4040	$< 0.0001^{***}$	0.1642
$H_{0,t=6}$	0.0636	0.4080	0.0714
$H_{0,e=0}$	0.5652	0.0630	0.0410^{*}
$H_{0,e=1}$	0.9480	0.0718	0.5604
$H_{0,e=2}$	0.6690	0.1092	0.4950

Table 2: P-values from hypothesis tests assessing the homogeneity of $ATT(g,t)$ for the three outcome variables. Asterisks denote statistical significance ($^{*} : p < 0.05$, $^{**} : p < 0.01$, $^{***} : p < 0.001$).

Overview

Summary

Key Takeaways

- We introduced a **design-based framework** for inference of group-time average treament effects. In particular, the nested stratification design enables simultaneous inference to assess the heterogeneity and dynamics of $ATT(g, t)$s.
- We proposed the **Reverse-Time Nested Matching algorithm** which recovers the nested design from observational time-series data.
  - The method is easy to implement using optimal full matching and enables intuitive analysis with reduced model dependence.
  - This improves upon past work in the matching literature by incorporating heterogeneous effects and fully exploiting the time series dataset.

Thank you!

Email: suehyunkim@snu.ac.kr
Preprint link: https://arxiv.org/abs/2511.23208

Supplementary Materials

Bootstrap estimation for Σ

Step 1. Suppose there are n_1 outermost strata, i.e. strata with respect to $\mathbf{X}_{T_0:0}$.
For each bootstrap iteration $b = 1, \ldots, B$, draw n_1 blocks with replacement to obtain an index set $\mathcal{I}^{(b)}$ and compute

$$\widehat{ATT}(\mathbf{g}, \mathbf{t})^{(b)} = \frac{1}{n_1} \sum_{i \in \mathcal{I}^{(b)}} \hat{\tau}(\mathbf{g}, \mathbf{t})_m$$

Step 2. Compute the bootstrap covariance matrix $\hat{\Sigma} = \text{Cov}\left(\widehat{ATT}(\mathbf{g}, \mathbf{t})^{(b)}\right)$.

Supplementary Materials

Hypothesis tests of the form $H_0 : ATT(g_1, t_1) = \cdots = ATT(g_k, t_k)$ can be rewritten as $H_0 : R\tau = 0$, where $\tau = ATT(\mathbf{g}, \mathbf{t})$. Using the estimated covariance matrix $\widehat{\Sigma}$, we present a **null-restricted bootstrap** procedure for testing such a hypothesis:

Null-restricted bootstrap for hypothesis testing

Step 1. Formulate the hypothesis as $H_0 : R\tau = 0$ and compute the null-restricted estimator

$$\widehat{\tau}_0 = \widehat{\tau} - R^\top (RR^\top)^{-1} R\widehat{\tau}.$$

Step 2. For each bootstrap draw $\widehat{\tau}^{*(b)}$, construct $\widehat{\tau}^{*(b)}_{\text{null}} = \widehat{\tau}_0 + (\widehat{\tau}^{*(b)} - \widehat{\tau})$ and compute

$$W_{\mathrm{obs}} = (R\widehat{\tau})^\top (R\widehat{\Sigma}R^\top)^{-1}(R\widehat{\tau}), \quad W^{*(b)} = (R\widehat{\tau}^{*(b)}_{\text{null}})^\top (R\widehat{\Sigma}R^\top)^{-1}(R\widehat{\tau}^{*(b)}_{\text{null}}).$$

Step 3. Compute the bootstrap p-value $\widehat{p}_B = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{W^{*(b)} \geq W_{\mathrm{obs}}\right\}$.