Synthetic Blips:

Generalizing Synthetic Controls for Dynamic Treatment Effects*

Anish Agarwal	Sukjin Han	Dwaipayan Saha
Columbia University	University of Bristol	Columbia University
aa5194@columbia.edu	vincent.han@bristol.ac.uk	ds4386@columbia.edu

Vasilis Syrgkanis	Haeyeon Yoon
Stanford University	University of Bristol
vsyrgk@stanford.edu	haeyeon.yoon@bristol.ac.uk

October 2, 2025

Abstract

We propose a generalization of the synthetic control and interventions methods to the setting with dynamic treatment effects. We consider the estimation of unit-specific treatment effects from panel data collected under a general treatment sequence. Here, each unit receives multiple treatments sequentially, according to an adaptive policy that depends on a latent, endogenously time-varying confounding state. Under a low-rank latent factor model assumption, we develop an identification strategy for any unit-specific mean outcome under any sequence of interventions. The latent factor model we propose admits linear time-varying and time-invariant dynamical systems as special cases. Our approach can be viewed as an identification strategy for structural nested mean models—a widely used framework for dynamic treatment effects—under a low-rank latent factor assumption on the blip effects. Unlike these models, however, it is more permissive in observational settings, thereby broadening its applicability. Our method, which we term *synthetic blip effects*, is a backwards induction process in which the blip effect of a treatment at each period and for a target unit is recursively expressed as a linear combination of the blip effects of a group of other units that received the designated treatment. This strategy avoids the combinatorial explosion in the number of units that would otherwise be required by a naive application of prior synthetic control and intervention methods in dynamic treatment settings. We provide estimation algorithms that are easy to implement in practice and yield estimators with desirable properties. Using unique Korean firm-level panel data, we demonstrate how the proposed framework can be used to estimate individualized dynamic treatment effects and to derive optimal treatment allocation rules in the context of financial support for exporting firms.

1 Introduction

In many observational studies, units undergo multiple treatments sequentially over time—for example, patients receive multiple therapies, customers are exposed to multiple advertising campaigns, and governments implement multiple policies. The treatment sequence often follows a general pattern rather than being restricted to a staggered adoption design, and interventions typically occur in a data-adaptive manner, with treatment assignment depending on the current (potentially unobserved) state of the treated unit and its past treatments. Furthermore, temporal spillovers across treatments and intermediate outcomes make treatment effects inherently dynamic. A common policy question is what the expected outcome would have been under an alternative policy or course of action. Counterfactual analysis using observational data with multiple sequentially and adaptively assigned treatments is the focus of a long line of research in causal inference.

Typical approaches for identification with time-varying treatments require a strong sequential exogeneity assumption, where the treatment decision at each period is exogenous conditional on an observable state that comprises the history of outcomes and treatments. This assumption is a generalization of the standard conditional exogeneity assumption in static settings. However, most observational datasets are plagued with unobserved confounding, and endogeneity can take complex form especially in dynamic settings. Many techniques exist for addressing unobserved confounding in one-shot treatment settings, such as instrumental variables, difference-in-differences, regression discontinuity designs, and synthetic controls, some of which have been extended to dynamic contexts. For example, event studies and difference-in-differences have been generalized to accommodate sequences of treatments (see below), but most studies assume staggered designs (i.e., irreversible treatment sequences), with few exceptions (Shahn et al., 2022; De Chaisemartin and d'Haultfoeuille, 2024; de Chaisemartin and D'Haultfœuille, 2025). Instrumental variables and regression discontinuity have also been extended to dynamic settings (Han, 2021, 2024; Hsu and Shen, 2024; Sojitra and Syrgkanis, 2024), which requires the existence of sequences of instruments or running variables over time. Beyond these contributions, methods for handling unobserved confounding in settings with general time-varying treatments remain largely underexplored.

In this work, we present the first extension of the synthetic controls literature to handle dynamic treatment effects. Synthetic controls (Abadie and Gardeazabal, 2003; Abadie et al., 2010)—and its generalization to synthetic interventions Agarwal et al. (2020b)—are widely used empirical approaches for handling unobserved confounding from observational panel data. However, the existing literature assumes that units are treated only once or in a non-adaptive manner. This limits the applicability of the technique to policy-relevant settings where multiple interventions occur sequentially over time. We propose an extension of the synthetic controls and synthetic interventions framework that enables identification of mean counterfactual outcomes under arbitrary treatment sequences, even when the observational data arise from an adaptive dynamic treatment policy. As in the synthetic interventions framework, we assume that the panel data stem from a low-rank data generation model, with latent factors capturing unobserved confounding signals. In static settings, the low-rank assumption, together with a technical overlap condition, allows each unit's mean

outcomes under any sequence of interventions to be expressed as linear combinations of observed outcomes from a carefully chosen sub-group of other units. We generalize this idea to dynamic contexts under a low-rank linear structural nested mean model assumption. Our work can also be viewed as extending the g-estimation framework for structural nested mean models (Robins, 2004; Vansteelandt and Joffe, 2014; Lewis and Syrgkanis, 2020) to accommodate unobserved confounding under a low-rank structure. In doing so, our work helps connect the econometric literature on synthetic controls with the biostatistics literature on structural nested mean models.

The key idea of our identification strategy is to express the mean outcome for a unit under a sequence of interventions as an additive function of "blip" effects corresponding to that sequence. The blip effect of an intervention at a given period can be interpreted as the treatment effect of that intervention, relative to a baseline intervention for that specific period, assuming a common sequence of interventions for all other periods. Subsequently, under our low-rank assumption and by applying a recursive argument, we can identify the blip effect of each treatment for each unit and time period. Our procedure can be viewed as a dynamic programming approach, in which a synthetic-control-type procedure is used to compute "synthetic blip effects" at each step of the dynamic program. These step-specific causal quantities are then combined to build the overall counterfactual outcome of any unit under any sequence of interventions.

We illustrate the usefulness of the proposed framework by estimating individualized dynamic treatment effects and optimal treatment allocation rules in the context of providing financial support to exporting firms. Exporting is inherently risky, and thus government agencies play an important role to provide insurance and loans to promote export activities. Using novel Korean firm-level data, we first estimate the effects of insurance and loans as two distinct treatments on firm performances, such as export values. In particular, we recover individualized counterfactual outcomes for all hypothetical intervention sequences. Aggregating across firms yields average effects, which reveal the sequencing of treatments matters for improving export values over time. For example, we find that concentrating interventions to early or later periods on average tends to be more effective than smoothing them across periods. We use firm-specific dynamic treatment effect estimates to learn targeting rules that maximize expected firm-level performance—a step we call *retrospective policy learning*. We show that these rules can significantly improve outcomes while requiring *less* public spending. Finally, we construct interpretable decision trees to guide policymakers in selecting new firms for financial support and in scheduling interventions—that is, *prospective policy learning*.

The paper is organized as follows. We close this section by discussing related work and introducing the setting and notation. Section 2 presents the latent factor model for time-varying treatments, and Section 3 discusses the limitations of the synthetic interventions approach in our setting. Sections 4 and 5 introduce our main models—the time-varying and time-invariant latent factor models—which involve modeling trade-offs. Each section establishes identification, develops an estimation algorithm, and provides the asymptotic theory for the resulting estimator. Section 6 contains our empirical application, and Section 7 concludes. The appendix includes all proofs and additional remarks on the models and assumptions.

1.1 Related Work

Panel data methods in econometrics. Consider a setting where one observes repeated measurements of multiple heterogeneous units over T time steps. Prominent approaches for this setting include difference-in-differences (Ashenfelter and Card, 1984; Bertrand et al., 2004; Angrist and Pischke, 2009) and synthetic controls (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Hsiao et al., 2012; Doudchenko and Imbens, 2016; Li and Bell, 2017; Xu, 2017; Amjad et al., 2018, 2019; Li, 2018; Arkhangelsky et al., 2020; Bai and Ng, 2020; Ben-Michael et al., 2020; Chan and Kwok, 2020; Chernozhukov et al., 2020; Fernández-Val et al., 2020; Agarwal et al., 2021b, 2020a; Athey et al., 2021). These frameworks estimate what would have happened to a unit that undergoes an intervention (i.e., a "treated" unit) had it remained under control (i.e., no intervention), potentially in the presence of unobserved confounding. That is, they estimate the counterfactual outcome of a treated unit if it had remained under control for all T time steps. Recently, the difference-in-differences literature has advanced by taking heterogeneity seriously under staggered designs (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024, among others). Staggered intervention has also been examined in the synthetic controls literature (Shaikh and Toulis, 2021; Ben-Michael et al., 2022; Powell, 2022; Cattaneo et al., 2025). These approaches typically estimate the counterfactual trajectory of treated units had they remained not-yet-treated.

Both one-shot and staggered designs can be viewed as special cases of the general problem we study in this paper: estimating counterfactual outcomes for a unit under *any* hypothetical sequence of interventions over the T time steps. A critical aspect underlying the above methods is the structure assumed between units and time under "control." One elegant way of encoding this structure is through a latent factor model (also known as an interactive fixed effects model), Chamberlain (1984); Liang and Zeger (1986); Arellano and Honore (2000); Bai (2003, 2009); Pesaran (2006); Moon and Weidner (2015, 2017). In such models, it is posited that there exist low-dimensional latent unit and time factors that capture unit- and time-specific heterogeneity, respectively, in the potential outcomes. Since the goal in these works is to estimate outcomes under "control," no structure is imposed on the potential outcomes under intervention.

In Agarwal et al. (2020b, 2021a), the latent factor model is extended to incorporate latent factorization across interventions as well, which allows for identification and estimation of counterfactual mean outcomes under intervention rather than just under control. In Section 3, we provide a detailed comparison with the synthetic interventions framework introduced in Agarwal et al. (2020b). That framework, however, is designed for static regimes and faces two key limitations in the dynamic treatment setting: (i) it does not allow for adaptive treatment assignment over time, and (ii) if there are A possible interventions at each of the T time steps, the sample complexity of the synthetic interventions estimator scales as A^T in order to estimate all possible intervention sequences. The non-adaptivity requirement and the exponential dependence on T make this estimator ill-suited for dynamic treatments, especially as T grows. We show that by imposing that an intervention at a given time step has an additive effect on future outcomes—i.e., an additive latent factor model—we achieve

Linear Factor Models (LFM)	Donor Granularity	Donor Sample Complexity	Adaptivity of Intervention Policy
Naive LFM (Synthetic Interventions)	$\bar{d}^T$	$O(A^T)$	Non-adaptive
Additive Time-Varying LFM (This Work)	(d, t)	$O(A \times T)$	Adaptive after some periods (i.e., staggered adoption of adaptive policy)
Additive Time-Invariant LFM (This Work)	d	$O(A)$	Adaptive after period 1

Table 1: Comparison of Donor Sample Complexity and Adaptivity Across Models.

significant gains in what can be identified and estimated. We study two variants, time-varying and time-invariant versions, which respectively nest the classical linear time-varying and linear time-invariant dynamical system models as special cases. We establish identification results and propose associated estimators to infer all A^T counterfactual trajectories per unit. Importantly, our identification result allows the interventions to be selected in an adaptive manner, and the sample complexity of the estimator no longer exhibits exponential dependence on T; see Table 1.

Another extension of such factor models is the class of "dynamic factor models", originally proposed in Geweke (1976). We refer the reader to Stock and Watson (2011); Chamberlain (2022) for extensive surveys, and to Imbens et al. (2021) for a recent analysis of such time-varying factor models in the context of synthetic controls. These models are similar in spirit to our setting in that they allow outcomes for a given time period to be dependent on outcomes from lagged time periods in an autoregressive manner. To capture this phenomenon, dynamic factor models explicitly represent the time-varying factor as an autoregressive process. However, the target causal parameter in these works is significantly different—they focus on identifying the latent factors and/or forecasting. There is relatively less emphasis on estimating counterfactual mean outcomes for a given unit under different sequences of interventions.

Linear dynamical systems in machine learning. Linear dynamical systems are an extensively studied class of models in the machine learning and applied mathematics literature, and are widely used as linear approximations to many nonlinear systems that nevertheless perform well in practice. A seminal work in this area is Kalman (1960), which introduces the Kalman filter as a robust solution for identifying and estimating the linear parameters that define the system. We refer the reader to the classic survey in Ljung (1999) and the more recent survey in Hardt et al. (2016). Previous works typically assume that (i) the system is driven by independent, and identically distributed (i.i.d.) mean-zero sub-Gaussian noise at each time step, and (ii) both the outcome variable and a meaningful per-time step state are observed and used in estimation. In contrast, we allow for confounding—i.e., the per-time-step actions chosen can be correlated with the system's state in

an unknown manner—and we do not assume access to a per-time-step state, only to the outcome variable. To tackle this setting, we show that linear dynamical systems, both time-varying and time-invariant, are special cases of the latent factor model that we propose. Our recursive "synthetic blip effects" identification strategy enables estimation of mean counterfactual outcomes under any sequence of interventions *without* first performing system identification, and despite unobserved confounding.

Public financial support for exports. Financial frictions play a central role in shaping firms' export performance, particularly in times of crisis (Amiti and Weinstein, 2011; Chor and Manova, 2012; Paravisini et al., 2015). To mitigate financing barriers and sustain exports, governments provide public financial support via export credit agencies (ECAs), mainly in the form of insurance and loans. Public support can generate different effects depending not only on its scale but also on how it is allocated and structured (Criscuolo et al., 2019; Rotemberg, 2019). Empirical studies of ECAs are typically limited to a single treatment (mostly insurance) due to data constraint (Felbermayr and Yalcin, 2013, among others), leaving the broader impact of combined support largely unexplored. This paper considers the entire set of support programs and analyzes how the timing and sequencing of interventions influence firm performance. By going beyond estimating treatment effects, it provides evidence on allocation strategies that enhance the effectiveness of public funds.

1.2 Setting and Notation

Notation. $[R]$ denotes $\{1, \ldots, R\}$ for $R \in \mathbb{N}$. $[R_1, R_2]$ denotes $\{R_1, \ldots, R_2\}$ for $R_1, R_2 \in \mathbb{N}$, with $R_1 < R_2$. $[R]_0$ denotes $\{0, \ldots, R\}$ for $R \in \mathbb{N}$. For a vector a, we define $a^\top$ as its transpose. For vectors $a, b \in \mathbb{R}^d$, we define the inner product of a and b as $\langle a, b \rangle = a^\top b = \sum_{\ell=1}^{d} a_\ell b_\ell$. For a matrix $M \in \mathbb{R}^{m \times n}$, we denote its Frobenius norm as $\|M\|_F$. Let O_p and o_p denote the probabilistic versions of the deterministic big-O and little-o notations.

Setup. Let there be N heterogeneous units. We collect data over T time steps for each unit.

Observed outcomes. For each unit and time period n, t, we observe $Y_{n,t} \in \mathbb{R}$, which is the outcome of interest.

Treatments. For each $n \in [N]$ and $t \in [T]$, we observe treatment actions $D_{n,t} \in [A]$, where $A \in \mathbb{N}$. We allow $D_{n,t}$ to be categorical, i.e., it can simply serve as a unique identifier for the action chosen. Denote a sequence of actions $(d_1, \ldots, d_t)$ by $\bar{d}^t \in [A]^t$; denote $(d_t, \ldots, d_T)$ by $\underline{d}^t \in [A]^{T-t}$. Define $\bar{D}_n^t, \underline{D}_n^t$ analogously to $\bar{d}^t, \underline{d}^t$, respectively, but now with respect to the observed sequence of actions $D_{n,t}$.

Control and interventional period. For each unit n, we assume there exists $t_n^* \in [T]$ before which it is in "control". We denote the control action at time step t as $0_t \in [A]$.[1] Note 0_ℓ and 0_t for $\ell \neq t$, do not necessarily equal each other. For $t \in [T]$, denote $\bar{0}^t = (0_1, \ldots, 0_t)$ and $\underline{0}^t = (0_t, \ldots, 0_T)$. For $t < t_n^*$, we assume $D_{n,t} = 0_t$, i.e., $\bar{D}_n^{t_n^*-1} = \bar{0}^{t_n^*-1}$. That is, during the control period all units are under a common sequence of actions, but for $t \geq t_n^*$, each unit n can undergo a possibly different

---

[1]The notation 0_t is introduced to allow a general control action that is not necessarily "no treatment."

sequence of actions from all other units, denoted by $\underline{D}_n^{t_n^*}$. Note that if $t_n^* = 1$, then unit n is never in the control period.

Counterfactual outcomes. As stated earlier, for each unit and time period n, t, we observe $Y_{n,t} \in \mathbb{R}$, which is the outcome of interest. We denote the potential outcome if unit n had instead undergone $\bar{d}^t$ as $Y_{n,t}^{(\bar{d}^t)}$. More generally, we denote the potential outcome $Y_{n,t}^{(\bar{D}_n^\ell, \underline{d}^{\ell+1})}$ if unit n receives the observed sequence of actions $\bar{D}_n^\ell$ till time step ℓ, and then instead undergoes $\underline{d}^{\ell+1}$ for the remaining $t - \ell$ time steps. [2]

We make the standard "stable unit treatment value assumption" (SUTVA) as follows.

Assumption 1 (Sequential Action SUTVA). *For all $n \in [N], t \in [T], \ell \in [t], \bar{d}^t \in [A]^t$:*

$$Y_{n,t}^{(\bar{D}_n^\ell, \underline{d}^{\ell+1})} = \sum_{\bar{\delta}^\ell \in [A]^\ell} Y_{n,t}^{(\bar{\delta}^\ell, \underline{d}^{\ell+1})} \cdot \mathbb{1}(\bar{D}_n^\ell = \bar{\delta}^\ell).$$

Further, for all $\bar{D}_n^t \in [A]^t$:

$$Y_{n,t}^{(\bar{D}_n^t)} = Y_{n,t}.$$

As an immediate implication, $Y_{n,t}^{(\bar{d}^\ell, \underline{d}^{\ell+1})} \mid \bar{D}_n^\ell = \bar{d}^\ell$ equals $Y_{n,t}^{(\bar{D}_n^\ell, \underline{d}^{\ell+1})} \mid \bar{D}_n^\ell = \bar{d}^\ell$, and $Y_{n,t}^{(\bar{d}^t)} \mid \bar{D}_n^t = \bar{d}^t$ equals $Y_{n,t} \mid \bar{D}_n^t = \bar{d}^t$.

Goal. Our goal is to accurately estimate the potential outcome if a given unit n had instead undergone $\bar{d}^T$ (instead of the actual observed sequence $\bar{D}_n^T$), for any given sequence of actions $\bar{d}^T$ over T time steps. That is, for all $n \in [N]$ $\bar{d}^T \in [A]^T$, our goal is to estimate $Y_{n,T}^{(\bar{d}^T)}$. We more formally define the target causal parameter in Section 2.

2 Latent Factor Model for Dynamic Treatments

We now present a novel latent factor model for causal inference with dynamic treatments. Towards that, we first define the collection of latent factors that are of interest.

Definition 1 (Latent factors). *For a given unit n and time step t, denote its latent factor as $v_{n,t}$. For a given sequence of actions over t time steps, $\bar{d}^t$, denote its associated latent factor as $w_{\bar{d}^t}$. Denote the collection of latent factors as*

$$\mathcal{LF} := \{v_{n,t}\}_{n \in [N], t \in [T]} \cup \{w_{\bar{d}^t}\}_{\bar{d}^t \in [A]^t, \ t \in [T]}.$$

Here $v_{n,t}, w_{\bar{d}^t} \in \mathbb{R}^{m(t)}$, where $m(t)$ is allowed to depend on t.

---

[2] We are slightly abusing notation as the potential outcome $Y_{n,t}^{(\bar{D}_n^\ell, \underline{d}^{\ell+1})}$ is only a function of the first $t - \ell$ components of $\underline{d}^{\ell+1}$, which is actually a vector of length $T - \ell$.

Assumption 2 (General factor model). *Assume* $\forall\, n \in [N],\, t \in [T], \bar{d}^t \in [A]^t,$

$$Y_{n,t}^{(\bar{d}^t)} = \langle v_{n,t}, w_{\bar{d}^t} \rangle + \varepsilon_{n,t}^{(\bar{d}^t)}. \tag{1}$$

Further,

$$\mathbb{E}[\varepsilon_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}] = 0. \tag{2}$$

In (1), the key assumption made is that $v_{n,t}$ does not depend on the action sequence $\bar{d}^t$, while $w_{\bar{d}^t}$ does not depend on unit n. That is, $v_{n,t}$ captures the unit n specific latent heterogeneity in determining the expected conditional potential outcome $\mathbb{E}[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}]$; $w_{\bar{d}^t}$ follows a similar intuition but with respect to the action sequence $\bar{d}^t$. Importantly, the factors can be correlated with the treatment sequence $\bar{D}^t$, making them unobserved confounders. This latent factorization will be key in all our identification and estimation algorithms, and the associated theoretical results. An interpretation of $\varepsilon_{n,t}^{(\bar{d}^t)}$ is that it represents the component of the potential outcome $Y_{n,T}^{(\bar{d}^T)}$ that is not factorizable into the latent factors represented by $\mathcal{LF}$; moreover, it helps model the inherent randomness in the potential outcomes $Y_{n,T}^{(\bar{d}^T)}$. In Sections 4 and 5 below, we show how various standard models of dynamical systems are a special case of our proposed factor model in Assumption 2.

Target Causal Parameter Our target causal parameter to estimate is, for all units $n \in [N]$ and any action sequence $\bar{d}^T \in [A]^T$,

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}], \qquad\qquad \text{(target causal parameter)}$$

i.e., the expected potential outcome conditional on the latent factors, $\mathcal{LF}$. In total this amounts to estimating $N \times A^T$ different (expected) potential outcomes, which we note grows exponentially in T.

3   Limitations of Synthetic Interventions Approach

Given that our goal is to bring to bear a novel factor model perspective to the dynamic treatment effects literature, we first exposit on some of the limitations of the current methods from the factor model literature that were designed for the static interventions regime, i.e., where an intervention is done only once at a particular time step. We focus on the synthetic interventions (SI) framework Agarwal et al. (2020b), which is a recent generalization of the popular synthetic controls framework. In particular, we provide an identification argument which builds upon the SI framework Agarwal et al. (2020b) and then discuss its limitations.

3.1 Identification Strategy via SI Framework

3.1.1 Notation and Assumptions

Donor units. To explain the identification strategy, we first need to define a collection of subsets of units based on: (i) the action sequence they receive; (ii) the correlation between their potential outcomes and the chosen actions. These subsets are defined as follows.

Definition 2 (SI donor units). *For $\bar{d}^T \in [A]^T$,*

$$\mathcal{I}^{\bar{d}^T} := \{j \in [N] : (i) \ \bar{D}_j^T = \bar{d}^T, \ (ii) \ \forall \bar{\delta}^T \in [A]^T, \ \mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \bar{D}_j^T, \mathcal{LF}] = 0\}. \tag{3}$$

The donor set $\mathcal{I}^{\bar{d}^T}$ contains units that receive exactly the sequence $\bar{d}^T$. Further, we require that for these particular units, the action sequence was chosen such that $\forall \bar{\delta}^T \in [A]^T$, $\mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \bar{D}_j^T, \mathcal{LF}] = \mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}] = 0$, i.e., $\varepsilon_{j,T}^{(\bar{\delta}^T)}$ is conditionally mean independent of the action sequence $\bar{D}_j^T$ unit j receives. Note a sufficient condition for property (ii) above is that $\forall \bar{\delta}^T \in [A]^T$, $Y_{j,T}^{(\bar{\delta}^T)} \perp \bar{D}_j^T \mid \mathcal{LF}$. That is, for these units, the action sequence for the entire time period T is chosen at $t = 0$ conditional on the latent factors, i.e., the policy for these units is *not adaptive* (cannot depend on observed outcomes $Y_{j,t}$ for $t \in [T]$).

Assumption 3. *$\forall n \in [N], \bar{d}^T \in [A]^T$ suppose that $v_{n,T}$ satisfies a well-supported condition, i.e., there exists linear weights $\beta^{n,\mathcal{I}^{\bar{d}^T}} \in \mathbb{R}^{|\mathcal{I}^{\bar{d}^T}|}$ such that:*

$$v_{n,T} = \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n,\mathcal{I}^{\bar{d}^T}} v_{j,T}. \qquad \text{(well-supported factors)}$$

Assumption 3 essentially states that for a given sequence of interventions $\bar{d}^T \in [A]^T$, the latent factor for the target unit $v_{n,T}$ lies in the linear span of the latent factors $v_{j,T}$ associated with the "donor" units in $\mathcal{I}^{\bar{d}^T}$. Note by Theorem 4.6.1 of Vershynin (2018), if the $\{v_{j,T}\}_{j \in [N]}$ are sampled as independent, mean zero, sub-Gaussian vectors, then $\text{span}(\{v_{j,T} : j \in \mathcal{I}^{\bar{d}^T}\}) = \mathbb{R}^{m(T)}$ with high probability as $|\mathcal{I}^{\bar{d}^T}|$ grows, and if $|\mathcal{I}^{\bar{d}^T}| \gg m(T)$ (recall $m(T)$ is the dimension of $v_{n,T}$).

3.1.2 Identification Result

We then have identification for the target parameter, which states that the (target causal parameter) can expressed as a function of observed outcomes. It is an adaptation of the identification argument in Agarwal et al. (2020b).

Theorem 1 (SI Identification Strategy). *Let Assumptions 1, 2, and 3 hold. Then, for $\forall n \in [N], \bar{d}^T \in [A]^T$, the mean counterfactual outcome can be expressed as:*

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \mathbb{E}\left[\sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n,\mathcal{I}^{\bar{d}^T}} Y_{j,T} \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T}\right].$$
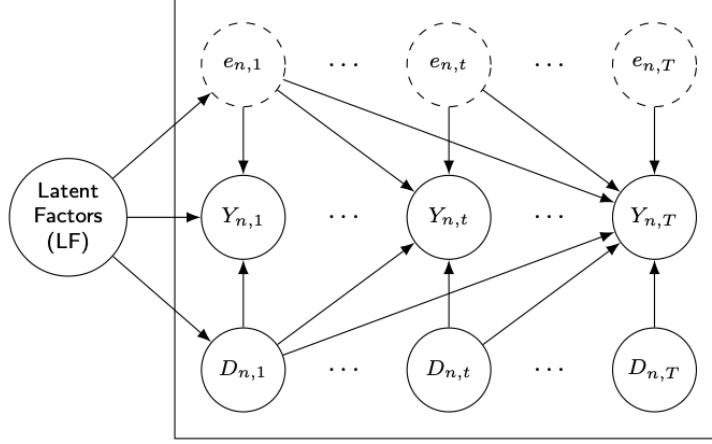
Figure 1: DAG that is consistent with the exogeneity conditions implied by the definition of $\mathcal{I}^{\bar{d}^T}$.

Interpretation of identification result. Theorem 1 establishes that to estimate the mean counterfactual outcome of unit n under the action sequence $\bar{d}^T$, select all donors that received that sequence, i.e., $\bar{D}^T = \bar{d}^T$, and for whom we know that their action sequence was not adaptive. The target causal parameter then is simply a linear re-weighting of the observed outcomes $(Y_{j,T})_{j \in \mathcal{I}^{\bar{d}^T}}$, where these linear weights $\beta_j^{n,\mathcal{I}^{\bar{d}^T}}$ express the latent factor $v_{n,T}$ for unit n as a linear combination of $\{v_{j,T}\}_{j \in \mathcal{I}^{\bar{d}^T}}$.

3.1.3 Discussions: SI Identification Strategy

Donor sample complexity. To estimate $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$ for all units $n \in [N]$ and any action sequence $\bar{d}^T \in [A]^T$, this SI identification strategy requires the existence of a sufficiently large subset of donor units $\mathcal{I}^{\bar{d}^T}$ for every $\bar{d}^T \in [A]^T$. That is, the number of donor units we require will need to scale at the order of A^T, which grows exponentially in T.

Donor exogeneity conditions. Further, the actions picked for these donor units cannot be adaptive as we require $\forall \bar{\delta}^T \in [A]^T$, $\mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \bar{D}_j^T, \mathcal{LF}] = 0$ for them. See Figure 1 for a directed acyclic graph (DAG) that is consistent with the exogeneity conditions implied by the definition of $\mathcal{I}^{\bar{d}^T}$ in (3).

Overcoming limitations of SI identification strategy. Given this combinatorial explosion in the number of donor units and the stringent non-adaptivity requirements on these donor units, in the following sections we study how additional structure on the latent factor model gives rise to novel identification strategies, which allows us to reduce the donor sample complexity and remove the exogeneity requirements between the chosen actions and the donor units.

4 Linear Time-Varying Latent Factor Model

Motivated by the limitation of the identification strategy in Section 3, we now impose additional structure on the latent factor model.

Assumption 4 (Linear time-varying (LTV) factor model)**.** *Assume* $\forall\ n \in [N]$, $t \in [T], \bar{d}^t \in [A]^t$,

$$Y_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t} \left\langle \psi_n^{t,\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t}^{(\bar{d}^t)}, \tag{4}$$

where $\psi_n^{t,\ell}, w_{d_\ell} \in \mathbb{R}^m$ *for* $\ell \in [t]$. *Further, let* $\mathcal{LF} = \{\psi_n^{t,\ell}\}_{n\in[N],t\in[T],\ell\in[t]} \cup \{w_d\}_{d\in[A]}$. *Assume*

$$\mathbb{E}[\varepsilon_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}] = 0. \tag{5}$$

Remark. *Note Assumption 4 implies Assumption 2 holds with*

$$v_{n,t} = [\psi_n^{t,1}, \ldots, \psi_n^{t,t}], \ \ w_{\bar{d}^t} = [w_{d_1}, \ldots, w_{d_t}].$$

Further $m(t) = m \times t$ *for* $m(t)$ *in Definition 1.*

We see that there is additional structure in the latent factors. In particular, the effect of action d_ℓ on $Y_{n,t}^{(\bar{d}^t)}$ for $\ell \in [t]$ is additive, given by $\langle \psi_n^{t,\ell}, w_{d_\ell} \rangle$. Intuitively, $\psi_n^{t,\ell}$ captures the latent unit specific heterogeneity in the potential outcome for unit n at a given time step t for an action taken at time step $\ell \leq t$; analogously w_{d_ℓ} captures the latent effect of action d_ℓ. This additional structure along will be useful in the identification strategy we employ in Section 4.2.

4.1 Motivating Example

A time-varying dynamical system is useful in modeling the dynamic evolution of treatment and outcome sequences. We show that the classical linear time-varying dynamical system model satisfies Assumption 4. Suppose for all $t \in [T]$, all units $n \in [N]$ obey the following dynamic triangular model for a sequence of actions $\bar{D}_n^t$ and counterfactual outcomes $Y_{n,t}^{(\bar{D}_n^t)}$:

$$D_{n,t} = f_n(w_{D_{n,t-1}}, z_{n,t-1}^{(\bar{D}_n^{t-1})}), \tag{6}$$

$$Y_{n,t}^{(\bar{D}_n^t)} = \left\langle \theta_{n,t}, z_{n,t}^{(\bar{D}_n^t)} \right\rangle + \left\langle \tilde{\theta}_{n,t}, w_{D_{n,t}} \right\rangle + \tilde{\eta}_{n,t}, \tag{7}$$

where $z_{n,t}^{(\bar{D}_n^t)} = \boldsymbol{B}_{n,t}\ z_{n,t-1}^{(\bar{D}_n^{t-1})} + \boldsymbol{C}_{n,t}\ w_{D_{n,t}} + \eta_{n,t}$ and $z_{n,0} = w_{D_{n,0}} = 0$. Here, $z_{n,t} \in \mathbb{R}^m$ is the latent state associated with unit n at time t and $w_{D_{n,t-1}} \in \mathbb{R}^m$ is the chosen action at time $t-1$. $\eta_{n,t} \in \mathbb{R}^m$ and $\tilde{\eta}_{n,t} \in \mathbb{R}$ represent independent mean-zero random innovations at each time step t. $\boldsymbol{B}_{n,t}, \boldsymbol{C}_{n,t} \in \mathbb{R}^{m\times m}$ are matrices governing the linear dynamics of $z_{n,t}^{(\bar{D}_n^t)}$. Note $\boldsymbol{B}_{n,t}, \boldsymbol{C}_{n,t}$ are specific to time step t and this is what makes this model a time-varying dynamical system. In contrast, in the classic linear time-invariant dynamical system described in Section 5.1 below, $\boldsymbol{B}_{n,t} = \boldsymbol{B}_n$

and $\boldsymbol{C}_{n,t} = \boldsymbol{C}_n$ for all $t \in [T]$. $\theta_{n,t}, \tilde{\theta}_{n,t} \in \mathbb{R}^m$ are parameters governing how the outcome of interest $Y_{n,t}^{(\bar{D}_n^t)}$ is a linear function of $z_{n,t}^{(\bar{D}_n^t)}$ and w_{d_t}, respectively. $f_n(\cdot)$ is a function which decides how the next action $w_{D_{n,t}}$ is chosen as a function of the previous action $w_{D_{n,t-1}}$, and current state $z_{n,t}$. We see that due to the input of $z_{n,t}^{(\bar{D}_n^t)}$ in $f_n(\cdot)$, i.e., the action sequence is *adaptive*. As a result, $\eta_{n,\ell}$ is correlated with $D_{n,t}$ for $\ell < t$.

Proposition 1. *Suppose the dynamic triangular model* (6)–(7) *holds. Then we have the following representation,*

$$Y_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t} \left(\left\langle \psi_n^{t,\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t,\ell} \right), \tag{8}$$

where $\psi_n^{t,\ell}, w_{d_\ell} \in \mathbb{R}^m$ for $\ell \in [t]$; here,

$$\psi_n^{t,\ell} = \left(\left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \boldsymbol{C}_{n,\ell} \right)' \theta_{n,t} \quad for \quad \ell \in [t-1],$$

$$\psi_n^{t,t} = \boldsymbol{C}_{n,t}' \theta_{n,t} + \tilde{\theta}_{n,t},$$

$$\varepsilon_{n,t,\ell} = \left(\left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \eta_{n,\ell} \right)' \theta_{n,t} \quad for \quad \ell \in [t-1],$$

$$\varepsilon_{n,t,t} = \theta_{n,t}' \eta_{n,t} + \tilde{\eta}_{n,t}.$$

Therefore, Assumption 4 holds with the additional structure that $\varepsilon_{n,t}^{(\bar{d}^t)}$ has an additive factorization as $\sum_{\ell=1}^{t} \varepsilon_{n,t,\ell}$, and it is not a function of d_ℓ.

In this example, our target parameter $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$ defined in (target causal parameter) translates to the expected potential once we condition on the latent parameters $\psi_n^{t,\ell}, w_{d_\ell}$, which are a function of $\boldsymbol{B}_{n,t}, \boldsymbol{C}_{n,t}, \theta_{n,t}, \tilde{\theta}_{n,t}$. Here the expectation is take with respect to the per-step independent mean-zero random innovations, $\varepsilon_{n,t,\ell}$, which are a function of $\{\eta_{n,q}, \tilde{\eta}_{n,q}\}_{q \geq \ell}$ (and $\boldsymbol{B}_{n,t}, \boldsymbol{C}_{n,t}, \theta_{n,t}, \tilde{\theta}_{n,t}$).

4.2 LTV Identification Strategy

In this section we identify $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$, that is, we represent this expected potential outcome for a target unit n and action sequence $\bar{d}^T$ as some function of observed outcomes.

4.2.1 Notation and Assumptions

Notation. We define the following useful notation for any unit $n \in [N]$:

$$\gamma_{n,T,t}(d_t) := \left\langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \right\rangle.$$

Note that $\gamma_{n,T,t}(d_t)$ can be interpreted as a "blip effect"—the expected difference in potential outcomes if unit n undergoes the sequence $(\bar{d}^t, \underline{0}^{t+1})$ instead of $(\bar{d}^{t-1}, \underline{0}^t)$. In particular, note that

Assumption 4 implies

$$\mathbb{E}\left[Y_{n,T}^{(\bar{d}^t,\underline{0}^{t+1})} - Y_{n,T}^{(\bar{d}^{t-1},\underline{0}^t)} \mid \mathcal{LF}\right] = \mathbb{E}\left[\left\langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \right\rangle + \varepsilon_{n,T}^{(\bar{d}^t,\underline{0}^{t+1})} - \varepsilon_{n,T}^{(\bar{d}^{t-1},\underline{0}^t)} \mid \mathcal{LF}\right]$$
$$= \left\langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \right\rangle \mid \mathcal{LF}.$$

Further, let

$$b_{n,T} := \sum_{t=1}^{T} \left\langle \psi_n^{T,t}, w_{0_t} \right\rangle.$$

This can be interpreted as the expected potential outcome if unit j remains under the control sequence $\bar{0}^T$ till time step T. Again, Assumption 4 implies

$$\mathbb{E}\left[Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}\right] = \mathbb{E}\left[\sum_{t=1}^{T} \left\langle \psi_n^{T,t}, w_{0_t} \right\rangle + \varepsilon_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}\right] = \sum_{t=1}^{T} \left\langle \psi_n^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF}. \tag{9}$$

Assumptions. We now state assumptions we need for the identification strategy that we propose.
Donor sets. We define different subsets of units based on the treatment sequence they receive:

$$\mathcal{I}_t^d := \{j \in [N] : (i) \ \bar{D}_j^t = (0_1, \ldots, 0_{t-1}, d),$$
$$(ii) \ \forall \ \bar{\delta}^T \in [A]^T, \ \mathbb{E}[Y_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}, \bar{D}_j^t] = \mathbb{E}[Y_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}]\}. \tag{10}$$

The donor set $\mathcal{I}_t^d$ contain units that remain under the control sequence $(0_1, \ldots, 0_{t-1})$ till time step $t-1$, and at time step t receive action d (i.e., $t_n^* \geq t-1$). Further, we require that for these particular units, the action sequence, $\bar{D}_j^t$, till time step t was chosen such that $\mathbb{E}[Y_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}, \bar{D}_j^t] = \mathbb{E}[Y_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}]$, i.e., the potential outcomes are conditionally mean independent of the action sequence $\bar{D}_j^t$ unit j receives till time step t. Of course, a sufficient condition for property (ii) above is that $\forall \ \bar{\delta}^T \in [A]^T, \ Y_{j,T}^{(\bar{\delta}^T)} \perp \bar{D}_j^t \mid \mathcal{LF}$. That is, for these units, the action sequence till time step t is chosen at $t=0$ conditional on the latent factors, i.e., the policy for these units can only be adaptive from time step $t+1$. Note, given Assumption 4, this property (ii) can be equivalently stated as $\mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}, \bar{D}_j^t] = \mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}] = 0.$

Assumption 5. *For $n \in [N]$, let $v_{n,T} := [\psi_n^{T,1}, \ldots, \psi_n^{T,T}]$. We assume that for all $n \in [N]$, $v_{n,T}$ satisfies a well-supported condition with respect to the various donor sets, i.e., for all $d \in [A]$ and $t \in [T]$, there exists $\beta^{n,\mathcal{I}_t^d} \in \mathbb{R}^{|\mathcal{I}_t^d|}$ such that*

$$v_{n,T} = \sum_{k \in \mathcal{I}_t^d} \beta_k^{n,\mathcal{I}_t^d} v_{k,T}. \qquad \text{(LTV well-supported factors)}$$

Assumption 5 requires that for units $n \in [N]$, their latent factors $[\psi_n^{T,1}, \ldots, \psi_n^{T,T}]$ are expressible as a linear combination of the units in the donor sets $\mathcal{I}_t^d$. See the discussion under Assumption 3 in Section 3 justifying such an assumption for settings when $\mathcal{I}_t^d$ is sufficiently large.

13

Assumption 6. *For all $n \in \mathcal{I}_t^d, t \in [T], \bar{d}^t \in [A]^t$,*

$$\mathbb{E}\left[Y_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})} - Y_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)} \mid \bar{D}_n^t = \bar{d}^t, \mathcal{LF}\right] = \gamma_{n,T,t}(d_t) \mid \mathcal{LF}.$$

Note that given Assumption 4, this condition can be equivalently written as

$$\mathbb{E}\left[\varepsilon_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})} - \varepsilon_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)} \mid \bar{D}_n^t = \bar{d}^t, \mathcal{LF}\right] = 0.$$

Below we give two sufficient conditions under which Assumption 6 holds.

1. Sufficient condition: Non-action dependent noise. Assumption 6 holds if $\varepsilon_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})} = \varepsilon_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)}$, which occurs if $\varepsilon_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})}$ and $\varepsilon_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)}$ are not a function of $(\bar{d}^t, \underline{0}^{t+1})$, and $(\bar{d}^{t-1}, \underline{0}^t)$, respectively. The motivating example of a classic linear time-varying dynamical system given in Section 4.1 satisfies this property.

2. Sufficient condition: Additive action-dependent noise. We now relax the sufficient condition above that $\varepsilon_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})}$ and $\varepsilon_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)}$ are not a function of the action sequence. Instead, suppose for all $\bar{d}^T \in [A]^T$, $\varepsilon_{n,T}^{(\bar{d}^T)} = \sum_{t=1}^T \eta_{n,t}^{(d_t)}$, where we assume that conditional on $\mathcal{LF}$, $\eta_{n,t}^{(d_t)}$ are mutually independent for all $t \in [T]$, and $d_t \in [A]$. Then $\varepsilon_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})} - \varepsilon_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)} = \eta_{n,t}^{(d_t)} - \eta_{n,t}^{(0_t)}$. In this case, a sufficient condition for Assumption 6 is that

$$\eta_{n,t}^{(d_t)}, \eta_{n,t}^{(0_t)} \perp\!\!\!\perp D_{n,t} \mid \mathcal{LF}.$$

That is, conditional on the latent factors, the action $D_{n,t}$ at time step t is independent of the additional noise $\eta_{n,T,t}^{(d_t)}, \eta_{n,T,t}^{(0_t)}$ generated at time step t. Note, however that $\varepsilon_{n,t}^{(\bar{d}^t)} \not\perp\!\!\!\perp D_{n,t} \mid \mathcal{LF}$. This is because $\varepsilon_{n,t-1}^{(\bar{d}^{t-1})}$ and $\varepsilon_{n,t}^{(\bar{d}^t)}$ remain auto-correlated, i.e., $. \ \varepsilon_{n,t}^{(\bar{d}^{t-1})} \not\perp\!\!\!\perp \varepsilon_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}$. Also, $\varepsilon_{n,t}^{(\bar{d}^{t-1})} \not\perp\!\!\!\perp D_{n,t} \mid \mathcal{LF}$, as the action $D_{n,t}$ can be a function of the observed outcomes $Y_{n,t-1}$.

Sequential conditional exogeneity, SNMMs and MSMs. We now connect our assumptions more closely to the notation and assumptions used in the structural nested mean model (SNMM) and the marginal structural model (MSM) in the statistics literature on dynamic treatment effects. A typical assumption in these literatures is sequential conditional exogeneity, which states that for some sequence of random state variables $S_{n,t}$, the treatments are sequentially conditionally exogenous, i.e.:

$$\forall \bar{d} \in [A]^T : Y_{n,T}^{(\bar{d})} \perp\!\!\!\perp D_{n,t} \mid \bar{S}_n^{t-1}, \bar{D}_n^t = \bar{d}_t, \mathcal{LF}, \tag{11}$$

where $\bar{S}_n^{t-1} = (S_{n,0}, \ldots, S_{n,t-1})$. Moreover, assume that the blip effects admit the following factor model representation:

$$\mathbb{E}\left[Y_{n,T}^{(\bar{d}^t, \underline{0}^{t+1})} - Y_{n,T}^{(\bar{d}^{t-1}, \underline{0}^t)} \mid \bar{S}_n^{t-1}, \bar{D}_n^t = \bar{d}^t, \mathcal{LF}\right] = \langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \rangle \mid \mathcal{LF}. \tag{12}$$

(12) implies that the conditional mean of the blip effect is invariant of the past states and actions. Lastly, assume that the baseline potential outcome has a factor model representation, i.e.:

$$\mathbb{E}\left[Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}\right] = \sum_{t=1}^{T} \langle \psi_n^{T,t}, w_{0_t} \rangle \mid \mathcal{LF}. \tag{13}$$

Then we have the following proposition,

Proposition 2. *Let* (11), (12) *and* (13) *hold. Then Assumptions 4 and 6 hold.*

The proof of Proposition 2 can be found in Appendix A. The proof, which is an inductive argument, is in essence known in the literature, i.e., SNMM models that are past action and state independent also imply a marginal structural model, i.e. Assumption 4, (see e.g. Technical Point 21.4 of Hernán and Robins (2020)). We include it in our appendix for completeness and to abide to our notation. Thus instead of Assumptions 4 and 6, one could impose (11)–(13), which are more in line with the dynamic treatment effect literature. Our identification argument would then immediately apply. However, our assumptions are more permissive and flexible in their current form. For instance, unlike a full SNMM specification, our blip definition in Assumption 6 only requires that the blip effect is not modified by past actions, but potentially allows for modification conditional on past states that confound the treatment. However, the full SNMM model presented above precludes such effect modifications.

4.2.2 Identification Result

Given these assumptions, we now present the main identification results. We first illustrate the key intuition behind the identification analysis in a simple two-period setting. Note that, for given unit n and action sequence $(d_1, d_2) \in [A] \times [A]$, the expected potential outcome—the main causal object of interest—can be decomposed into two blip effects and a baseline outcome:

$$\mathbb{E}\left[Y_n^{(d_1,d_2)}\right] = \underbrace{\mathbb{E}\left[Y_n^{(d_1,d_2)}\right] - \mathbb{E}\left[Y_n^{(0_1,d_2)}\right]}_{\text{Blip}_1(d_1)} + \underbrace{\mathbb{E}\left[Y_n^{(0_1,d_2)}\right] - \mathbb{E}\left[Y_n^{(0_1,0_2)}\right]}_{\text{Blip}_2(d_2)} + \underbrace{\mathbb{E}\left[Y_n^{(0_1,0_2)}\right]}_{\text{Baseline}}, \tag{14}$$

where we suppress for simplicity the subscript T of terminal period ($T = 2$) from the expression $Y_n^{(d_1,d_2)}$ and conditioning of $\mathcal{LF}$. Here $\text{Blip}_t(d_t)$ is unit n's treatment effect of intervention d_t compared to baseline 0_t for time step t with common interventions for all other time steps and the baseline is unit n's expected potential outcome if the unit is remained under the control sequence for all T time steps. Under Assumption 4, $\mathbb{E}[Y_n^{(d_1,d_2)}] = \sum_{\ell=1}^{2} \langle v_n^\ell, w_{d_\ell} \rangle$. In other words, Assumption 4 imposes latent factor structure on blip effects and baseline outcomes: $\text{Blip}_2(d_2) = \langle v_n^2, w_{d_2} - w_{0_2} \rangle$, $\text{Blip}_1(d_1) = \langle v_n^1, w_{d_1} - w_{0_1} \rangle$, and $\text{Baseline} = \sum_{\ell=1}^{2} \langle v_n^\ell, w_{0_\ell} \rangle$. This way, we assume blips effects and baseline of units can be represented as linear combinations of one another, making the identification problem akin to those in synthetic control and synthetic intervention.

Based on the latent factor structure, we now demonstrate how each component in (14) can be

Donor Type	Donor Set	Treatment Sequence
1	$\mathcal{I}_2^0$	$(0_1, 0_2)$
2	$\mathcal{I}_2^{d_2}$	$(0_1, d_2)$
3	$\mathcal{I}_1^{d_1}$	(d_1, D_2)

Table 2: Donor Types with $T = 2$.

recursively identified by constructing synthetic units based on an appropriate type of donors (Table 2). In Step 1, starting with the baseline $\mathbb{E}[Y_n^{(0_1,0_2)}]$, we create a synthetic baseline for all units via linear combination of the observed outcomes of Type 1 donors, which identifies the baseline. Next in Step 2, for $\text{Blip}_2(d_2)$, we create a synthetic $\text{Blip}_2(d_2)$ for all units via linear combination of $\text{Blip}_2(d_2) = \mathbb{E}[Y_n^{(0_1,d_2)} - Y_n^{(0_1,0_2)}]$ for Type 2 donors. Note that the blip effects of these donors are "observed" (i.e., already identified), as $\mathbb{E}[Y_n^{(0_1,d_2)}] = \mathbb{E}[Y_n]$ for these donors and their baselines $\mathbb{E}[Y_n^{(0_1,0_2)}]$ are identified in Step 1. Finally in Step 3, to identify $\text{Blip}_1(d_1)$, we first construct "observed" $\text{Blip}_1(d_1)$ for Type 3 donors. Note that, under Assumption 4, $\text{Blip}_1(d_1)$ can be expressed as, with the observed action D_2 (n suppressed) of the unit,

$$\mathbb{E}\left[Y_n^{(d_1,D_2)}\right] - \mathbb{E}\left[Y_n^{(0_1,D_2)}\right] = \mathbb{E}\left[Y_n^{(d_1,D_2)}\right] - \mathbb{E}\left[Y_n^{(0_1,D_2)} - Y_n^{(0_1,0_2)}\right] - \mathbb{E}\left[Y_n^{(0_1,0_2)}\right].$$

On the right-hand side, the first term satisfies $\mathbb{E}[Y_n^{(d_1,D_2)}] = \mathbb{E}[Y_n]$ for these donors, the second term is $\text{Blip}_2(D_2)$ identified for these donors in Step 2, and the third term is the baseline identified for these donors in Step 1. Now that $\text{Blip}_1(d_1)$ for Type 3 donors are identified, linear combination of them identifies $\text{Blip}_1(d_1)$ for all units. Therefore, we identify $\mathbb{E}[Y_n^{(d_1,d_2)}]$ for all units and any given $(d_1, d_2) \in [A]^2$.

We now present the formal identification results for general T.

Theorem 2. *Let Assumptions 1,4, 5, and 6 hold. Then, for any unit $n \in [N]$ and action sequence $\bar{d}^T \in [A]^T$, the expected counterfactual outcome can be expressed as:*

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \sum_{t=1}^{T} \gamma_{n,T,t}(d_t) + b_{n,T} \mid \mathcal{LF}, \qquad \text{(identification)}$$

where quantities on the right-hand side are identified as follows:

(i) We have the following representations of the baseline outcomes

$$\forall\, j \in \mathcal{I}_T^0 \;:\; b_{j,T} \mid \mathcal{LF} = \mathbb{E}[Y_{j,T} \mid \mathcal{LF},\, j \in \mathcal{I}_T^0], \qquad \text{(observed control)}$$

$$\forall\, i \notin \mathcal{I}_T^0 \;:\; b_{i,T} \mid \mathcal{LF} = \sum_{j \in \mathcal{I}_T^0} \beta_j^{i,\mathcal{I}_T^0}\, b_{j,T} \mid \mathcal{LF}, \mathcal{I}_T^0. \qquad \text{(synthetic control)}$$

16

(ii) We have the following representations of the blip effect at time T for $\forall d \in [A]$:

$$\forall\, j \in \mathcal{I}_T^d \;:\; \gamma_{j,T,T}(d) \mid \mathcal{LF} = \; \mathbb{E}[Y_{j,T} \mid \mathcal{LF},\, j \in \mathcal{I}_T^d] - b_{j,T} \mid \mathcal{LF}, \qquad (\text{``observed'' blip at time } T)$$

$$\forall\, i \notin \mathcal{I}_T^d \;:\; \gamma_{i,T,T}(d) \mid \mathcal{LF} = \sum_{j \in \mathcal{I}_T^d} \beta_j^{i,\mathcal{I}_T^d} \gamma_{j,T,T}(d) \mid \mathcal{LF}, \mathcal{I}_T^d. \qquad (\text{synthetic blip at time } T)$$

(iii) We have the following recursive representations of the blip effect $\forall\, t < T,\ d \in [A]$:

$$\forall\, j \in \mathcal{I}_t^d \;:\; \gamma_{j,T,t}(d) \mid \mathcal{LF} = \; \mathbb{E}[Y_{j,T} \mid \mathcal{LF}, \mathcal{I}_t^d] - b_{j,T} \mid \mathcal{LF} - \sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,\ell}) \mid \mathcal{LF},$$

$$(\text{``observed'' blip at time } t)$$

$$\forall\, i \notin \mathcal{I}_t^d \;:\; \gamma_{i,T,t}(d) \mid \mathcal{LF} = \sum_{j \in \mathcal{I}_t^d} \beta_j^{i,\mathcal{I}_t^d} \gamma_{j,T,t}(d) \mid \mathcal{LF}, \mathcal{I}_t^d. \qquad (\text{synthetic blip at time } t)$$

Interpretation of identification result. (identification) states that our target causal parameter of interest can be written as an additive function of $b_{n,T}$ and $\gamma_{n,T,t}(d_t)$ for $t \in [T]$ and $d_t \in [A]$. Theorem 2 establishes that these various quantities are expressible as functions of observed outcomes $\{Y_{j,T}\}_{j \in [N]}$. We give an interpretation below.

Identifying baseline outcomes. For units $j \in \mathcal{I}_T^0$, (observed control) states that their baseline outcome $b_{j,T}$ is simply their expected observed outcome at time step T, i.e., $Y_{j,T}$. For units $i \notin \mathcal{I}_T^0$, (synthetic control) states that we can identify $b_{i,T}$ by appropriately re-weighting the baseline outcomes $b_{j,T}$ of the units $j \in \mathcal{I}_T^0$ (identified via (observed control)).

Identifying blip effects at time T. For any given $d \in [A]$: For units $j \in \mathcal{I}_T^d$, (``observed'' blip at time T) states that their blip effect $\gamma_{j,T,T}(d)$ is equal to their observed outcome $Y_{j,T}$ minus the baseline outcome $b_{j,T}$ (identified via (synthetic control)). For units $i \notin \mathcal{I}_T^d$, (synthetic blip at time T) states that we can identify $\gamma_{i,T,T}(d)$ by appropriately re-weighting the blip effects $\gamma_{j,T,T}(d)$ of units $j \in \mathcal{I}_T^d$ (identified via (``observed'' blip at time T)).

Identifying blip effects at time $t < T$. Suppose by induction $\gamma_{n,T,\ell}(d)$ is identified for every $\ell \in [t+1, T]$, $n \in [N]$, $d \in [A]$, i.e., can be expressed in terms of observed outcomes. Then for any given $d \in [A]$: For units $j \in \mathcal{I}_t^d$, (``observed'' blip at time t) states that their blip effect $\gamma_{j,T,t}(d)$ is equal to their their observed outcome $Y_{j,T}$ minus the baseline outcome $b_{j,T}$ (identified via (synthetic control)) minus the sum of blip effects $\sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,t})$ (identified via the inductive hypothesis). For units $i \notin \mathcal{I}_t^d$, (synthetic blip at time t) states that we can identify $\gamma_{i,T,t}(d)$ by appropriately re-weighting the blip effects $\gamma_{j,T,t}(d)$ of units $j \in \mathcal{I}_T^d$ (identified via (``observed'' blip at time t)).

4.2.3 Discussions: LTV Identification Strategy

Donor sample complexity. To estimate $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$ for all units $n \in [N]$ and any action sequence $\bar{d}^T \in [A]^T$, the LTV identification strategy requires the existence of a sufficiently large subset of donor units $\mathcal{I}_t^d$ for every $d \in [A]$ and $t \in [T]$. That is, the number of donor units we
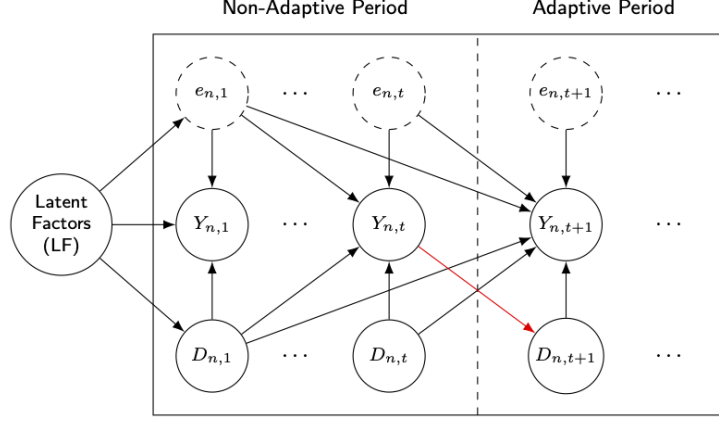
Figure 2: DAG that is consistent with the exogeneity conditions implied by the definition of $\mathcal{I}_t^d$. From time step $t+1$, the action sequence $(D_{n,t+1}, \ldots, D_{n,T})$ can be adaptive, i.e., dependent on the observed outcomes $\{Y_{n,t}\}_{t \in [T]}$ (depicted by the red arrow).

require will need to scale at the order of $A \times T$, which grows linearly in both A and T increases. Thus we see the the additional structure imposed by the time-varying factor model introduced in Assumption 4 leads to a decrease in sample complexity from A^T to $A \times T$, when compared with the general factor model introduced in Assumption 2.

Remark. *The additive structure in Assumption 4 can be relaxed to a hybrid structure that allows for flexible interaction among treatments for a fixed number (e.g., h) of consecutive periods, while maintaining additivity across the fixed windows: the latent factor $w_{\bar{d}^t}$ in Assumption 2 satisfies*

$$w_{\bar{d}^t} = [w_{\tilde{d}_1}, \ldots, w_{\tilde{d}_t}], \tilde{d}_\ell = (d_{\ell-h+1}, ..., d_\ell), \ \ell \geq h.$$

Then the sample complexity can be bounded by $A^h \times T$. This remark highlights the trade-off between sample complexity and model flexibility.

Donor exogeneity conditions. Further, for $j \in \mathcal{I}_t^d$, we require that $\forall \ \delta^T \in [A]^T$, $\mathbb{E}[\varepsilon_{j,T}^{(\bar{\delta}^T)} \mid \mathcal{LF}, \bar{D}_j^t] = 0$. That is, the actions picked for these donor units are only required to be non-adaptive till time step t as opposed to being non-adaptive for the entire time period T, which was required for the SI identification strategy in Section 3. See Figure 2 for a DAG that is consistent with the exogeneity conditions implied by the definition of $\mathcal{I}_t^d$ in (10).

Overcoming limitations of LTV identification strategy. We have shown that this additional linear time-varying latent factor structure, motivated by a linear time-varying dynamical system, yields substantial gains in terms of the number of donor units required and the flexibility of their action sequences. This begs the question of how much more can be gained if we instead consider a linear *time-invariant* latent factor structure, motivated by a linear time-invariant dynamical system. In Section 5, we show that this additional structure surprisingly implies far better donor sample complexity and less stringent exogeneity conditions on the donor units.

4.3 `SBE-PCR` Estimator in LTV Setting

Here we detail the specific algorithm that yields the `SBE-PCR` estimator. To do so we consider the following additional covariates.

Assumption 7 (Additional Covariates). *For each unit $n \in [N]$, we assume access to covariates $X_n \in \mathbb{R}^p$ such that each element satisfies*

$$X_{n,k} = \langle v_{n,T}, \rho_k \rangle + \varepsilon_{n,k}, \tag{15}$$

where $v_{n,T}$ is the unit latent factor defined in Assumptions 2 and $\varepsilon_{n,k}$ is independent mean-zero noise. Denote $X \in \mathbb{R}^{p \times N} = [X_1, \ldots, X_N]$. We can also design more general time-varying covariates as detailed in Appendix C.3.

We make an additional assumption regarding control factors in order develop an algorithm with consistent control estimators as will be seen in later sections.

Assumption 8. *For any donor set, i.e., any $t \in [T]$, $d \in [A]$, and unit $n \in \mathcal{I}_t^d$ there exist weights $\phi^{n,\mathcal{I}_t^d} \in \mathbb{R}^{|\mathcal{I}_t^d|-1}$ such that*

$$v_{n,T} = \sum_{k \in \mathcal{I}_t^d \setminus n} \phi_k^{n,\mathcal{I}_t^d} v_{k,T}.$$

This assumption allows us to detail the algorithm for estimating weights using Principal Component Regression (PCR). Specifically, for each $d \in [A]$, $t \in [T]$, and unit $n \in [N]$ we consider the donor set $\mathcal{I}_t^d$ and estimate weights to express the response vector $X_n \in \mathbb{R}^p$ as a linear combination of the covariates from other donor units. The corresponding matrix of covariates is $X_{\mathcal{I}_t^d \setminus n} = X_{:,\mathcal{I}_t^d \setminus n} \in \mathbb{R}^{p \times |\mathcal{I}_t^d \setminus n|}$, which only chooses the relevant donor columns.

We will apply PCR by regressing $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_t^d \setminus n}$-approximation $X_{\mathcal{I}_t^d \setminus n}$ with $k_{\mathcal{I}_t^d \setminus n} = \mathrm{rank}(\mathbb{E}[X_{\mathcal{I}_t^d \setminus n}])$, i.e., conducting PCR with parameter $k_{\mathcal{I}_t^d \setminus n}$. Denote the Singular Value Decomposition (SVD) of $X_{\mathcal{I}_t^d \setminus n}$ as

$$X_{\mathcal{I}_t^d \setminus n} = \sum_{l \geq 1} \sigma_l u_l v_l',$$

where $u_l \in \mathbb{R}^p$ and $v_l \in \mathbb{R}^{|\mathcal{I}_t^d \setminus n|}$ are the left and right singular vectors arranged in decreasing order of corresponding singular values σ_l.[3] At this point we know if $n \in \mathcal{I}_t^d$

$$\hat{\phi}^{n,\mathcal{I}_t^d} = \left(\sum_{l=1}^{k_{\mathcal{I}_t^d \setminus n}} (1/\sigma_l) v_l u_l' \right) X_n \in \mathbb{R}^{|\mathcal{I}_t^d|-1},$$

and if $n \notin \mathcal{I}_t^d$

$$\hat{\beta}^{n,\mathcal{I}_t^d} = \left(\sum_{l=1}^{k_{\mathcal{I}_t^d}} (1/\sigma_l) v_l u_l' \right) X_n \in \mathbb{R}^{|\mathcal{I}_t^d|}.$$

---

[3]Notice that depending on if unit n is in the donor set $\mathcal{I}_t^d$ our covariate matrix size varies. This is intentional in order to unify notation between units in donor sets and those not in donor sets, since $\mathcal{I}_t^d \setminus n = \mathcal{I}_t^d$ if $n \notin \mathcal{I}_t^d$.

The distinction between using β and ϕ is to emphasize the difference in dimension. For justification of using PCR in our context, refer to Agarwal et al. (2021b, 2020a). Given our weight estimation algorithm above we are ready for our `SBE-PCR` algorithm.

Step 1: Estimate baseline outcomes.

1. For $j \in \mathcal{I}_T^0$

$$\hat{b}_{j,T} = \sum_{k \in \mathcal{I}_T^0 \setminus j} \hat{\phi}_k^{j,\mathcal{I}_T^0} Y_{k,T}.$$

2. For $i \notin \mathcal{I}_T^0$

$$\hat{b}_{i,T} = \sum_{j \in \mathcal{I}_T^0} \hat{\beta}_j^{i,\mathcal{I}_T^0} \hat{b}_{j,T}.$$

Step 2: Estimate blip effects at time T.
For $d \in [A]$:

1. For $j \in \mathcal{I}_T^d$

$$\hat{\gamma}_{j,T,T}(d) = \sum_{k \in \mathcal{I}_T^d \setminus j} \hat{\phi}_k^{j,\mathcal{I}_T^d} Y_{k,T} - \hat{b}_{j,T}.$$

2. For $i \notin \mathcal{I}_T^d$

$$\hat{\gamma}_{i,T,T}(d) = \sum_{j \in \mathcal{I}_T^d} \hat{\beta}_j^{i,\mathcal{I}_T^d} \hat{\gamma}_{j,T,T}(d).$$

Step 3: Recursively estimate blip effects for time $t < T$.
For $d \in [A]$ and $t \in \{T-1, \ldots, 1\}$, recursively estimate as follows:

1. For $j \in \mathcal{I}_t^d$

$$\hat{\gamma}_{j,T,t}(d) = \sum_{k \in \mathcal{I}_t^d \setminus j} \hat{\phi}_k^{j,\mathcal{I}_t^d} \left(Y_{k,T} - \hat{b}_{k,T} - \sum_{\ell=t+1}^{T} \hat{\gamma}_{k,T,\ell}(D_{k,\ell}) \right).$$

2. For $i \notin \mathcal{I}_t^d$

$$\hat{\gamma}_{i,T,t}(d) = \sum_{j \in \mathcal{I}_t^d} \hat{\beta}_j^{i,\mathcal{I}_t^d} \hat{\gamma}_{j,T,t}(d).$$

Step 4: Estimate target causal parameter. For $n \in [N]$, and $\bar{d}^T \in [A]^T$, estimate the

20

causal parameter as follows:

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \sum_{t=1}^{T} \hat{\gamma}_{n,T,t}(d_t) + \hat{b}_{n,T}. \tag{16}$$

All the relevant weights in the above algorithm as computed via the previous PCR based algorithm.

4.4 SBE-PCR Consistency in LTV Setting

4.4.1 Additional Assumptions for Consistency

We state additional assumptions required to establish the consistency of the SBE-PCR estimator.

Assumption 9 (Sub-Gaussian Noise). *For all $n \in [N]$ and $\bar{d}^T \in [A]^T$, $\varepsilon_{n,T}^{(\bar{d}^T)}$ are independent sub-Gaussian random variables with $\mathrm{Var}(\varepsilon_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}) = \sigma^2$ and $\|\varepsilon_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}\|_{\psi_2} \leq C\sigma$ for some constant $C > 0$.*

Assumption 10 (Bounded Expected Potential Outcomes). *For all $n \in [N]$ and $\bar{d}^T \in [A]^T$, we have $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] \in [-1,1]$.*

Assumption 11 (Well-Balanced Singular Values). *For all $d \in [A]$ and $t \in [T]$ we have $\|\mathbb{E}[X_{\mathcal{I}_t^d}|\mathcal{LF}]\|_F \geq c'p|\mathcal{I}_t^d|$ where $X_{\mathcal{I}_t^d} \in \mathbb{R}^{p \times |\mathcal{I}_t^d|}$ is the relevant data matrix of observed covariates and $\kappa^{-1} \geq c$ where κ is the condition number of $\mathbb{E}[X_{\mathcal{I}_t^d}|\mathcal{LF}]$ for constants $c, c' > 0$.*

Assumption 12 (Row-Space Inclusion). *For all $d \in [A]$ and $t \in [T]$ there exist $\{\xi_i^{(d,t)}\}_{i \in [p]}$ such that for any $j \in \mathcal{I}_t^d$*

$$\mathbb{E}[Y_{j,T}|\mathcal{LF}, j \in \mathcal{I}_t^d] = \sum_{i=1}^{p} \xi_i^{(d,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^d})_{ij}|\mathcal{LF}, j \in \mathcal{I}_t^d],$$

where $X_{\mathcal{I}_t^d} \in \mathbb{R}^{p \times |\mathcal{I}_t^d|}$ is the relevant data matrix of observed covariates.[4]

The first three assumptions are standard and identical to those presented in (Agarwal et al., 2020b, Section 4.3). The final assumption facilitates consistency by ensuring that the test data lies within the subspace spanned by the training data—specifically, within its row space—thereby enabling generalization of SBE-PCR. It turns out that this is not a very restrictive assumption and standard within the literature. Appendix C.3 lists a sufficient condition for it and an implication that will help us later on.

---

[4]Here the latent factors we condition upon include the feature vectors $\{\rho_i\}_{i \in [p]}$. We also assume this for donor sets of the form $\mathcal{I}_t^d \setminus n$.

4.4.2 Consistency Results

Theorem 3. *Let assumption 1 to 12 hold. Consider the* `SBE-PCR` *estimator in Section 4.3 and suppose* $k = \max_{d \in [A], t \in [T]} rank(\mathbb{E}[X_{\mathcal{I}_t^d}])$. *Then conditional on the treatment assignments,* $\mathcal{LF}$, *and* $\{\rho_i\}_{i \in [p]}$ *we have:*

(i) Baseline Consistency: For any $n \in [N]$

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} = O_p \left(\sqrt{\log(p|\mathcal{I}_T^0|)} \left(\frac{k^{5/4}}{p^{1/4}} + k^{5/2} \max \left\{ \frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_T^0| - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right).$$

(ii) Terminal Blip Consistency: For any $d \in [A]$ *and unit* $n \in [N]$

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{7/4}}{p^{1/4}} + k^3 \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$ *and* $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$.

(iii) Non-Terminal Blip Consistency: For any $d \in [A]$, *unit* $n \in [N]$, *and* $t \in [1, \ldots, T-1]$:

$$\hat{\gamma}_{n,T,t}(d) - \gamma_{n,T,t}(d) \mid \mathcal{LF}$$
$$= O_p \left((T-t) \sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{(T-t)}}{p^{1/4}} + k^{(T-t)} \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{C} = \{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_q^{D_{n,q}}|)_{n \in [N], q \in [t+1, \ldots, T]}\}$ *with* $\pi_{\mathcal{I}} = \max \mathcal{C}, \alpha_{\mathcal{I}} = \min \mathcal{C}$.

(iv) Target Causal Parameter Consistency: For $n \in [N]$, *and* $\bar{d}^T \in [A]^T$:

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)}] - \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = O_p \left(T \sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^T}{p^{1/4}} + k^T \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{C} = \{|\mathcal{I}_T^0|, (|\mathcal{I}_t^{d_t}|)_{t \in [T]}, (|\mathcal{I}_t^{D_{n,t}}|)_{n \in [N], t \in [2, \ldots, T]}\}$ *with* $\pi_{\mathcal{I}} = \max \mathcal{C}$ *and* $\alpha_{\mathcal{I}} = \min \mathcal{C}$. *Here, each* $O_p(\cdot)$ *is defined with respect to the sequence* $\min\{p, \alpha_{\mathcal{I}}\}$.[5]

Theorem 3 concludes that the `SBE-PCR` estimator is consistent for the causal estimand. More precisely, for a fixed k and T, the estimation error decays as donor set cardinalities and p grow, provided $p = \omega(\pi_{\mathcal{I}}^{1/3})$.[6] Notably, the theorem establishes point-wise consistency, i.e., there is *no* average across units to establish the result. The proof can be found in Appendix C.4.

Assumption 13. *Let the setup of Assumption 4 holds. We further assume the counterfactual potential outcomes depends on the most recent constant* q *blips, namely, for all units* $n \in [N]$ *and* $t \in [q+1, T]$ *we have* $\psi_n^{t, t-q-i} = 0$ *for all* $i \in [t-q-1]$. *Notably, this implies that for any* $n \in [N]$

---

[5]Notice that $\alpha_{\mathcal{I}} \leq \pi_{\mathcal{I}}$ by definition.

[6]To be explicit we are taking $N \to \infty$, $p \to \infty$ and with the additional assumption that each donor set grows at the same rate there is a regime, i.e., relationship between p and N where estimation errors decays.

and $\bar{d}^T \in [A]^T$ we have

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)}|\mathcal{LF}] = \sum_{\ell=T-q}^{T} \langle \psi_n^{T,\ell}, w_{d_\ell} \rangle + \varepsilon_{n,T}^{(\bar{d}^T)}.$$

Theorem 4. *Let the setup of Theorem 3 and Assumption 13 hold. Then modifying the SBE-PCR to only estimate the baseline, terminal blip, and previous q blips we have for any $n \in [N]$, and $\bar{d}^T \in [A]^T$:*

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)}] - \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = O_p\left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^q}{p^{1/4}} + k^q \max\left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{C} = \{|\mathcal{I}_T^0|, (|\mathcal{I}_t^{d_t}|)_{t \in [T-q,...,T]}, (|\mathcal{I}_t^{D_{n,t}}|)_{n \in [N], t \in [T-q+1,...,T]}\}$ with $\pi_\mathcal{I} = \max \mathcal{C}$ and $\alpha_\mathcal{I} = \min \mathcal{C}$.

Theorem 4 concludes that upon modifying the SBE-PCR estimator to account for the system only depending on a *constant q* lags we have a consistent estimate of the causal estimand. More precisely, for fixed k, the estimation error decays as donor set cardinalities and number of covariates p grow, provided $p = \omega(\pi_\mathcal{I}^{1/3})$. Once again we have established pointwise consistency. However, the *key difference* from the previous theorem is that now we allow $T \to \infty$ as well, which justifies the growing number of covariates by including time-varying covariates, i.e., p can now depend on T asymptotically. In the empirical application of Section 6, we include time-varying covariates in estimation. The proof follows immediately from that of Theorem 3 and is included in Appendix C.5.

5 Linear Time-Invariant Latent Factor Model

Next, we introduce a linear time-invariant factor model, which is analogous to the factor model introduced in Assumption 4 in the previous section, but which further exploits the modeling trade-off discussed in Section 4.2.3.

Assumption 14 (Linear time-invariant (LTI) factor model). *Assume $\forall\ n \in [N],\ t \in [T], \bar{d}^t \in [A]^t$,*

$$Y_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t} \left\langle \psi_n^{t-\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t}^{(\bar{d}^t)}, \tag{17}$$

where $\psi_n^{t-\ell}, w_{d_\ell} \in \mathbb{R}^m$ for $\ell \in [t]$. Further, let $\mathcal{LF} = \{\psi_n^{t-\ell}\}_{n \in [N], t \in [T], \ell \in [0, t-1]} \cup \{w_d\}_{d \in [A]}$. Assume

$$\mathbb{E}[\varepsilon_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}] = 0. \tag{18}$$

Remark. *Note Assumption 14 implies Assumption 2 holds with*

$$v_{n,t} = [\psi_n^{t-1}, \ldots, \psi_n^{t-t}], \ w_{\bar{d}^t} = [w_{d_1}, \ldots, w_{d_t}].$$

Further $m(t) = m \times t$ for $m(t)$ in Definition 1.

Note that the effect of action d_ℓ on $Y_{n,t}^{(\bar{d}^t)}$ for $\ell \in [t]$ is additive, given by $\langle \psi_n^{t-\ell}, w_{d_\ell} \rangle$. Intuitively, $\psi_n^{t-\ell}$ captures the latent unit specific heterogeneity in the potential outcome for unit n, at a given time step t, for an action taken at time step $\ell \leq t$; analogously w_{d_ℓ} captures the latent effect of action d_ℓ. Further, compared to Assumption 4, we now have the additional structure that, rather than being dependent on the specific time steps ℓ and t, $\psi_n^{t-\ell}$ is only dependent on the lag $t - \ell$. As a result, the effect of action taken at time ℓ on the outcome at time t is only a function of the lag $t - \ell$. Hence we call this a "time-invariant" latent factor model, as opposed to a "time-varying" latent factor model. This additional structure will be crucial in the identification strategy we employ in Section 5.2.

Non-varying control sequence. For this identification strategy, we make an additional assumption that the control sequence is also time invariant.

Assumption 15. *There exists $\tilde{0} \in [A]$ such that the control sequence $0_t = \tilde{0}$ for all $t \in [T]$.*

5.1 Motivating Example

We show that the classical linear time-invariant dynamical system model satisfies Assumption 14. Suppose for all $t \in [T]$, all units $n \in [N]$ obey the following dynamic triangular model for a sequence of actions $\bar{D}_n^t$ and counterfactual outcomes $Y_{n,t}^{(\bar{D}_n^t)}$:

$$D_{n,t} = f_n(w_{D_{n,t-1}}, z_{n,t-1}^{(\bar{D}_n^{t-1})}), \tag{19}$$

$$Y_{n,t}^{(\bar{D}_n^t)} = \left\langle \theta_n, z_{n,t}^{(\bar{D}_n^t)} \right\rangle + \left\langle \tilde{\theta}_n, w_{D_{n,t}} \right\rangle + \tilde{\eta}_{n,t}, \tag{20}$$

where $z_{n,t}^{(\bar{D}_n^t)} = \boldsymbol{B}_n \, z_{n,t-1}^{(\bar{D}_n^{t-1})} + \boldsymbol{C}_n \, w_{D_{n,t}} + \eta_{n,t}$ and $z_{n,0} = w_{D_{n,0}} = 0$. Here, $z_{n,t}^{(\bar{D}_n^t)} \in \mathbb{R}^m$ is the latent state associated with unit n at time t and $w_{D_{n,t-1}} \in \mathbb{R}^m$ is the chosen action at time $t - 1$. $\eta_{n,t} \in \mathbb{R}^m$ and $\tilde{\eta}_{n,t} \in \mathbb{R}$ represent independent mean-zero random innovations at each time step t. $\boldsymbol{B}_n, \boldsymbol{C}_n \in \mathbb{R}^{m \times m}$ are matrices governing the linear dynamics of $z_{n,t}^{(\bar{D}_n^t)}$. In contrast to the linear time-varying dynamical system described in Section 4.1 above, these transition matrices are invariant across all $t \in [T]$. $\theta_n, \tilde{\theta}_n \in \mathbb{R}^m$ are parameters governing how the outcome of interest $Y_{n,t}^{(\bar{D}_n^t)}$ is a linear function of $z_{n,t}^{(\bar{D}_n^t)}$ and w_{d_t}, respectively. $f_n(\cdot)$ is a function which decides how the next action $w_{D_{n,t}}$ is chosen as a function of the previous action $w_{D_{n,t-1}}$, and current state $z_{n,t}^{(\bar{D}_n^t)}$. We see that due to the input of $z_{n,t}^{(\bar{D}_n^t)}$ in $f_n(\cdot)$, i.e., the action sequence is *adaptive*. As a result, $\eta_{n,\ell}$ is correlated with $D_{n,t}$ for $\ell < t$.

Proposition 3. *Suppose the dynamic triangular model* (19)–(20) *holds. Then we have the following representation,*

$$Y_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t} \left(\left\langle \psi_n^{t-\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t,\ell} \right), \tag{21}$$

where $\psi_n^{t,\ell}, w_{d_\ell} \in \mathbb{R}^m$ for $\ell \in [t]$; here,

$$\psi_n^{t-\ell} = \left(\boldsymbol{B}_n^{t-\ell}\boldsymbol{C}_n\right)' \theta_n \quad for \quad \ell \in [t-1],$$

$$\psi_n^0 = (\boldsymbol{C}_n)'\theta_n + \tilde{\theta}_n,$$

$$\varepsilon_{n,t,\ell} = \left(\boldsymbol{B}_n^{t-\ell}\eta_{n,\ell}\right)' \theta_n \quad for \quad \ell \in [t-1],$$

$$\varepsilon_{n,t,t} = (\eta_{n,t})'\theta_n + \tilde{\eta}_{n,t}.$$

Therefore, Assumption 14 holds with the additional structure that $\varepsilon_{n,t}^{(\bar{d}^t)}$ has an additive factorization as $\sum_{\ell=1}^t \varepsilon_{n,t,\ell}$, and it is not a function of d_ℓ.

In this example, our target parameter $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$ defined in (target causal parameter) translates to the expected potential once we condition on the latent parameters $\psi_n^{t,\ell}, w_{d_\ell}$, which are a function of $\boldsymbol{B}_n, \boldsymbol{C}_n, \theta_n, \tilde{\theta}_n$, and we take the average over the per-step independent mean-zero random innovations, $\varepsilon_{n,t,\ell}$, which is a function of $\eta_{n,\ell}$ (and $\boldsymbol{B}_n, \boldsymbol{C}_n, \theta_n$).

5.2 Identification Strategy

Our goal in this section is to identify $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$, namely represent this expected potential outcome for a target unit n and action sequence $\bar{d}^T$ as some function of observed outcomes.

5.2.1 Notation and Assumptions

Notation. We define the following useful notation for any unit $n \in [N]$ and $t \in [T]$:

$$\gamma_{n,T-t}(d) := \left\langle \psi_n^{T-t}, w_d - w_{\tilde{0}} \right\rangle.$$

The quantity $\gamma_{n,T-t}(d)$ can be interpreted as a "blip effect"—the expected difference in potential outcomes if unit n undergoes the sequence $(\bar{d}^{T-t}, \underline{0}^{T-t+1})$ instead of $(\bar{d}^{T-t-1}, \underline{0}^{T-t})$. This is because, Assumption 14 and 15 imply

$$\mathbb{E}\left[Y_{n,T}^{(\bar{d}^{T-t}, \underline{0}^{T-t+1})} - Y_{n,T}^{(\bar{d}^{T-t-1}, \underline{0}^{T-t})} \mid \mathcal{LF}\right]$$
$$= \mathbb{E}\left[\left\langle \psi_n^{T-t}, w_{d_t} - w_{\tilde{0}} \right\rangle + \varepsilon_{n,T}^{(\bar{d}^{T-t}, \underline{0}^{T-t+1})} - \varepsilon_{n,T}^{(\bar{d}^{T-t-1}, \underline{0}^{T-t})} \mid \mathcal{LF}\right]$$
$$= \left\langle \psi_n^{T-t}, w_{d_t} - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}.$$

Further, let

$$b_{n,T} := \sum_{\ell=1}^T \left\langle \psi_n^{T-\ell}, w_{\tilde{0}} \right\rangle.$$

This can be interpreted as the expected potential outcome if unit n remains under the control sequence $\bar{0}^T$ till time step T. Again, Assumption 14 and 15 imply

$$\mathbb{E}\left[Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}\right] = \mathbb{E}\left[\sum_{t=1}^{T}\left\langle \psi_n^{T,t}, w_{\tilde{0}}\right\rangle + \varepsilon_{n,t}^{(\bar{0}^T)} \mid \mathcal{LF}\right] = \sum_{t=1}^{T}\left\langle \psi_n^{T,t}, w_{\tilde{0}}\right\rangle \mid \mathcal{LF}. \tag{22}$$

Assumptions. We now state identifying assumptions.

Donor sets. We define two distinct subsets of units based on the treatment sequence they receive:

$$
\begin{aligned}
\mathcal{I}^d := \{j \in [N] : (i)\ &\bar{D}_j^{t_j^*} = (\tilde{0}, \ldots, \tilde{0}, d),\\
(ii)\ &\forall\ \bar{\delta}^t \in [A]^t, t \in [T],\ \mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}, \bar{D}_j^{t_j^*}] = \mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}]\}, \tag{23}\\
\mathcal{I}_t^0 := \{j \in [N] : (i)\ &\bar{D}_j^t = (\tilde{0}, \ldots, \tilde{0}),\\
(ii)\ &\forall\ \bar{\delta}^\ell \in [A]^\ell, \ell \in [T],\ \mathbb{E}[Y_{j,\ell}^{(\bar{\delta}^\ell)} \mid \mathcal{LF}, \bar{D}_j^t] = \mathbb{E}[Y_{j,\ell}^{(\bar{\delta}^\ell)} \mid \mathcal{LF}]\}. \tag{24}
\end{aligned}
$$

The donor set $\mathcal{I}^d$ contains units that remain under the control sequence $(\tilde{0}, \ldots, \tilde{0})$ till time step $t_j^* - 1$, and at time step t_j^* receive action d. Further, we require that for these particular units, the action sequence, $\bar{D}_j^{t_j^*}$, till time step t_j^* was chosen such that $\mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}, \bar{D}_j^{t_j^*}] = \mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}]$ for all $\bar{\delta}^t \in [A]^t$, i.e., the potential outcomes are conditionally mean independent of the action sequence $\bar{D}_j^{t_j^*}$ unit j receives till time step t_j^*. Of course, a sufficient condition for property (ii) above is that $\forall\ \bar{\delta}^t \in [A]^t$, $Y_{j,t}^{(\bar{\delta}^t)} \perp \bar{D}_j^{t_j^*} \mid \mathcal{LF}$. That is, for these units, the action sequence till time step t_j^* is chosen at $t = 0$ conditional on the latent factors, i.e., the policy for these units can only be adaptive from time step $t_j^* + 1$. Note, given Assumption 14, (23) can be equivalently stated as $\mathbb{E}[\varepsilon_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}, \bar{D}_j^{t_j^*}] = \mathbb{E}[\varepsilon_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}] = 0$. The donor set $\mathcal{I}_t^0$ follows a similar intuition to that of $\mathcal{I}^d$.

Assumption 16. *For $n \in [N]$, let $v_{n,T} := [\psi_n^0, \ldots, \psi_n^{T-1}]$. We assume that for all $n \in [N]$, $v_{n,T}$ satisfies a well-supported condition with respect to the various donor sets, i.e., for all $d \in [A]$ there exists $\beta^{n,\mathcal{I}^d} \in \mathbb{R}^{|\mathcal{I}^d|}$, and $\beta^{n,\mathcal{I}_t^0} \in \mathbb{R}^{|\mathcal{I}_t^0|}$ such that*

$$v_{n,T} = \sum_{k \in \mathcal{I}^d} \beta_k^{n,\mathcal{I}^d} v_{k,T}, \quad v_{n,T} = \sum_{k \in \mathcal{I}_t^0} \beta_k^{n,\mathcal{I}_t^0} v_{k,T}. \qquad \text{(LTI well-supported factors)}$$

Assumption 16 essentially states that for the various units $n \in [N]$, their latent factors $[\psi_n^{T-1}, \ldots, \psi_n^{T-T}]$ are expressible as a linear combination of the units in the donor sets $\mathcal{I}^d$ and $\mathcal{I}_t^0$. See the discussion under Assumption 3 in Section 3 justifying such an assumption for settings where $\mathcal{I}^d$ and $\mathcal{I}_t^0$ are sufficiently large.

Assumption 17. *For all $n \in [N], t \in [T], \bar{\delta}^t \in [A]^t$,*

$$\mathbb{E}\left[\varepsilon_{n,t}^{(\bar{\delta}^t)} - \varepsilon_{n,t}^{(\bar{\delta}^{t-1}, \tilde{0})} \mid \bar{D}_n^t = \bar{\delta}^t,\ \mathcal{LF}\right] = 0. \tag{25}$$

For sufficient conditions under which Assumption 17 holds, see the discussion under Assumption

6 in Section 4.2—an analogous argument holds here. Also an analogous version of Proposition 2 holds for the linear time-invariant setting using an identical argument.

5.2.2 Identification Result

Given these assumptions, we now present our identification theorem.

Theorem 5. *Let Assumptions 1, 14, 15, 16, and 17 hold. Then, for any unit $n \in [N]$ and action sequence $\bar{d}^T \in [A]^T$, the expected counterfactual outcome can be expressed as:*

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \sum_{t=1}^{T} \gamma_{n,T-t}(d_t) + b_{n,T} \mid \mathcal{LF}, \qquad \text{(identification)}$$

where quantities on the right-hand side are identified as follows:

(i) We have the following representations of the baseline outcomes for all $t \in [T]$

$$\forall\, j \in \mathcal{I}_t^0 \;:\; b_{j,t} \mid \mathcal{LF} = \mathbb{E}[Y_{j,t} \mid \mathcal{LF},\; j \in \mathcal{I}_t^0], \qquad \text{(observed control)}$$

$$\forall\, i \notin \mathcal{I}_t^0 \;:\; b_{i,t} \mid \mathcal{LF} = \sum_{j \in \mathcal{I}_t^0} \beta_j^{i,\mathcal{I}_t^0}\, b_{j,t} \mid \mathcal{LF}, \mathcal{I}_T^0. \qquad \text{(synthetic control)}$$

(ii) We have the following representations of the blip effect with 0 lag, for $\forall d \in [A]$:

$$\forall\, j \in \mathcal{I}^d \;:\; \gamma_{j,0}(d) \mid \mathcal{LF} = \mathbb{E}[Y_{j,t_j^*} \mid \mathcal{LF},\; j \in \mathcal{I}^d] - b_{j,t_j^*} \mid \mathcal{LF}, \qquad \text{(``observed'' lag 0 blip)}$$

$$\forall\, i \notin \mathcal{I}^d \;:\; \gamma_{i,0}(d) \mid \mathcal{LF} = \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \gamma_{j,0}(d) \mid \mathcal{LF}, \mathcal{I}^d. \qquad \text{(synthetic lag 0 blip)}$$

(iii) We have the following recursive representations of the blip effect $\forall\, t \in [T-1],\; d \in [A]$: [7]

$$\forall\, j \in \mathcal{I}^d \;:\; \gamma_{j,t}(d) \mid \mathcal{LF} = \mathbb{E}[Y_{j,t_j^*+t} \mid \mathcal{LF}, \mathcal{I}^d] - b_{j,t_j^*+t} \mid \mathcal{LF} - \sum_{\ell=0}^{t-1} \gamma_{j,\ell}(D_{j,t_j^*+t-\ell}) \mid \mathcal{LF},$$

$$\text{(``observed'' lag } t \text{ blip)}$$

$$\forall\, i \notin \mathcal{I}^d \;:\; \gamma_{i,t}(d) \mid \mathcal{LF} = \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \gamma_{j,t}(d) \mid \mathcal{LF}, \mathcal{I}^d. \qquad \text{(synthetic lag } t \text{ blip)}$$

Interpretation of identification result. (identification) states that our target causal parameter of interest can be written as an additive function of $b_{n,T}$ and $\gamma_{n,T-t}(d_t)$ for $t \in [T]$ and $d_t \in [A]$. Theorem 5 establishes that these various quantities are expressible as functions of observed outcomes $\{Y_{j,t}\}_{j \in [N], t \in [T]}$. We give an interpretation below.

Identifying baseline outcomes. Similar to the intuition for Theorem 2, for units $j \in \mathcal{I}_t^0$, (observed control) states that their baseline outcome $b_{j,t}$ is simply their expected observed outcome at time step t, i.e., $Y_{j,t}$. For units $i \notin \mathcal{I}_t^0$, (synthetic control) states that we can identify

---

[7] We implicitly assume we have access to outcomes till time step $2T - 1$. which we assume to be true without loss of generality. To see why consider $t_n^* = T$ and $t = T - 1$.

$b_{i,t}$ by appropriately re-weighting the baseline outcomes $b_{j,t}$ of the units $j \in \mathcal{I}_t^0$ (identified via (observed control)).

Identifying blip effects for lag 0. For any given $d \in [A]$: For units $j \in \mathcal{I}^d$, ("observed" lag 0 blip) states that their blip effect $\gamma_{j,0}(d)$ is equal to their observed outcome Y_{j,t_j^*} minus the baseline outcome b_{j,t_j^*} (identified via (synthetic control)). Recall t_j^* is equal to the first time step that unit j is no longer in the control sequence. For units $i \notin \mathcal{I}^d$, (synthetic lag 0 blip) states that we can identify $\gamma_{i,0}(d)$ by appropriately re-weighting the blip effects $\gamma_{j,0}(d)$ of units $j \in \mathcal{I}^d$ (identified via ("observed" lag 0 blip)).

Identifying blip effects for lag t *with* $t \in [T-1]$. Suppose by induction $\gamma_{n,\ell}(d)$ is identified for every lag $\ell < t$, $n \in [N]$, $d \in [A]$, i.e., can be expressed in terms of observed outcomes. Then for any given $d \in [A]$: For units $j \in \mathcal{I}^d$, ("observed" lag t blip) states that their blip effect $\gamma_{j,t}(d)$ is equal to their their observed outcome at time step $t_j^* + t$, Y_{j,t_j^*+t}, minus the baseline outcome b_{j,t_j^*+t} (identified via (synthetic control)) minus the sum of blip effects for smaller lags, $\sum_{\ell=0}^{t-1} \gamma_{j,\ell}(D_{j,t_j^*+t-\ell})$ (identified via the inductive hypothesis). For units $i \notin \mathcal{I}^d$, (synthetic lag t blip) states that we can identify $\gamma_{i,t}(d)$ by appropriately re-weighting the blip effects $\gamma_{j,t}(d)$ of units $j \in \mathcal{I}^d$ (identified via ("observed" lag t blip)).

5.2.3 Discussions: LTI Identification Strategy

Donor sample complexity. To estimate $\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}]$ for all units $n \in [N]$ and any action sequence $\bar{d}^T \in [A]^T$, the LTI identification strategy requires the existence of a sufficiently large subset of donor units $\mathcal{I}^d$ for every $d \in [A]$ and $\mathcal{I}_t^0$ for $t \in [T]$. That is, the number of donor units we require will need to scale at the order of A to ensure sufficient number of units for the donor sets $\{\mathcal{I}^d\}_{d \in [A]}$. To ensure that we have sufficient number of donors units for $\mathcal{I}_t^0$ for $t \in [T]$. But notice from the definition of $\mathcal{I}_t^0$ that for all $t \in [T-1]$, $\mathcal{I}_t^0 \subset \mathcal{I}_T^0$. Hence, we just require that $\mathcal{I}_T^0$ is sufficiently large. As a result the total donor sample complexity needs to scale at the order of $A + 1$. Thus we see the the additional structure imposed by the time-invariant factor model introduced in Assumption 14 leads to a decrease in sample complexity from $A \times T$ to $A + 1$, when compared with the time-varying factor model factor model introduced in Assumption 4. The other major assumption made is that the control sequence is also not time varying, see Assumption 15.

Donor exogeneity conditions. Further, for $j \in \mathcal{I}^d$, we require that $\forall \bar{\delta}^t \in [A]^t, t \in [T]$, $\mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}, \bar{D}_j^{t_j^*}] = \mathbb{E}[Y_{j,t}^{(\bar{\delta}^t)} \mid \mathcal{LF}]$. That is, the actions picked for these donor units are only required to be non-adaptive till time step t_j^*. As a special case, if we restrict ourselves to units

$$\tilde{\mathcal{I}}^d := \{j \in \mathcal{I}^d : t_j^* = 1\}, \tag{26}$$

then we actually impose no exogeneity conditions. That is, for these donor units, their entire action sequence can be adaptive. In contrast for the identification strategy in Section 4, we require that the donor units in $\mathcal{I}_t^d$ are non-adaptive till time step t. See Figure 3 for a DAG that is consistent with the exogeneity conditions implied by the definition of $\tilde{\mathcal{I}}^d$ in (10).
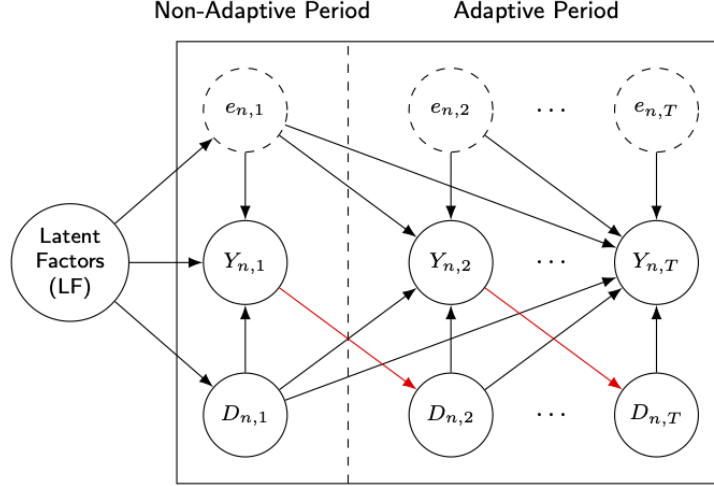
Figure 3: DAG that is consistent with the exogeneity conditions implied by the definition of $\tilde{\mathcal{I}}^d$. From time step 2, the action sequence $(D_{n,2}, \ldots, D_{n,T})$ can be adaptive, i.e., dependent on the observed outcomes $\{Y_{n,t}\}_{t\in[T]}$ (depicted by the red arrows). Hence, there is *no* non-adaptive period for these units.

5.3 `SBE-PCR` Estimator in LTI Setting

Now we detail the specific algorithm that yields the `SBE-PCR` estimator within the linear time-invariant setting. Analogous to the LTV case, we consider additional covariates with the usual factor decomposition (Assumption 7) and make an additional well supported assumption regarding control factors for consistency.

Assumption 18. *For any donor set, i.e., any $d \in [A]$ and $n \in \mathcal{I}^d$ there exist weights $\phi^{n,\mathcal{I}^d}$ such that*

$$v_{n,T} = \sum_{k\in\mathcal{I}^d\backslash n} \phi_k^{n,\mathcal{I}^d} \cdot v_{k,T},$$

and for any $t \in [T]$, there exist weights $\phi^{n,\mathcal{I}_t^0}$ such that

$$v_{n,T} = \sum_{k\in\mathcal{I}_t^0\backslash n} \phi_k^{n,\mathcal{I}_t^0} \cdot v_{k,T},$$

where $v_{n,T} = \left[\psi_n^{T-1}, \ldots, \psi_n^0\right]$.

This assumption allows us to detail the algorithm for estimating weights using PCR. Specifically, for each unit $n \in [N]$ and each donor set $\mathcal{I} \in \{\mathcal{I}^d, \mathcal{I}_t^0\}$, we estimate weights to express X_n as a linear combination of the covariates from other donor units. Let $X_n \in \mathbb{R}^p$ be the observed covariate vector for unit n, and let $X_{\mathcal{I}\backslash n} \in \mathbb{R}^{p\times(|\mathcal{I}|-1)}$ denote the matrix of covariates for the other units in the donor set.

We perform PCR by computing the rank-k approximation of $X_{\mathcal{I}\backslash n}$, where $k = \text{rank}(\mathbb{E}[X_{\mathcal{I}\backslash n}])$.

Denote the SVD as

$$X_{\mathcal{I}\backslash n} = \sum_{l\geq 1} \sigma_l u_l v_l^\top,$$

where $u_l \in \mathbb{R}^p$, $v_l \in \mathbb{R}^{|\mathcal{I}|-1}$, and σ_l are sorted in descending order. If $n \in \mathcal{I}$,

$$\hat{\phi}^{n,\mathcal{I}} = \left(\sum_{l=1}^{k}(1/\sigma_l)v_l u_l^\top\right) X_n \in \mathbb{R}^{|\mathcal{I}|-1},$$

and if $n \notin \mathcal{I}$,

$$\hat{\beta}^{n,\mathcal{I}} = \left(\sum_{l=1}^{k}(1/\sigma_l)v_l u_l^\top\right) X_n \in \mathbb{R}^{|\mathcal{I}|}.$$

The distinction between $\hat{\phi}^{n,\mathcal{I}}$ and $\hat{\beta}^{n,\mathcal{I}}$ lies in whether the unit is part of the donor set (interpolation) or not (extrapolation), which has implications for estimator variance.

Step 1: Estimate baseline outcomes.

For $t \in [T]$:

1. For $j \in \mathcal{I}_t^0$

$$\hat{b}_{j,t} = \sum_{k\in\mathcal{I}_t^0\backslash j} \hat{\phi}_k^{j,\mathcal{I}_t^0} Y_{k,t}.$$

2. For $i \notin \mathcal{I}_t^0$

$$\hat{b}_{i,t} = \sum_{j\in\mathcal{I}_t^0} \hat{\beta}_j^{i,\mathcal{I}_t^0} Y_{j,t}.$$

Step 2: Estimate blip effects for lag 0.

For $d \in [A]$:

1. For $j \in \mathcal{I}^d$

$$\hat{\gamma}_{j,0}(d) = \sum_{k\in\mathcal{I}^d\backslash j} \hat{\phi}_k^{j,\mathcal{I}^d}\left(Y_{k,t_k^*} - \hat{b}_{k,t_k^*}\right).$$

2. For $i \notin \mathcal{I}^d$

$$\hat{\gamma}_{i,0}(d) = \sum_{j\in\mathcal{I}^d} \hat{\beta}_j^{i,\mathcal{I}^d}\hat{\gamma}_{j,0}(d).$$

Step 3: Recursively estimate blip effects for time $t < T$.

For $d \in [A]$ and $t \in \{1,\ldots,T-1\}$, recursively estimate as follows:

1. For $j \in \mathcal{I}^d$

$$\hat{\gamma}_{j,t}(d) = \sum_{k \in \mathcal{I}^d \setminus j} \hat{\phi}_k^{j,\mathcal{I}^d} \left(Y_{k,t_k^*+t} - \hat{b}_{k,t_k^*+t} - \sum_{\ell=0}^{t-1} \hat{\gamma}_{k,\ell}(D_{k,t_k^*+t-\ell}) \right).$$

2. For $i \notin \mathcal{I}^d$

$$\hat{\gamma}_{i,t}(d) = \sum_{j \in \mathcal{I}^d} \hat{\beta}_j^{i,\mathcal{I}^d} \hat{\gamma}_{j,t}(d).$$

Step 4: Estimate target causal parameter. For $n \in [N]$, and $\bar{d}^T \in [A]^T$, estimate the causal parameter as follows:

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \sum_{t=1}^{T} \hat{\gamma}_{n,T-t}(d_t) + \hat{b}_{n,T}. \tag{27}$$

All the relevant weights in the above algorithm as computed via the previous PCR based algorithm.

5.4 `SBE-PCR` Consistency in LTI Setting

5.4.1 Additional Assumptions for Consistency

We now state the assumptions required for consistency of the `SBE-PCR` estimator under the LTI latent factor model. These assumptions parallel those in the LTV setting, with simplifications reflecting the time-invariant latent structure. We unify the donor set notation by writing $\mathcal{I} \in \{\mathcal{I}^d : d \in [A]\} \cup \{\mathcal{I}_t^0 : t \in [T]\}$, and refer to the relevant donor set generically as $\mathcal{I}$. Assumptions 9 and 10 in Section 4.4 are maintained here.

Assumption 19 (Well-Balanced Singular Values)**.** *For each donor set $\mathcal{I}$, the covariate matrix $X_\mathcal{I} \in \mathbb{R}^{p \times |\mathcal{I}|}$ satisfies:*
$$\|\mathbb{E}[X_\mathcal{I}|\mathcal{LF}]\|_F \geq c'p|\mathcal{I}|, \quad and \quad \kappa^{-1} \geq c,$$

where κ is the condition number of $\mathbb{E}[X_\mathcal{I}|\mathcal{LF}]$, and $c, c' > 0$ are constants.

Assumption 20 (Row-Space Inclusion)**.** *For any $t \in [T]$ we require existence of the weights $\xi^{(0,t)} \in \mathbb{R}^p$ such that for any $j \in \mathcal{I}_t^0$*

$$\mathbb{E}[Y_{j,t}|\mathcal{LF}, j \in \mathcal{I}_t^0] = \sum_{i=1}^{p} \xi_i^{(0,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^0})_{ij}|\mathcal{LF}, j \in \mathcal{I}_t^0],$$

for any $t \geq 0$ and $j \in \mathcal{I}^d$ there exist $\xi^{(d,t)} \in \mathbb{R}^p$ such that

$$\mathbb{E}[Y_{j,t_j^*+t}|\mathcal{LF}, j \in \mathcal{I}^d] = \sum_{i=1}^{p} \xi_i^{(d,t)} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij}|\mathcal{LF}, j \in \mathcal{I}^d],$$

31

and for any $t \geq 0$ and $j \in \mathcal{I}^d$ there exists $\alpha^{(0,t)} \in \mathbb{R}^p$ such that

$$\mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{0}^{t_j^*+t})}\big|\mathcal{LF}, j \in \mathcal{I}^d\right] = \sum_{i=1}^{p} \alpha_i^{(0,t)} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij}|\mathcal{LF}, j \in \mathcal{I}^d].$$

Assumption 20 is similar to Assumption 12, but using donor sets $\mathcal{I} \in \{\mathcal{I}^d\} \cup \{\mathcal{I}_t^0\}$ relevant to the LTI setting.

5.4.2 Consistency Results

We only present the main consistency theorem that allows T to grow; consistent results for fixed T that serve as preliminaries for proving this theorem are contained in Appendix D.4.

Assumption 21. *Let the setup of Assumption 14 hold. We further assume the counterfactual potential outcomes depends on the most recent constant q blips, namely, for all units $n \in [N]$ we have $\psi_n^{q+i} = 0$ for all $i \in [T - q - 1]$. Notably, this implies that for any $n \in [N]$ and $\bar{d}^T \in [A]^T$ we have*

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)}|\mathcal{LF}] = \sum_{\ell=T-q}^{T} \langle \psi_n^{T-\ell}, w_{d_\ell} \rangle + \varepsilon_{n,T}^{(\bar{d}^T)}.$$

Theorem 6. *Let Assumption 1 to 7, 9, 10, and 14 to 21 hold. Consider the `SBE-PCR` estimator in Section 5.3 modified to only estimate the baseline, terminal blip, and previous q blips, and suppose $k = \max_{\mathcal{I} \in \{\mathcal{I}^d\} \cup \{\mathcal{I}_t^0\}} rank(\mathbb{E}[X_{\mathcal{I}}])$. Then we have for any $n \in [N]$, and $\bar{d}^T \in [A]^T$:*

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)}] - \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^q}{p^{1/4}} + k^q \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{C} = \{|\mathcal{I}_T^0|, |\mathcal{I}_1^0|, (|\mathcal{I}^{d_t}|)_{t \in [T-q,...,T]}, (|\mathcal{I}^{D_{n,t_n^*+t}}|)_{n \in [N], t \in [1,...,q]}\}$ with $\pi_{\mathcal{I}} = \max \mathcal{C}$ and $\alpha_{\mathcal{I}} = \min \mathcal{C}$.

Theorem 6 concludes that upon modifying the `SBE-PCR` estimator to account for the system only depending on a *constant q* lags we have a consistent estimator of the causal estimand. More precisely, for fixed k, the estimation error decays as donor set cardinalities and number of covariates p grow, provided $p = \omega(\pi_{\mathcal{I}}^{1/3})$. The growing number of covariates can be justified by including time varying covariates with $T \to \infty$. Again, we have established pointwise consistency, i.e., there is *no* average across units to establish the result. The theorem's proof is included in Appendix D.5.

Remark. To gain some intuition about the difference in order between the minimum donor set cardinality and the maximum donor set cardinality appearing in the error rate bound, observe the following. Suppose the maximum $\pi_{\mathcal{I}}$ is attained at $|\mathcal{I}_T^0|$ while the minimum $\alpha_{\mathcal{I}}$ is attained at $|\mathcal{I}_1^0|$. Since the donor sets are strictly nested, $\mathcal{I}_1^0 \subset \mathcal{I}_2^0 \subset \cdots \subset \mathcal{I}_T^0$, their cardinalities are strictly increasing in t. It follows that the ratio between the maximum and minimum cardinalities grows by at least a multiplicative factor of order T. Even if the extremal values are not realized at the time dependent donor sets, the strict nesting still guarantees a gap of order at least T between the minimum and maximum donor set cardinalities.

6 Application: Export Financial Support

The goal of this section is to showcase the usefulness of our approach in understanding individual dynamic treatment effects and developing optimal targeting rules in a real-world application where panel data are available. We first introduce the backgrounds on financial credit support for exporting firms and data (Sections 6.1 and 6.2) and report the synthetic blip estimates of support impacts (Section 6.3). We then investigate the extent of possible improvement of support allocation for each firm (Section 6.4.1) and develop an optimal targeting rule for allocating support based on firm characteristics (Section 6.4.2).

6.1 Backgrounds

Exporting is inherently risky, requiring firms to secure upfront working capital, offer extended payment terms, and protect themselves against non-payment or foreign market shocks. When trade finance dried up during the Great Recession, the resulting contraction disproportionately hit firms reliant on weak banks or operating in finance-dependent sectors (Amiti and Weinstein, 2011; Chor and Manova, 2012; Paravisini et al., 2015). These vulnerabilities matter particularly in economies that rely heavily on international trade. The Korean economy is a great example of this, as exports accounted for 45–58% of Gross Domestic Product (GDP) between 2006 and 2015, making the economy highly sensitive to fluctuations in global trade and financial conditions. This dependence heightens exposure to geopolitical frictions, as Korea sits between China and the United States, where tariff disputes, supply-chain tensions, and restrictions on key sectors regularly generate uncertainty. In this environment, ECAs play a critical role by using public funds to provide insurance and loans that enable firms to sustain and expand their export activities.

Korea has two independent ECAs: the Korea Trade Insurance Corporation (K-SURE), which specializes in export insurance, and the Export-Import Bank of Korea (EXIM), which provides export loans. Firms seeking support apply through the relevant agency: K-SURE assesses the creditworthiness of exporters and their foreign buyers, while EXIM evaluates financial stability and contract documents. In practice, these agencies evaluate applications and select firms to support based on firm characteristics—an approach that connects to our later analyses of heterogeneity and optimal treatment allocation. Also, the agencies' selections are independent of one another, leaving room for potential improvements through communication and collaboration.

Using Korean firm-level data described below, we empirically examine how export credit support shapes firm performance and how its allocation can be improved. Estimating treatment effects is a necessary first step, but the policy challenge goes further: agencies must decide which firms to support, through which instruments, and at what point in time. By comparing observed allocation patterns with the counterfactual benchmark predicted by our model, we demonstrate how more efficient targeting could deliver greater export growth with the same or fewer government resources. We further extend the optimal allocation analysis by allowing the rule to depend on firm characteristics. This motivation stems from heterogeneity in impacts across firm size, productivity,

and financial constraints. Such a framework is particularly useful for making allocation decisions about newly entering firms and closely mirrors the agencies' own selection processes.

6.2 Data and Variables

Our empirical analysis relies on a novel Korean firm-level panel dataset for 2006–2015 that links three sources of firm-level data: (i) the Survey of Business Activities (SBA) from Statistics Korea,[8] which provides detailed firm characteristics; (ii) export insurance data from the K-SURE; and (iii) export loan data from the EXIM. Combining these sources allows us to track which firms received support, the form and timing of support, and their subsequent performances. We define the treatment group as firms that did not receive support in the first five years (2006–2010) but received at least one form of support in the later period (2011–2015). The control group consists of firms that were never supported during the sample period. Out of 2,052 unique firms, 167 received support at least once in the later period.

The outcome of interest is the export value of firm n in year t, $Y_{n,t}$. The vector of firm-level covariates, $X_{n,t}$, consists of eleven time-varying and two time-invariant variables. The time-varying covariates include exports relative to sales (export share), sales, number of workers, tangible capital stock, value-added, total factor productivity (TFP), total wage bill, R&D expenditure, debt-to-asset ratio, current assets over current liabilities (liquidity ratio), and a dummy for foreign direct investment (FDI).[9] The time-invariant firm covariates include an indicator for parent-company affiliation and the firm's age. At the industry level, we control for Z_m, a vector of two indicators for whether industry m has an above average capital intensity and above average wage per worker.[10]

In each year $t \in \{T_0 + 1, ..., T\}$, firm n receives treatment $D_n \in [A]_0 = \{0, 1, 2, 3\}$: $D_n = 1$ if firm n receives insurance, $D_n = 2$ if it receives loans, $D_n = 3$ if it receives both, and $D_n = 0$ indicates it receives none. In this application, $t_n^* = T_0 = 5$ for all n and $T = 10$ (i.e., no firm receives treatment until period T_0) and we redefine $\bar{d}^t = (d_{T_0+1}, ..., d_t)$ for notational simplicity. We assume the LTV latent factor model (Section 4) and the number of lags to be $q = 1$ in Assumption 13. Using the SBE-PCR algorithm, we estimate $\mathbb{E}[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}]$ for given $\bar{d}^t \in [A]_0^t$ and for each firm n and $t \geq T_0 + q + 1 = 7$ (i.e., the last four periods) in the data. These quantities are the crucial ingredient for all the analyses below: they are used to calculate average counterfactual outcomes and average treatment effects and to conduct policy learning.

---

[8]This annual survey provides detailed information on inputs, outputs, and trade activities of all firms with at least 50 employees and annual sales exceeding 300 million KRW (around 215K USD).

[9]TFP is measured as value-added divided by $K^{1/3}L^{2/3}$, where K denotes tangible capital stock and L the number of workers. Sales, tangible capital stock, value-added, total wage bill, and R&D expenditure are expressed in natural logarithms, with the underlying unit being million KRW (around 720 USD).

[10]Capital intensity is defined as tangible capital stock per worker, and wage per worker is measured as the total wage bill divided by the number of workers.
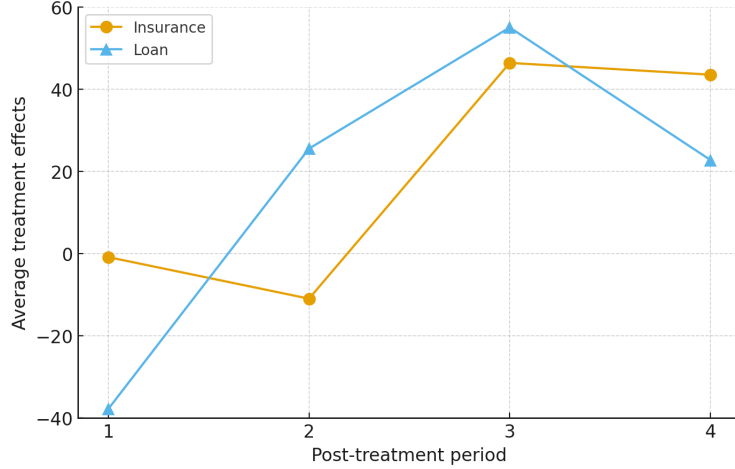
Figure 4: Dynamic Treatment Effects on Export Values

Notes: The figure depicts the trajectory of average treatment effects across firms measured in export value, with units in billions of KRW (around 720K USD). The cumulative average treatment effect amounts to 78.2 for insurance and 65.6 for loans.

6.3 Dynamic Effects of Financial Support

To understand time-varying effects of the financial support, we report the trajectories of various average counterfactual outcomes and average dynamic treatment effects. First, we estimate the average treatment effects of financial support relative to no intervention, where *the average is taken across firms*. Each support sequence is defined as $\bar{d}^T = (d, d, d, d, d)$ for $d = 1$ or 2 (again, 1 being insurance support and 2 being loans support). Figure 4 shows distinct patterns between the two treatments over post-treatment periods. Insurance has little effect initially but generates sizable gains from the third year onward, consistent with insurance stabilizing performance and supporting longer-run growth. Evidence on export credit insurance similarly shows that risk-mitigation instruments help sustain trade by reducing uncertainty rather than generating immediate effects (Niepmann and Schmidt-Eisenlohr, 2017). Loans show a negative effect in the first year, followed by positive effects that bring cumulative gains close to insurance. A natural interpretation is that early loan use covers input costs before output materializes, depressing short-run outcomes but enabling later expansion. This dynamic is consistent with the working-capital channel, where financing upfront input costs can depress short-run outcomes before revenues are realized (Schmidt-Eisenlohr, 2013; Antras and Foley, 2015; Paravisini et al., 2015). The cumulative average treatment effect—i.e., the sum of effects across four years—amounts to 78.2 billion KRW for insurance and 65.6 billion KRW for loans.

We next turn to a sequence-specific analysis. We attempt to understand how timing (but not the total amount of support) affects the trajectory of outcomes by considering the following: (i) a front-loading treatment $\bar{d}^T = (d, d, d, 0, 0)$ for either $d = 1$ or 2, which concentrates support at the start; (ii) an even-loading treatment $(d, 0, d, 0, d)$, which spreads it evenly; and (iii) a back-loading
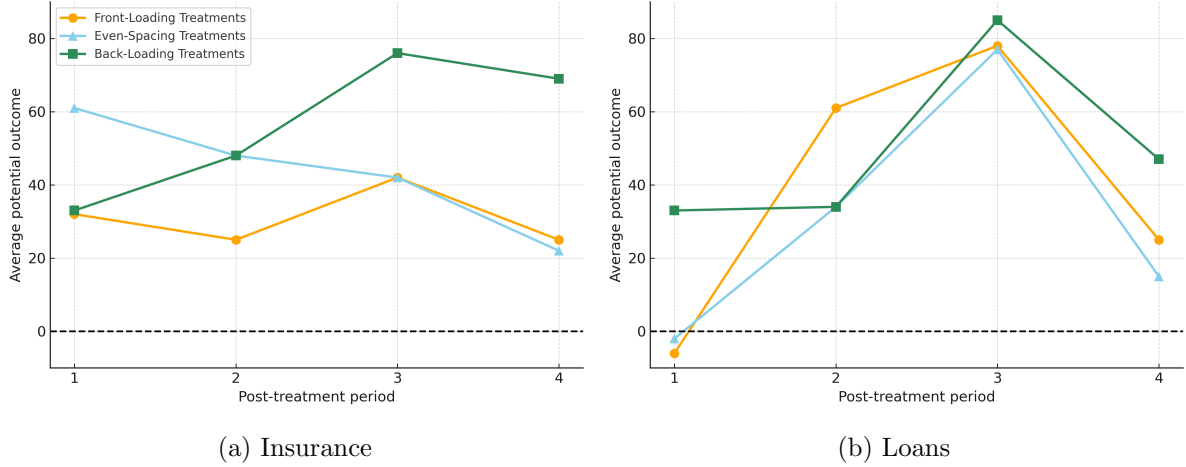
(a) Insurance   (b) Loans

Figure 5: Potential Export Values Under Front-, Even- and Back-Loading Treatment Schedules

Notes: The figure depicts the trajectory of counterfactual export values under front-, even- and back-loading hypothetical treatment schedules. The outcome is export value, with units in billions of KRW (around 720K USD). The cumulative average potential export value amounts to 123.36 for front-loading, 173.09 for evenly-loading, and 225.21 for back-loading insurance. For loans, the corresponding values are 158.06 for front-loading, 123.73 for evenly-loading, and 199.13 for back-loading support.

treatment $(0,0,d,d,d)$, which defers it to the end.[11] Figure 5 reports average potential outcomes under these strategies, with panel (a) presenting insurance and panel (b) loans. For insurance, the cumulative average potential export value is 123.36 for front, 173.09 for even, and 225.21 for back insurance, all in billion KRW. For loans, the corresponding values are 158.06 for front, 123.73 for even, and 199.13 for back support. For insurance, back-loading produces the largest cumulative gains, while even-loading underperforms, suggesting that distributing support thinly is less effective than concentrating it. For loans, back-loading again dominates, with even-loading weaker than front-loading, consistent with credit being most valuable when timed around production peaks. Overall, the results indicate that not only timing but also spacing matters: smoothing interventions across periods is generally less effective than concentrating them, though the optimal pattern varies by treatment types.

6.4 Optimal Targeting of Financial Support

In providing financial support, each ECA has its own rules for selecting export firms. It would be interesting to investigate (i) whether better (statistical) selection rules could have been used for each support program compared to the observed selections, (ii) whether collaboration among agencies in the selection process would have led to gains, and (iii) what selection rule could be implemented for new firms. Questions (i) and (ii) relate to *retrospective policy learning*, while (iii) corresponds to *prospective policy learning*.

---

[11]The analysis with $\bar{d}^T = (d,d,0,0,0)$, $(0,d,0,d,0)$, and $(0,0,0,d,d)$ produces a similar result, especially for insurance as support.

6.4.1 Retrospective Policy Learning

For each firm n in the data, we consider the optimal treatment schedule $\bar{d}^{T*}(n) \in \mathcal{D}$ that maximizes the aggregate outcome:

$$\bar{d}^{T*}(n) \in \operatorname*{argmax}_{\bar{d}^t \in \mathcal{D}} \sum_{t=T_0+q+1}^{T} \mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}\right], \tag{28}$$

where $\mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}\right]$ is estimated using the `SBE-PCR` algorithm. Here the set of possible schedules $\mathcal{D}$ can be restricted for institutional reasons or due to budget constraints. The example of the latter would be $\mathcal{D} = \{\bar{d}^T : \sum_{t=T_0+q+1}^{T} p_{d_t} \cdot d_t \leq B\}$ where p_{d_t} is the price of treatment d_t and B is the budget. The example of the former is the independent selection process of each ECA, in which case K-SURE is equipped with $\mathcal{D} = \{\bar{d}^T : d_t \in \{0,1\}\}$ and EXIM is equipped with $\mathcal{D} = \{\bar{d}^T : d_t \in \{0,2\}\}$; this example is investigated below.

Using this policy learning framework, we calculate the best counterfactual allocation subject to the budget not exceeding the observed one. Figure 6 reports the average counterfactual trajectories across firms under the observed and optimal treatment schedules. Relative to the observed allocations, the optimal (cost-constrained) paths yield systematically higher outcomes in every post-intervention period. Across the four post-intervention periods, the optimal allocation raises average outcomes by roughly 25–40%. Moreover, these gains are achieved with lower resource use. The total cost of the optimal allocation is 354 supports, compared with 365 under the observed allocation.[12] Overall, the findings indicate that the current targeting rules employed by the ECAs have substantial scope for improvement in terms of sequencing and timing of support. Our framework shows that policymakers can achieve better outcomes with fewer resources, highlighting the potential for more effective programs.

We next examine outcomes when insurance and loan are allocated independently, with the agencies acting separately, versus jointly, where decisions are coordinated as if by a single agency. As illustrated in Figure 7, average potential outcomes are higher under joint allocation in every period. Independent allocation uses 11,914 supports, while joint allocation requires 13,932. Despite the higher cost, efficiency is greater under joint allocation, with gains per unit cost rising from 37.52 to 51.15.[13] These results indicate that coordination among agencies reduces misallocation across treatments and leverages complementarities, allowing resources to generate higher returns per cost.

6.4.2 Prospective Policy Learning

Policymakers may want to estimate optimal targeting rules for new firms that are *not* observed in the data. To that end, we consider allocating support based on firms' observed covariates. Specifically, we consider a targeting rule $\bar{\delta}^T : \mathcal{X} \to [A]^T$ where $\mathcal{X}$ is the support of pre-treatment covariate

---

[12]One support is counted as one unit (e.g., the cost of giving insurance or a loan is one, and giving both is two).
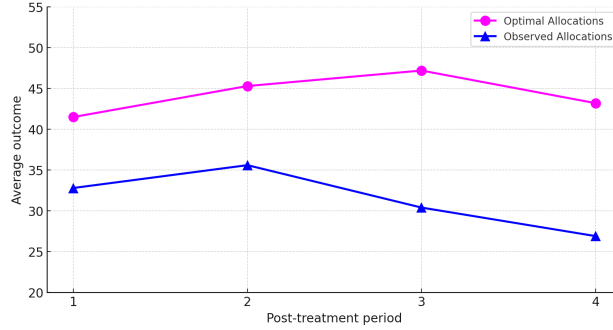
[13]In millions of KRW

Figure 6: Potential Export Values Under Optimal vs. Observed Treatment Schedules

Notes: The figure depicts the trajectory of counterfactual export values under optimal versus observed treatment schedules, with units in billions of KRW (around 720K USD). The cumulative average outcome amounts to 125.8 under the observed treatment schedules, compared with 177.3 under the optimal schedules.
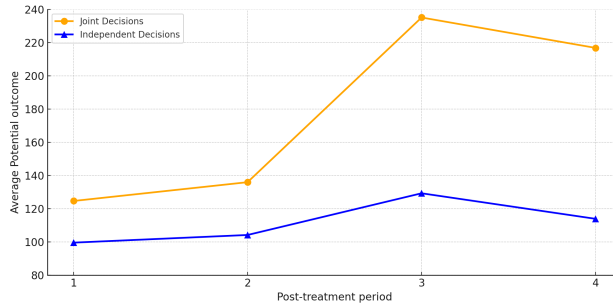

Figure 7: Potential Export Values Under Treatment Schedules Jointly vs. Independently Optimized

Notes: The figure depicts the trajectory of counterfactual export values under treatment schedules jointly versus independently optimized by the agencies, with units in billions of KRW around 720K USD). The cumulative average outcome amounts to 712.68 under joint optimization and 447.08 under independent optimization, a difference of 265.59.

38

vector X_n and

$$\bar{\delta}^{T*} \in \underset{\bar{\delta}^T \in \tilde{\mathcal{D}}}{\operatorname{argmax}} \sum_{t=T_0+q+1}^{T} \mathbb{E}\left[Y_{n,t}^{(\bar{\delta}^t(X_n))}\right], \tag{29}$$

where $\tilde{\mathcal{D}}$ is the (possibly restricted) class of targeting rules. Note that

$$\mathbb{E}\left[Y_{n,t}^{(\bar{\delta}^t(X_n))}\right] = \mathbb{E}\left[\sum_{\bar{d}^t} 1\{\bar{\delta}^t(X_n) = \bar{d}^t\} Y_{n,t}^{(\bar{d}^t)}\right] = \mathbb{E}\left[\sum_{\bar{d}^t} 1\{\bar{\delta}^t(X_n) = \bar{d}^t\} E\left[Y_{n,t}^{(\bar{d}^t)}|X_n\right]\right].$$

Under Assumption 7 (that $X_{n,k}$ has the latent factor structure),

$$\mathbb{E}\left[\mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}\right] \mid X_n\right] = \mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid X_n\right],$$

and thus,

$$\mathbb{E}\left[Y_{n,t}^{(\bar{\delta}^t(X_n))}\right] = \mathbb{E}\left[\sum_{\bar{d}^t} 1\{\bar{\delta}^t(X_n) = \bar{d}^t\} \mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}\right]\right].$$

Therefore, $\bar{\delta}^{T*}$ is identified as we identify $\mathbb{E}[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}]$ for all n, t and $\bar{d}^t$ from Theorem 2 or 5. This argument is also useful in estimating $\bar{\delta}^{T*}$ as we can take $\mathbb{E}[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}]$ as a pseudo-outcome variable for a prediction problem with predictors X_n and for subsequent policy learning.

Based on this framework, we implement a tree-based policy learning algorithm that yields interpretable decision rules. Based on fourteen firm characteristics,[14] the algorithm selects the most predictive variables and thresholds, partitioning firms into subgroups with distinct optimal treatment sequences. Each leaf in a decision tree corresponds to one recommended sequence, providing a transparent mapping from firm characteristics to intervention timing. Since considering all possible targeting rules in $\tilde{\mathcal{D}}$ is computationally and practically infeasible, we restrict $\tilde{\mathcal{D}}$ in the analysis.

First, we consider $\tilde{\mathcal{D}}$ to be the set of early treatment $(d_1, d_2, 0, 0, 0)$ and late treatment $(0, 0, 0, d_4, d_5)$ for $d_t \in \{1, 2\}$, yielding eight possible targeting rules in total. Figure 8 shows the optimal policy tree. We restrict attention to insurance or loan only, reflecting computational tractability and the budget constraints of policymakers, for whom providing multiple supports simultaneously is costly. The optimal decision tree suggests that firms with smaller wage bills are assigned to late support. Within this group, the number of workers determines the sequencing of instruments. Fewer workers lead to insurance then loan, while more workers lead to loan then insurance, reflecting payroll-driven liquidity needs. Among firms with larger wage bills, capital stock is decisive. Those with relatively low tangible assets are directed to early support, while those with stronger asset positions can defer to late support. It means firms with high labor costs but little

---

[14]Pre-treatment averages of export share, sales, employment, tangible capital, value added, TFP, total wage bill, R&D expenditure, debt-to-asset ratio, liquidity ratio, and an indicator for FDI status, as well as parent-company affiliation, firm age, and industry dummies for above-average capital intensity and wage per worker.
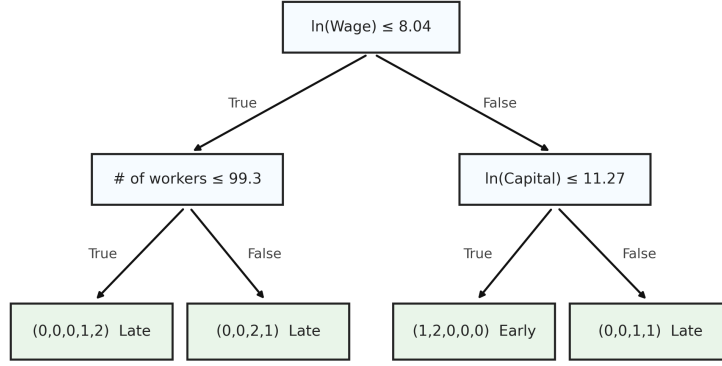
Figure 8: Decision Tree with Early and Late Treatment Options

Notes: In the optimal treatment sequences, $d_t = 1$ indicates insurance and 2 indicates loans support.
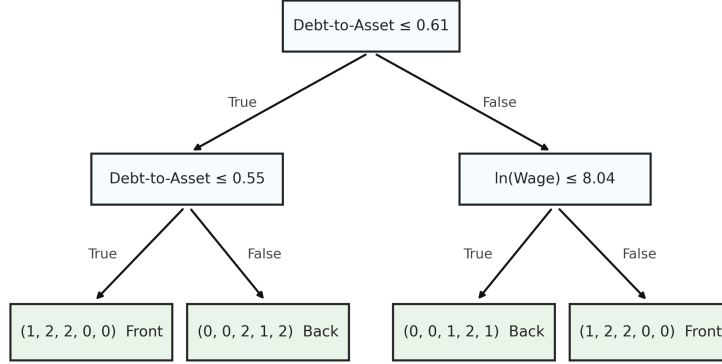

Figure 9: Decision Tree with Front-, Even- and Back-Loading Treatment Options

Notes: In the optimal treatment sequences, $d_t = 1$ indicates insurance and 2 indicates loans support.

collateralizable capital cannot easily finance wages internally and thus require earlier intervention.

Next, we consider targeting rules that not only concern timing but also spacing over time. In particular, we restrict $\tilde{\mathcal{D}}$ to be the set of front-loaded $(d_1, d_2, d_3, 0, 0)$, evenly-spaced $(d_1, 0, d_3, 0, d_5)$, and back-loaded $(0, 0, d_3, d_4, d_5)$ treatments for $d_t \in \{1, 2\}$, yielding twenty-four sequences in total. : high wages lead to front-loading with more loans, while lower wages lead to back-loading. Firms with moderate debt (around 0.55–0.61) are also routed to back-loading, consistent with temporary liquidity management. Interestingly, some low-debt firms are also assigned to front-loading, reflecting the model's prediction that these firms gain more from early expansionary financing than from delayed support.

Remark. *Note that, by equations* (16) *and* (27), $\mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} \mid \mathcal{LF}\right] = \langle \omega(\bar{d}^T), Y_n \rangle$ *for an appropriate vector* Y_n *of observed outcomes and a vector* $\omega(\bar{d}^T)$ *of parameters, which implies that our objective function has the outcome-weighted form (Zhao et al., 2012):* $\mathbb{E}[\sum_{\bar{d}^T \in [A]^T} 1\{\bar{\delta}^T(X_n) =$

$\bar{d}^T\} \langle \omega(\bar{d}^T), Y_n \rangle]$. *Therefore, analogous to* Zhao et al. (2015) *among others, we can show consistency of and bounds on the excess risk of the estimated policy by (i) using the convex surrogate version of the objective function and (ii) under the condition that $\hat{\omega}(d)$ converges to $\omega(d)$ at a certain rate. The condition (ii) can be guaranteed by our convergence rates in Theorem 4 or 6. We omit this analysis for succinctness.*

7   Conclusion

In this work, we formulate a causal framework for dynamic treatment effects under unobserved confounding using panel data. We propose a latent factor model, which admits linear time-varying and time-invariant dynamical systems as special cases. Depending on the structure placed on this factor model, we quantify the trade-off on the sample complexity and the level of adaptivity allowed in the intervention policy, for estimating counterfactual mean outcomes. The estimated counterfactual outcomes are useful in estimating the impact of particular treatment schedule relative to another, as well as the optimal rules of allocating treatment schedules. We showcase this usefulness in the context of government's financial support. We hope this work spurs further research connecting the growing fields of synthetic controls and panel data methods with dynamic treatment models studied in econometrics, and potentially sequential learning methods such as reinforcement learning studied in computer science.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californiaâs tobacco control program. *Journal of the American Statistical Association.*

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review.*

Agarwal, A., Dahleh, M., Shah, D., and Shen, D. (2021a). Causal matrix completion. *arXiv preprint arXiv:2109.15154.*

Agarwal, A., Shah, D., and Shen, D. (2020a). On principal component regression in a high-dimensional error-in-variables setting. *arXiv preprint arXiv:2010.14449.*

Agarwal, A., Shah, D., and Shen, D. (2020b). Synthetic interventions. *arXiv preprint arXiv:2006.07691.*

Agarwal, A., Shah, D., Shen, D., and Song, D. (2021b). On robustness of principal component regression. *Journal of the American Statistical Association*, 116(536):1731–1745.

Amiti, M. and Weinstein, D. E. (2011). Exports and financial shocks. *The Quarterly Journal of Economics*, 126(4):1841–1877.

Amjad, M., Mishra, V., Shah, D., and Shen, D. (2019). mrsc: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2).

Amjad, M. J., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of Machine Learning Research*, 19:1–51.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Antras, P. and Foley, C. F. (2015). Poultry in motion: a study of international trade finance practices. *Journal of Political Economy*, 123(4):853–901.

Arellano, M. and Honore, B. (2000). Panel data models: Some recent developments. *Handbook of Econometrics.*

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2020). Synthetic difference in differences.

Ashenfelter, O. C. and Card, D. (1984). Using the longitudinal structure of earnings to estimate the effect of training programs.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2020). Matrix completion, counterfactuals, and factor analysis of missing data.

Ben-Michael, E., Feller, A., and Rothstein, J. (2020). The augmented synthetic control method.

Ben-Michael, E., Feller, A., and Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):351–381.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

Cattaneo, M. D., Feng, Y., Palomba, F., and Titiunik, R. (2025). Uncertainty quantification in synthetic controls with staggered treatment adoption. *Review of Economics and Statistics*, pages 1–46.

Chamberlain, G. (1984). Panel data. In Griliches†, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition.

Chamberlain, G. (2022). Feedback in panel data models. *Journal of Econometrics*, 226(1):4–20. Annals Issue in Honor of Gary Chamberlain.

Chan, M. K. and Kwok, S. (2020). The PCDID Approach: Difference-in-Differences when Trends are Potentially Unparallel and Stochastic. Working Papers 2020-03, University of Sydney, School of Economics.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2020). Practical and robust t-test based inference for synthetic control and related methods.

Chor, D. and Manova, K. (2012). Off the cliff and back? credit conditions and international trade during the global financial crisis. *Journal of International Economics*, 87(1):117–133.

Criscuolo, C., Martin, R., Overman, H. G., and Van Reenen, J. (2019). Some causal effects of an industrial policy. *American Economic Review*, 109(1):48–85.

De Chaisemartin, C. and d'Haultfoeuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, pages 1–45.

de Chaisemartin, C. and D'Haultfœuille, X. (2025). Treatment-effect estimation in complex designs under a parallel-trends assumption. *arXiv preprint arXiv:2508.07808*.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review*, 110(9):2964–2996.

Doudchenko, N. and Imbens, G. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *NBER Working Paper No. 22791*.

Felbermayr, G. J. and Yalcin, E. (2013). Export credit guarantees and export performance: An empirical analysis for germany. *The World Economy*, 36(8):967–999.

Fernández-Val, I., Freeman, H., and Weidner, M. (2020). Low-rank approximations of nonseparable panel models.

Geweke, J. (1976). *The Dynamic Factor Analysis of Economic Time Series Models*. Workshop series / Social Systems Research Institute, University of Wisconsin. University of Wisconsin.

Han, S. (2021). Identification in nonparametric models for dynamic treatment effects. *Journal of Econometrics*, 225(2):132–147.

Han, S. (2024). Optimal dynamic treatment regimes and partial welfare ordering. *Journal of the American Statistical Association*, 119(547):2000–2010.

Hardt, M., Ma, T., and Recht, B. (2016). Gradient descent learns linear dynamical systems. *arXiv preprint arXiv:1609.05191*.

Hernán, M. and Robins, J. (2020). Causal inference: What if. *Boca Raton: Chapman & Hill/CRC*.

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of hong kong with mainland china. *Journal of Applied Econometrics*, 27(5):705–740.

Hsu, Y.-C. and Shen, S. (2024). Dynamic regression discontinuity under treatment effect heterogeneity. *Quantitative Economics*, 15(4):1035–1064.

Imbens, G., Kallus, N., and Mao, X. (2021). Controlling for unmeasured confounding in panel data using minimal bridge functions: From two-way fixed effects to factor models. *arXiv preprint arXiv:2108.03849*.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

Lewis, G. and Syrgkanis, V. (2020). Double/debiased machine learning for dynamic treatment effects via g-estimation. *arXiv preprint arXiv:2002.07285*.

Li, K. T. (2018). Inference for factor model based average treatment effects. *Available at SSRN 3112775*.

Li, K. T. and Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65 – 75.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall information and system sciences series. Prentice Hall PTR.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

Niepmann, F. and Schmidt-Eisenlohr, T. (2017). International trade, risk and the role of banks. *Journal of International Economics*, 107:111–126.

Paravisini, D., Rappoport, V., Schnabl, P., and Wolfenzon, D. (2015). Dissecting the effect of credit supply on trade: Evidence from matched credit-export data. *The Review of Economic Studies*, 82(1):333–359.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Powell, D. (2022). Synthetic control estimation beyond comparative case studies: does the minimum wage reduce employment? *Journal of Business & Economic Statistics*, 40(3):1302–1314.

Robins, J. M. (2004). *Optimal Structural Nested Models for Optimal Sequential Decisions*, pages 189–326. Springer New York, New York, NY.

Rotemberg, M. (2019). Equilibrium effects of firm subsidies. *American Economic Review*, 109(10):3475–3513.

Schmidt-Eisenlohr, T. (2013). Towards a theory of trade finance. *Journal of International Economics*, 91(1):96–112.

Shahn, Z., Dukes, O., Richardson, D., Tchetgen Tchetgen, E., and Robins, J. (2022). Structural nested mean models under parallel trends assumptions. *arXiv preprint arXiv:2204.10291*.

Shaikh, A. M. and Toulis, P. (2021). Randomization tests in observational studies with staggered adoption of treatment. *Journal of the American Statistical Association*, 116(536):1835–1848.

Sojitra, R. B. and Syrgkanis, V. (2024). Dynamic local average treatment effects. *arXiv preprint arXiv:2405.01463*.

Stock, J. H. and Watson, M. W. (2011). Dynamic Factor Models. In *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Vansteelandt, S. and Joffe, M. (2014). Structural Nested Models and G-estimation: The Partially Realized Promise. *Statistical Science*, 29(4):707 – 731.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168.

A   Connection to SNMM and MSM: Proof of Proposition 2

Verifying Assumption 4 holds. In what follows, all the conditional expectations are also conditioned on the latent factors $\mathcal{LF}$. However, for shorthand notation, we omit that conditioning. Note that:

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)} - Y_{n,t}^{(\bar{0}_t)}\right] = \sum_{\ell=1}^{t} \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)}\right] \tag{30}$$

We now prove that:

$$Q_{n,t} := \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)}\right] = \left\langle \psi_n^{t,\ell}, w_{d^\ell} - w_{0^\ell}\right\rangle$$

We establish this via a nested mean argument. Note

$$
\begin{aligned}
Q_{n,t} &= \mathbb{E}\left[\mathbb{E}\left[Y_{n,t}^{(\bar{d}_n^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}_n^{\ell-1},\underline{0}^\ell)} \mid S_n^0\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{n,t}^{(\bar{d}_n^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}_n^{\ell-1},\underline{0}^\ell)} \mid S_n^0, D_n^1 = d^1\right]\right]
\end{aligned}
\tag{31}
$$

where in (31), we have used (11). Now as our inductive step, suppose that we have shown:

$$Q_{n,t} = \mathbb{E}\left[\mathbb{E}\left[\dots \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)} \mid \bar{S}_n^{q-1}, \bar{D}_n^q = \bar{d}^q\right] \dots \mid S_n^0, D_n^1 = d^1\right]\right]$$

Then,

$$
\begin{aligned}
Q_{n,t} &= \mathbb{E}\left[\mathbb{E}\left[\dots \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)} \mid \bar{S}_n^q, \bar{D}_n^q = \bar{d}^q\right] \mid \bar{S}_n^{q-1}, \bar{D}_n^q = \bar{d}^q\right] \dots \mid S_n^0, D_n^1 = d^1\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\dots \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)} \mid \bar{S}_n^q, \bar{D}_n^{q+1} = \bar{d}^{q+1}\right] \mid \bar{S}_n^{q-1}, \bar{D}_n^q = \bar{d}^q\right] \dots \mid S_n^0, D_n^1 = d^1\right]\right],
\end{aligned}
\tag{32}
$$

where in (32), we have again used (11). This concludes the inductive proof. Thus, we have

$$
\begin{aligned}
Q_{n,t} &= \mathbb{E}\left[\mathbb{E}\left[\dots \mathbb{E}\left[Y_{n,t}^{(\bar{d}^\ell,\underline{0}^{\ell+1})} - Y_{n,t}^{(\bar{d}^{\ell-1},\underline{0}^\ell)} \mid \bar{S}_n^{\ell-1}, \bar{D}_n^\ell = \bar{d}^\ell\right] \dots \mid S_n^0, D_n^1 = d^1\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\dots \gamma_{n,t,\ell}(d^\ell) \dots \mid S_n^0, D_n^1 = d^1\right]\right] = \left\langle \psi_n^{t,\ell}, w_{d^\ell} - w_{0^\ell}\right\rangle
\end{aligned}
\tag{33}
$$

where in (33), we have used (12) and the fact that $\left\langle \psi_n^{t,\ell}, w_{d^\ell} - w_{0^\ell}\right\rangle$ is independent of S_n^ℓ and $\bar{D}_n^{\ell-1}$. Re-arranging (30) and (33), we have:

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^t)}\right] = \mathbb{E}\left[Y_{n,t}^{(\bar{0}^t)}\right] + \sum_{\ell=1}^{t} \gamma_{n,t,\ell}(d^\ell). \tag{34}$$

Combining (34) and (13) implies Assumption 4 holds.

    Verifying Assumption 6 holds. Assumption 6 is immediately implied by (12) and a simple

application of the tower law of expectations. In particular, we integrate $\bar{S}_{n,t-1}$ out of both sides of (12).

B   Proof of Theorem 1

By Assumption 2,

$$
\begin{aligned}
\mathbb{E}\left[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}\right] &= \mathbb{E}\left[\langle v_{n,T},\ w_{\bar{d}^T}\rangle + \varepsilon_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF}\right] \\
&= \mathbb{E}\left[\langle v_{n,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}\right] \\
&= \langle v_{n,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF} \\
&= \langle v_{n,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T},
\end{aligned}
$$

(35)

(36)

where (35) and (36) follow since $v_{n,T}, w_{\bar{d}^T}$ are deterministic conditional on the latent factors.

Then by Assumption 3,

$$
\langle v_{n,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T} = \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \langle v_{j,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T}
$$

Then by appealing to the conditional mean exogeneity of $\varepsilon_{j,T}^{(\bar{d}^T)}$ in Definition 2, we have

$$
\begin{aligned}
&\sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \langle v_{j,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T} \\
&= \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \langle v_{j,T},\ w_{\bar{d}^T}\rangle \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T} + \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \mathbb{E}[\varepsilon_{j,T}^{(\bar{d}^T)} \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T}] \\
&= \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \mathbb{E}[Y_{j,T}^{(\bar{d}^T)} \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T}], \\
&= \sum_{j \in \mathcal{I}^{\bar{d}^T}} \beta_j^{n, \mathcal{I}^{\bar{d}^T}} \mathbb{E}[Y_{j,T} \mid \mathcal{LF}, \mathcal{I}^{\bar{d}^T}],
\end{aligned}
$$

(37)

(38)

where (37) follows from Assumption 2; (38) follows from Assumption 1 and Definition 2.

This completes the proof.

C Proofs and Remarks for Time-Varying Linear Dynamical System

C.1 Proof of Proposition 1

Recall $z_{n,t}$ is the latent state of unit n if it undergoes action sequence $\bar{d}^t$. By a simple recursion we have

$$z_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t-1} \left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \boldsymbol{C}_{n,\ell} \, w_{d_\ell} + \boldsymbol{C}_{n,t} \, w_{d_t} + \sum_{\ell=1}^{t-1} \left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \eta_{n,\ell} + \eta_{n,t}$$

Hence,

$$
\begin{aligned}
Y_{n,t}^{(\bar{d}^t)} & \\
&= \left\langle \theta_{n,t}, \ \sum_{\ell=1}^{t-1} \left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \boldsymbol{C}_{n,\ell} \, w_{d_\ell} + \boldsymbol{C}_{n,t} \, w_{d_t} + \sum_{\ell=1}^{t-1} \left(\prod_{k=\ell+1}^{t} \boldsymbol{B}_{n,k} \right) \eta_{n,\ell} + \eta_{n,t} \right\rangle + \langle \tilde{\theta}_{n,t}, w_{d_t} \rangle + \tilde{\eta}_{n,t} \\
&= \sum_{\ell=1}^{t} \left(\left\langle \psi_n^{t,\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t,\ell} \right),
\end{aligned}
$$

where in the last line we use the definitions of $\psi_n^{t,\ell}$ and $\varepsilon_{n,t,\ell}$ in the proposition statement. This completes the proof.

C.2 Proof of Theorem 2

For simplicity, we omit the conditioning on $\mathcal{LF}$ in all derivations; all expectations are conditioned on $\mathcal{LF}$.

1. **Verifying** (identification). First, we verify (identification) holds, which allows us to express the counterfactual outcomes, in terms of the blips and the baseline. For all $n \in [N]$, using Assumption 4 we have:

$$
\begin{aligned}
\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] &= \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} - Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}] + \mathbb{E}[Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \left\langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \right\rangle + \varepsilon_{n,T}^{(\bar{d}^T)} - \varepsilon_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF} \right] + \mathbb{E}\left[\sum_{t=1}^{T} \left\langle \psi_n^{T,t}, w_{0_t} \right\rangle + \varepsilon_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF} \right] \\
&= \sum_{t=1}^{T} \gamma_{n,T,t}(d_t) \mid \mathcal{LF} + b_{n,T} \mid \mathcal{LF}
\end{aligned}
$$

2. **Verifying** (observed control) **&** (synthetic control):

We first show (observed control) holds. For $j \in \mathcal{I}_T^0$:

$$b_{j,T} \mid \mathcal{LF} = \sum_{t=1}^{T} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF} = \mathbb{E}\left[\sum_{t=1}^{T} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle + \varepsilon_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF} \right] \tag{39}$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle + \varepsilon_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF}, \mathcal{I}_T^0 \right] \tag{40}$$

$$= \mathbb{E}\left[Y_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF}, j \in \mathcal{I}_T^0 \right] \tag{41}$$

$$= \mathbb{E}\left[Y_{j,T} \mid \mathcal{LF}, j \in \mathcal{I}_T^0 \right], \tag{42}$$

where (39) and (41) follow from Assumption 4; (40) follows from the fact that $\left\langle \psi_j^{T,t}, w_{0_t} \right\rangle$ is deterministic conditional on $\mathcal{LF}$, and that $\mathbb{E}[\varepsilon_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF}, \mathcal{I}_T^0] = \mathbb{E}[\varepsilon_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF}]$ as seen in the definition of $\mathcal{I}_T^0$; (42) follows from Assumption 1.

Next we show (synthetic control) holds. For $i \notin \mathcal{I}_T^0$:

$$b_{i,T} \mid \mathcal{LF} = \sum_{t=1}^{T} \left\langle \psi_i^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF}$$

$$= \sum_{t=1}^{T} \left\langle \psi_i^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF}, \mathcal{I}_T^0 \tag{43}$$

$$= \sum_{t=1}^{T} \sum_{j \in \mathcal{I}_T^0} \beta_j^{i, \mathcal{I}_T^0} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF}, \mathcal{I}_T^0 \tag{44}$$

$$= \sum_{j \in \mathcal{I}_T^0} \beta_j^{i, \mathcal{I}_T^0} b_{j,T} \mid \mathcal{LF}, \mathcal{I}_T^0$$

where (43) follows from the fact that $\left\langle \psi_i^{T,t}, w_{0_t} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (44) follows from Assumption 5;

3. Verifying ("observed" blip at time T) **&** (synthetic blip at time T):

We first show ("observed" blip at time T) holds. For all $d \in [A]$ and $j \in \mathcal{I}_T^d$:

$$\gamma_{j,T,T}(d) \mid \mathcal{LF} = \left\langle \psi_j^{T,T}, w_d - w_{0_T} \right\rangle \mid \mathcal{LF}$$

$$= \mathbb{E}\left[\left\langle \psi_j^{T,T}, w_d - w_{0_T} \right\rangle + \varepsilon_{j,T}^{(0^{T-1},d)} \pm \sum_{t=1}^{T-1} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF} \right] \tag{45}$$

$$= \mathbb{E}\left[\left\langle \psi_j^{T,T}, w_d \right\rangle + \varepsilon_{j,T}^{(0^{T-1},d)} + \sum_{t=1}^{T-1} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF} \right] - \sum_{t=1}^{T} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF} \tag{46}$$

$$= \mathbb{E}\left[\left\langle \psi_j^{T,T}, w_d \right\rangle + \varepsilon_{j,T}^{(0^{T-1},d)} + \sum_{t=1}^{T-1} \left\langle \psi_j^{T,t}, w_{0_t} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_T^d \right] - b_{j,T} \mid \mathcal{LF}$$

$$= \mathbb{E}[Y_{j,T}^{(\bar{D}_j^T)} \mid \mathcal{LF},\ j \in \mathcal{I}_T^d] - b_{j,T} \mid \mathcal{LF} \tag{47}$$

$$= \mathbb{E}[Y_{j,T} \mid \mathcal{LF},\ j \in \mathcal{I}_T^d] - b_{j,T} \mid \mathcal{LF} \tag{48}$$

where (45), (46) follow from Assumption 4; (47) follows from the definition of $\mathcal{I}_t^d$ and Assumption 4; (48) follows from Assumption 1.

Next we show (synthetic blip at time T) holds. For $i \notin \mathcal{I}_T^d$

$$\gamma_{i,T,T}(d) \mid \mathcal{LF} = \left\langle \psi_i^{T,T}, w_d - w_{0_T} \right\rangle \mid \mathcal{LF} = \left\langle \psi_i^{T,T}, w_d - w_{0_T} \right\rangle \mid \mathcal{LF}, \mathcal{I}_T^d \tag{49}$$

$$= \sum_{j \in \mathcal{I}_T^d} \beta_j^{i,\mathcal{I}_T^d} \left\langle \psi_j^{T,T}, w_d - w_{0_T} \right\rangle \mid \mathcal{LF}, \mathcal{I}_T^d \tag{50}$$

$$= \sum_{j \in \mathcal{I}_T^d} \beta_j^{i,\mathcal{I}_T^d} \gamma_{j,T,T}(d) \mid \mathcal{LF}, \mathcal{I}_T^d$$

where (49) follows from the fact that $\left\langle \psi_i^{T,T}, w_d - w_{0_T} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (50) follows from Assumption 5.

4. **Verifying** ("observed" blip at time t) **&** (synthetic blip at time t):

We first show ("observed" blip at time t) holds. For all $d \in [A]$, $t < T$, $j \in \mathcal{I}_t^d$:

$$\mathbb{E}\left[Y_{j,T} - Y_{j,T}^{(\bar{0}_T)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] = \mathbb{E}\left[Y_{j,T}^{(\bar{D}_j^T)} - Y_{j,T}^{(\bar{0}_T)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] \tag{51}$$

$$= \mathbb{E}\left[Y_{j,T}^{(\bar{D}_j^T)} - Y_{j,T}^{(\bar{D}^{t-1}, 0^t)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] \tag{52}$$

$$= \sum_{\ell=t}^{T} \mathbb{E}\left[Y_{j,T}^{(\bar{D}_j^\ell, 0^{\ell+1})} - Y_{j,T}^{(\bar{D}_j^{\ell-1}, 0^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] \tag{53}$$

where (51) follows from Assumption 1; (52) uses that for $j \in \mathcal{I}_t^d$, $\bar{D}_n^t = (0_1, \ldots, 0_{t-1}, d_t)$, and

51

Assumption 1. Then,

$$
\sum_{\ell=t}^{T} \mathbb{E}\left[Y_{j,T}^{(\bar{D}_j^\ell, \underline{0}^{\ell+1})} - Y_{j,T}^{(\bar{D}_j^{\ell-1}, \underline{0}^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
= \sum_{\ell=t}^{T} \mathbb{E}\left[\left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle + \varepsilon_{j,T}^{(\bar{D}_j^\ell, \underline{0}^{\ell+1})} - \varepsilon_{j,T}^{(\bar{D}_j^{\ell-1}, \underline{0}^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] \tag{54}
$$

$$
= \mathbb{E}\left[\left\langle \psi_j^{T,t}, w_{D_{j,t}} - w_{0_t} \right\rangle + \varepsilon_{j,T}^{(\bar{D}_j^\ell, \underline{0}^{\ell+1})} - \varepsilon_{j,T}^{(\bar{D}_j^{\ell-1}, \underline{0}^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] + \sum_{\ell=t+1}^{T} \mathbb{E}\left[\left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
= \mathbb{E}\left[\left\langle \psi_j^{T,t}, w_d - w_{0_t} \right\rangle + \varepsilon_{j,T}^{(\bar{D}_j^\ell, \underline{0}^{\ell+1})} - \varepsilon_{j,T}^{(\bar{D}_j^{\ell-1}, \underline{0}^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] + \sum_{\ell=t+1}^{T} \mathbb{E}\left[\left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
\tag{55}
$$

$$
= \left\langle \psi_j^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF} + \sum_{\ell=t+1}^{T} \mathbb{E}\left[\left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
+ \mathbb{E}\left[\varepsilon_{j,T}^{(\bar{D}_j^\ell, \underline{0}^{\ell+1})} - \varepsilon_{j,T}^{(\bar{D}_j^{\ell-1}, \underline{0}^\ell)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
= \left\langle \psi_j^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF} + \sum_{\ell=t+1}^{T} \mathbb{E}\left[\left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right]
$$

$$
+ \mathbb{E}\left[\mathbb{E}\left[\varepsilon_{j,T}^{(\bar{\delta}^\ell, \underline{0}^{\ell+1})} - \varepsilon_{j,T}^{(\bar{\delta}^{\ell-1}, \underline{0}^\ell)} \mid \bar{D}_j^\ell = (\bar{\delta}^\ell, \mathcal{LF}, j \in \mathcal{I}_t^d \right] \right]
$$

$$
= \left\langle \psi_j^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF} + \sum_{\ell=t+1}^{T} \left\langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0_\ell} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \tag{56}
$$

$$
= \gamma_{j,T,t}(d) \mid \mathcal{LF} + \sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,\ell}) \mid \mathcal{LF} \tag{57}
$$

where (54) follows from Assumption 4; (55) follows from the definition of $\mathcal{I}_t^d$, i.e., for $j \in \mathcal{I}_t^d$, $\bar{D}_j^t = (\bar{0}^{t-1}, d)$ and that $\forall \, \delta \in [A], \ell \in [T]$, $\mathbb{E}[\varepsilon_{j,T,\ell}^{(\delta)} \mid \mathcal{LF}, \bar{D}_j^t] = \mathbb{E}[\varepsilon_{j,T,\ell}^{(\delta)} \mid \mathcal{LF}]$; (56) follows from Assumption 6, where we require the last term on the l.h.s. of the equality to be zero only for $j \in \mathcal{I}_t^d$.

Re-arranging (57) we have that,

$$
\gamma_{j,T,t}(d) \mid \mathcal{LF} = \mathbb{E}\left[Y_{j,t} - Y_{j,T}^{(\bar{0}_T)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] - \sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,\ell}) \mid \mathcal{LF}
$$

$$
= \mathbb{E}\left[Y_{j,t} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] - \mathbb{E}\left[Y_{j,T}^{(\bar{0}_T)} \mid \mathcal{LF} \right] - \sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,\ell}) \mid \mathcal{LF} \tag{58}
$$

$$
= \mathbb{E}\left[Y_{j,t} \mid \mathcal{LF}, j \in \mathcal{I}_t^d \right] - b_{j,T} \mid \mathcal{LF} - \sum_{\ell=t+1}^{T} \gamma_{j,T,\ell}(D_{j,\ell}) \mid \mathcal{LF} \tag{59}
$$

where (58) follows from the definition of $\mathcal{I}_t^d$; (59) follows from Assumption 4.

Next we show (synthetic blip at time t) holds. For all $d \in [A]$, $t < T$, $i \notin \mathcal{I}_t^d$:

$$\gamma_{i,T,t}(d) \mid \mathcal{LF} = \left\langle \psi_i^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF} \tag{60}$$

$$= \left\langle \psi_i^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF}, \mathcal{I}_t^d \tag{61}$$

$$= \sum_{j \in \mathcal{I}_t^d} \beta_j^{i, \mathcal{I}_t^d} \left\langle \psi_j^{T,t}, w_d - w_{0_t} \right\rangle \mid \mathcal{LF}, \mathcal{I}_T^d \tag{62}$$

$$= \sum_{j \in \mathcal{I}_t^d} \beta_j^{i, \mathcal{I}_t^d} \gamma_{j,T,t}(d) \mid \mathcal{LF}, \mathcal{I}_t^d$$

where (61) follows from the the fact that $\left\langle \psi_i^{T,t}, w_d - w_{0_t} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (62) follows from Assumption 5;

C.3 General Remarks on LTV Setting

C.3.1 Covariate Design

We point out that Assumption 7 in its base form does not allow for time-varying covariates. Specifically, it assumes access to p covariates for each unit, each with respect to their unit factor at the terminal time, i.e., $v_{n,T}$. Furthermore, as seen in Theorem 3 we require $p \to \infty$ and fixed T for consistency which seems highly unlikely in practice.

However, notice in Theorem 4 we are able to send $T \to \infty$ as well. As such, inclusion of time varying covariates allow for $p \to \infty$ be justified. To that end, here would be a general construction of such covariates.

Assumption 22. *For each unit $n \in [N]$, we have covariates $X_n = (X_{n,1}^\top, X_{n,2}^\top, \ldots, X_{n,T}^\top)^\top \in \mathbb{R}^{pT}$. Specifically, for any $t \in [T]$ we have $X_{n,t} \in \mathbb{R}^p$ where*

$$X_{n,t,k} = \langle v_{n,t}, \rho_k^t \rangle + \varepsilon_{n,t,k}$$

for any $k \in [p]$ where $v_{n,t}$ is the unit latent factor defined in Assumptions 2 and ε_{ntk} is mean-zero noise. Specifically, we collect p features at each time step $t \in [T]$. Denote $X \in \mathbb{R}^{pT \times N} = [X_1, \ldots, X_N]$.

Notice that with the added flexibility in the above formulation we can use observed values as covariates as well.

C.3.2 Row-Space Inclusion

Here is a sufficient condition on which we have row-space inclusion.

Lemma 1. *Let the setup of Assumption 12 hold and denote*

$$V_{\mathcal{I}_t^d} = ([v_{j,T}]_{j \in \mathcal{I}_t^d})^\top \in \mathbb{R}^{|\mathcal{I}_t^d| \times mT} \quad and \quad b = ([\langle v_{j,T}, w_{(\bar{D}_j^T)} \rangle]_{j \in \mathcal{I}_t^d})^\top \in \mathbb{R}^{|\mathcal{I}_t^d|}.$$

If $span(\{\rho_i\}_{i\in[p]}) \cap \{y \in \mathbb{R}^{mT} : V_{\mathcal{I}_t^d}y = b\}$ is non-empty, then Assumption 12 holds for ρ_i as defined in Assumption 7.

Proof. We require weights $\xi_i^{(d,t)}$ such that following holds for any $j \in \mathcal{I}_t^d$

$$\mathbb{E}[Y_{j,T}|\mathcal{LF}] = \sum_{i=1}^{p} \xi_i^{(d,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^d})_{ij}|\mathcal{LF}].$$

Note that $\mathbb{E}[Y_{j,T}|\mathcal{LF}] = \langle v_{j,T}, w_{(\bar{D}_j^T)} \rangle$ for any $j \in \mathcal{I}_t^d$. Furthermore, $\mathbb{E}[(X_{\mathcal{I}_t^d})_{ij}|\mathcal{LF}] = \langle v_{j,T}, \rho_i \rangle$ under the formulation presented in Assumption 7. As such, if we define $B_{ij} = \langle v_{i,T}, \rho_j \rangle$ for all $i \in \mathcal{I}_t^d$ and $j \in [p]$ then our problem is equivalent to there being a solution $\xi^{(d,t)} \in \mathbb{R}^p$ to the linear system $B\xi^{(d,t)} = b$. To conclude notice that $B = V_{\mathcal{I}_T^d}[\rho_1, \dots, \rho_p]$. □

In general, the point is if the covariates defined by $\{\rho_i\}_{i\in[p]}$ are sufficiently expressive then row-space inclusion holds. Specifically, we seek to maximize the dimension of their span, as would occur if they were linearly independent. The next result is a consequence of Assumption 12 and will be essential in establishing consistency.

Lemma 2. *Let Assumption 12 hold. Then for any $d \in [A]$ and $t \in [T]$ there exists $\alpha^{(d,t)} \in \mathbb{R}^p$ such that*

$$w_{(0_1,\dots,0_{t-1},d,0_{t+1},\dots,0_T)} = \sum_{i=1}^{p} \alpha_i^{(d,t)} \cdot \rho_i.$$

That is $w_{(0_1,\dots,0_{t-1},d,0_{t+1},\dots,0_T)} \in span(\{\rho_i\}_{i\in[p]})$.

Proof. By Assumption 12, for any unit $j \in \mathcal{I}_t^d$ and any collection of treatment sequences $(D_{j,t+1}, \dots, D_{j,T})_{j\in\mathcal{I}_t^d}$, there exists a solution to the following system:

$$V_{\mathcal{I}_t^d}[\rho_1, \dots, \rho_p]\xi = [\langle v_{j,T}, w_{(0_1,\dots,0_{t-1},d,D_{j,t+1},\dots,D_{j,T})} \rangle]_{j\in\mathcal{I}_t^d}^{\top}.$$

As such, there exists a solution for the following set of sequences $(0_{t+1}, \dots, 0_T)_{j\in\mathcal{I}_t^d}$ as well. In that case the system can be written as

$$V_{\mathcal{I}_t^d}[\rho_1, \dots, \rho_p]\xi = V_{\mathcal{I}_t^d}w_{(0_1,\dots,0_{t-1},d,0_{t+1},\dots,0_T)},$$

which we know to have a solution ξ. This implies $[\rho_1, \dots, \rho_p]\xi - w_{(0_1,\dots,0_{t-1},d,0_{t+1},\dots,0_T)} \in \ker(V_{\mathcal{I}_t^d})$.

By assumption we know $rank(V_{\mathcal{I}_t^d}) = mT$ or equivalently the matrix has full column rank. See the discussion under Assumption 3 in Section 3 justifying such an assumption for settings when $\mathcal{I}_t^d$ is sufficiently large. This implies that $\ker(V_{\mathcal{I}_t^d}) = \{0\}$, by Rank-Nullity Theorem.

Combining the above results we know $[\rho_1, \dots, \rho_p]\xi - w_{(0_1,\dots,0_{t-1},d,0_{t+1},\dots,0_T)} = 0$. Since, this is true for any $d \in [A]$ and $t \in [T]$ we have the desired result upon rearranging. □

Lemma 3. *Let Assumption 12 hold. Then for all $d \in [A]$ and $t \in [T]$ there exist $\xi_i^{(d,t)'} \in \mathbb{R}^p$ such that*

$$\mathbb{E}[Y_{j,T}^{(\overline{D_{j,-\ell} \cup 0_\ell})}|\mathcal{LF}] = \sum_{i=1}^{p} \xi_i^{(d,t)'} \cdot \mathbb{E}[(X_{\mathcal{I}_t^d})_{ij}|\mathcal{LF}],$$

where $\overline{D_{j,-\ell} \cup 0_\ell} = (0_1, \ldots, 0_{t-1}, d, D_{j,t+1}, \ldots, D_{j,\ell-1}, 0_\ell, D_{j,\ell+1}, \ldots, D_{j,T})$ *for any $\ell > t$.*

Proof. This holds as an immediate consequence of Assumption 12 where we consider $D_{j,\ell} = 0_\ell$ instead. $\square$

C.3.3 Linear Factor Model Assumption

Assumption 13 is not restrictive. Recalling the Linear Dynamical System setting from Proposition 1, we present a few sufficient conditions for the above to hold true.

1. **Hard Memory Cutoff**

$$\exists q \in \mathbb{N}, \quad \forall T, \quad \prod_{j=T-q}^{T} \boldsymbol{B}_{n,j} = 0. \tag{63}$$

2. **Exponential Forgetting (Spectral Decay Condition)**

$$\exists C > 0, \rho \in (0,1), \text{ such that for all } T, t, \quad \left\| \prod_{j=t}^{T} \boldsymbol{B}_{n,j} \right\|_2 \leq C\rho^{T-t}. \tag{64}$$

3. **Soft Memory Cutoff (Higher-Order Markov Property)**

$$\mathbb{P}(z_{n,T} \mid z_{n,T-1}, z_{n,T-2}, \ldots, z_{n,0}) = \mathbb{P}(z_{n,T} \mid z_{n,T-1}, \ldots, z_{n,T-q}). \tag{65}$$

Clearly, the first condition is the *strongest* and implies the other two. In general, this shows that our assumption of fixed memory is a reasonable one proving the effectiveness of our methodology within the dynamic treatment regime from a statistical perspective.

C.4 Proof of Theorem 3

1. Verifying Baseline Consistency: We first check the units not in control the entire time.

Donor Set Baseline Consistency: Consider unit $n \in \mathcal{I}_T^0$. Denote $X_{\mathcal{I}_T^0 \setminus n} = X_{:, \mathcal{I}_T^0 \setminus n} \in \mathbb{R}^{p \times |\mathcal{I}_T^0 \setminus n|}$. We know the baseline outcome admits the representation

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} = \left\langle \hat{\phi}^{n,\mathcal{I}_T^0}, Y_{\mathcal{I}_T^0 \setminus n} \right\rangle - \left\langle \phi^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n} \mid \mathcal{LF}] \right\rangle,$$

where $\hat{\phi}^{n,\mathcal{I}_T^0}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_T^0 \setminus n}$-approximation $X_{\mathcal{I}_T^0 \setminus n}$ with $k_{\mathcal{I}_T^0 \setminus n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_T^0 \setminus n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_T^0 \setminus n}$.

Lemma 4. *We claim the following*

$$\left\langle \phi^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle,$$

where $\tilde{\phi}^{n,\mathcal{I}_T^0} = VV^\top \phi^{n,\mathcal{I}_T^0}$ where $V \in \mathbb{R}^{|\mathcal{I}_T^0 \setminus n| \times k_{\mathcal{I}_T^0 \setminus n}}$ denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_T^0 \setminus n}]$ and $k_{\mathcal{I}_T^0 \setminus n} = rank(\mathbb{E}[X_{\mathcal{I}_T^0 \setminus n}])$, i.e,

$$\mathbb{E}[X_{\mathcal{I}_T^0 \setminus n}] = \sum_{l=1}^{k_{\mathcal{I}_T^0 \setminus n}} \sigma_l u_l v_l^\top = U\Sigma V^T,$$

where $u_\ell \in \mathbb{R}^p$ and $v_\ell \in \mathbb{R}^{|\mathcal{I}_T^0 \setminus n|}$

Proof. By Assumption 12 there exists $\xi^{(0,T)}$ such that for any $j \in \mathcal{I}_T^0 \setminus n$

$$\mathbb{E}[Y_{j,T}|\mathcal{LF}, j \in \mathcal{I}_T^0 \setminus n] = \sum_{i=1}^p \xi_i^{(0,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^0 \setminus n})_{ij}|\mathcal{LF}, j \in \mathcal{I}_T^0 \setminus n].$$

As such, the row-space of $\mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}]^\top \in \mathbb{R}^{1 \times |\mathcal{I}_T^0 \setminus n|}$ is included in row space of $\mathbb{E}[X_{\mathcal{I}_T^0 \setminus n}] \in \mathbb{R}^{p \times |\mathcal{I}_T^0 \setminus n|}$.[15] This yields

$$\mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] = VV^T \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}],$$

which gives us

$$\left\langle \tilde{\phi}^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle = \left\langle VV^\top \phi^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle = \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}]^\top VV^\top \cdot \phi^{n,\mathcal{I}_T^0} = \left\langle \phi^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle$$

proving the desired result. $\square$

Using Lemma 4, we can now lift the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency for $n \in \mathcal{I}_T^0$

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} = \left\langle \hat{\phi}^{n,\mathcal{I}_T^0}, Y_{\mathcal{I}_T^0 \setminus n} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_T^0}, \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \right\rangle$$

$$= O_p \left(\sqrt{\log(p|\mathcal{I}_T^0|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max \left\{ \frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_T^0| - 1}} \right\} \right] \right), \quad (66)$$

where we set $T_1 = 1$, $\tilde{w}^{(i,d)} = \tilde{\phi}^{n,\mathcal{I}_T^0}$, $\hat{w}^{(i,d)} = \hat{\phi}^{n,\mathcal{I}_T^0}$, $Y_{t,\mathcal{I}^{(d)}} = Y_{\mathcal{I}_T^0 \setminus n}$, $\mathbb{E}[Y_{t,\mathcal{I}^{(d)}}] = \mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n} \mid \mathcal{LF}]$, and $\mathcal{P}_{V_{\mathrm{pre}}} = VV^\top$. Furthermore, in the final rate we set $T_0 = p$, $N_d = |\mathcal{I}_T^0 \setminus n|$, and $r_{\mathrm{pre}} = k_{\mathcal{I}_T^0 \setminus n}$. To conclude, we used that $|\mathcal{I}_T^0 \setminus n| = |\mathcal{I}_T^0| - 1$ and $k_{\mathcal{I}_T^0 \setminus n} \leq k$ where k is the uniform upper bound on the rank on all possible expected covariate matrices, i.e., $k = \max_{d \in [A], t \in [T]} \mathrm{rank}(\mathbb{E}[X_{\mathcal{I}_t^d}])$.

---

[15] This is equivalent to the column space of the right singular vectors of $\mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}]^\top$ being included in the column space of V, or equivalently $\mathbb{E}[Y_{\mathcal{I}_T^0 \setminus n}] \in \mathrm{span}(\{v_1, \ldots, v_{k_{\mathcal{I}_T^0 \setminus n}}\})$.

Non-Donor Set Baseline Consistency: Consider unit $n \notin \mathcal{I}_T^0$. Denote $X_{\mathcal{I}_T^0} = X_{:, \mathcal{I}_T^0} \in \mathbb{R}^{p \times |\mathcal{I}_T^0|}$. We know the baseline outcome admits the representation

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}_T^0}, \hat{b}_{\mathcal{I}_T^0} \right\rangle - \left\langle \beta^{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \right\rangle,$$

where $\hat{\beta}^{n,\mathcal{I}_T^0}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_T^0}$-approximation of $X_{\mathcal{I}_T^0}$ with $k_{\mathcal{I}_T^0} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_T^0}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_T^0}$.

Lemma 5. *We have that*

$$\left\langle \beta^{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \right\rangle = \left\langle \tilde{\beta}^{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \right\rangle$$

with $\tilde{\beta}^{n,\mathcal{I}_T^0} = VV^\top \beta^{n,\mathcal{I}_T^0}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_T^0}]$.

Proof. It would suffice to prove that

$$VV^\top b_{\mathcal{I}_T^0} = b_{\mathcal{I}_T^0},$$

which is equivalent to $b_{\mathcal{I}_T^0}^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_T^0}]$. By definition, for any $j \in \mathcal{I}_T^0$ we know $b_{j,T} = \mathbb{E}[Y_{j,T}|\mathcal{LF}, j \in \mathcal{I}_T^0]$. Lastly, by Assumption 12 there exists $\xi^{(0,T)} \in \mathbb{R}^p$ such that

$$\mathbb{E}[Y_{j,T}|\mathcal{LF}, j \in \mathcal{I}_T^0] = \sum_{i=1}^p \xi_i^{(0,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^0})_{ij}|\mathcal{LF}, j \in \mathcal{I}_T^0].$$

This concludes the proof. □

Lemma 5 allows us to write

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}_T^0}, \hat{b}_{\mathcal{I}_T^0} \right\rangle - \left\langle \tilde{\beta}^{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \right\rangle$$
$$= \underbrace{\langle \tilde{\beta}^{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}_T^0} = \hat{b}_{\mathcal{I}_T^0} - b_{\mathcal{I}_T^0}$ and $\Delta_{n,\mathcal{I}_T^0} = \hat{\beta}^{n,\mathcal{I}_T^0} - \tilde{\beta}^{n,\mathcal{I}_T^0}$.[16]

Bounding term 1a: For this term, we first state the following result without proof.

Lemma 6 (Appendix B.4, Lemma 8 of Agarwal et al. (2020b)). *Given any $n \in [N]$, $d \in [A]$, and $t \in [T]$ let all relevant assumptions hold. Then conditioned on the latent factors and treatment assignments, we have that*

$$\|\tilde{\beta}^{n,\mathcal{I}_t^d}\|_2 \le C \cdot \sqrt{\frac{k_{\mathcal{I}_t^d}}{|\mathcal{I}_t^d|}}$$

for some constant $C > 0$. This immediately implies that $\|\tilde{\beta}^{n,\mathcal{I}_t^d}\|_1 \le C\sqrt{k_{\mathcal{I}_t^d}}$.

---

[16]Notice that the analysis from Agarwal et al. (2020b) that resolved the donor unit analysis no longer applies since $\eta_{\mathcal{I}_T^0}$ is not composed of independent σ^2-subgaussian random variables.

Note that by Hölder and Cauchy-Schwarz Inequalities,

$$\langle \tilde{\beta}^{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle \leq \|\tilde{\beta}^{n,\mathcal{I}_T^0}\|_1 \cdot \|\eta_{\mathcal{I}_T^0}\|_\infty$$

$$\leq \sqrt{|\mathcal{I}_T^0|} \cdot \|\tilde{\beta}^{n,\mathcal{I}_T^0}\|_2 \cdot \|\eta_{\mathcal{I}_T^0}\|_\infty.$$

Lemma 6 gives us that $\|\tilde{\beta}^{n,\mathcal{I}_T^0}\|_2 \leq C\sqrt{k/|\mathcal{I}_T^0|}$. Donor Baseline Consistency (Equation 66) yields

$$\|\eta_{\mathcal{I}_T^0}\|_\infty = O_p\left(\sqrt{\log(p|\mathcal{I}_T^0|)}\left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{\frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_T^0| - 1}}\right\}\right]\right). \tag{67}$$

Combining both results, we know

$$\langle \tilde{\beta}^{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle = O_p\left(\sqrt{k\log(p|\mathcal{I}_T^0|)}\left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{\frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_T^0| - 1}}\right\}\right]\right). \tag{68}$$

Bounding term 2b: Once again we state the following without proof

Lemma 7 (Appendix B.4, Lemma 7 of Agarwal et al. (2020b)). *Let the setup of Lemma 6 hold, the w.p. at least $1 - O(1/(p|\mathcal{I}_t^d|)^{10})$,*

$$\|\tilde{\beta}^{n,\mathcal{I}_t^d} - \hat{\beta}^{n,\mathcal{I}_t^d}\|_2^2 \leq C(\sigma) \cdot \log(p|\mathcal{I}_t^d|)\left(\frac{k^{3/2}}{p^{1/2}|\mathcal{I}_t^d|} + \frac{k^3}{\min\{p, |\mathcal{I}_t^d|\}^2}\right),$$

where $C(\sigma)$ is a constant that only depends on σ, which appears in Assumption 9.

Once again note that

$$\langle \Delta_{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle \leq \sqrt{|\mathcal{I}_T^0|} \cdot \|\Delta_{n,\mathcal{I}_T^0}\|_2 \cdot \|\eta_{\mathcal{I}_T^0}\|_\infty,$$

where using Lemma 7 and Equation 67 gives us

$$\langle \Delta_{n,\mathcal{I}_T^0}, \eta_{\mathcal{I}_T^0} \rangle \tag{69}$$
$$= O_p\left(\sqrt{|\mathcal{I}_T^0|} \cdot \sqrt{\log(p|\mathcal{I}_T^0|)}\left(\frac{k^{3/4}}{p^{1/4}|\mathcal{I}_T^0|^{1/2}} + \frac{k^{3/2}}{\min\{p, |\mathcal{I}_T^0|\}}\right) \cdot \sqrt{\log(p|\mathcal{I}_T^0|)}\left(\frac{k^{3/4}}{p^{1/4}}\right.\right.$$
$$\left.\left. + k^2 \max\left\{\frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_T^0| - 1}}\right\}\right)\right).$$

Bounding term 2c: Lemma 5 and Cauchy–Schwarz gives

$$\langle \Delta_{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0} \rangle = \langle b_{\mathcal{I}_T^0}, VV^\top \Delta_{n,\mathcal{I}_T^0} \rangle \leq \|b_{\mathcal{I}_T^0}\|_2 \cdot \|VV^\top \Delta_{n,\mathcal{I}_T^0}\|_2.$$

We introduce the following result without proof.

Lemma 8 (Appendix C, Lemma 9 of Agarwal et al. (2020b))**.** *Let the setup from Lemma 6 hold, then*

$$VV^\top \Delta_{n,\mathcal{I}_t^d} = O_p\left(\frac{\sqrt{k}}{\sqrt{|\mathcal{I}_t^d|}p^{1/4}} + \frac{k^{3/2}\sqrt{\log(p|\mathcal{I}_t^d|)}}{\sqrt{|\mathcal{I}_t^d|}\cdot\min\{\sqrt{p},\sqrt{|\mathcal{I}_t^d|}\}} + \frac{k^2\sqrt{\log(p|\mathcal{I}_t^d|)}}{\min\{p^{3/2},|\mathcal{I}_t^d|^{3/2}\}}\right).$$

Assumption 10 gives for any $j \in \mathcal{I}_T^0$

$$|b_{j,T}| = \left|\mathbb{E}[Y_{j,T}^{(\bar{0}^T)}]\right| \leq 1,$$

which lets us conclude

$$\|b_{\mathcal{I}_T^0}\|_2 \leq \sqrt{|\mathcal{I}_T^0|}.$$

Together we know

$$\langle \Delta_{n,\mathcal{I}_T^0}, b_{\mathcal{I}_T^0}\rangle = O_p\left(\frac{\sqrt{k}}{p^{1/4}} + \frac{k^{3/2}\sqrt{\log(p|\mathcal{I}_T^0|)}}{\min\{\sqrt{p},\sqrt{|\mathcal{I}_T^0|}\}} + \frac{k^2\sqrt{|\mathcal{I}_T^0|\log(p|\mathcal{I}_T^0|)}}{\min\{p^{3/2},|\mathcal{I}_T^0|^{3/2}\}}\right). \tag{70}$$

Combining the Equations 68, 69, and 70 gives the following final rate for units $n \notin \mathcal{I}_T^0$:[17]

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF}$$
$$= O_p\left(\sqrt{k\log(p|\mathcal{I}_T^0|)}\left(\frac{k^{3/4}}{p^{1/4}} + k^2\max\left\{\frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_T^0|-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right).$$

Baseline Consistency: The above two sections allows us to conclude that for any $n \in [N]$

$$\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} \tag{71}$$
$$= O_p\left(\sqrt{\log(p|\mathcal{I}_T^0|)}\left(\frac{k^{5/4}}{p^{1/4}} + k^{5/2}\max\left\{\frac{\sqrt{|\mathcal{I}_T^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_T^0|-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right).$$

2. Verifying Terminal Blip Consistency:

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}_T^d$. Denote $X_{\mathcal{I}_T^d\setminus n} = X_{:,\mathcal{I}_T^d\setminus n} \in \mathbb{R}^{p\times|\mathcal{I}_T^d\setminus n|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = \underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}_T^d}, Y_{\mathcal{I}_T^d\setminus n}\right\rangle - \left\langle \phi^{n,\mathcal{I}_T^d}, \mathbb{E}[Y_{\mathcal{I}_T^d\setminus n} \mid \mathcal{LF}]\right\rangle}_{\text{Term 1}} + \underbrace{b_{n,T} \mid \mathcal{LF} - \hat{b}_{n,T}}_{\text{Term 2}},$$

where $\hat{\phi}^{n,\mathcal{I}_T^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank

---

[17]In order to get this final rate we made some assumptions on how $|\mathcal{I}_T^0|$ and p grow relative to each other.

$k_{\mathcal{I}_T^d \backslash n}$-approximation $X_{\mathcal{I}_T^d \backslash n}$ with $k_{\mathcal{I}_T^d \backslash n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_T^d \backslash n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_T^d \backslash n}$.

Bounding Term 1: This argument is nearly identical to that for *Donor Set Baseline Consistency.*

Lemma 9. *We have that*

$$\left\langle \phi^{n, \mathcal{I}_T^d}, \mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}] \right\rangle = \left\langle \tilde{\phi}^{n, \mathcal{I}_T^d}, \mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}] \right\rangle$$

with $\tilde{\phi}^{n, \mathcal{I}_T^d} = VV^\top \phi^{n, \mathcal{I}_T^d}$, *where* V *denotes the right singular vectors of* $\mathbb{E}[X_{\mathcal{I}_T^d \backslash n}]$.

Proof. It would suffice to prove that

$$VV^\top \mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}] = \mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}],$$

which is equivalent to $\mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}]^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_T^d \backslash n}]$. By Assumption 12 there exists $\xi^{(d,T)}$ such that for any $j \in \mathcal{I}_T^d \backslash n$

$$\mathbb{E}[Y_{j,T} | \mathcal{LF}, j \in \mathcal{I}_T^d \backslash n] = \sum_{i=1}^p \xi_i^{(d,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^d \backslash n})_{ij} | \mathcal{LF}, j \in \mathcal{I}_T^d \backslash n].$$

This concludes the proof.

$\square$

Using Lemma 9, we can once again use the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$\text{Term } 1 = \left\langle \hat{\phi}^{n, \mathcal{I}_T^d}, Y_{\mathcal{I}_T^d \backslash n} \right\rangle - \left\langle \tilde{\phi}^{n, \mathcal{I}_T^d}, \mathbb{E}[Y_{\mathcal{I}_T^d \backslash n}] \right\rangle \tag{72}$$

$$= O_p \left(\sqrt{\log(p|\mathcal{I}_T^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max \left\{ \frac{\sqrt{|\mathcal{I}_T^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_T^d| - 1}} \right\} \right] \right).$$

Bounding Term 2: This rate is exactly as listed in Equation 71.

Combining Term 1 and 2 rates, we find for any $n \in \mathcal{I}_T^d$

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = O_p \left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{5/4}}{p^{1/4}} + k^{5/2} \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \tag{73}$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$.

Non-Donor Set Consistency: Consider unit $n \notin \mathcal{I}_T^d$. Denote $X_{\mathcal{I}_T^d} = X_{:,\mathcal{I}_T^d} \in \mathbb{R}^{p \times |\mathcal{I}_T^d|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n, \mathcal{I}_T^d}, \hat{\gamma}_{\mathcal{I}_T^d, T, T}(d) \right\rangle - \left\langle \beta^{n, \mathcal{I}_T^d}, \gamma_{\mathcal{I}_T^d, T, T}(d) \right\rangle,$$

where $\hat{\beta}^{n,\mathcal{I}_T^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_T^d}$-approximation $X_{\mathcal{I}_T^d}$ with $k_{\mathcal{I}_T^d} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_T^d}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_T^d}$.

We use an essentially identical argument to that established in *Non-Donor Set Baseline Consistency.*

Lemma 10. *We have that*

$$\left\langle \beta^{n,\mathcal{I}_T^d}, \gamma_{\mathcal{I}_T^d,T,T}(d) \right\rangle = \left\langle \tilde{\beta}^{n,\mathcal{I}_T^d}, \gamma_{\mathcal{I}_T^d,T,T}(d) \right\rangle$$

with $\tilde{\beta}^{n,\mathcal{I}_T^d} = VV^\top \beta^{n,\mathcal{I}_T^d}$, *where* V *denotes the right singular vectors of* $\mathbb{E}[X_{\mathcal{I}_T^d}]$.

Proof. It would suffice to prove that

$$VV^\top \gamma_{\mathcal{I}_T^d,T,T}(d) = \gamma_{\mathcal{I}_T^d,T,T}(d),$$

which is equivalent to $\gamma_{\mathcal{I}_T^d,T,T}(d)^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_T^d}]$. To that end, recall for any $j \in \mathcal{I}_T^d$

$$
\begin{aligned}
\gamma_{j,T,T}(d) &= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \mathbb{E}[Y_{j,T}^{(\bar{0}^T)}|\mathcal{LF}, j \in \mathcal{I}_T^d] \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \mathbb{E}[\langle v_{j,T}, w_{(\bar{0}^T)}\rangle + \varepsilon_{j,T}^{(\bar{0}^T)}|\mathcal{LF}, j \in \mathcal{I}_T^d] \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \langle v_{j,T}, w_{(\bar{0}^T)}\rangle|\mathcal{LF}, j \in \mathcal{I}_T^d \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \langle v_{j,T}, w_{(\bar{0}^T)}\rangle|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \left\langle v_{j,T}, \sum_{i=1}^p \alpha_i^{(0,T)} \cdot \rho_i \right\rangle|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \sum_{i=1}^p \alpha_i^{(0,T)} \cdot \langle v_{j,T}, \rho_i \rangle|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \sum_{i=1}^p \alpha_i^{(0,T)} \cdot \mathbb{E}[\langle v_{j,T}, \rho_i \rangle + \varepsilon_{ji}|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d] \\
&= \mathbb{E}[Y_{j,T}^{(0_1,\ldots,0_{T-1},d)}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \sum_{i=1}^p \alpha_i^{(0,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^d})_{ij}|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d] \\
&= \sum_{i=1}^p \xi_i^{(d,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^d})_{ij}|\mathcal{LF}, j \in \mathcal{I}_T^d] - \sum_{i=1}^p \alpha_i^{(0,T)} \cdot \mathbb{E}[(X_{\mathcal{I}_T^d})_{ij}|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d] \\
&= \sum_{i=1}^p (\xi_i^{(d,T)} - \alpha_i^{(0,T)}) \cdot \mathbb{E}[(X_{\mathcal{I}_T^d})_{ij}|\mathcal{LF}, \{\rho_i\}_{i\in[p]}, j \in \mathcal{I}_T^d].
\end{aligned}
$$

We use Lemma 2 in the fifth equality and Assumption 12 in the second to last equality. The remainder of the steps follows from relevant definitions and standard manipulations. This completes the proof. $\square$

Lemma [10] allows us to write

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}_T^d}, \hat{\gamma}_{\mathcal{I}_T^d,T,T}(d) \right\rangle - \left\langle \beta^{n,\mathcal{I}_T^d}, \gamma_{\mathcal{I}_T^d,T,T}(d) \right\rangle$$

$$= \underbrace{\langle \tilde{\beta}^{n,\mathcal{I}_T^d}, \eta_{\mathcal{I}_T^d}(d) \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n,\mathcal{I}_T^d}, \eta_{\mathcal{I}_T^d}(d) \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n,\mathcal{I}_T^d}, \gamma_{\mathcal{I}_T^d,T,T}(d) \rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}_T^d}(d) = \hat{\gamma}_{\mathcal{I}_T^d,T,T}(d) - \gamma_{\mathcal{I}_T^d,T,T}(d)$ and $\Delta_{n,\mathcal{I}_T^d} = \hat{\beta}^{n,\mathcal{I}_T^d} - \tilde{\beta}^{n,\mathcal{I}_T^d}$. Using the previously referenced argument and applying the appropriate version of Lemmas [6], [7], and [8] allows us to claim for $n \notin \mathcal{I}_T^d$

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = O_p \left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{7/4}}{p^{1/4}} + k^3 \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \quad (74)$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$.

Terminal Blip Consistency: The above two sections allows us to conclude that for any $n \in [N]$

$$\hat{\gamma}_{n,T,T}(d) - \gamma_{n,T,T}(d) \mid \mathcal{LF} = O_p \left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{7/4}}{p^{1/4}} + k^3 \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \quad (75)$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_T^d|\}$.

3. Verifying Non-Terminal Blip Consistency:

For any unit $n \in [N]$, treatment $d \in [A]$, and $t \in [1, \ldots, T-1]$, consider the statement $P_{d,n}(t)$:

$$\hat{\gamma}_{n,T,t}(d) - \gamma_{n,T,t}(d) \mid \mathcal{LF}$$

$$= O_p \left((T-t)\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{(T-t)}}{p^{1/4}} + k^{(T-t)} \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_q^{D_{n,q}}|)_{n \in [N], q \in [t+1, \ldots, T]}\}$ with $\pi_\mathcal{I} = \max \mathcal{F}, \alpha_\mathcal{I} = \min \mathcal{F}$.

We proceed by strong induction.

To that end, consider the base case $t = T-1$, i.e., proving $P_{d,n}(T-1)$:

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}_{T-1}^d$. Denote $X_{\mathcal{I}_{T-1}^d \backslash n} = X_{:,\mathcal{I}_{T-1}^d \backslash n} \in \mathbb{R}^{p \times |\mathcal{I}_{T-1}^d \backslash n|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,T,T-1}(d) - \gamma_{n,T,T-1}(d) \mid \mathcal{LF} = \underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, Y_{\mathcal{I}_{T-1}^d \backslash n} \right\rangle - \left\langle \phi^{n,\mathcal{I}_{T-1}^d}, \mathbb{E}[Y_{\mathcal{I}_{T-1}^d \backslash n} \mid \mathcal{LF}] \right\rangle}_{\text{Term 1}}$$

$$+ \underbrace{\left\langle \phi^{n,\mathcal{I}_{T-1}^d}, b_{\mathcal{I}_{T-1}^d \backslash n} \mid \mathcal{LF} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{b}_{\mathcal{I}_{T-1}^d \backslash n} \right\rangle}_{\text{Term 2}}$$

$$+ \underbrace{\left\langle \phi^{n,\mathcal{I}_{T-1}^d}, \gamma_{\mathcal{I}_{T-1}^d \backslash n,T,T}(D_{\mathcal{I}_{T-1}^d \backslash n,T}) \mid \mathcal{LF} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{\gamma}_{\mathcal{I}_{T-1}^d \backslash n,T,T}(D_{\mathcal{I}_{T-1}^d \backslash n,T}) \right\rangle}_{\text{Term 3}}.$$

where $\gamma_{\mathcal{I}_{T-1}^d\backslash n,T,T}(D_{\mathcal{I}_{T-1}^d\backslash n,T}) = [(\gamma_{j,T,T}(D_{j,T}))_{j\in\mathcal{I}_{T-1}^d}]^\top$ and $\hat{\phi}^{n,\mathcal{I}_{T-1}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_{T-1}^d\backslash n}$-approximation $X_{\mathcal{I}_{T-1}^d\backslash n}$ with $k_{\mathcal{I}_{T-1}^d\backslash n} = \mathrm{rank}(\mathbb{E}[X_{\mathcal{I}_{T-1}^d\backslash n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_{T-1}^d\backslash n}$.

Bounding Term 1: We prove a similar row space result.

Lemma 11. *We have for any $t \in [T-1]$*

$$\left\langle \phi^{n,\mathcal{I}_t^d}, \mathbb{E}[Y_{\mathcal{I}_t^d\backslash n}] \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_t^d}, \mathbb{E}[Y_{\mathcal{I}_t^d\backslash n}] \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}_t^d} = VV^\top \phi^{n,\mathcal{I}_t^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_t^d\backslash n}]$.

Proof. It would suffice to prove that

$$VV^\top \mathbb{E}[Y_{\mathcal{I}_t^d\backslash n}] = \mathbb{E}[Y_{\mathcal{I}_t^d\backslash n}],$$

which is equivalent to $\mathbb{E}[Y_{\mathcal{I}_t^d\backslash n}]^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_t^d\backslash n}]$. By Assumption 12 there exists $\xi^{(d,t)}$ such that for any $j \in \mathcal{I}_t^d \backslash n$

$$\mathbb{E}[Y_{j,T}|\mathcal{LF}, j \in \mathcal{I}_t^d \backslash n] = \sum_{i=1}^p \xi_i^{(d,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^d\backslash n})_{ij}|\mathcal{LF}, j \in \mathcal{I}_t^d \backslash n].$$

This concludes the proof. $\qquad\square$

Using Lemma 11 for $t = T-1$, we use the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$\text{Term } 1 = \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, Y_{\mathcal{I}_{T-1}^d\backslash n} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, \mathbb{E}[Y_{\mathcal{I}_{T-1}^d\backslash n}] \right\rangle \tag{76}$$

$$= O_p\left(\sqrt{\log(p|\mathcal{I}_{T-1}^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{ \frac{\sqrt{|\mathcal{I}_{T-1}^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_{T-1}^d|-1}} \right\} \right] \right).$$

Bounding Term 2:

Lemma 12. *We have for any $t \in [T-1]$*

$$\left\langle \phi^{n,\mathcal{I}_t^d}, b_{\mathcal{I}_t^d\backslash n} \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_t^d}, b_{\mathcal{I}_t^d\backslash n} \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}_t^d} = VV^\top \phi^{n,\mathcal{I}_t^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_t^d\backslash n}]$.

Proof. It would suffice to prove that

$$VV^\top b_{\mathcal{I}_t^d\backslash n} = b_{\mathcal{I}_t^d\backslash n},$$

which is equivalent to $(b_{\mathcal{I}_t^d\backslash n,T})^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_t^d\backslash n}]$. Applying Lemma 2 we know

for any $j \in \mathcal{I}_t^d \setminus n$

$$b_{j,T} = \langle v_{j,T}, w_{(\bar{0}^T)} \rangle = \sum_{i=1}^p \alpha_i^{(0_T, T)} \cdot \langle v_{j,T}, \rho_i \rangle = \sum_{i=1}^p \xi_i \cdot \mathbb{E}[(X_{\mathcal{I}_t^d \setminus n})_{ij} | \mathcal{LF}, j \in \mathcal{I}_t^d \setminus n].$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Using Lemma 12 for $t = T - 1$ we can write

$$\left\langle \phi^{n,\mathcal{I}_{T-1}^d}, b_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{b}_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, b_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{b}_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle$$

Next we negate the RHS and decompose as follows:[18]

$$\left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{b}_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, b_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle$$
$$= \underbrace{\left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, \eta_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle}_{\text{Term 1a}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}_{T-1}^d \setminus n}, \eta_{\mathcal{I}_{T-1}^d \setminus n} \right\rangle}_{\text{Term 1b}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}_{T-1}^d \setminus n}, b_{\mathcal{I}_{T-1}^d \setminus n, T} \right\rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}_{T-1}^d \setminus n} = \hat{b}_{\mathcal{I}_{T-1}^d \setminus n} - b_{\mathcal{I}_{T-1}^d \setminus n}$ and $\Delta_{n,\mathcal{I}_{T-1}^d \setminus n} = \hat{\phi}^{n,\mathcal{I}_{T-1}^d} - \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}$. Using the previously referenced argument by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 71 for Terms 1a, 1b, and 1c respectively allows to claim

$$\text{Term 2} = \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{b}_{\mathcal{I}_{T-1}^d \setminus n, T} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, b_{\mathcal{I}_{T-1}^d \setminus n, T} \right\rangle \tag{77}$$
$$= O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left[\frac{k^{7/4}}{p^{1/4}} + k^3 \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}} \right\} \right] \right),$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|\}$.

Bounding Term 3:

Lemma 13. *We have for any $t \in [T-1]$ and $\ell > t$*

$$\left\langle \phi^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \setminus n, T, \ell}(D_{\mathcal{I}_t^d \setminus n, \ell}) \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \setminus n, T, \ell}(D_{\mathcal{I}_t^d \setminus n, \ell}) \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}_t^d} = VV^\top \phi^{n,\mathcal{I}_t^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_t^d \setminus n}]$.

Proof. It would suffice to prove that

$$VV^\top \gamma_{\mathcal{I}_t^d \setminus n, T, \ell}(D_{\mathcal{I}_t^d \setminus n, \ell}) = \gamma_{\mathcal{I}_t^d \setminus n, T, \ell}(D_{\mathcal{I}_t^d \setminus n, \ell}),$$

which is equivalent to $(\gamma_{\mathcal{I}_t^d \setminus n, T, \ell}(D_{\mathcal{I}_t^d \setminus n, \ell}))^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_t^d \setminus n}]$. Assumption 12 and

---

[18]The negation is used primarily for convenience sake as it makes no difference in the final rate.

Lemma 2 give the existence of $\xi^{(d,t)}$ and $\xi^{(d,t)'}$ such that for any $j \in \mathcal{I}_t^d \setminus n$

$$\gamma_{j,T,\ell}(D_{j,\ell}) = \langle \psi_j^{T,\ell}, w_{D_{j,\ell}} - w_{0\ell} \rangle \pm \sum_{t \neq \ell} \langle \psi_j^{T,t}, w_{D_{j,t}} \rangle$$

$$= \mathbb{E}[Y_{j,T}] - \mathbb{E}[Y_{j,T}^{(\overline{D_{j,-\ell} \cup 0_\ell})}]$$

$$= \sum_{i=1}^{p} (\xi_i^{(d,t)} - \xi^{(d,t)'}) \cdot \mathbb{E}[(X_{\mathcal{I}_t^d \setminus n})_{ij} | \mathcal{LF}, j \in \mathcal{I}_t^d \setminus n].$$

This concludes the proof. $\square$

Using Lemma 13 for $t = T - 1$ and $\ell = T$ we can write

$$\left\langle \phi^{n,\mathcal{I}_{T-1}^d}, \gamma_{\mathcal{I}_{T-1}^d \setminus n, T, T}(D_{\mathcal{I}_{T-1}^d \setminus n, T}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{\gamma}_{\mathcal{I}_{T-1}^d \setminus n, T, T}(D_{\mathcal{I}_{T-1}^d \setminus n, T}) \right\rangle$$

$$= \left\langle \tilde{\phi}^{n,\mathcal{I}_{T-1}^d}, \gamma_{\mathcal{I}_{T-1}^d \setminus n, T, T}(D_{\mathcal{I}_{T-1}^d \setminus n, T}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_{T-1}^d}, \hat{\gamma}_{\mathcal{I}_{T-1}^d \setminus n, T, T}(D_{\mathcal{I}_{T-1}^d \setminus n, T}) \right\rangle$$

At this point we can follow the earlier approach for Term 2 by negating, using the same decomposition, and applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 75 to write

$$\text{Term } 3 = O_p \left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{9/4}}{p^{1/4}} + k^{7/2} \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right)$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$. Notice that this dominates the rates for Terms 1 and 2 and as such we also have for any $n \in \mathcal{I}_{T-1}^d$

$$\hat{\gamma}_{n,T,T-1}(d) - \gamma_{n,T,T-1}(d) \mid \mathcal{LF} \tag{78}$$

$$= O_p \left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{9/4}}{p^{1/4}} + k^{7/2} \max \left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$.

Non-Donor Set Consistency: Consider any $t \in [T-1]$ and unit $n \notin \mathcal{I}_t^d$. Denote $X_{\mathcal{I}_t^d} = X_{:, \mathcal{I}_t^d} \in \mathbb{R}^{p \times |\mathcal{I}_t^d|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,T,t}(d) - \gamma_{n,T,t}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}_t^d}, \hat{\gamma}_{\mathcal{I}_t^d, T, t}(d) \right\rangle - \left\langle \beta^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d, T, t}(d) \right\rangle,$$

where $\hat{\beta}^{n,\mathcal{I}_t^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_t^d}$-approximation $X_{\mathcal{I}_t^d}$ with $k_{\mathcal{I}_t^d} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_t^d}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_t^d}$.

We use an identical argument to that established in *Baseline Consistency – Non-Donor Set.*

Lemma 14. *We have that*

$$\left\langle \beta^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d, T, t}(d) \right\rangle = \left\langle \tilde{\beta}^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d, T, t}(d) \right\rangle$$

with $\tilde{\beta}^{n,\mathcal{I}_t^d} = VV^\top \beta^{n,\mathcal{I}_t^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_t^d}]$.

Proof. It would suffice to prove that

$$VV^\top \gamma_{\mathcal{I}_t^d, T, t}(d) = \gamma_{\mathcal{I}_t^d, T, t}(d),$$

which is equivalent to $\gamma_{\mathcal{I}_t^d, T, t}(d)^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_t^d}]$. To that end, recall for any $j \in \mathcal{I}_t^d$,

$$
\begin{aligned}
\gamma_{j,T,t}(d) &= \langle \psi_j^{T,t}, w_d - w_{0_t} \rangle \mid \mathcal{LF} \\
&= \langle \psi_j^{T,t}, w_d - w_{0_t} \rangle \pm \sum_{\ell \neq t} \langle \psi_j^{T,\ell}, w_{0_\ell} \rangle \mid \mathcal{LF} \\
&= \mathbb{E}[\langle v_{j,T}, w_{(\bar{0}^{t-1}, d, \underline{0}^{t+1})} \rangle + \varepsilon_{j,T}^{(\bar{0}^{t-1}, d, \underline{0}^{t+1})} \mid \mathcal{LF}, j \in \mathcal{I}_t^d] - \mathbb{E}[\langle v_{j,T}, w_{(\bar{0}^T)} \rangle + \varepsilon_{j,T}^{(\bar{0}^T)} \mid \mathcal{LF}, j \in \mathcal{I}_t^d] \\
&= \langle v_{j,T}, w_{(\bar{0}^{t-1}, d, \underline{0}^{t+1})} \rangle - \langle v_{j,T}, w_{(\bar{0}^T)} \rangle \mid \mathcal{LF}, j \in \mathcal{I}_t^d \\
&= \langle v_{j,T}, w_{(\bar{0}^{t-1}, d, \underline{0}^{t+1})} - w_{(\bar{0}^T)} \rangle \mid \mathcal{LF}, \{\rho_i\}_{i \in [p]}, j \in \mathcal{I}_t^d \\
&= \left\langle v_{j,T}, \sum_{i=1}^p \alpha_i^{(d,t)} \rho_i - \sum_{i=1}^p \alpha_i^{(0,t)} \rho_i \right\rangle \mid \mathcal{LF}, \{\rho_i\}_{i \in [p]}, j \in \mathcal{I}_t^d \\
&= \sum_{i=1}^p (\alpha_i^{(d,t)} - \alpha_i^{(0,t)}) \cdot \mathbb{E}[\langle v_{j,T}, \rho_i \rangle + \varepsilon_{ji} \mid \mathcal{LF}, \{\rho_i\}_{i \in [p]}, j \in \mathcal{I}_t^d] \\
&= \sum_{i=1}^p (\alpha_i^{(d,t)} - \alpha_i^{(0,t)}) \cdot \mathbb{E}[(X_{\mathcal{I}_t^d})_{ij} \mid \mathcal{LF}, \{\rho_i\}_{i \in [p]}, j \in \mathcal{I}_t^d].
\end{aligned}
$$

The sixth equality is due to Assumption 12 being applied to each term. The remainder of the steps follows from relevant definitions and standard manipulations. This completes the proof. $\qquad\square$

Using the above framework and Lemma 14 with $t = T-1$ allows us to write

$$
\begin{aligned}
\hat{\gamma}_{n,T,T-1}(d) - \gamma_{n,T,T-1}(d) \mid \mathcal{LF} &= \left\langle \hat{\beta}^{n,\mathcal{I}_{T-1}^d}, \hat{\gamma}_{\mathcal{I}_t^d, T, T-1}(d) \right\rangle - \left\langle \tilde{\beta}^{n,\mathcal{I}_{T-1}^d}, \gamma_{\mathcal{I}_t^d, T, T-1}(d) \right\rangle \\
&= \underbrace{\langle \tilde{\beta}^{n,\mathcal{I}_{T-1}^d}, \eta_{\mathcal{I}_{T-1}^d}(d) \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n,\mathcal{I}_{T-1}^d}, \eta_{\mathcal{I}_{T-1}^d}(d) \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n,\mathcal{I}_{T-1}^d}, \gamma_{\mathcal{I}_t^d, T, T-1}(d) \rangle}_{\text{Term 1c}},
\end{aligned}
$$

where $\eta_{\mathcal{I}_{T-1}^d}(d) = \hat{\gamma}_{\mathcal{I}_t^d, T, T-1}(d) - \gamma_{\mathcal{I}_t^d, T, T-1}(d)$ and $\Delta_{n,\mathcal{I}_{T-1}^d} = \hat{\beta}^{n,\mathcal{I}_{T-1}^d} - \tilde{\beta}^{n,\mathcal{I}_{T-1}^d}$. Using the previously referenced argument by applying the appropriate version of Lemmas 6, 7, and 8 allows to claim for $n \notin \mathcal{I}_{T-1}^d$

$$\hat{\gamma}_{n,T,T-1}(d) - \gamma_{n,T,T-1}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{11/4}}{p^{1/4}} + k^4 \max\left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \tag{79}$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_{T-1}^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$. Combining equations 78 and 79 yields the base case.

Inductive Step: We assume $P_{d,n}(\ell)$ for $\ell \in [t+1, \ldots, T-1]$ and prove $P_{d,n}(t)$.

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}_t^d$. Denote $X_{\mathcal{I}_t^d \backslash n} = X_{:, \mathcal{I}_t^d \backslash n} \in \mathbb{R}^{p \times |\mathcal{I}_t^d \backslash n|}$. We know the baseline outcome admits the representation

$$
\begin{aligned}
\hat{\gamma}_{n,T,t}(d) - \gamma_{n,T,t}(d) \mid \mathcal{LF} = &\underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, Y_{\mathcal{I}_t^d \backslash n} \right\rangle - \left\langle \phi^{n,\mathcal{I}_t^d}, \mathbb{E}[Y_{\mathcal{I}_t^d \backslash n} \mid \mathcal{LF}] \right\rangle}_{\text{Term } 1} \\
&+ \underbrace{\left\langle \phi^{n,\mathcal{I}_t^d}, b_{\mathcal{I}_t^d \backslash n} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{b}_{\mathcal{I}_t^d \backslash n} \right\rangle}_{\text{Term } 2} \\
&+ \underbrace{\left\langle \phi^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \backslash n, T, T}(D_{\mathcal{I}_t^d \backslash n, T}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{\gamma}_{\mathcal{I}_t^d \backslash n, T, T}(D_{\mathcal{I}_t^d \backslash n, T}) \right\rangle}_{\text{Term } 3} \\
&+ \sum_{\ell=t+1}^{T-1} \left(\underbrace{\left\langle \phi^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \backslash n, T, \ell}(D_{\mathcal{I}_t^d \backslash n, \ell}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{\gamma}_{\mathcal{I}_t^d \backslash n, T, \ell}(D_{\mathcal{I}_t^d \backslash n, \ell}) \right\rangle}_{\text{Term } \ell} \right).
\end{aligned}
$$

where $\hat{\phi}^{n,\mathcal{I}_t^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_t^d \backslash n}$-approximation $X_{\mathcal{I}_t^d \backslash n}$ with $k_{\mathcal{I}_t^d \backslash n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_t^d \backslash n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_t^d \backslash n}$.

Bounding Term 1: We simply use Lemma 11 which holds for any $t \in [T-1]$ to leverage the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$
\begin{aligned}
\text{Term } 1 = &\left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, Y_{\mathcal{I}_t^d \backslash n} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_t^d}, \mathbb{E}[Y_{\mathcal{I}_t^d \backslash n}] \right\rangle \\
= &O_p \left(\sqrt{\log(p|\mathcal{I}_t^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max \left\{ \frac{\sqrt{|\mathcal{I}_t^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_t^d|-1}} \right\} \right] \right).
\end{aligned}
\tag{80}
$$

Bounding Term 2: Using the previously referenced argument for Term 2 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 71 and Lemma 12 we know

$$
\begin{aligned}
\text{Term } 2 = &\left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{b}_{\mathcal{I}_t^d \backslash n} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_t^d}, b_{\mathcal{I}_t^d \backslash n} \right\rangle \\
= &O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left[\frac{k^{7/4}}{p^{1/4}} + k^3 \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}} \right\} \right] \right),
\end{aligned}
\tag{81}
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|\}$.

Bounding Term 3: Using the previously referenced argument for Term 3 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 75 for any $d \in \{D_{n,T}\}_{n \in [N]}$ and Lemma 13 with $\ell = T$ to write

$$\text{Term 3} = \left\langle \phi^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \backslash n, T, T}(D_{\mathcal{I}_t^d \backslash n, T}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{\gamma}_{\mathcal{I}_t^d \backslash n, T, T}(D_{\mathcal{I}_t^d \backslash n, T}) \right\rangle \tag{82}$$

$$= O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{9/4}}{p^{1/4}} + k^{7/2} \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_T^{D_{n,T}}|)_{n \in [N]}\}$.

Bounding Term ℓ for $\ell \in [t+1, \ldots, T-1]$: For any such ℓ, we use an argument similar to Term 3 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside the inductive hypothesis $P_{d,n}(\ell)$ for all $d \in \{D_{n,\ell}\}_{n \in [N]}$ and Lemma 13 to write

$$\text{Term } \ell = \left\langle \phi^{n,\mathcal{I}_t^d}, \gamma_{\mathcal{I}_t^d \backslash n, T, \ell}(D_{\mathcal{I}_t^d \backslash n, \ell}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}_t^d}, \hat{\gamma}_{\mathcal{I}_t^d \backslash n, T, \ell}(D_{\mathcal{I}_t^d \backslash n, \ell}) \right\rangle \tag{83}$$

$$= O_p \left((T - \ell)\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{(T-\ell)}}{p^{1/4}} + k^{(T-\ell)} \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_q^{D_{n,q}}|)_{n \in [N], q \in [\ell, \ldots, T]}\}$ with $\pi_{\mathcal{I}} = \max \mathcal{F}, \alpha_{\mathcal{I}} = \min \mathcal{F}$.

Note that Terms 1-3 are dominated by the summation, as such it suffices to analyze the latter. To that end,

$$\sum_{\ell=t+1}^{T-1} \text{Term } \ell = O_p \left(\sum_{\ell=t+1}^{T-1} (T - \ell)\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{(T-\ell)}}{p^{1/4}} + k^{(T-\ell)} \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_q^{D_{n,q}}|)_{n \in [N], q \in [t+1, \ldots, T]}\}$ with $\pi_{\mathcal{I}} = \max \mathcal{F}, \alpha_{\mathcal{I}} = \min \mathcal{F}$. Notice we bounded the smaller donor set cardinalates by the largest one, i.e., when $\ell = t+1$. We analyze the time dependent terms and denote

$$C := \sqrt{\log(p\pi_{\mathcal{I}})}, \quad C' := \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\}.$$

Upon substitution and reindexing we have

$$C \sum_{m=1}^{T-t-1} m \left(\frac{k^m}{p^{1/4}} + C' k^m \right) = C \left(\frac{1}{p^{1/4}} + C' \right) \sum_{m=1}^{T-t-1} m k^m.$$

We apply the geometric sum derivative trick for $k \geq 1$

$$\sum_{m=1}^{M} m k^m = \frac{k(1 - (M+1)k^M + M k^{M+1})}{(1-k)^2} = \Theta(M k^{M+1})$$

Taking $M = T - t - 1$, we conclude

$$\sum_{\ell=t+1}^{T-1} \text{Term } \ell = O_p\left((T-t)\sqrt{\log(p\pi_\mathcal{I})}\left(\frac{k^{(T-t)}}{p^{1/4}} + k^{(T-t)}\max\left\{\frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

Combining every term yields for any $n \in \mathcal{I}_t^d$

$$\hat{\gamma}_{n,T,t}(d) - \gamma_{n,T,t}(d) \mid \mathcal{LF} \tag{84}$$
$$= O_p\left((T-t)\sqrt{\log(p\pi_\mathcal{I})}\left(\frac{k^{(T-t)}}{p^{1/4}} + k^{(T-t)}\max\left\{\frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_t^d|, (|\mathcal{I}_q^{D_{n,q}}|)_{n\in[N], q\in[t+1,...,T]}\}$ with $\pi_\mathcal{I} = \max \mathcal{F}, \alpha_\mathcal{I} = \min \mathcal{F}$.

Non-Donor Set Consistency: Applying the *Non-Donor Set Consistency* argument written for the Base Case for general t, specifically Lemma 14 for any $t \in [T-1]$, proves $P_{d,n}(t)$.

4. Verifying Target Causal Parameter Consistency: For any unit $n \in [N]$ and $\bar{d}^T \in [A]^T$ we recall the SBE-PCR estimator and the corresponding causal estimand.

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)}\right] = \sum_{t=1}^{T} \hat{\gamma}_{n,T,t}(d_t) + \hat{b}_{n,T} \quad \text{and} \quad \mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF}\right] = \sum_{t=1}^{T} \gamma_{n,T,t}(d_t) + b_{n,T} \mid \mathcal{LF}.$$

The difference is exactly

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)}\right] - \mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF}\right] = \left(\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF}\right) + \sum_{t=1}^{T} \left(\hat{\gamma}_{n,T,t}(d_t) - \gamma_{n,T,t}(d_t) \mid \mathcal{LF}\right)$$

We apply the known bound for each term, specifically Equation 71, Equation 75 with $d = d_T$, and $P_{d_t,n}(t)$ for every $t \in [T-1]$. Once again we encounter the same geometric sum, which gives the desired result upon noting that the baseline rate is dominated by that of the sum.

C.5 Proof of Theorem 4

We recall that for any unit $n \in [N]$ and $\bar{d}^T \in [A]^T$

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF}\right] = \sum_{t=1}^{T} \gamma_{n,T,t}(d_t) + b_{n,T} \mid \mathcal{LF} = \sum_{t=1}^{T} \langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \rangle + b_{n,T} \mid \mathcal{LF}.$$

Given Assumption 13 we know that $\psi_n^{T,T-q-i} = 0$ for all $i \geq 1$. As such,

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF}\right] = \sum_{t=T-q}^{T} \langle \psi_n^{T,t}, w_{d_t} - w_{0_t} \rangle + b_{n,T} \mid \mathcal{LF} = \sum_{t=T-q}^{T} \gamma_{n,T,t}(d_t) + b_{n,T} \mid \mathcal{LF}$$

We modify the SBE-PCR estimator accordingly

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}\right] := \sum_{t=T-q}^{T} \hat{\gamma}_{n,T,t}(d_t) + \hat{b}_{n,T}.$$

Applying the analysis from the proof of Theorem 3 yields the desired result.

D   Proofs and Remarks for Time-Invariant Linear Dynamical System

D.1   Proof of Proposition 3

Recall $z_{n,t}$ is the latent state of unit n if it undergoes action sequence $\bar{d}^t$. By a simple recursion we have

$$z_{n,t}^{(\bar{d}^t)} = \sum_{\ell=1}^{t} \boldsymbol{B}_n^{t-\ell} \boldsymbol{C}_n \, w_{d_\ell} + \sum_{\ell=1}^{t} \boldsymbol{B}_n^{t-\ell} \eta_{n,\ell} + \eta_{n,t}$$

Hence,

$$Y_{n,t}^{(\bar{d}^t)} = \left\langle \theta_n, \sum_{\ell=1}^{t} \boldsymbol{B}_n^{t-\ell} \boldsymbol{C}_n \, w_{d_\ell} + \sum_{\ell=1}^{t} \boldsymbol{B}_n^{t-\ell} \eta_{n,\ell} \right\rangle + \langle \tilde{\theta}_n, w_{d_t} \rangle + \tilde{\eta}_{n,t}$$

$$= \sum_{\ell=1}^{t} \left(\left\langle \psi_n^{t-\ell}, w_{d_\ell} \right\rangle + \varepsilon_{n,t,\ell} \right),$$

where in the last line we use the definitions of $\psi_n^{t-\ell}$ and $\varepsilon_{n,t,\ell}$ in the proposition statement. This completes the proof.

D.2   Proof of Theorem 5

For simplicity, we omit the conditioning on $\mathcal{LF}$ in all derivations; all expectations are conditioned on $\mathcal{LF}$.

1. Verifying (identification)**.** First, we verify (identification) holds, which allows us to express the counterfactual outcomes, in terms of the blips and the baseline. For all $n \in [N]$, using Assumption 14 we have:

$$\mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} - Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}] + \mathbb{E}[Y_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF}]$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} \langle \psi_n^{T-t}, w_{d_t} - w_{\tilde{0}} \rangle + \varepsilon_{n,T}^{(\bar{d}^T)} - \varepsilon_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF} \right] + \mathbb{E}\left[\sum_{t=1}^{T} \langle \psi_n^{T-t}, w_{\tilde{0}} \rangle + \varepsilon_{n,T}^{(\bar{0}^T)} \mid \mathcal{LF} \right]$$

$$= \sum_{t=1}^{T} \gamma_{n,T-t}(d_t) \mid \mathcal{LF} + b_{n,T} \mid \mathcal{LF}$$

2. Verifying (observed control) **&** (synthetic control)**:**

We first show (observed control) holds. For $j \in \mathcal{I}_t^0$:

$$b_{j,t} \mid \mathcal{LF} = \sum_{\ell=1}^{t} \left\langle \psi_j^{t-\ell}, w_{\tilde{0}} \right\rangle \mid \mathcal{LF} = \mathbb{E}\left[\sum_{\ell=1}^{t} \left\langle \psi_j^{t-\ell}, w_{\tilde{0}} \right\rangle + \varepsilon_{j,t}^{(\bar{0}^t)} \mid \mathcal{LF} \right] \tag{85}$$

$$= \mathbb{E}\left[\sum_{\ell=1}^{t} \left\langle \psi_j^{t-\ell}, w_{\tilde{0}} \right\rangle + \varepsilon_{j,t}^{(\bar{0}^t)} \mid \mathcal{LF}, \mathcal{I}_t^0 \right] \tag{86}$$

$$= \mathbb{E}\left[Y_{j,t}^{(\bar{0}^t)} \mid \mathcal{LF}, j \in \mathcal{I}_t^0 \right] \tag{87}$$

$$= \mathbb{E}\left[Y_{j,t} \mid \mathcal{LF}, j \in \mathcal{I}_t^0 \right], \tag{88}$$

where (85) and (87) follow from Assumption 14; (86) follows from the fact that $\left\langle \psi_j^{t-\ell}, w_{\tilde{0}} \right\rangle$ is deterministic conditional on $\mathcal{LF}$, and that $\mathbb{E}[\varepsilon_{j,t}^{(\bar{0}^t)} \mid \mathcal{LF}, \mathcal{I}_t^0] = \mathbb{E}[\varepsilon_{j,t}^{(\bar{0}^t)} \mid \mathcal{LF}]$ as seen in the definition of $\mathcal{I}_t^0$; (88) follows from Assumption 1.

Next we show (synthetic control) holds. For $i \notin \mathcal{I}_t^0$:

$$b_{i,t} \mid \mathcal{LF} = \sum_{\ell=1}^{t} \left\langle \psi_i^{t-\ell}, w_{\tilde{0}} \right\rangle \mid \mathcal{LF}$$

$$= \sum_{\ell=1}^{t} \left\langle \psi_i^{t-\ell}, w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}_t^0 \tag{89}$$

$$= \sum_{\ell=1}^{t} \sum_{j \in \mathcal{I}_t^0} \beta_j^{i,\mathcal{I}_t^0} \left\langle \psi_j^{t-\ell}, w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}_t^0 \tag{90}$$

$$= \sum_{j \in \mathcal{I}_t^0} \beta_j^{i,\mathcal{I}_t^0} b_{j,t} \mid \mathcal{LF}, \mathcal{I}_t^0$$

where (89) follows from the fact that $\left\langle \psi_i^{T-t}, w_{\tilde{0}} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (90) follows from Assumption 16;

3. Verifying ("observed" lag 0 blip) **&** (synthetic lag 0 blip)**:**

We first show ("observed" lag 0 blip) holds. For all $d \in [A]$ and $j \in \mathcal{I}^d$:

$$\gamma_{j,0}(d) \mid \mathcal{LF} = \left\langle \psi_j^0, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}$$

$$= \mathbb{E}\left[\left\langle \psi_j^0, w_d - w_{\tilde{0}} \right\rangle + \varepsilon_{j,t_j^*}^{(\bar{0}^{t_j^*-1},d)} \pm \sum_{\ell=1}^{t_j^*-1} \left\langle \psi_j^\ell, w_{\tilde{0}} \right\rangle \mid \mathcal{LF} \right] \tag{91}$$

$$= \mathbb{E}\left[\left\langle \psi_j^0, w_d \right\rangle + \varepsilon_{j,t_j^*}^{(\bar{0}^{t_j^*-1},d)} + \sum_{\ell=1}^{t_j^*-1} \left\langle \psi_j^\ell, w_{\tilde{0}} \right\rangle \mid \mathcal{LF} \right] - \sum_{\ell=0}^{t_j^*-1} \left\langle \psi_j^\ell, w_{\tilde{0}} \right\rangle \mid \mathcal{LF}$$

$$= \mathbb{E}\left[\left\langle \psi_j^0, w_d \right\rangle + \varepsilon_{j,t_j^*}^{(\bar{0}^{t_j^*-1},d)} + \sum_{t=1}^{t_j^*-1} \left\langle \psi_j^\ell, w_{\tilde{0}} \right\rangle \mid \mathcal{LF} \right] - b_{j,t_j^*} \mid \mathcal{LF}$$

$$= \mathbb{E}[Y_{j,t_j^*}^{(\bar{D}_j^{t_j^*})} \mid \mathcal{LF}, \ j \in \mathcal{I}^d] - b_{j,t_j^*} \mid \mathcal{LF} \tag{92}$$

$$= \mathbb{E}[Y_{j,t_j^*} \mid \mathcal{LF}, \ j \in \mathcal{I}^d] - b_{j,t_j^*} \mid \mathcal{LF} \tag{93}$$

where (91) follows from Assumption 14; (92) follows from the definition of $\mathcal{I}^d$ and Assumption 14; (93) follows from Assumption 1.

Next we show (synthetic lag 0 blip) holds. For $i \notin \mathcal{I}^d$

$$\gamma_{i,0}(d) \mid \mathcal{LF} = \left\langle \psi_i^0, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF} = \left\langle \psi_i^0, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}^d \tag{94}$$

$$= \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \left\langle \psi_j^0, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}^d \tag{95}$$

$$= \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \gamma_{j,0}(d) \mid \mathcal{LF}, \mathcal{I}^d$$

(94) follows from the fact that $\left\langle \psi_i^0, w_d - w_{\tilde{0}} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (95) follows from Assumption 16.

4. Verifying ("observed" lag t blip) **&** (synthetic lag t blip)**:**

We first show ("observed" lag t blip) holds. For all $d \in [A]$, $t \in [T-1]$, $j \in \mathcal{I}^d$:

$$\mathbb{E}\left[Y_{j,t_j^*+t} - Y_{j,t_j^*+t}^{(\bar{0}^{t_j^*+t})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] = \mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t})} - Y_{j,t_j^*+t}^{(\bar{0}^{t_j^*+t})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] \tag{96}$$

$$= \mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t})} - Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*-1},\underline{0}^{t_j^*})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] \tag{97}$$

$$= \sum_{\ell=0}^{t} \mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell},\underline{0}^{t_j^*+t-\ell+1})} - Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell-1},\underline{0}^{t_j^*+t-\ell})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] \tag{98}$$

where (96) follows from Assumption 1; (97) uses that for $j \in \mathcal{I}^d$, $\bar{D}_n^{t_j^*-1} = (\tilde{0}, \ldots, \tilde{0})$, and Assumption

72

1. Then,

$$\sum_{\ell=0}^{t} \mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell}, \underline{0}^{t_j^*+t-\ell+1})} - Y_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell-1}, \underline{0}^{t_j^*+t-\ell})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right]$$

$$= \sum_{\ell=0}^{t} \mathbb{E}\left[\left\langle \psi_j^{\ell}, w_{D_{j,t_j^*+t-\ell}} - w_{\tilde{0}} \right\rangle + \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell}, \underline{0}^{t_j^*+t-\ell+1})} - \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell-1}, \underline{0}^{t_j^*+t-\ell})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] \quad (99)$$

$$= \mathbb{E}\left[\left\langle \psi_j^{t}, w_{D_{j,t_j^*+t}} - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}^d \right]$$

$$\quad + \sum_{\ell=0}^{t-1} \mathbb{E}\left[\left\langle \psi_j^{\ell}, w_{D_{j,t_j^*+t-\ell}} - w_{\tilde{0}} \right\rangle + \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell}, \underline{0}^{t_j^*+t-\ell+1})} - \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell-1}, \underline{0}^{t_j^*+t-\ell})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right]$$

$$= \left\langle \psi_j^{t}, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}$$

$$\quad + \sum_{\ell=0}^{t-1} \mathbb{E}\left[\left\langle \psi_j^{\ell}, w_{D_{j,t_j^*+t-\ell}} - w_{\tilde{0}} \right\rangle + \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell}, \underline{0}^{t_j^*+t-\ell+1})} - \varepsilon_{j,t_j^*+t}^{(\bar{D}_j^{t_j^*+t-\ell-1}, \underline{0}^{t_j^*+t-\ell})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] \quad (100)$$

$$= \left\langle \psi_j^{t}, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF} + \sum_{\ell=0}^{t-1} \mathbb{E}\left[\left\langle \psi_j^{\ell}, w_{D_{j,t_j^*+t-\ell}} - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}^d \right].$$

$$\quad + \sum_{\ell=0}^{t-1} \mathbb{E}\left[\mathbb{E}\left[\varepsilon_{j,t_j^*+t}^{(\bar{\delta}^{t_j^*+t-\ell}, \underline{0}^{t_j^*+t-\ell+1})} - \varepsilon_{j,t_j^*+t}^{(\bar{\delta}^{t_j^*+t-\ell-1}, \underline{0}^{t_j^*+t-\ell})} \mid \bar{D}_j^{t_j^*+t-\ell} = \bar{\delta}^{t_j^*+t-\ell} \right] \mid \mathcal{LF}, j \in \mathcal{I}^d \right]$$

$$= \left\langle \psi_j^{t}, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF} + \sum_{\ell=0}^{t-1} \left\langle \psi_j^{\ell}, w_{D_{j,t_j^*+t-\ell}} - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, j \in \mathcal{I}^d \quad (101)$$

$$= \gamma_{j,t}(d) \mid \mathcal{LF} + \sum_{\ell=0}^{t-1} \gamma_{j,t}(D_{j,t_j^*+t-\ell}) \mid \mathcal{LF} \quad (102)$$

where (99) follows from Assumption 14; (100) follows from the definition of $\mathcal{I}^d$, i.e., for $j \in \mathcal{I}^d$, $\bar{D}_j^{t_j^*} = (\bar{0}^{t_j^*-1}, d)$; (101) follows from Assumption 17. Re-arranging (102) we have that,

$$\gamma_{j,t}(d) \mid \mathcal{LF} = \mathbb{E}\left[Y_{j,t_j^*+t} - Y_{j,t_j^*+t}^{(\bar{0}_{t_j^*+t})} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] - \sum_{\ell=0}^{t-1} \gamma_{j,t}(D_{j,t_j^*+t-\ell}) \mid \mathcal{LF}$$

$$= \mathbb{E}\left[Y_{j,t_j^*+t} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] - \mathbb{E}\left[Y_{j,t_j^*+t}^{(\bar{0}_{t_j^*+t})} \mid \mathcal{LF} \right] - \sum_{\ell=0}^{t-1} \gamma_{j,t}(D_{j,t_j^*+t-\ell}) \mid \mathcal{LF} \quad (103)$$

$$= \mathbb{E}\left[Y_{j,t_j^*+t} \mid \mathcal{LF}, j \in \mathcal{I}^d \right] - b_{j,t_j^*+t} \mid \mathcal{LF} - \sum_{\ell=0}^{t-1} \gamma_{j,t}(D_{j,t_j^*+t-\ell}) \mid \mathcal{LF} \quad (104)$$

where (103) follows from the definition of $\mathcal{I}^d$; (104) follows from Assumption 14.

Next we show (synthetic lag t blip) holds. For all $d \in [A]$, $t < T$, $i \notin \mathcal{I}^d$:

$$\gamma_{i,t}(d) \mid \mathcal{LF} = \left\langle \psi_i^t, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF} \tag{105}$$

$$= \left\langle \psi_i^t, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}^d \tag{106}$$

$$= \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \left\langle \psi_j^t, w_d - w_{\tilde{0}} \right\rangle \mid \mathcal{LF}, \mathcal{I}^d \tag{107}$$

$$= \sum_{j \in \mathcal{I}^d} \beta_j^{i,\mathcal{I}^d} \gamma_{j,t}(d) \mid \mathcal{LF}, \mathcal{I}^d$$

where (106) follows from the the fact that $\left\langle \psi_i^t, w_d - w_{\tilde{0}} \right\rangle$ is deterministic conditional on $\mathcal{LF}$; (107) follows from Assumption 16;

D.3 General Remarks on LTI Setting

D.3.1 Linear Factor Model Assumption

Assumption 14 is not restrictive. Recalling the Linear Dynamical System setting from Proposition 3, we present a few sufficient conditions for the above to hold true.

1. **Hard Memory Cutoff**

$$\exists q \in \mathbb{N}, \boldsymbol{B}_n^{q+1} = 0. \tag{108}$$

2. **Exponential Forgetting (Spectral Decay Condition)**

$$\exists C > 0, \rho \in (0,1), \text{ such that for all } t, \quad \left\| \boldsymbol{B}_n^t \right\|_2 \leq C\rho^t. \tag{109}$$

3. **Soft Memory Cutoff (Higher-Order Markov Property)**

$$\mathbb{P}(z_{n,T} \mid z_{n,T-1}, z_{n,T-2}, \ldots, z_{n,0}) = \mathbb{P}(z_{n,T} \mid z_{n,T-1}, \ldots, z_{n,T-q}). \tag{110}$$

Clearly, the first condition is the *strongest* and implies the other two. In general, this shows that our assumption of fixed memory is a reasonable one proving the effectiveness of our methodology within the dynamic treatment regime from a statistical perspective.

D.3.2 Row-Space Inclusion

The next result is a consequence of Assumption 20 and will be essential in establishing consistency.

Lemma 15. *Let Assumption 20 hold. Then for all $d \in [A]$ and $t \in [T]$ and $\ell < t$ there exists $\alpha^{(d,t,\ell)} \in \mathbb{R}^p$ such that*

$$\mathbb{E}\left[Y_{j,t_j^*+t}^{(\overline{D_{j,t_j^*+t-\ell}}, \tilde{0}^{t_j^*+t-\ell+1})} \Big| \mathcal{LF} \right] = \sum_{i=1}^{p} \alpha_i^{(d,t,\ell)} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij} | \mathcal{LF}]$$

and there exists $\alpha^{(d,t,\ell)'} \in \mathbb{R}^p$ such that

$$\mathbb{E}\left[Y_{j,t_j^*+t}^{(\overline{D_{j,t_j^*+t-\ell-1}}, \overline{\tilde{0}^{t_j^*+t-\ell}})} \Big| \mathcal{LF}\right] = \sum_{i=1}^p \alpha_i^{(d,t,\ell)'} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij}|\mathcal{LF}].$$

Proof. This holds as an immediate consequence of Assumption 20 where we consider $D_{j,t_j^*+t-\ell+i} = 0_\ell$ for any $i \in [\ell]$ for the first term and $D_{j,t_j^*+t-\ell} = 0_\ell$ as well for the second term which is fine given our assumption that $\ell < t$. $\qquad\square$

D.4   Preliminary Results for Proof of Theorem 6

We present consistent results (and their proofs) that serve as preliminaries for proving Theorem 6. This is analogous to Theorem 3, which serve as a preliminary result for Theorem 4.

Theorem 7. *Let Assumption 1 to 20 hold.*[19] *Consider the* `SBE-PCR` *estimator in Section 5.3 and suppose $k = \max_{\mathcal{I} \in \{\mathcal{I}^d\} \cup \{\mathcal{I}_t^0\}} rank(\mathbb{E}[X_\mathcal{I}])$. Then conditional on the treatment assignments, $\mathcal{LF}$, and $\{\rho_i\}_{i \in [p]}$ we have:*

   *(i) **Baseline Consistency:** For any $n \in [N]$ and $t \in [T]$*

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = O_p\left(\sqrt{\log(p|\mathcal{I}_t^0|)} \left(\frac{k^{5/4}}{p^{1/4}} + k^{5/2} \max\left\{\frac{\sqrt{|\mathcal{I}_t^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_t^0| - 1}}, \frac{1}{\sqrt{p}}\right\}\right)\right).$$

   *(ii) **Terminal Blip Consistency:** For any $d \in [A]$ and unit $n \in [N]$*

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^{9/4}}{p^{1/4}} + k^{7/2} \max\left\{\frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\pi_\mathcal{I} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|\}$ and $\alpha_\mathcal{I} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|\}$.
   *(iii) **Non-Terminal Blip Consistency:** For any $d \in [A]$, unit $n \in [N]$, and $t \in [1, \ldots, T-1]$:*

$$\hat{\gamma}_{n,t}(d) - \gamma_{n,t}(d) \mid \mathcal{LF}$$
$$= O_p\left(t\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^t}{p^{1/4}} + k^t \max\left\{\frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{C} = \{|\mathcal{I}^d|, |\mathcal{I}_T^0|, |\mathcal{I}_1^0|, (\mathcal{I}^{D_{n,t_n^*+q}})_{n \in [N], q \in [1,\ldots,t]}\}$ with $\pi_\mathcal{I} = \max \mathcal{C}, \alpha_\mathcal{I} = \min \mathcal{C}$.
   *(iv) **Target Causal Parameter Consistency:** For $n \in [N]$, and $\bar{d}^T \in [A]^T$:*

$$\widehat{\mathbb{E}}[Y_{n,T}^{(\bar{d}^T)}] - \mathbb{E}[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF}] = O_p\left(T\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^T}{p^{1/4}} + k^T \max\left\{\frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{C} = \{|\mathcal{I}_T^0|, |\mathcal{I}_1^0|, (|\mathcal{I}^{d_t}|)_{t \in [T]}, (\mathcal{I}^{D_{n,t_n^*+t}})_{n \in [N], t \in [1,\ldots,T-1]}\}$ with $\pi_\mathcal{I} = \max \mathcal{C}$ and $\alpha_\mathcal{I} = \min \mathcal{C}$. Here, each $O_p(\cdot)$ is defined with respect to the sequence $\min\{p, \alpha_\mathcal{I}\}$.[20]

---

[19]To be precise this theorem statement does not require any of the corresponding assumptions in Section 4.
[20]Notice that $\alpha_\mathcal{I} \leq \pi_\mathcal{I}$ by definition.

Below we provide a full proof of Theorem 7, which is quite similar to that of Theorem 3.

1. Verifying Baseline Consistency:

For any $t \in [T]$:

Donor Set Baseline Consistency: Consider unit $n \in \mathcal{I}_t^0$. Denote $X_{\mathcal{I}_t^0 \setminus n} = X_{:,\mathcal{I}_t^0 \setminus n} \in \mathbb{R}^{p \times |\mathcal{I}_t^0 \setminus n|}$. We know the baseline outcome admits the representation

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = \left\langle \hat{\phi}^{n,\mathcal{I}_t^0}, Y_{\mathcal{I}_t^0 \setminus n, t} \right\rangle - \left\langle \phi^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t} \mid \mathcal{LF}] \right\rangle,$$

where $\hat{\phi}^{n,\mathcal{I}_t^0}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_t^0 \setminus n}$-approximation $X_{\mathcal{I}_t^0 \setminus n}$ with $k_{\mathcal{I}_t^0 \setminus n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}_t^0 \setminus n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_t^0 \setminus n}$.

Lemma 16. *We claim the following*

$$\left\langle \phi^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle,$$

where $\tilde{\phi}^{n,\mathcal{I}_t^0} = VV^\top \phi^{n,\mathcal{I}_t^0}$ where $V \in \mathbb{R}^{|\mathcal{I}_t^0 \setminus n| \times k_{\mathcal{I}_t^0 \setminus n}}$ denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}_t^0 \setminus n}]$.

Proof. By Assumption 20 there exists $\xi^{(0,t)}$ such that for any $j \in \mathcal{I}_t^0 \setminus n$

$$\mathbb{E}[Y_{j,t} \mid \mathcal{LF}, j \in \mathcal{I}_t^0 \setminus n] = \sum_{i=1}^p \xi_i^{(0,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^0 \setminus n})_{ij} \mid \mathcal{LF}, j \in \mathcal{I}_t^0 \setminus n].$$

As such,

$$\mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] = VV^T \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}],$$

which gives us

$$\left\langle \tilde{\phi}^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle = \left\langle VV^\top \phi^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle = \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}]^\top VV^\top \cdot \phi^{n,\mathcal{I}_t^0} = \left\langle \phi^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle$$

proving the desired result. ☐

Using Lemma 16, we can now lift the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency for $n \in \mathcal{I}_t^0$

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = \left\langle \hat{\phi}^{n,\mathcal{I}_t^0}, Y_{\mathcal{I}_t^0 \setminus n, t} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}_t^0}, \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n, t}] \right\rangle$$

$$= O_p \left(\sqrt{\log(p|\mathcal{I}_t^0|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max \left\{ \frac{\sqrt{|\mathcal{I}_t^0|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}_t^0| - 1}} \right\} \right] \right), \quad (111)$$

where we set $T_1 = 1$, $\tilde{w}^{(i,d)} = \tilde{\phi}^{n,\mathcal{I}_t^0}$, $\hat{w}^{(i,d)} = \hat{\phi}^{n,\mathcal{I}_t^0}$, $Y_{t,\mathcal{I}^{(d)}} = Y_{\mathcal{I}_t^0 \setminus n}$, $\mathbb{E}[Y_{t,\mathcal{I}^{(d)}}] = \mathbb{E}[Y_{\mathcal{I}_t^0 \setminus n} \mid \mathcal{LF}]$, and $\mathcal{P}_{V_{\text{pre}}} = VV^\top$. Furthermore, in the final rate we set $T_0 = p$, $N_d = |\mathcal{I}_t^0 \setminus n|$, and $r_{\text{pre}} = k_{\mathcal{I}_t^0 \setminus n}$. To conclude, we used that $|\mathcal{I}_t^0 \setminus n| = |\mathcal{I}_t^0| - 1$ and $k_{\mathcal{I}_t^0 \setminus n} \leq k$ where k is the uniform upper bound on the rank on all possible expected covariate matrices, i.e., $k = \max_{\mathcal{I} \in \{\mathcal{I}^d\} \cup \{\mathcal{I}_t^0\}} \text{rank}(\mathbb{E}[X_\mathcal{I}])$.

Non-Donor Set Baseline Consistency: Consider unit $n \notin \mathcal{I}_t^0$. Denote $X_{\mathcal{I}_t^0} = X_{:,\mathcal{I}_t^0} \in \mathbb{R}^{p \times |\mathcal{I}_t^0|}$. We know the baseline outcome admits the representation

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}_t^0}, \hat{b}_{\mathcal{I}_t^0,t} \right\rangle - \left\langle \beta^{n,\mathcal{I}_t^0}, b_{\mathcal{I}_t^0,t} \right\rangle,$$

where $\hat{\beta}^{n,\mathcal{I}_t^0}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}_t^0}$-approximation of $X_{\mathcal{I}_t^0}$ with $k_{\mathcal{I}_t^0} = \mathrm{rank}(\mathbb{E}[X_{\mathcal{I}_t^0}])$, i.e., doing PCR with parameter $k_{\mathcal{I}_t^0}$.

Lemma 17. *We have that*

$$\left\langle \beta^{n,\mathcal{I}_t^0}, b_{\mathcal{I}_t^0,t} \right\rangle = \left\langle \tilde{\beta}^{n,\mathcal{I}_t^0}, b_{\mathcal{I}_t^0,t} \right\rangle$$

with $\tilde{\beta}^{n,\mathcal{I}_t^0} = VV^\top \beta^{n,\mathcal{I}_t^0}$, *where* V *denotes the right singular vectors of* $\mathbb{E}[X_{\mathcal{I}_t^0}]$.

Proof. It suffices to prove that

$$VV^\top b_{\mathcal{I}_t^0,t} = b_{\mathcal{I}_t^0,t},$$

which is equivalent to $(b_{\mathcal{I}_t^0,t})^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}_t^0}]$. By definition, for any $j \in \mathcal{I}_t^0$ we know $b_{j,t} = \mathbb{E}[Y_{j,t}|\mathcal{LF}, j \in \mathcal{I}_t^0]$. Lastly, by Assumption 20 there exists $\xi^{(0,t)} \in \mathbb{R}^p$ such that

$$\mathbb{E}[Y_{j,t}|\mathcal{LF}, j \in \mathcal{I}_t^0] = \sum_{i=1}^{p} \xi_i^{(0,t)} \cdot \mathbb{E}[(X_{\mathcal{I}_t^0})_{ij}|\mathcal{LF}, j \in \mathcal{I}_t^0].$$

This concludes the proof. $\square$

Lemma 17 allows us to write

$$\begin{aligned}
\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} &= \left\langle \hat{\beta}^{n,\mathcal{I}_t^0}, \hat{b}_{\mathcal{I}_t^0,t} \right\rangle - \left\langle \tilde{\beta}^{n,\mathcal{I}_t^0}, b_{\mathcal{I}_t^0,t} \right\rangle \\
&= \underbrace{\langle \tilde{\beta}^{n,\mathcal{I}_t^0}, \eta_{\mathcal{I}_t^0} \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n,\mathcal{I}_t^0}, \eta_{\mathcal{I}_t^0} \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n,\mathcal{I}_t^0}, b_{\mathcal{I}_t^0} \rangle}_{\text{Term 1c}},
\end{aligned}$$

where $\eta_{\mathcal{I}_t^0} = \hat{b}_{\mathcal{I}_t^0,t} - b_{\mathcal{I}_t^0,t}$ and $\Delta_{n,\mathcal{I}_t^0} = \hat{\beta}^{n,\mathcal{I}_t^0} - \tilde{\beta}^{n,\mathcal{I}_t^0}$. Using the previously referenced argument in Section C.4 for any *Non-Donor Set Component* and applying the appropriate version of Lemmas 6, 7, and 8 allows us to claim for $n \notin \mathcal{I}_t^0$

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = O_p\left(\sqrt{k \log(p|\mathcal{I}_t^0|)} \left(\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{ \frac{\sqrt{|\mathcal{I}_t^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_t^0|-1}}, \frac{1}{\sqrt{p}} \right\} \right) \right).$$

Baseline Consistency: The donor and non-donor cases together imply that for any $n \in [N]$

$$\hat{b}_{n,t} - b_{n,t} \mid \mathcal{LF} = O_p\left(\sqrt{\log(p|\mathcal{I}_t^0|)} \left(\frac{k^{5/4}}{p^{1/4}} + k^{5/2} \max\left\{ \frac{\sqrt{|\mathcal{I}_t^0|}}{p^{3/2}}, \frac{1}{\sqrt{|\mathcal{I}_t^0|-1}}, \frac{1}{\sqrt{p}} \right\} \right) \right). \quad (112)$$

2. Verifying Terminal Blip Consistency:

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}^d$. Denote $X_{\mathcal{I}^d \setminus n} = X_{:,\mathcal{I}^d \setminus n} \in \mathbb{R}^{p \times |\mathcal{I}^d \setminus n|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = \underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle - \left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \mid \mathcal{LF}] \right\rangle}_{\text{Term 1}}$$
$$+ \underbrace{\left\langle \phi^{n,\mathcal{I}^d \setminus n}, b_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d \setminus n}, \hat{b}_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle}_{\text{Term 2}},$$

where $Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} = [(Y_{j,t^*_j})_{j \in \mathcal{I}^d \setminus n}]^\top$ and $\hat{\phi}^{n,\mathcal{I}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}^d \setminus n}$-approximation of $X_{\mathcal{I}^d \setminus n}$ with $k_{\mathcal{I}^d \setminus n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}^d \setminus n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}^d \setminus n}$.[21]

Bounding Term 1: This argument is nearly identical to that for *Donor Set Baseline Consistency.*

Lemma 18. *We have that*

$$\left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}}] \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}}] \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}^d} = VV^\top \phi^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d \setminus n}]$.

Proof. It would suffice to prove that

$$VV^\top \mathbb{E}[Y_{\mathcal{I}^d \setminus n}] = \mathbb{E}[Y_{\mathcal{I}^d \setminus n}],$$

which is equivalent to $\mathbb{E}[Y_{\mathcal{I}^d \setminus n}]^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d \setminus n}]$. By Assumption 20 there exists $\xi^{(d,0)}$ such that for any $j \in \mathcal{I}^d \setminus n$

$$\mathbb{E}[Y_{j,t^*_j} \mid \mathcal{LF}, j \in \mathcal{I}^d \setminus n] = \sum_{i=1}^{p} \xi_i^{(d,0)} \cdot \mathbb{E}[(X_{\mathcal{I}^d \setminus n})_{ij} \mid \mathcal{LF}, j \in \mathcal{I}^d \setminus n].$$

This concludes the proof. $\square$

Using Lemma 18, we can once again use the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$\text{Term 1} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}}] \right\rangle \qquad (113)$$
$$= O_p \left(\sqrt{\log(p|\mathcal{I}^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max \left\{ \frac{\sqrt{|\mathcal{I}^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}^d| - 1}} \right\} \right] \right).$$

Bounding Term 2:

Lemma 19. *We have*

$$\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}} \right\rangle$$

---

[21]The vectorized baseline term is defined similarly to outcome as shown above.

with $\tilde{\phi}^{n,\mathcal{I}^d} = VV^\top \phi^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$.

Proof. It would suffice to prove that

$$VV^\top b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}} = b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}},$$

which is equivalent to $(b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}})^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$. Applying the third conclusion of Assumption 20 with $t = 0$ we know for any $j \in \mathcal{I}^d \setminus n$

$$b_{j,t^*_j} = \mathbb{E}\left[Y^{(\bar{0}^{t^*_j})}_{j,t^*_j} \big| \mathcal{LF}, j \in \mathcal{I}^d\right] = \sum_{i=1}^p \alpha_i^{(0,0)} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij} \mid \mathcal{LF} j \in \mathcal{I}^d \setminus n].$$

This concludes the proof.

$\square$

Using Lemma 19 we can write

$$\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle$$

Next we negate the RHS and decompose as follows:[22]

$$\left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle$$
$$= \underbrace{\left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d\setminus n}\right\rangle}_{\text{Term 1a}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d\setminus n}\right\rangle}_{\text{Term 1b}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}^d\setminus n} = \hat{b}_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}} - b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}$ and $\Delta_{n,\mathcal{I}^d} = \hat{\phi}^{n,\mathcal{I}^d} - \tilde{\phi}^{n,\mathcal{I}^d}$. Using the previously referenced argument by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 112 for Terms 1a, 1b, and 1c respectively allows to claim

$$\text{Term 2} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n, t^*_{\mathcal{I}^d}}\right\rangle \tag{114}$$
$$= O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})}\left[\frac{k^{7/4}}{p^{1/4}} + k^3 \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}\right\}\right]\right),$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}^0_T|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}^0_1|, |\mathcal{I}^d|\}$. To be precise the both collection of donor sets above should include $(\mathcal{I}^0_t)_{t\in[T]}$, but note that $\mathcal{I}^0_1 \subset \cdots \subset \mathcal{I}^0_T$.

Combining Term 1 and 2 rates, we find for any $n \in \mathcal{I}^d$

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^{7/4}}{p^{1/4}} + k^3 \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right), \tag{115}$$

---

[22] The negation is used primarily for convenience sake as it makes no difference in the final rate.

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|\}$.

Non-Donor Set Consistency: Consider unit $n \notin \mathcal{I}^d$. Denote $X_{\mathcal{I}^d} = X_{\cdot, \mathcal{I}^d} \in \mathbb{R}^{p \times |\mathcal{I}^d|}$. We know the baseline outcome admits the representation

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n, \mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d, 0}(d) \right\rangle - \left\langle \beta^{n, \mathcal{I}^d}, \gamma_{\mathcal{I}_T^d, 0}(d) \right\rangle,$$

where $\hat{\beta}^{n, \mathcal{I}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}^d}$-approximation $X_{\mathcal{I}^d}$ with $k_{\mathcal{I}^d} = \text{rank}(\mathbb{E}[X_{\mathcal{I}^d}])$, i.e., doing PCR with parameter $k_{\mathcal{I}^d}$.

We use an essentially identical argument to that established in *Non-Donor Set Baseline Consistency.*

Lemma 20. *We have that*

$$\left\langle \beta^{n, \mathcal{I}^d}, \gamma_{\mathcal{I}^d, 0}(d) \right\rangle = \left\langle \tilde{\beta}^{n, \mathcal{I}^d}, \gamma_{\mathcal{I}^d, 0}(d) \right\rangle$$

with $\tilde{\beta}^{n, \mathcal{I}^d} = V V^\top \beta^{n, \mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d}]$.

Proof. It would suffice to prove that

$$V V^\top \gamma_{\mathcal{I}^d, 0}(d) = \gamma_{\mathcal{I}^d, 0}(d),$$

which is equivalent to $\gamma_{\mathcal{I}^d, 0}(d)^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d}]$. To that end, recall for any $j \in \mathcal{I}^d$

$$\gamma_{j,0}(d) = \langle \psi_j^0, w_d - w_{\tilde{0}} \rangle$$
$$= \mathbb{E}\left[Y_{j, t_j^*}^{(\overline{D_{j, t_j^*}})} \right] - \mathbb{E}\left[Y_{j, t_j^*}^{(\overline{D_{j, t_j^* - 1}}, \tilde{0}_j^{t_j^*})} \right]$$
$$= \mathbb{E}\left[Y_{j, t_j^*} - Y_{j, t_j^* + t}^{(\bar{\tilde{0}}^{t_j^*})} \mid j \in \mathcal{I}^d \right]$$
$$\sum_{i=1}^p (\xi_i^{(d,0)} - \alpha_i^{(0,t)}) \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij} \mid \mathcal{LF}, j \in \mathcal{I}^d].$$

The first two equalities follow by the definition of blips, the third follows from $\overline{D_{j, t_j^*}} = (\tilde{0}, \ldots, \tilde{0}, d)$ where d occurs in the t_j^* index. The last equality is due to the second and third conclusions of Assumption 20 being applied to each term respectively. $\square$

Lemma 20 allows us to write

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n, \mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d, 0}(d) \right\rangle - \left\langle \beta^{n, \mathcal{I}^d}, \gamma_{\mathcal{I}^d, 0}(d) \right\rangle$$
$$= \underbrace{\langle \tilde{\beta}^{n, \mathcal{I}^d}, \eta_{\mathcal{I}^d}(d) \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n, \mathcal{I}^d}, \eta_{\mathcal{I}^d}(d) \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n, \mathcal{I}^d}, \gamma_{\mathcal{I}^d, 0}(d) \rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}^d}(d) = \hat{\gamma}_{\mathcal{I}^d, 0}(d) - \gamma_{\mathcal{I}^d, 0}(d)$ and $\Delta_{n, \mathcal{I}^d} = \hat{\beta}^{n, \mathcal{I}^d} - \tilde{\beta}^{n, \mathcal{I}^d}$. Using the previously referenced

argument and applying the appropriate version of Lemmas 6, 7, and 8 allows us to claim for $n \notin \mathcal{I}_T^d$

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^{9/4}}{p^{1/4}} + k^{3/2}\max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{7/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right), \quad (116)$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|\}$.

Terminal Blip Consistency: The above two sections allows us to conclude that for any $n \in [N]$

$$\hat{\gamma}_{n,0}(d) - \gamma_{n,0}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^{9/4}}{p^{1/4}} + k^{7/2}\max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right), \quad (117)$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|\}$.

3. Verifying Non-Terminal Blip Consistency:

For any unit $n \in [N]$, treatment $d \in [A]$, and $t \in [1, \ldots, T-1]$, consider the statement $P_{d,n}(t)$:

$$\hat{\gamma}_{n,t}(d) - \gamma_{n,t}(d) \mid \mathcal{LF}$$
$$= O_p\left(t\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^t}{p^{1/4}} + k^t \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{F} = \{|\mathcal{I}^d|, |\mathcal{I}_T^0|, |\mathcal{I}_1^0|, (\mathcal{I}^{D_{n,t_n^*+q}})_{n\in[N],q\in[1,\ldots,t]}\}$ with $\pi_{\mathcal{I}} = \max \mathcal{F}, \alpha_{\mathcal{I}} = \min \mathcal{F}$.

We proceed by strong induction.

To that end, consider the base case $t = 1$, i.e., proving $P_{d,n}(1)$:

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}^d$. Denote $X_{\mathcal{I}^d \setminus n} = X_{:,\mathcal{I}^d \setminus n} \in \mathbb{R}^{p \times |\mathcal{I}^d \setminus n|}$. We know the blip admits the representation

$$\hat{\gamma}_{n,1}(d) - \gamma_{n,1}(d) \mid \mathcal{LF} = \underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1}\right\rangle - \left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1} \mid \mathcal{LF}]\right\rangle}_{\text{Term 1}}$$
$$+ \underbrace{\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1} \mid \mathcal{LF}\right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1}\right\rangle}_{\text{Term 2}}$$
$$+ \underbrace{\left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1}) \mid \mathcal{LF}\right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1})\right\rangle}_{\text{Term 3}}.$$

where $\gamma_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+1}) = [(\gamma_{j,0}(D_{j,t_j^*+1}))_{j\in\mathcal{I}^d \setminus n}]^\top$ and $\hat{\phi}^{n,\mathcal{I}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}^d \setminus n}$-approximation of $X_{\mathcal{I}^d \setminus n}$ with $k_{\mathcal{I}^d \setminus n} = \text{rank}(\mathbb{E}[X_{\mathcal{I}^d \setminus n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}^d \setminus n}$.

Bounding Term 1: We prove a similar row space result.

Lemma 21. *We have for any $t \in [T-1]$*

$$\left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+t}]\right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d \setminus n, t_{\mathcal{I}^d}^*+t}]\right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}^d} = VV^\top \phi^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$.

Proof. It would suffice to prove that

$$VV^\top \mathbb{E}[Y_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t}] = \mathbb{E}[Y_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t}],$$

which is equivalent to $\mathbb{E}[Y_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t}]^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$. By Assumption 20 there exists $\xi^{(d,t)}$ such that for any $j \in \mathcal{I}^d \setminus n$

$$\mathbb{E}[Y_{j,t^*_j+t}|\mathcal{LF}, j \in \mathcal{I}^d \setminus n] = \sum_{i=1}^{p} \xi_i^{(d,t)} \cdot \mathbb{E}[(X_{\mathcal{I}^d\setminus n})_{ij}|\mathcal{LF}, j \in \mathcal{I}^d \setminus n].$$

This concludes the proof. $\qquad\qquad\square$

Using Lemma 21 for $t = 1$, we use the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$\text{Term 1} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1} \right\rangle - \left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1}] \right\rangle \qquad (118)$$
$$= O_p\left(\sqrt{\log(p|\mathcal{I}^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{ \frac{\sqrt{|\mathcal{I}^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}^d|-1}} \right\} \right] \right).$$

Bounding Term 2:

Lemma 22. *We have for any $t \in [T-1]$*

$$\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t} \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t} \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}^d} = VV^\top \phi^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$.

Proof. It would suffice to prove that

$$VV^\top b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t} = b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t},$$

which is equivalent to $(b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+t})^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d\setminus n}]$. Applying the third conclusion of Assumption 20 we know for any $j \in \mathcal{I}^d \setminus n$

$$b_{j,t^*_j+t} = \mathbb{E}\left[Y_{j,t^*_j+t}^{(\bar{0}^{t^*_j+t})} \Big| \mathcal{LF}, j \in \mathcal{I}^d \right] = \sum_{i=1}^{p} \alpha_i^{(0,t)} \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij} \mid \mathcal{LF} j \in \mathcal{I}^d \setminus n].$$

This concludes the proof.

$\qquad\qquad\square$

Using Lemma 22 for $t = 1$ we can write

$$\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1} \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\setminus n,t^*_{\mathcal{I}^d}+1} \right\rangle$$

Next we negate the RHS and decompose as follows:[23]

$$\left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} \right\rangle$$
$$= \underbrace{\left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d\backslash n} \right\rangle}_{\text{Term 1a}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d\backslash n} \right\rangle}_{\text{Term 1b}} + \underbrace{\left\langle \Delta_{n,\mathcal{I}^d}, b_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} \right\rangle}_{\text{Term 1c}},$$

where $\eta_{\mathcal{I}^d\backslash n} = \hat{b}_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} - b_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1}$ and $\Delta_{n,\mathcal{I}^d} = \hat{\phi}^{n,\mathcal{I}^d} - \tilde{\phi}^{n,\mathcal{I}^d}$. Using the previously referenced argument by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 112 for Terms 1a, 1b, and 1c respectively allows to claim

$$\text{Term 2} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+1} \right\rangle \tag{119}$$
$$= O_p\left(\sqrt{\log(p\pi_{\mathcal{I}}|)} \left[\frac{k^{7/4}}{p^{1/4}} + k^3 \max\left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}} \right\} \right] \right),$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}^0_T|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}^0_1|, |\mathcal{I}^d|\}$.

Bounding Term 3:

Lemma 23. *We have for any $t \in [T-1]$ and $\ell < t$*

$$\left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+t-\ell}) \right\rangle = \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+t-\ell}) \right\rangle$$

with $\tilde{\phi}^{n,\mathcal{I}^d} = VV^\top \phi^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d\backslash n}]$.

Proof. It would suffice to prove that

$$VV^\top \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+t-\ell}) = \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+t-\ell}),$$

which is equivalent to $(\gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t^*_{\mathcal{I}^d}+t-\ell}))^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d\backslash n}]$. Notice that for any $j \in \mathcal{I}^d \backslash n$

$$\gamma_{j,\ell}(D_{j,t^*_j+t-\ell}) = \langle \psi^\ell_j, w_{D_{j,t^*_j+t-\ell}} - w_{\tilde{0}} \rangle$$
$$= \mathbb{E}\left[Y^{(\overline{D_{j,t^*_j+t-\ell}}, \tilde{0}^{t^*_j+t-\ell+1})}_{j,t^*_j+t} \right] - \mathbb{E}\left[Y^{(\overline{D_{j,t^*_j+t-\ell-1}}, \tilde{0}^{t^*_j+t-\ell})}_{j,t^*_j+t} \right]$$
$$= \sum_{i=1}^p (\alpha^{(0,t,\ell)} - \alpha^{(0,t,\ell)'}) \cdot \mathbb{E}[(X_{\mathcal{I}^d\backslash n})_{ij}], j \in \mathcal{I}^d \backslash n,$$

where we use the definition of blips in the first two equalities and both conclusions of Lemma 15 yield the last equality. □

---

[23]The negation is used primarily for convenience sake as it makes no difference in the final rate.

Using Lemma 23 for $t = 1$ and $\ell = 0$ we can write

$$
\left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}+1}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d_{T-1}}, \hat{\gamma}_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}+1}) \right\rangle
$$
$$
= \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}+1}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d_{T-1}}, \hat{\gamma}_{\mathcal{I}^d \setminus n, 0}(D_{\mathcal{I}^d \setminus n, t^*_{\mathcal{I}^d}+1}) \right\rangle
$$

At this point, we can follow the earlier approach for Term 2 by negating, using the same decomposition, and applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 117 to write

$$
\text{Term 3} = O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{11/4}}{p^{1/4}} + k^4 \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}^0_T|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t^*_n+1}}|)_{n \in [N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}^0_1|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t^*_n+1}}|)_{n \in [N]}\}$. Notice that this dominates the rates for Terms 1 and 2 and as such we also have for any $n \in \mathcal{I}^d$

$$
\hat{\gamma}_{n,1}(d) - \gamma_{n,1}(d) \mid \mathcal{LF} = O_p \left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{11/4}}{p^{1/4}} + k^4 \max \left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \quad (120)
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}^0_T|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t^*_n+1}}|)_{n \in [N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}^0_1|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t^*_n+1}}|)_{n \in [N]}\}$.

Non-Donor Set Consistency: Consider any $t \in [T-1]$ and unit $n \notin \mathcal{I}^d$. Denote $X_{\mathcal{I}^d} = X_{:,\mathcal{I}^d} \in \mathbb{R}^{p \times |\mathcal{I}^d|}$. We know the blip effect admits the representation

$$
\hat{\gamma}_{n,t}(d) - \gamma_{n,t}(d) \mid \mathcal{LF} = \left\langle \hat{\beta}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d,t}(d) \right\rangle - \left\langle \beta^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d,t}(d) \right\rangle,
$$

where $\hat{\beta}^{n,\mathcal{I}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}^d}$-approximation $X_{\mathcal{I}^d}$ with $k_{\mathcal{I}^d} = \text{rank}(\mathbb{E}[X_{\mathcal{I}^d}])$, i.e., doing PCR with parameter $k_{\mathcal{I}^d}$.

We use an identical argument to that established in *Baseline Consistency – Non-Donor Set*.

Lemma 24. *We have that*

$$
\left\langle \beta^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d,t}(d) \right\rangle = \left\langle \tilde{\beta}^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d,t}(d) \right\rangle
$$

with $\tilde{\beta}^{n,\mathcal{I}^d} = VV^\top \beta^{n,\mathcal{I}^d}$, where V denotes the right singular vectors of $\mathbb{E}[X_{\mathcal{I}^d}]$.

Proof. It would suffice to prove that

$$
VV^\top \gamma_{\mathcal{I}^d,t}(d) = \gamma_{\mathcal{I}^d,t}(d),
$$

which is equivalent to $\gamma_{\mathcal{I}^d,t}(d)^\top$ being in the rowspace of $\mathbb{E}[X_{\mathcal{I}^d}]$. To that end, recall for any $j \in \mathcal{I}^d$,

$$
\begin{aligned}
\gamma_{j,t}(d) &= \langle \psi_j^t, w_d - w_{\tilde{0}} \rangle \\
&= \mathbb{E}\left[Y_{j,t_j^*+t}^{(\overline{D_{j,t_j^*}},\tilde{0}^{t_j^*+1})} \right] - \mathbb{E}\left[Y_{j,t_j^*+t}^{(\overline{D_{j,t_j^*-1}},\tilde{0}^{t_j^*})} \right] \\
&= \mathbb{E}\left[Y_{j,t_j^*+t} - Y_{j,t_j^*+t}^{(\bar{0}^{t_j^*+t})} \big| j \in \mathcal{I}^d \right] \\
&\sum_{i=1}^p (\xi_i^{(d,t)} - \alpha_i^{(0,t)}) \cdot \mathbb{E}[(X_{\mathcal{I}^d})_{ij} | \mathcal{LF}, j \in \mathcal{I}^d].
\end{aligned}
$$

The first two equalities follow by the definition of blips, the third follows from $\overline{D_{j,t_j^*+t}} = (\tilde{0}, \ldots, \tilde{0}, d, \tilde{0}, \ldots, \tilde{0})$ where d occurs in the t_j^* index. The last equality is due to the second and third conclusions of Assumption 20 being applied to each term respectively. $\qquad\square$

Using the above framework and Lemma 24 with $t = 1$ allows us to write

$$
\begin{aligned}
\hat{\gamma}_{n,1}(d) - \gamma_{n,1}(d) \mid \mathcal{LF} &= \left\langle \hat{\beta}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d,1}(d) \right\rangle - \left\langle \tilde{\beta}^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d,1}(d) \right\rangle \\
&= \underbrace{\langle \tilde{\beta}^{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d}(d) \rangle}_{\text{Term 1a}} + \underbrace{\langle \Delta_{n,\mathcal{I}^d}, \eta_{\mathcal{I}^d}(d) \rangle}_{\text{Term 1b}} + \underbrace{\langle \Delta_{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d,1}(d) \rangle}_{\text{Term 1c}},
\end{aligned}
$$

where $\eta_{\mathcal{I}^d}(d) = \hat{\gamma}_{\mathcal{I}^d,1}(d) - \gamma_{\mathcal{I}^d,1}(d)$ and $\Delta_{n,\mathcal{I}^d} = \hat{\beta}^{n,\mathcal{I}^d} - \tilde{\beta}^{n,\mathcal{I}^d}$. Using the previously referenced argument by applying the appropriate version of Lemmas 6, 7, and 8 allows to claim for $n \notin \mathcal{I}^d$

$$
\hat{\gamma}_{n,1}(d) - \gamma_{n,1}(d) \mid \mathcal{LF} = O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{13/4}}{p^{1/4}} + k^{9/2} \max\left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right), \quad (121)
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+1}}|)_{n\in[N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+1}}|)_{n\in[N]}\}$. Combining equations 120 and 121 yields the base case.

Inductive Step: We assume $P_{d,n}(\ell)$ for $\ell \in [1, \ldots, t-1]$ and prove $P_{d,n}(t)$.

For any $d \in [A]$:

Donor Set Consistency: Consider unit $n \in \mathcal{I}^d$. Denote $X_{\mathcal{I}^d\backslash n} = X_{:,\mathcal{I}^d\backslash n} \in \mathbb{R}^{p \times |\mathcal{I}^d\backslash n|}$. We know

the baseline outcome admits the representation

$$
\hat{\gamma}_{n,t}(d) - \gamma_{n,t}(d) \mid \mathcal{LF} = \underbrace{\left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle - \left\langle \phi^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}] \right\rangle}_{\text{Term 1}}
$$

$$
+ \underbrace{\left\langle \phi^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle}_{\text{Term 2}}
$$

$$
+ \underbrace{\left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,0}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d\backslash n,0}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}) \right\rangle}_{\text{Term 3}}
$$

$$
+ \sum_{\ell=1}^{t-1} \left(\underbrace{\left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t-\ell}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t-\ell}) \right\rangle}_{\text{Term } \ell} \right).
$$

where $\hat{\phi}^{n,\mathcal{I}^d}$ are the regression coefficients from regressing additional covariates $X_n \in \mathbb{R}^p$ on the rank $k_{\mathcal{I}^d\backslash n}$-approximation of $X_{\mathcal{I}^d\backslash n}$ with $k_{\mathcal{I}^d\backslash n} = \mathrm{rank}(\mathbb{E}[X_{\mathcal{I}^d\backslash n}])$, i.e., doing PCR with parameter $k_{\mathcal{I}^d\backslash n}$.

Bounding Term 1: We simply use Lemma 21 which holds for any $t \in [T-1]$ to leverage the proof technique in Agarwal et al. (2020b) Theorem 2 (Appendix C) to show consistency of

$$
\text{Term 1} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, Y_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \mathbb{E}[Y_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}] \right\rangle \tag{122}
$$
$$
= O_p\left(\sqrt{\log(p|\mathcal{I}^d|)} \left[\frac{k^{3/4}}{p^{1/4}} + k^2 \max\left\{ \frac{\sqrt{|\mathcal{I}^d|}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{|\mathcal{I}^d|-1}} \right\} \right] \right).
$$

Bounding Term 2: Using the previously referenced argument for Term 2 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 112 and Lemma 22 we know

$$
\text{Term 2} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{b}_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, b_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t} \right\rangle \tag{123}
$$
$$
= O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})} \left[\frac{k^{7/4}}{p^{1/4}} + k^3 \max\left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{p}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}} \right\} \right] \right),
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}^0_T|, |\mathcal{I}^d|\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}^0_1|, |\mathcal{I}^d|\}$.

Bounding Term 3: Using the previously referenced argument for Term 3 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside Equation 117 for any $d \in \{D_{n,t^*_n+t}\}_{n\in[N]}$ and Lemma 23 with $\ell = 0$ to write

$$
\text{Term 3} = \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d\backslash n,0}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}) \right\rangle - \left\langle \tilde{\phi}^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,0}(D_{\mathcal{I}^d\backslash n, t^*_{\mathcal{I}^d}+t}) \right\rangle \tag{124}
$$
$$
= O_p\left(\sqrt{\log(p\pi_{\mathcal{I}})} \left(\frac{k^{11/4}}{p^{1/4}} + k^4 \max\left\{ \frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),
$$

where $\pi_{\mathcal{I}} = \max\{|\mathcal{I}_T^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+t}}|)_{n\in[N]}\}$ and $\alpha_{\mathcal{I}} = \min\{|\mathcal{I}_1^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+t}}|)_{n\in[N]}\}$.

Bounding Term ℓ for $\ell \in [1, \ldots, t-1]$: For any such ℓ, we use an argument similar to Term 3 in the base case by applying the appropriate version of Lemma 6, 7, and 8 alongside the inductive hypothesis $P_{d,n}(\ell)$ for all $d \in \{D_{n,t_n^*+t-\ell}\}_{n\in[N]}$ and Lemma 23 to write

$$\text{Term } \ell = \left\langle \phi^{n,\mathcal{I}^d}, \gamma_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t_{\mathcal{I}^d}^*+t-\ell}) \right\rangle - \left\langle \hat{\phi}^{n,\mathcal{I}^d}, \hat{\gamma}_{\mathcal{I}^d\backslash n,\ell}(D_{\mathcal{I}^d\backslash n,t_{\mathcal{I}^d}^*+t-\ell}) \right\rangle \tag{125}$$
$$= O_p\left(\ell\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^\ell}{p^{1/4}} + k^\ell \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_1^T|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+q}}|)_{n\in[N],q\in[1,\ldots,\ell]}, (|\mathcal{I}^{D_{n,t_n^*+t-\ell}}|)_{n\in[N]}\}$ with $\pi_{\mathcal{I}} = \max\mathcal{F}, \alpha_{\mathcal{I}} = \min\mathcal{F}$.

Note that Terms 1-2 are dominated by the summation, as such it suffices to analyze the latter and Term-3. To that end for the summation,

$$\sum_{\ell=1}^{t-1} \text{Term } \ell = O_p\left(\sum_{\ell=1}^{t-1}\ell\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^\ell}{p^{1/4}} + k^\ell \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_1^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+q}}|)_{n\in[N],q\in[1,\ldots,t-1]}\}$ with $\pi_{\mathcal{I}} = \max\mathcal{F}, \alpha_{\mathcal{I}} = \min\mathcal{F}$. Notice we bounded the smaller donor set cardinalates by the largest one, i.e., when $\ell = t-1$. We analyze the time dependent terms and denote

$$C := \sqrt{\log(p\pi_{\mathcal{I}})}, \quad C' := \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}.$$

Upon substitution we have

$$C\sum_{m=1}^{t-1} m\left(\frac{k^m}{p^{1/4}} + C'k^m\right) = C\left(\frac{1}{p^{1/4}} + C'\right)\sum_{m=1}^{t-1} mk^m.$$

We apply the geometric sum derivative trick for $k \geq 1$

$$\sum_{m=1}^{M} mk^m = \frac{k(1-(M+1)k^M+Mk^{M+1})}{(1-k)^2} = \Theta(Mk^{M+1})$$

Taking $M = t-1$, we conclude

$$\sum_{\ell=1}^{t-1} \text{Term } \ell = O_p\left(t\sqrt{\log(p\pi_{\mathcal{I}})}\left(\frac{k^t}{p^{1/4}} + k^t \max\left\{\frac{\sqrt{\pi_{\mathcal{I}}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_{\mathcal{I}}-1}}, \frac{1}{\sqrt{p}}\right\}\right)\right),$$

Combining this with Term-3 yields for any $n \in \mathcal{I}^d$

$$\hat{\gamma}_{n,t}(d) - \gamma_{n,t}(d) \mid \mathcal{LF} \tag{126}$$
$$= O_p \left(t\sqrt{\log(p\pi_\mathcal{I})} \left(\frac{k^t}{p^{1/4}} + k^t \max\left\{ \frac{\sqrt{\pi_\mathcal{I}}}{p^{3/2}}, \frac{1}{\sqrt{\alpha_\mathcal{I} - 1}}, \frac{1}{\sqrt{p}} \right\} \right) \right),$$

where $\mathcal{F} = \{|\mathcal{I}_T^0|, |\mathcal{I}_1^0|, |\mathcal{I}^d|, (|\mathcal{I}^{D_{n,t_n^*+q}}|)_{n\in[N], q\in[1,...,t]}\}$ with $\pi_\mathcal{I} = \max\mathcal{F}, \alpha_\mathcal{I} = \min\mathcal{F}$.

Non-Donor Set Consistency: Applying the *Non-Donor Set Consistency* argument written for the Base Case for general t, specifically Lemma 24 for any $t \in [T-1]$, proves $P_{d,n}(t)$.

4. Verifying Target Causal Parameter Consistency: For any unit $n \in [N]$ and $\bar{d}^T \in [A]^T$ we recall the SBE-PCR estimator and the corresponding causal estimand.

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)} \right] = \sum_{t=1}^{T} \hat{\gamma}_{n,T-t}(d_t) + \hat{b}_{n,T} \quad \text{and} \quad \mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF} \right] = \sum_{t=1}^{T} \gamma_{n,T-t}(d_t) + b_{n,T} \mid \mathcal{LF}.$$

The difference is exactly

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)} \right] - \mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF} \right] = \left(\hat{b}_{n,T} - b_{n,T} \mid \mathcal{LF} \right) + \sum_{t=1}^{T} \left(\hat{\gamma}_{n,T-t}(d_t) - \gamma_{n,T-t}(d_t) \mid \mathcal{LF} \right)$$

We apply the known bound for each term, specifically Equation 112, Equation 117 with $d = d_T$, and $P_{d_t,n}(T-t)$ for every $t \in [T-1]$. Once again we encounter the same geometric sum, which gives the desired result upon noting that the baseline rate is dominated by that of the sum.

D.5 Proof of Theorem 6

We recall that for any unit $n \in [N]$ and $\bar{d}^T \in [A]^T$

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF} \right] = \sum_{t=1}^{T} \gamma_{n,T-t}(d_t) + b_{n,T} \mid \mathcal{LF} = \sum_{t=1}^{T} \langle \psi_n^{T-t}, w_{d_t} - w_{0_t} \rangle + b_{n,T} \mid \mathcal{LF}.$$

Given Assumption 21 we know that $\psi_n^{q+i} = 0$ for all $i \in [T - q - 1]$. As such,

$$\mathbb{E}\left[Y_{n,t}^{(\bar{d}^T)} \mid \mathcal{LF} \right] = \sum_{t=T-q}^{T} \langle \psi_n^{T-t}, w_{d_t} - w_{0_t} \rangle + b_{n,T} \mid \mathcal{LF} = \sum_{t=T-q}^{T} \gamma_{n,T-t}(d_t) + b_{n,T} \mid \mathcal{LF}$$

We modify the SBE-PCR estimator accordingly

$$\hat{\mathbb{E}}\left[Y_{n,T}^{(\bar{d}^T)} \mid \mathcal{LF} \right] := \sum_{t=T-q}^{T} \hat{\gamma}_{n,T-t}(d_t) + \hat{b}_{n,T}.$$

Applying the analysis from the proof of Theorem 7 yields the desired result.