Causal Mediation Analysis with High-Dimensional Surrogates: Identification, Estimation and Inference under Latent Factor Models

Himani Yadav, Shi Bo, AmirEmad Ghassami[†], Debarghya Mukherjee[†]

Abstract

Causal mediation analysis is an important tool for understanding how a treatment affects an outcome through intermediate variables referred to as mediators. However, in many applications, these mediators and pre-treatment covariates are not directly observed and are instead measured through noisy, high-dimensional surrogates. Mediation analysis when such latent variables are observed only through these surrogates remains less explored. In this paper, we propose Factor-Adjusted Regression for Mediation Analysis (FARMA), a factor-based framework for estimating direct and indirect effects when either one or both of the mediators and relevant pre-treatment covariates are latent. FARMA uses the surrogate measurements to recover information about the underlying latent variables. Under a latent factor structure, we show that the direct and indirect effects are approximately identified, with an approximation error that vanishes as the surrogate dimensions grow. We further show that our estimator is $\sqrt{n}$-consistent and asymptotically normal. We combine the theory with extensive simulation studies and apply it to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, where we investigate whether the effect of depressive symptoms on cognitive impairment is mediated through the latent process underlying high-dimensional DNA methylation measurements.

Keywords— Mediation Analysis; Latent Factor Models; Ultra-High-Dimensionality; Identification

[†] Corresponding authors.

Contents

1 Introduction **3**

2 Model Setting **4**

3 Identification **6**

4 FARMA Algorithm for Mediation Analysis **8**

5 Theoretical Analysis **10**

6 Simulation Studies **12**

7 Real Data Experiments **15**

8 Conclusion and Future Work **17**

A Appendix **21**
A.1 Notation. 21
A.2 Supporting Results 22

B Proof of Theorem 3.1: Identification **24**
B.1 Proof of Corollary 3.2 38

C Asymptotic Analysis **38**
C.1 Supporting Results 39
C.2 Proof of Theorem 5.2 65

D Auxiliary Material **92**
D.1 Technical Lemmas 92
D.2 Calculation of Effects 93
D.3 More on Simulation Study and Sensitivity Analysis 94
D.4 Covariates 101

E Comparison to Related Literature **101**
E.1 Comparison to Zhang et al. (2026a) 101

1 Introduction

Causal mediation analysis provides a principled framework for understanding the mechanisms through which a treatment affects an outcome (Baron and Kenny, 1986; Pearl, 2001; Imai et al., 2010; VanderWeele and Vansteelandt, 2014). It moves beyond testing whether an overall effect exists to ask how that effect operates by separating the total effect into a direct effect, representing the effect through other pathways, and an indirect effect, representing the effect through an intermediate variable (referred to as the *mediator*). With advances in biostatistics, epidemiologic study designs, and causal inference methods, mediation analysis has evolved considerably and is now applied across a wide range of disciplines (Valeri and VanderWeele, 2015). For example, in biomedical research, mediation analysis has become a powerful tool for identifying mechanisms of disease progression and for uncovering potential medicinal targets (MacKinnon, 2012; VanderWeele, 2015; Jung, 2021). The statistical literature on mediation analysis has developed considerably from its early regression-based formulations. The product-of-coefficients method of Baron and Kenny (1986) offered an intuitive way to study mediated pathways. Later, the potential-outcomes framework of Rubin (1974) allowed mediation to be defined in formal causal terms through the natural direct and indirect effects (Robins and Greenland, 1992; Pearl, 2001), along with the conditions for their identification (Imai et al., 2010). Subsequent work developed semiparametric, weighting, and multiply robust methods for their estimation and inference (Tchetgen Tchetgen and Shpitser, 2012; VanderWeele and Vansteelandt, 2014; VanderWeele, 2015). Most of these methods consider settings in which the mediator and the covariates needed for adjustment are directly observed and can be represented using a relatively small number of variables. More recent work has considered settings with mediation with unobserved structure (Albert et al., 2016; Derkach et al., 2019; Ghassami et al., 2025) as well as high-dimensional mediators or covariates using different approaches (Zhang et al., 2016; Huang and Pan, 2016; Chén et al., 2018; Zhang et al., 2018; Zhao and Luo, 2021; Perera et al., 2022; Guo et al., 2022, 2023, 2024; Rakshit and Guo, 2024; Bo et al., 2024). Factor-model approaches have also been widely used for dimension reduction and regression with high-dimensional data (Bai and Ng, 2002; Bai, 2003; Fan et al., 2013, 2020; Fan and Liao, 2022).

However, in many real-world applications, the true mediators are not directly observed. Instead, researchers may rely on a potentially large number of surrogate measurements that provide indirect information about the underlying mediator. This creates the challenge of identifying and estimating natural direct and indirect effects using only these observed surrogates. A motivating example arises in psychiatric and neurodegenerative disorders, where mediating mechanisms involve complex molecular processes that cannot be directly quantified. For instance, oxidative stress has been implicated as a key pathway linking depression to Alzheimer's disease (AD), the leading cause of dementia (Livingston et al., 2024; Buccellato et al., 2021). The most direct way to quantify oxidative stress is by measuring reactive oxygen species (ROS), but this is difficult due to their short lifespan and high reactivity (Katerji et al., 2019). Therefore, researchers rely on more stable molecular signatures that reflect cumulative oxidative activity. Among them is DNA methylation, which is altered in response to oxidative stress and can be measured at scale (Miranda et al., 2000; Pandya et al., 2013; Black et al., 2015). While DNA methylation provides a useful surrogate to investigate the mediating role of oxidative stress, its measurement across hundreds of thousands of CpG sites using high-throughput assays results in a high-dimensional data structure where individual sites capture only partial aspects of the underlying biological process, leaving the true latent mechanism distributed across the entire profile. The key question, therefore, is how to identify and estimate mediated effects in such settings

Our Contribution. In this paper, we propose **F**actor-**A**djusted **R**egression for **M**ediation **A**nalysis (**FARMA**), a factor-based framework for causal mediation analysis when the mediator and relevant pre-treatment covariates may be latent and represented by high-dimensional surrogates. Our contributions are threefold.

Identification. We establish identification of the causal natural direct and indirect effects when both the mediator and the relevant covariate factors are latent. With a fixed number of surrogates, the target effects are generally not point identified from the observed data alone without strong restrictions on the data-generating process. We show that suitable projections of the high-dimensional surrogates yield an observed-data parameter that approaches the latent target as the number of measurements increases, that leads to the asymptotic point identification. The method does not require the true factor dimensions to be known exactly and remains valid when the working dimensions are chosen as upper bounds on the true

dimensions (Bai and Ng, 2002; Onatski, 2010; Fan and Liao, 2022).

Estimation. Building on this identification result, we propose FARMA to estimate the causal effects. FARMA constructs low-dimensional summaries of the surrogate covariates and mediators using data-driven projection directions, and then uses these summaries in the mediation regressions. The method does not require the true factor dimensions to be known exactly. It remains valid when the working dimensions are chosen as upper bounds on the true dimensions, which is helpful in practice when the exact number of factors is difficult to determine (Bai and Ng, 2002; Onatski, 2010; Fan and Liao, 2022). Under regularity and suitable growth conditions on the surrogate dimensions, we show that the resulting estimator is $\sqrt{n}$-consistent and asymptotically normal for the parameter of interest. This result provides the basis for statistical inference on the natural direct and indirect effects even when mediator are not observed directly.

Application. We apply FARMA to investigate the effect of depression on Alzheimer's disease (AD) mediated by oxidative stress, where the mediating process is not directly observed and is instead represented by high-dimensional DNA methylation measurements. We illustrate the approach using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) Study.

The paper is organized as follows. Section 2 presents the problem setting and model. Section 3 establishes identification of the target parameter. Section 4 introduces the FARMA algorithm. Section 5 gives the theoretical justification. Section 6 reports the simulation results. Section 7 presents a real-data study using the ADNI dataset, where we investigate the effect of depression on Alzheimer's disease mediated through indirect pathways. Section 8 concludes and discusses future work.

2 Model Setting

We consider a binary treatment variable $A \in \{0,1\}$ (also called *exposure* or *action* variable), and a real-valued outcome $Y \in \mathbb{R}$. For each individual, we consider mediators $\mathbf{f}_M = [f_{M_1}, \ldots, f_{M_{\bar{q}}}] \in \mathbb{R}^{\bar{q}}$ through which treatment may influence the outcome, and pre-treatment covariates $\mathbf{f}_X = [f_{X_1}, \ldots, f_{X_{\bar{p}}}] \in \mathbb{R}^{\bar{p}}$, which capture baseline characteristics that may confound the relationship among treatment, mediators and outcome. We use lowercase notation, such as f_{M_i} and f_{X_j}, to denote the individual coordinates within these factors, respectively. The treatment can influence the outcome in two ways, indirectly through the mediators, or directly without passing through them. Formally, the variables are assumed to be generated from the following structural equations:

$$\begin{aligned} \mathbf{f}_M &= (1-A)\Big(\delta_0 + \mathbf{B}_0 \mathbf{f}_X + \mathbf{u}_{XM,0}\Big) + A\Big(\delta_1 + \mathbf{B}_1 \mathbf{f}_X + \mathbf{u}_{XM,1}\Big), \\ Y &= (1-A)\Big(\alpha_0 + \beta_0^\top \mathbf{f}_X + \gamma_0^\top \mathbf{f}_M + \epsilon_0\Big) + A\Big(\alpha_1 + \beta_1^\top \mathbf{f}_X + \gamma_1^\top \mathbf{f}_M + \epsilon_1\Big), \end{aligned} \tag{2.1}$$

where ϵ_0 and ϵ_1 are centered outcome errors on the control and treated arms, $\mathbf{u}_{XM,0} \in \mathbb{R}^{\bar{q}}$ and $\mathbf{u}_{XM,1} \in \mathbb{R}^{\bar{q}}$ are the corresponding centered mediator structural noises, $\mathbf{B}_0, \mathbf{B}_1 \in \mathbb{R}^{\bar{q} \times \bar{p}}$, $\beta_0, \beta_1 \in \mathbb{R}^{\bar{p}}$, and $\gamma_0, \gamma_1 \in \mathbb{R}^{\bar{q}}$ are unknown coefficients, and $\delta_0, \delta_1 \in \mathbb{R}^{\bar{q}}$ and $\alpha_0, \alpha_1 \in \mathbb{R}$ are intercepts. Note that we allow for an $A - \mathbf{f}_X$ interaction in the generation of $\mathbf{f}_M$, and $A - \mathbf{f}_X$ and $A - \mathbf{f}_M$ interactions in the generation of the outcome Y.

In many real-world applications, $\mathbf{f}_X$ and $\mathbf{f}_M$ are *latent* (i.e., unobserved). Instead, we observe measurements $X \in \mathbb{R}^p$ and $M \in \mathbb{R}^q$ that serve as noisy surrogates for these latent variables. We focus on high-dimensional measurements with $p \gg \bar{p}$ and $q \gg \bar{q}$. We adopt the following factor model for the observed measurements:

$$X = \mathbf{\Gamma}_X \mathbf{f}_X + \mathbf{u}_X, \qquad M = \mathbf{\Gamma}_M \mathbf{f}_M + \mathbf{u}_M, \tag{2.2}$$

where $\mathbf{\Gamma}_X \in \mathbb{R}^{p \times \bar{p}}$ and $\mathbf{\Gamma}_M \in \mathbb{R}^{q \times \bar{q}}$ are unknown loading matrices, and $\mathbf{u}_X$ and $\mathbf{u}_M$ are mean-zero errors. Henceforth, we refer to $\mathbf{f}_X$ and $\mathbf{f}_M$ as latent factors. Figure 1 summarizes the causal relationships among the latent factors, treatment, and outcome specified in (2.1) with the measurement relationships in (2.2). The standard mediation model with directly observed covariates and mediators is recovered as a special case by setting $p = \bar{p}$, $q = \bar{q}$, $\mathbf{\Gamma}_X = I_p$, and $\mathbf{\Gamma}_M = I_q$, with vanishing errors. In this case, $X = \mathbf{f}_X$ and $M = \mathbf{f}_M$, so the structural equations in (2.1) can be written directly in terms of the observed variables X and M.

We first consider the setting in which $\mathbf{f}_X$ and $\mathbf{f}_M$ are directly observed, and define the causal estimands using the potential-outcomes framework (Rubin, 1974; Robins and Greenland, 1992). For each individual,

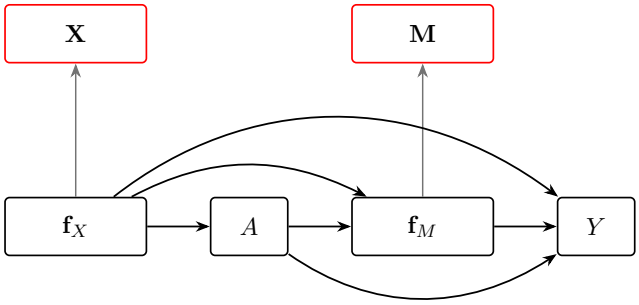


Figure 1: Causal diagram for the factor-augmented mediation model. The observed high-dimensional variables $\mathbf{X}$ and $\mathbf{M}$ are generated from the latent factors $\mathbf{f}_X$ and $\mathbf{f}_M$, respectively.

let $Y^{(a,f_M)}$ denote the potential outcome under treatment $A = a$ and mediator value $\mathbf{f}_M = f_M$, and let $\mathbf{f}_M^{(a)}$ denote the potential mediator value under treatment $A = a$, where $a \in \{0, 1\}$. The potential outcome under treatment a is then $Y^{(a)} = Y^{(a,\mathbf{f}_M^{(a)})}$. The average treatment effect (ATE) admits the following decomposition:

$$\underbrace{\mathbb{E}\left[Y^{(1)} - Y^{(0)}\right]}_{\text{Total Effect}} = \underbrace{\mathbb{E}\left[Y^{(1,\mathbf{f}_M^{(1)})} - Y^{(1,\mathbf{f}_M^{(0)})}\right]}_{\text{Indirect Effect}} + \underbrace{\mathbb{E}\left[Y^{(1,\mathbf{f}_M^{(0)})} - Y^{(0,\mathbf{f}_M^{(0)})}\right]}_{\text{Direct Effect}}. \tag{2.3}$$

The two terms on the right-hand side are the total indirect effect and the pure direct effect, respectively, in the terminology of Robins and Greenland (1992). Following Pearl (2001), we refer to them as the natural indirect effect (NIE) and the natural direct effect (NDE). In our setting, these effects are defined with respect to $\mathbf{f}_M$. The NIE measures the effect of changing the mediator from its potential value under control to its potential value under treatment, while holding treatment fixed at $A = 1$. The NDE measures the effect of changing treatment from $A = 0$ to $A = 1$, while holding the mediator at its potential value under control.

The decomposition in (2.3) depends on three counterfactual means, $\mathbb{E}[Y^{(1)}]$ ($\equiv \mathbb{E}[Y^{(1,\mathbf{f}_M^{(1)})}]$), $\mathbb{E}[Y^{(0)}]$ ($\equiv \mathbb{E}[Y^{(0,\mathbf{f}_M^{(0)})}]$), and $\mathbb{E}[Y^{(1,\mathbf{f}_M^{(0)})}]$. Under consistency, positivity and sequential ignorability, the first two can be identified using some standard covariate adjustment. $\mathbb{E}[Y^{(1,\mathbf{f}_M^{(0)})}]$ is a cross-world mean that combines the outcome under treatment with the mediator value that would arise under control. Since its identification requires additional assumptions that links the outcome and mediator across treatment regimes, we consider this as our parameter of interest:

$$\theta_0 := \mathbb{E}\left[Y^{(1,\mathbf{f}_M^{(0)})}\right]. \tag{2.4}$$

To identify θ_0, we impose the following standard assumptions.

Assumption 2.1 (Consistency). $Y^{(a,f_M)} = Y$ *if* $A = a$, $\mathbf{f}_M = f_M$; $\quad \mathbf{f}_M^{(a)} = \mathbf{f}_M$ *if* $A = a \quad \forall \; a, f_M$.

Assumption 2.2 (Positivity). *For all strata of f_X, the probability of being in the treatment and control group is positive, i.e., $0 < P(A = a \mid \mathbf{f}_X = f_X)$, and the conditional density of the mediator is positive, i.e., $0 < P(f_M \mid A = a, \mathbf{f}_X = f_X)$ for all A, f_X and f_M.*

In addition to positivity and consistency, we operate under the assumption of sequential exchangeability (Imai et al., 2010), which implies that there are no unobserved confounders for the treatment-outcome, mediator-outcome, and treatment-mediator relationships:

Assumption 2.3 (Sequential exchangeability). *Let $X_1 \perp\!\!\!\perp X_2 \mid X_3$ indicate that the random variables X_1 and X_2 are conditionally independent given the random variable X_3.*

1. $Y^{(a,f_M)} \perp\!\!\!\perp \{A, \mathbf{f}_M\} \mid \mathbf{f}_X, \forall\, a, f_M,$
2. $\mathbf{f}_M^{(a)} \perp\!\!\!\perp A \mid \mathbf{f}_X, \forall\, a,$
3. $Y^{(a,f_M)} \perp\!\!\!\perp \mathbf{f}_M^{(a')} \mid \mathbf{f}_X, \forall\, a, a', f_M.$

These assumptions concern the latent covariates $\mathbf{f}_X$ and latent mediators $\mathbf{f}_M$. Under consistency, positivity, and sequential exchangeability, the mediation formula of Imai et al. (2010) expresses θ_0 as a functional of the joint distribution of $(Y, A, \mathbf{f}_X, \mathbf{f}_M)$:

$$\theta_0 = \iint \mathbb{E}\left[Y \mid \mathbf{f}_X = f_X,\, \mathbf{f}_M = f_M,\, A = 1\right] dP_{\mathbf{f}_M \mid \mathbf{f}_X, A}(f_M \mid f_X, 0)\, dP_{\mathbf{f}_X}(f_X). \tag{2.5}$$

Here, the mediation functional combines three components: the outcome regression under treatment, the conditional distribution of the mediator factors under control, and the marginal distribution of the covariate factors. This representation identifies θ_0 from the distribution involving the latent factors. In our setting, however, $\mathbf{f}_X$ and $\mathbf{f}_M$ are unobserved, so these components cannot be estimated directly from the available measurements X and M. Under the linear structural model, (2.5) yields the closed form

$$\theta_0 = \mathbb{E}\left[Y^{(1,\mathbf{f}_M^{(0)})}\right] = \alpha_1 + \gamma_1^\top \delta_0, \tag{2.6}$$

where the covariate contribution vanishes under $\mathbb{E}[\mathbf{f}_X] = \mathbf{0}$. Building on this, we can more generally define the mediation functional for any $a, a' \in \{0, 1\}$ as $\theta_{aa'} = \mathbb{E}[Y^{(a,\mathbf{f}_M^{(a')})}]$. In particular, $\theta_{10} = \theta_0$, where the two baseline potential outcomes are as follows:

$$\theta_{00} = \mathbb{E}[Y^{(0)}] = \alpha_0 + \gamma_0^\top \delta_0, \quad \theta_{11} = \mathbb{E}[Y^{(1)}] = \alpha_1 + \gamma_1^\top \delta_1. \tag{2.7}$$

Following the decomposition in (2.3), the natural direct and indirect effects are given by

$$\theta_{\text{NDE}} = \theta_{10} - \theta_{00}, \quad \theta_{\text{NIE}} = \theta_{11} - \theta_{10}, \tag{2.8}$$

so that the total effect is $\theta_{11} - \theta_{00} = \theta_{\text{NDE}} + \theta_{\text{NIE}}$. Hence, evaluating these effects requires recovering θ_{10}, θ_{00}, and θ_{11} from the observed data. However, a central challenge that still remains is that their coefficients are defined through models involving unobserved latent factors rather than direct measurements. In Section 3, we show how suitable projections of the high-dimensional surrogates allow us to overcome this and identify the parameters from observed data under appropriate conditions as the surrogate dimensions increase.

Notations. Throughout the paper, let n denote the sample size, and let p and q denote the dimensions of the surrogate vectors X and M, respectively. For any random variable, vector, or matrix $\mathbf{X}$, let $\mathbb{E}[\mathbf{X}]$ denote its expectation and $\mathbb{E}[\mathbf{X} \mid \cdot]$ its conditional expectation. Let $P(\cdot)$ denote probability and $P(\cdot \mid \cdot)$ conditional probability. Let $\mathrm{Cov}(\mathbf{X})$ (or $\Sigma_{\mathbf{X}}$) denote its covariance matrix, and let $\mathrm{Cov}(\mathbf{X} \mid \cdot)$ represent its conditional covariance matrix. Let $\mathcal{N}(\mu, \Sigma)$ denote the multivariate Gaussian distribution with mean μ and covariance Σ. For any vector $\mathbf{x}$, let $\|\mathbf{x}\|_2$ denote its Euclidean norm. For any matrix $\mathbf{X}$, let $\mathbf{X}^\top$ denote its transpose, and let $\|\mathbf{X}\|_{\text{op}}$ and $\|\mathbf{X}\|_F$ denote its operator norm and Frobenius norm, respectively. Let $\mathbf{1}_d$ and $\mathbf{0}_d$ denote vectors of all ones and all zeros in dimension d (such as $\mathbf{1}_p$ or $\mathbf{1}_q$), and let I_d denote the $d \times d$ identity matrix. For real numbers a and b, define $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We denote the smallest and largest eigenvalues of a symmetric matrix $\mathbf{X}$ by $\lambda_{\min}(\mathbf{X})$ and $\lambda_{\max}(\mathbf{X})$, and the smallest and largest singular values of a rectangular matrix A by $\nu_{\min}(A)$ and $\nu_{\max}(A)$. For two possibly random sequences a_n and b_n, write $a_n \ll b_n$ (equivalently $b_n \gg a_n$) if $a_n = o_p(b_n)$, meaning $a_n / b_n \xrightarrow{p} 0$ whenever the ratio is well defined.

3 Identification

In this section, we show that high-dimensional surrogate measurements provide the information needed to identify the mediation functional when both the covariates and mediators of interest are latent. Recall from Section 2 that, under Assumptions 2.1–2.3, the structural models, and $\mathbb{E}[\mathbf{f}_X] = \mathbf{0}$, the causal estimand reduces to $\theta_0 = \mathbb{E}[Y^{(1,\mathbf{f}_M^{(0)})}] = \alpha_1 + \gamma_1^\top \delta_0$. When the surrogate dimensions p and q are fixed and the latent factors are unobserved, θ_0 is not generally identifiable. Since the surrogates are noisy measurements, their idiosyncratic errors do not vanish at fixed dimension, and the map from observed moments to structural coefficients is confounded by the error covariances $\Sigma_{\mathbf{u}_X}$ and $\Sigma_{\mathbf{u}_M}$.

To overcome this, we establish approximate identification results for θ_0, where the target is recovered up to an error that vanishes as $p, q \to \infty$. Since the latent factors are pervasive, information about each factor is

shared across many surrogate measurements, which allows low-dimensional projections to preserve common factor information while filtering out idiosyncratic noise. We show that, under the structural models in (2.1)–(2.2), θ_0 becomes point identifiable as the dimensions $p, q \to \infty$. We begin by stating the assumptions needed for identification of θ_0. These assumptions ensure that the latent factors and the structural and proxy errors are well-behaved.

Assumption 3.1 (Sampling, latent factors, and errors)**.** *There exist constants $0 < c < C < \infty$, independent of n, p, q, such that the following conditions hold for each $a \in \{0, 1\}$.*

(i) The sequence $\{(\mathbf{f}_{X,i}, A_i, \epsilon_{a,i}, \mathbf{u}_{X,i}, \mathbf{u}_{M,i}, \mathbf{u}_{XM,a,i})\}_{i=1}^n$ is independent and identically distributed.

(ii) For each $a \in \{0, 1\}$ and $i = 1, \ldots, n$, the errors $\epsilon_{a,i}$, $\mathbf{u}_{X,i}$, $\mathbf{u}_{M,i}$, and $\mathbf{u}_{XM,a,i}$ are mean zero and mutually independent, and $(\epsilon_{a,i}, \mathbf{u}_{X,i}, \mathbf{u}_{M,i}, \mathbf{u}_{XM,a,i}) \perp\!\!\!\perp (\mathbf{f}_{X,i}, A_i)$.

(iii) The factor dimensions $\bar{p}$ and $\bar{q}$ are fixed, and $\mathbb{E}[\mathbf{f}_X] = \mathbf{0}$. For $\mathbf{v} \in \{\mathbf{f}_X, \mathbf{f}_M\}$, $\|\mu_{\mathbf{v}|a}\|_2 \le C$.

(iv) Let $\mathbf{f} = (\mathbf{f}_X^\top, \mathbf{f}_M^\top)^\top$. The conditional covariance matrices satisfy $c \le \lambda_{\min}(\Sigma_{\mathbf{f}|A=a}) \le \lambda_{\max}(\Sigma_{\mathbf{f}|A=a}) \le C$.

(v) For $\mathbf{u} \in \{\mathbf{u}_X, \mathbf{u}_M, \mathbf{u}_{XM,a}\}$. The error covariance matrices satisfy $c \le \lambda_{\min}(\Sigma_{\mathbf{u}}) \le \lambda_{\max}(\Sigma_{\mathbf{u}}) \le C$.

Assumption 3.1 provides baseline conditions which are common in the factor model literature and ensure the data-generating process is well-behaved. It provides the independence and moment conditions that are used in the identification argument. Assumption 3.1(ii) yields the orthogonality relations between the latent factors and idiosyncratic errors, while Assumption 3.1(iii)–(v) control the conditional means and keep the relevant covariance matrices uniformly bounded and nondegenerate. Moreover, the independence from A also implies that the proxy-error covariance matrices are the same in both treatment groups.

Assumption 3.2 (Model Parameters)**.** *There exist universal constants $0 < c < C < \infty$ such that, for $a \in \{0, 1\}$, the following conditions hold.*

(i) The structural parameters are uniformly bounded, $\max\{|\alpha_a|, \|\delta_a\|_2, \|\beta_a\|_2, \|\gamma_a\|_2, \|\mathbf{B}_a\|_{\text{op}}\} \le C$.

(ii) The loading matrices satisfy

$$c \le \lambda_{\min}\left(\frac{\mathbf{\Gamma}_X^\top \mathbf{\Gamma}_X}{p}\right) \le \lambda_{\max}\left(\frac{\mathbf{\Gamma}_X^\top \mathbf{\Gamma}_X}{p}\right) \le C,$$

and

$$c \le \lambda_{\min}\left(\frac{\mathbf{\Gamma}_M^\top \mathbf{\Gamma}_M}{q}\right) \le \lambda_{\max}\left(\frac{\mathbf{\Gamma}_M^\top \mathbf{\Gamma}_M}{q}\right) \le C.$$

Assumption 3.2 (i) requires the structural parameters to remain uniformly bounded, which rules out models in which the strength of the structural relationships diverges with the dimensions of the observed surrogates and allows the approximation error to be controlled uniformly in p and q. Assumption 3.2 (ii) requires the latent factors to be pervasive, in the sense that their loadings do not become asymptotically sparse as $p, q \to \infty$. This ensures that each factor contributes a non-vanishing amount of variation to X and M. A sufficient condition for Assumption 3.2(ii) is that the rows of Γ_X and Γ_M are independently generated from common distributions with finite second moments and positive definite second-moment matrices. Then, by SLLN, $\mathbf{\Gamma}_X^\top \mathbf{\Gamma}_X / p$ and $\mathbf{\Gamma}_M^\top \mathbf{\Gamma}_M / q$ converge to fixed matrices, so that the eigenvalue bounds in Assumption 3.2(ii) hold. This condition is widely used in literature (Stock and Watson, 2002; Bai, 2003; Fan et al., 2013) and see Example 2.1 in Fan et al. (2020) for more discussion.

To exploit the high-dimensional structure, we introduce projection matrices that map the observed variables onto a lower-dimensional space while preserving information about the latent factors.

Definition 3.1 (Full-Projection Matrices)**.** *The matrices $\mathbf{W}_X \in \mathbb{R}^{p \times \tilde{p}}$ and $\mathbf{W}_M \in \mathbb{R}^{q \times \tilde{q}}$ are called full-projection matrices with $\mathbf{W}_X = \sqrt{p} V_X$ and $\mathbf{W}_M = \sqrt{q} V_M$ where $V_X \in \mathbb{R}^{p \times \tilde{p}}$ and $V_M \in \mathbb{R}^{q \times \tilde{q}}$ have minimum and maximum bounded eigenvalues, i.e., $c \le \lambda_{\min}(V_X^\top V_X) \le \lambda_{\max}(V_X^\top V_X) \le C$ and $c \le \lambda_{\min}(V_M^\top V_M) \le \lambda_{\max}(V_M^\top V_M) \le C$ for some constant C, if the resulting matrices*

$$\mathbf{H}_X = \tfrac{1}{p} \mathbf{W}_X^\top \Gamma_X; \quad \mathbf{H}_M = \tfrac{1}{q} \mathbf{W}_M^\top \Gamma_M \tag{3.1}$$

have full column rank.

In Section 4, we will provide a specific choice of full-projection matrices. We next show by leveraging full-projection matrices, the parameter θ_0 is (approximately) identifiable with fixed p and q. Define

$$\begin{aligned}\theta_{0,p,q} &= e_1^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1] \\ &\quad + \left(e_1^\top \mathbb{E}[QQ^\top \mid A=0]^{-1}\mathbb{E}[QN^\top \mid A=0]\right)\left(\mathbf{S}_{\tilde{q}}\mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1]\right).\end{aligned}$$

where $L = \left(1,\, p^{-1}X^\top \mathbf{W}_X,\, q^{-1}M^\top \mathbf{W}_M\right)^\top$, $Q = \left(1,\, p^{-1}X^\top \mathbf{W}_X\right)^\top$, and $N = q^{-1}\mathbf{W}_M^\top M$. Here, $\mathbf{S}_{\tilde{q}} = \begin{bmatrix} 0_{\tilde{q}\times(1+\tilde{p})} & I_{\tilde{q}} \end{bmatrix} \in \mathbb{R}^{\tilde{q}\times(1+\tilde{p}+\tilde{q})}$ be the selection matrix that selects last $\tilde{q}$ elements.

Theorem 3.1. *Suppose Assumptions 2.1-2.3 and 3.1–3.2 hold. Let $\mathbf{W}_X$ and $\mathbf{W}_M$ be any full-projection matrices, and θ_0 be defined as in Equation (2.6). Then, there exists a universal constant $C > 0$ independent of p and q, such that*

$$|\theta_0 - \theta_{0,p,q}| \;\le\; C\left(\tfrac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)} + \tfrac{1}{q\,\nu^2_{\min}(\mathbf{H}_M)}\right)\left(1 + \tfrac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)}\right).$$

If, in addition, $p\nu^2_{\min}(\mathbf{H}_X) \uparrow \infty$ and $q\nu^2_{\min}(\mathbf{H}_M) \uparrow \infty$ as $p, q \to \infty$, then $\theta_{0,p,q} \to \theta_0$, i.e., θ_0 is point identifiable as $p, q \to \infty$.

Remark 3.1. *The theorem assumes that both X and M are high-dimensional. A similar argument can be developed when either X or M is treated as fixed (i.e., no factor structure is imposed), we can replace $\mathbf{f}_X$ with X in the model and identification can still be achieved through the remaining high-dimensional component. For instance, if X is fixed, then $|\theta_0 - \theta_{0,q}| \le C\left(\tfrac{1}{q\,\nu^2_{\min}(\mathbf{H}_M)}\right)$. An analogous result holds when M is fixed.*

Theorem 3.1 provides an upper bound for the approximation error. As a consequence, it yields identification of θ_0 when p and q grow large. To the best of our knowledge, identification results of this type, have not been explicitly established in this form in the literature, including parameters such as the average treatment effect. We formalize this in the Corollary 3.2 where we define $\theta_{\mathrm{ATE},p}$ as follows:

$$\begin{aligned}\theta_{\mathrm{ATE},p,q} = &\Big(e_1^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1] \\ &+ \left(e_1^\top \mathbb{E}[QQ^\top \mid A=1]^{-1}\mathbb{E}[QN^\top \mid A=1]\right)\left(\mathbf{S}_{\tilde{q}}\mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1]\right)\Big) \\ &- \Big(e_1^\top \mathbb{E}[LL^\top \mid A=0]^{-1}\mathbb{E}[LY \mid A=0] \\ &+ \left(e_1^\top \mathbb{E}[QQ^\top \mid A=0]^{-1}\mathbb{E}[QN^\top \mid A=0]\right)\left(\mathbf{S}_{\tilde{q}}\mathbb{E}[LL^\top \mid A=0]^{-1}\mathbb{E}[LY \mid A=0]\right)\Big)\end{aligned}$$

Corollary 3.2. *Consider the structural models in (2.1)–(2.2). Suppose Assumptions 2.1–2.3 and 3.1–3.2 hold. Define $\theta_{\mathrm{ATE}} := (\alpha_1 + \gamma_1^\top \delta_1) - (\alpha_0 + \gamma_0^\top \delta_0)$. Then, θ_{ATE} satisfies the conditions of Theorem 3.1, there exists a universal constant $C > 0$, independent of p and q, such that*

$$\left|\theta_{\mathrm{ATE}} - \theta_{\mathrm{ATE},p,q}\right| \le C\left(\tfrac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)} + \tfrac{1}{q\,\nu^2_{\min}(\mathbf{H}_M)}\right)\left(1 + \tfrac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)}\right).$$

Moreover, if $p\,\nu^2_{\min}(\mathbf{H}_X) \to \infty$ and $q\,\nu^2_{\min}(\mathbf{H}_M) \to \infty$, then $\theta_{\mathrm{ATE},p,q} \to \theta_{\mathrm{ATE}}$, and hence θ_{ATE} is point identifiable as $p, q \to \infty$.

4 FARMA Algorithm for Mediation Analysis

In this section, we present FARMA for estimating the mediation functional $\theta_0 = \alpha_1 + \gamma_1^\top \delta_0$. The procedure constructs projection directions on one subsample and fits two low-dimensional regressions on the other, then interchanges the subsamples and averages the resulting estimates. Let $\tilde{p}$ and $\tilde{q}$ denote the working dimensions of the covariate and mediator projections, respectively.

Algorithm 1 Cross-fitted FARMA estimator of the mediation functional

Step 0: Randomly partition the sample indices into two approximately equal folds, $\mathcal{D}_1$ and $\mathcal{D}_2$.

Step 1: Let $\hat{\Sigma}_X$ and $\hat{\Sigma}_M$ be the sample covariance matrices of X_i and M_i computed on $\mathcal{D}_1$. Form $\hat{\mathbf{V}}_X \in \mathbb{R}^{p\times\tilde{p}}$ and $\hat{\mathbf{V}}_M \in \mathbb{R}^{q\times\tilde{q}}$ from their leading orthonormal eigenvectors, and set $\mathbf{W}_X = \sqrt{p}\,\hat{\mathbf{V}}_X$ and $\mathbf{W}_M = \sqrt{q}\,\hat{\mathbf{V}}_M$.

Step 2: For each $i \in \mathcal{D}_2$, compute the projected factors

$$\hat{\mathbf{f}}_{X,i} = p^{-1}\mathbf{W}_X^\top X_i, \quad \hat{\mathbf{f}}_{M,i} = q^{-1}\mathbf{W}_M^\top M_i,$$

and define $\hat{Z}_i = (1, \hat{\mathbf{f}}_{X,i}^\top, \hat{\mathbf{f}}_{M,i}^\top)^\top$ and $\hat{R}_i = (1, \hat{\mathbf{f}}_{X,i}^\top)^\top$.

Step 3: Let $\mathcal{I}_a = \{i \in \mathcal{D}_2 : A_i = a\}$ for $a \in \{0,1\}$. Regress Y_i on $\hat{Z}_i$ by ordinary least squares using $i \in \mathcal{I}_1$. Denote the fitted intercept and mediator-score coefficient by $\hat{\alpha}_1$ and $\hat{\gamma}_1 \in \mathbb{R}^{\tilde{q}}$, respectively.

Step 4: Let $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{1+\tilde{p}}$. Regress $\hat{\mathbf{f}}_{M,i}^\top\hat{\gamma}_1$ on $\hat{R}_i$ using $i \in \mathcal{I}_0$, and define the fitted intercept as

$$\widehat{\gamma_1^\top\delta_0} = e_1^\top\big(\textstyle\sum_{i\in\mathcal{I}_0}\hat{R}_i\hat{R}_i^\top\big)^{-1}\sum_{i\in\mathcal{I}_0}\hat{R}_i\big(\hat{\mathbf{f}}_{M,i}^\top\hat{\gamma}_1\big).$$

Step 5: Set $\hat{\theta}_0^{(1)} = \hat{\alpha}_1 + \widehat{\gamma_1^\top\delta_0}$. Interchange $\mathcal{D}_1$ and $\mathcal{D}_2$ and repeat Steps 1–4 to obtain $\hat{\theta}_0^{(2)}$.

Return: $\hat{\theta}_0^{\mathrm{cf}} = \frac{1}{2}\big(\hat{\theta}_0^{(1)} + \hat{\theta}_0^{(2)}\big)$.

In Step 1, we use $\mathcal{D}_1$ to construct the projection matrices from the leading eigenvectors of the sample covariance matrices. In Step 2, we apply these directions to the surrogates in $\mathcal{D}_2$ to obtain the projected covariate and mediator scores. This separation allows the regression observations to be analyzed conditional on the projection directions, while retaining the factor-space information needed for estimation.

Steps 3 and 4 implement the two regression stages underlying the mediation functional, a treated outcome regression and a control bridge regression. In Step 3, the treated outcome regression yields $\hat{\alpha}_1$ and $\hat{\gamma}_1$.

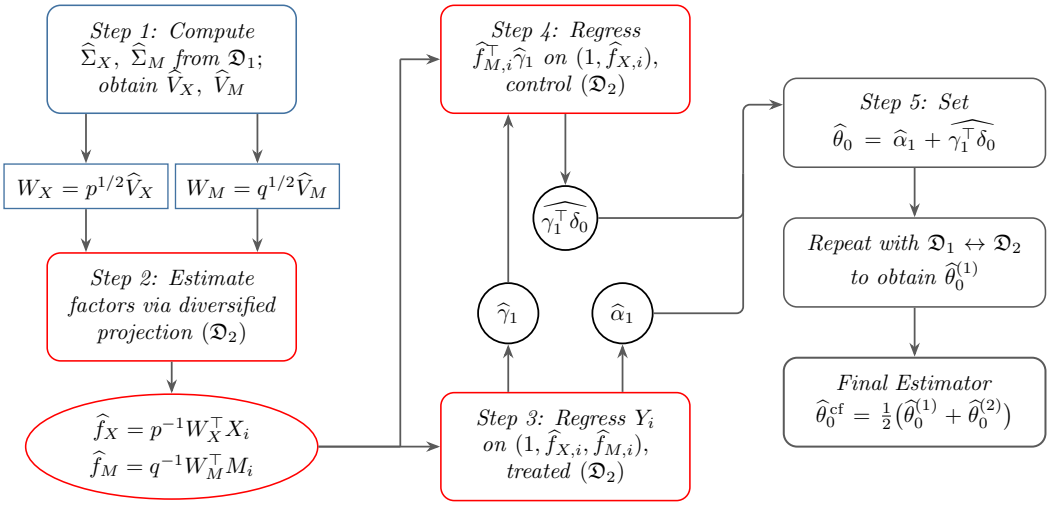


Figure 2: Schematic overview of FARMA. The estimates $\widehat{\alpha}_1$ and $\widehat{\gamma_1^\top\delta_0}$ combine into $\widehat{\theta}_0$, and interchanging the subsets gives the cross-fitted $\widehat{\theta}_0^{\mathrm{cf}}$.

In Step 4, the bridge regression uses $\hat{\gamma}_1$ to form the scalar response $\hat{\mathbf{f}}_{M,i}^\top\hat{\gamma}_1$ and regresses it on the projected covariates among control observations. The fitted intercept, denoted by $\widehat{\gamma_1^\top\delta_0}$, estimates the contribution of the control mediator relation to the treated outcome.

In Step 5, we add the two intercepts to obtain $\hat{\theta}_0^{(1)}$, repeat the construction with the folds interchanged, and average the two estimates. Thus, every observation contributes to the regression steps in one split, and

the procedure only requires recovery of the relevant factor spaces.

Natural direct and indirect effects. Under the centering convention $\mathbb{E}(\mathbf{f}_X) = 0$, thepotential outcomes in (2.7) satisfy $\theta_{aa} = \alpha_a + \gamma_a^\top \delta_a$ for $a \in \{0,1\}$. Thus, the effects defined in (2.8) have the structural representations as follows:

$$\theta_{\mathrm{NDE}} = \alpha_1 - \alpha_0 + (\gamma_1 - \gamma_0)^\top \delta_0,$$
$$\theta_{\mathrm{NIE}} = \gamma_1^\top (\delta_1 - \delta_0).$$

Algorithm 2 estimates these expressions using the same projections and scalar-regression construction as Algorithm 1. The projected covariates remain in each regression, since the marginal centering condition does not imply $\mathbb{E}(\mathbf{f}_X \mid A = a) = 0$. In Step 1 of Algorithm 2, we reuse the treated outcome fit and estimate the corresponding control outcome coefficients. In Step 2, the two within-arm mediator regressions provide $\widehat{\gamma_0^\top \delta_0}$ and $\widehat{\gamma_1^\top \delta_1}$. Step 3 combines these with $\widehat{\gamma_1^\top \delta_0}$ from Algorithm 1 to estimate the two effects. Finally, Step 4 interchanges the folds, and the two estimates of each effect are averaged.

Algorithm 2 Cross-fitted FARMA estimators of the NDE and NIE

Step 1: Reuse the projected factors, design vectors, treatment groups, and fitted quantities from the first split of Algorithm 1. Regress Y_i on $\hat{Z}_i$ using $i \in \mathcal{I}_0$, and denote the fitted intercept and mediator-score coefficient by $\hat{\alpha}_0$ and $\hat{\gamma}_0 \in \mathbb{R}^{\tilde{q}}$, respectively.

Step 2: For each $a \in \{0,1\}$, regress $\hat{\mathbf{f}}_{M,i}^\top \hat{\gamma}_a$ on $\hat{R}_i$ using $i \in \mathcal{I}_a$, and define the fitted intercept as

$$\widehat{\gamma_a^\top \delta_a} = e_1^\top \Big(\textstyle\sum_{i \in \mathcal{I}_a} \hat{R}_i \hat{R}_i^\top\Big)^{-1} \sum_{i \in \mathcal{I}_a} \hat{R}_i \big(\hat{\mathbf{f}}_{M,i}^\top \hat{\gamma}_a\big).$$

Step 3: Form the fold-specific estimates

$$\hat{\theta}_{\mathrm{NDE}}^{(1)} = \hat{\alpha}_1 - \hat{\alpha}_0 + \widehat{\gamma_1^\top \delta_0} - \widehat{\gamma_0^\top \delta_0},$$
$$\hat{\theta}_{\mathrm{NIE}}^{(1)} = \widehat{\gamma_1^\top \delta_1} - \widehat{\gamma_1^\top \delta_0}.$$

Step 4: Interchange $\mathcal{D}_1$ and $\mathcal{D}_2$ and repeat Steps 1–3 using the corresponding fitted quantities from Algorithm 1, yielding $\hat{\theta}_{\mathrm{NDE}}^{(2)}$ and $\hat{\theta}_{\mathrm{NIE}}^{(2)}$.

Return:

$$\hat{\theta}_{\mathrm{NDE}}^{\mathrm{cf}} = \tfrac{1}{2}\big(\hat{\theta}_{\mathrm{NDE}}^{(1)} + \hat{\theta}_{\mathrm{NDE}}^{(2)}\big), \quad \hat{\theta}_{\mathrm{NIE}}^{\mathrm{cf}} = \tfrac{1}{2}\big(\hat{\theta}_{\mathrm{NIE}}^{(1)} + \hat{\theta}_{\mathrm{NIE}}^{(2)}\big).$$

Remark 4.1 (Cross-fitting)**.** *The two-fold construction extends to any fixed number of folds $K \geq 2$. For each fold, construct the projection directions from the remaining observations, run the regressions on that fold, and average the resulting K fold-specific estimates of each target.*

5 Theoretical Analysis

In this section, we study the asymptotic behavior of the cross-fitted FARMA estimator $\hat{\theta}_0^{\mathrm{cf}}$ and establish the $\sqrt{n}$-consistency and asymptotic normality. Recall that Algorithm 1 constructs the estimator using two-fold cross-fitting, with the projection directions estimated on one fold and the regression steps carried out on the other. The roles of the two folds are then interchanged and the resulting estimates are averaged. We derive an asymptotic linear expansion for one fold-specific estimator, and the same argument applies after interchanging the folds.

Let $\pi_1 = \mathbb{P}(A_i = 1)$ and $\pi_0 = \mathbb{P}(A_i = 0) = 1 - \pi_1$ denote the treatment probabilities with $0 < \pi_1, \pi_0 < 1$. For $k \in \{1,2\}$, let $n_k = |\mathcal{D}_k|$, $n_k^t = \sum_{i \in \mathcal{D}_k} \mathbf{1}(A_i = 1)$, and $n_k^c = \sum_{i \in \mathcal{D}_k} \mathbf{1}(A_i = 0)$ denote the total, treated, and control sample sizes in fold k, respectively. Under an approximately equal split, we have that $n_k \sim n/2$, and

$$\frac{n_k^t}{n_k} \xrightarrow{p} \pi_1, \qquad \frac{n_k^c}{n_k} \xrightarrow{p} \pi_0. \tag{5.1}$$

Thus, the treated and control sample sizes in each fold are of the same order as n. The main challenge is to control the effect of estimating the latent factor spaces on the two regression stages. In particular, the regressions are carried out using projected factors rather than the latent factors themselves, and the working dimensions may be larger than the true factor dimensions. We show that the error from estimating the factor spaces here is asymptotically negligible. Under the structural models (2.2), the projected factors in the regression fold satisfy

$$\begin{aligned}\hat{\mathbf{f}}_{X,i} &= \frac{1}{p}\mathbf{W}_X^\top X_i = \underbrace{\left(\frac{1}{p}\mathbf{W}_X^\top \Gamma_X\right)}_{:=\mathbf{H}_X}\mathbf{f}_{X,i} + \underbrace{\frac{1}{p}\mathbf{W}_X^\top \mathbf{u}_{X,i}}_{\mathbf{e}_{X,i}},\\ \hat{\mathbf{f}}_{M,i} &= \frac{1}{q}\mathbf{W}_M^\top M_i = \underbrace{\left(\frac{1}{q}\mathbf{W}_M^\top \Gamma_M\right)}_{:=\mathbf{H}_M}\mathbf{f}_{M,i} + \underbrace{\frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_{M,i}}_{\mathbf{e}_{M,i}}.\end{aligned} \tag{5.2}$$

Here, the matrices $\mathbf{H}_X \in \mathbb{R}^{\tilde{p}\times\bar{p}}$ and $\mathbf{H}_M \in \mathbb{R}^{\tilde{q}\times\bar{q}}$ map the latent factors into the projected coordinates, while $\mathbf{e}_{X,i}$ and $\mathbf{e}_{M,i}$ collect the projected idiosyncratic errors. We work under the overestimated factor regime described earlier, where the working dimensions may be larger than the true factor dimensions.

For the rest of the paper, we use the following notation. Let $\mathbf{f}_{1,i} = (1, \mathbf{f}_{X,i}^\top, \mathbf{f}_{M,i}^\top)^\top$ and $\mathbf{f}_{X1,i} = (1, \mathbf{f}_{X,i}^\top)^\top$. Define the conditional second-moment matrices $\Omega_1 = \mathbb{E}\big[\mathbf{f}_{1,i}\mathbf{f}_{1,i}^\top \mid A_i = 1\big]$ and $\Omega_{X0} = \mathbb{E}\big[\mathbf{f}_{X1,i}\mathbf{f}_{X1,i}^\top \mid A_i = 0\big]$. These matrices correspond to the treated outcome regression and the control bridge regression, respectively.

We first state the regularity conditions used in the asymptotic analysis.

Assumption 5.1 (Sub-Gaussian tails and nondegeneracy)**.** *There exist constants $C < \infty$ and $\underline{\sigma}^2 > 0$, independent of n, p, q, such that, for $a \in \{0, 1\}$ and $i = 1, \dots, n$, the following hold.*

(i) All projections of $\mathbf{f}_{X,i}$, $\mathbf{u}_{X,i}$, $\mathbf{u}_{M,i}$, $\mathbf{u}_{XM,a,i}$ and the scalar error $\epsilon_{a,i}$ are sub-Gaussian with a common parameter bounded by C.

(ii) The structural residual variance is bounded below, $\mathrm{Var}(\epsilon_{a,i}) \geq \underline{\sigma}^2 > 0$.

Assumption 5.1 provides uniform sub-Gaussian concentration for sample moments and bounds the structural noise variance away from zero.

Assumption 5.2 (Dimensions and loadings)**.** *There exists a constant $C < \infty$, independent of n, p, q, such that the following hold.*

(i) The loading rows are uniformly bounded, $\max_{j\leq p} \|\mathbf{\Gamma}_{X,j\cdot}\|_2 \leq C$ and $\max_{j\leq q} \|\mathbf{\Gamma}_{M,j\cdot}\|_2 \leq C$.

(ii) The working dimensions $\tilde{p}$ and $\tilde{q}$ are fixed, $\log(p \vee q)/n \to 0$ and $(p \wedge q)/n \to \infty$ as $n, p, q \to \infty$.

Assumption 5.2(i) bounds the contribution of each individual surrogate to the factor structure. The first condition in Assumption 5.2(ii), $\log(p \vee q)/n \to 0$, is mild and allows the surrogate dimensions to grow rapidly with the sample size, as long as they do not grow exponentially fast. The second condition, $(p \wedge q)/n \to \infty$, requires the number of surrogate measurements to grow faster than the sample size. This condition ensures that the error from estimating the factor spaces is negligible at the $\sqrt{n}$ scale. It may be possible to weaken the condition to $(p \wedge q)/\sqrt{n} \to \infty$, which we leave for future work.

Proposition 5.1. *Let $\mathbf{V}_X \in \mathbb{R}^{p\times\tilde{p}}$ and $\mathbf{V}_M \in \mathbb{R}^{q\times\tilde{q}}$ have orthonormal columns corresponding to the leading $\tilde{p}$ and $\tilde{q}$ eigenvectors of the sample covariance matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_M$, respectively. Take $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$ and $\mathbf{W}_M = \sqrt{q}\,\mathbf{V}_M$, where $\tilde{p} \geq \bar{p}$ and $\tilde{q} \geq \bar{q}$, and let $\mathbf{H}_X$ and $\mathbf{H}_M$ be defined as in Definition 3.1. Suppose Assumptions 3.1–3.2 and 5.1–5.2 hold. Then there exist constants $0 < c \leq C < \infty$ and $C_1 > 0$, independent of n, p, q. For $t > 0$, define*

$$\delta_X = \sqrt{\tfrac{\log p + t}{n}} + \tfrac{1}{\sqrt{p}}, \quad \delta_M = \sqrt{\tfrac{\log q + t}{n}} + \tfrac{1}{\sqrt{q}}.$$

There is a constant $c_\star > 0$ such that if $\delta_\star = \max\{\delta_X, \delta_M\} \leq c_\star$, then, with probability at least $1 - C_1 e^{-t}$,

$$c \leq \nu_{\min}(\mathbf{H}_X) \leq \nu_{\max}(\mathbf{H}_X) \leq C, \quad c \leq \nu_{\min}(\mathbf{H}_M) \leq \nu_{\max}(\mathbf{H}_M) \leq C.$$

Proposition 5.1 ensures that the projected factor maps remain well-conditioned with probability tending to one, which is important for the stability of the regression steps and the asymptotic analysis of the FARMA estimator.

Let $a_1 = (1, \mathbf{0}_{\bar{p}}^\top, \delta_0^\top)^\top$, $e_{1,1+\bar{p}} = (1, \mathbf{0}_{\bar{p}}^\top)^\top$, and $\eta_i = \gamma_1^\top \mathbf{u}_{XM,0,i}$. Define $\xi_{1,i} = \left(\Omega_1^{-1} a_1\right)^\top \mathbf{f}_{1,i}\, \epsilon_{1,i}$, and $\xi_{2,i} = \left(\Omega_{X0}^{-1} e_{1,1+\bar{p}}\right)^\top \mathbf{f}_{X1,i}\, \eta_i$,. Here, $\xi_{1,i}$ and $\xi_{2,i}$ are the first-order terms from the treated outcome regression and the control bridge regression, respectively.

Theorem 5.2 (Asymptotic normality of $\hat{\theta}_0^{\mathrm{cf}}$)**.** *Suppose Assumptions 2.1–2.3, 3.1–3.2, and 5.1–5.2 hold, and let $\hat{\theta}_0^{\mathrm{cf}}$ be the cross-fitted FARMA estimator defined in Algorithm 1. Then*

$$\sqrt{n}\left(\hat{\theta}_0^{\mathrm{cf}} - \theta_0\right) \xrightarrow{d} N(0, \sigma^2), \quad \sigma^2 = \frac{\sigma_1^2}{\pi_1} + \frac{\sigma_2^2}{\pi_0}$$

where $\sigma_1^2 = \mathbb{E}[\xi_1^2 \mid A = 1]$ and $\sigma_2^2 = \mathbb{E}[\xi_2^2 \mid A = 0]$.

Theorem 5.2 shows that the error from estimating the high-dimensional factor spaces does not affect the first-order limiting distribution of $\hat{\theta}_0^{cf}$. We can choose the working dimensions $\tilde{p}$ and $\tilde{q}$ to be larger than the true factor dimensions, as long as they remain fixed and the projected spaces contain the true factor spaces. The estimator still has the same $\sqrt{n}$ convergence rate and asymptotic variance as if the true latent factors were observed. The proof is given in Appendix A.

6 Simulation Studies

In this section, we design simulation experiments to evaluate the performance of our estimator ($\hat{\theta}_0^{\mathrm{cf}}$ as defined in Section 2) under high-dimensional surrogates for the covariates and mediators, in terms of root mean squared error (RMSE), standard deviation (SD), confidence interval length, and coverage. We consider a data-generating process where both covariates and mediators are low-dimensional latent factors as stated in Section 2. The data is generated as follows:

1. We generate a latent factor $\mathbf{f}_X \in \mathbb{R}^{\bar{p}}$ for each sample, with $\mathbf{f}_X \sim \mathcal{N}(0, \Sigma_X)$. The covariance matrix is constructed as $\Sigma_X = U \Lambda U^\top$, where $U \in \mathbb{R}^{\bar{p} \times \bar{p}}$ is an orthonormal matrix, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{\bar{p}})$ is diagonal with $\lambda_j \sim \mathcal{U}(1, 2)$ for $1 \le j \le \bar{p}$. Given covariates $\mathbf{f}_X$, the observed surrogates $X \in \mathbb{R}^p$ are generated as:
$$X = \mathbf{\Gamma}_X \mathbf{f}_X + \mathbf{u}_X, \quad \mathbf{u}_X \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_p),$$
where $\mathbf{\Gamma}_X \in \mathbb{R}^{p \times \bar{p}}$ has entries sampled from $\mathcal{U}(0.5, 1.5)$.

2. The treatment assignment follows a logistic model:
$$\mathbb{P}(A = 1 \mid \mathbf{f}_X) = \mathrm{logit}^{-1}(\mathbf{f}_X^\top \xi), \quad \text{where } \xi \in \mathbb{R}^{\bar{p}} \text{ with } \xi_j \sim \mathcal{U}(0, 2).$$
Then, $A \sim \mathrm{Bernoulli}(\mathbb{P}(A = 1 \mid \mathbf{f}_X))$.

3. The latent mediator $\mathbf{f}_M \in \mathbb{R}^{\bar{q}}$ is generated according to:
$$\mathbf{f}_M = A(\delta_1 + \mathbf{B}_1 \mathbf{f}_X + \mathbf{U}_{XM,0}) + (1 - A)(\delta_0 + \mathbf{B}_0 \mathbf{f}_X + \mathbf{U}_{XM,1}), \quad \mathbf{U}_{XM,0}, \mathbf{U}_{XM,1} \sim \mathcal{N}(0, \Sigma_U),$$
where $\Sigma_U = R_U \Lambda_U R_U^\top$, $R_U \in \mathbb{R}^{\bar{q} \times \bar{q}}$ is an orthonormal matrix, $\Lambda_U = \mathrm{diag}(\lambda_1, \ldots, \lambda_{\bar{q}})$ with $\lambda_j \overset{i.i.d.}{\sim} \mathcal{U}(1, 2)$, $\delta_0, \delta_1 \in \mathbb{R}^{\bar{q}}$ have entries drawn from $\mathcal{U}(0.5, 1.5)$, and $\mathbf{B}_0, \mathbf{B}_1 \in \mathbb{R}^{\bar{q} \times \bar{p}}$ have entries sampled from $\mathcal{U}(0.5, 1.5)$. Given mediators $\mathbf{f}_M$, we generate the observed surrogates $M \in \mathbb{R}^q$ by:
$$M = \mathbf{\Gamma}_M \mathbf{f}_M + \mathbf{u}_M, \quad \mathbf{u}_M \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_q),$$
where $\mathbf{\Gamma}_M \in \mathbb{R}^{q \times \bar{q}}$ has entries from $\mathcal{U}(0.5, 1.5)$.

4. We generate response variable Y from the model:

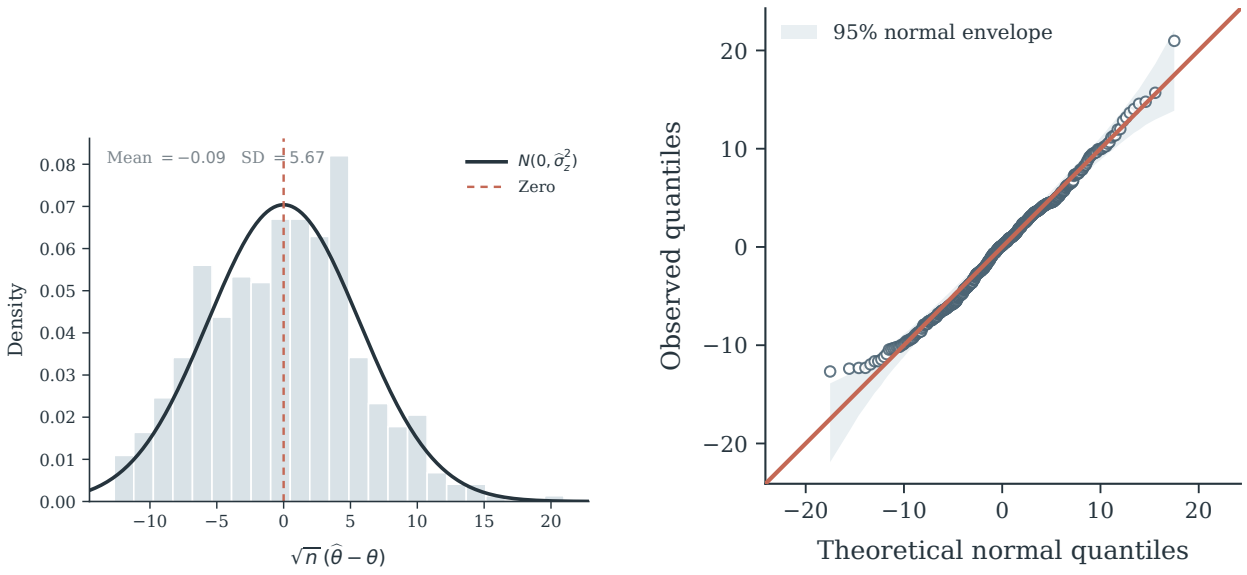
$$Y = A(\alpha_1 + \beta_1^\top \mathbf{f}_X + \gamma_1^\top \mathbf{f}_M + \epsilon_0) + (1 - A)(\alpha_0 + \beta_0^\top \mathbf{f}_X + \gamma_0^\top \mathbf{f}_M + \epsilon_1), \quad \epsilon_0, \epsilon_1 \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2),$$

where we set $\alpha_0 = \alpha_1 = 1$, and vectors $\beta_0, \beta_1 \in \mathbb{R}^{\bar{p}}$, $\gamma_0, \gamma_1 \in \mathbb{R}^{\bar{q}}$ are sampled from $\mathcal{U}(0.5, 1.5)$.

We vary the number of samples n over $\{500, 1000, 1500, 2000, 2500\}$ and the dimensionality $p = q \in \{500, 1000, 2500, 5000\}$ to reflect different high-dimensional scenarios. For all simulations, we vary $\sigma \in \{0.1, 1\}$, and set $\bar{p} = \bar{q} = 3$ with $\tilde{p} = \tilde{q} = 7$. This choice of $\tilde{p}$ (resp. $\tilde{q}$) comes from the lowest-RMSE rule in large samples according to our sensitivity analysis (Appendix D.3). Our estimator ($\hat{\theta}_0^{\text{cf}} = \hat{\alpha}_1 + \widehat{\gamma_1^\top \delta_0}$ as defined in Section 2) is computed as in Algorithm 1. The weight matrices in Step 2 (i.e., $\mathbf{W}_X$, $\mathbf{W}_M$) from principal component analysis are obtained via the R package `irlba`, and the OLS estimators in Steps 4–6 via the `lm` function in R. All results are averaged over 500 replications.

		$\sigma = 0.1$				$\sigma = 1$			
$p = q$	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
500	500	0.2068	0.2065	0.8094	0.950	0.2291	0.2294	0.8991	0.958
	1000	0.1451	0.1452	0.5692	0.944	0.1461	0.1462	0.5732	0.956
	1500	0.1154	0.1153	0.4520	0.952	0.1239	0.1236	0.4846	0.960
	2000	0.1009	0.1001	0.3925	0.964	0.1018	0.1017	0.3986	0.942
	2500	0.0830	0.0831	0.3256	0.956	0.0927	0.0918	0.3597	0.946
1000	500	0.1642	0.1639	0.6424	0.952	0.1348	0.1346	0.5278	0.958
	1000	0.1109	0.1110	0.4350	0.954	0.0902	0.0903	0.3538	0.946
	1500	0.0937	0.0938	0.3677	0.946	0.0783	0.0784	0.3072	0.954
	2000	0.0793	0.0793	0.3110	0.950	0.0632	0.0632	0.2476	0.952
	2500	0.0703	0.0703	0.2755	0.948	0.0574	0.0572	0.2241	0.946
2500	500	0.1754	0.1751	0.6865	0.954	0.2526	0.2527	0.9907	0.954
	1000	0.1200	0.1201	0.4709	0.956	0.1895	0.1893	0.7420	0.952
	1500	0.1039	0.1040	0.4077	0.944	0.1462	0.1463	0.5736	0.956
	2000	0.0810	0.0811	0.3179	0.954	0.1242	0.1244	0.4875	0.952
	2500	0.0764	0.0765	0.2997	0.938	0.1159	0.1161	0.4549	0.956
5000	500	0.1700	0.1701	0.6670	0.934	0.2072	0.2074	0.8129	0.940
	1000	0.1223	0.1223	0.4793	0.956	0.1350	0.1352	0.5298	0.952
	1500	0.0989	0.0989	0.3878	0.954	0.1095	0.1089	0.4269	0.946
	2000	0.0856	0.0856	0.3356	0.946	0.0945	0.0946	0.3706	0.952
	2500	0.0794	0.0795	0.3116	0.944	0.0848	0.0847	0.3322	0.950

Table 1: Performance of $\hat{\theta}_0^{\text{cf}}$ under the baseline design, over 500 replications.



(a) Histogram

(b) Normal Q–Q plot

Figure 3: Histogram and Q–Q plot of $\sqrt{n}(\widehat{\theta}_0^{\mathrm{cf}} - \theta_0)$ under the base design with $p = q = 2500$, $n = 1500$, $\sigma = 1$, and $\widetilde{p} = \widetilde{q} = 7$, based on 500 Monte Carlo replications.

We evaluate the estimator across four simulation designs:

- **Baseline design.** This applies the data-generating process described above over the full grid of n, $p = q \in \{500, 1000, 2500, 5000\}$ and $\sigma \in \{0.1, 1\}$.
- **Signal strength.** The baseline design fixes a single non-zero value for every parameter. To let the mediated pathway range from absent to strong, we proceed as follows. Fix a direction vector $g \in \mathbb{R}^{\bar{q}}$ with entries drawn once from $\mathcal{U}(0.5, 1.5)$ and set

$$\gamma_1 = \gamma_0 = c\, g, \qquad c \in \{0, 0.1, 0.25, 0.5, 1\},$$

  leaving all other components of the data-generating process unchanged, so that $\theta_0 = 1 + c\,\delta_0^\top g$ and the mediated contribution $\delta_0^\top \gamma_1$ vanishes at $c = 0$ and grows linearly in c. We report results for $p = q \in \{1000, 2500, 5000\}$ at $\sigma = 0.1$, so that the accuracy of $\hat{\theta}_0$ can be traced as the mediator signal weakens relative to the noise.
- **Heavy-tailed errors.** Separately, we replace all three error terms by variance-matched Student-t variates: writing $t \sim t_\nu$ with $\nu = 6$, each coordinate of $\mathbf{u}_X$ and $\mathbf{u}_M$ and the outcome error ϵ is set to $\sigma\, t/\sqrt{\nu/(\nu-2)} = \sigma\, t/\sqrt{3/2}$, which has variance exactly σ^2 and so isolates tail thickness from any change of scale. All other components of the design and the full grid of n, $p = q$ and σ are unchanged.
- **Sensitivity to the working factor dimensions.** Algorithm 1 requires the numbers $\tilde{p}$ and $\tilde{q}$ of principal components extracted from X and from M, and $\bar{p}$, $\bar{q}$ are unknown in any application. Setting $\kappa = \tilde{p} = \tilde{q}$, we vary

$$\kappa \in \{1, 2, 3, 5, 7, 10, 15, 20, 30\}$$

  across all five sample sizes and $p = q \in \{1000, 2500, 5000\}$ at $\sigma = 0.1$, retaining the baseline design in all other respects. This brackets the truth $\bar{p} = \bar{q} = 3$ on both sides and reaches an order of magnitude above it. We also additionally vary $\tilde{p}$ and $\tilde{q}$ separately over $\{3, 7, 20\}$ at $n = 2500$.

Overall, the simulation results provide consistent support for the accuracy of $\widehat{\theta}_0^{\mathrm{cf}}$ and the validity of the associated inference. Under the baseline design, the RMSE closely tracks the empirical standard deviation, both decrease as n increases, and the confidence intervals become shorter while maintaining coverage near

the nominal 95% level, even when $p = q$ substantially exceeds n. The same pattern persists under variance-matched t_6 errors; see Table 3. The histogram and Q–Q plot are also broadly consistent with the Gaussian distribution; see Figure 3a. As the mediated pathway varies from absent to strong, the estimator continues to exhibit decreasing estimation error with increasing n and coverage close to the nominal level; see Table 4. The sensitivity analysis further shows that the estimator is relatively insensitive to moderate or substantial overestimation of the common working factor dimension; see Table 5 and Figure 4. Similar stability is observed when $\widetilde{p}$ and $\widetilde{q}$ are varied separately; see Table 6. In contrast, severe underestimation of the working dimensions can produce persistent bias and substantial undercoverage in the more high-dimensional settings, as shown in Table 5. These findings are consistent with the asymptotic theory and indicate that reliable estimation and inference require the working representations to capture the relevant latent factor spaces.

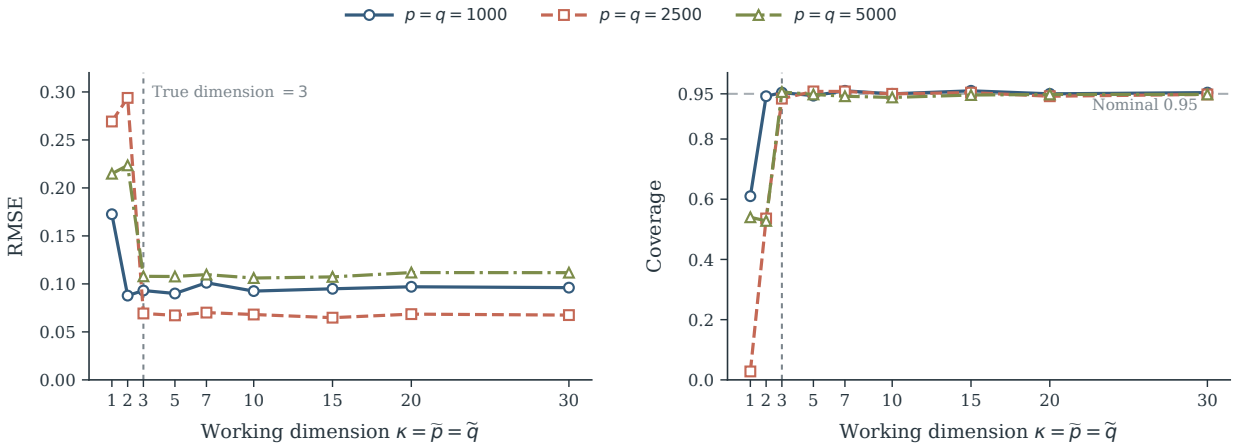


Figure 4: Sensitivity Analysis of the working factor dimensions.

7 Real Data Experiments

Depression is among the most consistently reported risk factors for cognitive decline in Alzheimer's disease, yet the biological route through which it acts remains poorly understood (Livingston et al., 2024). A hypothesis is that part of the effect is biological, transmitted through epigenetic changes such as DNA methylation, yet direct evidence for such mediation remains scarce (Zhang et al., 2026b; Fransquet et al., 2018). We apply our proposed method to investigate how depressive symptoms influence cognitive decline in Alzheimer's disease, and whether part of this relationship may operate through an underlying biological process represented by high-dimensional DNA methylation measurements.

As mentioned in Section 1, we view oxidative stress as the underlying biological process of interest and DNA methylation as a high-dimensional representation of this process. Since reactive oxygen species are short lived and highly reactive, oxidative stress is difficult to measure directly (Katerji et al., 2019). DNA methylation provides more stable molecular signals that may reflect changes associated with oxidative activity (Miranda et al., 2000; Pandya et al., 2013; Black et al., 2015). These changes are not expected to be represented by a single CpG site, but rather by patterns distributed across many sites. We therefore use the joint methylation profile as a collection of high-dimensional surrogates for the latent mediator, allowing the underlying biological signal to be summarized across the methylation measurements.

We conduct our analysis using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a longitudinal, multi-center study designed to validate biomarkers for Alzheimer's disease clinical trials (Mueller et al., 2005). Our treatment is the Geriatric Depression Scale (GDS), a self-reported measure of depressive symptoms ranging from 0 to 15, with higher scores indicating more severe symptoms. Our outcome is the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), which ranges from 0 to 85, where higher scores reflect greater cognitive impairment. The mediator surrogates are DNA methylation measure-

ments recorded at hundreds of thousands of CpG sites across the genome. Further details of the study are available at https://adni.loni.usc.edu/.

Data preprocessing. We study whether part of the relationship between depressive symptoms and cognition may operate through the latent biological process represented by DNA methylation, using measurements from the earliest available visit date during the ADNI-GO and ADNI-2 phases, for which methylation data are available. The dataset[1] includes DNA methylation data for 653 individuals, ADAS-Cog scores for 1,212 individuals, GDS scores for 1,800 individuals, demographic information for 1,895 individuals, and medical history data for 1,673 individuals. Of these, 649 subjects had complete data across all components and form our analysis sample. We preprocess the data as follows:

- The treatment A is a binarized Geriatric Depression Scale, where a GDS score below 5 indicates a normal mood state (Greenberg, 2012), so subjects with GDS > 5 are classified as depressed and the rest as non-depressed. We include ten clinically relevant covariates X, such as gender, weight, and education level, of which seven are numerical and three are categorical (see Appendix D.4 for the full list).
- We impute missing covariate values with `mice` (Van Buuren and Groothuis-Oudshoorn, 2011). To increase model expressiveness, we add second- through sixth-order polynomial terms and interactions between the categorical and numerical covariates, giving a covariate vector of dimension 171.
- The raw DNA methylation data give marker intensities in red and green channels for each sample, from the Illumina HumanMethylationEPIC BeadChip array. We compute the methylation ratio, ranging continuously from 0 to 1, using the `minfi` package (Aryee et al., 2014).
- The full methylation vector has dimension 866,836. We retain only sites with no missing values and CpG sites reported as significantly associated with Alzheimer's disease in large-scale EWAS (Zhang et al., 2020; Lunnon et al., 2014; Battram et al., 2022), which reduces the mediators to dimension 3,206.

After preprocessing, we retain $n = 649$ subjects, each with a 171-dimensional covariate-surrogate vector X and 3,206 DNA methylation measurements forming the high-dimensional mediator-surrogate vector M.

Estimation. To estimate the natural indirect effect, natural direct effect, and total effect of GDS on ADAS-Cog, we apply Algorithm 1 to the counterfactual means $\mathbb{E}[Y^{(1,1)}]$, $\mathbb{E}[Y^{(0,0)}]$, and $\mathbb{E}[Y^{(1,0)}]$ (see Appendix D.2 for the decomposition). We set the projection dimensions $\tilde{p}$ and $\tilde{q}$ to the number of leading principal components that explain 50% of the cumulative variance, yielding $(\tilde{p}, \tilde{q}) = (4, 11)$. We quantify uncertainty with a nonparametric bootstrap. We resample the $n = 649$ subjects with replacement. For each resample, we apply FARMA using ten independent random two-fold splits and average the resulting direct, indirect, and total effect estimates to obtain one bootstrap replicate. Repeating this procedure across B resamples yields a bootstrap distribution for each effect, from which we report the bootstrap mean and the 95% percentile confidence interval defined by the empirical 2.5% and 97.5% quantiles.

Results. The full-sample point estimates and the bootstrap summaries are given in Table 2.

Effect	Estimate	Bootstrap Mean	Quantile CI (95%)
Direct	3.5891	3.7306	$[1.5699, 6.1491]$
Indirect	0.0354	0.0043	$[-0.5131, 0.4899]$
Total	3.6244	3.7349	$[1.6816, 6.1864]$

Table 2: Estimated natural direct, natural indirect, and total effects with $B = 300$.

The estimated total effect of GDS-defined depressive status on ADAS-Cog was 3.6244, with a 95% percentile bootstrap confidence interval of $[1.6816, 6.1864]$. The positive estimate indicates that participants

[1]The data are accessible at https://ida.loni.usc.edu/login.jsp.

with elevated depressive symptoms had an estimated ADAS-Cog score approximately 3.62 points higher than those without elevated symptoms, corresponding to worse cognitive function. The natural direct effect was estimated at 3.5891 with a 95% confidence interval of $[1.5699, 6.1491]$, whereas the natural indirect effect through the latent pathway represented by the DNA methylation measurements was 0.0354 with an interval of $[-0.5131, 0.4899]$. Thus, the intervals for the total and direct effects excluded zero, while the interval for the indirect effect included zero. Based on the point estimates, the indirect effect accounted for approximately 1.0% of the total effect.

Overall, the estimated total effect was primarily attributable to the direct effect. The near-zero indirect-effect estimate and its confidence interval provide little evidence of mediation through the latent biological pathway represented by the DNA methylation measurements considered in this analysis. However, these results do not establish the absence of a mediated effect and do not rule out mediation through biological processes not adequately captured by the selected methylation features or through other biological and behavioral pathways.

8 Conclusion and Future Work

In this paper, we introduced Factor-Adjusted Regression for Mediation Analysis (FARMA), a factor-based framework for estimating mediation functionals when the covariates and mediators are latent and observed only through high-dimensional noisy surrogates. The main idea is that when the factor structure is pervasive, low-dimensional projections of the observed surrogates keep the information needed for causal mediation analysis. Within this framework, we showed that the mediation functional is approximately identified at finite proxy dimensions, and that the approximation error goes to zero as the numbers of surrogate covariates and mediators grow. We then built a cross-fitted estimator that separates the construction of the projection directions from the regression steps and does not need the true numbers of latent factors to be known. Under suitable regularity and growth conditions, this estimator is $\sqrt{n}$-consistent and asymptotically normal, so we can form valid Wald confidence intervals for the mediation functional, and hence for the natural direct and indirect effects. The simulations support the finite-sample accuracy of the estimator and its confidence intervals. In the ADNI application, we find that depressive symptoms are associated with greater cognitive impairment, with little evidence of mediation through the selected DNA methylation measurements.

Several directions remain open for future research. The present theory requires the dimensions of the observed variables to grow faster than the sample size, and extending it to the proportional regime, in which $\frac{p}{n} \to c_X$ and $\frac{q}{n} \to c_M$ for finite positive constants c_X and c_M that would require accounting for projection errors that persist asymptotically rather than vanish. A second direction is to relax sequential exchangeability to accommodate confounding beyond the latent structure captured by the available surrogates, together with the sensitivity analysis needed to assess robustness to such departures. Finally, we focused on linear structural models, and extending FARMA to generalized linear and other nonlinear outcome and mediator models would widen the range of problems it can handle.

References

Xinyu Zhang, Wei Zhou, Jingyuan Liu, and Jian Kang. Statistical inference for mediation models with high dimensional exposures and mediators. *Journal of the American Statistical Association*, 0(0):1–22, 2026a.

Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

Judea Pearl. Direct and indirect effects. In John Breese and Daphne Koller, editors, *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 411–420. Morgan Kaufmann, 2001.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, 2010. doi: 10.1214/10-STS321.

Tyler VanderWeele and Stijn Vansteelandt. Mediation analysis with multiple mediators. *Epidemiologic methods*, 2(1):95–115, 2014.

Linda Valeri and Tyler J VanderWeele. Sas macro for causal mediation analysis with survival data. *Epidemiology*, 26(2):e23–e24, 2015.

David MacKinnon. *Introduction to statistical mediation analysis*. Routledge, 2012.

Tyler VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.

Sun Jae Jung. Introduction to mediation analysis and examples of its application to real-world data. *Journal of Preventive Medicine and Public Health*, 54(3):166, 2021.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

James M Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992.

Eric J Tchetgen Tchetgen and Ilya Shpitser. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*, 40(3):1816, 2012.

Jeffrey M Albert, Cuiyu Geng, and Suchitra Nelson. Causal mediation analysis with a latent mediator. *Biometrical Journal*, 58(3):535–548, 2016.

Andriy Derkach, Ruth M Pfeiffer, Ting-Huei Chen, and Joshua N Sampson. High dimensional mediation analysis with latent variables. *Biometrics*, 75(3):745–756, 2019.

Amiremad Ghassami, Alan Yang, Ilya Shpitser, and Eric Tchetgen Tchetgen. Causal inference with hidden mediators. *Biometrika*, 112(1):asae037, 2025.

Haixiang Zhang, Yinan Zheng, Zhou Zhang, Tao Gao, Brian Joyce, Grace Yoon, Wei Zhang, Joel Schwartz, Allan Just, Elena Colicino, et al. Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, 32(20):3150–3154, 2016.

Yen-Tsung Huang and Wen-Chi Pan. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, 72(2):402–413, 2016.

Oliver Y Chén, Ciprian Crainiceanu, Elizabeth L Ogburn, Brian S Caffo, Tor D Wager, and Martin A Lindquist. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19(2):121–136, 2018.

Jie Zhang, Zhi Wei, and Jun Chen. A distance-based approach for testing the mediation effect of the human microbiome. *Bioinformatics*, 34(11):1875–1883, 2018.

Yi Zhao and Xi Luo. Pathway lasso: pathway estimation and selection with high-dimensional mediators. *Statistics and its Interface*, 15(1):39, 2021.

Chamila Perera, Haixiang Zhang, Yinan Zheng, Lifang Hou, Annie Qu, Cheng Zheng, Ke Xie, and Lei Liu. Hima2: high-dimensional mediation analysis and its application in epigenome-wide dna methylation data. *BMC bioinformatics*, 23(1):296, 2022.

Xu Guo, Runze Li, Jingyuan Liu, and Mudong Zeng. High-dimensional mediation analysis for selecting dna methylation loci mediating childhood trauma and cortisol stress reactivity. *Journal of the American Statistical Association*, 117(539):1110–1121, 2022.

Xu Guo, Runze Li, Jingyuan Liu, and Mudong Zeng. Statistical inference for linear mediation models with high-dimensional mediators and application to studying stock reaction to covid-19 pandemic. *Journal of Econometrics*, 235(1):166–179, 2023.

Xu Guo, Runze Li, Jingyuan Liu, and Mudong Zeng. Estimations and tests for generalized mediation models with high-dimensional potential mediators. *Journal of Business & Economic Statistics*, 42(1):243–256, 2024.

Prabrisha Rakshit and Zijian Guo. Statistical inference in high-dimensional poisson regression with applications to mediation analysis. *arXiv preprint arXiv:2410.20671*, 2024.

Shi Bo, AmirEmad Ghassami, and Debarghya Mukherjee. A debiased estimator for the mediation functional in an ultra-high-dimensional setting in the presence of interaction effects, 2024.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

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

Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4): 603–680, 2013.

Jianqing Fan, Yuan Ke, and Kaizheng Wang. Factor-adjusted regularized model selection. *Journal of Econometrics*, 216(1):71–85, 2020.

Jianqing Fan and Yuan Liao. Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *Journal of the American Statistical Association*, 117(538):909–924, 2022.

Gill Livingston, Jonathan Huntley, Kathy Y Liu, Sergi G Costafreda, Geir Selbæk, Suvarna Alladi, David Ames, Sube Banerjee, Alistair Burns, Carol Brayne, et al. Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. *The lancet*, 404(10452):572–628, 2024.

Francesca Romana Buccellato, Marianna D'Anca, Chiara Fenoglio, Elio Scarpini, and Daniela Galimberti. Role of oxidative damage in alzheimer's disease and neurodegeneration: From pathogenic mechanisms to biomarker discovery. *Antioxidants*, 10(9):1353, 2021.

Meghri Katerji, Maria Filippova, and Penelope Duerksen-Hughes. Approaches and methods to measure oxidative stress in clinical samples: Research applications in the cancer field. *Oxidative medicine and cellular longevity*, 2019(1):1279250, 2019.

Soledad Miranda, Carlos Opazo, Luis F Larrondo, Francisco J Muñoz, Francisca Ruiz, Federico Leighton, and Nibaldo C Inestrosa. The role of oxidative stress in the toxicity induced by amyloid β-peptide in alzheimer's disease. *Progress in neurobiology*, 62(6):633–648, 2000.

Chirayu D Pandya, Kristy R Howell, and Anilkumar Pillai. Antioxidants as potential therapeutics for neuropsychiatric disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 46:214–223, 2013.

Catherine N Black, Mariska Bot, Peter G Scheffer, Pim Cuijpers, and Brenda WJH Penninx. Is depression associated with increased oxidative stress? a systematic review and meta-analysis. *Psychoneuroendocrinology*, 51:164–175, 2015.

Alexei Onatski. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016, 2010.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002.

Yijiao Zhang, Yubai Yuan, Yuexia Zhang, Zhongyi Zhu, and Annie Qu. Individualized dynamic mediation analysis using latent factor models. *Journal of the American Statistical Association*, pages 1–13, 2026b.

Peter D Fransquet, Paul Lacaze, Richard Saffery, John McNeil, Robyn Woods, and Joanne Ryan. Blood dna methylation as a potential biomarker of dementia: a systematic review. *Alzheimer's & Dementia*, 14(1): 81–103, 2018.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005.

Sherry A. Greenberg. The geriatric depression scale (GDS). Try This: Best Practices in Nursing Care to Older Adults 4, Hartford Institute for Geriatric Nursing, New York University, 2012. Pages 1–2.

Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.

Martin J Aryee, Andrew E Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium dna methylation microarrays. *Bioinformatics*, 30(10):1363–1369, 2014.

Lanyu Zhang, Tiago C Silva, Juan I Young, Lissette Gomez, Michael A Schmidt, Kara L Hamilton-Nelson, Brian W Kunkle, Xi Chen, Eden R Martin, and Lily Wang. Epigenome-wide meta-analysis of dna methylation differences in prefrontal cortex implicates the immune processes in alzheimer's disease. *Nature communications*, 11(1):6114, 2020.

Katie Lunnon, Rebecca Smith, Eilis Hannon, Philip L De Jager, Gyan Srivastava, Manuela Volta, Claire Troakes, Safa Al-Sarraj, Joe Burrage, Ruby Macdonald, et al. Methylomic profiling implicates cortical deregulation of ank1 in alzheimer's disease. *Nature neuroscience*, 17(9):1164–1170, 2014.

Thomas Battram, Paul Yousefi, Gemma Crawford, Claire Prince, Mahsa Sheikhali Babaei, Gemma Sharp, Charlie Hatcher, María Jesús Vega-Salas, Sahar Khodabakhsh, Oliver Whitehurst, et al. The ewas catalog: a database of epigenome-wide association studies. *Wellcome open research*, 7:41, 2022.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2 edition, 2026.

A Appendix

This appendix consists the proofs of the two main results, along with notation and some auxiliary lemmas they rely on. Section B proves the identification theorem, Theorem 3.1 from Section 3. The argument is carried out at the population level and is built around the projected functional $\theta_{0,p,q} = T_1 + T_3T_2$, whose approximation error we control through three remainder terms R_α, R_γ, and R_δ. Section C proves the asymptotic normality theorem, Theorem 5.2 from Section 5, which gives the sampling theory for our cross-fitted estimator. The supporting results for identification and asymptotic analysis are gathered in Sections A.2 and C.1, respectively. The two matrix identities used in the identification proof are collected in Section D.1. At the start of each proof section, we include a dependency chart showing how the lemmas fit together.

The notation of Section A.1 is used throughout both proofs. Any change is specified at the point where it is introduced. The identification argument involves no sampling. The asymptotic analysis uses the two-fold cross-fitting scheme of Algorithm 1, where the projection matrices and the regression coefficients are estimated on separate folds and then the two fold-specific estimators are then averaged. We describe full details in the Section C.

A.1 Notation.

For notational convenience, define $\mathbf{f}_i = (\mathbf{f}_{X_i}^\top, \mathbf{f}_{M_i}^\top)^\top \in \mathbb{R}^{\bar{p}+\bar{q}}$ and $\mathbf{u}_i = (\mathbf{u}_{X_i}^\top, \mathbf{u}_{M_i}^\top)^\top \in \mathbb{R}^{p+q}$ be the vectors. Let $\mathbf{F}_X \in \mathbb{R}^{n\times\bar{p}}$, $\mathbf{F}_M \in \mathbb{R}^{n\times\bar{q}}$ be the matrices, and write $\mathbf{F} = (\mathbf{F}_X, \mathbf{F}_M) \in \mathbb{R}^{n\times(\bar{p}+\bar{q})}$. The augmented factor matrix is $\mathbf{F}_1 = (\mathbf{1}, \mathbf{F}) \in \mathbb{R}^{n\times(1+\bar{p}+\bar{q})}$, where $\mathbf{1} \in \mathbb{R}^n$ is the vector of ones. Let $\mathbf{H}_X \in \mathbb{R}^{\tilde{p}\times\bar{p}}$ and $\mathbf{H}_M \in \mathbb{R}^{\tilde{q}\times\bar{q}}$ be the left rectangular matrices. Define $\mathbf{H} = \mathrm{diag}(\mathbf{H}_X, \mathbf{H}_M) \in \mathbb{R}^{(\tilde{p}+\tilde{q})\times(\bar{p}+\bar{q})}$ and $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}) = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M) \in \mathbb{R}^{(1+\tilde{p}+\tilde{q})\times(1+\bar{p}+\bar{q})}$. The true design matrix is $\mathbf{Z} = \mathbf{F}_1\mathbf{H}_1^\top \in \mathbb{R}^{n\times(1+\tilde{p}+\tilde{q})}$, or equivalently $\mathbf{Z}^\top = \mathbf{H}_1\mathbf{F}_1^\top \in \mathbb{R}^{(1+\tilde{p}+\tilde{q})\times n}$.

Let $\Delta = (\alpha_1, \beta_1^\top, \gamma_1^\top)^\top \in \mathbb{R}^{1+\bar{p}+\bar{q}}$, where $\alpha_1 \in \mathbb{R}$, $\beta_1 \in \mathbb{R}^{\bar{p}}$, and $\gamma_1 \in \mathbb{R}^{\bar{q}}$ be true coefficients. The corresponding coefficient vector in the estimated factor scale is

$$\delta = (\mathbf{H}_1^\dagger)^\top\Delta = \begin{pmatrix} \alpha_1 \\ (\mathbf{H}_X^\dagger)^\top\beta_1 \\ (\mathbf{H}_M^\dagger)^\top\gamma_1 \end{pmatrix} \in \mathbb{R}^{1+\tilde{p}+\tilde{q}}.$$

The estimated coefficient vector is $\hat{\delta} = (\hat{\alpha}_1, \hat{\beta}_1^\top, \hat{\gamma}_1^\top)^\top \in \mathbb{R}^{1+\tilde{p}+\tilde{q}}$, where $\hat{\alpha}_1 \in \mathbb{R}$, $\hat{\beta}_1 \in \mathbb{R}^{\tilde{p}}$, and $\hat{\gamma}_1 \in \mathbb{R}^{\tilde{q}}$.

Let $\mathbf{V}_X \in \mathbb{R}^{p\times\tilde{p}}$ and $\mathbf{V}_M \in \mathbb{R}^{q\times\tilde{q}}$ be as defined in Sections 3 and 5 accordingly. Define $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X \in \mathbb{R}^{p\times\tilde{p}}$, $\mathbf{W}_M = \sqrt{q}\,\mathbf{V}_M \in \mathbb{R}^{q\times\tilde{q}}$, and $\mathbf{W} = \mathrm{diag}(p^{-1}\mathbf{W}_X, q^{-1}\mathbf{W}_M) \in \mathbb{R}^{(p+q)\times(\tilde{p}+\tilde{q})}$. Let $\mathbf{U} = (\mathbf{U}_X, \mathbf{U}_M) \in \mathbb{R}^{n\times(p+q)}$. Define the projected estimation error matrix $\mathbf{E} = \mathbf{U}\mathbf{W} \in \mathbb{R}^{n\times(\tilde{p}+\tilde{q})}$, and its augmented version $\mathbf{E}_0 = (\mathbf{0}_n, \mathbf{E}) \in \mathbb{R}^{n\times(1+\tilde{p}+\tilde{q})}$, where $\mathbf{0}_n \in \mathbb{R}^n$ is the zero column vector. For each observation, define $\mathbf{e}_i = \mathbf{W}^\top\mathbf{u}_i \in \mathbb{R}^{\tilde{p}+\tilde{q}}$ and $\mathbf{e}_{0i} = (0, \mathbf{e}_i^\top)^\top \in \mathbb{R}^{1+\tilde{p}+\tilde{q}}$. Similarly, $\mathbf{e}_{X_i} = p^{-1}\mathbf{W}_X^\top\mathbf{u}_{X_i} \in \mathbb{R}^{\tilde{p}}$, and $\mathbf{e}_{X0,i} = (0, \mathbf{e}_{X_i}^\top)^\top \in \mathbb{R}^{1+\tilde{p}}$.

The factor estimator is $\hat{\mathbf{F}} = (\hat{\mathbf{F}}_X, \hat{\mathbf{F}}_M) \in \mathbb{R}^{n\times(\tilde{p}+\tilde{q})}$, where $\hat{\mathbf{F}}_X \in \mathbb{R}^{n\times\tilde{p}}$ and $\hat{\mathbf{F}}_M \in \mathbb{R}^{n\times\tilde{q}}$. The estimated design matrix is $\hat{\mathbf{Z}} = (\mathbf{1}, \hat{\mathbf{F}}) \in \mathbb{R}^{n\times(1+\tilde{p}+\tilde{q})}$. Throughout, we use the decomposition

$$\hat{\mathbf{Z}} = \mathbf{Z} + \mathbf{E}_0, \tag{A.1}$$

where $\hat{\mathbf{Z}}$, $\mathbf{Z}$, and $\mathbf{E}_0$ all lie in $\mathbb{R}^{n\times(1+\tilde{p}+\tilde{q})}$.

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ and $\boldsymbol{\epsilon}_1 = (\epsilon_{1,1}, \ldots, \epsilon_{1,n})^\top \in \mathbb{R}^n$. The treatment-specific outcome equation is

$$\mathbf{Y} = \mathbf{Z}\delta + \boldsymbol{\epsilon}_1, \tag{A.2}$$

where $\mathbf{Z} \in \mathbb{R}^{n\times(1+\tilde{p}+\tilde{q})}$ and $\delta \in \mathbb{R}^{1+\tilde{p}+\tilde{q}}$.

For the mediator block, define the selector matrix $\mathbf{S}_{\tilde{q}} = \begin{bmatrix} 0_{\tilde{q}\times(1+\tilde{p})} & I_{\tilde{q}} \end{bmatrix} \in \mathbb{R}^{\tilde{q}\times(1+\tilde{p}+\tilde{q})}$, so that $\mathbf{S}_{\tilde{q}}$ extracts the last $\tilde{q}$ coordinates of any vector in the projected design scale. Also, define $\mathbf{P}_{\bar{q}} = \begin{bmatrix} 0_{\bar{q}\times(1+\bar{p})} & I_{\bar{q}} \end{bmatrix} \in \mathbb{R}^{\bar{q}\times(1+\bar{p}+\bar{q})}$, so that $\mathbf{H}_M^\top\mathbf{S}_{\tilde{q}} = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top$. Throughout, $e_{\mathbf{1}}$ denotes the first standard basis vector of whatever dimension the surrounding matrix product requires.

A.2 Supporting Results

This subsection collects the invariance lemma and the well-posedness lemma for the population regressions used in the proof of Theorem 3.1. The two matrix identities, the intercept-augmented block inversion formula and the Woodbury identity, are given in Section D.1. We first fix notation and write the moment identities that the lemmas and the main proof in Section B draw on repeatedly. Define

$$L = \begin{pmatrix} 1 \\ \frac{1}{p}\mathbf{W}_X^\top X \\ \frac{1}{q}\mathbf{W}_M^\top M \end{pmatrix}; \quad Q = \begin{pmatrix} 1 \\ \frac{1}{p}\mathbf{W}_X^\top X \end{pmatrix}; \quad N = \frac{1}{q}\mathbf{W}_M^\top M; \quad Z = \begin{pmatrix} 1 \\ \mathbf{H}_X\mathbf{f}_X \\ \mathbf{H}_M\mathbf{f}_M \end{pmatrix}.$$

Substituting the factor model $X = \mathbf{\Gamma}_X\mathbf{f}_X + \mathbf{u}_X$ and $M = \mathbf{\Gamma}_M\mathbf{f}_M + \mathbf{u}_M$ from Equation (2.1) into L, and using $\mathbf{H}_X = \frac{1}{p}\mathbf{W}_X^\top\mathbf{\Gamma}_X$, $\mathbf{H}_M = \frac{1}{q}\mathbf{W}_M^\top\mathbf{\Gamma}_M$ from Definition 3.1, gives the signal-plus-noise decomposition as follows:

$$L = Z + \mathbf{e}_0 = \mathbf{H}_1\mathbf{f}_1 + \mathbf{e}_0, \tag{A.3}$$

where $\mathbf{f}_1 = (1, \mathbf{f}_X^\top, \mathbf{f}_M^\top)^\top$, $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$, $e_0 = (0, \mathbf{e}_X^\top, \mathbf{e}_M^\top)^\top$, $\mathbf{e}_X = \frac{1}{p}\mathbf{W}_X^\top\mathbf{u}_X \in \mathbb{R}^{\tilde{p}}$, and $\mathbf{e}_M = \frac{1}{q}\mathbf{W}_M^\top\mathbf{u}_M \in \mathbb{R}^{\tilde{q}}$. Since Q consists of the leading $1+\tilde{p}$ coordinates of L, the same computation restricted to the X-block gives

$$Q = \mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0}, \quad \mathbf{f}_{X1} = (1, \mathbf{f}_X^\top)^\top, \quad \mathbf{H}_{X1} = \mathrm{diag}(1, \mathbf{H}_X), \quad \mathbf{e}_{X0} = (0, \mathbf{e}_X^\top)^\top. \tag{A.4}$$

We record the following moment identities, which hold for both treatment groups and are used throughout the Section B. By Assumption 3.1(ii), $\mathbb{E}(\mathbf{u}_X) = \mathbb{E}(\mathbf{u}_M) = 0$, so $\mathbb{E}(\mathbf{e}_0) = 0$, and the independence property gives $\mathbb{E}[\mathbf{f}_1\mathbf{e}_0^\top \mid A] = 0$. Also by Assumption 3.1(ii), $\mathbb{E}(\mathbf{u}_X\mathbf{u}_M^\top) = 0$, hence $\mathbb{E}(\mathbf{e}_X\mathbf{e}_M^\top) = 0$, so

$$\Sigma_{\mathbf{e}_0} := \mathbb{E}(\mathbf{e}_0\mathbf{e}_0^\top) = \mathrm{diag}(0, \Sigma_{\mathbf{e}_X}, \Sigma_{\mathbf{e}_M}), \quad \Sigma_{\mathbf{e}_X} = \tfrac{1}{p}\mathbf{V}_X^\top\Sigma_{\mathbf{u}_X}\mathbf{V}_X, \quad \Sigma_{\mathbf{e}_M} = \tfrac{1}{q}\mathbf{V}_M^\top\Sigma_{\mathbf{u}_M}\mathbf{V}_M,$$

using $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$ and $\mathbf{W}_M = \sqrt{q}\,\mathbf{V}_M$ from Definition 3.1. The conditional second moments of $\mathbf{u}_X, \mathbf{u}_M$ in Assumption 3.1(i) and (ii) do not depend on $(\mathbf{f}_X, \mathbf{f}_M, A)$, so these identities hold unconditionally. Now, restricting to $A = 0$ gives the X-block analogues

$$\mathbb{E}[\mathbf{f}_{X1}\mathbf{e}_{X0}^\top \mid A = 0] = 0, \quad \mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top] = \mathrm{diag}(0, \Sigma_{\mathbf{e}_X}).$$

The first lemma establishes that the target functional depends on the latent factors only through their diversified projections.

Lemma A.1 (Invariance under linear transformation of the factors)**.** *Assume the diversified weight matrices $\mathbf{H}_X$ and $\mathbf{H}_M$ have full column rank as in Definition 3.1. Then, θ_0 is invariant under linear transformation of the latent factors, in the sense that*

$$\mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[Y \mid \mathbf{f}_M, \mathbf{f}_X, A=1] \,\big|\, \mathbf{f}_X, A=0\big]\Big] = \mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[Y \mid \mathbf{H}_M\mathbf{f}_M, \mathbf{H}_X\mathbf{f}_X, A=1] \,\big|\, \mathbf{H}_X\mathbf{f}_X, A=0\big]\Big].$$

Proof. The result follows once we show that the two conditioning sigma-fields are equal.

$$\sigma(\mathbf{f}_X, \mathbf{f}_M) = \sigma(\mathbf{H}_X\mathbf{f}_X, \mathbf{H}_M\mathbf{f}_M), \quad \sigma(\mathbf{f}_X) = \sigma(\mathbf{H}_X\mathbf{f}_X). \tag{A.5}$$

We consider that $\mathbf{H}_X$ and $\mathbf{H}_M$ are defined as in Definition 3.1. For the first equality, since $\mathbf{H}_X\mathbf{f}_X$ and $\mathbf{H}_M\mathbf{f}_M$ are measurable functions of $\mathbf{f}_X$ and $\mathbf{f}_M$ respectively, we have that

$$\sigma(\mathbf{H}_X\mathbf{f}_X, \mathbf{H}_M\mathbf{f}_M) \subseteq \sigma(\mathbf{f}_X, \mathbf{f}_M). \tag{A.6}$$

For the reverse inclusion, the full column rank of $\mathbf{H}_X$ and $\mathbf{H}_M$ gives Moore-Penrose left inverses $\mathbf{H}_X^\dagger\mathbf{H}_X = I_{\tilde{p}}$ and $\mathbf{H}_M^\dagger\mathbf{H}_M = I_{\tilde{q}}$, so $\mathbf{f}_X = \mathbf{H}_X^\dagger(\mathbf{H}_X\mathbf{f}_X)$ and $\mathbf{f}_M = \mathbf{H}_M^\dagger(\mathbf{H}_M\mathbf{f}_M)$ are measurable functions of $\mathbf{H}_X\mathbf{f}_X$ and $\mathbf{H}_M\mathbf{f}_M$ that gives us

$$\sigma(\mathbf{f}_X, \mathbf{f}_M) \subseteq \sigma(\mathbf{H}_X\mathbf{f}_X, \mathbf{H}_M\mathbf{f}_M). \tag{A.7}$$

Here, combining (A.6) and (A.7) gives the first equality in (A.5). The second argument is the identical rather applied to the X-block alone.

Now, conditional expectation depends on the conditioning variables only through the sigma-field they generate. By equality (A.5), conditioning on $\{\mathbf{H}_M\mathbf{f}_M, \mathbf{H}_X\mathbf{f}_X\}$ is same as conditioning on $\{\mathbf{f}_M, \mathbf{f}_X\}$, and conditioning on $\mathbf{H}_X\mathbf{f}_X$ is the same as conditioning on $\mathbf{f}_X$. Thus, applying this to the inner and outer conditional expectations, we get

$$\mathbb{E}[Y \mid \mathbf{H}_M\mathbf{f}_M, \mathbf{H}_X\mathbf{f}_X, A=1] = \mathbb{E}[Y \mid \mathbf{f}_M, \mathbf{f}_X, A=1] \quad \text{a.s.},$$

and then

$$\mathbb{E}\big[\mathbb{E}[Y \mid \mathbf{H}_M\mathbf{f}_M, \mathbf{H}_X\mathbf{f}_X, A=1] \mid \mathbf{H}_X\mathbf{f}_X, A=0\big] = \mathbb{E}\big[\mathbb{E}[Y \mid \mathbf{f}_M, \mathbf{f}_X, A=1] \mid \mathbf{f}_X, A=0\big].$$

Now, taking the outer expectation of each side gives the stated equality, and we have as follows:

$$\begin{aligned}\theta_0 &= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y \mid \mathbf{H}_M\mathbf{f}_M, \mathbf{H}_X\mathbf{f}_X, A=1] \mid \mathbf{H}_X\mathbf{f}_X, A=0]] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y \mid \mathbf{f}_M, \mathbf{f}_X, A=1] \mid \mathbf{f}_X, A=0]] \\ &= \int\int \mathbb{E}[Y \mid \mathbf{f}_X = f_X, \mathbf{f}_M = f_M, A=1]\, dP(f_M \mid \mathbf{f}_X = f_X, A=0)\, dP(f_X).\end{aligned}$$

□

Recall the definition of $\theta_{0,p,q}$ in Section 3:

$$T_1 = e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1], \quad T_2 = \mathbf{S}_{\tilde{q}}\mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1] \in \mathbb{R}^{\tilde{q}},$$

$$T_3 = e_{\mathbf{1}}^\top \mathbb{E}[QQ^\top \mid A=0]^{-1}\mathbb{E}[QN^\top \mid A=0] \in \mathbb{R}^{1\times\tilde{q}}.$$

The next lemma guarantees that these population regression coefficients in the projected functional are well defined.

Lemma A.2 (Well-posedness of the population regressions)**.** *Let $\mathbf{W}_X$ and $\mathbf{W}_M$ be full-projection matrices as defined in Definition 3.1. Under Assumptions 3.1 and 3.2, the matrices $\mathbb{E}[LL^\top \mid A=1]$ and $\mathbb{E}[QQ^\top \mid A=0]$ are positive definite, and hence invertible. Thus, T_1, T_2 and T_3 are well defined.*

Proof. For any nonzero $w \in \mathbb{R}^{\tilde{p}}$, we have that

$$w^\top \Sigma_{\mathbf{e}_X} w = \frac{1}{p}(\mathbf{V}_X w)^\top \Sigma_{\mathbf{u}_X}(\mathbf{V}_X w) \geq \frac{\lambda_{\min}(\Sigma_{\mathbf{u}_X})}{p}\|\mathbf{V}_X w\|^2 \geq \frac{\lambda_{\min}(\mathbf{V}_X^\top\mathbf{V}_X)\,\lambda_{\min}(\Sigma_{\mathbf{u}_X})}{p} > 0,$$

using $\lambda_{\min}(\mathbf{V}_X^\top\mathbf{V}_X) \geq C$ from Definition 3.1 and $\lambda_{\min}(\Sigma_{\mathbf{u}_X}) \geq c > 0$ from Assumption 3.1(v). Hence, $\Sigma_{\mathbf{e}_X} \succ 0$ and the same argument gives $\Sigma_{\mathbf{e}_M} \succ 0$.

Next, we show that the augmented factor second-moment matrices are positive definite. Recall $\Omega_1 = \mathbb{E}[\mathbf{f}_1\mathbf{f}_1^\top \mid A=1]$. Now, take any nonzero vector $v = (v_0, b^\top)^\top \in \mathbb{R}^{1+\bar{p}+\bar{q}}$, where $b = (b_X^\top, b_M^\top)^\top$. We have that $v^\top\Omega_1 v = \mathbb{E}\big[(v_0 + b^\top\mathbf{f})^2 \mid A=1\big]$. If $b \neq 0$, let $\mu_{\mathbf{f}|1} = \mathbb{E}[\mathbf{f} \mid A=1]$ be the conditional mean under treatment group of the factors. Now, by the definition of variance, we have that

$$\mathrm{Var}(b^\top\mathbf{f} \mid A=1) = \mathbb{E}\Big[\big(b^\top\mathbf{f} - \mathbb{E}[b^\top\mathbf{f} \mid A=1]\big)^2 \;\Big|\; A=1\Big] = \mathbb{E}\Big[\big(b^\top(\mathbf{f} - \mu_{\mathbf{f}|1})\big)^2 \;\Big|\; A=1\Big]$$

Since $b^\top(\mathbf{f} - \mu_{\mathbf{f}|1})$ is a scalar, so we can write its square as $\big(b^\top(\mathbf{f} - \mu_{\mathbf{f}|1})\big)^2 = b^\top(\mathbf{f} - \mu_{\mathbf{f}|1})(\mathbf{f} - \mu_{\mathbf{f}|1})^\top b$. When we take the conditional expectation, we get that

$$b^\top\,\mathbb{E}\Big[(\mathbf{f} - \mu_{\mathbf{f}|1})(\mathbf{f} - \mu_{\mathbf{f}|1})^\top \;\Big|\; A=1\Big]\,b = b^\top\,\mathrm{Cov}(\mathbf{f} \mid A=1)b = b^\top\Sigma_{\mathbf{f}|A=1}\,b,$$

By Assumption 3.1(iv), we have that $\lambda_{\min}(\Sigma_{\mathbf{f}|A=1}) \geq c > 0$, and so

$$\mathrm{Var}(b^\top \mathbf{f} \mid A = 1) = b^\top \Sigma_{\mathbf{f}|A=1}\, b \geq \lambda_{\min}(\Sigma_{\mathbf{f}|A=1})\, \|b\|^2 \geq c\, \|b\|^2 > 0.$$

Since $\mathbb{E}[Z^2] \geq \mathrm{Var}(Z)$ for any Z and adding the constant v_0 does not change the variance, we get that

$$v^\top \Omega_1 v = \mathbb{E}\big[(v_0 + b^\top \mathbf{f})^2 \mid A = 1\big] \geq \mathrm{Var}(b^\top \mathbf{f} \mid A = 1) > 0.$$

If $b = 0$, then $v \neq 0$ implies $v_0 \neq 0$, so $v^\top \Omega_1 v = v_0^2 > 0$. Hence, $\Omega_1 \succ 0$. The same argument applied to $\mathbf{f}_{X1}$ shows that $\Omega_{X0} := \mathbb{E}[\mathbf{f}_{X1}\mathbf{f}_{X1}^\top \mid A = 0] \succ 0$.
By Equation A.3 and the moment identities in Section A.2, we have

$$\mathbb{E}[LL^\top \mid A = 1] = \mathbf{H}_1 \Omega_1 \mathbf{H}_1^\top + \Sigma_{\mathbf{e}_0}.$$

Let $\Omega_1^{1/2}$ denote the symmetric positive definite square root of Ω_1. For any $v = (v_0, v_X^\top, v_M^\top)^\top \in \mathbb{R}^{1+\tilde{p}+\tilde{q}}$, we obtain

$$v^\top \mathbb{E}[LL^\top \mid A = 1] v = \left\|\Omega_1^{1/2} \mathbf{H}_1^\top v\right\|^2 + v_X^\top \Sigma_{\mathbf{e}_X} v_X + v_M^\top \Sigma_{\mathbf{e}_M} v_M \geq 0.$$

Suppose this sum vanishes. Since each term is nonnegative, all three terms must be zero. The positive definiteness of $\Sigma_{\mathbf{e}_X}$ and $\Sigma_{\mathbf{e}_M}$ implies that the equalities $v_X^\top \Sigma_{\mathbf{e}_X} v_X = 0$ and $v_M^\top \Sigma_{\mathbf{e}_M} v_M = 0$ yield $v_X = 0$ and $v_M = 0$, respectively. The first term gives $\|\Omega_1^{1/2} \mathbf{H}_1^\top v\|^2 = 0$, which implies $\Omega_1^{1/2} \mathbf{H}_1^\top v = 0$. Since $\Omega_1^{1/2}$ is invertible, we obtain $\mathbf{H}_1^\top v = 0$.

Since $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$ and $v_X = v_M = 0$,

$$\mathbf{H}_1^\top v = (v_0,\, \mathbf{H}_X^\top v_X,\, \mathbf{H}_M^\top v_M)^\top = (v_0, 0, 0)^\top,$$

so $\mathbf{H}_1^\top v = 0$ forces $v_0 = 0$ as well, and hence $v = 0$. Therefore, we get that $\mathbb{E}[LL^\top \mid A = 1] \succ 0$.

For the control part, Equation A.4 gives $\mathbb{E}[QQ^\top \mid A = 0] = \mathbf{H}_{X1}\, \Omega_{X0}\, \mathbf{H}_{X1}^\top + \mathrm{diag}(0, \Sigma_{\mathbf{e}_X})$. The same argument with $v = (v_0, v_X^\top)^\top$ gives

$$v^\top \mathbb{E}[QQ^\top \mid A = 0] v = \left\|\Omega_{X0}^{1/2} \mathbf{H}_{X1}^\top v\right\|^2 + v_X^\top \Sigma_{\mathbf{e}_X} v_X \geq 0,$$

which vanishes only if $v_X = 0$ and $\mathbf{H}_{X1}^\top v = (v_0, 0)^\top = 0$, forcing $v_0 = 0$ and hence $v = 0$. Therefore, $\mathbb{E}[QQ^\top \mid A = 0] \succ 0$. Therefore, both matrices are invertible, and thus, T_1, T_2 and T_3 are well defined. □

The block-inversion formula for the intercept and noise signal shared by $\mathbb{E}[LL^\top \mid A = 1]$ and $\mathbb{E}[QQ^\top \mid A = 0]$ is given in Lemma D.1 in Section D.1.

B Proof of Theorem 3.1: Identification

We now prove the identification theorem. We express our target functional through the three regression quantities T_1, T_2, T_3, combined as $\theta_{0,p} = T_1 + T_3 T_2$, and bound the approximation error of each part. The argument draws on several auxiliary results, invariance of the functional under linear transformation of the latent factors, which lets us work throughout with the projected factors, well-posedness of the regression quantities, its Woodbury form, and the augmented inverse. Figure 5 shows how these lemmas enter the proof.

Proof. We now prove the identification theorem. Lemma A.1 shows that $\theta_0 = \alpha_1 + \gamma_1^\top \delta_0$ is invariant under linear transformations of the latent factors, which lets us pass to the projected functional $\theta_{0,p} = T_1 + T_3 T_2$. We bound the approximation error $\theta_0 - \theta_{0,p}$ through three remainder terms R_α, R_γ and R_δ, combining their rates to obtain the rate stated in the theorem.

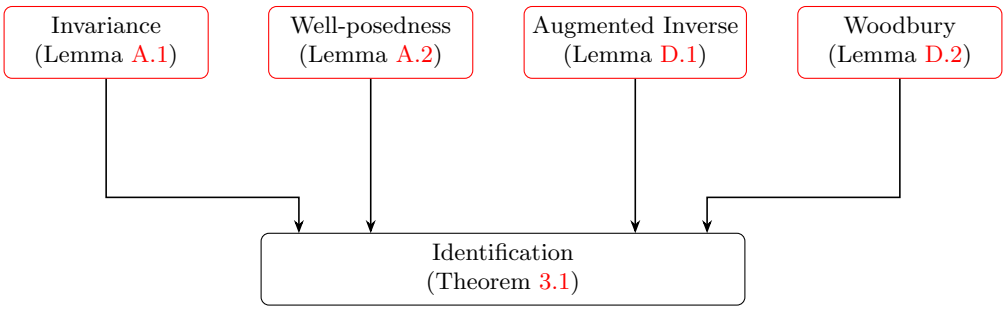


Figure 5: Proof dependency for Theorem 3.1. An arrow from one result to another means the first is used in proving the second.

The proof has five steps. Step 1 sets out the functional $\theta_{0,p} = T_1 + T_3T_2$ and the roles of its three pieces. Step 2 reduces the approximation error to the three remainders R_α, R_γ and R_δ, states the rate for each, and shows how the rates combine. Steps 3, 4, and 5 then prove the rates for R_α (via T_1), R_γ (via T_2), and R_δ (via T_3) goes to zero.

By Lemma A.1, with $\mathbf{H}_X$ and $\mathbf{H}_M$ full column rank by Definition 3.1, it suffices to work with the transformed factors $\mathbf{H}_X\mathbf{f}_X$ and $\mathbf{H}_M\mathbf{f}_M$ throughout the remainder of the proof. The notations $L, Q, N, Z, \mathbf{e}_0$, $\mathbf{H}_1, \mathbf{f}_1, \mathbf{H}_{X1}, \mathbf{f}_{X1}, \mathbf{e}_{X0}$ and the decompositions (A.3)–(A.4), together with the moment identities $\mathbb{E}[\mathbf{f}_1\mathbf{e}_0^\top \mid A] = 0$, $\Sigma_{\mathbf{e}_0} = \mathrm{diag}(0, \Sigma_{\mathbf{e}_X}, \Sigma_{\mathbf{e}_M})$, and $\mathbb{E}[\mathbf{f}_{X1}\mathbf{e}_{X0}^\top \mid A = 0] = 0$, are all as established in Section A.2.

Step 1: The projected functional.
We write $\theta_{0,p}$ in terms of the distributions of L, Q, and N as

$$\begin{aligned}\theta_{0,p} &= e_1^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1] \\ &\quad + (e_1^\top \mathbb{E}[QQ^\top \mid A=0]^{-1}\mathbb{E}[QN^\top \mid A=0])(\mathbf{S}_{\tilde{q}}\mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1]) \\ &:= T_1 + T_3T_2\ ,\end{aligned}$$

By Lemma A.2, the matrices $\mathbb{E}[LL^\top \mid A=1]$ and $\mathbb{E}[QQ^\top \mid A=0]$ are invertible, so T_1, T_2 and T_3 are well defined. The product T_3T_2 is scalar. The term T_1 corresponds to the identification of α_1. The vector T_2 lies in the projected mediator scale, so the object estimating the original coefficient γ_1 is $\mathbf{H}_M^\top T_2$, and T_3 identifies the projected intercept $\delta_0^\top \mathbf{H}_M^\top$.

Step 2: Reduction of the approximation error to three remainder terms.
We will show that

$$\begin{aligned} T_1 &= \alpha_1 + R_\alpha, \qquad |R_\alpha| \lesssim C_1 \frac{1}{\min\{p\nu^2_{\min}(\mathbf{H}_X), q\nu^2_{\min}(\mathbf{H}_M)\}}, \\ \mathbf{H}_M^\top T_2 &= \gamma_1 + R_\gamma, \qquad \|R_\gamma\| \lesssim C_{2,X}\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + C_{2,M}\frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}, \\ T_3 &= \delta_0^\top \mathbf{H}_M^\top + R_\delta,\ \|R_\delta\| \lesssim C_3\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)}, \end{aligned}$$

and that R_δ factors as $R_\delta = \widetilde{R}_\delta \mathbf{H}_M^\top$ for a matrix $\widetilde{R}_\delta$ satisfying $\|\widetilde{R}_\delta\| \lesssim C_3/(p\nu^2_{\min}(\mathbf{H}_X))$. These are proved in Steps 3, 4, and 5 respectively. The constants are

$$\begin{aligned} C_1 &:= \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\frac{\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\|\,\bar{\lambda}}{\lambda_{\min}(\Sigma_{\mathbf{f}|1})}, \quad C_{2,X} := C_{H_M}\sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\,\frac{\|\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\|\,\|\beta_1\|\,\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})\,\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})}, \\ C_{2,M} &:= \sqrt{\kappa(D)}\,\frac{\|\gamma_1\|\,\bar{\lambda}_{V_M}\bar{\lambda}_{\mathbf{u}_M}}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})}, \qquad C_3 := \sqrt{\kappa(\Sigma_{\mathbf{f}_X|0})}\,\frac{\|\mu_{\mathbf{f}_{X1}|0}\|\,\|\Phi\|\,\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_{X1}|0})}. \end{aligned}$$

Here $\bar{\lambda}_{\mathbf{u}_X} := \lambda_{\max}(\Sigma_{\mathbf{u}_X})$, $\bar{\lambda}_{\mathbf{u}_M} := \lambda_{\max}(\Sigma_{\mathbf{u}_M})$, $\bar{\lambda}_{V_X} := \lambda_{\max}(\mathbf{V}_X^\top \mathbf{V}_X)$, $\bar{\lambda}_{V_M} := \lambda_{\max}(\mathbf{V}_M^\top \mathbf{V}_M)$, and $\bar{\lambda} := \max\{\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}, \bar{\lambda}_{V_M}\bar{\lambda}_{\mathbf{u}_M}\}$, all defined in Step 3. The factor C_{H_M} is the $O(1)$ bound on $\|\mathbf{H}_M^\top(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M})^{-1}\mathbf{H}_M\|$ established in Step 4, the matrix D is defined in Equation (B.3) of Step 4, and Φ is defined in Equation (B.9) of Step 5. Each constant is a fixed population quantity is bounded using Assumptions 3.1(i)-(v), and does not depend on p or q. Since $\mathbf{H}_M^\top T_2 = \gamma_1 + R_\gamma$, the error term expands as

$$\begin{aligned}
\theta_{0,p} - \theta_0 &= T_1 + T_3 T_2 - (\alpha_1 + \gamma_1^\top \delta_0) \\
&= \alpha_1 + R_\alpha + (\delta_0^\top \mathbf{H}_M^\top + R_\delta) T_2 - (\alpha_1 + \delta_0^\top \gamma_1) \\
&= \alpha_1 + R_\alpha + \delta_0^\top \mathbf{H}_M^\top T_2 + R_\delta T_2 - (\alpha_1 + \delta_0^\top \gamma_1) \\
&= \alpha_1 + R_\alpha + \delta_0^\top (\gamma_1 + R_\gamma) + R_\delta T_2 - (\alpha_1 + \delta_0^\top \gamma_1) \\
&= R_\alpha + \delta_0^\top R_\gamma + R_\delta T_2.
\end{aligned}$$

The matrix $\mathbf{H}_M^\top$ acts on $T_2 \in \mathbb{R}^{\tilde{q}}$, and since $\mathbf{H}_M$ has full column rank with $\bar{q}$, the kernel of $\mathbf{H}_M^\top$ is nontrivial. The relation $\mathbf{H}_M^\top T_2 = \gamma_1 + R_\gamma$ determines only the component of T_2 lying in range$(\mathbf{H}_M)$, and leaves the component of T_2 in $\mathsf{ker}(\mathbf{H}_M^\top)$ unconstrained. Thus, the final term $R_\delta T_2$ must be controlled without separately bounding $\|T_2\|$. This can shown using the factorization $R_\delta = \widetilde{R}_\delta \mathbf{H}_M^\top$ established in Step 5. After substituting this factorization gives

$$R_\delta T_2 = \widetilde{R}_\delta \, \mathbf{H}_M^\top T_2 = \widetilde{R}_\delta \, (\gamma_1 + R_\gamma)$$

and since $\|\gamma_1\| \le C$ by Assumption 3.1(ii) and $\|R_\gamma\| = O\left(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)}\right) + \left(\frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\right)$ by Step 4, we get that

$$|R_\delta T_2| \le \|\widetilde{R}_\delta\| \, (\|\gamma_1\| + \|R_\gamma\|) \lesssim \frac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)}\left(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\right)$$

Therefore,

$$\begin{aligned}
|\theta_0 - \theta_{0,p}| &\lesssim |R_\alpha| + \|\delta_0\| \, \|R_\gamma\| + |R_\delta T_2| \\
&\lesssim C_1 \frac{1}{\min\{p\nu^2_{\min}(\mathbf{H}_X),\, q\nu^2_{\min}(\mathbf{H}_M)\}} + C_2\left(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\right) \\
&\quad + C_3\left(\frac{1}{p\,\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{p\nu^2_{\min}(\mathbf{H}_X)}\left(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\right)\right) \\
&\le C_*\left(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\right)\left(1 + \frac{1}{p\nu^2_{\min}(\mathbf{H}_X)}\right).
\end{aligned}$$

Since C_1, C_2, and C_3 are independent of p and q, the stated rate in Theorem 3.1 follows. It remains to establish the three remainder bounds, which we do in Steps 3-5.

Step 3: T_1**.**
Recall the outcome regression (2.1) under the treatment group. The intercept α_1 admits the representation

$$e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A = 1]^{-1} \mathbb{E}[LY \mid A = 1] = \alpha_1 + \underbrace{(-K_1 + K_2)}_{R_\alpha},$$

where $\|R_\alpha\| = o(1)$ because $\|K_1\| = o(1)$ and $K_2 = 0$. We derive this representation as follows:

$$
\begin{aligned}
& e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY \mid A=1] \\
&\overset{(1)}{=} e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY - LL^\top\delta + LL^\top\delta \mid A=1] \\
&= e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LL^\top\delta \mid A=1] + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY - LL^\top\delta \mid A=1] \\
&\overset{(2)}{=} e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LL^\top \mid A=1]\delta + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY - LL^\top\delta \mid A=1] \\
&= e_{\mathbf{1}}^\top \delta + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY - LL^\top\delta \mid A=1] \\
&= \alpha_1 + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[LY - LL^\top\delta \mid A=1] \\
&= \alpha_1 + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L(Z^\top\delta + \epsilon_1) - LL^\top\delta \mid A=1] \\
&= \alpha_1 + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L(Z-L)^\top\delta \mid A=1] + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\epsilon_1 \mid A=1] \\
&= \alpha_1 - e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\mathbf{e}_0^\top \mid A=1]\delta + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\epsilon_1 \mid A=1] \\
&= \alpha_1 - \underbrace{e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\mathbf{e}_0^\top\delta \mid A=1]}_{K_1} + \underbrace{e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\epsilon_1 \mid A=1]}_{K_2}. \qquad \text{(B.1)}
\end{aligned}
$$

Equality (1) uses the outcome model $Y = Z^\top\delta + \epsilon_1$ from (2.1) under $A = 1$. Here, (2) uses that δ is a fixed vector, so $\mathbb{E}[LL^\top\delta \mid A=1] = \mathbb{E}[LL^\top \mid A=1]\delta$. The remaining steps use $Z - L = -\mathbf{e}_0$ from (A.3). The error term is $R_\alpha = -K_1 + K_2$. We show that K_1 and K_2 are each asymptotically negligible.

Negligibility of K_1. From $L = \mathbf{H}_1\mathbf{f}_1 + \mathbf{e}_0$,

$$
\begin{aligned}
K_1 &= e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[L\mathbf{e}_0^\top\delta \mid A=1] \\
&= e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[(\mathbf{H}_1\mathbf{f}_1 + \mathbf{e}_0)\mathbf{e}_0^\top \mid A=1]\delta \\
&= e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\mathbb{E}[\mathbf{H}_1\mathbf{f}_1\mathbf{e}_0^\top \mid A=1]\delta + e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\Sigma_{\mathbf{e}_0}\delta \\
&= e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A=1]^{-1}\Sigma_{\mathbf{e}_0}\delta,
\end{aligned}
$$

where the first term vanishes by $\mathbb{E}[\mathbf{f}_1\mathbf{e}_0^\top \mid A=1] = 0$ from Section A.2. With $\Lambda_{\mathbf{f}_1|1} := \mathbb{E}[\mathbf{f}_1\mathbf{f}_1^\top \mid A=1]$ for the uncentered second moment and $\Delta' = (\beta_1^\top, \gamma_1^\top)^\top$, we get

$$
\|K_1\| = \left\| e_{\mathbf{1}}^\top \left(\mathbf{H}_1\Lambda_{\mathbf{f}_1|1}\mathbf{H}_1^\top + \Sigma_{\mathbf{e}_0}\right)^{-1}\Sigma_{\mathbf{e}_0}\mathbf{H}_1(\mathbf{H}_1^\top\mathbf{H}_1)^{-1}\Delta \right\|.
$$

Using Lemma D.1, whose formula is stated in the mean $\mu_{\mathbf{f}|1}$ and covariance $\Sigma_{\mathbf{f}|1}$, and noting that $\Sigma_{\mathbf{e}_0} = \mathrm{diag}(0, \Sigma_{\mathbf{e}})$ be the intercept row of Δ so that only Δ' survives, we get

$$
\begin{aligned}
\|K_1\| &= \left\| -\mu_{\mathbf{f}|1}^\top\mathbf{H}^\top(\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}\Sigma_{\mathbf{e}}\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\Delta' \right\| \\
&\overset{(1)}{=} \left\| -\mu_{\mathbf{f}|1}^\top\mathbf{H}^\top(\Sigma_{\mathbf{e}}^{-1} - \Sigma_{\mathbf{e}}^{-1}\mathbf{H}(\Sigma_{\mathbf{f}|1}^{-1} + \mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1})\Sigma_{\mathbf{e}}\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\Delta' \right\| \\
&= \left\| \mu_{\mathbf{f}|1}^\top\Delta' - \mu_{\mathbf{f}|1}^\top\mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H}(\Sigma_{\mathbf{f}|1}^{-1} + \mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\Delta' \right\| \\
&= \left\| \mu_{\mathbf{f}|1}^\top\Sigma_{\mathbf{f}|1}^{-1}(\Sigma_{\mathbf{f}|1}^{-1} + \mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H})^{-1}\Delta' \right\| \\
&= \left\| \mu_{\mathbf{f}|1}^\top(\mathbf{I}_{\bar{p}+\bar{q}} + \mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H}\Sigma_{\mathbf{f}|1})^{-1}\Delta' \right\|
\end{aligned}
$$

where (1) uses the Woodbury identity of Lemma D.2. The matrix $A := \mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H}$ is symmetric positive semidefinite, but $A\Sigma_{\mathbf{f}|1}$ inside $(\mathbf{I} + A\Sigma_{\mathbf{f}|1})^{-1}$ is not symmetric. So, we set $S := \Sigma_{\mathbf{f}|1}^{1/2}A\Sigma_{\mathbf{f}|1}^{1/2}$, which is symmetric positive semidefinite,

$$
(\mathbf{I} + A\Sigma_{\mathbf{f}|1})^{-1} = \Sigma_{\mathbf{f}|1}^{-1/2}(\mathbf{I} + S)^{-1}\Sigma_{\mathbf{f}|1}^{1/2}, \quad \|(\mathbf{I} + A\Sigma_{\mathbf{f}|1})^{-1}\| \le \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\frac{1}{1 + \lambda_{\min}(S)},
$$

with $\lambda_{\min}(S) \ge \lambda_{\min}(\Sigma_{\mathbf{f}|1})\,\lambda_{\min}(A)$ and $\kappa(\Sigma_{\mathbf{f}|1}) = \lambda_{\max}(\Sigma_{\mathbf{f}|1})/\lambda_{\min}(\Sigma_{\mathbf{f}|1}) = O(1)$ by Assumption 3.1(iii). Thus, this gives us that

$$
\|K_1\| \le \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\frac{\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\|}{1 + \lambda_{\min}(\Sigma_{\mathbf{f}|1})\,\lambda_{\min}(\mathbf{H}^\top\Sigma_{\mathbf{e}}^{-1}\mathbf{H})}.
$$

It remains to bound $\lambda_{\min}(\mathbf{H}^\top \Sigma_{\mathbf{e}}^{-1}\mathbf{H})$ from below. Since $\mathbf{H} = \operatorname{diag}(\mathbf{H}_X, \mathbf{H}_M)$ and $\Sigma_{\mathbf{e}} = \operatorname{diag}(\Sigma_{\mathbf{e}_X}, \Sigma_{\mathbf{e}_M})$ are block diagonal,

$$\mathbf{H}^\top \Sigma_{\mathbf{e}}^{-1}\mathbf{H} = \operatorname{diag}\Big(\mathbf{H}_X^\top \big(\tfrac{1}{p}\mathbf{V}_X^\top \Sigma_{\mathbf{u}_X}\mathbf{V}_X\big)^{-1}\mathbf{H}_X,\ \mathbf{H}_M^\top \big(\tfrac{1}{q}\mathbf{V}_M^\top \Sigma_{\mathbf{u}_M}\mathbf{V}_M\big)^{-1}\mathbf{H}_M\Big),$$

so $\lambda_{\min}(\mathbf{H}^\top \Sigma_{\mathbf{e}}^{-1}\mathbf{H})$ is the smaller of the two blocks minimum eigenvalues. For the first block,

$$\begin{aligned}
\lambda_{\min}\big(\mathbf{H}_X^\top \big(\tfrac{1}{p}\mathbf{V}_X^\top \Sigma_{\mathbf{u}_X}\mathbf{V}_X\big)^{-1}\mathbf{H}_X\big) &\geq \lambda_{\min}(\mathbf{H}_X^\top \mathbf{H}_X)\,\lambda_{\min}\big(\tfrac{1}{p}\mathbf{V}_X^\top \Sigma_{\mathbf{u}_X}\mathbf{V}_X\big)^{-1} \\
&\geq \frac{p\,\lambda_{\min}(\mathbf{H}_X^\top \mathbf{H}_X)}{\lambda_{\max}(\mathbf{V}_X^\top \mathbf{V}_X)\,\lambda_{\max}(\Sigma_{\mathbf{u}_X})} \\
&= \frac{p\,\nu_{\min}^2(\mathbf{H}_X)}{\lambda_{\max}(\mathbf{V}_X^\top \mathbf{V}_X)\,\lambda_{\max}(\Sigma_{\mathbf{u}_X})},
\end{aligned}$$

using $\lambda_{\max}(\mathbf{V}_X^\top \mathbf{V}_X) \leq C$ from Definition 3.1, and the identical computation gives

$$\lambda_{\min}(\mathbf{H}_M^\top (\tfrac{1}{q}\mathbf{V}_M^\top \Sigma_{\mathbf{u}_M}\mathbf{V}_M)^{-1}\mathbf{H}_M) \geq \frac{(q\,\nu_{\min}^2(\mathbf{H}_M))}{\lambda_{\max}(\mathbf{V}_M^\top \mathbf{V}_M)\,\lambda_{\max}(\Sigma_{\mathbf{u}_M})}.$$

We write $\bar{\lambda}_{\mathbf{u}_X} := \lambda_{\max}(\Sigma_{\mathbf{u}_X})$, $\bar{\lambda}_{\mathbf{u}_M} := \lambda_{\max}(\Sigma_{\mathbf{u}_M})$, $\bar{\lambda}_{V_X} := \lambda_{\max}(\mathbf{V}_X^\top \mathbf{V}_X)$, and $\bar{\lambda}_{V_M} := \lambda_{\max}(\mathbf{V}_M^\top \mathbf{V}_M)$. Then

$$\begin{aligned}
\lambda_{\min}(\mathbf{H}^\top \Sigma_{\mathbf{e}}^{-1}\mathbf{H}) &\geq \min\left\{\frac{p\nu_{\min}^2(\mathbf{H}_X)}{\bar{\lambda}_{\mathbf{u}_X}\bar{\lambda}_{V_X}}, \frac{q\nu_{\min}^2(\mathbf{H}_M)}{\bar{\lambda}_{\mathbf{u}_M}\bar{\lambda}_{V_M}}\right\} \\
&\geq \frac{\min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}}{\max\{\bar{\lambda}_{\mathbf{u}_X}\bar{\lambda}_{V_X},\, \bar{\lambda}_{\mathbf{u}_M}\bar{\lambda}_{V_M}\}}.
\end{aligned}$$

We write $\bar{\lambda} := \max\{\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X},\, \bar{\lambda}_{V_M}\bar{\lambda}_{\mathbf{u}_M}\}$ and $\underline{\lambda}_f := \lambda_{\min}(\Sigma_{\mathbf{f}|1})$, and then substituting the lower bound gives

$$\begin{aligned}
\|K_1\| &\leq \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\| \cdot \frac{1}{1 + \underline{\lambda}_f \min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}/\bar{\lambda}} \\
&= \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\| \cdot \frac{\bar{\lambda}}{\bar{\lambda} + \underline{\lambda}_f \min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}} \\
&\leq \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\frac{\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\|\,\bar{\lambda}}{\underline{\lambda}_f} \cdot \frac{1}{\min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}} \\
&= C_1 \frac{1}{\min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}},
\end{aligned}$$

where the third inequality drops the positive term $\bar{\lambda}$ from the denominator, and

$$C_1 = \sqrt{\kappa(\Sigma_{\mathbf{f}|1})}\,\|\mu_{\mathbf{f}|1}\|\,\|\Delta'\|\,\bar{\lambda}/\underline{\lambda}_f$$

does not depend on p or q. Hence, we have $|K_1| = \mathcal{O}\left(\frac{1}{\min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}}\right)$ as $p\nu_{\min}^2(\mathbf{H}_X) \to \infty$ and $q\nu_{\min}^2(\mathbf{H}_M) \to \infty$.

Negligibility of K_2. From $L = Z + \mathbf{e}_0$, $\mathbb{E}[L\epsilon_1 \mid A = 1] = \mathbb{E}[Z\epsilon_1 \mid A = 1] + \mathbb{E}[\mathbf{e}_0\epsilon_1 \mid A = 1]$. The first term vanishes because $\mathbb{E}[\epsilon_1] = 0$ by Assumption 3.1(iv), so

$$\mathbb{E}[Z\epsilon_1 \mid A = 1] = \mathbb{E}\big[Z\,\mathbb{E}[\epsilon_1] \mid A = 1\big] = 0.$$

The second term vanishes by Assumption 3.1(vi), since $\mathbf{e}_0$ is a linear function of $(\mathbf{u}_X, \mathbf{u}_M)$, which are orthogonal to ϵ_1 under $A = 1$. Hence $K_2 = e_{\mathbf{1}}^\top \mathbb{E}[LL^\top \mid A = 1]^{-1} \cdot 0 = 0$ exactly. Since $R_\alpha = -K_1 + K_2$, we have that

$$|R_\alpha| \leq |K_1| + |K_2| = \mathcal{O}\left(\frac{1}{\min\{p\nu_{\min}^2(\mathbf{H}_X),\, q\nu_{\min}^2(\mathbf{H}_M)\}}\right).$$

Step 4: T_2.
Recall that the latent factors are identified only up to linear transformation. Since $\mathbf{H}_M$ has full column rank, the deterministic transformation $\mathbf{H}_M^\top$ recovers γ_1 from T_2 up to a remainder term whose rate we now establish. Using the outcome model exactly as in Step 3, we get

$$\begin{aligned}
&\mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[LY \mid A=1] \\
&\quad = \mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[LL^\top \delta \mid A=1] + \mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[L(Z-L)^\top \delta \mid A=1] \\
&\qquad + \mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[L\epsilon_1 \mid A=1] \\
&\quad = \gamma_1 - \underbrace{\mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[L\mathbf{e}_0^\top \delta \mid A=1]}_{K_3} + \underbrace{\mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \mathbb{E}[L\epsilon_1 \mid A=1]}_{K_4},
\end{aligned}$$

so that $\mathbf{H}_M^\top T_2 = \gamma_1 + \underbrace{(-K_3 + K_4)}_{R_\gamma}$. It suffices to prove the negligibility of K_3 and K_4.

Negligibility of K_3. From $L = \mathbf{H}_1 \mathbf{f}_1 + \mathbf{e}_0$ and the cancellation $\mathbb{E}[\mathbf{f}_1 \mathbf{e}_0^\top \mid A=1] = 0$ of Section A.2,

$$K_3 = \mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top \mid A=1]^{-1} \Sigma_{\mathbf{e}_0} \delta.$$

To bound the norm of K_3, write the block inverse of $\mathbb{E}[LL^\top \mid A=1]$, partitioned according to the intercept, X-block, and M-block, as

$$\begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}.$$

Since $\mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} = \mathbf{P}_M \mathbf{H}_1^\top$, where $\mathbf{S}_{\tilde{q}}$ selects the last $\tilde{q}$ projected coordinates and $\mathbf{P}_M$ selects the last $\bar{q}$ true-factor coordinates, and using $\Sigma_{\mathbf{e}_0} = \operatorname{diag}(0, \Sigma_{\mathbf{e}_X}, \Sigma_{\mathbf{e}_M})$ together with $\delta = (\alpha_1, (\mathbf{H}_X^\dagger)^\top \delta_1', (\mathbf{H}_M^\dagger)^\top \gamma_1)$ gives us that

$$\begin{aligned}
\|K_3\| &= \|\mathbf{H}_M^\top \mathbf{S}_{\tilde{q}} \mathbb{E}[LL^\top | A=1]^{-1} \Sigma_{\mathbf{e}_0} \delta\| \\
&= \left\| \mathbf{P}_{\tilde{q}} \mathbf{H}_1^\top \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \Sigma_{\mathbf{e}_0} \delta \right\| \\
&= \left\| \mathbf{P}_{\tilde{q}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mathbf{H}_X^\top & 0 \\ 0 & 0 & \mathbf{H}_M^\top \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_{\mathbf{e}_X} & 0 \\ 0 & 0 & \Sigma_{\mathbf{e}_M} \end{bmatrix} \begin{bmatrix} 0 \\ (\mathbf{H}_X^\dagger)^\top \beta_1 \\ (\mathbf{H}_M^\dagger)^\top \gamma_1 \end{bmatrix} \right\| \\
&= \left\| \mathbf{P}_{\tilde{q}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mathbf{H}_X^\top & 0 \\ 0 & 0 & \mathbf{H}_M^\top \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \begin{bmatrix} 0 \\ \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1 \\ \Sigma_{\mathbf{e}_M} (\mathbf{H}_M^\dagger)^\top \gamma_1 \end{bmatrix} \right\| \\
&= \left\| \mathbf{H}_M^\top B_{32} \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1 + \mathbf{H}_M^\top B_{33} \Sigma_{\mathbf{e}_M} (\mathbf{H}_M^\dagger)^\top \gamma_1 \right\| \\
&\le \underbrace{\|\mathbf{H}_M^\top B_{32} \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1\|}_{M_1} + \underbrace{\|\mathbf{H}_M^\top B_{33} \Sigma_{\mathbf{e}_M} (\mathbf{H}_M^\dagger)^\top \gamma_1\|}_{M_2}
\end{aligned}$$

The forms of B_{32} and B_{33} follow from Lemma D.1 and a further application of Block Inverse formula to the (X, M) partition of $\mathbf{H} \Sigma_{\mathbf{f}|1} \mathbf{H}^\top + \Sigma_{\mathbf{e}}$. Setting $A_{11} = \mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X}$, $A_{12} = \mathbf{H}_X \Sigma_{\mathbf{f}_X \mathbf{f}_M^\top|1} \mathbf{H}_M^\top$, and $A_{22} = \mathbf{H}_M \Sigma_{\mathbf{f}_M|1} \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}$, so the Block Inverse formula gives

$$\begin{aligned}
B_{32} &= -B_{33}\, \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top (\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}, \\
B_{33} &= \left(\mathbf{H}_M \Sigma_{\mathbf{f}_M|1} \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M} - \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top (\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \mathbf{H}_X \Sigma_{\mathbf{f}_X \mathbf{f}_M^\top|1} \mathbf{H}_M^\top\right)^{-1}.
\end{aligned}$$

Since M_1 and M_2 are norms, the leading minus sign in B_{32} does not affect the bounds.

Bound on M_2. We have

$$
\begin{aligned}
&\|M_2\| \qquad \text{(B.2)}\\
&= \left\|\mathbf{H}_M^\top B_{33}\Sigma_{\mathbf{e}_M}(\mathbf{H}_M^\dagger)^\top\gamma_1\right\|\\
&= \left\|\mathbf{H}_M^\top\left(\mathbf{H}_M\Sigma_{\mathbf{f}_M|1}\mathbf{H}_M^\top+\Sigma_{\mathbf{e}_M}-\mathbf{H}_M\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top+\Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1}\mathbf{H}_M^\top\right)^{-1}\Sigma_{\mathbf{e}_M}(\mathbf{H}_M^\dagger)^\top\gamma_1\right\|\\
&= \left\|\mathbf{H}_M^\top\left(\mathbf{H}_M\underbrace{\left(\Sigma_{\mathbf{f}_M|1}-\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top+\Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1}\right)}_{D}\mathbf{H}_M^\top+\Sigma_{\mathbf{e}_M}\right)^{-1}\Sigma_{\mathbf{e}_M}(\mathbf{H}_M^\dagger)^\top\gamma_1\right\|\\
&= \left\|\mathbf{H}_M^\top\left(\mathbf{H}_M D\mathbf{H}_M^\top+\Sigma_{\mathbf{e}_M}\right)^{-1}\Sigma_{\mathbf{e}_M}(\mathbf{H}_M^\dagger)^\top\gamma_1\right\|\\
&\overset{(1)}{=} \left\|\mathbf{H}_M^\top\left(\Sigma_{\mathbf{e}_M}^{-1}-\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1}+\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\right)\Sigma_{\mathbf{e}_M}\mathbf{H}_M(\mathbf{H}_M^\top\mathbf{H}_M)^{-1}\gamma_1\right\|\\
&= \left\|(\mathbf{I}_{\bar q}+\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M D)^{-1}\gamma_1\right\| \qquad \text{(B.3)}
\end{aligned}
$$

where step (1) applies the Woodbury identity of Lemma D.2 to $(\mathbf{H}_M D\mathbf{H}_M^\top+\Sigma_{\mathbf{e}_M})^{-1}$, and the remaining equalities collect terms as in the bound for $\|K_1\|$ in Step 3, using $(\mathbf{H}_M^\dagger)^\top=\mathbf{H}_M(\mathbf{H}_M^\top\mathbf{H}_M)^{-1}$, and then applying the symmetric-similarity reduction used for K_1 with $A_M:=\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M$, we get

$$
\|M_2\|\le\sqrt{\kappa(D)}\,\frac{\|\gamma_1\|}{1+\lambda_{\min}(D)\lambda_{\min}(A_M)}.
$$

The two-sided bound on D established below gives $\lambda_{\min}(D)\ge\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})$ and $\kappa(D)=O(1)$. For $\lambda_{\min}(A_M)$, since $\lambda_{\max}(\mathbf{V}_M^\top\mathbf{V}_M)\le C$, for any unit vector v, $v^\top\mathbf{V}_M^\top\Sigma_{\mathbf{u}_M}\mathbf{V}_M v\le\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}$, so $\lambda_{\max}(\Sigma_{\mathbf{e}_M})\le\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}/q$, giving $\lambda_{\min}(\Sigma_{\mathbf{e}_M}^{-1})\ge q/(\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M})$ and hence

$$
\lambda_{\min}(A_M)\ge\frac{q\,\nu_{\min}^2(\mathbf{H}_M)}{\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}}. \tag{B.4}
$$

Now, combining these,

$$
\lambda_{\min}(D)\,\lambda_{\min}(A_M)\ge\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})\cdot\frac{q\,\nu_{\min}^2(\mathbf{H}_M)}{\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}},
$$

so, since $t\mapsto 1/(1+t)$ is decreasing and dropping the positive term $\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}$ from the denominator only enlarges the bound,

$$
\|M_2\|\le C_{M_2}\,\frac{1}{q\,\nu_{\min}^2(\mathbf{H}_M)},\quad C_{M_2}:=\frac{\sqrt{\kappa(D)}\,\|\gamma_1\|\,\bar\lambda_{V_M}\bar\lambda_{\mathbf{u}_M}}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})},
$$

and C_{M_2} does not depend on p or q.

Two-sided bound on D. We now bound

$$
D=(\Sigma_{\mathbf{f}_M|1}-\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top+\Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1})
$$

above and below, uniformly in p,q. For the lower bound of D, we can simplify this expression by simply substituting the relations $\Sigma_{\mathbf{f}_M|1}=\mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top+\Sigma_{\mathbf{u}_{XM,1}}$ and $\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}=\mathbf{B}_1\Sigma_{\mathbf{f}_X|1}$ that can be easily computed

from Equation (2.1) under treatment.

$$
\begin{aligned}
\Sigma_{\mathbf{f}_M|1} &= \mathbb{E}[\mathbf{f}_M\mathbf{f}_M^\top|A=1] - \mathbb{E}[\mathbf{f}_M|A=1]\mathbb{E}[\mathbf{f}_M^\top|A=1] \\
&= \mathbb{E}[(\delta_1 + \mathbf{B}_1\mathbf{f}_X + \mathbf{u}_{XM,1})(\delta_1 + \mathbf{B}_1\mathbf{f}_X + \mathbf{u}_{XM,1})^\top|A=1] \\
&\quad - \mathbb{E}[(\delta_1 + \mathbf{B}_1\mathbf{f}_X + \mathbf{u}_{XM,1})|A=1]\mathbb{E}[(\delta_1 + \mathbf{B}_1\mathbf{f}_X + \mathbf{u}_{XM,1})^\top|A=1] \\
&= \mathbb{E}[\delta_1\delta_1^\top + \delta_1\mathbf{f}_X^\top\mathbf{B}_1^\top + \delta_1(\mathbf{u}_{XM,1})^\top + \mathbf{B}_1\mathbf{f}_X\delta_1^\top + \mathbf{B}_1\mathbf{f}_X\mathbf{f}_X^\top\mathbf{B}_1^\top + \mathbf{B}_1\mathbf{f}_X(\mathbf{u}_{XM,1})^\top + \mathbf{u}_{XM,1}\delta_1^\top \\
&\quad + \mathbf{u}_{XM,1}\mathbf{f}_X^\top\mathbf{B}_1^\top + \mathbf{u}_{XM,1}(\mathbf{u}_{XM,1})^\top|A=1] - \delta_1\delta_1^\top - \delta_1\mathbb{E}[\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top \\
&\quad - \delta_1\mathbb{E}[(\mathbf{u}_{XM,1})^\top|A=1] - \mathbf{B}_1\mathbb{E}[\mathbf{f}_X|A=1]\delta_1^\top - \mathbf{B}_1\mathbb{E}[\mathbf{f}_X|A=1]\mathbb{E}[\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top \\
&\quad - \mathbf{B}_1\mathbb{E}[\mathbf{f}_X|A=1]\mathbb{E}[(\mathbf{u}_{XM,1})^\top|A=1] + \mathbb{E}[\mathbf{u}_{XM,1}|A=1]\delta_1^\top \\
&\quad - \mathbb{E}[\mathbf{u}_{XM,1}|A=1]\mathbb{E}[\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top - \mathbb{E}[\mathbf{u}_{XM,1}|A=1]\mathbb{E}[(\mathbf{u}_{XM,1})^\top|A=1] \\
&= \mathbb{E}[\mathbf{B}_1\mathbf{f}_X\mathbf{f}_X^\top\mathbf{B}_1^\top + \mathbf{u}_{XM,1}(\mathbf{u}_{XM,1})^\top|A=1] - \mathbf{B}_1\mathbb{E}[\mathbf{f}_X|A=1]\mathbb{E}[\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top \\
&= \mathbf{B}_1\mathbb{E}[\mathbf{f}_X\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top + \mathbb{E}[\mathbf{u}_{XM,1}(\mathbf{u}_{XM,1})^\top|A=1] - \mathbf{B}_1\mathbb{E}[\mathbf{f}_X|A=1]\mathbb{E}[\mathbf{f}_X^\top|A=1]\mathbf{B}_1^\top \\
&= \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top + \Sigma_{\mathbf{u}_{XM,1}}
\end{aligned}
$$

Similarly, $\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1} = \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}$. Therefore,

$$
\begin{aligned}
&\Sigma_{\mathbf{f}_M|1} - \Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\Sigma_{\mathbf{f}_X|1}^{-1}\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1} \\
&= \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top + \Sigma_{\mathbf{u}_{XM,1}} - \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\Sigma_{\mathbf{f}_X|1}^{-1}\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top \\
&= \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top + \Sigma_{\mathbf{u}_{XM,1}} - \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top \\
&= \Sigma_{\mathbf{u}_{XM,1}}.
\end{aligned}
$$

Also, since $\Sigma_{\mathbf{f}_X|1}$ is symmetric, we have

$$
\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1} = \Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}^\top = \Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top.
$$

Now, substituting these identities into

$$
D = \Sigma_{\mathbf{f}_M|1} - \Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X\mathbf{f}_M^\top|1},
$$

we get

$$
\begin{aligned}
D &= \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top + \Sigma_{\mathbf{u}_{XM,1}} - \mathbf{B}_1\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{B}_1^\top \\
&= \Sigma_{\mathbf{u}_{XM,1}} + \mathbf{B}_1\left(\underbrace{\Sigma_{\mathbf{f}_X|1} - \Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}}_{E}\right)\mathbf{B}_1^\top \\
&= \Sigma_{\mathbf{u}_{XM,1}} + \mathbf{B}_1 E\mathbf{B}_1^\top. \qquad \text{(B.5)}
\end{aligned}
$$

Now, we first show that $E \succeq 0$. We know that $\mathbf{H}_X\mathbf{f}_X + \mathbf{e}_X$ is the X-block of L from (A.3). By Assumption 3.1(v) and the tower property, we have

$$
\mathrm{Cov}(\mathbf{H}_X\mathbf{f}_X + \mathbf{e}_X \mid A=1) = \mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X}, \qquad \mathrm{Cov}(\mathbf{f}_X, \mathbf{H}_X\mathbf{f}_X + \mathbf{e}_X \mid A=1) = \Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top.
$$

Therefore,

$$
\Omega := \mathrm{Cov}\left(\begin{pmatrix}\mathbf{f}_X \\ \mathbf{H}_X\mathbf{f}_X + \mathbf{e}_X\end{pmatrix} \,\middle|\, A=1\right) = \begin{pmatrix}\Sigma_{\mathbf{f}_X|1} & \Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top \\ \mathbf{H}_X\Sigma_{\mathbf{f}_X|1} & \mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X}\end{pmatrix} \succeq 0.
$$

The bottom-right block is invertible since $\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top \succeq 0$ and $\Sigma_{\mathbf{e}_X} \succ 0$. In particular, for any $w \neq 0$,

$$
w^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|1}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})w \geq w^\top\Sigma_{\mathbf{e}_X}w > 0.
$$

Now, fix a and let

$$b := -(\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \mathbf{H}_X \Sigma_{\mathbf{f}_X|1} a.$$

Since $\Omega \succeq 0$, we have

$$\begin{aligned} 0 &\leq \begin{pmatrix} a \\ b \end{pmatrix}^\top \Omega \begin{pmatrix} a \\ b \end{pmatrix} \\ &= a^\top \Sigma_{\mathbf{f}_X|1} a - a^\top \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top (\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \mathbf{H}_X \Sigma_{\mathbf{f}_X|1} a \\ &= a^\top E a. \end{aligned}$$

Since a was arbitrary, $E \succeq 0$. Also, for any w,

$$w^\top (\mathbf{B}_1 E \mathbf{B}_1^\top) w = (\mathbf{B}_1^\top w)^\top E (\mathbf{B}_1^\top w) \geq 0,$$

so $\mathbf{B}_1 E \mathbf{B}_1^\top \succeq 0$. Thus, from (B.5),

$$D - \Sigma_{\mathbf{u}_{XM,1}} = \mathbf{B}_1 E \mathbf{B}_1^\top \succeq 0.$$

For the lower bound, let v be a unit eigenvector of D corresponding to $\lambda_{\min}(D)$. Then,

$$\begin{aligned} \lambda_{\min}(D) &= v^\top D v \\ &= v^\top \Sigma_{\mathbf{u}_{XM,1}} v + v^\top (D - \Sigma_{\mathbf{u}_{XM,1}}) v \\ &\geq v^\top \Sigma_{\mathbf{u}_{XM,1}} v \\ &\geq \lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}}). \end{aligned}$$

Hence, using Assumption 3.1(v).

$$\lambda_{\min}(D) \geq \lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}}) \geq c > 0,$$

For the upper bound, for any matrix $\mathbf{C}$ and symmetric positive definite K,

$$a^\top \mathbf{C}^\top K^{-1} \mathbf{C} a = (\mathbf{C}a)^\top K^{-1} (\mathbf{C}a) \geq 0,$$

so $\mathbf{C}^\top K^{-1} \mathbf{C} \succeq 0$. Now, take $\mathbf{C} = \mathbf{H}_X \Sigma_{\mathbf{f}_X|1}$ and $K = \mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X}$. Then the subtracted term in E is positive semidefinite, and hence $E \preceq \Sigma_{\mathbf{f}_X|1}$. Therefore,

$$\mathbf{B}_1 E \mathbf{B}_1^\top \preceq \mathbf{B}_1 \Sigma_{\mathbf{f}_X|1} \mathbf{B}_1^\top.$$

Using (B.5) and $\Sigma_{\mathbf{f}_M|1} = \mathbf{B}_1 \Sigma_{\mathbf{f}_X|1} \mathbf{B}_1^\top + \Sigma_{\mathbf{u}_{XM,1}}$, we have

$$\begin{aligned} D &= \Sigma_{\mathbf{u}_{XM,1}} + \mathbf{B}_1 E \mathbf{B}_1^\top \\ &\preceq \Sigma_{\mathbf{u}_{XM,1}} + \mathbf{B}_1 \Sigma_{\mathbf{f}_X|1} \mathbf{B}_1^\top \\ &= \Sigma_{\mathbf{f}_M|1}. \end{aligned}$$

Let u be a unit eigenvector of D corresponding to $\lambda_{\max}(D)$. Then, using Assumption 3.1(iii),

$$\begin{aligned} \lambda_{\max}(D) = u^\top D u &= u^\top \Sigma_{\mathbf{f}_M|1} u - u^\top (\Sigma_{\mathbf{f}_M|1} - D) u \\ &\leq u^\top \Sigma_{\mathbf{f}_M|1} u \\ &\leq \lambda_{\max}(\Sigma_{\mathbf{f}_M|1}) \\ &\leq C, \end{aligned}$$

Lastly, combining the two bounds, we get

$$0 < c \leq \lambda_{\min}(D) \leq \lambda_{\max}(D) \leq C < \infty, \quad \kappa(D) = \frac{\lambda_{\max}(D)}{\lambda_{\min}(D)} \leq \frac{C}{c} = O(1). \tag{B.6}$$

Since $\lambda_{\min}(D) > 0$, $D \succ 0$, which justifies the use of $D^{1/2}$ and $D^{-1/2}$ above.

Bound on M_1. We proceed in similar way as for M_2. First, we substitute the expression for B_{32} and applying the Woodbury identity of Lemma D.2 to $(\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}$, and then collecting terms exactly as in the bound for $\|K_1\|$ in Step 3, we get that

$$\begin{aligned}
\|M_1\| &= \left\|\mathbf{H}_M^\top B_{32} \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \left(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\right)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top (\mathbf{H}_X \Sigma_{\mathbf{f}_X|1} \mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1\right\| \\
&\overset{(1)}{=} \left\|\mathbf{H}_M^\top \left(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\right)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top \big(\Sigma_{\mathbf{e}_X}^{-1} - \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X (\Sigma_{\mathbf{f}_X|1}^{-1} \right. \\
&\quad \left. + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1}\big) \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1\right\|.
\end{aligned}$$

Now, expanding the inner factor and collecting terms exactly as in the derivation of $\|K_1\|$ in Step 3, we get that

$$\begin{aligned}
&\|M_1\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \Sigma_{\mathbf{e}_X} (\mathbf{H}_X^\dagger)^\top \beta_1 - \mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \right. \\
&\quad \left. \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \Sigma_{\mathbf{e}_X} \mathbf{H}_X (\mathbf{H}_X^\top \mathbf{H}_X)^{-1} \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \beta_1 \right. \\
&\quad \left. - \mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \beta_1 - \mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \right. \\
&\quad \left. \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X - \Sigma_{\mathbf{f}_X|1}^{-1}) (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \beta_1 - \mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \right. \\
&\quad (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X) (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \beta_1 + \mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \\
&\quad \left. \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \Sigma_{\mathbf{f}_X|1}^{-1} (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} \Sigma_{\mathbf{f}_X|1}^{-1} (\Sigma_{\mathbf{f}_X|1}^{-1} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X)^{-1} \beta_1\right\| \\
&= \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M \Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} (\mathbf{I}_{\bar{p}} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X \Sigma_{\mathbf{f}_X|1})^{-1} \beta_1\right\| \\
&\leq \left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M\right\| \cdot \left\|\Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1} (\mathbf{I}_{\bar{p}} + \mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X \Sigma_{\mathbf{f}_X|1})^{-1} \beta_1\right\| \qquad \text{(B.7)}
\end{aligned}$$

We'll apply the same symmetric similarity argument as for K_1 above, we get that

$$\begin{aligned}
\|M_1\| \leq \left\|\mathbf{H}_M^\top (\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M})^{-1} \mathbf{H}_M\right\| \cdot \sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\, \| \\
\Sigma_{\mathbf{f}_M \mathbf{f}_X^\top|1}\| \, \|\beta_1\| \, \frac{\bar{\lambda}_{V_X} \bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})} \cdot \frac{1}{p \nu_{\min}^2(\mathbf{H}_X)}.
\end{aligned}$$

The last equality comes from $\lambda_{\min}(\mathbf{H}_X^\top \Sigma_{\mathbf{e}_X}^{-1} \mathbf{H}_X) \geq \frac{p \nu_{\min}^2(\mathbf{H}_X)}{\bar{\lambda}_{V_X} \bar{\lambda}_{\mathbf{u}_X}}$ as derived earlier. It remains to show the leading factor is $\left\|\mathbf{H}_M^\top \Big(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1} \mathbf{H}_M\right\| = O(1)$. Applying the Woodbury identity to $(\mathbf{H}_M D \mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M})^{-1}$

and collecting terms, we get that

$$
\begin{aligned}
&\|\mathbf{H}_M^\top\big(\mathbf{H}_M D\mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\big)^{-1}\mathbf{H}_M\| \\
&= \|\mathbf{H}_M^\top(\Sigma_{\mathbf{e}_M}^{-1} - \Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1})\mathbf{H}_M\| \\
&= \|\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M - \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M\| \\
&= \|\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M - \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}(D^{-1} - D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)\| \\
&= \|\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M - \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M) \\
&\quad + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(D^{-1} + \mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}D^{-1}\| = \|\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \\
&= \|D^{-1}D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \\
&= \|D^{-1}(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M - \mathbf{I}_{\bar{q}})(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \\
&= \|D^{-1}(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1} - D^{-1}\mathbf{I}_{\bar{q}}(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \\
&\leq \|D^{-1}\| + \|D^{-1}(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \leq \frac{1}{\lambda_{\min}(D)} + \|D^{-1}(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\| \\
&\leq \frac{1}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})} + \frac{1}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})}\cdot\|(\mathbf{I}_{\bar{q}} + D\mathbf{H}_M^\top\Sigma_{\mathbf{e}_M}^{-1}\mathbf{H}_M)^{-1}\|
\end{aligned}
$$

Now, using the similar steps as for K_1, we get that

$$
\begin{aligned}
\|\mathbf{H}_M^\top\big(\mathbf{H}_M D\mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\big)^{-1}\mathbf{H}_M\| &\leq \frac{1}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})} \\
&+ \frac{1}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})}\cdot\sqrt{\kappa(D)}\cdot\frac{1}{1+\lambda_{\min}(D)\frac{(q\nu^2_{\min}(\mathbf{H}_M))}{\bar{\lambda}_{\mathbf{u}_M}\bar{\lambda}_{V_M}}}
\end{aligned}
$$

We name this upper bound as C_{D_M}. Thus, we have that $\left\|\mathbf{H}_M^\top\Big(\mathbf{H}_M D\mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M}\Big)^{-1}\mathbf{H}_M\right\| = O(1)$. Substituting this bound into the inequality (B.7), we get

$$
\begin{aligned}
\|M_1\| &\leq \|\mathbf{H}_M^\top(\mathbf{H}_M D\mathbf{H}_M^\top + \Sigma_{\mathbf{e}_M})^{-1}\mathbf{H}_M\|\cdot\sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\cdot\| \\
&\quad \Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\|\cdot\|\beta_1\|\cdot\frac{\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})}\cdot\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} \\
&\leq C_{D_M}\sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\frac{1}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})}\cdot\frac{\|\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\|\|\beta_1\|\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})}\cdot\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} \\
&\leq C_{D_M}\sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\frac{\|\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\|\|\beta_1\|\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})}\cdot\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} \\
&= C_{2,X}\cdot\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)}
\end{aligned}
$$

where $C_{2,X} = C_{D_M}\sqrt{\kappa(\Sigma_{\mathbf{f}_X|1})}\frac{\|\Sigma_{\mathbf{f}_M\mathbf{f}_X^\top|1}\|\|\beta_1\|\bar{\lambda}_{V_X}\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{u}_{XM,1}})\lambda_{\min}(\Sigma_{\mathbf{f}_X|1})}$ which does not depend on p and q. Using the bounds established above for M_1 and M_2 into $\|K_3\| \leq M_1 + M_2$, it follows that

$$
\|K_3\| \leq M_1 + M_2 = \mathcal{O}\Big(\frac{1}{p\nu^2_{\min}(\mathbf{H}_X)} + \frac{1}{q\nu^2_{\min}(\mathbf{H}_M)}\Big).
$$

For K_4. We have that

$$
K_4 = \mathbf{H}_M^\top\mathbf{S}_{\bar{q}}\mathbb{E}[LL^\top|A=1]^{-1}\mathbb{E}[L\epsilon_1|A=1] = 0
$$

By the same decomposition $L = Z + \mathbf{e}_0$ and the same two conditional-mean-zero arguments used for K_2 ($\mathbb{E}[\epsilon_1] = 0$ from Assumption 3.1(iv), and the orthogonality of $\mathbf{e}_0$ to ϵ_1 from Assumption 3.1(vi), $\mathbb{E}[L\epsilon_1 \mid$

$A = 1] = 0$ exactly. Now, left-multiplying by $\mathbf{H}_M^\top \mathbf{S}_{\bar{q}} \mathbb{E}[LL^\top \mid A = 1]^{-1}$, then gives $K_4 = 0$. Recall that $R_\gamma = -K_3 + K_4$. When we combine the bounds for both the terms, it gives us when $p\nu_{\min}^2(\mathbf{H}_X) \to \infty$ and $q\nu_{\min}^2(\mathbf{H}_M) \to \infty$,

$$\|R_\gamma\| \leq \|K_3\| + \|K_4\| = \mathcal{O}(\frac{1}{p\nu_{\min}^2(\mathbf{H}_X)} + \frac{1}{q\nu_{\min}^2(\mathbf{H}_M)})$$

Step 5: T_3**.**
To identify δ_0 up to the linear transformation $\mathbf{H}_M$, we decompose N under the control group. Using $M = \mathbf{\Gamma}_M \mathbf{f}_M + \mathbf{u}_M$ and the mediator regression $\mathbf{f}_M = \delta_0 + \mathbf{B}_0 \mathbf{f}_X + \mathbf{u}_{XM,0}$ under $A = 0$,

$$\begin{aligned}
N &= \frac{1}{q}\mathbf{W}_M^\top M = \frac{1}{q}\mathbf{W}_M^\top \Gamma_M \mathbf{f}_M + \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M \\
&= \frac{1}{q}\mathbf{W}_M^\top \Gamma_M(\delta_0 + \mathbf{B}_0\mathbf{f}_X + \mathbf{u}_{XM,0}) + \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M = \mathbf{H}_M(\delta_0 + \mathbf{B}_0\mathbf{f}_X + \mathbf{u}_{XM,0}) + \mathbf{e}_M \\
&= \mathbf{H}_M\delta_0 + \mathbf{H}_M\mathbf{B}_0\mathbf{H}_X^\dagger \mathbf{H}_X\mathbf{f}_X + \mathbf{H}_M\mathbf{u}_{XM,0} + \mathbf{e}_M \\
&= \begin{pmatrix}\mathbf{H}_M\delta_0 & \mathbf{H}_M\mathbf{B}_0\mathbf{H}_X^\dagger\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \mathbf{H}_X\end{pmatrix}\begin{pmatrix}1\\ \mathbf{f}_X\end{pmatrix} + \mathbf{H}_M\mathbf{u}_{XM,0} + \mathbf{e}_M \\
&= \phi_{XM,0}^\top \mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{XM,0}, \qquad \text{(B.8)}
\end{aligned}$$

so that $N^\top = \mathbf{f}_{X1}^\top \mathbf{H}_{X1}^\top \phi_{XM,0} + \mathbf{e}_{XM,0}^\top$, where

$$\phi_{XM,0} = \begin{pmatrix}\delta_0^\top \mathbf{H}_M^\top \\ (\mathbf{H}_M\mathbf{B}_0\mathbf{H}_X^\dagger)^\top\end{pmatrix} = (\mathbf{H}_{X1}^\dagger)^\top \Phi \mathbf{H}_M^\top, \quad \mathbf{e}_{XM,0}^\top = \mathbf{u}_{XM,0}^\top \mathbf{H}_M^\top + \mathbf{e}_M^\top,$$

and where, since $(\mathbf{H}_M\mathbf{B}_0\mathbf{H}_X^\dagger)^\top = (\mathbf{H}_X^\dagger)^\top \mathbf{B}_0^\top \mathbf{H}_M^\top$,

$$\Phi := \begin{pmatrix}\delta_0^\top \\ \mathbf{B}_0^\top\end{pmatrix} \in \mathbb{R}^{(1+\bar{p})\times\bar{q}}, \quad \Phi' := \mathbf{B}_0, \quad \|\Phi'\| = \|\mathbf{B}_0\|_{\text{op}} \leq C \qquad \text{(B.9)}$$

by Assumption 3.1(ii). The matrix Φ' denotes the lower $\bar{p}$-row block of Φ, the analogue of the truncation $\Delta' = (\beta_1^\top, \gamma_1^\top)^\top$ of $\Delta = (\alpha_1, \beta_1^\top, \gamma_1^\top)^\top$ used in Step B be as $\mathbf{H}_{X1} = \text{diag}(1, \mathbf{H}_X)$ is block diagonal, $\mathbf{H}_{X1}(\mathbf{H}_{X1}^\top \mathbf{H}_{X1})^{-1} = \text{diag}(1, \mathbf{H}_X(\mathbf{H}_X^\top \mathbf{H}_X)^{-1})$ acts on Φ block by block, and left multiplication by $\Sigma_{\mathbf{e}_{X0}} = \text{diag}(0, \Sigma_{\mathbf{e}_X})$ gives the leading $1 \times \bar{q}$ row block, leaving only the contribution of Φ'. The quantity Φ' is what appears below wherever $\Sigma_{\mathbf{e}_X}$ acts on Φ.
Define $Q := (1, p^{-1}\mathbf{W}_X^\top X)^\top$, as above. Then, we have

$$
\begin{aligned}
&e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[QN^\top|A=0] \\
&= e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[QN^\top + QQ^\top\phi_{XM,0} - QQ^\top\phi_{XM,0}|A=0] \\
&= e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[QQ^\top\phi_{XM,0}|A=0] + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[QN^\top - QQ^\top\phi_{XM,0}|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q(\mathbf{f}_{X1}^\top \mathbf{H}_{X1}^\top \phi_{XM,0} + \mathbf{e}_{XM,0}^\top) - QQ^\top\phi_{XM,0}|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q\mathbf{f}_{X1}^\top \mathbf{H}_{X1}^\top \phi_{XM,0} - QQ^\top\phi_{XM,0}|A=0] \\
&\quad + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q\mathbf{e}_{XM,0}^\top|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q(\mathbf{f}_{X1}^\top \mathbf{H}_{X1}^\top - Q^\top)\phi_{XM,0}|A=0] \\
&\quad + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q\mathbf{e}_{XM,0}^\top|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q(-\mathbf{e}_{X0}^\top)\phi_{XM,0}|A=0] \\
&\quad + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[Q(\mathbf{e}_{XM,0}^\top)|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top - e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[(\mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0})\mathbf{e}_{X0}^\top\phi_{XM,0}|A=0] \\
&\quad + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[(\mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0})(\mathbf{e}_{XM,0}^\top)|A=0] \\
&\overset{(1)}{=} \delta_0^\top \mathbf{H}_M^\top - \underbrace{e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{H}_{X1}\mathbf{f}_{X1}\mathbf{e}_{X0}^\top\phi_{XM,0}|A=0]}_{=0} \\
&\quad - e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top\phi_{XM,0}|A=0] \\
&\quad + \underbrace{e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{H}_{X1}\mathbf{f}_{X1}(\mathbf{e}_{XM,0}^\top)|A=0]}_{=0} + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{XM,0}^\top|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top - e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top\phi_{XM,0}|A=0] \\
&\quad + e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{XM,0}^\top|A=0] \\
&= \delta_0^\top \mathbf{H}_M^\top - \underbrace{e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0}}_{K_5} + \underbrace{e_1^\top \mathbb{E}[QQ^\top|A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{XM,0}^\top]}_{K_6}
\end{aligned}
$$

Here, (1) involves $\mathbf{e}_{X0}^\top\phi_{XM,0}$ or $\mathbf{e}_{XM,0}^\top$, and $\mathbf{f}_{X1}\mathbf{e}_{X0}^\top$, $\mathbf{f}_{X1}\mathbf{e}_{XM,0}^\top$ are handled identically, $\mathbb{E}[\mathbf{f}_{X1}\mathbf{e}_{X0}^\top \mid A=0]=0$ and $\mathbb{E}[\mathbf{f}_{X1}\mathbf{e}_{XM,0}^\top \mid A=0]=0$ both follow from $\mathbb{E}[\mathbf{e}_{X0} \mid \mathbf{f}_X, A=0]=0$ and $\mathbb{E}[\mathbf{e}_{XM,0} \mid \mathbf{f}_X, A=0]=0$, respectively, via the tower property. Since $\mathbf{e}_{XM,0} = \mathbf{H}_M\mathbf{u}_{XM,0} + \mathbf{e}_M$, the second of these requires, in addition to $\mathbb{E}[\mathbf{u}_X \mid \mathbf{f}_X, \mathbf{f}_M, A]=0$ (which gives $\mathbb{E}[\mathbf{e}_{X0} \mid \mathbf{f}_X, A=0]=0$ and $\mathbb{E}[\mathbf{e}_M \mid \mathbf{f}_X, A=0]=0$), the mediator-error condition $\mathbb{E}[\mathbf{u}_{XM,0} \mid \mathbf{f}_X, A=0]=0$ from Assumption 3.1(iv). So, we have

$$\mathbb{E}[\mathbf{e}_{XM,0}^\top \mid \mathbf{f}_X, A=0] = \mathbb{E}[\mathbf{u}_{XM,0}^\top \mid \mathbf{f}_X, A=0]\mathbf{H}_M^\top + \mathbb{E}[\mathbf{e}_M^\top \mid \mathbf{f}_X, A=0] = 0,$$

using Assumptions 3.1(iv) and (v). Consider that $R_\delta = -K_5 + K_6$, and we need to bound both the terms.

Negligibility of R_δ.
Recall that, we need to consider the neglibility of $\tilde{R}_\delta$ when $R_\delta = \tilde{R}_\delta \mathbf{H}_M^\top$ as mentioned in Step B. We start

as follows using the factor representation $Q = \mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0}$, we have

$$
\begin{aligned}
K_5 &= e_{\mathbf{1}}^\top \mathbb{E}[QQ^\top \mid A=0]^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0} \\
&= e_{\mathbf{1}}^\top \big(\mathbb{E}[(\mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0})(\mathbf{H}_{X1}\mathbf{f}_{X1} + \mathbf{e}_{X0})^\top \mid A=0]\big)^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0} \\
&= e_{\mathbf{1}}^\top \big(\mathbb{E}[\mathbf{H}_{X1}\mathbf{f}_{X1}\mathbf{f}_{X1}^\top\mathbf{H}_{X1}^\top + \mathbf{H}_{X1}\mathbf{f}_{X1}\mathbf{e}_{X0}^\top + \mathbf{e}_{X0}\mathbf{f}_{X1}^\top\mathbf{H}_{X1}^\top + \mathbf{e}_{X0}\mathbf{e}_{X0}^\top \mid A=0]\big)^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0} \\
&\overset{(1)}{=} e_{\mathbf{1}}^\top \big(\mathbb{E}[\mathbf{H}_{X1}\mathbf{f}_{X1}\mathbf{f}_{X1}^\top\mathbf{H}_{X1}^\top \mid A=0] + \mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top \mid A=0]\big)^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0} \\
&= e_{\mathbf{1}}^\top \big(\mathbf{H}_{X1}\mathbb{E}[\mathbf{f}_{X1}\mathbf{f}_{X1}^\top \mid A=0]\mathbf{H}_{X1}^\top + \mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top \mid A=0]\big)^{-1}\mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{X0}^\top]\phi_{XM,0} \\
&= e_{\mathbf{1}}^\top \big(\mathbf{H}_{X1}\Lambda_{\mathbf{f}_{X1}|0}\mathbf{H}_{X1}^\top + \Sigma_{\mathbf{e}_{X0}}\big)^{-1}\Sigma_{\mathbf{e}_{X0}}\phi_{XM,0} \\
&= e_{\mathbf{1}}^\top \big(\mathbf{H}_{X1}\Lambda_{\mathbf{f}_{X1}|0}\mathbf{H}_{X1}^\top + \Sigma_{\mathbf{e}_{X0}}\big)^{-1}\Sigma_{\mathbf{e}_{X0}}(\mathbf{H}_{X1}^\dagger)^\top\Phi\mathbf{H}_M^\top \\
&= e_{\mathbf{1}}^\top \big(\mathbf{H}_{X1}\Lambda_{\mathbf{f}_{X1}|0}\mathbf{H}_{X1}^\top + \Sigma_{\mathbf{e}_{X0}}\big)^{-1}\Sigma_{\mathbf{e}_{X0}}\mathbf{H}_{X1}(\mathbf{H}_{X1}^\top\mathbf{H}_{X1})^{-1}\Phi\mathbf{H}_M^\top,
\end{aligned}
$$

Here, (1) follows the conditional mean-zero condition $\mathbb{E}[\mathbf{e}_{X0} \mid \mathbf{f}_X, A=0] = 0$, which by the tower property gives $\mathbb{E}[\mathbf{f}_{X1}\mathbf{e}_{X0}^\top \mid A=0] = 0$, so the two cross terms vanish, and where $\Lambda_{\mathbf{f}_{X1}|0} := \mathbb{E}[\mathbf{f}_{X1}\mathbf{f}_{X1}^\top \mid A=0]$ is the uncentered second moment and $\phi_{XM,0} = (\mathbf{H}_{X1}^\dagger)^\top\Phi\mathbf{H}_M^\top$. We now rewrite the expression for K_5.

$$
K_5 = \underbrace{e_1^\top(\mathbf{H}_{X1}\mu_{\mathbf{f}_{X1}\mathbf{f}_{X1}^\top|0}\mathbf{H}_{X1}^\top + \Sigma_{\mathbf{e}_{X0}})^{-1}\Sigma_{\mathbf{e}_{X0}}\mathbf{H}_{X1}(\mathbf{H}_{X1}^\top\mathbf{H}_{X1})^{-1}\Phi}_{\tilde{K}_5}\mathbf{H}_M^\top = \tilde{K}_5\mathbf{H}_M^\top
$$

Now, we bound the $\tilde{K}_5$,

$$
\begin{aligned}
&\|\tilde{K}_5\| \\
&= \|e_1^\top(\mathbf{H}_{X1}\mu_{\mathbf{f}_{X1}\mathbf{f}_{X1}^\top|0}\mathbf{H}_{X1}^\top + \Sigma_{\mathbf{e}_{X0}})^{-1}\Sigma_{\mathbf{e}_{X0}}\mathbf{H}_{X1}(\mathbf{H}_{X1}^\top\mathbf{H}_{X1})^{-1}\Phi\| \\
&= \left\| e_1^\top \begin{bmatrix} (1+\mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|0}\mathbf{H}_X^\top \\ \quad +\Sigma_{\mathbf{e}_X})^{-1}\mathbf{H}_X\mu_{\mathbf{f}_X|0} & -\mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|0}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \\ (-\mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|0}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1})^\top & (\mathbf{H}_X\Sigma_{\mathbf{f}_X|0}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \Sigma_{\mathbf{e}_X}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi \end{bmatrix} \right\| \\
&= \| - \mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top(\mathbf{H}_X\Sigma_{\mathbf{f}_X|0}\mathbf{H}_X^\top + \Sigma_{\mathbf{e}_X})^{-1}\Sigma_{\mathbf{e}_X}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi\| \\
&= \| - \mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top(\Sigma_{\mathbf{e}_X}^{-1} - \Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X(\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1})\Sigma_{\mathbf{e}_X}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi\| \\
&= \big\|\mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\Sigma_{\mathbf{e}_X}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi - \mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X(\Sigma_{\mathbf{f}_X|0}^{-1} \\
&\quad + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\Sigma_{\mathbf{e}_X}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi\big\| \\
&= \|\mu_{\mathbf{f}_X|0}^\top\Phi - \mu_{\mathbf{f}_X|0}^\top\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X(\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X(\mathbf{H}_X^\top\mathbf{H}_X)^{-1}\Phi\| \\
&= \|\mu_{\mathbf{f}_X|0}^\top\Phi - \mu_{\mathbf{f}_X|0}^\top((\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X) - \Sigma_{\mathbf{f}_X|0}^{-1})(\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\Phi\| \\
&= \|\mu_{\mathbf{f}_X|0}^\top\Phi - \mu_{\mathbf{f}_X|0}^\top\Phi + \mu_{\mathbf{f}_X|0}^\top\Sigma_{\mathbf{f}_X|0}^{-1}(\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\Phi\| \\
&= \|\mu_{\mathbf{f}_X|0}^\top\Sigma_{\mathbf{f}_X|0}^{-1}(\Sigma_{\mathbf{f}_X|0}^{-1} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)^{-1}\Phi\| = \|\mu_{\mathbf{f}_X|0}^\top(\mathbf{I}_{\bar{p}} + \mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X\Sigma_{\mathbf{f}_X|0})^{-1}\Phi\| \\
&\le \sqrt{\kappa(\Sigma_{\mathbf{f}_X|0})}\frac{\|\mu_{\mathbf{f}_X|0}^\top\|\|\Phi\|}{1 + \lambda_{\min}(\Sigma_{\mathbf{f}_X|0})\lambda_{\min}(\mathbf{H}_X^\top\Sigma_{\mathbf{e}_X}^{-1}\mathbf{H}_X)}
\end{aligned}
$$

Now, we have

$$
\|\tilde{K}_5\| \le \sqrt{\kappa(\Sigma_{\mathbf{f}_X|0})}\frac{\|\mu_{\mathbf{f}_{X1}|0}^\top\|\|\Phi\|\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_{X1}|0})} \cdot \frac{1}{p\nu_{\min}^2(\mathbf{H}_X)} = C_{K_5} \cdot \frac{1}{p\nu_{\min}^2(\mathbf{H}_X)}
$$

Using the similar steps as above, we have $\|\tilde{K}_5\| = \mathcal{O}\left(\frac{1}{p\nu_{\min}^2(\mathbf{H}_X)}\right)$ where

$$
C_{K_5} = \sqrt{\kappa(\Sigma_{\mathbf{f}_X|0})}\frac{\|\mu_{\mathbf{f}_{X1}|0}^\top\|\|\Phi\|\bar{\lambda}_{\mathbf{u}_X}}{\lambda_{\min}(\Sigma_{\mathbf{f}_{X1}|0})}
$$

which doesn't depend on p.

For K_6. Since $\mathbf{e}_{X0}$ is a linear function of $\mathbf{u}_X$,

$$\begin{aligned} K_6 &= e_1^\top \mathbb{E}[QQ^\top | A=0]^{-1} \mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_{XM,0}^\top] = e_1^\top \mathbb{E}[QQ^\top | A=0]^{-1} \mathbb{E}[\mathbf{e}_{X0}(\mathbf{u}_{XM,0}^\top \mathbf{H}_M^\top + \mathbf{e}_M^\top)] \\ &= e_1^\top \mathbb{E}[QQ^\top | A=0]^{-1} \mathbb{E}[\mathbf{e}_{X0}\mathbf{u}_{XM,0}^\top \mathbf{H}_M^\top] + e_1^\top \mathbb{E}[QQ^\top | A=0]^{-1} \mathbb{E}[\mathbf{e}_{X0}\mathbf{e}_M^\top] = 0 \end{aligned}$$

the first term vanishing by Assumption 3.1(vi)(b) and the second by Assumption 3.1(vi)(a).

Recall that $R_\delta = -K_5 + K_6$, so

$$R_\delta = -K_5 = -\tilde{K}_5 \mathbf{H}_M^\top$$

For notational convenience, take $\tilde{R}_\delta = -\tilde{K}_5$, and we have $\|\tilde{R}_\delta\| = \mathcal{O}\left(\frac{1}{p\nu_{\min}^2(\mathbf{H}_X)}\right)$ as $p\nu_{\min}^2(\mathbf{H}_X) \to \infty$. □

B.1 Proof of Corollary 3.2

Proof. For $a \in \{0,1\}$, define $\mu_a := \mathbb{E}[Y^{(a)}] = \alpha_a + \gamma_a^\top \delta_a$Here each $\mu_a = \alpha_a + \gamma_a^\top \delta_a$ takes both its outcome coefficients (α_a, γ_a) and its mediator coefficient δ_a within the same arm a, so that μ_1 and μ_0 are the within-arm cases $(a_Y, a_M) = (1,1)$ and $(0,0)$, whereas $\theta_0 = \alpha_1 + \gamma_1^\top \delta_0$ of Theorem 3.1 is the cross-world case $(1,0)$. Since Assumptions 3.1(iv)–(v) and 3.1 hold for each arm separately, the argument of Theorem 3.1 applies to each with the arm index relabeled. We therefore write μ_a for this quantity in Theorem 3.1. Then, we have $\theta_{\text{ATE}} = \mu_1 - \mu_0, \quad \theta_{\text{ATE},p} = \mu_{1,p} - \mu_{0,p}$. Using the same notation as in Theorem 3.1, let

$$L = \begin{pmatrix} 1 \\ \frac{1}{p}\mathbf{W}_X^\top X \\ \frac{1}{q}\mathbf{W}_M^\top M \end{pmatrix}, \quad Q = \begin{pmatrix} 1 \\ \frac{1}{p}\mathbf{W}_X^\top X \end{pmatrix}, \quad N = \frac{1}{q}\mathbf{W}_M^\top M.$$

By the same population-level argument as in Theorem 3.1, applied separately under $A=1$ and $A=0$, $\mu_{a,p}$ admits the representation given in the corollary statement for each $a \in \{0,1\}$. It follows that

$$\theta_{\text{ATE}} - \theta_{\text{ATE},p} = (\mu_1 - \mu_{1,p}) - (\mu_0 - \mu_{0,p}),$$

so, by the triangle inequality,

$$\left|\theta_{\text{ATE}} - \theta_{\text{ATE},p}\right| \le |\mu_1 - \mu_{1,p}| + |\mu_0 - \mu_{0,p}|.$$

Applying Theorem 3.1 to $a=1$ and $a=0$, there exist constants $C_1', C_0' > 0$, independent of p and q, such that

$$|\mu_a - \mu_{a,p}| \le C_a'\left(\frac{1}{p\,\nu_{\min}^2(\mathbf{H}_X)} + \frac{1}{q\,\nu_{\min}^2(\mathbf{H}_M)}\right)\left(1 + \frac{1}{p\,\nu_{\min}^2(\mathbf{H}_X)}\right), \quad a \in \{0,1\}.$$

Combining the two bounds yields

$$\left|\theta_{\text{ATE}} - \theta_{\text{ATE},p}\right| \le (C_1' + C_0')\left(\frac{1}{p\,\nu_{\min}^2(\mathbf{H}_X)} + \frac{1}{q\,\nu_{\min}^2(\mathbf{H}_M)}\right)\left(1 + \frac{1}{p\,\nu_{\min}^2(\mathbf{H}_X)}\right).$$

As p and q grow large, the right-hand side converges to zero, so $\theta_{\text{ATE},p} \to \theta_{\text{ATE}}$ as $p\nu_{\min}^2(\mathbf{H}_X) \to \infty$ and $q\nu_{\min}^2(\mathbf{H}_M) \to \infty$. □

C Asymptotic Analysis

To prove Theorem 5.2, we first establish a few results that will be used in the main proof. Recall that FARMA uses two-fold cross-fitting. We consider one regression fold at a time, with the projection matrices

estimated using the other fold. Within the regression fold, the treated units are used to estimate $\hat{\alpha}_1$ and $\hat{\gamma}_1$, while the control units are used for the second-stage bridge regression. We derive the expansion for one fold and combine the two fold-specific estimators at the end.

For fold k, let n_k^t and n_k^c denote the numbers of treated and control units respectively. Since the two folds have approximately equal size and the treatment probabilities are bounded away from zero and one, $n_k^t \asymp n, n_k^c \asymp n$ and $n_k \asymp n$. Thus, rates written in terms of n_k^t or n_k^c are of the same order as rates in n.

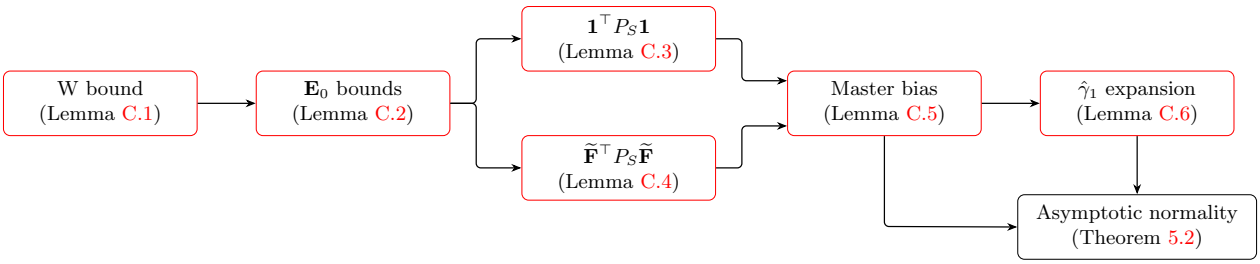


Figure 6: Proof dependency for Theorem 5.2. An arrow from one result to another means the first is used in proving the second.

The supporting results below are used in the main proof. Lemmas C.1-C.5 give the bounds needed to control the error terms, while Lemma C.6 gives the expansion of $\hat{\gamma}_1$. We use these results in the proof of Theorem 5.2.

C.1 Supporting Results

Throughout this subsection, we work conditional on the training fold, so $\mathbf{W}_X$ and $\mathbf{W}_M$ are treated as fixed. We use the notation from Section A.1.

We first prove Proposition 5.1, which shows that our normalized loading matrices are well-conditioned.

Proof of Proposition 5.1

Proof of Proposition 5.1. We first prove the result for $\mathbf{H}_X$. We write $S_X = \Gamma_X \Sigma_{\mathbf{f}_X} \Gamma_X^\top$ and $\Sigma_X = S_X + \Sigma_{\mathbf{u}_X}$. Since $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$,

$$\mathbf{H}_X = \frac{1}{p}\mathbf{W}_X^\top \Gamma_X = \frac{1}{\sqrt{p}}\mathbf{V}_X^\top \Gamma_X, \quad \mathbf{H}_X^\top \mathbf{H}_X = \frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_X} \Gamma_X,$$

where $P_{\mathbf{V}_X} = \mathbf{V}_X \mathbf{V}_X^\top$. Therefore, it is enough to bound the minimum and maximum eigenvalues of $\mathbf{H}_X^\top \mathbf{H}_X$. For notational simplicity, throughout the proof we write $n = n_k$ and relabel the observations in $\mathcal{D}_k$ as $i = 1, 2$.

Step 1: Upper bound. Since the projection statisfies $P_{\mathbf{V}_X} \preceq I_p$, we have that $\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_X} \Gamma_X \preceq \frac{1}{p}\Gamma_X^\top \Gamma_X$. Therefore, by Assumption 3.2,

$$\nu_{\max}^2(\mathbf{H}_X) = \lambda_{\max}\left(\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_X} \Gamma_X\right) \le \lambda_{\max}\left(\frac{1}{p}\Gamma_X^\top \Gamma_X\right) \le C.$$

Hence, $\nu_{\max}(\mathbf{H}_X) \le \sqrt{C}$. The rest of the proof establishes the lower bound.

Step 2: Signal spectrum and eigengap. By the law of total covariance,

$$\Sigma_{\mathbf{f}_X} = \mathbb{E}[\mathrm{Cov}(\mathbf{f}_X \mid A)] + \mathrm{Cov}[\mathbb{E}(\mathbf{f}_X \mid A)].$$

Since $A \in \{0, 1\}$, the first term is the weighted average of $\mathrm{Cov}(\mathbf{f}_X \mid A = 0)$ and $\mathrm{Cov}(\mathbf{f}_X \mid A = 1)$. Recall that $\mathbf{f} = (\mathbf{f}_X^\top, \mathbf{f}_M^\top)^\top$, so $\mathrm{Cov}(\mathbf{f}_X \mid A = a)$ is a principal submatrix of $\Sigma_{\mathbf{f}|A=a}$. Thus, by Assumption 3.1(iv),

$$cI_{\bar{p}} \preceq \mathrm{Cov}(\mathbf{f}_X \mid A = a) \preceq CI_{\bar{p}}, \qquad a \in \{0, 1\}.$$

Now, taking the weighted average over $A = 0$ and $A = 1$, we therefore have

$$cI_{\bar{p}} \preceq \mathbb{E}[\mathrm{Cov}(\mathbf{f}_X \mid A)] \preceq CI_{\bar{p}}.$$

Now consider $\mathrm{Cov}[\mathbb{E}(\mathbf{f}_X \mid A)]$. Let $\pi = \mathbb{P}(A = 1)$. Since A is binary, we have

$$\mathrm{Cov}[\mathbb{E}(\mathbf{f}_X \mid A)] = \pi(1-\pi)(\mu_{\mathbf{f}_X|1} - \mu_{\mathbf{f}_X|0})(\mu_{\mathbf{f}_X|1} - \mu_{\mathbf{f}_X|0})^\top \succeq 0.$$

By Assumption 3.1(iii), $\|\mu_{\mathbf{f}_X|a}\|_2 \leq C$ for $a \in \{0, 1\}$, and hence

$$\|\mu_{\mathbf{f}_X|1} - \mu_{\mathbf{f}_X|0}\|_2 \leq \|\mu_{\mathbf{f}_X|1}\|_2 + \|\mu_{\mathbf{f}_X|0}\|_2 \leq 2C.$$

Therefore, $\|\mathrm{Cov}[\mathbb{E}(\mathbf{f}_X \mid A)]\|_{\mathrm{op}} \leq C$, where C be the universal constant. Now, combining the two terms, we have

$$\begin{aligned}
&\lambda_{\min}(\Sigma_{\mathbf{f}_X}) \geq \lambda_{\min}\left(\mathbb{E}[\mathrm{Cov}(\mathbf{f}_X \mid A)]\right) \geq c, \\
&\lambda_{\max}(\Sigma_{\mathbf{f}_X}) \leq \lambda_{\max}\left(\mathbb{E}[\mathrm{Cov}(\mathbf{f}_X \mid A)]\right) + \lambda_{\max}\left(\mathrm{Cov}[\mathbb{E}(\mathbf{f}_X \mid A)]\right) \leq C.
\end{aligned}$$

Therefore, $c \leq \lambda_{\min}(\Sigma_{\mathbf{f}_X}) \leq \lambda_{\max}(\Sigma_{\mathbf{f}_X}) \leq C$.

Recall that $S_X = \Gamma_X \Sigma_{\mathbf{f}_X} \Gamma_X^\top$. Let $A_X = \Gamma_X \Sigma_{\mathbf{f}_X}^{1/2}$ and $B_X = \Sigma_{\mathbf{f}_X}^{1/2} \Gamma_X^\top$. Then $A_X B_X = S_X$, while $B_X A_X = \Sigma_{\mathbf{f}_X}^{1/2} \Gamma_X^\top \Gamma_X \Sigma_{\mathbf{f}_X}^{1/2}$. Since $A_X B_X$ and $B_X A_X$ have the same nonzero eigenvalues, the nonzero eigenvalues of S_X are the same as those of $\Sigma_{\mathbf{f}_X}^{1/2} \Gamma_X^\top \Gamma_X \Sigma_{\mathbf{f}_X}^{1/2}$. By Assumption 3.2, $\Gamma_X^\top \Gamma_X \succeq pc_\Gamma I_{\bar{p}}$ for some $c_\Gamma > 0$. Therefore, we get

$$\Sigma_{\mathbf{f}_X}^{1/2} \Gamma_X^\top \Gamma_X \Sigma_{\mathbf{f}_X}^{1/2} \succeq pc_\Gamma \Sigma_{\mathbf{f}_X} \succeq pc_\Gamma c_f I_{\bar{p}},$$

where $\lambda_{\min}(\Sigma_{\mathbf{f}_X}) \geq c_f > 0$ for some $c_f > 0$. Hence, $\lambda_{\min}(\Sigma_{\mathbf{f}_X}^{1/2} \Gamma_X^\top \Gamma_X \Sigma_{\mathbf{f}_X}^{1/2}) \geq pc_\Gamma c_f$. Since these are the nonzero eigenvalues of S_X, we have $\lambda_{\bar{p}}(S_X) \geq pc_\Gamma c_f$. After $c_\Gamma c_f$ into some general positive constant c, we obtain $\lambda_{\bar{p}}(S_X) \geq pc$. Also, Assumption 3.2 implies that Γ_X has full column rank $\bar{p}$. Since $\Sigma_{\mathbf{f}_X} \succ 0$, $\Sigma_{\mathbf{f}_X}^{1/2}$ is nonsingular, and therefore

$$\mathrm{rank}(S_X) = \mathrm{rank}\left(\Gamma_X \Sigma_{\mathbf{f}_X}^{1/2}\right) = \bar{p}.$$

Hence, S_X has exactly $\bar{p}$ nonzero eigenvalues, so $\lambda_{\bar{p}+1}(S_X) = 0$. Therefore, the eigengap at $\bar{p}$ satisfies

$$\lambda_{\bar{p}}(S_X) - \lambda_{\bar{p}+1}(S_X) = \lambda_{\bar{p}}(S_X) \geq pc.$$

Finally, since $S_X = \Gamma_X(\Sigma_{\mathbf{f}_X} \Gamma_X^\top)$, we have $\mathrm{Col}(S_X) \subseteq \mathrm{Col}(\Gamma_X)$. Since both spaces have dimension $\bar{p}$, it follows that $\mathrm{Col}(S_X) = \mathrm{Col}(\Gamma_X)$. Hence, the leading $\bar{p}$-dimensional eigenspace of S_X is $\mathrm{Col}(\Gamma_X)$.

Step 3: Operator-norm perturbation bound. The proxy measurement satisfies $\mathbf{x}_i = \Gamma_X \mathbf{f}_{X,i} + \mathbf{u}_{X,i}$. By Assumption 3.1(iii), $\mathbb{E}[\mathbf{f}_{X,i}] = 0$ and by Assumption 3.1(ii), $\mathbb{E}(\mathbf{u}_{X,i}) = 0$. Writing $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ and the sample covariance matrix is

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top.$$

Recall that $\delta_X = \sqrt{\frac{\log p + t}{n}} + \frac{1}{\sqrt{p}}$. Substituting the factor model and using $\Sigma_X = S_X + \Sigma_{\mathbf{u}_X}$, we have

$$
\begin{aligned}
\hat{\Sigma}_X - S_X &= \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top - S_X \\
&= \frac{1}{n}\sum_{i=1}^n \big(\Gamma_X \mathbf{f}_{X,i} + \mathbf{u}_{X,i}\big)\big(\Gamma_X \mathbf{f}_{X,i} + \mathbf{u}_{X,i}\big)^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top - S_X \\
&= \Gamma_X\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{f}_{X,i}^\top\Big)\Gamma_X^\top + \Gamma_X\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{u}_{X,i}^\top\Big) \\
&\quad + \Big(\frac{1}{n}\sum_{i=1}^n \mathbf{u}_{X,i}\mathbf{f}_{X,i}^\top\Big)\Gamma_X^\top + \frac{1}{n}\sum_{i=1}^n \mathbf{u}_{X,i}\mathbf{u}_{X,i}^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top - \Gamma_X\Sigma_{\mathbf{f}_X}\Gamma_X^\top \\
&= \Gamma_X\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{f}_{X,i}^\top - \Sigma_{\mathbf{f}_X}\Big)\Gamma_X^\top + \Big(\frac{1}{n}\sum_{i=1}^n \mathbf{u}_{X,i}\mathbf{u}_{X,i}^\top - \Sigma_{\mathbf{u}_X}\Big) + \Sigma_{\mathbf{u}_X} \\
&\quad + \Gamma_X\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{u}_{X,i}^\top\Big) + \Big(\Gamma_X\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{u}_{X,i}^\top\Big)^\top - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top \\
&= \underbrace{\Gamma_X\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{f}_{X,i}^\top - \Sigma_{\mathbf{f}_X}\Big)\Gamma_X^\top}_{S^{(1)}} + \underbrace{\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{u}_{X,i}\mathbf{u}_{X,i}^\top - \Sigma_{\mathbf{u}_X}\Big)}_{S^{(2)}} \\
&\quad + \underbrace{\Gamma_X\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{u}_{X,i}^\top}_{S^{(3)}} + \Big(\Gamma_X\frac{1}{n}\sum_{i=1}^n \mathbf{f}_{X,i}\mathbf{u}_{X,i}^\top\Big)^\top + \Sigma_{\mathbf{u}_X} - \bar{\mathbf{x}}\bar{\mathbf{x}}^\top
\end{aligned} \tag{C.1}
$$

Let $\mathbf{F}_X = [\mathbf{f}_{X,1}, \ldots, \mathbf{f}_{X,n}]^\top$ and $\mathbf{U}_X = [\mathbf{u}_{X,1}, \ldots, \mathbf{u}_{X,n}]^\top$. We bound each term in (C.1) separately.

For term $S^{(1)}$. The eigenvalues of $\Sigma_{\mathbf{f}_X}$ are bounded away from zero and infinity using Step 2. Define $\mathbf{k}_i = \Sigma_{\mathbf{f}_X}^{-1/2}\mathbf{f}_{X,i}$. By Assumption 3.1(i), the vectors $\mathbf{k}_i$ are independent. Also, by Assumption 3.1 (iii), $\mathbb{E}(\mathbf{k}_i) = 0$ and $\mathsf{var}(\mathbf{k}_i) = I_p$. Thus, the vectors $\mathbf{k}_i$ are isotropic. To check sub-Gaussianity, for any unit vector $\mathbf{v} \in \mathbb{R}^{\bar{p}}$,

$$
\mathbf{v}^\top\mathbf{k}_i = \mathbf{v}^\top\Sigma_{\mathbf{f}_X}^{-1/2}\mathbf{f}_{X,i} = (\Sigma_{\mathbf{f}_X}^{-1/2}\mathbf{v})^\top\mathbf{f}_{X,i}.
$$

By Assumption 5.1(i), we have that

$$
\|\mathbf{v}^\top\mathbf{k}_i\|_{\psi_2} \le C\|\Sigma_{\mathbf{f}_X}^{-1/2}\mathbf{v}\|_2 \le C\|\Sigma_{\mathbf{f}_X}^{-1/2}\|_{\mathrm{op}}\|\mathbf{v}\|_2 \le C,
$$

where the last inequality follows from $\lambda_{\min}(\Sigma_{\mathbf{f}_X}) \ge c > 0$ established in Step 2. Therefore, the vectors $\mathbf{k}_i$ are independent, mean-zero, isotropic sub-Gaussian vectors with uniformly bounded sub-Gaussian norm. Let $\mathbf{K}_X = [\mathbf{k}_1, \ldots, \mathbf{k}_n]^\top$. Since $\mathbf{f}_{X,i} = \Sigma_{\mathbf{f}_X}^{1/2}\mathbf{k}_i$, we have $\mathbf{F}_X = \mathbf{K}_X\Sigma_{\mathbf{f}_X}^{1/2}$. Hence, we have

$$
\begin{aligned}
\Big\|\frac{1}{n}\mathbf{F}_X^\top\mathbf{F}_X - \Sigma_{\mathbf{f}_X}\Big\|_{\mathrm{op}} &= \Big\|\Sigma_{\mathbf{f}_X}^{1/2}\Big(\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\Big)\Sigma_{\mathbf{f}_X}^{1/2}\Big\|_{\mathrm{op}} && \text{(C.2)} \\
&\le \|\Sigma_{\mathbf{f}_X}^{1/2}\|_{\mathrm{op}}^2\Big\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\Big\|_{\mathrm{op}} && \text{(C.3)} \\
&= \|\Sigma_{\mathbf{f}_X}\|_{\mathrm{op}}\Big\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\Big\|_{\mathrm{op}} && \text{(C.4)} \\
&\le C_f\Big\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\Big\|_{\mathrm{op}}, && \text{(C.5)}
\end{aligned}
$$

where C_f is a constant such that $\|\Sigma_{\mathbf{f}_X}\|_{\mathrm{op}} \le C_f$. Now, $\mathbf{K}_X$ is an $n \times \bar{p}$ random matrix whose rows are independent, mean-zero, isotropic sub-Gaussian vectors. Using the Equation (4.27) in the proof of Theorem 4.6.1

of Vershynin (2026) applied to $\mathbf{K}_X$, we have

$$\left\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\right\|_{\text{op}} \le C\max(\delta_X, \delta_X^2), \qquad \delta_X = C\left(\sqrt{\frac{\bar{p}}{n}} + \frac{s}{\sqrt{n}}\right).$$

Now, taking $s = \sqrt{t}$, the probability in Theorem 4.6.1 becomes $1 - 2e^{-t}$, and

$$\delta_X = C\left(\sqrt{\frac{\bar{p}}{n}} + \sqrt{\frac{t}{n}}\right) \le C\sqrt{\frac{\bar{p}+t}{n}}.$$

Therefore, with probability at least $1 - 2e^{-t}$,

$$\left\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\right\|_{\text{op}} \le C\left(\sqrt{\frac{\bar{p}+t}{n}} + \frac{\bar{p}+t}{n}\right).$$

Since $\bar{p}$ is fixed, we have $\bar{p} + t \le C(\log p + t)$ for sufficiently large p, we have:

$$\left\|\frac{1}{n}\mathbf{K}_X^\top\mathbf{K}_X - I_{\bar{p}}\right\|_{\text{op}} \le C\sqrt{\frac{\log p + t}{n}} \implies \left\|\frac{1}{n}\mathbf{F}_X^\top\mathbf{F}_X - \Sigma_{\mathbf{f}_X}\right\|_{\text{op}} \le C\sqrt{\frac{\log p + t}{n}}.$$

which follows from Equation C.2 and we absorb C_f in C. Also, by Assumption 3.2,

$$\left\|\frac{1}{p}\Gamma_X^\top\Gamma_X\right\|_{\text{op}} = \lambda_{\max}\left(\frac{1}{p}\Gamma_X^\top\Gamma_X\right) \le C \implies \|\Gamma_X\|_{\text{op}}^2 = \lambda_{\max}(\Gamma_X^\top\Gamma_X) \le Cp.$$

Finally, we have that

$$\begin{aligned}\|S^{(1)}\|_{\text{op}} = \left\|\Gamma_X\left(\frac{1}{n}\mathbf{F}_X^\top\mathbf{F}_X - \Sigma_{\mathbf{f}_X}\right)\Gamma_X^\top\right\|_{\text{op}} &\le \|\Gamma_X\|_{\text{op}}^2\left\|\frac{1}{n}\mathbf{F}_X^\top\mathbf{F}_X - \Sigma_{\mathbf{f}_X}\right\|_{\text{op}} \\ &\le Cp\sqrt{\frac{\log p + t}{n}} \le Cp\delta_X.\end{aligned}$$

For term $S^{(2)}$. Define $\mathbf{l}_i = \Sigma_{\mathbf{u}_X}^{-1/2}\mathbf{u}_{X,i}$. By Assumptions 3.1(i)-(ii), 3.1(v), and 5.1(i), the vectors $\mathbf{l}_i$ are independent, mean-zero, isotropic sub-Gaussian vectors. Let $\mathbf{L}_X = [\mathbf{l}_1, \ldots, \mathbf{l}_n]^\top$. Then

$$\frac{1}{n}\mathbf{U}_X^\top\mathbf{U}_X = \Sigma_{\mathbf{u}_X}^{1/2}\left(\frac{1}{n}\mathbf{L}_X^\top\mathbf{L}_X\right)\Sigma_{\mathbf{u}_X}^{1/2},$$

and hence

$$\begin{aligned}\|S^{(2)}\|_{\text{op}} &= \left\|\frac{1}{n}\mathbf{U}_X^\top\mathbf{U}_X - \Sigma_{\mathbf{u}_X}\right\|_{\text{op}} \\ &= \left\|\Sigma_{\mathbf{u}_X}^{1/2}\left(\frac{1}{n}\mathbf{L}_X^\top\mathbf{L}_X - I_p\right)\Sigma_{\mathbf{u}_X}^{1/2}\right\|_{\text{op}} \\ &\le \|\Sigma_{\mathbf{u}_X}\|_{\text{op}}\left\|\frac{1}{n}\mathbf{L}_X^\top\mathbf{L}_X - I_p\right\|_{\text{op}} \\ &\le C\left\|\frac{1}{n}\mathbf{L}_X^\top\mathbf{L}_X - I_p\right\|_{\text{op}}.\end{aligned}$$

By Equation (4.27) in the proof of Theorem 4.6.1 of Vershynin (2026), applied to the $n \times p$ matrix $\mathbf{L}_X$ and taking the tail parameter there to be $\sqrt{t}$, with probability at least $1 - 2e^{-t}$, we have

$$\left\|\frac{1}{n}\mathbf{L}_X^\top\mathbf{L}_X - I_p\right\|_{\text{op}} \le C\left(\sqrt{\frac{p+t}{n}} + \frac{p+t}{n}\right).$$

Using $\|\Sigma_{\mathbf{u}_X}\|_{\mathrm{op}} \le C$ from Assumption 3.1(v), we have

$$\|S^{(2)}\|_{\mathrm{op}} \le C\left(\sqrt{\frac{p+t}{n}} + \frac{p+t}{n}\right).$$

Since $\delta_X \le c_\star \le 1$, we have $(\log p + t)/n \le 1$, and hence $t \le n$. Therefore,

$$\sqrt{\frac{p+t}{n}} \le \sqrt{\frac{p}{n}} + \sqrt{\frac{t}{n}} \le C\sqrt{p} \le Cp\delta_X.$$

Also, we have

$$\frac{p+t}{n} \le \frac{p}{n} + 1 \le Cp\delta_X.$$

Thus, we get that

$$\|S^{(2)}\|_{\mathrm{op}} \le Cp\delta_X.$$

For term $S^{(3)}$. By Theorem 4.6.1 of Vershynin (2026) applied to the independent, mean-zero, isotropic sub-Gaussian rows of $\mathbf{K}_X$ and $\mathbf{L}_X$, and taking the tail parameter there to be $\sqrt{t}$, with probability at least $1 - Ce^{-t}$,

$$\|\mathbf{K}_X\|_{\mathrm{op}} \le C(\sqrt{n} + \sqrt{\bar{p}} + \sqrt{t}), \qquad \|\mathbf{L}_X\|_{\mathrm{op}} \le C(\sqrt{n} + \sqrt{p} + \sqrt{t}).$$

Since $\bar{p}$ is fixed, we have $\|\mathbf{K}_X\|_{\mathrm{op}} \le C(\sqrt{n} + \sqrt{t})$. Also, since $\mathbf{F}_X = \mathbf{K}_X\Sigma_{\mathbf{f}_X}^{1/2}$ and $\mathbf{U}_X = \mathbf{L}_X\Sigma_{\mathbf{u}_X}^{1/2}$, we have

$$\|\mathbf{F}_X\|_{\mathrm{op}} \le \|\mathbf{K}_X\|_{\mathrm{op}}\|\Sigma_{\mathbf{f}_X}^{1/2}\|_{\mathrm{op}} \le C(\sqrt{n} + \sqrt{t}),$$

and

$$\|\mathbf{U}_X\|_{\mathrm{op}} \le \|\mathbf{L}_X\|_{\mathrm{op}}\|\Sigma_{\mathbf{u}_X}^{1/2}\|_{\mathrm{op}} \le C(\sqrt{n} + \sqrt{p} + \sqrt{t}),$$

where we used the bounded eigenvalues of $\Sigma_{\mathbf{f}_X}$ and $\Sigma_{\mathbf{u}_X}$. Therefore,

$$\begin{aligned}
\|S^{(3)}\|_{\mathrm{op}} &= \left\|\Gamma_X \frac{1}{n}\mathbf{F}_X^\top \mathbf{U}_X\right\|_{\mathrm{op}} \\
&\le \|\Gamma_X\|_{\mathrm{op}} \frac{1}{n}\|\mathbf{F}_X\|_{\mathrm{op}}\|\mathbf{U}_X\|_{\mathrm{op}} \\
&\le C\frac{\sqrt{p}}{n}(\sqrt{n} + \sqrt{t})(\sqrt{n} + \sqrt{p} + \sqrt{t}),
\end{aligned}$$

where we used $\|\Gamma_X\|_{\mathrm{op}} \le C\sqrt{p}$. Since $\delta_X \le c_\star \le 1$, we have $(\log p + t)/n \le 1$, and hence $t \le n$. Therefore, $\sqrt{t} \le \sqrt{n}$, and

$$\begin{aligned}
\|S^{(3)}\|_{\mathrm{op}} &\le C\frac{\sqrt{p}}{n}\sqrt{n}(\sqrt{n} + \sqrt{p}) \\
&\le C\left(\sqrt{p} + \frac{p}{\sqrt{n}}\right).
\end{aligned}$$

For sufficiently large p, $\log p + t \ge 1$, so

$$\frac{1}{\sqrt{n}} \le \sqrt{\frac{\log p + t}{n}}.$$

Hence, we get

$$\|S^{(3)}\|_{\mathrm{op}} \le C\left\{\sqrt{p} + p\sqrt{\frac{\log p + t}{n}}\right\} = Cp\delta_X.$$

Since the operator norm is unchanged by transposition, $\|(S^{(3)})^\top\|_{\mathrm{op}} = \|S^{(3)}\|_{\mathrm{op}} \le Cp\delta_X$.

For the remaining covariance term, Assumption 3.1(v) gives $\|\Sigma_{\mathbf{u}_X}\|_{\mathrm{op}} \le C$. Since

$$p\delta_X = p\sqrt{\frac{\log p + t}{n}} + \sqrt{p} \ge \sqrt{p},$$

we have $\|\Sigma_{\mathbf{u}_X}\|_{\mathrm{op}} \le Cp\delta_X$ for $p \ge 1$.

Finally, we bound the sample mean term. Since

$$\bar{\mathbf{x}} = \Gamma_X \bar{\mathbf{f}}_X + \bar{\mathbf{u}}_X, \qquad \bar{\mathbf{f}}_X = \frac{1}{n}\sum_{i=1}^{n} \mathbf{f}_{X,i}, \qquad \bar{\mathbf{u}}_X = \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_{X,i},$$

we first bound $\bar{\mathbf{f}}_X$ and $\bar{\mathbf{u}}_X$ separately. By Assumptions 3.1(i), 3.1(iii), and 5.1(i), the vectors $\mathbf{f}_{X,i}$ are independent, mean-zero, and sub-Gaussian. Since $\bar{p}$ is fixed, with probability at least $1 - Ce^{-t}$,

$$\|\bar{\mathbf{f}}_X\|_2 \le C\sqrt{\frac{\bar{p}+t}{n}} \le C\sqrt{\frac{\log p + t}{n}}.$$

Now consider $\bar{\mathbf{u}}_X$. For each $j = 1, \dots, p$,

$$\bar{u}_{X,j} = \frac{1}{n}\sum_{i=1}^{n} u_{X,ij}.$$

By Assumptions 3.1(i)–(ii) and 5.1(i), the variables $u_{X,1j}, \dots, u_{X,nj}$ are independent, mean-zero, and sub-Gaussian. Therefore, for each fixed j,

$$\mathbb{P}\left(|\bar{u}_{X,j}| > C\sqrt{\frac{\log p + t}{n}}\right) \le 2e^{-(\log p + t)} = \frac{2e^{-t}}{p}.$$

Now, taking a union bound over $j = 1, \dots, p$, we have that

$$\mathbb{P}\left(\max_{j \le p} |\bar{u}_{X,j}| > C\sqrt{\frac{\log p + t}{n}}\right) \le 2e^{-t}.$$

Hence, with probability at least $1 - 2e^{-t}$,

$$\max_{j \le p} |\bar{u}_{X,j}| \le C\sqrt{\frac{\log p + t}{n}}.$$

Using $\|\mathbf{v}\|_2 \le \sqrt{p}\,\|\mathbf{v}\|_\infty$ for any $\mathbf{v} \in \mathbb{R}^p$, we have

$$\|\bar{\mathbf{u}}_X\|_2 \le \sqrt{p}\max_{j \le p} |\bar{u}_{X,j}| \le C\sqrt{\frac{p(\log p + t)}{n}}.$$

Now, using $\bar{\mathbf{x}} = \Gamma_X \bar{\mathbf{f}}_X + \bar{\mathbf{u}}_X$, we have

$$\begin{aligned}
\|\bar{\mathbf{x}}\|_2 &\le \|\Gamma_X \bar{\mathbf{f}}_X\|_2 + \|\bar{\mathbf{u}}_X\|_2 \\
&\le \|\Gamma_X\|_{\mathrm{op}} \|\bar{\mathbf{f}}_X\|_2 + \|\bar{\mathbf{u}}_X\|_2 \\
&\le C\sqrt{p}\sqrt{\frac{\log p + t}{n}} + C\sqrt{\frac{p(\log p + t)}{n}} \\
&\le C\sqrt{\frac{p(\log p + t)}{n}},
\end{aligned}$$

where we used $\|\Gamma_X\|_{\mathrm{op}} \le C\sqrt{p}$. Since $\bar{\mathbf{x}}\bar{\mathbf{x}}^\top$ is a rank-one matrix,

$$\|\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\|_{\mathrm{op}} = \|\bar{\mathbf{x}}\|_2^2 \le Cp\frac{\log p + t}{n}.$$

Recall that

$$\delta_X = \sqrt{\frac{\log p + t}{n}} + \frac{1}{\sqrt{p}}.$$

Since $\delta_X \le c_\star \le 1$, we have $(\log p + t)/n \le 1$. Therefore,

$$\frac{\log p + t}{n} \le \sqrt{\frac{\log p + t}{n}} \le \delta_X.$$

Hence,

$$\|\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\|_{\mathrm{op}} \le Cp\delta_X.$$

Now, let $A_{X,t}$ be the intersection of the events used in the bounds for $S^{(1)}$, $S^{(2)}$, $S^{(3)}$, $\bar{\mathbf{f}}_X$, and $\bar{\mathbf{u}}_X$. Since there are only a fixed number of such events and each has failure probability bounded by Ce^{-t}, a union bound gives

$$\mathbb{P}(A_{X,t}) \ge 1 - C_0 e^{-t}.$$

Therefore, on $A_{X,t}$, all the bounds hold simultaneously. Using (C.1) and the triangle inequality, we obtain

$$\|\hat{\Sigma}_X - S_X\|_{\mathrm{op}} \le Cp\delta_X. \tag{C.6}$$

Step 4: Eigenspace perturbation.
By Step 2, the eigengap of S_X at $\bar{p}$ satisfies

$$\lambda_{\bar{p}}(S_X) - \lambda_{\bar{p}+1}(S_X) \ge pc_\Gamma c_f,$$

and the leading $\bar{p}$-dimensional eigenspace of S_X is $\mathrm{Col}(\Gamma_X)$. Let $P_{\mathbf{V}_{X,\bar{p}}}$ denote the projection onto the leading $\bar{p}$-dimensional eigenspace of $\hat{\Sigma}_X$, and let P_{Γ_X} denote the projection onto $\mathrm{Col}(\Gamma_X)$. On the event $A_{X,t}$, Step 3 gives

$$\|\hat{\Sigma}_X - S_X\|_{\mathrm{op}} \le Cp\delta_X.$$

If $\delta_X \le c_\Gamma c_f/(2C)$, then we have

$$\|\hat{\Sigma}_X - S_X\|_{\mathrm{op}} \le \frac{1}{2}pc_\Gamma c_f \le \frac{1}{2}\{\lambda_{\bar{p}}(S_X) - \lambda_{\bar{p}+1}(S_X)\}.$$

Therefore, by the Davis–Kahan $\sin\Theta$ theorem applied to S_X and $\hat{\Sigma}_X$, we get

$$\begin{aligned}\left\|P_{\mathbf{V}_{X,\bar{p}}} - P_{\Gamma_X}\right\|_{\mathrm{op}} &\le \frac{2\|\hat{\Sigma}_X - S_X\|_{\mathrm{op}}}{\lambda_{\bar{p}}(S_X) - \lambda_{\bar{p}+1}(S_X)} \\ &\le \frac{2Cp\delta_X}{pc_\Gamma c_f} \\ &= C_6\delta_X,\end{aligned}$$

where $C_6 = 2C/(c_\Gamma c_f)$ depends only on the constants.

Step 5: Working projection.
Since $\tilde{p} \ge \bar{p}$, the columns of $\mathbf{V}_{X,\bar{p}}$ and $\mathbf{V}_X$ are the leading $\bar{p}$ and $\tilde{p}$ eigenvectors of the same matrix $\hat{\Sigma}_X$. We first show that there is a gap at $\bar{p}$ for $\hat{\Sigma}_X$. By Weyl's inequality and the bounds of the Steps 2-3, on $A_{X,t}$,

$$\begin{aligned}\lambda_{\bar{p}}(\hat{\Sigma}_X) &\ge \lambda_{\bar{p}}(S_X) - \|\hat{\Sigma}_X - S_X\|_{\mathrm{op}} \\ &\ge pc\,c_f - Cp\delta_X,\end{aligned}$$

and also,

$$\begin{aligned}\lambda_{\bar{p}+1}(\hat{\Sigma}_X) &\le \lambda_{\bar{p}+1}(S_X) + \|\hat{\Sigma}_X - S_X\|_{\mathrm{op}} \\ &\le Cp\delta_X.\end{aligned}$$

If $\delta_X \le c\,c_f/(4C)$, then

$$\lambda_{\bar{p}}(\hat{\Sigma}_X) \ge \frac{3}{4}pc\,c_f, \quad \lambda_{\bar{p}+1}(\hat{\Sigma}_X) \le \frac{1}{4}pc\,c_f.$$

Since $\tilde{p} \geq \bar{p}$, the leading $\bar{p}$-dimensional eigenspace is contained in the leading $\tilde{p}$-dimensional eigenspace, so $\mathrm{Col}(\mathbf{V}_{X,\bar{p}}) \subseteq \mathrm{Col}(\mathbf{V}_X)$. Now, writing $\mathbf{V}_X = [\mathbf{v}_1, \ldots, \mathbf{v}_{\tilde{p}}]$ and $\mathbf{V}_{X,\bar{p}} = [\mathbf{v}_1, \ldots, \mathbf{v}_{\bar{p}}]$, we have

$$P_{\mathbf{V}_X} - P_{\mathbf{V}_{X,\bar{p}}} = \sum_{j=\bar{p}+1}^{\tilde{p}} \mathbf{v}_j \mathbf{v}_j^\top.$$

For any vector $\mathbf{a}$, we have

$$\mathbf{a}^\top \left(P_{\mathbf{V}_X} - P_{\mathbf{V}_{X,\bar{p}}}\right) \mathbf{a} = \sum_{j=\bar{p}+1}^{\tilde{p}} (\mathbf{v}_j^\top \mathbf{a})^2 \geq 0.$$

Therefore, we have

$$P_{\mathbf{V}_X} - P_{\mathbf{V}_{X,\bar{p}}} \succeq 0.$$

Therefore, we get that

$$\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_X} \Gamma_X \succeq \frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_{X,\bar{p}}} \Gamma_X.$$

Since $P_{\Gamma_X}\Gamma_X = \Gamma_X$, we can write that

$$\begin{aligned}\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_{X,\bar{p}}} \Gamma_X &= \frac{1}{p}\Gamma_X^\top \Gamma_X + \mathcal{E}_X, \\ \mathcal{E}_X &= \frac{1}{p}\Gamma_X^\top \left(P_{\mathbf{V}_{X,\bar{p}}} - P_{\Gamma_X}\right)\Gamma_X.\end{aligned}$$

By Assumption 3.2 and the bound of Step 4,

$$\|\mathcal{E}_X\|_{\mathrm{op}} \leq \frac{1}{p}\|\Gamma_X\|_{\mathrm{op}}^2 \left\|P_{\mathbf{V}_{X,\bar{p}}} - P_{\Gamma_X}\right\|_{\mathrm{op}} \leq CC_6\delta_X.$$

Step 6: Lower bound. Now, combining Step 5 with Weyl's inequality and Assumption 3.2,

$$\begin{aligned}\nu_{\min}^2(\mathbf{H}_X) &= \lambda_{\min}\left(\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_X}\Gamma_X\right) \\ &\geq \lambda_{\min}\left(\frac{1}{p}\Gamma_X^\top P_{\mathbf{V}_{X,\bar{p}}}\Gamma_X\right) \\ &\geq \lambda_{\min}\left(\frac{1}{p}\Gamma_X^\top \Gamma_X\right) - \|\mathcal{E}_X\|_{\mathrm{op}} \\ &\geq c - CC_6\delta_X.\end{aligned}$$

We set $c_\star = \min\{1, cc_f/(4C), c/(2CC_6)\}$. When $\delta_X \leq c_\star$, we get that $\nu_{\min}^2(\mathbf{H}_X) \geq c/2$, hence, we have $\nu_{\min}(\mathbf{H}_X) \geq \sqrt{c/2}$. Therefore, on $A_{X,t}$,

$$\sqrt{\frac{c}{2}} \leq \nu_{\min}(\mathbf{H}_X) \leq \nu_{\max}(\mathbf{H}_X) \leq \sqrt{C}.$$

For the mediator block, $\mathbb{E}[\mathbf{f}_M]$ need not be zero. Since $\hat{\Sigma}_M$ is computed after demeaning, we apply the same argument to $\mathbf{f}_M - \mathbb{E}[\mathbf{f}_M]$. By the same argument as in Step 2 and Assumption 3.1(iii)-(iv), $\mathrm{Cov}(\mathbf{f}_M)$ has eigenvalues bounded away from zero and infinity. Thus, Steps 1–6 apply with $(q, \Gamma_M, \mathrm{Cov}(\mathbf{f}_M), \Sigma_{\mathbf{u}_M}, \bar{q}, \delta_M)$ in place of $(p, \Gamma_X, \Sigma_{\mathbf{f}_X}, \Sigma_{\mathbf{u}_X}, \bar{p}, \delta_X)$. Hence, on an event $A_{M,t}$ with probability at least $1 - C_0 e^{-t}$,

$$\sqrt{\frac{c}{2}} \leq \nu_{\min}(\mathbf{H}_M) \leq \nu_{\max}(\mathbf{H}_M) \leq \sqrt{C}.$$

Let $A_t = A_{X,t} \cap A_{M,t}$. By a union bound, $\mathbb{P}(A_t) \geq 1 - C_1 e^{-t}$ for some universal constant $C_1 > 0$. Therefore, whenever $\delta_\star = \max\{\delta_X, \delta_M\} \leq c_\star$, we have

$$c \leq \nu_{\min}(\mathbf{H}_X) \leq \nu_{\max}(\mathbf{H}_X) \leq C, \qquad c \leq \nu_{\min}(\mathbf{H}_M) \leq \nu_{\max}(\mathbf{H}_M) \leq C,$$

□

Lemma C.1 (Operator norm of the weighting matrix)**.** *Let* $\mathbf{W} = \mathrm{diag}(p^{-1}\mathbf{W}_X, q^{-1}\mathbf{W}_M)$*, where* $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$ *and* $\mathbf{W}_M = \sqrt{q}\,\mathbf{V}_M$*, with* $\mathbf{V}_X^\top \mathbf{V}_X = I_{\tilde{p}}$ *and* $\mathbf{V}_M^\top \mathbf{V}_M = I_{\tilde{q}}$*. Then*

$$\|\mathbf{W}\|_{\mathrm{op}}^2 = \frac{1}{p \wedge q}.$$

Proof. Since $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$ and $\mathbf{W}_M = \sqrt{q}\,\mathbf{V}_M$, we have $\mathbf{W} = \mathrm{diag}(p^{-1/2}\mathbf{V}_X, q^{-1/2}\mathbf{V}_M)$. Because $\mathbf{V}_X$ and $\mathbf{V}_M$ have orthonormal columns, we have that

$$\|p^{-1/2}\mathbf{V}_X\|_{\mathrm{op}}^2 = \frac{1}{p}, \quad \|q^{-1/2}\mathbf{V}_M\|_{\mathrm{op}}^2 = \frac{1}{q}.$$

Also, since the operator norm of a block-diagonal matrix is the maximum of the operator norms of its diagonal blocks, so

$$\|\mathbf{W}\|_{\mathrm{op}}^2 = \max\left\{\frac{1}{p}, \frac{1}{q}\right\} = \frac{1}{p \wedge q}.$$

□

Lemma C.2. *Suppose Assumptions 3.1–3.2 and 5.1–5.2 hold. Then, conditional on the training fold, as* $n, p, q \to \infty$*,*

(i) $\left\|\frac{1}{n}\mathbb{E}(\mathbf{E}_0^\top \mathbf{E}_0)\right\|_{op} = O\left(\frac{1}{p\wedge q}\right), \|\mathbf{E}_0\|_{op} = O_p\left(\sqrt{\frac{n}{p\wedge q}}\right).$

(ii) $\left\|\frac{1}{\sqrt{n}}\mathbf{F}_1^\top \mathbf{E}_0\right\|_{op} = O_p\left(\frac{1}{\sqrt{p\wedge q}}\right), \left\|\frac{1}{n}\mathbf{1}^\top \mathbf{E}_0\right\|_2 = O_p\left(\frac{1}{\sqrt{n(p\wedge q)}}\right).$

(iii) $\left\|\frac{1}{\sqrt{n}}\left\{\mathbf{E}_0^\top \mathbf{E}_0 - \mathbb{E}(\mathbf{E}_0^\top \mathbf{E}_0)\right\}\right\|_{op} = O_p\left(\frac{1}{p\wedge q}\right), \left\|\frac{1}{\sqrt{n}}\mathbf{E}_0^\top \mathbf{1}\mathbf{1}^\top \mathbf{E}_0\right\|_{op} = O_p\left(\frac{\sqrt{n}}{p\wedge q}\right).$

(iv) $\left\|\frac{1}{\sqrt{n}}\mathbf{E}_0^\top \boldsymbol{\epsilon}_1\right\|_2 = O_p\left(\frac{1}{\sqrt{p\wedge q}}\right).$

Here n *denotes the number of rows of the relevant sub-sample. When the lemma is applied to the treated outcome regression,* n *is replaced by* n_k^t*, and when it is applied to the control bridge regression,* n *is replaced by* n_k^c*. Since* $n_k^t \asymp n_k^c \asymp n_k \asymp n$*, every rate above is order-equivalent under these substitutions.*

Proof. Recall from Section A.1 that $\mathbf{E}_0 = (\mathbf{0}_n, \mathbf{E})$ with $\mathbf{E} = \mathbf{U}\mathbf{W}$. Since the first column of $\mathbf{E}_0$ is zero, it is enough to prove the corresponding bounds for $\mathbf{E}$. We use the inequalities $\|\mathbf{A}\|_{\mathrm{op}} \le \|\mathbf{A}\|_F$, $\|\mathbf{A}\mathbf{B}\|_{\mathrm{op}} \le \|\mathbf{A}\|_{\mathrm{op}}\|\mathbf{B}\|_{\mathrm{op}}$, and $\mathrm{tr}(\mathbf{A}^\top\mathbf{B}\mathbf{A}) \le \|\mathbf{B}\|_{\mathrm{op}}\|\mathbf{A}\|_F^2$ for positive semidefinite $\mathbf{B}$. By Lemma C.1, $\|\mathbf{W}\|_{\mathrm{op}}^2 = 1/(p \wedge q)$. Since $\mathbf{W}$ has at most $\tilde{p} + \tilde{q}$ columns and $\tilde{p}, \tilde{q}$ are fixed,

$$\|\mathbf{W}\|_F^2 \le (\tilde{p} + \tilde{q})\|\mathbf{W}\|_{\mathrm{op}}^2 = O\left(\frac{1}{p \wedge q}\right).$$

Because $\mathbf{W}$ is fixed conditional on the training fold, the rows of $\mathbf{E} = \mathbf{U}\mathbf{W}$ are independent across i within the regression fold.

Proof of part (i). Let $\Sigma_{\mathbf{u}} = \mathbb{E}(\mathbf{u}_i\mathbf{u}_i^\top)$. By Assumption 3.1(ii) and (v), $\|\Sigma_{\mathbf{u}}\|_{\mathrm{op}} \le C$. Since $\mathbf{E} = \mathbf{U}\mathbf{W}$,

$$\mathbf{E}^\top\mathbf{E} = \mathbf{W}^\top\mathbf{U}^\top\mathbf{U}\mathbf{W}.$$

The rows of $\mathbf{U}$ are i.i.d., so $\mathbb{E}(\mathbf{U}^\top\mathbf{U}) = n\Sigma_u$, and therefore $\mathbb{E}(\mathbf{E}^\top\mathbf{E}) = n\mathbf{W}^\top\Sigma_u\mathbf{W}$. It follows that

$$\left\|\frac{1}{n}\mathbb{E}(\mathbf{E}^\top\mathbf{E})\right\|_{\mathrm{op}} = \left\|\mathbf{W}^\top\Sigma_u\mathbf{W}\right\|_{\mathrm{op}} \le \|\mathbf{W}\|_{\mathrm{op}}^2\|\Sigma_u\|_{\mathrm{op}} = O\left(\frac{1}{p \wedge q}\right),$$

Since $\mathbf{E}_0$ only adds a zero column to $\mathbf{E}$, the same bound holds for $\mathbf{E}_0$.

Next, using $\|\mathbf{E}\|_{\text{op}} \le \|\mathbf{E}\|_F$,

$$\begin{aligned}
\mathbb{E}\|\mathbf{E}\|_{\text{op}}^2 &\le \mathbb{E}\|\mathbf{E}\|_F^2 \\
&= \mathbb{E}\,\text{tr}(\mathbf{E}^\top \mathbf{E}) \\
&= n\,\text{tr}(\mathbf{W}^\top \Sigma_{\mathbf{u}} \mathbf{W}) \\
&\le n\|\Sigma_{\mathbf{u}}\|_{\text{op}}\|\mathbf{W}\|_F^2 \\
&= O\left(\frac{n}{p \wedge q}\right).
\end{aligned}$$

Therefore, by Markov's inequality,

$$\|\mathbf{E}\|_{\text{op}} = O_p\left(\sqrt{\frac{n}{p \wedge q}}\right).$$

Since adding the zero column does not change the operator norm, we have

$$\|\mathbf{E}_0\|_{\text{op}} = \|\mathbf{E}\|_{\text{op}} = O_p\left(\sqrt{\frac{n}{p \wedge q}}\right).$$

This proves part (i).

Proof of part (ii). Since $\mathbf{F}_1 = (\mathbf{1} \quad \mathbf{F})$, we first bound $\frac{1}{\sqrt{n}}\mathbf{1}^\top \mathbf{E}$ and $\frac{1}{\sqrt{n}}\mathbf{F}^\top \mathbf{E}$ in order to find the bound of $\frac{1}{\sqrt{n}}\mathbf{F}_1^\top \mathbf{E}_0$. For the first term,

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{\sqrt{n}}\mathbf{1}^\top \mathbf{E}\right\|_2^2 &= \mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{W}^\top \mathbf{u}_i\right\|_2^2 \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\|\mathbf{W}^\top \mathbf{u}_i\|_2^2 \\
&= \text{tr}(\mathbf{W}^\top \Sigma_{\mathbf{u}} \mathbf{W}) \\
&\le \|\Sigma_{\mathbf{u}}\|_{\text{op}}\|\mathbf{W}\|_F^2 \\
&= O\left(\frac{1}{p \wedge q}\right),
\end{aligned}$$

where the cross terms vanish because the observations are independent and $\mathbb{E}(\mathbf{u}_i) = 0$. Thus, we have

$$\left\|\frac{1}{\sqrt{n}}\mathbf{1}^\top \mathbf{E}\right\|_2 = O_p\left(\frac{1}{\sqrt{p \wedge q}}\right).$$

For the second term, note that

$$\mathbf{F}^\top \mathbf{E} = \sum_{i=1}^n \mathbf{f}_i \mathbf{u}_i^\top \mathbf{W}.$$

By Assumption 3.1, $\mathbb{E}(\mathbf{f}_i \mathbf{u}_i^\top) = 0$, so the cross terms vanish. Hence, we have

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{\sqrt{n}}\mathbf{F}^\top \mathbf{E}\right\|_F^2 &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\|\mathbf{f}_i \mathbf{u}_i^\top \mathbf{W}\|_F^2 \\
&= \mathbb{E}\left[\|\mathbf{f}_i\|_2^2 \|\mathbf{W}^\top \mathbf{u}_i\|_2^2\right] \\
&\overset{(1)}{\le} \{\mathbb{E}\|\mathbf{f}_i\|_2^4\}^{1/2}\{\mathbb{E}\|\mathbf{W}^\top \mathbf{u}_i\|_2^4\}^{1/2} \\
&\overset{(2)}{=} O\left(\frac{1}{p \wedge q}\right),
\end{aligned}$$

where (1) uses Cauchy-Schwarz and (2) uses Assumption 5.1 and Lemma C.1. Since sub-Gaussian random variables have finite moments of all orders, the fourth moments above are finite.

Now, combining the bounds for $\mathbf{1}^\top \mathbf{E}$ and $\mathbf{F}^\top \mathbf{E}$, and using $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_F$, Markov's inequality gives

$$\left\| \frac{1}{\sqrt{n}} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\left(\frac{1}{\sqrt{p \wedge q}} \right).$$

Moreover,

$$\mathbb{E} \left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2^2 = \frac{1}{n} \mathbb{E} \left\| \frac{1}{\sqrt{n}} \mathbf{1}^\top \mathbf{E} \right\|_2^2 = O\left(\frac{1}{n(p \wedge q)} \right),$$

Thus, we simply have

$$\left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E}_0 \right\|_2 = O_p\left(\frac{1}{\sqrt{n(p \wedge q)}} \right).$$

This proves part (ii).

Proof of part (iii). Let $\mathbf{w}_a$ denote the ath column of $\mathbf{W}$, for $a = 1, \dots, \tilde{p} + \tilde{q}$. Since $\mathbf{E} = \mathbf{UW}$, the (a, b) entry of $\mathbf{E}^\top \mathbf{E}$ is

$$(\mathbf{E}^\top \mathbf{E})_{ab} = \sum_{i=1}^{n} (\mathbf{w}_a^\top \mathbf{u}_i)(\mathbf{w}_b^\top \mathbf{u}_i).$$

Hence, we have

$$\frac{1}{\sqrt{n}} \left\{ (\mathbf{E}^\top \mathbf{E})_{ab} - \mathbb{E}[(\mathbf{E}^\top \mathbf{E})_{ab}] \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ (\mathbf{w}_a^\top \mathbf{u}_i)(\mathbf{w}_b^\top \mathbf{u}_i) - \mathbb{E}[(\mathbf{w}_a^\top \mathbf{u}_i)(\mathbf{w}_b^\top \mathbf{u}_i)] \right\}.$$

The summands are independent and mean zero. Therefore,

$$\begin{aligned}
&\mathrm{Var}\left[\frac{1}{\sqrt{n}} \left\{ (\mathbf{E}^\top \mathbf{E})_{ab} - \mathbb{E}[(\mathbf{E}^\top \mathbf{E})_{ab}] \right\} \right] \\
&= \mathrm{Var}\left[(\mathbf{w}_a^\top \mathbf{u}_1)(\mathbf{w}_b^\top \mathbf{u}_1) \right] \\
&\le \mathbb{E}\left[(\mathbf{w}_a^\top \mathbf{u}_1)^2 (\mathbf{w}_b^\top \mathbf{u}_1)^2 \right] \\
&\le \{ \mathbb{E}(\mathbf{w}_a^\top \mathbf{u}_1)^4 \}^{1/2} \{ \mathbb{E}(\mathbf{w}_b^\top \mathbf{u}_1)^4 \}^{1/2} \\
&\le C \|\mathbf{w}_a\|_2^2 \|\mathbf{w}_b\|_2^2 \\
&\le \frac{C}{(p \wedge q)^2},
\end{aligned}$$

where the fourth-moment bound follows from Assumption 5.1(i), and the last inequality uses $\|\mathbf{w}_a\|_2, \|\mathbf{w}_b\|_2 \le \|\mathbf{W}\|_{\mathrm{op}}$ together with Lemma C.1.
By Chebyshev's inequality, for every a, b,

$$\frac{1}{\sqrt{n}} \left\{ (\mathbf{E}^\top \mathbf{E})_{ab} - \mathbb{E}[(\mathbf{E}^\top \mathbf{E})_{ab}] \right\} = O_p\left(\frac{1}{p \wedge q} \right).$$

Since $\tilde{p} + \tilde{q}$ is fixed, the matrix has a fixed number of entries. Also, for a fixed-dimensional matrix $\mathbf{A}$,

$$\|\mathbf{A}\|_{\mathrm{op}} \le \|\mathbf{A}\|_F \le (\tilde{p} + \tilde{q}) \max_{a,b} |A_{ab}|.$$

Therefore,

$$\left\| \frac{1}{\sqrt{n}} \left\{ \mathbf{E}^\top \mathbf{E} - \mathbb{E}(\mathbf{E}^\top \mathbf{E}) \right\} \right\|_{\mathrm{op}} = O_p\left(\frac{1}{p \wedge q} \right).$$

Since $\mathbf{E}_0$ only adds a zero column to $\mathbf{E}$, the same bound holds for $\mathbf{E}_0$. This proves the first half of part (ii).

For the second part, note that $\mathbf{E}^\top \mathbf{1}\mathbf{1}^\top \mathbf{E} = (\mathbf{E}^\top \mathbf{1})(\mathbf{E}^\top \mathbf{1})^\top$ is a rank-one matrix (an outer product of the vector $\mathbf{E}^\top \mathbf{1} \in \mathbb{R}^{\tilde{p}+\tilde{q}}$ with itself), so its operator norm is exactly the squared Euclidean norm of that vector, so we have

$$\left\| \frac{1}{\sqrt{n}} \mathbf{E}^\top \mathbf{1}\mathbf{1}^\top \mathbf{E} \right\|_{\mathrm{op}} = \frac{1}{\sqrt{n}} \|\mathbf{E}^\top \mathbf{1}\|_2^2.$$

Now we connect $\|\mathbf{E}^\top \mathbf{1}\|_2$ to the quantity already bounded in part (ii). Since $\mathbf{E}^\top \mathbf{1} = n \cdot \left(\frac{1}{n}\mathbf{1}^\top \mathbf{E}\right)^\top$, we get

$$\|\mathbf{E}^\top \mathbf{1}\|_2^2 = n^2 \left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2^2,$$

Now, substituting this we get that

$$\left\| \frac{1}{\sqrt{n}} \mathbf{E}^\top \mathbf{1}\mathbf{1}^\top \mathbf{E} \right\|_{\mathrm{op}} = \frac{n^2}{\sqrt{n}} \left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2^2 = n^{3/2} \left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2^2.$$

By part (ii),

$$\left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2 = O_p\left(\frac{1}{\sqrt{n(p \wedge q)}} \right),$$

and hence

$$\left\| \frac{1}{n} \mathbf{1}^\top \mathbf{E} \right\|_2^2 = O_p\left(\frac{1}{n(p \wedge q)} \right).$$

Therefore, we have

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}} \mathbf{E}^\top \mathbf{1}\mathbf{1}^\top \mathbf{E} \right\|_{\mathrm{op}} &= n^{3/2} O_p\left(\frac{1}{n(p \wedge q)} \right) \\ &= O_p\left(\frac{\sqrt{n}}{p \wedge q} \right). \end{aligned}$$

Since $\mathbf{E}_0$ only adds a zero column to $\mathbf{E}$, the same bound holds for $\mathbf{E}_0$. This proves part (iii).

Proof of part (iv). Finally,

$$\mathbf{E}^\top \boldsymbol{\epsilon}_1 = \sum_{i=1}^n \mathbf{W}^\top \mathbf{u}_i \epsilon_{1,i}.$$

Using independence across observations and the tower-property condition $\mathbb{E}(\mathbf{u}_i \epsilon_{1,i}) = 0$, the cross terms vanish. Hence, we have

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{\sqrt{n}} \mathbf{E}^\top \boldsymbol{\epsilon}_1 \right\|_2^2 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \|\mathbf{W}^\top \mathbf{u}_i \epsilon_{1,i}\|_2^2 \\ &= \mathbb{E}\left[\epsilon_{1,i}^2 \|\mathbf{W}^\top \mathbf{u}_i\|_2^2 \right] \\ &\leq \{\mathbb{E}|\epsilon_{1,i}|^4\}^{1/2} \{\mathbb{E}\|\mathbf{W}^\top \mathbf{u}_i\|_2^4\}^{1/2} \\ &= O\left(\frac{1}{p \wedge q} \right), \end{aligned}$$

by Cauchy–Schwarz, Assumption 5.1(ii), and Lemma C.1. Therefore, by Markov's inequality,

$$\left\| \frac{1}{\sqrt{n}} \mathbf{E}_0^\top \boldsymbol{\epsilon}_1 \right\|_2 = O_p\left(\frac{1}{\sqrt{p \wedge q}} \right).$$

This proves part (iv). □

A projection bound.
We next bound the projection term $\mathbf{1}^\top P_S \mathbf{1}$, which is used in the bias bounds below.

Lemma C.3. *Suppose Assumptions 3.1–3.2 and 5.1–5.2 hold. Then, conditional on the training fold,*

$$\mathbf{1}^\top P_S \mathbf{1} = O_p(1),$$

where $P_S = \mathbf{E}_0 S (S^\top \mathbf{E}_0^\top \mathbf{E}_0 S)^{-1} S^\top \mathbf{E}_0^\top$.

Proof. The idea is to use the special form of $\mathbf{E}_0$. Since the first block of $\mathbf{E}_0$ is zero, $\mathbf{E}_0 S = \mathbf{E}\bar{S}$, where $\bar{S}$ is the lower block of S. We then normalize $\mathbf{E}\bar{S}$ by its population covariance and work with the resulting whitened matrix. Let n be the sample size of the regression sub-sample. The quantities $p = p_n$ and $q = q_n$ may grow with n, while $\tilde{p}, \tilde{q}, \bar{p}, \bar{q}$ are fixed. Define $d = (\tilde{p} - \bar{p}) + (\tilde{q} - \bar{q})$, which is fixed. For the X part, let $\mathbf{U}_X \in \mathbb{R}^{n \times p}$ and $\mathbf{W}_X \in \mathbb{R}^{p \times \tilde{p}}$, and recall $\mathbf{W}_X = \sqrt{p}\,\mathbf{V}_X$ with $\mathbf{V}_X^\top \mathbf{V}_X = I_{\tilde{p}}$, so that

$$\mathbf{E}_X = \frac{1}{p}\mathbf{U}_X W_X = \frac{1}{p}\mathbf{U}_X(\sqrt{p}\,V_X) = \frac{1}{\sqrt{p}}\mathbf{U}_X V_X \in \mathbb{R}^{n \times \tilde{p}}.$$

Similarly, for the M part,

$$\mathbf{E}_M = \frac{1}{q}\mathbf{U}_M W_M = \frac{1}{\sqrt{q}}\mathbf{U}_M V_M \in \mathbb{R}^{n \times \tilde{q}},$$

where $W_M = \sqrt{q}\,V_M$ and $V_M^\top V_M = I_{\tilde{q}}$. Also, $\mathbf{E} = \begin{pmatrix}\mathbf{E}_X & \mathbf{E}_M\end{pmatrix} \in \mathbb{R}^{n \times (\tilde{p}+\tilde{q})}$ and $\mathbf{E}_0 = \begin{pmatrix}0 & \mathbf{E}\end{pmatrix} \in \mathbb{R}^{n \times (1+\tilde{p}+\tilde{q})}$. Let $S_X \in \mathbb{R}^{\tilde{p} \times (\tilde{p}-\bar{p})}$ be an orthonormal basis for the orthogonal complement of $\mathrm{Col}(\mathbf{H}_X)$, and let $S_M \in \mathbb{R}^{\tilde{q} \times (\tilde{q}-\bar{q})}$ be an orthonormal basis for the orthogonal complement of $\mathrm{Col}(\mathbf{H}_M)$. Hence $S_X^\top S_X = I_{\tilde{p}-\bar{p}}$ and $S_M^\top S_M = I_{\tilde{q}-\bar{q}}$. We take

$$\bar{S} = \begin{pmatrix} S_X & 0 \\ 0 & S_M \end{pmatrix} \in \mathbb{R}^{(\tilde{p}+\tilde{q}) \times d}.$$

If a different orthonormal basis is used for the same space, the projection is unchanged because only the column space matters. By Assumption 3.1, the second moments $\Sigma_{u_X} = \mathbb{E}[\frac{1}{n}\mathbf{U}_X^\top \mathbf{U}_X]$ and $\Sigma_{u_M} = \mathbb{E}[\frac{1}{n}\mathbf{U}_M^\top \mathbf{U}_M]$ satisfy $\mathbb{E}[\mathbf{U}_X^\top \mathbf{U}_M] = 0$ and

$$\lambda_{\min}(\Sigma_{u_X}) \geq c_X > 0, \quad \lambda_{\min}(\Sigma_{u_M}) \geq c_M > 0.$$

Assumption 5.1 provides the fourth-moment bounds used below.

Step 1. Reduce $\mathbf{E}_0 S$ to $\mathbf{E}\bar{S}$. Write S by separating its first row from the lower block

$$S = \begin{pmatrix} s_0^\top \\ \bar{S} \end{pmatrix}, \quad s_0 \in \mathbb{R}^d, \quad \bar{S} \in \mathbb{R}^{(\tilde{p}+\tilde{q}) \times d}.$$

Since $\mathbf{E}_0 = (0\ \mathbf{E})$, we have

$$\begin{aligned} \mathbf{E}_0 S &= (0\ \mathbf{E}) \begin{pmatrix} s_0^\top \\ \bar{S} \end{pmatrix} \\ &= 0 \cdot s_0^\top + \mathbf{E}\bar{S} \\ &= \mathbf{E}\bar{S}. \end{aligned}$$

So $\mathbf{E}_0 S$ and $\mathbf{E}\bar{S}$ are the same matrix, giving $P_S = P_{\mathbf{E}\bar{S}}$ and $\mathbf{1}^\top P_S \mathbf{1} = \mathbf{1}^\top P_{\mathbf{E}\bar{S}} \mathbf{1}$. It is therefore enough to prove $\mathbf{1}^\top P_{\mathbf{E}\bar{S}} \mathbf{1} = \mathcal{O}_P(1)$.

Step 2. Compute the population covariance matrix using block matrices. Define

$$\Sigma = \mathbb{E}\left[\frac{1}{n}\bar{S}^\top \mathbf{E}^\top \mathbf{E}\bar{S}\right].$$

We compute this directly from the block form of $\mathbf{E}$. Since

$$\mathbf{E} = \begin{pmatrix}\mathbf{E}_X & \mathbf{E}_M\end{pmatrix},$$

we have

$$\mathbf{E}^\top \mathbf{E} = \begin{pmatrix} \mathbf{E}_X^\top \mathbf{E}_X & \mathbf{E}_X^\top \mathbf{E}_M \\ \mathbf{E}_M^\top \mathbf{E}_X & \mathbf{E}_M^\top \mathbf{E}_M \end{pmatrix} \quad \text{and} \quad \bar{S} = \begin{pmatrix} S_X & 0 \\ 0 & S_M \end{pmatrix},$$

we get

$$\begin{aligned} \frac{1}{n}\bar{S}^\top \mathbf{E}^\top \mathbf{E} \bar{S} &= \frac{1}{n} \begin{pmatrix} S_X^\top & 0 \\ 0 & S_M^\top \end{pmatrix} \begin{pmatrix} \mathbf{E}_X^\top \mathbf{E}_X & \mathbf{E}_X^\top \mathbf{E}_M \\ \mathbf{E}_M^\top \mathbf{E}_X & \mathbf{E}_M^\top \mathbf{E}_M \end{pmatrix} \begin{pmatrix} S_X & 0 \\ 0 & S_M \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n} S_X^\top \mathbf{E}_X^\top \mathbf{E}_X S_X & \frac{1}{n} S_X^\top \mathbf{E}_X^\top \mathbf{E}_M S_M \\ \frac{1}{n} S_M^\top \mathbf{E}_M^\top \mathbf{E}_X S_X & \frac{1}{n} S_M^\top \mathbf{E}_M^\top \mathbf{E}_M S_M \end{pmatrix}. \end{aligned}$$

Now we compute each block. Since $\mathbf{E}_X = \frac{1}{\sqrt{p}} \mathbf{U}_X V_X$ and $\mathbf{E}_M = \frac{1}{\sqrt{q}} \mathbf{U}_M V_M$, we have

$$\begin{aligned} \frac{1}{n}\mathbf{E}_X^\top \mathbf{E}_X &= \frac{1}{p} V_X^\top \left(\frac{1}{n} \mathbf{U}_X^\top \mathbf{U}_X \right) V_X, \\ \frac{1}{n}\mathbf{E}_M^\top \mathbf{E}_M &= \frac{1}{q} V_M^\top \left(\frac{1}{n} \mathbf{U}_M^\top \mathbf{U}_M \right) V_M, \\ \frac{1}{n}\mathbf{E}_X^\top \mathbf{E}_M &= \frac{1}{\sqrt{pq}} V_X^\top \left(\frac{1}{n} \mathbf{U}_X^\top \mathbf{U}_M \right) V_M. \end{aligned}$$

Taking expectations gives

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\mathbf{E}_X^\top \mathbf{E}_X \right] &= \frac{1}{p} \mathbf{V}_X^\top \Sigma_{\mathbf{u}_X} \mathbf{V}_X, \\ \mathbb{E}\left[\frac{1}{n}\mathbf{E}_M^\top \mathbf{E}_M \right] &= \frac{1}{q} \mathbf{V}_M^\top \Sigma_{\mathbf{u}_M} \mathbf{V}_M, \\ \mathbb{E}\left[\frac{1}{n}\mathbf{E}_X^\top \mathbf{E}_M \right] &= 0. \end{aligned}$$

Therefore,

$$\Sigma = \begin{pmatrix} p^{-1}\Omega_X & 0 \\ 0 & q^{-1}\Omega_M \end{pmatrix},$$

where

$$\Omega_X = S_X^\top \mathbf{V}_X^\top \Sigma_{\mathbf{u}_X} \mathbf{V}_X S_X, \qquad \Omega_M = S_M^\top \mathbf{V}_M^\top \Sigma_{\mathbf{u}_M} \mathbf{V}_M S_M.$$

We now show that Ω_X and Ω_M are uniformly positive definite. For any $a \in \mathbb{R}^{\tilde{p}-\bar{p}}$,

$$\begin{aligned} a^\top \Omega_X a &= (\mathbf{V}_X S_X a)^\top \Sigma_{\mathbf{u}_X} (\mathbf{V}_X S_X a) \\ &\geq c \|\mathbf{V}_X S_X a\|_2^2, \end{aligned}$$

where the last inequality follows from Assumption 3.1. Since $\mathbf{V}_X^\top \mathbf{V}_X = I_{\tilde{p}}$ and $S_X^\top S_X = I_{\tilde{p}-\bar{p}}$,

$$\begin{aligned} \|\mathbf{V}_X S_X a\|_2^2 &= a^\top S_X^\top \mathbf{V}_X^\top \mathbf{V}_X S_X a \\ &= a^\top S_X^\top S_X a = \|a\|_2^2. \end{aligned}$$

Hence, $a^\top \Omega_X a \geq c\|a\|_2^2$, so $\lambda_{\min}(\Omega_X) \geq c$. The same argument gives $\lambda_{\min}(\Omega_M) \geq c$. Therefore,

$$\lambda_{\min}(p^{-1}\Omega_X) \geq \frac{c}{p}, \qquad \lambda_{\min}(q^{-1}\Omega_M) \geq \frac{c}{q},$$

and hence Σ is positive definite and invertible.

Step 3: Whiten the matrix and rewrite the projection. Define

$$\widetilde{\mathbf{E}} = \mathbf{E}\bar{S}\Sigma^{-1/2}.$$

Since Σ is positive definite, $\Sigma^{-1/2}$ is invertible. Therefore, multiplying $\mathbf{E}\bar{S}$ on the right by $\Sigma^{-1/2}$ does not change its column space, so $\mathrm{Col}(\widetilde{\mathbf{E}}) = \mathrm{Col}(\mathbf{E}\bar{S})$ and $P_{\widetilde{\mathbf{E}}} = P_{\mathbf{E}\bar{S}}$. Hence, we have

$$\begin{aligned}\mathbf{1}^\top P_{\mathbf{E}\bar{S}}\mathbf{1} &= \mathbf{1}^\top P_{\widetilde{\mathbf{E}}}\mathbf{1} \\ &= \mathbf{1}^\top \widetilde{\mathbf{E}}(\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{E}})^{-1}\widetilde{\mathbf{E}}^\top\mathbf{1}.\end{aligned}$$

Now, using the normalization by n, we get

$$\mathbf{1}^\top P_{\mathbf{E}\bar{S}}\mathbf{1} = \left(\frac{1}{\sqrt{n}}\mathbf{1}^\top\widetilde{\mathbf{E}}\right)\left(\frac{1}{n}\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{E}}\right)^{-1}\left(\frac{1}{\sqrt{n}}\widetilde{\mathbf{E}}^\top\mathbf{1}\right).$$

Define

$$A_n = \frac{1}{\sqrt{n}}\mathbf{1}^\top\widetilde{\mathbf{E}}, \qquad B_n = \frac{1}{n}\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{E}}.$$

Then, we get

$$\mathbf{1}^\top P_{\mathbf{E}\bar{S}}\mathbf{1} = A_n B_n^{-1} A_n^\top.$$

Thus, it is enough to show that $\|A_n\|_2 = O_p(1)$ and $\|B_n^{-1}\|_{\mathrm{op}} = O_p(1)$.

Step 4: The whitened rows have covariance I_d. Let $\widetilde{e}_i^\top$ denote the ith row of $\widetilde{\mathbf{E}}$ and let $e_i^\top$ denote the ith row of $\mathbf{E}$. From the definition of $\widetilde{\mathbf{E}}$, $\widetilde{e}_i^\top = e_i^\top\bar{S}\Sigma^{-1/2}$. Since $\mathbb{E}(e_i) = 0$, we also have $\mathbb{E}(\widetilde{e}_i) = 0$.
Recall that

$$\Sigma = \mathbb{E}\left[\frac{1}{n}\bar{S}^\top\mathbf{E}^\top\mathbf{E}\bar{S}\right].$$

Since $\mathbf{E}^\top\mathbf{E} = \sum_{i=1}^n e_i e_i^\top$, and the rows are identically distributed within the regression sub-sample,

$$\begin{aligned}\Sigma &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\bar{S}^\top e_i e_i^\top \bar{S}\right] \\ &= \mathbb{E}\left[\bar{S}^\top e_i e_i^\top \bar{S}\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Cov}(\widetilde{e}_i) &= \mathbb{E}(\widetilde{e}_i\widetilde{e}_i^\top) \\ &= \Sigma^{-1/2}\mathbb{E}\left[\bar{S}^\top e_i e_i^\top \bar{S}\right]\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma^{-1/2} \\ &= I_d.\end{aligned}$$

Thus, the rows of $\widetilde{\mathbf{E}}$ are mean zero with covariance I_d.

Step 5. The fourth moments remain bounded after whitening. Since

$$\Sigma = \begin{pmatrix} p^{-1}\Omega_X & 0 \\ 0 & q^{-1}\Omega_M \end{pmatrix},$$

we have

$$\Sigma^{-1/2} = \begin{pmatrix} \sqrt{p}\,\Omega_X^{-1/2} & 0 \\ 0 & \sqrt{q}\,\Omega_M^{-1/2} \end{pmatrix}.$$

Also, we have

$$\begin{aligned}\mathbf{E}\bar{S} &= \begin{pmatrix}\mathbf{E}_X S_X & \mathbf{E}_M S_M\end{pmatrix} \\ &= \begin{pmatrix}\frac{1}{\sqrt{p}}\mathbf{U}_X V_X S_X & \frac{1}{\sqrt{q}}\mathbf{U}_M V_M S_M\end{pmatrix}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\widetilde{E} &= \mathbf{E}\bar{S}\Sigma^{-1/2} \\ &= \begin{pmatrix} \frac{1}{\sqrt{p}}\mathbf{U}_X V_X S_X & \frac{1}{\sqrt{q}}\mathbf{U}_M V_M S_M \end{pmatrix} \begin{pmatrix} \sqrt{p}\,\Omega_X^{-1/2} & 0 \\ 0 & \sqrt{q}\,\Omega_M^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U}_X V_X S_X \Omega_X^{-1/2} & \mathbf{U}_M V_M S_M \Omega_M^{-1/2} \end{pmatrix}.\end{aligned}$$

Write the ith whitened row as $\widetilde{e}_i = \begin{pmatrix} \widetilde{g}_{X,i}^\top & \widetilde{g}_{M,i}^\top \end{pmatrix}^\top$, where $\widetilde{g}_{X,i}$ and $\widetilde{g}_{M,i}$ denote the X and M blocks, respectively. Consider one coordinate of $\widetilde{g}_{X,i}$. It can be written as $a^\top \mathbf{u}_{X,i}$, where a is a column of $\mathbf{V}_X S_X \Omega_X^{-1/2}$. Since $\mathbf{V}_X$ and S_X have orthonormal columns,

$$\begin{aligned}\|a\|_2 &\le \|\mathbf{V}_X\|_{\mathrm{op}} \|S_X\|_{\mathrm{op}} \|\Omega_X^{-1/2}\|_{\mathrm{op}} \\ &= \|\Omega_X^{-1/2}\|_{\mathrm{op}} \\ &= \lambda_{\min}(\Omega_X)^{-1/2} \le C,\end{aligned}$$

where the last inequality follows from Step 2.

By Assumption 5.1(i), linear forms of $\mathbf{u}_{X,i}$ have uniformly bounded fourth moments. Hence, $\mathbb{E}(a^\top \mathbf{u}_{X,i})^4 \le C$. Since $\tilde{p} - \bar{p}$ is fixed, it follows that $\mathbb{E}\|\widetilde{g}_{X,i}\|_2^4 \le C$. The same argument applied to the M block gives $\mathbb{E}\|\widetilde{g}_{M,i}\|_2^4 \le C$. Finally, since

$$\|\widetilde{e}_i\|_2^2 = \|\widetilde{g}_{X,i}\|_2^2 + \|\widetilde{g}_{M,i}\|_2^2,$$

we have

$$\|\widetilde{e}_i\|_2^4 \le 2\|\widetilde{g}_{X,i}\|_2^4 + 2\|\widetilde{g}_{M,i}\|_2^4.$$

Therefore, $\mathbb{E}\|\widetilde{e}_i\|_2^4 \le C$ uniformly in n, p, q.

Step 6. Show that $B_n^{-1} = \mathcal{O}_P(1)$. Recall that

$$B_n = \frac{1}{n}\widetilde{E}^\top \widetilde{E} = \frac{1}{n}\sum_{i=1}^n \widetilde{e}_i \widetilde{e}_i^\top.$$

From Step 4, $\mathbb{E}[\widetilde{e}_i \widetilde{e}_i^\top] = I_d$. For the (j,k) entry, $(B_n)_{jk} = \frac{1}{n}\sum_{i=1}^n \widetilde{e}_{ij}\widetilde{e}_{ik}$. Hence, we have $\mathbb{E}[(B_n)_{jk}] = (I_d)_{jk}$. Also, using independence across i and the uniform fourth moment bound,

$$\begin{aligned}\mathrm{Var}\{(B_n)_{jk}\} &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \widetilde{e}_{ij}\widetilde{e}_{ik}\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(\widetilde{e}_{ij}\widetilde{e}_{ik}) \\ &\le \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[\widetilde{e}_{ij}^2 \widetilde{e}_{ik}^2] \\ &\le \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\|\widetilde{e}_i\|^4 \\ &\le \frac{C_4}{n} \to 0.\end{aligned}$$

Therefore, we have $(B_n)_{jk} \xrightarrow{p} (I_d)_{jk}$ for each fixed pair (j,k). Since d is fixed, this gives $B_n \xrightarrow{p} I_d$. By continuity of matrix inversion at I_d, $B_n^{-1} \xrightarrow{p} I_d$. Therefore $B_n^{-1} = \mathcal{O}_P(1)$.

Step 7. Show that $A_n = \mathcal{O}_P(1)$. Recall that

$$A_n^\top = \frac{1}{\sqrt{n}}\widetilde{E}^\top \mathbf{1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{e}_i.$$

Since the rows are mean zero and $\mathrm{Cov}(\widetilde{e}_i) = I_d$, we have

$$\mathbb{E}\|A_n^\top\|^2 = \mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{e}_i\right\|^2 = \mathrm{tr}\left[\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{e}_i\right)\right] = \mathrm{tr}(I_d) = d.$$

Therefore, by Markov's inequality applied to $\|A_n\|^2$,

$$\mathbb{P}(\|A_n\| > M) = \mathbb{P}(\|A_n\|^2 > M^2) \le \frac{\mathbb{E}\|A_n\|^2}{M^2} = \frac{d}{M^2}.$$

Now, letting $M \to \infty$, we get $A_n = \mathcal{O}_P(1)$. From Step 3,

$$\mathbf{1}^\top P_S \mathbf{1} = \mathbf{1}^\top P_{\mathbf{E}\bar{S}} \mathbf{1} = A_n B_n^{-1} A_n^\top.$$

Using the operator norm bound,

$$\mathbf{1}^\top P_S \mathbf{1} = A_n B_n^{-1} A_n^\top \le \|A_n\|^2 \|B_n^{-1}\|_{\mathrm{op}}.$$

From Step 6, $B_n^{-1} = \mathcal{O}_P(1)$, and from Step 7, $A_n = \mathcal{O}_P(1)$. Thus, we have that

$$\|A_n\|^2 \|B_n^{-1}\|_{\mathrm{op}} = \mathcal{O}_P(1)\mathcal{O}_P(1) = \mathcal{O}_P(1).$$

Therefore, $\mathbf{1}^\top P_S \mathbf{1} = \mathcal{O}_P(1)$. This proves the result. □

Lemma C.4 (Trace bound for a fixed-rank projection)**.** *Suppose the conditions of Lemma C.3 hold. Let*

$$P_{S^*} = \mathbf{E}^* S^* (S^{*\top} \mathbf{E}^{*\top} \mathbf{E}^* S^*)^{-1} S^{*\top} \mathbf{E}^{*\top},$$

be the projection onto $\mathrm{Col}(\mathbf{E}^* S^*)$ *for an error matrix* $\mathbf{E}^*$ *and* S^* *has a fixed number of columns, and let* $\widetilde{\mathbf{F}}^*$ *be centered with bounded second moments and independent of* $\mathbf{E}^*$*. Then,*

$$\left\|\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^*\right\|_{\mathrm{op}} = O_p(1).$$

Proof. Using the same whitening argument as in Lemma C.3, write

$$P_{S^*} = \widetilde{\mathbf{E}}(\widetilde{\mathbf{E}}^\top \widetilde{\mathbf{E}})^{-1} \widetilde{\mathbf{E}}^\top,$$

where the rows $\widetilde{e}_i$ of $\widetilde{\mathbf{E}}$ are mean zero with covariance I_d. Define $C_n = (\widetilde{\mathbf{F}}^{*\top} \widetilde{\mathbf{E}})/\sqrt{n}$ and $B_n = (\widetilde{\mathbf{E}}^\top \widetilde{\mathbf{E}})/n$. Then, we have

$$\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^* = C_n B_n^{-1} C_n^\top.$$

We first show that $\|C_n\|_F = O_p(1)$. For each entry, we have

$$(C_n)_{jk} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{f}_{ij}^* \widetilde{e}_{ik}.$$

Since $\widetilde{\mathbf{F}}^*$ is centered and independent of $\widetilde{\mathbf{E}}$, $\mathbb{E}(\widetilde{f}_{ij}^* \widetilde{e}_{ik}) = 0$. Also, using independence across i,

$$\begin{aligned}
\mathrm{Var}\{(C_n)_{jk}\} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[(\widetilde{f}_{ij}^*)^2 \widetilde{e}_{ik}^2\right] \\
&= \mathbb{E}(\widetilde{f}_{ij}^*)^2 \, \mathbb{E}(\widetilde{e}_{ik}^2) \\
&\le C,
\end{aligned}$$

where $\mathbb{E}(\widetilde{e}_{ik}^2) = 1$ by Lemma C.3. Hence $(C_n)_{jk} = O_p(1)$. Since both dimensions are fixed, $\|C_n\|_F = O_p(1)$. By Lemma C.3, $\|B_n^{-1}\|_{\text{op}} = O_p(1)$. Therefore,

$$\begin{aligned}\operatorname{tr}\left(\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^*\right) &= \operatorname{tr}\left(C_n B_n^{-1} C_n^\top\right) \\ &\le \|C_n\|_F^2 \|B_n^{-1}\|_{\text{op}} \\ &= O_p(1).\end{aligned}$$

Since $\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^*$ is positive semidefinite, we get

$$\left\|\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^*\right\|_{\text{op}} \le \operatorname{tr}\left(\widetilde{\mathbf{F}}^{*\top} P_{S^*} \widetilde{\mathbf{F}}^*\right) = O_p(1).$$

□

Lemmas C.3 and C.4 will be used to control the projection terms in the bias bounds below. We next establish the change of basis result needed for the main proof.

A lemma for the bias terms.

Lemma C.5 (Change of basis for the bias terms)**.** *Let $\hat{\mathbf{Z}}$ be defined as in Equation A.1. Suppose Assumptions 3.1-3.2 and 5.1-5.2 hold. Then, conditional on the training fold,*

$$\left\|\mathbf{H}_1^\top \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}}\right)^{-1} \mathbf{H}_1\right\|_{\text{op}} = O_p(1).$$

Proof. Recall $M = \frac{1}{n_t}\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}}$, with all matrices restricted to treated rows.

Step 1: Change of basis.
Since $\mathbf{H}_1$ is full column rank, write $\mathbf{H}_1 = UR$, where $U \in \mathbb{R}^{(1+\tilde{p}+\tilde{q})\times(1+\bar{p}+\bar{q})}$ has orthonormal columns and R is invertible. By Proposition 5.1, $\nu_{\min}(R) \ge c$, and $\|R\|_{\text{op}} \le C$. Now, we extend U to a full orthonormal basis of $\mathbb{R}^{1+\tilde{p}+\tilde{q}}$ using a matrix $S \in \mathbb{R}^{(1+\tilde{p}+\tilde{q})\times(\tilde{p}-\bar{p}+\tilde{q}-\bar{q})}$ whose columns are orthonormal to U, so that $J = [U \;\; S]$ is square and orthogonal. Thus, $S^\top \mathbf{H}_1 = S^\top UR = 0$.

Since J is orthogonal, $J^{-1} = J^\top$ and $(J^\top M J)^{-1} = J^\top M^{-1} J$, so $M^{-1} = J(J^\top M J)^{-1} J^\top$. Also,

$$\begin{aligned}J^\top \mathbf{H}_1 &= J^\top UR \\ &= \begin{pmatrix} U^\top U \\ S^\top U \end{pmatrix} R \\ &= \begin{pmatrix} I \\ 0 \end{pmatrix} R \\ &= \begin{pmatrix} R \\ 0 \end{pmatrix},\end{aligned}$$

Similarly, we have that $\mathbf{H}_1^\top J = \begin{pmatrix} R^\top & 0 \end{pmatrix}$.

Step 2: Matrix $J^\top M J$.
We write $M = \mathbf{H}_1 \Sigma_{\mathbf{F}} \mathbf{H}_1^\top + \mathbf{H}_1 B + B^\top \mathbf{H}_1^\top + C$, where

$$\Sigma_{\mathbf{F}} = \frac{1}{n_t}\mathbf{F}_1^\top \mathbf{F}_1, \quad B = \frac{1}{n_t}\mathbf{F}_1^\top \mathbf{E}_0, \quad C = \frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0.$$

Now, using $\mathbf{H}_1 = UR$ and the orthogonality of $S^\top U = 0$, we get that

$$J^\top M J = \begin{bmatrix} \underbrace{R\Sigma_{\mathbf{F}} R^\top}_{=D} + \underbrace{RBU + U^\top B^\top R^\top + U^\top C U}_{=K_{11}} & \underbrace{RBS + U^\top C S}_{=K_{12}} \\ \underbrace{S^\top B^\top R^\top + S^\top C U}_{=K_{21}} & \underbrace{S^\top C S}_{=K_{22}} \end{bmatrix}.$$

Step 3: Schur complement and E_n.
By the block matrix inversion formula, the $(1,1)$ block of $(J^\top MJ)^{-1}$ is the inverse of the Schur complement

$$D + \underbrace{K_{11} - K_{12}K_{22}^{-1}K_{21}}_{=E_n} = D + E_n.$$

Now, substituting the explicit expressions for B and C, the cross term $K_{12}K_{22}^{-1}K_{21}$ simplifies as follows. First, factor as

$$\begin{aligned}
K_{12} &= \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top)\mathbf{E}_0 S, \\
K_{21} &= \frac{1}{n_t} S^\top \mathbf{E}_0^\top (\mathbf{F}_1 R^\top + \mathbf{E}_0 U), \\
K_{22} &= \frac{1}{n_t} S^\top \mathbf{E}_0^\top \mathbf{E}_0 S.
\end{aligned}$$

Therefore, we will get

$$\begin{aligned}
K_{12}K_{22}^{-1}K_{21} &= \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top)\mathbf{E}_0 S(S^\top \mathbf{E}_0^\top \mathbf{E}_0 S)^{-1} S^\top \mathbf{E}_0^\top (\mathbf{F}_1 R^\top + \mathbf{E}_0 U) \\
&= \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top) P_S (\mathbf{F}_1 R^\top + \mathbf{E}_0 U),
\end{aligned}$$

where $P_S = \mathbf{E}_0 S(S^\top \mathbf{E}_0^\top \mathbf{E}_0 S)^{-1} S^\top \mathbf{E}_0^\top$ is the orthogonal projection onto the column space of $\mathbf{E}_0 S$. Thus,

$$\begin{aligned}
E_n = K_{11} - K_{12}K_{22}^{-1}K_{21} &= \frac{1}{n_t} R\mathbf{F}_1^\top \mathbf{E}_0 U + \frac{1}{n_t} U^\top \mathbf{E}_0^\top \mathbf{F}_1 R^\top + \frac{1}{n_t} U^\top \mathbf{E}_0^\top \mathbf{E}_0 U \\
&\quad - \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top) P_S (\mathbf{F}_1 R^\top + \mathbf{E}_0 U).
\end{aligned}$$

To simplify, we observe that the first three terms can be written as

$$\frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top)(\mathbf{F}_1 R^\top + \mathbf{E}_0 U) - \frac{1}{n_t} R\mathbf{F}_1^\top \mathbf{F}_1 R^\top,$$

so that

$$E_n = \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top)(I - P_S)(\mathbf{F}_1 R^\top + \mathbf{E}_0 U) - \frac{1}{n_t} R\mathbf{F}_1^\top \mathbf{F}_1 R^\top.$$

Now, expanding $(I - P_S)$ and collecting terms by whether they involve P_S or not, this becomes

$$\begin{aligned}
E_n = &-\frac{1}{n_t} R\mathbf{F}_1^\top P_S \mathbf{F}_1 R^\top + \frac{1}{n_t} R\mathbf{F}_1^\top (I - P_S)\mathbf{E}_0 U \\
&+ \frac{1}{n_t} U^\top \mathbf{E}_0^\top (I - P_S)\mathbf{F}_1 R^\top + \frac{1}{n_t} U^\top \mathbf{E}_0^\top (I - P_S)\mathbf{E}_0 U. \qquad \text{(C.7)}
\end{aligned}$$

Step 4: Invertibility of $D + E_n$.
We show that $\lambda_{\min}(D + E_n)$ is bounded away from zero with probability tending to one. Thus, we apply Weyl's inequality

$$\lambda_{\min}(D + E_n) \geq \lambda_{\min}(D) - \|E_n\|.$$

The argument shows that D has smallest eigenvalue bounded away from zero, while $\|E_n\| = o_p(1)$.

Lower bound on $\lambda_{\min}(D)$. Since $D = \frac{1}{n_t} R\mathbf{F}_1^\top \mathbf{F}_1 R^\top$,

$$\lambda_{\min}(D) \geq \nu_{\min}^2(R)\, \lambda_{\min}\left(\frac{1}{n_t}\mathbf{F}_1^\top \mathbf{F}_1\right).$$

The treated-unit factor Gram matrix concentrates around the conditional second-moment matrix $\Omega_1 = \mathbb{E}[\mathbf{f}_{1,i}\mathbf{f}_{1,i}^\top \mid A_i = 1]$, which is positive definite by Assumption 3.1, so $\lambda_{\min}(\frac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1) \geq c > 0$ with probability tending to one. With $\nu_{\min}(R) \geq c_R > 0$ by construction, so

$$\lambda_{\min}(D) \geq c \cdot c_R^2 > 0$$

with probability tending to one.

Upper bound on $\|E_n\|$. From the expansion (C.7), the triangle inequality gives

$$\begin{aligned}\|E_n\| \leq \underbrace{\left\|\frac{1}{n_t}R\mathbf{F}_1^\top P_S\mathbf{F}_1 R^\top\right\|}_{T_1} + \underbrace{\left\|\frac{1}{n_t}R\mathbf{F}_1^\top(I-P_S)\mathbf{E}_0 U\right\|}_{T_2} \\ + \underbrace{\left\|\frac{1}{n_t}U^\top\mathbf{E}_0^\top(I-P_S)\mathbf{F}_1 R^\top\right\|}_{T_3} + \underbrace{\left\|\frac{1}{n_t}U^\top\mathbf{E}_0^\top(I-P_S)\mathbf{E}_0 U\right\|}_{T_4}.\end{aligned}$$

We bound each term in turn.

Bound on T_1. Since R has bounded operator norm,

$$T_1 \leq \|R\|^2 \left\|\frac{1}{n_t}\mathbf{F}_1^\top P_S\mathbf{F}_1\right\|.$$

Write $\mathbf{F}_1 = \mathbf{1}\mu_\mathbf{F}^\top + \widetilde{\mathbf{F}}_1$, where $\widetilde{\mathbf{F}}_1$ is the centred part. Then

$$\mathbf{F}_1^\top P_S\mathbf{F}_1 = \mu_\mathbf{F}\mathbf{1}^\top P_S\mathbf{1}\mu_\mathbf{F}^\top + \mu_\mathbf{F}\mathbf{1}^\top P_S\widetilde{\mathbf{F}}_1 + \widetilde{\mathbf{F}}_1^\top P_S\mathbf{1}\mu_\mathbf{F}^\top + \widetilde{\mathbf{F}}_1^\top P_S\widetilde{\mathbf{F}}_1,$$

so by the triangle inequality,

$$\left\|\frac{1}{n_t}\mathbf{F}_1^\top P_S\mathbf{F}_1\right\| \leq \underbrace{\frac{1}{n_t}\left\|\mu_\mathbf{F}\mathbf{1}^\top P_S\mathbf{1}\mu_\mathbf{F}^\top\right\|}_{A_1} + \underbrace{\frac{1}{n_t}\left\|\mu_\mathbf{F}\mathbf{1}^\top P_S\widetilde{\mathbf{F}}_1\right\|}_{A_2} + \underbrace{\frac{1}{n_t}\left\|\widetilde{\mathbf{F}}_1^\top P_S\mathbf{1}\mu_\mathbf{F}^\top\right\|}_{A_3} + \underbrace{\frac{1}{n_t}\left\|\widetilde{\mathbf{F}}_1^\top P_S\widetilde{\mathbf{F}}_1\right\|}_{A_4}.$$

For A_1, by Lemma C.3, $\mathbf{1}^\top P_S\mathbf{1} = O_p(1)$. Therefore,

$$\begin{aligned}A_1 &= \frac{1}{n_t}\left\|\mu_\mathbf{F}\mathbf{1}^\top P_S\mathbf{1}\mu_\mathbf{F}^\top\right\|_{\text{op}} \\ &= \frac{1}{n_t}\|\mu_\mathbf{F}\|_2^2\,\mathbf{1}^\top P_S\mathbf{1} \\ &= O_p\left(\frac{\|\mu_\mathbf{F}\|_2^2}{n_t}\right).\end{aligned}$$

For A_1, the only random quantity that needs to be controlled is the projection term $\mathbf{1}^\top P_S\mathbf{1}$. By Lemma C.3, we have

$$\mathbf{1}^\top P_S\mathbf{1} = O_p(1).$$

Therefore,

$$\begin{aligned}A_1 &= \frac{1}{n_t}\left\|\mu_\mathbf{F}\mathbf{1}^\top P_S\mathbf{1}\mu_\mathbf{F}^\top\right\|_{\text{op}} \\ &= \frac{1}{n_t}\|\mu_\mathbf{F}\|_2^2\,\mathbf{1}^\top P_S\mathbf{1} \\ &= O_p\left(\frac{\|\mu_\mathbf{F}\|_2^2}{n_t}\right).\end{aligned}$$

Thus, the contribution of the mean part is of order $\|\mu_{\mathbf{F}}\|_2^2/n_t$.
For A_2, we need to control the mixed term $\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1$. Since P_S is an orthogonal projection, it is symmetric and satisfies $P_S^2 = P_S$. Hence, we have

$$\begin{aligned}\left\|\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1\right\|_2^2 &= \left\|(P_S\mathbf{1})^\top (P_S\widetilde{\mathbf{F}}_1)\right\|_2^2 \\ &\le \|P_S\mathbf{1}\|_2^2 \|P_S\widetilde{\mathbf{F}}_1\|_{\mathrm{op}}^2.\end{aligned}$$

The first factor satisfies

$$\|P_S\mathbf{1}\|_2^2 = \mathbf{1}^\top P_S \mathbf{1} = O_p(1)$$

by Lemma C.3. For the second factor, we have

$$\begin{aligned}\|P_S\widetilde{\mathbf{F}}_1\|_{\mathrm{op}}^2 &= \left\|(P_S\widetilde{\mathbf{F}}_1)^\top (P_S\widetilde{\mathbf{F}}_1)\right\|_{\mathrm{op}} \\ &= \left\|\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1\right\|_{\mathrm{op}} \\ &= O_p(1),\end{aligned}$$

where the last equality follows from Lemma C.4. Combining the two bounds gives $\left\|\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1\right\|_2 = O_p(1)$. Therefore,

$$\begin{aligned}A_2 &= \frac{1}{n_t}\left\|\mu_{\mathbf{F}}\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1\right\|_{\mathrm{op}} \\ &\le \frac{1}{n_t}\|\mu_{\mathbf{F}}\|_2 \left\|\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1\right\|_2 \\ &= O_p\left(\frac{\|\mu_{\mathbf{F}}\|_2}{n_t}\right).\end{aligned}$$

Thus, the mixed term is also of order n_t^{-1}, up to the factor $\|\mu_{\mathbf{F}}\|_2$.
The term A_3 is the transpose of the term in A_2. Since taking the transpose does not change the operator norm, so we have

$$\begin{aligned}A_3 &= \frac{1}{n_t}\left\|\widetilde{\mathbf{F}}_1^\top P_S \mathbf{1}\mu_{\mathbf{F}}^\top\right\|_{\mathrm{op}} \\ &= \frac{1}{n_t}\left\|\mu_{\mathbf{F}}\mathbf{1}^\top P_S \widetilde{\mathbf{F}}_1\right\|_{\mathrm{op}} \\ &= O_p\left(\frac{\|\mu_{\mathbf{F}}\|_2}{n_t}\right).\end{aligned}$$

Lastly, for A_4, both sides of the projection contain the centered factor matrix. Lemma C.4 gives directly that

$$\left\|\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1\right\|_{\mathrm{op}} = O_p(1).$$

Therefore, we have

$$\begin{aligned}A_4 &= \frac{1}{n_t}\left\|\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1\right\|_{\mathrm{op}} \\ &= O_p\left(\frac{1}{n_t}\right).\end{aligned}$$

Finally, combining the four terms, we obtain

$$\begin{aligned}\left\|\frac{1}{n_t}\mathbf{F}_1^\top P_S \mathbf{F}_1\right\|_{\mathrm{op}} &\le A_1 + A_2 + A_3 + A_4 \\ &= O_p\left(\frac{1 + \|\mu_{\mathbf{F}}\|_2 + \|\mu_{\mathbf{F}}\|_2^2}{n_t}\right).\end{aligned}$$

Since $\|R\|_{\mathrm{op}} \le C$, it follows that

$$T_1 = O_p\left(\frac{1 + \|\mu_{\mathbf{F}}\|_2 + \|\mu_{\mathbf{F}}\|_2^2}{n_t}\right). \tag{C.8}$$

Bound on T_2 and T_3. Since $\|U\| = 1$, $\|R\| = \mathcal{O}(1)$, and using $\|I - P_S\| \le 1$,

$$T_2 \le \|R\| \left\| \frac{1}{n_t} \mathbf{F}_1^\top \mathbf{E}_0 \right\| \|U\| \le \left\| \frac{1}{n_t} \mathbf{F}_1^\top \mathbf{E}_0 \right\|.$$

By Lemma C.2(ii), $\left\| \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\{(p \wedge q)^{-1/2}\}$, and the $(1/n_t)$-scaled quantity above is smaller by one further factor of $n_t^{-1/2}$,

$$\left\| \frac{1}{n_t} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = \frac{1}{\sqrt{n_t}} \left\| \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\left\{ \left(\frac{1}{n_t(p \wedge q)} \right)^{1/2} \right\},$$

so

$$T_2, T_3 = \mathcal{O}_p\left(\frac{1}{\sqrt{n_t(p \wedge q)}} \right). \tag{C.9}$$

In particular, $T_2, T_3 = o_p(1)$ once $n_t, p \wedge q \to \infty$ separately, with no relative growth rate between n_t and $p \wedge q$ required.

Bound on T_4. Using $\|U\| = 1$ and $\|I - P_S\| \le 1$,

$$T_4 \le \left\| \frac{1}{n_t} \mathbf{E}_0^\top \mathbf{E}_0 \right\| = \mathcal{O}_p\left(\frac{1}{p \wedge q} \right),$$

where the last equality follows from Lemma C.2. Hence,

$$T_4 = \mathcal{O}_p\left(\frac{1}{p \wedge q} \right). \tag{C.10}$$

Combining the bounds. Now, collecting (C.8)-(C.10),

$$\|E_n\| = \mathcal{O}_p\left(\frac{1 + \|\mu_{\mathbf{F}}\|_2 + \|\mu_{\mathbf{F}}\|_2^2}{n_t} + \frac{1}{\sqrt{n_t(p \wedge q)}} + \frac{1}{p \wedge q} \right). \tag{C.11}$$

Each term vanishes once $n_t \to \infty$ and $p \wedge q \to \infty$, so $\|E_n\| = o_p(1)$ under the growth conditions of the theorem, with no relative rate between n_t and $p \wedge q$ required. Weyl's inequality then gives

$$\begin{aligned} \lambda_{\min}(D + E_n) &\ge c\, c_R^2 - \|E_n\| \\ &\ge c\, c_R^2 - o_p(1) \\ &\ge \frac{c\, c_R^2}{2} > 0 \end{aligned}$$

with probability tending to one, and therefore

$$\|(D + E_n)^{-1}\|_{\mathrm{op}} = O_p(1). \tag{C.12}$$

In conclusion, the change of basis gives $M^{-1} = J(J^\top M J)^{-1} J^\top$ and $\mathbf{H}_1^\top J = \begin{pmatrix} R^\top & 0 \end{pmatrix}$, and the $(1,1)$ block of $(J^\top M J)^{-1}$ is the inverse Schur complement $(D + E_n)^{-1}$. Putting these together,

$$\begin{aligned} \mathbf{H}_1^\top M^{-1} \mathbf{H}_1 &= \mathbf{H}_1^\top J (J^\top M J)^{-1} J^\top \mathbf{H}_1 \\ &= \begin{pmatrix} R^\top & 0 \end{pmatrix} (J^\top M J)^{-1} \begin{pmatrix} R \\ 0 \end{pmatrix} \\ &= R^\top (D + E_n)^{-1} R. \end{aligned}$$

Since $\|R\|_{\mathrm{op}} = \mathcal{O}(1)$ and $\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ by (C.12),

$$\left\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\right\|_{\mathrm{op}} \le \|R\|_{\mathrm{op}}^2 \,\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1),$$

which is what the lemma claims. The proof also gives two facts we use later, the bound $\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ and the orthogonality $e_{\mathbf{1}}^\top S = 0$. The second holds because the first column of $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$ is $e_{\mathbf{1}}$, so $e_{\mathbf{1}}$ lies in $\mathrm{Col}(\mathbf{H}_1) = \mathrm{Col}(U)$, and the columns of S are orthogonal to that space by construction. □

The proof also gives $\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ and $e_{\mathbf{1}}^\top S = 0$, which will be used in the bias bounds below.

Consistency and linear expansion of $\hat{\gamma}_1$

We next establish the linear expansion of the first-stage mediator coefficient $\hat{\gamma}_1$. The quantity $\mathbf{H}_M^\top \hat{\gamma}_1$ enters the second-stage bridge regression, so we need both its convergence rate and its leading term.

For this lemma and its proof, every sum over i runs over the treated units $\{i : A_i = 1\}$, the matrices $\hat{\mathbf{Z}}, \mathbf{F}_1, \mathbf{E}_0, \boldsymbol{\epsilon}_1$ collect treated rows only, and $M = \frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}$. The treated regression is by the conditional second-moment matrix $\Omega_1 = \mathbb{E}[\mathbf{f}_{1,i}\mathbf{f}_{1,i}^\top \mid A_i = 1]$, positive definite by Assumption 3.1(ii).

Lemma C.6. *Suppose Assumptions 3.1–3.2 and 5.1–5.2 hold. Then,*

$$\mathbf{H}_M^\top\hat{\gamma}_1 - \gamma_1 = \mathbf{P}_{\bar{q}}\Omega_1^{-1}\frac{1}{n_t}\sum_{i:A_i=1}\mathbf{f}_{1,i}\epsilon_{1,i} + r_\gamma, \qquad \|r_\gamma\|_2 = o_p(n_t^{-1/2}).$$

Proof. We want to show that

$$\sqrt{n_t}\left(\mathbf{H}_M^\top\hat{\gamma}_1 - \gamma_1\right) = \mathcal{O}_p(1),$$

and identify the leading linear term. Let $\mathbf{S}_{\tilde{q}}$ be the projected mediator selector and $\mathbf{P}_{\bar{q}}$ be the true mediator selector defined in the notation section, so that $\mathbf{H}_M^\top\mathbf{S}_{\tilde{q}} = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top$. From the least-squares expression for $\hat{\gamma}_1$ on the treated units, we have

$$\mathbf{H}_M^\top\hat{\gamma}_1 = \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top Y.$$

Now, using the outcome model $Y = \mathbf{Z}\delta + \epsilon_1$ on the treated units, we get that

$$\begin{aligned}
\mathbf{H}_M^\top\hat{\gamma}_1 &= \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top(\mathbf{Z}\delta+\epsilon_1)\\
&= \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top\{\hat{\mathbf{Z}}\delta + (\mathbf{Z}-\hat{\mathbf{Z}})\delta+\epsilon_1\}\\
&= \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\delta + \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top(\mathbf{Z}-\hat{\mathbf{Z}})\delta + \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top\epsilon_1\\
&= \gamma_1 + \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top(\mathbf{Z}-\hat{\mathbf{Z}})\delta + \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{n_t}\hat{\mathbf{Z}}^\top\epsilon_1.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\sqrt{n_t}(\mathbf{H}_M^\top\hat{\gamma}_1 - \gamma_1) \qquad\qquad (\text{C.13})\\
&= -\mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\hat{\mathbf{Z}}^\top\mathbf{E}_0\delta + \mathbf{H}_M^\top\mathbf{S}_{\tilde{q}}\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\hat{\mathbf{Z}}^\top\epsilon_1\\
&= -\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}(\mathbf{H}_1\mathbf{F}_1^\top + \mathbf{E}_0^\top)\mathbf{E}_0\delta + \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}(\mathbf{H}_1\mathbf{F}_1^\top + \mathbf{E}_0^\top)\epsilon_1\\
&= -\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta - \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta\\
&\quad + \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1 + \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top\left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1
\end{aligned}$$

$$= \underbrace{-\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta}_{=J_1} \underbrace{-\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta}_{=J_2}$$
$$\underbrace{+\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1}_{=J_3} \underbrace{+\mathbf{P}_{\bar{q}}\mathbf{H}_1^\top \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1}_{=J_4}.$$

Then, the decomposition can be written as

$$\sqrt{n_t}\left(\mathbf{H}_M^\top\hat{\gamma}_1 - \gamma_1\right) = -J_1 - J_2 + J_3 + J_4.$$

Write $M = \frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}$. Since $\mathbf{P}_{\bar{q}}$ is only a selector matrix, it has bounded operator norm. Hence multiplying by $\mathbf{P}_{\bar{q}}$ does not change the order of the bounds. We also use $\|\delta\| = O_p(1)$.

Term J_1. We have

$$J_1 = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta.$$

By Lemma C.5, $\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\|_{\text{op}} = O_p(1)$, so

$$\|J_1\| \le \|\mathbf{P}_{\bar{q}}\|_{\text{op}}\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\|_{\text{op}}\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta\right\| = O_p(1)\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta\right\|.$$

By Lemma C.2(ii) and $\|\delta\| = O_p(1)$,

$$\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta\right\| \le \left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\right\|_{\text{op}}\|\delta\| = O_p\left\{\left(\frac{1}{p\wedge q}\right)^{1/2}\right\},$$

so

$$\|J_1\| = O_p\left\{\left(\frac{1}{p\wedge q}\right)^{1/2}\right\}.$$

Thus $J_1 = o_p(1)$ as soon as $p\wedge q\to\infty$, exactly as for the structurally identical treated- and control-side terms in Section C.2, with no relative rate between n_t and $p\wedge q$ required.

Term J_2.

$$J_2 = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top M^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta.$$

This term differs from J_1 because there is no $\mathbf{H}_1$ on the right of M^{-1}. After the change of basis the left S-block disappears, but the right side keeps an S-part, so we use the block-inverse identity directly. As in Lemma C.5, write $\mathbf{H}_1 = UR$ with $U^\top U = I$ and complete U to an orthogonal $J = [U \quad S]$, so that $\mathbf{H}_1^\top J = (R^\top \ 0)$ and $M^{-1} = J(J^\top MJ)^{-1}J^\top$. Thus, we have

$$\begin{aligned} J_2 &= \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top J(J^\top MJ)^{-1}J^\top\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta \\ &= \mathbf{P}_{\bar{q}}\begin{pmatrix}R^\top & 0\end{pmatrix}(J^\top MJ)^{-1}J^\top\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta. \end{aligned}$$

The left side is now simple because of the zero block. However, the vector on the right is not of the form $\mathbf{H}_1 a$ for some vector a. So when we multiply it by $J^\top$, it has both a U-part and an S-part

$$J^\top\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta = \begin{pmatrix}\frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top\mathbf{E}_0\delta \\ \frac{1}{\sqrt{n_t}}S^\top\mathbf{E}_0^\top\mathbf{E}_0\delta\end{pmatrix}.$$

Now we use the first row of the block inverse of $J^\top MJ$. From the block inverse formula, the first block row of $(J^\top MJ)^{-1}$ is

$$\begin{pmatrix}(D+E_n)^{-1} & -(D+E_n)^{-1}K_{12}K_{22}^{-1}\end{pmatrix}.$$

Therefore, using $\mathbf{H}_1^\top J = \begin{pmatrix} R^\top & 0 \end{pmatrix}$ and $J^\top = \begin{pmatrix} U^\top \\ S^\top \end{pmatrix}$, we have

$$\begin{aligned}\mathbf{H}_1^\top M^{-1} &= R^\top (D+E_n)^{-1}\left(U^\top - K_{12}K_{22}^{-1}S^\top\right) \\ &= R^\top (D+E_n)^{-1}\left[U^\top - \left\{\frac{1}{n_t}R\mathbf{F}_1^\top \mathbf{E}_0 S + \frac{1}{n_t}U^\top \mathbf{E}_0^\top \mathbf{E}_0 S\right\}\left\{\frac{1}{n_t}S^\top \mathbf{E}_0^\top \mathbf{E}_0 S\right\}^{-1} S^\top\right].\end{aligned}$$

with $\|(D+E_n)^{-1}\|_{\text{op}} = O_p(1)$ by Lemma C.5. Now, applying this identity to $(1/\sqrt{n_t})\mathbf{E}_0^\top \mathbf{E}_0\delta$, we get Recall that $P_S = \mathbf{E}_0 S(S^\top \mathbf{E}_0^\top \mathbf{E}_0 S)^{-1} S^\top \mathbf{E}_0^\top$. Then,

$$\begin{aligned}J_2 &= \mathbf{P}_{\bar{q}} R^\top (D+E_n)^{-1}\left[\frac{1}{\sqrt{n_t}}U^\top \mathbf{E}_0^\top \mathbf{E}_0\delta - \left(\frac{1}{n_t}R\mathbf{F}_1^\top \mathbf{E}_0 S + \frac{1}{n_t}U^\top \mathbf{E}_0^\top \mathbf{E}_0 S\right)\left(\frac{1}{n_t}S^\top \mathbf{E}_0^\top \mathbf{E}_0 S\right)^{-1}\frac{1}{\sqrt{n_t}}S^\top \mathbf{E}_0^\top \mathbf{E}_0\delta\right] \\ &= \mathbf{P}_{\bar{q}} R^\top (D+E_n)^{-1}\left[\frac{1}{\sqrt{n_t}}U^\top \mathbf{E}_0^\top \mathbf{E}_0\delta - \frac{1}{\sqrt{n_t}}\left(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top\right)P_S \mathbf{E}_0\delta\right] \\ &= \mathbf{P}_{\bar{q}} R^\top (D+E_n)^{-1}\left[\frac{1}{\sqrt{n_t}}U^\top \mathbf{E}_0^\top (I-P_S)\mathbf{E}_0\delta - \frac{1}{\sqrt{n_t}}R\mathbf{F}_1^\top P_S \mathbf{E}_0\delta\right].\end{aligned}$$

The outside factor is bounded because $\|\mathbf{P}_{\bar{q}}\| = 1$, $\|R\|_{\text{op}} = O(1)$, and $\|(D+E_n)^{-1}\|_{\text{op}} = O_p(1)$. Thus it is enough to bound the two terms inside the square bracket.

For the first term, using $\|U\|_{\text{op}} = 1$ and $\|I - P_S\|_{\text{op}} \le 1$, we get that

$$\left\|\frac{1}{\sqrt{n_t}}U^\top \mathbf{E}_0^\top (I-P_S)\mathbf{E}_0\delta\right\| \le \frac{1}{\sqrt{n_t}}\|\mathbf{E}_0^\top (I-P_S)\mathbf{E}_0\|_{\text{op}}\|\delta\| \le \sqrt{n_t}\left\|\frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\right\|_{\text{op}}\|\delta\|.$$

Since $\left\|\frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\right\|_{\text{op}} = O_p\left(\frac{1}{p\wedge q}\right)$, from Lemma C.2 the first term is $O_p\left(\frac{\sqrt{n_t}}{p\wedge q}\|\delta\|\right)$. For the second term, we use the simple bound $\|P_S\|_{\text{op}} \le 1$. Then

$$\left\|\frac{1}{\sqrt{n_t}}R\mathbf{F}_1^\top P_S \mathbf{E}_0\delta\right\| \le \|R\|_{\text{op}}\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\right\|_{\text{op}}\|P_S\|_{\text{op}}\|\mathbf{E}_0\delta\|.$$

Here $\|R\|_{\text{op}} = O(1)$, $\|P_S\|_{\text{op}} \le 1$, and $\|(1/\sqrt{n_t})\mathbf{F}_1^\top\|_{\text{op}} = O_p(1)$ because $(1/n_t)\mathbf{F}_1^\top \mathbf{F}_1$ has bounded largest eigenvalue using Assumption 3.1. Also,

$$\|\mathbf{E}_0\delta\| \le \|\mathbf{E}_0^\top \mathbf{E}_0\|_{\text{op}}^{1/2}\|\delta\| = \sqrt{n_t}\left\|\frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\right\|_{\text{op}}^{1/2}\|\delta\| = O_p\left(\sqrt{\frac{n_t}{p\wedge q}}\|\delta\|\right).$$

Therefore, the second term is $O_p\left(\sqrt{\frac{n_t}{p\wedge q}}\right)$. Now, combining the two pieces inside the bracket and using the bounded outside factor, we get

$$\|J_2\| = O_p\left(\frac{\sqrt{n_t}}{p\wedge q} + \sqrt{\frac{n_t}{p\wedge q}}\right).$$

Thus $J_2 = o_p(1)$ under the growth condition $p\wedge q \gg n_t$ of Assumption 5.2, and, unlike J_1 above, J_2 genuinely needs this stronger relative growth rate; the dominant rate $\sqrt{n_t/(p\wedge q)}$ does not vanish unless $n_t/(p\wedge q) \to 0$.

Term J_3. Double check both terms from here before arxiv We have

$$J_3 = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \epsilon_1.$$

Again this is a two-sided term, so we use $\left\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\right\|_{\text{op}} = O_p(1)$. Thus,

$$\|J_3\| \le O_p(1)\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \epsilon_1\right\|.$$

It remains to bound the last factor. By the treated-arm exogeneity and mean zero condition $\mathbb{E}(\epsilon_{1,i}) = \mathbb{E}(\epsilon_{1,i}|\mathbf{f}_{X,i}, \mathbf{f}_{M,i}, A_i = 1) = 0$ of Assumption 3.1(ii), we have $\mathbb{E}(\epsilon_1) = 0$, so

$$\mathbb{E}\left(\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \epsilon_1 \,\middle|\, \mathbf{F}_1\right) = 0.$$

Also, using the bounded conditional variance of Assumption 5.1,

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \epsilon_1\right\|^2 \,\middle|\, \mathbf{F}_1\right] = \operatorname{tr}\left\{\frac{1}{n_t}\mathbf{F}_1^\top \mathbb{E}(\epsilon_1\epsilon_1^\top)\mathbf{F}_1\right\} \le C \operatorname{tr}\left(\frac{1}{n_t}\mathbf{F}_1^\top \mathbf{F}_1\right).$$

The last term is $O_p(1)$ by the bounded factor second moments of Assumption 3.1. Chebyshev's inequality then gives $\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \epsilon_1\| = O_p(1)$, so $J_3 = O_p(1)$. This is the signal term. To get its linear form, note that $\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 = R^\top(D + E_n)^{-1}R$ from Lemma C.5, where $D = R\Sigma_{\mathbf{F}}R^\top$ with $\Sigma_{\mathbf{F}} = \frac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1 \to_p \Omega_1$ and $\|E_n\| = o_p(1)$. Hence, we have

$$\begin{aligned}\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 &= R^\top(D + E_n)^{-1}R \\ &\xrightarrow{p} R^\top(R\Omega_1 R^\top)^{-1}R \\ &= \Omega_1^{-1}.\end{aligned}$$

Therefore, we get

$$\begin{aligned}J_3 &= \mathbf{P}_{\bar{q}}\Omega_1^{-1}\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1 + \mathbf{P}_{\bar{q}}\big(\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 - \Omega_1^{-1}\big)\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1 \\ &= \mathbf{P}_{\bar{q}}\Omega_1^{-1}\frac{1}{\sqrt{n_t}}\sum_{i:A_i=1}\mathbf{f}_{1,i}\epsilon_{1,i} + o_p(1),\end{aligned}$$

since the first factor of the second term is $o_p(1)$ in operator norm and the second factor is $O_p(1)$.

Term J_4.

$$J_4 = \mathbf{P}_{\bar{q}}\mathbf{H}_1^\top M^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1.$$

As with J_2, this is a one-sided term because there is no $\mathbf{H}_1$ on the right. So, the block-inverse identity gives

$$\mathbf{H}_1^\top M^{-1}\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1 = R^\top(D + E_n)^{-1}\left[\frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top(I - P_S)\epsilon_1 - \frac{1}{\sqrt{n_t}}R\mathbf{F}_1^\top P_S\epsilon_1\right].$$

For the first term, conditional on $\mathbf{E}_0$, the matrix P_S is fixed. Also, by the independence and moment conditions on ϵ_1, we have $\mathbb{E}(\epsilon_1 \mid \mathbf{E}_0) = 0$ and $\mathbb{E}(\epsilon_1\epsilon_1^\top \mid \mathbf{E}_0) \preceq CI$. Therefore, we have

$$\begin{aligned}&\mathbb{E}\left[\left\|\frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top(I - P_S)\epsilon_1\right\|_2^2 \,\middle|\, \mathbf{E}_0\right] \\ &\quad= \frac{1}{n_t}\operatorname{tr}\left\{U^\top\mathbf{E}_0^\top(I - P_S)\mathbb{E}(\epsilon_1\epsilon_1^\top \mid \mathbf{E}_0)(I - P_S)\mathbf{E}_0U\right\} \\ &\quad\le \frac{C}{n_t}\operatorname{tr}\left\{U^\top\mathbf{E}_0^\top(I - P_S)\mathbf{E}_0U\right\} \\ &\quad\le \frac{C}{n_t}\operatorname{tr}\left(U^\top\mathbf{E}_0^\top\mathbf{E}_0U\right).\end{aligned}$$

Since U has a fixed number of orthonormal columns, so

$$\begin{aligned}\frac{1}{n_t}\operatorname{tr}\left(U^\top\mathbf{E}_0^\top\mathbf{E}_0U\right) &\le C\left\|\frac{1}{n_t}\mathbf{E}_0^\top\mathbf{E}_0\right\|_{\text{op}} \\ &= O_p\left(\frac{1}{p \wedge q}\right)\end{aligned}$$

by Lemma C.2(i). Hence, we have

$$\left\| \frac{1}{\sqrt{n_t}} U^\top \mathbf{E}_0^\top (I - P_S)\epsilon_1 \right\|_2 = O_p\left(\frac{1}{\sqrt{p \wedge q}}\right).$$

For the second term $\frac{1}{\sqrt{n_t}} R\mathbf{F}_1^\top P_S \epsilon_1$, we again write $\mathbf{F}_1 = \mathbf{1}\mu_{\mathbf{F}}^\top + \widetilde{\mathbf{F}}_1$. The mean part is $\frac{1}{\sqrt{n_t}} R\mu_{\mathbf{F}} \mathbf{1}^\top P_S \epsilon_1$. Since P_S is a function of $\mathbf{E}_0$, and the treated-arm exogeneity of Assumption 5.1(ii) gives $\mathbb{E}(\epsilon_{1,i} \mid \mathbf{f}_{1,i}, \mathbf{E}_0, A_i = 1) = 0$, we have $\mathbb{E}(\epsilon_1 \mid P_S) = 0$, hence $\mathbb{E}(\mathbf{1}^\top P_S \epsilon_1 \mid P_S) = 0$, and

$$\begin{aligned}
\mathbb{E}\left[(\mathbf{1}^\top P_S \epsilon_1)^2 \mid P_S\right] &= \mathbf{1}^\top P_S \mathbb{E}(\epsilon_1 \epsilon_1^\top \mid P_S) P_S \mathbf{1} \\
&\leq C \mathbf{1}^\top P_S^2 \mathbf{1} \\
&= C \mathbf{1}^\top P_S \mathbf{1}
\end{aligned}$$

By Lemma C.3, $\mathbf{1}^\top P_S \mathbf{1} = O_p(1)$. Therefore, we have $\mathbf{1}^\top P_S \epsilon_1 = O_p(1)$, and so

$$\left\| \frac{1}{\sqrt{n_t}} R\mu_{\mathbf{F}} \mathbf{1}^\top P_S \epsilon_1 \right\| = O_p\left(\frac{1}{\sqrt{n_t}}\right).$$

For the centered part, conditional on $\widetilde{\mathbf{F}}_1$ and $\mathbf{E}_0$,

$$\mathbb{E}\left[\left\| \frac{1}{\sqrt{n_t}} R\widetilde{\mathbf{F}}_1^\top P_S \epsilon_1 \right\|_2^2 \,\middle|\, \widetilde{\mathbf{F}}_1, \mathbf{E}_0 \right] \leq \frac{C}{n_t} \operatorname{tr}\left(\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1\right).$$

By Lemma C.4, we have

$$\left\| \widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1 \right\|_{\text{op}} = O_p(1).$$

Since the factor dimension is fixed, its trace is also $O_p(1)$. Hence,

$$\left\| \frac{1}{\sqrt{n_t}} R\widetilde{\mathbf{F}}_1^\top P_S \epsilon_1 \right\|_2 = O_p\left(\frac{1}{\sqrt{n_t}}\right).$$

Therefore, we have

$$\|J_4\| = O_p\left\{ \left(\frac{1}{\sqrt{p \wedge q}}\right) + \frac{1}{\sqrt{n_t}} \right\}.$$

Hence $J_4 = o_p(1)$ if $p \wedge q \to \infty$ and $n_t \to \infty$. Thus, we have

$$J_1 = o_p(1), \quad J_2 = o_p(1), \quad J_3 = O_p(1), \quad J_4 = o_p(1).$$

Now, combining, we obtain

$$\sqrt{n_t}\left(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1\right) = \mathbf{P}_{\bar{q}} \Omega_1^{-1} \frac{1}{\sqrt{n_t}} \sum_{i:A_i=1} \mathbf{f}_{1,i} \epsilon_{1,i} + o_p(1) = \mathcal{O}_p(1).$$

Now, dividing by $\sqrt{n_t}$ gives the stated linear expansion with $r_\gamma = o_p(n_t^{-1/2})$, and in particular

$$\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1 = \mathcal{O}_p(n_t^{-1/2}),$$

which proves the lemma. □

C.2 Proof of Theorem 5.2

Before proceeding with the proof, we first note how the estimation error in $\hat{\gamma}_1$ enters the expansion. From the treated regression expansion above, $\mathbf{H}_M^\top \hat{\gamma}_1$ is $\sqrt{n_t}$-consistent for γ_1. Since $\hat{\gamma}_1$ is the coefficient on the projected mediator factors, it is a $\tilde{q}$-dimensional vector, while γ_1 is recovered through $\mathbf{H}_M^\top \hat{\gamma}_1$. We write

$$\Delta_{\gamma,n} := \mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1 = O_p(n_t^{-1/2}).$$

When $\hat{\gamma}_1$ is plugged into the second-stage bridge regression, the corresponding term in the intercept is $\Delta_{\gamma,n}^\top \delta_0$. Since n_t is of the same order as n, $\sqrt{n}\,\Delta_{\gamma,n}^\top \delta_0 = O_p(1)$, so the term remains in the first-order expansion. We keep it below as L_7.

Proof. We begin with the estimator and the target. Recall that $\hat{\theta}_0 = \hat{\alpha}_1 + \widehat{\gamma_1^\top \delta_0}$ and $\theta_0 = \alpha_1 + \gamma_1^\top \delta_0$. Since the second-stage response depends on the first-stage estimator $\hat{\gamma}_1$, we separate the bridge estimation error from the error due to estimating γ_1. For one regression fold, $\hat{\alpha}_1$ and $\hat{\gamma}_1$ are estimated using the treated units, while $\widehat{\gamma_1^\top \delta_0}$ is estimated using the control units. The whitening matrices are estimated on the other fold and are therefore fixed conditional on the training fold. We write n_t and n_c for the treated and control sample sizes in the regression fold. Since $n_t, n_c \asymp n$, the ratios n/n_t and n/n_c are bounded in probability.

Since $\hat{\gamma}_1$ is the coefficient on the projected mediator factors, the population bridge intercept conditional on $\hat{\gamma}_1$ is $\eta_0(\hat{\gamma}_1) = \hat{\gamma}_1^\top \mathbf{H}_M \delta_0$. Also,

$$\hat{\gamma}_1^\top \mathbf{H}_M \delta_0 - \gamma_1^\top \delta_0 = \delta_0^\top (\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1).$$

Therefore,

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_0 - \theta_0) &= \sqrt{n}(\hat{\alpha}_1 + \hat{\eta}_0) - \sqrt{n}(\alpha_1 + \gamma_1^\top \delta_0) \\
&= \sqrt{n}(\hat{\alpha}_1 - \alpha_1) + \sqrt{n}(\hat{\eta}_0 - \gamma_1^\top \delta_0) \\
&= \sqrt{n}(\hat{\alpha}_1 - \alpha_1) + \sqrt{n}(\hat{\eta}_0 - \hat{\gamma}_1^\top \mathbf{H}_M \delta_0) + \sqrt{n}(\hat{\gamma}_1^\top \mathbf{H}_M \delta_0 - \gamma_1^\top \delta_0) \\
&= \sqrt{n}(\hat{\alpha}_1 - \alpha_1) + \sqrt{n}(\hat{\eta}_0 - \hat{\gamma}_1^\top \mathbf{H}_M \delta_0) + \sqrt{n}\, \delta_0^\top (\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1). \qquad \text{(C.14)}
\end{aligned}$$

We consider the three terms in (C.14) separately. The first term is expanded over the treated units on the $\sqrt{n_t}$ scale. From the expansion below,

$$\sqrt{n_t}(\hat{\alpha}_1 - \alpha_1) = -G_1 - G_2 + G_3 + G_4,$$

and therefore

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1) = \sqrt{\frac{n}{n_t}}\,(-G_1 - G_2 + G_3 + G_4)\,.$$

The second term is expanded over the control units on the $\sqrt{n_c}$ scale into the terms $L_1, \ldots, L_6$.
For the third term, Lemma C.6 gives

$$\sqrt{n_t}(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1) = -J_1 - J_2 + J_3 + J_4.$$

Hence, we have

$$\sqrt{n}\, \delta_0^\top (\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1) = \sqrt{\frac{n}{n_t}}\, \delta_0^\top (-J_1 - J_2 + J_3 + J_4).$$

Since $J_1, J_2, J_4 = o_p(1)$ and $\sqrt{n/n_t} = O_p(1)$, their contributions are of order $o_p(1)$. Therefore, we get that

$$\sqrt{n}\, \delta_0^\top (\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1) = \sqrt{\frac{n}{n_t}}\, \delta_0^\top J_3 + o_p(1).$$

We denote this term by L_7.

Part 1: Linear Expansions.
The goal is to write $\sqrt{n}(\hat{\alpha}_1 - \alpha_1)$ and $\sqrt{n}(\widehat{\gamma_1^\top \delta_0} - \gamma_1^\top \delta_0)$ each as a sum of explicitly named terms. The bias terms will later be shown negligible, while the signal terms will remain in the first-order expansion.

Expansion of $\hat{\alpha}_1$.
Under $A = 1$, the outcome model in (2.1) can be written as $Y = \mathbf{Z}\delta + \epsilon_1$. Recall that $\hat{\alpha}_1$ is the first coordinate

of the least-squares coefficient from regressing Y on $\hat{\mathbf{Z}}$ using the treated units. Let $e_{\mathbf{1}} = (1, 0, \ldots, 0)^\top$. Then,

$$\begin{aligned}
\hat{\alpha}_1 &= e_{\mathbf{1}}^\top (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top Y \\
&= e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top Y \\
&= e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top (\mathbf{Z}\delta + \epsilon_1) \\
&= e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top (\hat{\mathbf{Z}}\delta + \mathbf{Z}\delta - \hat{\mathbf{Z}}\delta + \epsilon_1) \\
&= e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}}\delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top (\mathbf{Z} - \hat{\mathbf{Z}})\delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top \epsilon_1.
\end{aligned}$$

The first term is $e_{\mathbf{1}}^\top \delta = \alpha_1$. Therefore, we have that

$$\hat{\alpha}_1 = \alpha_1 + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top (\mathbf{Z} - \hat{\mathbf{Z}})\delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top \epsilon_1.$$

Recall that $\hat{\mathbf{Z}} = \mathbf{Z} + \mathbf{E}_0$, so $\mathbf{Z} - \hat{\mathbf{Z}} = -\mathbf{E}_0$. Hence, we have

$$\hat{\alpha}_1 = \alpha_1 - e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top \mathbf{E}_0 \delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{n_t} \hat{\mathbf{Z}}^\top \epsilon_1.$$

Now, taking α_1 on the left side and multiplying the whole equation by $\sqrt{n_t}$ gives

$$\begin{aligned}
\sqrt{n_t}(\hat{\alpha}_1 - \alpha_1) &= -e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} \hat{\mathbf{Z}}^\top \mathbf{E}_0 \delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} \hat{\mathbf{Z}}^\top \epsilon_1 \\
&= -e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} (\mathbf{Z}^\top + \mathbf{E}_0^\top) \mathbf{E}_0 \delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} (\mathbf{Z}^\top + \mathbf{E}_0^\top) \epsilon_1.
\end{aligned} \tag{C.15}$$

Also, since $\mathbf{Z} = \mathbf{F}_1 \mathbf{H}_1^\top$, we have $\mathbf{Z}^\top = \mathbf{H}_1 \mathbf{F}_1^\top$. Therefore,

$$\begin{aligned}
\sqrt{n_t}(\hat{\alpha}_1 - \alpha_1) &= -e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} (\mathbf{H}_1 \mathbf{F}_1^\top + \mathbf{E}_0^\top) \mathbf{E}_0 \delta + e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} (\mathbf{H}_1 \mathbf{F}_1^\top + \mathbf{E}_0^\top) \epsilon_1 \\
&= \underbrace{- e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \mathbf{H}_1 \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \mathbf{E}_0 \delta}_{=G_1} \underbrace{- e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} \mathbf{E}_0^\top \mathbf{E}_0 \delta}_{=G_2} \\
&\quad + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \mathbf{H}_1 \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \epsilon_1}_{=G_3} + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_t} \hat{\mathbf{Z}}^\top \hat{\mathbf{Z}} \right)^{-1} \frac{1}{\sqrt{n_t}} \mathbf{E}_0^\top \epsilon_1}_{=G_4}.
\end{aligned}$$

Thus,

$$\sqrt{n_t}(\hat{\alpha}_1 - \alpha_1) = -G_1 - G_2 + G_3 + G_4.$$

Here, G_3 is the signal term, while G_1, G_2, and G_4 are the terms arising from the estimation error $\mathbf{E}_0$. In Part 2, we show that $G_1, G_2, G_4 = o_p(1)$, while G_3 remains. The corresponding $\sqrt{n}$ expansion follows by multiplying by $\sqrt{n/n_t}$, as noted above.

Bridge equation for the second stage.
Before running the second-stage regression on the control units, we first express the response $b = \hat{\gamma}_1^\top \hat{\mathbf{f}}_M$ in terms of the target $\gamma_1^\top \delta_0$ and the remaining error terms. Recall that

$$\Delta_{\gamma,n} = \mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1, \quad \hat{\gamma}_1^\top \mathbf{H}_M = \gamma_1^\top + \Delta_{\gamma,n}^\top,$$

where Lemma C.6 gives $\Delta_{\gamma,n} = O_p(n_t^{-1/2})$.

Since Lemma C.6 controls $\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1$ but not the full vector $\hat{\gamma}_1$, we split $\hat{\gamma}_1$ into its row-space and null-space parts. Let $\mathbf{P}_M = \mathbf{H}_M(\mathbf{H}_M^\top \mathbf{H}_M)^{-1}\mathbf{H}_M^\top$ be the projection onto $\mathrm{Col}(\mathbf{H}_M)$, and write

$$\hat{\gamma}_1 = \mathbf{P}_M \hat{\gamma}_1 + (I - \mathbf{P}_M)\hat{\gamma}_1 = \hat{\gamma}_{1,\parallel} + \hat{\gamma}_{1,\perp}.$$

Since $\mathbf{H}_M^\top \hat{\gamma}_1 = \gamma_1 + \Delta_{\gamma,n}$, the row-space part satisfies

$$\begin{aligned}\hat{\gamma}_{1,\parallel} &= \mathbf{H}_M(\mathbf{H}_M^\top \mathbf{H}_M)^{-1}\mathbf{H}_M^\top \hat{\gamma}_1 \\ &= (\mathbf{H}_M^\dagger)^\top(\gamma_1 + \Delta_{\gamma,n}),\end{aligned}$$

and therefore $\hat{\gamma}_{1,\parallel}^\top = (\gamma_1^\top + \Delta_{\gamma,n}^\top)\mathbf{H}_M^\dagger$. Recall that $\hat{\mathbf{f}}_M = \mathbf{H}_M \mathbf{f}_M + \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M$ for the second stage, the split of $\hat{\gamma}_1$, Lemma C.6, and the mediator equation $\mathbf{f}_M = \delta_0 + \mathbf{B}_0 \mathbf{f}_X + \mathbf{u}_{XM,0}$, we have

$$\begin{aligned}
b &= \hat{\gamma}_1^\top \hat{\mathbf{f}}_M \\
&= \hat{\gamma}_1^\top \left(\mathbf{H}_M \mathbf{f}_M + \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M\right) \quad \text{[Equation (5.2)]} \\
&\overset{(1)}{=} \hat{\gamma}_1^\top \mathbf{H}_M \mathbf{f}_M + \hat{\gamma}_{1,\parallel}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M + \hat{\gamma}_{1,\perp}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M \\
&= \left(\gamma_1^\top + \Delta_{\gamma,n}^\top\right)\mathbf{f}_M + \left(\gamma_1^\top \mathbf{H}_M^\dagger + \Delta_{\gamma,n}^\top \mathbf{H}_M^\dagger\right)\frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t}}\right) \quad \text{[Lemma C.6]} \\
&= \left(\gamma_1^\top + \Delta_{\gamma,n}^\top\right)\mathbf{f}_M + \gamma_1^\top \mathbf{H}_M^\dagger \mathcal{O}_p\left(\frac{1}{\sqrt{q}}\right) + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t q}}\right) + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t}}\right) \\
&= \left(\gamma_1^\top + \Delta_{\gamma,n}^\top\right)\mathbf{f}_M + \mathcal{O}_p\left(\frac{1}{\sqrt{q}}\right) + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t}}\right) \quad \text{[Proposition 5.1]} \\
&= \left(\gamma_1^\top + \Delta_{\gamma,n}^\top\right)(\delta_0 + \mathbf{B}_0 \mathbf{f}_X + \mathbf{u}_{XM,0}) + \mathcal{O}_p\left(\frac{1}{\sqrt{q}}\right) + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t}}\right) \\
&= \gamma_1^\top \delta_0 + \Delta_{\gamma,n}^\top \delta_0 + \gamma_1^\top \mathbf{B}_0 \mathbf{f}_X + \gamma_1^\top \mathbf{u}_{XM,0} + \Delta_{\gamma,n}^\top \mathbf{B}_0 \mathbf{f}_X + \Delta_{\gamma,n}^\top \mathbf{u}_{XM,0} + \mathcal{O}_p\left(\frac{1}{\sqrt{q}}\right) + \mathcal{O}_p\left(\frac{1}{\sqrt{n_t}}\right). \quad \text{(C.16)}
\end{aligned}$$

The three $\mathcal{O}_p$ contributions on the right of (C.16) come from the mediator-whitening term, which the split of $\hat{\gamma}_1$ in (1) writes exactly as

$$\begin{aligned}
\hat{\gamma}_1^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M &= \hat{\gamma}_{1,\parallel}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M + \hat{\gamma}_{1,\perp}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M \\
&= \gamma_1^\top \mathbf{H}_M^\dagger \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M + \Delta_{\gamma,n}^\top \mathbf{H}_M^\dagger \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M + \hat{\gamma}_{1,\perp}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M.
\end{aligned} \quad \text{(C.17)}$$

By Proposition 5.1, $\|\mathbf{H}_M^\dagger\|_{\mathrm{op}} = O(1)$, while $\frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M = O_p\left(\frac{1}{\sqrt{q}}\right)$. Therefore, the first two terms in (C.17) have orders

$$\gamma_1^\top \mathbf{H}_M^\dagger \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M = O_p\left(\frac{1}{\sqrt{q}}\right), \quad \Delta_{\gamma,n}^\top \mathbf{H}_M^\dagger \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_M = O_p\left(\frac{1}{\sqrt{n_t q}}\right).$$

It remains to control the null-space contribution $\hat{\gamma}_{1,\perp}^\top q^{-1}\mathbf{W}_M^\top \mathbf{u}_M$ in (C.17). We do so using the least-squares representation of the treated regression. Recall that

$$\hat{\delta} = \begin{pmatrix}\hat{\alpha}_1 \\ \hat{\beta}_1 \\ \hat{\gamma}_1\end{pmatrix} = \left(\frac{1}{n_t}\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}}\right)^{-1} \frac{1}{n_t}\hat{\mathbf{Z}}^\top Y, \quad \hat{\mathbf{Z}} = \mathbf{Z} + \mathbf{E}_0,$$

where $\mathbf{Z} = \mathbf{F}_1 \mathbf{H}_1^\top$ and $Y = \mathbf{Z}\delta + \epsilon_1$. As in Lemma C.5, write $\mathbf{H}_1 = UR, U^\top U = I$, and complete U to the orthogonal matrix $J = [U \;\; S]$. Since $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$, the null-space completion can be chosen blockwise as

$$S = \begin{pmatrix} 0 & 0 \\ S_X & 0 \\ 0 & S_M \end{pmatrix},$$

where $\mathrm{Col}(S_X) = \mathsf{ker}(\mathbf{H}_X^\top), \mathrm{Col}(S_M) = \mathsf{ker}(\mathbf{H}_M^\top)$, and the columns of S_X and S_M are orthonormal. Let $d_X = \tilde{p} - \bar{p}, d_M = \tilde{q} - \bar{q}, d = d_X + d_M$. Since $S^\top \mathbf{H}_1 = 0$ and also $\mathbf{H}_1$ has full column rank, $(\mathbf{H}_1^\dagger)^\top = \mathbf{H}_1(\mathbf{H}_1^\top \mathbf{H}_1)^{-1}$. Therefore,

$$\delta = (\mathbf{H}_1^\dagger)^\top \Delta = \mathbf{H}_1(\mathbf{H}_1^\top \mathbf{H}_1)^{-1}\Delta \in \mathrm{Col}(\mathbf{H}_1).$$

Also, $\mathbf{H}_1 = UR$ and $S^\top U = 0$. Hence, we have the $S^\top \delta = S^\top \mathbf{H}_1(\mathbf{H}_1^\top \mathbf{H}_1)^{-1}\Delta = 0$. Since $J = [\,U \quad S\,]$ is orthogonal. Hence, we have $JJ^\top = I$, and therefore every vector $\hat{\delta}$ can be decomposed as

$$\hat{\delta} = JJ^\top \hat{\delta} = [\,U \quad S\,] \begin{pmatrix} U^\top \hat{\delta} \\ S^\top \hat{\delta} \end{pmatrix}.$$

Now, we write the coefficient in the J-coordinates as

$$J^\top \hat{\delta} = \begin{pmatrix} \hat{a} \\ \hat{\kappa} \end{pmatrix}, \quad \hat{\kappa} = \begin{pmatrix} \hat{\kappa}_X \\ \hat{\kappa}_M \end{pmatrix}.$$

Here, we are defining that $\hat{a} = U^\top \hat{\delta}, \hat{\kappa} = S^\top \hat{\delta}$, so we have the exact decomposition $\hat{\delta} = U\hat{a} + S\hat{\kappa}$. The first term $U\hat{a}$ lies in $\mathrm{Col}(U) = \mathrm{Col}(\mathbf{H}_1)$, while $S\hat{\kappa}$ lies in its orthogonal complement. Since $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$, the matrix S can be chosen blockwise as

$$S = \begin{pmatrix} 0 & 0 \\ S_X & 0 \\ 0 & S_M \end{pmatrix},$$

Therefore,

$$S\hat{\kappa} = \begin{pmatrix} 0 & 0 \\ S_X & 0 \\ 0 & S_M \end{pmatrix} \begin{pmatrix} \hat{\kappa}_X \\ \hat{\kappa}_M \end{pmatrix} = \begin{pmatrix} 0 \\ S_X \hat{\kappa}_X \\ S_M \hat{\kappa}_M \end{pmatrix}.$$

Thus, the mediator component of the null-space part is therefore $\hat{\gamma}_{1,\perp} = S_M \hat{\kappa}_M$. We next look at the design matrix. To study the coefficient $\hat{\kappa}$, we also write the treated regression in the same J-coordinates. Since

$$\hat{\delta} = J \begin{pmatrix} \hat{a} \\ \hat{\kappa} \end{pmatrix},$$

the fitted value can be written as

$$\hat{\mathbf{Z}}\hat{\delta} = \hat{\mathbf{Z}} J \begin{pmatrix} \hat{a} \\ \hat{\kappa} \end{pmatrix} = \begin{bmatrix} \hat{\mathbf{Z}}U & \hat{\mathbf{Z}}S \end{bmatrix} \begin{pmatrix} \hat{a} \\ \hat{\kappa} \end{pmatrix}.$$

Since we want to control the null-space coefficient $\hat{\kappa}$, and hence $\hat{\gamma}_{1,\perp}$, we need to understand the behavior of these two blocks in the joint least-squares regression. Since $\hat{\mathbf{Z}} = \mathbf{Z} + \mathbf{E}_0$ and we have $\mathbf{Z} = \mathbf{F}_1 \mathbf{H}_1^\top$, we have $\hat{\mathbf{Z}}S = \mathbf{Z}S + \mathbf{E}_0 S$. but $\mathbf{Z}S = \mathbf{F}_1 \mathbf{H}_1^\top S = 0$, Therefore,$\hat{\mathbf{Z}}S = \mathbf{E}_0 S$. Here, the S-direction contain only the projected proxy errors. To see the different scales of these columns, recall that

$$\mathbf{E}_X = \frac{1}{p}\mathbf{U}_X \mathbf{W}_X = \frac{1}{\sqrt{p}}\mathbf{U}_X V_X,$$

since $\mathbf{W}_X = \sqrt{p}\, V_X$, and similarly

$$\mathbf{E}_M = \frac{1}{q}\mathbf{U}_M \mathbf{W}_M = \frac{1}{\sqrt{q}}\mathbf{U}_M V_M.$$

Hence, we have that

$$\mathbf{E}_X S_X = \frac{1}{\sqrt{p}}\mathbf{U}_X V_X S_X, \quad \mathbf{E}_M S_M = \frac{1}{\sqrt{q}}\mathbf{U}_M V_M S_M.$$

Therefore, a row of $\mathbf{E}_X S_X$ has scale $p^{-1/2}$ and a row of $\mathbf{E}_M S_M$ has scale $q^{-1/2}$. Their variances are also be of orders p^{-1} and q^{-1}, respectively. The issue here is that corresponding Gram matrix does not have an

$O_p(1)$ limit without rescaling. We therefore use the same normalization as in Lemma C.3. Conditional on the training fold, define $\Sigma_S = \mathbb{E}\left[\frac{1}{n_t} S^\top \mathbf{E}_0^\top \mathbf{E}_0 S\right]$. Using the block form of S and $\mathbf{E}_0$, we have

$$\Sigma_S = \begin{pmatrix} p^{-1}\Omega_X & 0 \\ 0 & q^{-1}\Omega_M \end{pmatrix},$$

where

$$\Omega_X = S_X^\top V_X^\top \Sigma_{\mathbf{u}_X} V_X S_X, \quad \Omega_M = S_M^\top V_M^\top \Sigma_{\mathbf{u}_M} V_M S_M.$$

The eigenvalues of Ω_X and Ω_M are bounded above and away from zero, so Σ_S is invertible. We then define the normalized error matrix $\widetilde{\mathbf{E}}_S = \mathbf{E}_0 S \Sigma_S^{-1/2}$. Since

$$\Sigma_S^{-1/2} = \begin{pmatrix} \sqrt{p}\,\Omega_X^{-1/2} & 0 \\ 0 & \sqrt{q}\,\Omega_M^{-1/2} \end{pmatrix},$$

this normalization removes the shrinking $p^{-1/2}$ and $q^{-1/2}$ scales of the original null-space columns. So, the normalized columns have order-one variance, and the argument of Lemma C.3 gives $\frac{1}{n_t}\widetilde{\mathbf{E}}_S^\top \widetilde{\mathbf{E}}_S \xrightarrow{p} I_d$. We now use this normalization only as an invertible reparameterization of the same least-squares problem. Since

$$\hat{\mathbf{Z}} U = \mathbf{Z} U + \mathbf{E}_0 U = \mathbf{F}_1 R^\top + \mathbf{E}_0 U$$

and $\hat{\mathbf{Z}} S = \mathbf{E}_0 S = \widetilde{\mathbf{E}}_S \Sigma_S^{1/2}$, we have

$$\hat{\mathbf{Z}} J = \left[\hat{\mathbf{Z}} U,\, \widetilde{\mathbf{E}}_S\right] \begin{pmatrix} I & 0 \\ 0 & \Sigma_S^{1/2} \end{pmatrix}.$$

Thus, we define $\bar{\mathbf{Z}} = \left[\hat{\mathbf{Z}} U,\, \widetilde{\mathbf{E}}_S\right]$, the regression on $\bar{\mathbf{Z}}$ is exactly the treated regression of our Algorithm 1, expressed in rescaled J-coordinates. Let $\hat{\xi} = (\hat{a}^\top \quad \hat{\lambda}^\top)^\top$, $\hat{\lambda} = \Sigma_S^{1/2}\hat{\kappa}$. Since $\delta \in \mathrm{Col}(U)$, write $\delta = U\theta$, $\theta = U^\top \delta$. The true coefficient in the normalized coordinates is then $\xi_0 = (\theta^\top \quad 0_d^\top)^\top$ Define also

$$\bar{\mathbf{H}}_1 = \begin{pmatrix} R \\ 0 \end{pmatrix}, \quad \bar{\mathbf{E}}_0 = \left[\mathbf{E}_0 U,\, \widetilde{\mathbf{E}}_S\right].$$

Then $\bar{\mathbf{Z}} = \mathbf{F}_1 \bar{\mathbf{H}}_1^\top + \bar{\mathbf{E}}_0$, while $\bar{\mathbf{E}}_0 \xi_0 = \mathbf{E}_0 U\theta = \mathbf{E}_0 \delta$. Thus, we have $Y = \mathbf{Z}\delta + \epsilon_1 = \bar{\mathbf{Z}}\xi_0 - \bar{\mathbf{E}}_0\xi_0 + \epsilon_1$. Before expanding the least-squares estimator here, we first have that the normalized Gram matrix is well conditioned. Let $\bar{M} = \frac{1}{n_t}\bar{\mathbf{Z}}^\top \bar{\mathbf{Z}}$. Its upper-left block is

$$\frac{1}{n_t}(\hat{\mathbf{Z}}U)^\top(\hat{\mathbf{Z}}U) = R\left(\frac{1}{n_t}\mathbf{F}_1^\top \mathbf{F}_1\right) R^\top + R\frac{1}{n_t}\mathbf{F}_1^\top \mathbf{E}_0 U + U^\top \frac{1}{n_t}\mathbf{E}_0^\top \mathbf{F}_1 R^\top + U^\top \frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0 U.$$

By Lemma C.2,

$$\left\| \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\left(\frac{1}{\sqrt{p \wedge q}}\right),$$

so

$$\left\| \frac{1}{n_t} \mathbf{F}_1^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\left(\frac{1}{\sqrt{n_t(p \wedge q)}}\right) = o_p(1).$$

Also, we have

$$\left\| \frac{1}{n_t} \mathbf{E}_0^\top \mathbf{E}_0 \right\|_{\mathrm{op}} = O_p\left(\frac{1}{p \wedge q}\right) = o_p(1).$$

Therefore, we get that

$$\begin{aligned} \frac{1}{n_t}(\hat{\mathbf{Z}}U)^\top(\hat{\mathbf{Z}}U) &= R\left(\frac{1}{n_t}\mathbf{F}_1^\top \mathbf{F}_1\right) R^\top + o_p(1) \\ &\xrightarrow{p} R\Omega_1 R^\top. \end{aligned}$$

For the lower-right block, Lemma C.3 gives $\frac{1}{n_t}\widetilde{\mathbf{E}}_S^\top \widetilde{\mathbf{E}}_S = I_d + o_p(1)$. For the off-diagonal block,

$$\frac{1}{n_t}(\hat{\mathbf{Z}}U)^\top \widetilde{\mathbf{E}}_S = R\frac{1}{n_t}\mathbf{F}_1^\top \widetilde{\mathbf{E}}_S + U^\top \frac{1}{n_t}\mathbf{E}_0^\top \widetilde{\mathbf{E}}_S.$$

The first term is $O_p(n_t^{-1/2})$ because the normalized error rows are mean zero, have bounded second moments, and are independent of the true factors. For the second term, if $\tilde{e}_{S,i}$ denotes the ith row of $\widetilde{\mathbf{E}}_S$ and $e_{0,i}$ the ith row of $\mathbf{E}_0$, then

$$\mathbb{E}\|e_{0,i}\|^2 = O\left(\frac{1}{p \wedge q}\right), \quad \mathbb{E}\|\tilde{e}_{S,i}\|^2 = O(1).$$

Hence, by Cauchy-Schwarz,

$$\left\|\mathbb{E}\left(e_{0,i}\tilde{e}_{S,i}^\top\right)\right\|_{\text{op}} = O\left(\frac{1}{\sqrt{p \wedge q}}\right) = o(1),$$

and the centered fluctuation is also $o_p(1)$ since the dimension is fixed. Thus $\frac{1}{n_t}(\hat{\mathbf{Z}}U)^\top \widetilde{\mathbf{E}}_S = o_p(1)$. Now, combining the three blocks,

$$\bar{M} \xrightarrow{p} \begin{pmatrix} R\Omega_1 R^\top & 0 \\ 0 & I_d \end{pmatrix}.$$

The limiting matrix is positive definite, since Ω_1 is positive definite and R is nonsingular. Therefore,

$$\|\bar{M}^{-1}\|_{\text{op}} = O_p(1). \tag{C.18}$$

We can now expand the null-space coefficient in the same way as the terms $J_1, \ldots, J_4$. Recall that

$$\bar{\mathbf{Z}} = \left[\hat{\mathbf{Z}}U,\, \widetilde{\mathbf{E}}_S\right] = \left[\mathbf{F}_1 R^\top + \mathbf{E}_0 U,\, \widetilde{\mathbf{E}}_S\right],$$

and $\bar{M} = \frac{1}{n_t}\bar{\mathbf{Z}}^\top\bar{\mathbf{Z}}$. Since $\mathbf{E}_0 S = \widetilde{\mathbf{E}}_S \Sigma_S^{1/2}$, the coefficient corresponding to the normalized null-space design $\widetilde{\mathbf{E}}_S$ is $\Sigma_S^{1/2}\hat{\kappa}$. Therefore, the least-squares estimator in the normalized coordinates satisfies

$$\begin{pmatrix} \hat{a} \\ \Sigma_S^{1/2}\hat{\kappa} \end{pmatrix} = \bar{M}^{-1}\frac{1}{n_t}\bar{\mathbf{Z}}^\top Y.$$

Since $\delta \in \text{Col}(U)$, $\delta = UU^\top\delta$. Thus, we have that

$$\bar{\mathbf{Z}}\begin{pmatrix} U^\top\delta \\ 0 \end{pmatrix} = \hat{\mathbf{Z}}UU^\top\delta = \hat{\mathbf{Z}}\delta = \mathbf{Z}\delta + \mathbf{E}_0\delta.$$

Since $Y = \mathbf{Z}\delta + \epsilon_1$, we can write

$$Y = \bar{\mathbf{Z}}\begin{pmatrix} U^\top\delta \\ 0 \end{pmatrix} - \mathbf{E}_0\delta + \epsilon_1.$$

Now, substituting this into the least-squares expression gives

$$\begin{pmatrix} \hat{a} - U^\top\delta \\ \Sigma_S^{1/2}\hat{\kappa} \end{pmatrix} = -\bar{M}^{-1}\frac{1}{n_t}\bar{\mathbf{Z}}^\top\mathbf{E}_0\delta + \bar{M}^{-1}\frac{1}{n_t}\bar{\mathbf{Z}}^\top\epsilon_1.$$

Using from earlier, $\bar{\mathbf{Z}} = \left[\mathbf{F}_1 R^\top + \mathbf{E}_0 U,\, \widetilde{\mathbf{E}}_S\right]$, we have

$$\bar{\mathbf{Z}}^\top\mathbf{E}_0\delta = \begin{pmatrix} R\mathbf{F}_1^\top\mathbf{E}_0\delta + U^\top\mathbf{E}_0^\top\mathbf{E}_0\delta \\ \widetilde{\mathbf{E}}_S^\top\mathbf{E}_0\delta \end{pmatrix},$$

and

$$\bar{\mathbf{Z}}^\top\epsilon_1 = \begin{pmatrix} R\mathbf{F}_1^\top\epsilon_1 + U^\top\mathbf{E}_0^\top\epsilon_1 \\ \widetilde{\mathbf{E}}_S^\top\epsilon_1 \end{pmatrix}.$$

Let $P_d = \begin{pmatrix} 0_{d\times(1+\bar{p}+\bar{q})} & I_d \end{pmatrix}$ select the last d coordinates. Taking the null-space coordinates and multiplying by $\sqrt{n_t}$ gives

$$\sqrt{n_t}\Sigma_S^{1/2}\hat{\kappa} = -\underbrace{P_d\bar{M}^{-1}\begin{pmatrix} R\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta \\ 0 \end{pmatrix}}_{K_1} - \underbrace{P_d\bar{M}^{-1}\begin{pmatrix} \frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top\mathbf{E}_0\delta \\ \frac{1}{\sqrt{n_t}}\widetilde{\mathbf{E}}_S^\top\mathbf{E}_0\delta \end{pmatrix}}_{K_2}$$
$$+\underbrace{P_d\bar{M}^{-1}\begin{pmatrix} R\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1 \\ 0 \end{pmatrix}}_{K_3} + \underbrace{P_d\bar{M}^{-1}\begin{pmatrix} \frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top\epsilon_1 \\ \frac{1}{\sqrt{n_t}}\widetilde{\mathbf{E}}_S^\top\epsilon_1 \end{pmatrix}}_{K_4}. \tag{C.19}$$

We now bound these four terms.

Term K_1. Using $\|\bar{M}^{-1}\|_{\mathrm{op}} = O_p(1)$, the bounded operator norm of R, and Lemma C.2,

$$\begin{aligned}\|K_1\| &\le O_p(1)\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta\right\| \\ &\le O_p(1)\left\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\right\|_{\mathrm{op}}\|\delta\| \\ &= O_p\left(\frac{1}{\sqrt{p\wedge q}}\right) = o_p(1).\end{aligned}$$

Term K_2. The first block of K_2 satisfies

$$\begin{aligned}\left\|\frac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top\mathbf{E}_0\delta\right\| &\le \sqrt{n_t}\left\|\frac{1}{n_t}\mathbf{E}_0^\top\mathbf{E}_0\right\|_{\mathrm{op}}\|\delta\| \\ &= O_p\left(\frac{\sqrt{n_t}}{p\wedge q}\right) \\ &= o_p(1).\end{aligned}$$

For the second block of K_2, instead of treating the product abstractly, we expand the normalized null-space error directly. By definition,

$$\widetilde{\mathbf{E}}_S = \mathbf{E}_0 S\Sigma_S^{-1/2}, \quad \widetilde{\mathbf{E}}_S^\top = \Sigma_S^{-1/2}S^\top\mathbf{E}_0^\top.$$

Therefore, we get that

$$\frac{1}{\sqrt{n_t}}\widetilde{\mathbf{E}}_S^\top\mathbf{E}_0\delta = \frac{1}{\sqrt{n_t}}\Sigma_S^{-1/2}S^\top\mathbf{E}_0^\top\mathbf{E}_0\delta.$$

Recall that

$$\mathbf{E}_0 = \begin{pmatrix} 0 & \mathbf{E}_X^\top & \mathbf{E}_M^\top \end{pmatrix}^\top, \quad S = \begin{pmatrix} 0 & 0 \\ S_X & 0 \\ 0 & S_M \end{pmatrix}, \quad \Sigma_S^{-1/2} = \begin{pmatrix} \sqrt{p}\,\Omega_X^{-1/2} & 0 \\ 0 & \sqrt{q}\,\Omega_M^{-1/2} \end{pmatrix}.$$

Also,

$$\mathbf{E}_0\delta = \mathbf{E}_X(\mathbf{H}_X^\dagger)^\top\beta_1 + \mathbf{E}_M(\mathbf{H}_M^\dagger)^\top\gamma_1.$$

Hence, we have

$$\begin{aligned}\frac{1}{\sqrt{n_t}}\widetilde{\mathbf{E}}_S^\top\mathbf{E}_0\delta &= \begin{pmatrix} \frac{\sqrt{p}}{\sqrt{n_t}}\Omega_X^{-1/2}S_X^\top\mathbf{E}_X^\top \\ \frac{\sqrt{q}}{\sqrt{n_t}}\Omega_M^{-1/2}S_M^\top\mathbf{E}_M^\top \end{pmatrix}\left\{\mathbf{E}_X(\mathbf{H}_X^\dagger)^\top\beta_1 + \mathbf{E}_M(\mathbf{H}_M^\dagger)^\top\gamma_1\right\} \\ &= \begin{pmatrix} \frac{\sqrt{p}}{\sqrt{n_t}}\Omega_X^{-1/2}S_X^\top\mathbf{E}_X^\top\mathbf{E}_X(\mathbf{H}_X^\dagger)^\top\beta_1 + \frac{\sqrt{p}}{\sqrt{n_t}}\Omega_X^{-1/2}S_X^\top\mathbf{E}_X^\top\mathbf{E}_M(\mathbf{H}_M^\dagger)^\top\gamma_1 \\ \frac{\sqrt{q}}{\sqrt{n_t}}\Omega_M^{-1/2}S_M^\top\mathbf{E}_M^\top\mathbf{E}_X(\mathbf{H}_X^\dagger)^\top\beta_1 + \frac{\sqrt{q}}{\sqrt{n_t}}\Omega_M^{-1/2}S_M^\top\mathbf{E}_M^\top\mathbf{E}_M(\mathbf{H}_M^\dagger)^\top\gamma_1 \end{pmatrix}.\end{aligned}$$

We bound these four terms separately. Firstly, we know that

$$\mathbb{E}(\mathbf{E}_X^\top \mathbf{E}_X) = \frac{n_t}{p} V_X^\top \Sigma_{\mathbf{u}_X} V_X, \quad \mathbb{E}(\mathbf{E}_M^\top \mathbf{E}_M) = \frac{n_t}{q} V_M^\top \Sigma_{\mathbf{u}_M} V_M.$$

Moreover, by the mutual independence of $\mathbf{u}_X$ and $\mathbf{u}_M$ using Assumption 3.1(i), $\mathbb{E}(\mathbf{E}_X^\top \mathbf{E}_M) = 0$ and $\mathbb{E}(\mathbf{E}_M^\top \mathbf{E}_X) = 0$. Since the projected dimensions are fixed and the proxy errors have uniformly bounded fourth moments,

$$\left\| \mathbf{E}_X^\top \mathbf{E}_X - \mathbb{E}(\mathbf{E}_X^\top \mathbf{E}_X) \right\|_{\text{op}} = O_p\left(\frac{\sqrt{n_t}}{p} \right),$$

and similarly to have

$$\left\| \mathbf{E}_M^\top \mathbf{E}_M - \mathbb{E}(\mathbf{E}_M^\top \mathbf{E}_M) \right\|_{\text{op}} = O_p\left(\frac{\sqrt{n_t}}{q} \right).$$

The cross-products satisfy

$$\|\mathbf{E}_X^\top \mathbf{E}_M\|_{\text{op}} = O_p\left(\sqrt{\frac{n_t}{pq}} \right), \quad \|\mathbf{E}_M^\top \mathbf{E}_X\|_{\text{op}} = O_p\left(\sqrt{\frac{n_t}{pq}} \right).$$

Consider first the $\mathbf{E}_X^\top \mathbf{E}_X$ term. Splitting it into its mean and centered parts,

$$\begin{aligned}
&\frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \mathbf{E}_X^\top \mathbf{E}_X (\mathbf{H}_X^\dagger)^\top \beta_1 \\
&\quad = \frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \left\{ \mathbf{E}_X^\top \mathbf{E}_X - \mathbb{E}(\mathbf{E}_X^\top \mathbf{E}_X) \right\} (\mathbf{H}_X^\dagger)^\top \beta_1 + \frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \mathbb{E}(\mathbf{E}_X^\top \mathbf{E}_X) (\mathbf{H}_X^\dagger)^\top \beta_1.
\end{aligned}$$

Since $\|\Omega_X^{-1/2}\|_{\text{op}} = O(1)$, $\|S_X\|_{\text{op}} = 1$, and $\|\mathbf{H}_X^\dagger\|_{\text{op}} = O_p(1)$, the centered part is

$$\frac{\sqrt{p}}{\sqrt{n_t}} O_p\left(\frac{\sqrt{n_t}}{p} \right) = O_p(p^{-1/2}),$$

while the mean part is

$$\frac{\sqrt{p}}{\sqrt{n_t}} O_p\left(\frac{n_t}{p} \right) = O_p\left(\sqrt{\frac{n_t}{p}} \right).$$

Thus, we have that

$$\frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \mathbf{E}_X^\top \mathbf{E}_X (\mathbf{H}_X^\dagger)^\top \beta_1 = O_p\left(\sqrt{\frac{n_t}{p}} + \frac{1}{\sqrt{p}} \right).$$

Now, exactly the same argument gives

$$\frac{\sqrt{q}}{\sqrt{n_t}} \Omega_M^{-1/2} S_M^\top \mathbf{E}_M^\top \mathbf{E}_M (\mathbf{H}_M^\dagger)^\top \gamma_1 = O_p\left(\sqrt{\frac{n_t}{q}} + \frac{1}{\sqrt{q}} \right).$$

For the first cross-product term, we have that

$$\begin{aligned}
\left\| \frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \mathbf{E}_X^\top \mathbf{E}_M (\mathbf{H}_M^\dagger)^\top \gamma_1 \right\| &\leq O_p(1) \frac{\sqrt{p}}{\sqrt{n_t}} \left\| \mathbf{E}_X^\top \mathbf{E}_M \right\|_{\text{op}} \\
&= O_p(1) \frac{\sqrt{p}}{\sqrt{n_t}} O_p\left(\sqrt{\frac{n_t}{pq}} \right) \\
&= O_p(q^{-1/2}).
\end{aligned}$$

Similarly, we ahve

$$\frac{\sqrt{q}}{\sqrt{n_t}} \Omega_M^{-1/2} S_M^\top \mathbf{E}_M^\top \mathbf{E}_X (\mathbf{H}_X^\dagger)^\top \beta_1 = O_p(p^{-1/2}).$$

Now, combining the four terms gives us

$$\left\| \frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \mathbf{E}_0 \delta \right\| = O_p\left(\sqrt{\frac{n_t}{p}} + \sqrt{\frac{n_t}{q}} + \frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}} \right).$$

Since $p \wedge q \gg n_t$ under Assumption 5.2, $\frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \mathbf{E}_0 \delta = o_p(1)$. Therefore, together with the bound for the first block, $K_2 = o_p(1)$.

Term K_3. By the same conditional second-moment argument used for J_3, $\left\| \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \epsilon_1 \right\| = O_p(1)$. Therefore,

$$\begin{aligned} \|K_3\| &\le \|P_d\|_{\mathrm{op}} \|\bar{M}^{-1}\|_{\mathrm{op}} \|R\|_{\mathrm{op}} \left\| \frac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \epsilon_1 \right\| \\ &= O_p(1). \end{aligned}$$

Term K_4. Recall that

$$K_4 = P_d \bar{M}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n_t}} U^\top \mathbf{E}_0^\top \epsilon_1 \\ \frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \epsilon_1 \end{pmatrix}.$$

For the first block, Lemma C.2 gives

$$\left\| \frac{1}{\sqrt{n_t}} U^\top \mathbf{E}_0^\top \epsilon_1 \right\| = O_p\left(\frac{1}{\sqrt{p \wedge q}} \right) = o_p(1).$$

For the second block, using $\widetilde{\mathbf{E}}_S = \mathbf{E}_0 S \Sigma_S^{-1/2}$,

$$\frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \epsilon_1 = \frac{1}{\sqrt{n_t}} \Sigma_S^{-1/2} S^\top \mathbf{E}_0^\top \epsilon_1.$$

Since

$$\Sigma_S^{-1/2} = \begin{pmatrix} \sqrt{p}\, \Omega_X^{-1/2} & 0 \\ 0 & \sqrt{q}\, \Omega_M^{-1/2} \end{pmatrix}, \quad S^\top \mathbf{E}_0^\top = \begin{pmatrix} S_X^\top \mathbf{E}_X^\top \\ S_M^\top \mathbf{E}_M^\top \end{pmatrix},$$

we get that

$$\frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \epsilon_1 = \begin{pmatrix} \frac{\sqrt{p}}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top \mathbf{E}_X^\top \epsilon_1 \\ \frac{\sqrt{q}}{\sqrt{n_t}} \Omega_M^{-1/2} S_M^\top \mathbf{E}_M^\top \epsilon_1 \end{pmatrix}.$$

Using $\mathbf{E}_X = (\mathbf{U}_X V_X)/\sqrt{p}$ and $\mathbf{E}_M = (\mathbf{U}_M V_M)/\sqrt{q}$, this becomes

$$\frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \epsilon_1 = \begin{pmatrix} \frac{1}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top V_X^\top \mathbf{U}_X^\top \epsilon_1 \\ \frac{1}{\sqrt{n_t}} \Omega_M^{-1/2} S_M^\top V_M^\top \mathbf{U}_M^\top \epsilon_1 \end{pmatrix}.$$

Now using the independence and mean-zero assumptions for the proxy errors and the outcome error, together with the bounded second moments and the bounded eigenvalues of Ω_X^{-1} and Ω_M^{-1}, we have that

$$\left\| \frac{1}{\sqrt{n_t}} \Omega_X^{-1/2} S_X^\top V_X^\top \mathbf{U}_X^\top \epsilon_1 \right\| = O_p(1),$$

and similarly,

$$\left\| \frac{1}{\sqrt{n_t}} \Omega_M^{-1/2} S_M^\top V_M^\top \mathbf{U}_M^\top \epsilon_1 \right\| = O_p(1).$$

Therefore, $\left\| \frac{1}{\sqrt{n_t}} \widetilde{\mathbf{E}}_S^\top \epsilon_1 \right\| = O_p(1)$.
Since $\|Q_S\|_{\mathrm{op}} = 1$ and $\|\bar{M}^{-1}\|_{\mathrm{op}} = O_p(1)$, it follows that $K_4 = O_p(1)$. Now, combining the four terms in (C.19),

$$K_1 = o_p(1), \quad K_2 = o_p(1), \quad K_3 = O_p(1), \quad K_4 = O_p(1).$$

It follows that $\sqrt{n_t} \left\| \Sigma_S^{1/2} \hat{\kappa} \right\| = O_p(1)$, or equivalently, $\left\| \Sigma_S^{1/2} \hat{\kappa} \right\| = O_p(n_t^{-1/2})$.

Finally,

$$\Sigma_S^{1/2} = \begin{pmatrix} p^{-1/2}\Omega_X^{1/2} & 0 \\ 0 & q^{-1/2}\Omega_M^{1/2} \end{pmatrix}.$$

Therefore, for the mediator null-space block, we have

$$\left\| q^{-1/2}\Omega_M^{1/2}\hat{\kappa}_M \right\| = O_p(n_t^{-1/2}).$$

Since the eigenvalues of Ω_M are bounded above and away from zero, $\|\hat{\kappa}_M\| = O_p\left(\sqrt{q/n_t}\right)$. Finally, since $\hat{\gamma}_{1,\perp} = S_M\hat{\kappa}_M$ and S_M has orthonormal columns, $\|\hat{\gamma}_{1,\perp}\| = O_p\left(\sqrt{q/n_t}\right)$. Finally, because S_M has orthonormal columns,

$$\|\hat{\gamma}_{1,\perp}\| = \|S_M\hat{\kappa}_M\| = O_p\left(\sqrt{\frac{q}{n_t}}\right). \tag{C.20}$$

Let $\mathcal{G}_t$ denote the sigma-field generated by the training fold and the treated observations used in the first-stage regression. When we condition on $\mathcal{G}_t$, $\hat{\gamma}_{1,\perp}$ and $\mathbf{W}_M$ are fixed, while the control proxy errors $\{\mathbf{u}_{M,i} : A_i = 0\}$ are independent, mean zero, and independent of the treated sample. For a control observation define $h_i = \hat{\gamma}_{1,\perp}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_{M,i}$. Using $\|\Sigma_{\mathbf{u}_M}\|_{\mathrm{op}} \le C$ and $\mathbf{W}_M^\top\mathbf{W}_M = qI_{\tilde{q}}$, we obtain

$$\begin{aligned} \mathrm{Var}\left(h_i \mid \mathcal{G}_t\right) &= \frac{1}{q^2}\hat{\gamma}_{1,\perp}^\top \mathbf{W}_M^\top \Sigma_{\mathbf{u}_M}\mathbf{W}_M\hat{\gamma}_{1,\perp} \\ &\le \frac{C}{q^2}\|\mathbf{W}_M\hat{\gamma}_{1,\perp}\|^2 \\ &= \frac{C}{q}\|\hat{\gamma}_{1,\perp}\|^2 \\ &= O_p\left(\frac{1}{n_t}\right), \end{aligned} \tag{C.21}$$

where the last equality follows from (C.20). Therefore, using the bounded second moments of the fixed-dimensional control factors and the independence of $\mathbf{f}_{X1,i}$ from $\mathbf{u}_{M,i}$,

$$\begin{aligned} \mathbb{E}\left[\left\| \frac{1}{\sqrt{n_c}} \sum_{i:A_i=0} \mathbf{f}_{X1,i}h_i \right\|^2 \,\middle|\, \mathcal{G}_t\right] &\le C\ \mathrm{Var}\left(h_i \mid \mathcal{G}_t\right) \\ &= O_p\left(\frac{1}{n_t}\right). \end{aligned} \tag{C.22}$$

Hence, we have

$$\frac{1}{\sqrt{n_c}} \sum_{i:A_i=0} \mathbf{f}_{X1,i}\hat{\gamma}_{1,\perp}^\top \frac{1}{q}\mathbf{W}_M^\top \mathbf{u}_{M,i} = O_p\left(\frac{1}{\sqrt{n_t}}\right) = o_p(1). \tag{C.23}$$

The same conclusion holds for the contraction with the control covariate projection error. Since

$$\mathbf{e}_{X,i} = \frac{1}{p}\mathbf{W}_X^\top \mathbf{u}_{X,i}, \quad \mathbb{E}\|\mathbf{e}_{X,i}\|^2 = O\left(\frac{1}{p}\right),$$

and $\mathbf{u}_{X,i}$ is independent of $\mathbf{u}_{M,i}$ by Assumption 3.1, the summands $\mathbf{e}_{X,i}h_i$ are mean zero across the control units. Moreover, conditional on $\mathcal{G}_t$,

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{e}_{X,i}h_i\|^2 \mid \mathcal{G}_t\right] &= \mathbb{E}\|\mathbf{e}_{X,i}\|^2\, \mathbb{E}\left[h_i^2 \mid \mathcal{G}_t\right] \\ &= O\left(\frac{1}{p}\right) O_p\left(\frac{1}{n_t}\right) \\ &= O_p\left(\frac{1}{pn_t}\right). \end{aligned}$$

Therefore,

$$\frac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top\left(\hat{\gamma}_{1,\perp}^\top\frac{1}{q}\mathbf{W}_M^\top\mathbf{U}_M\right)=O_p\left(\frac{1}{\sqrt{pn_t}}\right)=o_p(1). \tag{C.24}$$

Thus, although the coefficient in the mediator null space can have norm of order $\sqrt{q/n_t}$, it is multiplied in the second stage by a mediator projection error whose variance is of order q^{-1}. The two orders cancel: $\frac{1}{q}\|\hat{\gamma}_{1,\perp}\|^2 = O_p\left(1/n_t\right)$. Hence the null-space component contributes only $O_p(n_t^{-1/2})$ under the normalized contractions that enter the second-stage regression. Since, combining this with the two row-space pieces in (C.17), namely $\gamma_1^\top\mathbf{H}_M^\dagger\frac{1}{q}\mathbf{W}_M^\top\mathbf{u}_M = O_p\left(1/\sqrt{q}\right)$ and $\Delta_{\gamma,n}^\top\mathbf{H}_M^\dagger\frac{1}{q}\mathbf{W}_M^\top\mathbf{u}_M = O_p\left(\frac{1}{\sqrt{n_tq}}\right)$, the complete mediator-whitening contribution has order

$$O_p\left(\frac{1}{\sqrt{q}}+\frac{1}{\sqrt{n_tq}}+\frac{1}{\sqrt{n_t}}\right)=O_p\left(\frac{1}{\sqrt{q}}+\frac{1}{\sqrt{n_t}}\right)=o_p(1)$$

under the normalized contractions used in the second-stage regression.

Expansion of $\widehat{\gamma_1^\top\delta_0}$.
With the bridge equation in hand, the expansion proceeds as it did for $\hat{\alpha}_1$. Before running the regression, we remove from $\mathcal{R}_n$ the one piece that is linear in the regressor. The rotation slope term $\Delta_{\gamma,n}^\top\mathbf{B}_0\mathbf{H}_X^\dagger\hat{\mathbf{f}}_{X,i}$ equals $\tilde{c}^\top\hat{\mathbf{R}}_i$ with $\tilde{c}=(0,\ (\mathbf{H}_X^\dagger)^\top\mathbf{B}_0^\top\Delta_{\gamma,n})^\top$, a coefficient constant across the control units whose intercept slot is zero. Here, the least squares reproduces a term of the form $\hat{\mathbf{R}}\tilde{c}$ exactly, so its contribution to the extracted intercept is $e_\mathbf{1}^\top\tilde{c}=0$, and it is fit entirely by the discarded slope. The error term $-\Delta_{\gamma,n}^\top\mathbf{B}_0\mathbf{H}_X^\dagger\frac{1}{p}\mathbf{W}_X^\top\mathbf{u}_X$, of order $n_t^{-1/2}p^{-1/2}$, is not linear in $\hat{\mathbf{f}}_X$ and stays in the remainder. We write $\mathcal{R}_n$ for the remainder after removing the regressor-shaped term, its ith entry is now $(\mathcal{R}_n)_i=\Delta_{\gamma,n}^\top\mathbf{u}_{XM,0,i}+r_{n,i}$, containing only the two noise-type parts, since the rotation slope part has been fit exactly by the slope and contributes nothing to the intercept.

We also record the constant-shift. Since $\mathbf{b}$ contains the constant $\Delta_{\gamma,n}^\top\delta_0\cdot\mathbf{1}_{n_c}$, and $\mathbf{1}_{n_c}$ is the first column of $\hat{\mathbf{R}}$, this shift is absorbed entirely into the fitted intercept through $\hat{\mathbf{R}}^\top\mathbf{1}_{n_c}=\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\,e_\mathbf{1}$.

We estimate $\gamma_1^\top\delta_0$ by regressing $\mathbf{b}$ on $\hat{\mathbf{R}}$ over the control units, using the decomposition (C.16),

$$\begin{aligned}
\widehat{\gamma_1^\top\delta_0} &= e_\mathbf{1}^\top(\hat{\mathbf{R}}^\top\hat{\mathbf{R}})^{-1}\hat{\mathbf{R}}^\top b\\
&= e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top b\\
&= e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\left(\mathbf{R}\phi+\Delta_{\gamma,n}^\top\delta_0\cdot\mathbf{1}_{n_c}+\boldsymbol{\eta}+\mathcal{R}_n\right)\\
&= e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\left(\hat{\mathbf{R}}\phi+\mathbf{R}\phi-\hat{\mathbf{R}}\phi+\Delta_{\gamma,n}^\top\delta_0\cdot\mathbf{1}_{n_c}+\boldsymbol{\eta}+\mathcal{R}_n\right)\\
&= e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\phi+e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top(\mathbf{R}-\hat{\mathbf{R}})\phi\\
&\quad+e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\left(\Delta_{\gamma,n}^\top\delta_0\cdot\mathbf{1}_{n_c}\right)+e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\boldsymbol{\eta}\\
&\quad+e_\mathbf{1}^\top\left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\mathcal{R}_n.
\end{aligned}$$

The first term is $\phi_1=\gamma_1^\top\delta_0$, since $(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}})^{-1}\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}=I$. The third term simplifies exactly, using

$\mathbf{1}_{n_c} = \hat{\mathbf{R}} e_{\mathbf{1}}$,

$$
\begin{aligned}
& e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top \left(\Delta_{\gamma,n}^\top \delta_0 \cdot \mathbf{1}_{n_c}\right) \\
&= \Delta_{\gamma,n}^\top \delta_0 \cdot e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\, e_{\mathbf{1}} \\
&= \Delta_{\gamma,n}^\top \delta_0,
\end{aligned}
$$

so a constant added to every response of a regression with an intercept column is absorbed entirely and exactly into the fitted intercept. Hence

$$
\begin{aligned}
\widehat{\gamma_1^\top \delta_0} &= \gamma_1^\top \delta_0 + \Delta_{\gamma,n}^\top \delta_0 + e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top(\mathbf{R}-\hat{\mathbf{R}})\phi \\
&\quad + e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\boldsymbol{\eta} + e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\mathcal{R}_n \\
&= \gamma_1^\top \delta_0 + \Delta_{\gamma,n}^\top \delta_0 - e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\mathbf{E}_{X0}\phi \\
&\quad + e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\boldsymbol{\eta} + e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{n_c}\hat{\mathbf{R}}^\top\mathcal{R}_n,
\end{aligned}
$$

where the last equality uses $\mathbf{R}-\hat{\mathbf{R}} = -\mathbf{E}_{X0}$ from (B.8). Now, substituting $\hat{\mathbf{R}}^\top = \mathbf{R}^\top + \mathbf{E}_{X0}^\top = \mathbf{H}_{X1}\mathbf{F}_{X1}^\top + \mathbf{E}_{X0}^\top$ exactly as for $\hat{\alpha}_1$, and scaling by $\sqrt{n_c}$,

$$
\begin{aligned}
&\sqrt{n_c}\left(\widehat{\gamma_1^\top \delta_0} - \gamma_1^\top \delta_0\right) \qquad\qquad (C.25)\\
&= \sqrt{n_c}\,\Delta_{\gamma,n}^\top \delta_0 - \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \mathbf{H}_{X1} \frac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top \mathbf{E}_{X0}\phi}_{=L_1} - \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top \mathbf{E}_{X0}\phi}_{=L_2} \\
&\quad + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \mathbf{H}_{X1} \frac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top \boldsymbol{\eta}}_{=L_3} + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top \boldsymbol{\eta}}_{=L_4} \\
&\quad + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \mathbf{H}_{X1} \frac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top \mathcal{R}_n}_{=L_5} + \underbrace{e_{\mathbf{1}}^\top \left(\frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}\right)^{-1} \frac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top \mathcal{R}_n}_{=L_6}.
\end{aligned}
$$

We write $L_7 = \sqrt{n_c}\,\Delta_{\gamma,n}^\top \delta_0 = \sqrt{n_c}\,\delta_0^\top(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1)$,

$$
\sqrt{n_c}\left(\widehat{\gamma_1^\top \delta_0} - \gamma_1^\top \delta_0\right) = L_7 - L_1 - L_2 + L_3 + L_4 + L_5 + L_6.
$$

The terms $L_1, \ldots, L_4$ are the analogues of $G_1, \ldots, G_4$ under

$$
(\mathbf{Z}, \mathbf{H}_1, \delta, \epsilon_1, \mathbf{F}_1, \mathbf{E}_0, p \wedge q) \mapsto (\mathbf{R}, \mathbf{H}_{X1}, \phi, \boldsymbol{\eta}, \mathbf{F}_{X1}, \mathbf{E}_{X0}, p),
$$

the terms L_5, L_6 come from the remainder $\mathcal{R}_n$, and L_7 is the first-order cost of estimating γ_1.

Part 2: Asymptotic Negligibility of the Bias Terms.

We prove that the eight terms

$$
G_1,\ G_2,\ G_4, \quad L_1,\ L_2,\ L_4,\ L_5,\ L_6
$$

are each $o_p(1)$, and that the ninth term obeys $L_7 = \delta_0^\top J_3 + o_p(1)$. Throughout, $M = \frac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}$ denotes the treated Gram matrix and $\widetilde{M} = \frac{1}{n_c}\hat{\mathbf{R}}^\top\hat{\mathbf{R}}$ the control Gram matrix, and every operator-norm bound is

conditional on the training fold, on which the whitening matrices are fixed, so that the rows of $\mathbf{E}_0, \mathbf{E}_{X0}$ are independent within each stratum, while $\mathbf{E}_0, \mathbf{E}_{X0}$ and the projections $P_S, \widetilde{P}_S$ themselves remain random. Since $n_k^t \asymp n_k^c \asymp n_k \asymp n$, every rate written in n_k^t or n_k^c is order-equivalent to the corresponding rate in n. The bias terms fall into two structural patterns. The two-sided terms G_1 and L_1 carry a loading matrix on both sides of the Gram inverse, while the one-sided terms G_2, G_4, L_2, and L_4 carry a loading matrix only on the left. We record the two block-inverse identities used below.

The two reusable block-inverse identities.
We record in full the two identities from Lemma C.5 that drive every bound below.

(BI-1) Two-sided identity. Since $J = [U\ S]$ is orthogonal, $M^{-1} = J(J^\top M J)^{-1} J^\top$, and from $J^\top \mathbf{H}_1 = \binom{R}{0}$ and $\mathbf{H}_1^\top J = (R^\top\ 0)$,

$$\begin{aligned} \mathbf{H}_1^\top M^{-1} \mathbf{H}_1 &= \begin{pmatrix} R^\top & 0 \end{pmatrix} (J^\top M J)^{-1} \begin{pmatrix} R \\ 0 \end{pmatrix} \\ &= R^\top \left[(J^\top M J)^{-1} \right]_{11} R \\ &= R^\top (D + E_n)^{-1} R, \end{aligned} \tag{C.26}$$

where $\left[(J^\top M J)^{-1} \right]_{11} = (D + E_n)^{-1}$ is the top-left block of the inverse, equal to the inverse of the Schur complement $D + E_n$. With $\|R\|_{\mathrm{op}} = O(1)$ and $\|(D + E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ from Lemma C.5,

$$\left\| \mathbf{H}_1^\top M^{-1} \mathbf{H}_1 \right\|_{\mathrm{op}} \le \|R\|_{\mathrm{op}}^2 \, \|(D + E_n)^{-1}\|_{\mathrm{op}} = O_p(1). \tag{C.27}$$

(BI-2) One-sided identity. For the one-sided pattern we need the full first block-row of $(J^\top M J)^{-1}$. We write the partitioned matrix as follows:

$$J^\top M J = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}, \quad K_{11} = D + (\text{cross terms}),$$

the standard block-inverse formula gives, for the first block-row,

$$\begin{aligned} \left[(J^\top M J)^{-1} \right]_{1*} &= \left((D + E_n)^{-1},\ -(D + E_n)^{-1} K_{12} K_{22}^{-1} \right), \\ D + E_n &= K_{11} - K_{12} K_{22}^{-1} K_{21}. \end{aligned}$$

Therefore, using $\mathbf{H}_1^\top J = (R^\top\ 0)$,

$$\begin{aligned} \mathbf{H}_1^\top M^{-1} &= \begin{pmatrix} R^\top & 0 \end{pmatrix} (J^\top M J)^{-1} J^\top \\ &= R^\top (D + E_n)^{-1} \begin{pmatrix} I, & -K_{12} K_{22}^{-1} \end{pmatrix} J^\top \\ &= R^\top (D + E_n)^{-1} \left[U^\top - K_{12} K_{22}^{-1} S^\top \right]. \end{aligned} \tag{C.28}$$

Now, substituting the blocks $K_{12} = \frac{1}{n_t}(R\mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top)\mathbf{E}_0 S$ and $K_{22} = \frac{1}{n_t} S^\top \mathbf{E}_0^\top \mathbf{E}_0 S$, we get that

$$\mathbf{H}_1^\top M^{-1} = R^\top (D + E_n)^{-1} \left[U^\top - \left(\tfrac{1}{n_t} R \mathbf{F}_1^\top \mathbf{E}_0 S + \tfrac{1}{n_t} U^\top \mathbf{E}_0^\top \mathbf{E}_0 S \right) \left(\tfrac{1}{n_t} S^\top \mathbf{E}_0^\top \mathbf{E}_0 S \right)^{-1} S^\top \right]. \tag{C.29}$$

The two S-dependent factors combine into the projection P_S, so for any conformable vector w,

$$\left(\tfrac{1}{n_t} R \mathbf{F}_1^\top \mathbf{E}_0 S + \tfrac{1}{n_t} U^\top \mathbf{E}_0^\top \mathbf{E}_0 S \right) \left(\tfrac{1}{n_t} S^\top \mathbf{E}_0^\top \mathbf{E}_0 S \right)^{-1} \tfrac{1}{\sqrt{n_t}} S^\top \mathbf{E}_0^\top w = \tfrac{1}{\sqrt{n_t}} (R \mathbf{F}_1^\top + U^\top \mathbf{E}_0^\top) P_S w, \tag{C.30}$$

because $\mathbf{E}_0 S (S^\top \mathbf{E}_0^\top \mathbf{E}_0 S)^{-1} S^\top \mathbf{E}_0^\top = P_S$ and the factors of n_t cancel. Both identities have exact analogues on the control side, with $(\mathbf{H}_{X1}, \widetilde{U}, \widetilde{R}, \widetilde{S}, \widetilde{M}, \mathbf{F}_{X1}, \mathbf{E}_{X0}, \widetilde{P}_S, \widetilde{D} + \widetilde{E}_n, n_c)$ in place of $(\mathbf{H}_1, U, R, S, M, \mathbf{F}_1, \mathbf{E}_0, P_S, D + E_n, n_t)$.

Terms from $\hat{\alpha}_1$ (treated units).

$G_1 = o_p(1)$. With $v := \frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top \mathbf{E}_0\delta$, we have

$$\begin{aligned} G_1 &= e_{\mathbf{1}}^\top \left(\tfrac{1}{n_t}\hat{\mathbf{Z}}^\top\hat{\mathbf{Z}}\right)^{-1}\mathbf{H}_1 \tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta \\ &= e_{\mathbf{1}}^\top M^{-1}\mathbf{H}_1\mathbf{v} \end{aligned}$$

Now, use that $M^{-1} = J(J^\top MJ)^{-1}J^\top$ and $J^\top\mathbf{H}_1 = \binom{R}{0}$, and then simply split $e_{\mathbf{1}}^\top J = (a^\top, b^\top)$ with $a^\top = e_{\mathbf{1}}^\top U$ and $b^\top = e_{\mathbf{1}}^\top S$:

$$\begin{aligned} G_1 &= (a^\top, b^\top)(J^\top MJ)^{-1}\begin{pmatrix} R \\ 0 \end{pmatrix}\mathbf{v} \\ &= \Big[a^\top\big[(J^\top MJ)^{-1}\big]_{11} + b^\top\big[(J^\top MJ)^{-1}\big]_{21}\Big]R\mathbf{v}. \end{aligned}$$

Now use the block-inverse formula for the first block-column, and we obtain $\big[(J^\top MJ)^{-1}\big]_{11} = (D+E_n)^{-1}$ and $\big[(J^\top MJ)^{-1}\big]_{21} = -K_{22}^{-1}K_{21}(D+E_n)^{-1}$. Hence

$$G_1 = a^\top(D+E_n)^{-1}R\mathbf{v} \;-\; b^\top K_{22}^{-1}K_{21}(D+E_n)^{-1}R\mathbf{v}. \tag{C.31}$$

The second block vanishes identically. Now, as the first column of $\mathbf{H}_1 = \mathrm{diag}(1, \mathbf{H}_X, \mathbf{H}_M)$ is $e_{\mathbf{1}}$, so $e_{\mathbf{1}} \in \mathrm{Col}(\mathbf{H}_1) = \mathrm{Col}(U)$, and since $\mathrm{Col}(S) \perp \mathrm{Col}(U)$ by construction, we get that

$$b^\top = e_{\mathbf{1}}^\top S = 0. \tag{C.32}$$

Thus, we have $G_1 = a^\top(D+E_n)^{-1}R\mathbf{v}$. It remains to bound $\|\mathbf{v}\|$. By the sub-multiplicativity of the operator norm and Lemma C.2(ii), we get that

$$\begin{aligned} \|\mathbf{v}\| &= \Big\|\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\delta\Big\| \\ &\le \Big\|\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\mathbf{E}_0\Big\|_{\mathrm{op}}\|\delta\| \\ &= O_p\big((p\wedge q)^{-1/2}\big)\cdot O_p(1) \\ &= O_p\big((p\wedge q)^{-1/2}\big), \end{aligned} \tag{C.33}$$

using $\|\delta\| = O_p(1)$ (Assumption 3.1(ii)). Now, combining with $\|a\| = \|U^\top e_{\mathbf{1}}\| \le \|e_{\mathbf{1}}\| = 1$, $\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$, $\|R\|_{\mathrm{op}} = O(1)$, gives us

$$\begin{aligned} |G_1| &\le \|a\|\,\|(D+E_n)^{-1}\|_{\mathrm{op}}\,\|R\|_{\mathrm{op}}\,\|\mathbf{v}\| \\ &= O_p(1)\cdot O_p\big((p\wedge q)^{-1/2}\big) \\ &= O_p\big((p\wedge q)^{-1/2}\big) \\ &= o_p(1), \end{aligned}$$

which is negligible as $p\wedge q \to \infty$.

$G_2 = o_p(1)$. By definition, we have

$$G_2 = e_{\mathbf{1}}^\top\mathbf{H}_1^\top M^{-1}\tfrac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\mathbf{E}_0\delta.$$

Now, apply the one-sided identity (C.29) and then simplify the S-factors via (C.30) with $w = \mathbf{E}_0\delta$, we get that

$$\begin{aligned} G_2 &= e_{\mathbf{1}}^\top R^\top(D+E_n)^{-1}\Big[\tfrac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top\mathbf{E}_0\delta - \tfrac{1}{\sqrt{n_t}}\big(R\mathbf{F}_1^\top + U^\top\mathbf{E}_0^\top\big)P_S\mathbf{E}_0\delta\Big] \\ &= e_{\mathbf{1}}^\top R^\top(D+E_n)^{-1}\Big[\underbrace{\tfrac{1}{\sqrt{n_t}}U^\top\mathbf{E}_0^\top(I-P_S)\mathbf{E}_0\delta}_{=:G_{2,1}} \;-\; \underbrace{\tfrac{1}{\sqrt{n_t}}R\mathbf{F}_1^\top P_S\mathbf{E}_0\delta}_{=:G_{2,2}}\Big], \end{aligned} \tag{C.34}$$

Here, the outside scalar row $e_{\mathbf{1}}^\top R^\top (D+E_n)^{-1}$ has norm less than $\|e_{\mathbf{1}}\|\,\|R\|_{\mathrm{op}}\,\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$, so it suffices to bound $\|G_{2,1}\|$ and $\|G_{2,2}\|$.
For $G_{2,1}$, from using $\|U\|_{\mathrm{op}} = 1$, $\|I - P_S\|_{\mathrm{op}} \le 1$, and factoring out $\sqrt{n_t}$ will give us

$$\begin{aligned}
\|G_{2,1}\| &\le \tfrac{1}{\sqrt{n_t}}\,\|U\|_{\mathrm{op}}\,\big\|\mathbf{E}_0^\top (I-P_S)\mathbf{E}_0\big\|_{\mathrm{op}}\,\|\delta\| \\
&\le \tfrac{1}{\sqrt{n_t}}\,\|\mathbf{E}_0^\top \mathbf{E}_0\|_{\mathrm{op}}\,\|\delta\| \\
&= \sqrt{n_t}\,\Big\|\tfrac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\Big\|_{\mathrm{op}}\,\|\delta\| \\
&= \sqrt{n_t}\cdot O_p\Big(\tfrac{1}{p\wedge q}\Big)\cdot O_p(1) \\
&= O_p\Big(\tfrac{\sqrt{n_t}}{p\wedge q}\Big),
\end{aligned} \tag{C.35}$$

where $\|\frac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\|_{\mathrm{op}} = O_p((p\wedge q)^{-1})$ by Lemma C.2(i) and (iii).
For $G_{2,2}$, using $\|R\|_{\mathrm{op}} = O(1)$, $\|P_S\|_{\mathrm{op}} \le 1$, $\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\|_{\mathrm{op}} = O_p(1)$ (as $\frac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1$ has bounded top eigenvalue, Assumption 3.1(iii)–(iv)), and

$$\begin{aligned}
\|\mathbf{E}_0\delta\| &\le \|\mathbf{E}_0^\top \mathbf{E}_0\|_{\mathrm{op}}^{1/2}\|\delta\| \\
&= \sqrt{n_t}\,\Big\|\tfrac{1}{n_t}\mathbf{E}_0^\top \mathbf{E}_0\Big\|_{\mathrm{op}}^{1/2}\|\delta\| \\
&= \sqrt{n_t}\cdot O_p\big((p\wedge q)^{-1/2}\big)\cdot O_p(1) \\
&= O_p\Big(\sqrt{\tfrac{n_t}{p\wedge q}}\Big),
\end{aligned}$$

we obtain

$$\begin{aligned}
\|G_{2,2}\| &\le \|R\|_{\mathrm{op}}\,\Big\|\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\Big\|_{\mathrm{op}}\,\|P_S\|_{\mathrm{op}}\,\|\mathbf{E}_0\delta\| \\
&= O_p(1)\cdot O_p\Big(\sqrt{\tfrac{n_t}{p\wedge q}}\Big) \\
&= O_p\Big(\sqrt{\tfrac{n_t}{p\wedge q}}\Big).
\end{aligned} \tag{C.36}$$

Now, adding (C.35) and (C.36) and noting the second rate dominates once $p\wedge q \to \infty$ (because $\sqrt{n_t}/(p\wedge q) = \sqrt{n_t/(p\wedge q)}\cdot(p\wedge q)^{-1/2}$), we get that

$$\begin{aligned}
|G_2| &= O_p\Big(\tfrac{\sqrt{n_t}}{p\wedge q} + \sqrt{\tfrac{n_t}{p\wedge q}}\Big) \\
&= O_p\Big(\sqrt{\tfrac{n_t}{p\wedge q}}\Big).
\end{aligned}$$

Hence, $G_2 = o_p(1)$ as $n_t/(p\wedge q) \to 0$.

$G_4 = o_p(1)$. By definition, we have

$$G_4 = e_{\mathbf{1}}^\top \mathbf{H}_1^\top M^{-1}\,\tfrac{1}{\sqrt{n_t}}\mathbf{E}_0^\top \epsilon_1.$$

We apply one side identity from Equation (C.29), and fold via Equation (C.30) with $w = \epsilon_1$, we obtain

$$G_4 = e_{\mathbf{1}}^\top R^\top (D+E_n)^{-1}\Big[\underbrace{\tfrac{1}{\sqrt{n_t}}U^\top \mathbf{E}_0^\top (I-P_S)\epsilon_1}_{=:G_{4,1}} - \underbrace{\tfrac{1}{\sqrt{n_t}}R\mathbf{F}_1^\top P_S\epsilon_1}_{=:G_{4,2}}\Big]. \tag{C.37}$$

The outside factor is $O_p(1)$ as mentioned before.
Bound on $G_{4,1}$. Since $\|U\|_{\mathrm{op}} = 1$, $\|I - P_S\|_{\mathrm{op}} \le 1$, the projection $I - P_S$ cannot increase the Euclidean norm, so $G_{4,1}$ inherits the bound on $\frac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1$. Therefore, by Lemma C.2(iv),

$$\begin{aligned}
\|G_{4,1}\| &\le \Big\|\tfrac{1}{\sqrt{n_t}}\mathbf{E}_0^\top\epsilon_1\Big\|_2 \\
&= O_p\left(\frac{1}{\sqrt{p\wedge q}}\right).
\end{aligned} \tag{C.38}$$

Bound on $G_{4,2}$. First, we decompose $\mathbf{F}_1 = \mathbf{1}\mu_{\mathbf{F}}^\top + \widetilde{\mathbf{F}}_1$ into mean and centered parts, so we obtain

$$\frac{1}{\sqrt{n_t}} R\mathbf{F}_1^\top P_S \epsilon_1 = \frac{1}{\sqrt{n_t}} R\mu_{\mathbf{F}} \mathbf{1}^\top P_S \epsilon_1 + \frac{1}{\sqrt{n_t}} R\widetilde{\mathbf{F}}_1^\top P_S \epsilon_1$$

For the mean part, the treated-arm exogeneity of Assumption 3.1(iv) gives $\mathbb{E}(\epsilon_1) = 0$, and since P_S is a function of $\mathbf{E}_0$, we have $\mathbb{E}(\epsilon_1 \mid P_S) = E(\epsilon_1) = 0$. Hence, conditional on P_S, $\mathbf{1}^\top P_S \epsilon_1$ has mean zero and, with $\mathbb{E}(\epsilon_1 \epsilon_1^\top \mid P_S) \preceq C\, I$ (Assumption 5.1(ii)), we obtain

$$\begin{aligned} \mathbb{E}\big[(\mathbf{1}^\top P_S \epsilon_1)^2 \mid P_S\big] &= \mathbf{1}^\top P_S\, \mathbb{E}(\epsilon_1 \epsilon_1^\top \mid P_S)\, P_S \mathbf{1} \\ &\le C\, \mathbf{1}^\top P_S^2 \mathbf{1} \\ &= C\, \mathbf{1}^\top P_S \mathbf{1} \\ &= O_p(1) \end{aligned} \tag{C.39}$$

the last equality by Lemma C.3 and $P_S^2 = P_S$. Now, using Chebyshev inequality gives that $\mathbf{1}^\top P_S \epsilon_1 = O_p(1)$. Since $\|R\|_{\mathrm{op}} = O(1)$ and $\|\mu_{\mathbf{F}}\|_2 = O(1)$ (Assumption 3.1(ii)), we have

$$\begin{aligned} \left\| \tfrac{1}{\sqrt{n_t}} R\mu_{\mathbf{F}} \mathbf{1}^\top P_S \epsilon_1 \right\| &= \tfrac{1}{\sqrt{n_t}} \|R\|_{\mathrm{op}} \|\mu_{\mathbf{F}}\|_2\, |\mathbf{1}^\top P_S \epsilon_1| \\ &= O_p\Big(\frac{1}{\sqrt{n_t}}\Big). \end{aligned} \tag{C.40}$$

For the centered part, we have conditional on $(\widetilde{\mathbf{F}}_1, \mathbf{E}_0)$, ϵ_1 is mean zero, and

$$\begin{aligned} &\mathbb{E}\Big[\Big\| \tfrac{1}{\sqrt{n_t}} R\widetilde{\mathbf{F}}_1^\top P_S \epsilon_1 \Big\|^2 \,\Big|\, \widetilde{\mathbf{F}}_1, \mathbf{E}_0\Big] \\ &= \tfrac{1}{n_t} \operatorname{tr}\Big(R\widetilde{\mathbf{F}}_1^\top P_S\, \mathbb{E}(\epsilon_1 \epsilon_1^\top \mid \widetilde{\mathbf{F}}_1, \mathbf{E}_0)\, P_S \widetilde{\mathbf{F}}_1 R^\top \Big) \\ &\le \tfrac{C}{n_t} \|R\|_{\mathrm{op}}^2 \operatorname{tr}\big(\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1\big) \\ &= \tfrac{C}{n_t} \cdot O_p(1) \\ &= O_p\Big(\frac{1}{\sqrt{n_t}}\Big), \end{aligned} \tag{C.41}$$

where $\operatorname{tr}(\widetilde{\mathbf{F}}_1^\top P_S \widetilde{\mathbf{F}}_1) = O_p(1)$ by Lemma C.4. Now, we have $\|\frac{1}{\sqrt{n_t}} R\widetilde{\mathbf{F}}_1^\top P_S \epsilon_1\| = O_p(n_t^{-1/2})$. Thus, adding the two parts, $\|G_{4,2}\| = O_p(n_t^{-1/2})$ and (C.38) and combining with the $O_p(1)$ outside factor, we obtain

$$|G_4| = O_p\left(\frac{1}{\sqrt{p \wedge q}} + \frac{1}{\sqrt{n_t}} \right)$$

which needs only $p \wedge q \to \infty$ and $n_t \to \infty$ separately. Therefore, the three treated-side bias terms are negligible $G_1, G_2, G_4 = o_p(1)$, leaving G_3 as the only surviving treated term.

Terms from $\widehat{\gamma_1^\top \delta_0}$ (control units).
Before bounding L_1, L_2, L_4, we rewrite the control-side versions of the tools used for G_1, G_2, G_4. Each is obtained by repeating an earlier under the substitution as

$$(\mathbf{F}_1,\ \mathbf{E}_0,\ S,\ M,\ P_S,\ p \wedge q) \ \longrightarrow\ (\mathbf{F}_{X1},\ \mathbf{E}_{X0},\ \widetilde{S},\ \widetilde{M},\ \widetilde{P}_S,\ p),$$

where p replaces $p \wedge q$ because the second-stage regression involves only the X-side error block $\mathbf{E}_{X\cdot}$. Now, running the proof of Lemma C.5 under this substitution produces the control Schur complement $\widetilde{D} + \widetilde{E}_n$ of $\widetilde{J}^\top \widetilde{M} \widetilde{J}$, with leading block $\widetilde{D} = \frac{1}{n_c} \widetilde{R} \mathbf{F}_{X1}^\top \mathbf{F}_{X1} \widetilde{R}^\top$. This gives exactly as before, $\lambda_{\min}(\widetilde{D}) \ge c \cdot c_{\widetilde{R}}^2 > 0$ (Assumption 3.1(iii) on the control part), and the four-piece bound of (C.7) gives us that

$$\begin{aligned} \|\widetilde{E}_n\| &= O_p\Big(\tfrac{1 + \|\mu_{\mathbf{F}_X}\|_2 + \|\mu_{\mathbf{F}_X}\|_2^2}{n_c} + \tfrac{1}{\sqrt{n_c\, p}} + \tfrac{1}{p} \Big), \\ \|(\widetilde{D} + \widetilde{E}_n)^{-1}\|_{\mathrm{op}} &= O_p(1). \end{aligned} \tag{C.42}$$

The block-inverse identities (C.26)–(C.30) likewise hold verbatim in the substituted notation, and Lemma C.3 applies to $\widetilde{P}_S$ unchanged as $\mathbf{E}_{X0}$ carries only the single error block $\mathbf{E}_X$, the rates of Lemma C.2 hold with p in place of $p \wedge q$. Finally, the coefficient vector is bounded, $\|\phi\| = O_p(1)$, since $\phi = (\gamma_1^\top \delta_0,\ \gamma_1^\top \mathbf{B}_0 \mathbf{H}_X^\dagger)$ collects $\gamma_1^\top \delta_0$ together with a bounded linear combination of γ_1 and $\mathbf{B}_0$, all bounded by Assumption 3.1(ii), with $\nu_{\min}(\mathbf{H}_X)$ bounded below by Proposition 5.1.

$L_1 = o_p(1)$. By definition

$$L_1 = e_{\mathbf{1}}^\top \widetilde{M}^{-1} \mathbf{H}_{X1}\, \widetilde{\mathbf{v}}, \quad \widetilde{\mathbf{v}} := \tfrac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathbf{E}_{X0} \phi,$$

is identical in form to G_1. The control two-sided identity gives us exactly as in (C.31)-(C.32),

$$L_1 = \widetilde{a}^\top (\widetilde{D} + \widetilde{E}_n)^{-1} \widetilde{R}\, \widetilde{\mathbf{v}} \;-\; \widetilde{b}^\top \widetilde{K}_{22}^{-1} \widetilde{K}_{21} (\widetilde{D} + \widetilde{E}_n)^{-1} \widetilde{R}\, \widetilde{\mathbf{v}}, \quad \widetilde{a}^\top = e_{\mathbf{1}}^\top \widetilde{U}, \quad \widetilde{b}^\top = e_{\mathbf{1}}^\top \widetilde{S}.$$

The second block vanishes because the first column of $\mathbf{H}_{X1} = \mathrm{diag}(1, \mathbf{H}_X)$ is $e_{\mathbf{1}}$, so $e_{\mathbf{1}} \in \mathrm{Col}(\mathbf{H}_{X1}) = \mathrm{Col}(\widetilde{U}) \perp \mathrm{Col}(\widetilde{S})$, giving $\widetilde{b}^\top = e_{\mathbf{1}}^\top \widetilde{S} = 0$. Thus, $L_1 = \widetilde{a}^\top (\widetilde{D} + \widetilde{E}_n)^{-1} \widetilde{R}\, \widetilde{\mathbf{v}}$, and by the X-only Lemma C.2(ii),

$$\begin{aligned} \|\widetilde{\mathbf{v}}\| &\le \left\| \tfrac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathbf{E}_{X0} \right\|_{\mathrm{op}} \|\phi\| \\ &= O_p\Big(\frac{1}{\sqrt{p}}\Big) \cdot O_p(1) \\ &= O_p\Big(\frac{1}{\sqrt{p}}\Big). \end{aligned}$$

With $\|\widetilde{a}\| \le 1$, $\|(\widetilde{D} + \widetilde{E}_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ and $\|\widetilde{R}\|_{\mathrm{op}} = O(1)$, we have

$$\begin{aligned} |L_1| &\le \|\widetilde{a}\|\, \|(\widetilde{D} + \widetilde{E}_n)^{-1}\|_{\mathrm{op}}\, \|\widetilde{R}\|_{\mathrm{op}}\, \|\widetilde{\mathbf{v}}\| \\ &= O_p\Big(\frac{1}{\sqrt{p}}\Big) \\ &= o_p(1) \end{aligned}$$

needing $p \to \infty$.

$L_2 = o_p(1)$. By definition,

$$L_2 = e_{\mathbf{1}}^\top \mathbf{H}_{X1}^\top \widetilde{M}^{-1}\, \tfrac{1}{\sqrt{n_c}} \mathbf{E}_{X0}^\top \mathbf{E}_{X0} \phi.$$

This has the one-sided structure of G_2, with $\mathbf{H}_{X1}$ on the left only. Applying the control one-sided identity and folding the $\widetilde{S}$-factors via the control analogue of (C.30) with $w = \mathbf{E}_{X0}\phi$,

$$L_2 = e_{\mathbf{1}}^\top \widetilde{R}^\top (\widetilde{D} + \widetilde{E}_n)^{-1} \Big[\underbrace{\tfrac{1}{\sqrt{n_c}} \widetilde{U}^\top \mathbf{E}_{X0}^\top (I - \widetilde{P}_S) \mathbf{E}_{X0} \phi}_{=: L_{2,1}} - \underbrace{\tfrac{1}{\sqrt{n_c}} \widetilde{R} \mathbf{F}_{X1}^\top \widetilde{P}_S \mathbf{E}_{X0} \phi}_{=: L_{2,2}} \Big].$$

The outside factor $e_{\mathbf{1}}^\top \widetilde{R}^\top (\widetilde{D} + \widetilde{E}_n)^{-1}$ is $O_p(1)$, so it suffices to bound $L_{2,1}$ and $L_{2,2}$. Both rely on the X-only Lemma C.2(iii),

$$\left\| \tfrac{1}{n_c} \mathbf{E}_{X0}^\top \mathbf{E}_{X0} \right\|_{\mathrm{op}} = O_p\Big(\frac{1}{p}\Big).$$

For the bound on $L_{2,1}$. Since $\|\widetilde{U}\|_{\mathrm{op}} = 1$ and $\|I - \widetilde{P}_S\|_{\mathrm{op}} \le 1$, the projection cannot increase the norm, so $L_{2,1}$ reduces to a quadratic form in $\mathbf{E}_{X0}$. Factoring out $\sqrt{n_c}$,

$$\begin{aligned} \|L_{2,1}\| &\le \tfrac{1}{\sqrt{n_c}} \|\mathbf{E}_{X0}^\top \mathbf{E}_{X0}\|_{\mathrm{op}}\, \|\phi\| \\ &= \sqrt{n_c} \left\| \tfrac{1}{n_c} \mathbf{E}_{X0}^\top \mathbf{E}_{X0} \right\|_{\mathrm{op}} \|\phi\| \\ &= O_p\Big(\tfrac{\sqrt{n_c}}{p}\Big), \end{aligned}$$

exactly as in (C.35). For the bound on $L_{2,2}$. Here we use $\|\widetilde{R}\|_{\mathrm{op}} = O(1)$, $\|\frac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top\|_{\mathrm{op}} = O_p(1)$, and $\|\widetilde{P}_S\|_{\mathrm{op}} \le 1$, together with

$$\begin{aligned}\|\mathbf{E}_{X0}\phi\| &\le \|\mathbf{E}_{X0}^\top\mathbf{E}_{X0}\|_{\mathrm{op}}^{1/2}\,\|\phi\| \\ &= \sqrt{n_c}\left\|\tfrac{1}{n_c}\mathbf{E}_{X0}^\top\mathbf{E}_{X0}\right\|_{\mathrm{op}}^{1/2}\|\phi\| \\ &= O_p\Big(\sqrt{\tfrac{n_c}{p}}\Big).\end{aligned}$$

Here, multiplying the bounded factors gives us,

$$\begin{aligned}\|L_{2,2}\| &\le \|\widetilde{R}\|_{\mathrm{op}}\left\|\tfrac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top\right\|_{\mathrm{op}}\|\widetilde{P}_S\|_{\mathrm{op}}\,\|\mathbf{E}_{X0}\phi\| \\ &= O_p\Big(\sqrt{\tfrac{n_c}{p}}\Big).\end{aligned}$$

Now, adding the two bounds, and noting the second dominates once $p \to \infty$ (as $\sqrt{n_c}/p = \sqrt{n_c/p}\cdot p^{-1/2}$), we get that

$$\begin{aligned}|L_2| &= O_p\Big(\tfrac{\sqrt{n_c}}{p} + \sqrt{\tfrac{n_c}{p}}\Big) \\ &= O_p\Big(\sqrt{\tfrac{n_c}{p}}\Big) \\ &= o_p(1),\end{aligned}$$

under $n_c/p \to 0$, the control analogue of G_2's relative growth requirement.

$L_4 = o_p(1)$**.** (this and next one need to be checked again tomorrow) We have that,

$$L_4 = e_{\mathbf{1}}^\top\mathbf{H}_{X1}^\top\widetilde{M}^{-1}\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top\boldsymbol{\eta}, \quad \boldsymbol{\eta} = \gamma_1^\top\mathbf{u}_{XM,0} \ \text{ (control units)}.$$

This is the one-sided pattern of G_4, now with the control noise $\boldsymbol{\eta}$ in place of ϵ_1. Now, applying the control one-sided identity and folding the $\widetilde{S}$-factors via (C.30) with $w = \boldsymbol{\eta}$, we obtain

$$L_4 = e_{\mathbf{1}}^\top\widetilde{R}^\top(\widetilde{D}+\widetilde{E}_n)^{-1}\Big[\underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{U}^\top\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\boldsymbol{\eta}}_{=:\,L_{4,1}} - \underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{R}\mathbf{F}_{X1}^\top\widetilde{P}_S\boldsymbol{\eta}}_{=:\,L_{4,2}}\Big].$$

The outside factor $e_{\mathbf{1}}^\top\widetilde{R}^\top(\widetilde{D}+\widetilde{E}_n)^{-1}$ is $O_p(1)$, so it suffices to bound $L_{4,1}$ and $L_{4,2}$. For the bound on $L_{4,1}$, condition on $\mathbf{E}_{X0}$. Since $\boldsymbol{\eta}$ is independent of $\mathbf{E}_{X0}$, has mean zero, and has covariance bounded by CI_{n_c},

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\sqrt{n_c}}\widetilde{U}^\top\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\boldsymbol{\eta}\right\|^2\,\middle|\,\mathbf{E}_{X0}\right] &\le \frac{C}{n_c}\operatorname{tr}\left\{\widetilde{U}^\top\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathbf{E}_{X0}\widetilde{U}\right\} \\ &\le \frac{C}{n_c}\|\mathbf{E}_{X0}\|_{\mathrm{op}}^2 \\ &= O_p\left(\frac{1}{p}\right).\end{aligned}$$

Thus, $\|L_{4,1}\| = O_p(p^{-1/2})$.

For the bound on $L_{4,2}$. We split the factor matrix into mean and centered parts, $\mathbf{F}_{X1} = \mathbf{1}\mu_{\mathbf{F}_X}^\top + \widetilde{\mathbf{F}}_{X1}$, so that

$$\tfrac{1}{\sqrt{n_c}}\widetilde{R}\mathbf{F}_{X1}^\top\widetilde{P}_S\boldsymbol{\eta} = \underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{R}\mu_{\mathbf{F}_X}\mathbf{1}^\top\widetilde{P}_S\boldsymbol{\eta}}_{\text{mean part}} + \underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{R}\widetilde{\mathbf{F}}_{X1}^\top\widetilde{P}_S\boldsymbol{\eta}}_{\text{centered part}}.$$

Since the control-arm exogeneity of Assumption 3.1(ii) gives $\mathbb{E}(\mathbf{u}_{XM,0,i} \mid \mathbf{f}_{X,i}, \mathbf{u}_{X,i}, \mathbf{u}_{M,i}, A_i = 0) = 0$, and $\widetilde{P}_S$ is a function of $\mathbf{E}_{X0}$, we have $\mathbb{E}(\boldsymbol{\eta} \mid \widetilde{P}_S) = 0$. Hence, conditional on $\widetilde{P}_S$, the mean part has a bounded second moment, exactly as in (C.39):

$$\mathbb{E}\big[(\mathbf{1}^\top \widetilde{P}_S \boldsymbol{\eta})^2 \mid \widetilde{P}_S\big] \le C\, \mathbf{1}^\top \widetilde{P}_S \mathbf{1} = O_p(1) \quad \Longrightarrow \quad \mathbf{1}^\top \widetilde{P}_S \boldsymbol{\eta} = O_p(1),$$

the last step by Chebyshev inequality, so that $\big\| \frac{1}{\sqrt{n_c}} \widetilde{R} \mu_{\mathbf{F}_X} \mathbf{1}^\top \widetilde{P}_S \boldsymbol{\eta} \big\| = O_p\left(\frac{1}{\sqrt{n_c}}\right)$.

For the centered part, conditioning on $(\widetilde{\mathbf{F}}_{X1}, \mathbf{E}_{X0})$ and arguing exactly as in (C.41) gives as

$$\begin{aligned}\mathbb{E}\Big[\Big\| \tfrac{1}{\sqrt{n_c}} \widetilde{R} \widetilde{\mathbf{F}}_{X1}^\top \widetilde{P}_S \boldsymbol{\eta} \Big\|^2 \,\Big|\, \widetilde{\mathbf{F}}_{X1}, \mathbf{E}_{X0}\Big] &\le \tfrac{C}{n_c} \operatorname{tr}\left(\widetilde{\mathbf{F}}_{X1}^\top \widetilde{P}_S \widetilde{\mathbf{F}}_{X1}\right) \\ &= O_p\left(\frac{1}{n_c}\right),\end{aligned}$$

where $\operatorname{tr}(\widetilde{\mathbf{F}}_{X1}^\top \widetilde{P}_S \widetilde{\mathbf{F}}_{X1}) = O_p(1)$ by Lemma C.4. So, the centered part is $O_p(n_c^{-1/2})$ as well. Now, adding the two gives us

$$|L_4| = O_p\big(p^{-1/2} + n_c^{-1/2}\big) = o_p(1).$$

$L_5 = o_p(1)$**.** By definition, we have that

$$L_5 = e_{\mathbf{1}}^\top \mathbf{H}_{X1}^\top \widetilde{M}^{-1} \mathbf{H}_{X1} \tfrac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathcal{R}_n.$$

Here, the left factor is $\|e_{\mathbf{1}}^\top \mathbf{H}_{X1}^\top \widetilde{M}^{-1} \mathbf{H}_{X1}\|_{\mathrm{op}} = O_p(1)$ by (C.27), so it suffices to bound the contraction $\frac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathcal{R}_n$. Now, we split the remainder by noise source,

$$\mathcal{R}_n = \mathbf{E}_{X0} \mathbf{a}_n + \mathcal{R}_n^M, \quad \mathbf{a}_n = -(\mathbf{H}_X^\dagger)^\top B_0^\top (\gamma_1 + \Delta_{\gamma,n}), \quad \|\mathbf{a}_n\| = O_p(1),$$

where the covariate part $\mathbf{E}_{X0}\mathbf{a}_n$ depends only on $\mathbf{u}_X$, and the mediator part $\mathcal{R}_n^M$ depends only on $(\mathbf{u}_M, \mathbf{u}_{XM,0})$, with $r_n^M := (\max_i \mathbb{E}[(r_{n,i}^M)^2])^{1/2} = O_p(q^{-1/2} + n_t^{-1/2})$. Since $\mathcal{R}_n^M$ is independent of $\mathbf{E}_{X0}$, we condition on $\mathbf{E}_{X0}$ throughout, under which the covariate part is fixed. The rows of $\mathbf{F}_{X1}$ are the true factors and are independent of the control noise, and we split them into their mean and fluctuation, $\mathbf{F}_{X1} = \mathbf{1}\mu_X^\top + \mathbf{F}_X$ with $\mu_X = \mathbb{E}(\mathbf{f}_{X1})$.

Now, for the covariate part, conditioning on $\mathbf{E}_{X0}$, the deterministic contribution satisfies

$$\begin{aligned}\Big| \tfrac{1}{\sqrt{n_c}} \mu_X \mathbf{1}^\top \mathbf{E}_{X0} \mathbf{a}_n \Big| &\le \|\mu_X\| \tfrac{1}{\sqrt{n_c}} \|\mathbf{1}\| \, \|\mathbf{E}_{X0}\|_{\mathrm{op}} \, \|\mathbf{a}_n\| \\ &= \|\mu_X\| \tfrac{1}{\sqrt{n_c}} \sqrt{n_c}\, O_p\big(\sqrt{n_c/p}\big)\, O_p(1) \\ &= O_p\Big(\sqrt{\frac{n_c}{p}}\Big),\end{aligned}$$

where we have used $\|\mathbf{1}\| = \sqrt{n_c}$ and $\|\mathbf{E}_{X0}\|_{\mathrm{op}} = O_p(\sqrt{n_c/p})$ from Lemma C.2, and this is $o_p(1)$ under Assumption 5.2. For the fluctuation contribution, we have that it is conditionally mean zero, and

$$\begin{aligned}\mathbb{E}\Big[\Big\| \tfrac{1}{\sqrt{n_c}} \mathbf{F}_X^\top \mathbf{E}_{X0} \mathbf{a}_n \Big\|^2 \,\Big|\, \mathbf{E}_{X0}\Big] &\le \tfrac{C}{n_c} \|\mathbf{E}_{X0}\|_{\mathrm{op}}^2 \, \|\mathbf{a}_n\|^2 \\ &= O_p\left(\frac{1}{p}\right),\end{aligned}$$

so we get that the fluctuation contribution is $O_p(p^{-1/2})$.

Now, for the mediator part, since $\mathcal{R}_n^M$ is mean zero and independent of $\mathbf{E}_{X0}$, and the rows of $\mathbf{F}_{X1}$ are independent of the control noise, the sum $\frac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathcal{R}_n^M$ is mean zero, and we have that

$$\begin{aligned}\mathbb{E}\Big[\Big\| \tfrac{1}{\sqrt{n_c}} \mathbf{F}_{X1}^\top \mathcal{R}_n^M \Big\|^2\Big] &\le \big(\max_i \mathbb{E}\|\mathbf{f}_{X1,i}\|^2\big)\, (r_n^M)^2 \\ &= O_p\Big(\frac{1}{q} + \frac{1}{n_t}\Big),\end{aligned}$$

so it is $O_p(q^{-1/2}+n_t^{-1/2})$. Now, combining the covariate and mediator parts, we get that

$$\left\|\tfrac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top\mathcal{R}_n\right\| = O_p\Big(\tfrac{1}{\sqrt{p}}+\tfrac{1}{\sqrt{q}}+\tfrac{1}{\sqrt{n_t}}\Big),$$

so with the $O_p(1)$ left factor,

$$|L_5| = O_p\Big(\tfrac{1}{\sqrt{p}}+\tfrac{1}{\sqrt{q}}+\tfrac{1}{\sqrt{n_t}}\Big) = o_p(1).$$

$L_6 = o_p(1)$**.** By definition, we have that

$$L_6 = e_{\mathbf{1}}^\top \mathbf{H}_{X1}^\top \widetilde{M}^{-1}\,\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top\mathcal{R}_n.$$

This is the one-sided pattern of L_4, with the remainder $\mathcal{R}_n$ in place of the control noise $\boldsymbol{\eta}$. Now, keeping the fixed vector $e_{\mathbf{1}}^\top$ attached and folding the $\widetilde{S}$-factors through the control analogue of (C.30) with $w=\mathcal{R}_n$, we get that

$$L_6 = e_{\mathbf{1}}^\top \widetilde{R}^\top(\widetilde{D}+\widetilde{E}_n)^{-1}\Big[\underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{U}^\top\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathcal{R}_n}_{=:\,L_{6,1}} - \underbrace{\tfrac{1}{\sqrt{n_c}}\widetilde{R}\,\mathbf{F}_{X1}^\top\widetilde{P}_S\mathcal{R}_n}_{=:\,L_{6,2}}\Big].$$

Here, the outside factor is $O_p(1)$, since $\|e_{\mathbf{1}}^\top\widetilde{R}^\top\| = O(1)$ and $\|(\widetilde{D}+\widetilde{E}_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ by (C.42), so it suffices to bound $L_{6,1}$ and $L_{6,2}$.

Now, both the projection $\widetilde{P}_S$ and the contracting matrix $\mathbf{E}_{X0}^\top$ are functions of the control proxy errors $\mathbf{E}_{X0}$, and the remainder $\mathcal{R}_n$ carries a covariate part built from the same errors, so we condition on $\mathbf{E}_{X0}$ throughout. As in L_5, we have that $\mathcal{R}_n = \mathbf{E}_{X0}\mathbf{a}_n + \mathcal{R}_n^M$, and under this conditioning $\widetilde{P}_S$ and $\mathbf{E}_{X0}\mathbf{a}_n$ are fixed.

Bound on $L_{6,1}$**.** Since $\|\widetilde{U}\|_{\mathrm{op}}=1$, it suffices to bound $\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathcal{R}_n$. Now, for the covariate part, conditioning on $\mathbf{E}_{X0}$ makes $\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathbf{E}_{X0}\mathbf{a}_n$ deterministic, so we bound it directly in norm,

$$\begin{aligned}\left\|\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathbf{E}_{X0}\mathbf{a}_n\right\| &\le \tfrac{1}{\sqrt{n_c}}\,\|\mathbf{E}_{X0}\|_{\mathrm{op}}^2\,\|\mathbf{a}_n\| \\ &= \tfrac{1}{\sqrt{n_c}}\,O_p\Big(\tfrac{n_c}{p}\Big)\,O_p(1) \\ &= O_p\Big(\tfrac{\sqrt{n_c}}{p}\Big),\end{aligned}$$

where we have used $\|I-\widetilde{P}_S\|_{\mathrm{op}}\le 1$ and $\|\mathbf{E}_{X0}\|_{\mathrm{op}}^2 = O_p(n_c/p)$ from Lemma C.2. Now, under Assumption 5.2, $(p\wedge q)/n\to\infty$ gives $p\gg n\ge n_c$, so $\sqrt{n_c}/p\to 0$ and this part is $o_p(1)$. For the mediator part, since $\mathcal{R}_n^M$ is mean zero and independent of $\mathbf{E}_{X0}$, we have that $\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathcal{R}_n^M$ is conditionally mean zero, and

$$\begin{aligned}\mathbb{E}\Big[\Big\|\tfrac{1}{\sqrt{n_c}}\mathbf{E}_{X0}^\top(I-\widetilde{P}_S)\mathcal{R}_n^M\Big\|^2\,\Big|\,\mathbf{E}_{X0}\Big] &\le \tfrac{C}{n_c}\,\|\mathbf{E}_{X0}\|_{\mathrm{op}}^2\,(r_n^M)^2 \\ &= O_p\Big(\tfrac{1}{p}\Big)\,O_p\Big(\frac{1}{q}+\frac{1}{n_t}\Big),\end{aligned}$$

so this part is $O_p\big(p^{-1/2}(q^{-1/2}+n_t^{-1/2})\big)$, which is smaller than $q^{-1/2}+n_t^{-1/2}$. Now, the rotation part $\Delta_{\gamma,n}^\top\mathbf{u}_{XM,0,i}$, with $\Delta_{\gamma,n}=O_p(n_t^{-1/2})$ independent of $\mathbf{E}_{X0}$, contributes $O_p(n_t^{-1/2})$. Combining, we get that

$$\|L_{6,1}\| = O_p\Big(\tfrac{\sqrt{n_c}}{p}+\tfrac{1}{\sqrt{q}}+\tfrac{1}{\sqrt{n_t}}\Big) = o_p(1).$$

Bound on $L_{6,2}$**.** Using $\|\widetilde{R}\|_{\mathrm{op}}=O(1)$, we have that

$$\|L_{6,2}\| \le \|\widetilde{R}\|_{\mathrm{op}}\left\|\tfrac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top\widetilde{P}_S\mathcal{R}_n\right\|.$$

Now, the contraction is by $\mathbf{F}_{X1}^\top$, so we split the factors into mean and fluctuation, $\mathbf{F}_{X1} = \mathbf{1}\mu_X^\top + \mathbf{F}_X$ with $\mu_X = \mathbb{E}(\mathbf{f}_{X1})$. For the covariate part, the deterministic contribution uses the fixed-rank projection quadratic form $\mathbf{1}^\top \widetilde{P}_S \mathbf{1} \le \|\mathbf{1}\|^2 = n_c$, so we have that

$$\begin{aligned}\left|\tfrac{1}{\sqrt{n_c}}\mu_X \mathbf{1}^\top \widetilde{P}_S \mathbf{E}_{X0}\mathbf{a}_n\right| &\le \|\mu_X\| \{\mathbf{1}^\top \widetilde{P}_S \mathbf{1}\}^{1/2} \tfrac{1}{\sqrt{n_c}}\|\mathbf{E}_{X0}\mathbf{a}_n\| \\ &\le \|\mu_X\| \sqrt{n_c}\, \tfrac{1}{\sqrt{n_c}} \|\mathbf{E}_{X0}\|_{\mathrm{op}} \|\mathbf{a}_n\| \\ &= O_p\left(\frac{1}{\sqrt{p}}\right),\end{aligned}$$

where we have used $\|\mathbf{E}_{X0}\|_{\mathrm{op}} = O_p(\sqrt{n_c/p})$. For the fluctuation contribution, conditioning on $\mathbf{E}_{X0}$ makes $\widetilde{P}_S \mathbf{E}_{X0}\mathbf{a}_n$ fixed, and it is conditionally mean zero, so

$$\begin{aligned}&\mathbb{E}\Big[\Big\|\tfrac{1}{\sqrt{n_c}}\mathbf{F}_X^\top \widetilde{P}_S \mathbf{E}_{X0}\mathbf{a}_n\Big\|^2 \,\Big|\, \mathbf{E}_{X0}\Big] \\ &\le \tfrac{C}{n_c}\, \|\widetilde{P}_S \mathbf{E}_{X0}\mathbf{a}_n\|^2 \\ &\le \tfrac{C}{n_c}\, \|\mathbf{E}_{X0}\mathbf{a}_n\|^2 \\ &= O_p\left(\frac{1}{p}\right),\end{aligned}$$

so the fluctuation contribution is $O_p(p^{-1/2})$. Now, for the mediator part, since $\mathcal{R}_n^M$ is independent of $\mathbf{E}_{X0}$ and mean zero, we have that $\tfrac{1}{\sqrt{n_c}}\mathbf{F}_{X1}^\top \widetilde{P}_S \mathcal{R}_n^M$ is conditionally mean zero with second moment $O_p(q^{-1} + n_t^{-1})$, so it is $O_p(q^{-1/2} + n_t^{-1/2})$. Combining, we get that

$$\|L_{6,2}\| = O_p\Big(\tfrac{1}{\sqrt{p}} + \tfrac{1}{\sqrt{q}} + \tfrac{1}{\sqrt{n_t}}\Big) = o_p(1).$$

Now, combining the two pieces with the $O_p(1)$ outside factor, we get that

$$|L_6| = O_p\Big(\tfrac{1}{\sqrt{n_t}} + \tfrac{1}{\sqrt{p}} + \tfrac{1}{\sqrt{q}}\Big) = o_p(1),$$

as $n_t, p, q \to \infty$, with no relative growth condition between p and q.

L_7.
Unlike the bias terms above, L_7 does not vanish, so we identify its first-order form rather than bounding it away. Recall $L_7 = \sqrt{n_c}\,\delta_0^\top(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1)$, which lives on the control $\sqrt{n_c}$-scale, whereas Lemma C.6 expands the rotation error on the treated $\sqrt{n_t}$-scale,

$$\sqrt{n_t}\,\big(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1\big) = J_1 - J_2 + J_3 + J_4, \quad -J_1, J_2, J_4 = o_p(1), \quad J_3 = O_p(1),$$

where the surviving signal term is

$$\begin{aligned}J_3 &= \mathbf{P}_{\bar{q}}\, \mathbf{H}_1^\top M^{-1} \mathbf{H}_1\, \tfrac{1}{\sqrt{n_t}} \mathbf{F}_1^\top \epsilon_1 \\ &= \mathbf{P}_{\bar{q}}\, \Omega_1^{-1}\, \tfrac{1}{\sqrt{n_t}} \sum_{i:A_i=1} \mathbf{f}_{1,i}\epsilon_{1,i} + o_p(1),\end{aligned}$$

the second equality using $\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 - \Omega_1^{-1} = o_p(1)$ established in Lemma C.6.

Because the two scales differ, we do not equate L_7 with $\delta_0^\top J_3$ directly. Substituting the treated-scale expansion and keeping the scale factor,

$$L_7 = \sqrt{n_c}\,\delta_0^\top(\mathbf{H}_M^\top \hat{\gamma}_1 - \gamma_1) = \sqrt{\tfrac{n_c}{n_t}}\,\delta_0^\top\big(-J_1 - J_2 + J_3 + J_4\big).$$

For $k \in \{1,2,4\}$ the Cauchy-Schwarz inequality gives $|\delta_0^\top J_k| \le \|\delta_0\| \, \|J_k\| = o_p(1)$, since $\|\delta_0\| = O(1)$ by Assumption 3.1(iii) and $J_k = o_p(1)$, while $\sqrt{n_c/n_t} = O_p(1)$ by Assumption 5.2, so those three products are $o_p(1)$. Only the signal survives, and it carries the ratio explicitly,

$$L_7 = \sqrt{\tfrac{n_c}{n_t}}\, \delta_0^\top J_3 + o_p(1),$$
$$\delta_0^\top J_3 = \left(\mathbf{P}_{\bar{q}}^\top \delta_0\right)^\top \Omega_1^{-1} \tfrac{1}{\sqrt{n_t}} \sum_{i:A_i=1} \mathbf{f}_{1,i}\epsilon_{1,i} + o_p(1).$$

The factor $\sqrt{n_c/n_t}$ is $O_p(1)$.
For regression fold k, let n_k denote the fold size, with treated and control counts n_k^t and n_k^c. Since

$$\frac{n_k^t}{n_k} \xrightarrow{p} \pi_1, \qquad \frac{n_k^c}{n_k} \xrightarrow{p} \pi_0,$$

we place the treated and control expansions on the common $\sqrt{n_k}$ scale. Now, collecting the bounds of this part,

$$G_1 = G_2 = G_4 = o_p(1), \quad L_1 = L_2 = L_4 = L_5 = L_6 = o_p(1), \quad L_7 = \sqrt{\tfrac{n_c}{n_t}}\, \delta_0^\top J_3 + o_p(1),$$

Suppressing the fold. within the terms G_j, L_j, we have

$$\sqrt{n_k}(\hat{\alpha}_1 - \alpha_1) = \sqrt{\frac{n_k}{n_k^t}}\, G_3 + o_p(1),$$
$$\sqrt{n_k}\left(\widehat{\gamma_1^\top \delta_0} - \gamma_1^\top \delta_0\right) = \sqrt{\frac{n_k}{n_k^c}}\, L_3 + \sqrt{\frac{n_k}{n_k^t}}\, \delta_0^\top J_3 + o_p(1). \tag{C.43}$$

Therefore,

$$\sqrt{n_k}(\hat{\theta}_{0,k} - \theta_0) = \sqrt{\frac{n_k}{n_k^t}}\left(G_3 + \delta_0^\top J_3\right) + \sqrt{\frac{n_k}{n_k^c}}\, L_3 + o_p(1).$$

The remaining terms are therefore G_3, L_3, and $\delta_0^\top J_3$. We next derive their joint asymptotic distribution.

Part 3: Asymptotic Normality of $\hat{\theta}_0$.
Each surviving term carries a random sandwich matrix, $\mathbf{H}_1^\top M^{-1}\mathbf{H}_1$ for G_3 and $\delta_0^\top J_3$, and $\mathbf{H}_{X1}^\top \widetilde{M}^{-1}\mathbf{H}_{X1}$ for L_3. The argument proceeds in six steps. Step 1 shows these sandwich matrices converge in probability to the fixed conditional inverses Ω_1^{-1} and Ω_{X0}^{-1}. Steps 2–3 reduce $G_3 + \delta_0^\top J_3$ and L_3 to stratum averages of i.i.d. mean-zero summands on the fold scale. Step 4 assembles the single influence function with the correct π_1, π_0 scaling and shows the cross term vanishes exactly. Step 5 verifies the Lyapunov condition and concludes for one fold, and Step 6 lifts the result to the cross-fitted estimator. We close by bounding σ_n^2 above and below.

Throughout this part we work on one regression fold $\mathcal{D}$ of size $m = |\mathcal{D}|$, with treated and control counts $n_t = |\{i \in \mathcal{D} : A_i = 1\}|$ and $n_c = |\{i \in \mathcal{D} : A_i = 0\}|$, and $n_t/m \to_p \pi_1$, $n_c/m \to_p \pi_0$ by Assumption 5.2. The full-sample lift is carried out in Step 6.

Step 1: Sandwich matrices converge to conditional inverses.
By Lemma C.5, we have that $\mathbf{H}_1 = UR$ with $U^\top U = I$ and R square and invertible ($\nu_{\min}(R) \ge c_R > 0$), and

$$\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 = R^\top (D + E_n)^{-1} R, \quad D = \tfrac{1}{n_t} R \mathbf{F}_1^\top \mathbf{F}_1 R^\top.$$

Define the fixed matrix $D_\infty = R\,\Omega_1\,R^\top$. Since R is square and invertible and $\Omega_1 \succ 0$, D_∞ is invertible with $\lambda_{\min}(D_\infty) \ge \nu_{\min}^2(R)\,\lambda_{\min}(\Omega_1) \ge c\,c_R^2 > 0$, and

$$\begin{aligned} \mathbf{S}_1 :&= R^\top D_\infty^{-1} R \\ &= R^\top \left(R\Omega_1 R^\top\right)^{-1} R \\ &= R^\top R^{-\top} \Omega_1^{-1} R^{-1} R = \Omega_1^{-1}. \end{aligned} \tag{C.44}$$

This is an exact identity, using $(R\Omega_1 R^\top)^{-1} = R^{-\top}\Omega_1^{-1}R^{-1}$ for square invertible R. It is the limit toward which the random sandwich converges.

(a) $D \xrightarrow{p} D_\infty$. The treated outer products $\mathbf{f}_{1,i}\mathbf{f}_{1,i}^\top$ are i.i.d. across treated i, with conditional mean Ω_1 and finite fourth moments (Assumption 5.1(i)). Now, applying the multivariate central limit theorem entrywise to the mean-zero summands $\mathbf{f}_{1,i}\mathbf{f}_{1,i}^\top - \Omega_1$ gives, for each fixed pair (j,k),

$$\sqrt{n_t}\Big\{\Big(\tfrac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1\Big)_{jk} - (\Omega_1)_{jk}\Big\} = O_p(1).$$

As the dimension $1+\bar{p}+\bar{q}$ is fixed, entrywise convergence is operator-norm convergence at the same rate, so $\big\|\frac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1 - \Omega_1\big\|_{\mathrm{op}} = O_p(n_t^{-1/2})$, and hence

$$\begin{aligned}\|D - D_\infty\|_{\mathrm{op}} &= \Big\|R\Big(\tfrac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1 - \Omega_1\Big)R^\top\Big\|_{\mathrm{op}}\\ &\le \|R\|_{\mathrm{op}}^2\, O_p\left(\frac{1}{\sqrt{n_t}}\right)\\ &= O_p\left(\frac{1}{\sqrt{n_t}}\right).\end{aligned}$$

(b) $(D+E_n)^{-1} \to_p D_\infty^{-1}$. We begin with the identity $A^{-1}-B^{-1} = A^{-1}(B-A)B^{-1}$, with $A = D+E_n$ and $B = D_\infty$, giving

$$(D+E_n)^{-1} - D_\infty^{-1} = (D+E_n)^{-1}\big[(D_\infty - D) - E_n\big]D_\infty^{-1}.$$

Taking operator norms and using sub-multiplicativity together with $\|(D+E_n)^{-1}\|_{\mathrm{op}} = O_p(1)$ (Lemma C.5), $\|D_\infty^{-1}\|_{\mathrm{op}} \le (c\,c_R^2)^{-1} = O(1)$, $\|D-D_\infty\|_{\mathrm{op}} = O_p(n_t^{-1/2})$ from part (a), and $\|E_n\|_{\mathrm{op}} = o_p(1)$ at the rate (C.11),

$$\begin{aligned}\|(D+E_n)^{-1} - D_\infty^{-1}\|_{\mathrm{op}} &\le \|(D+E_n)^{-1}\|_{\mathrm{op}}\big(\|D-D_\infty\|_{\mathrm{op}} + \|E_n\|_{\mathrm{op}}\big)\|D_\infty^{-1}\|_{\mathrm{op}}\\ &= o_p(1).\end{aligned} \tag{C.45}$$

Combining with (C.44),

$$\begin{aligned}\big\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 - \Omega_1^{-1}\big\|_{\mathrm{op}} &= \big\|R^\top\big[(D+E_n)^{-1} - D_\infty^{-1}\big]R\big\|_{\mathrm{op}}\\ &\le \|R\|_{\mathrm{op}}^2\, o_p(1)\\ &= o_p(1).\end{aligned}$$

The identical argument on the control side, with $\mathbf{H}_{X1} = \widetilde{U}\widetilde{R}$, $\widetilde{D}_\infty = \widetilde{R}\,\Omega_{X0}\,\widetilde{R}^\top$, and $\mathbf{S}_2 = \widetilde{R}^\top\widetilde{D}_\infty^{-1}\widetilde{R} = \Omega_{X0}^{-1}$, gives

$$\big\|\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 - \Omega_1^{-1}\big\|_{\mathrm{op}} = o_p(1), \quad \big\|\mathbf{H}_{X1}^\top\widetilde{M}^{-1}\mathbf{H}_{X1} - \Omega_{X0}^{-1}\big\|_{\mathrm{op}} = o_p(1). \tag{C.46}$$

Step 2: $G_3 + \delta_0^\top J_3$ as a treated i.i.d. average.
Both terms share the same sandwich, since $\delta_0^\top J_3 = (\mathbf{P}_{\bar{q}}^\top\delta_0)^\top\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1$ has the same shape as $G_3 = e_{\mathbf{1}}^\top\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1$. The fixed vectors therefore add, giving $a_1 = e_{\mathbf{1}} + \mathbf{P}_{\bar{q}}^\top\delta_0$. Because $\mathbf{P}_{\bar{q}}^\top\delta_0$ lies in the mediator block, its intercept slot is zero, so

$$\|a_1\|^2 = 1 + \|\mathbf{P}_{\bar{q}}^\top\delta_0\|^2, \quad 1 \le \|a_1\| = O(1) \quad (\text{as } \|\delta_0\| = O(1)). \tag{C.47}$$

Now, combining them into a single object,

$$\begin{aligned}G_3 + \delta_0^\top J_3 &= a_1^\top\mathbf{H}_1^\top M^{-1}\mathbf{H}_1\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1\\ &= a_1^\top\Omega_1^{-1}\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1 + a_1^\top\big(\mathbf{H}_1^\top M^{-1}\mathbf{H}_1 - \Omega_1^{-1}\big)\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1.\end{aligned}$$

The second term is $o_p(1)$, since its matrix factor is $o_p(1)$ in operator norm by (C.46) and its vector factor is $O_p(1)$. By Assumption 3.1(ii), ϵ_1 is mean zero and independent of $\mathbf{F}_1$, while Assumption 5.1(i) gives uniformly bounded second moments. Therefore,

$$\begin{aligned}\mathbb{E}\Big[\Big\|\tfrac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1\Big\|^2\,\Big|\,\mathbf{F}_1\Big] &= \operatorname{tr}\Big(\tfrac{1}{n_t}\mathbf{F}_1^\top\,\mathbb{E}(\epsilon_1\epsilon_1^\top\mid\mathbf{F}_1)\,\mathbf{F}_1\Big)\\ &\le C\operatorname{tr}\Big(\tfrac{1}{n_t}\mathbf{F}_1^\top\mathbf{F}_1\Big)\\ &= O_p(1),\end{aligned} \tag{C.48}$$

so $\big\|\frac{1}{\sqrt{n_t}}\mathbf{F}_1^\top\epsilon_1\big\| = O_p(1)$ by the Chebyshev inequality. Therefore, we have that

$$G_3+\delta_0^\top J_3 = \frac{1}{\sqrt{n_t}}\sum_{i:A_i=1}\xi_{1,i}+o_p(1),\quad \xi_{1,i}=\big(\Omega_1^{-1}a_1\big)^\top\mathbf{f}_{1,i}\,\epsilon_{1,i}. \tag{C.49}$$

The conditional exogeneity $\mathbb{E}[\epsilon_{1,i}\mid\mathbf{f}_{1,i},A_i=1]=0$ follows from Assumption 3.1(ii) makes each $\xi_{1,i}$ mean zero, and the $\xi_{1,i}$ are i.i.d. across treated i because $\Omega_1^{-1}a_1$ is fixed and $(\mathbf{f}_{1,i},\epsilon_{1,i})$ are i.i.d.. Their variance is

$$\begin{aligned}\sigma_1^2 :&= \operatorname{Var}(\xi_{1,i}\mid A_i=1)\\ &= \mathbb{E}\big[\epsilon_{1,i}^2\,\{(\Omega_1^{-1}a_1)^\top\mathbf{f}_{1,i}\}^2\,\big|\,A_i=1\big],\end{aligned} \tag{C.50}$$

finite because $\|\Omega_1^{-1}\|_{\mathrm{op}}=O(1)$, $\|a_1\|=O(1)$, and the conditional fourth moment of $\mathbf{f}_{1,i}$ and second moment of $\epsilon_{1,i}$ are bounded (Assumption 5.1). When $\epsilon_{1,i}\perp\mathbf{f}_{1,i}$, it simplifies to $\sigma_1^2=\sigma_{\epsilon_1,1}^2\,a_1^\top\Omega_1^{-1}a_1$ with $\sigma_{\epsilon_1,1}^2=\mathbb{E}[\epsilon_{1,i}^2]$.

Step 3: L_3 as a control i.i.d. average.
Repeating Step 2 with $(\widetilde{M},\mathbf{H}_{X1},\mathbf{F}_{X1},\boldsymbol{\eta},\Omega_{X0})$ in place of $(M,\mathbf{H}_1,\mathbf{F}_1,\epsilon_1,\Omega_1)$, then using the second half of (C.46) and the control analogue of (C.48) with $\boldsymbol{\eta}=\gamma_1^\top\mathbf{u}_{XM,0}$ in place of ϵ_1,

$$L_3 = \frac{1}{\sqrt{n_c}}\sum_{i:A_i=0}\xi_{2,i}+o_p(1), \tag{C.51}$$

where $\xi_{2,i}=\big(\Omega_{X0}^{-1}e_{\mathbf{1}}\big)^\top\mathbf{f}_{X1,i}\,\eta_i$, and $\eta_i=\gamma_1^\top\mathbf{u}_{XM,0,i}$. By Assumption 3.1(ii), $\mathbf{u}_{XM,0,i}$ is mean zero and independent of $(\mathbf{f}_{X,i},A_i)$, so $\mathbb{E}[\eta_i\mid\mathbf{f}_{X1,i},A_i=0]=0$, so the $\xi_{2,i}$ are i.i.d. mean-zero across control i, with conditional variance

$$\begin{aligned}\sigma_2^2 :&= \operatorname{Var}(\xi_{2,i}\mid A_i=0)\\ &= \mathbb{E}\big[\eta_i^2\,\{(\Omega_{X0}^{-1}e_{\mathbf{1}})^\top\mathbf{f}_{X1,i}\}^2\,\big|\,A_i=0\big]\\ &< \infty,\end{aligned} \tag{C.52}$$

finite since

$$\begin{aligned}\mathbb{E}[\eta_i^2\mid A_i=0] &= \gamma_1^\top\Sigma_{u_{XM,0}}\gamma_1\\ &\le \|\gamma_1\|_2^2\|\Sigma_{u_{XM,0}}\|_{\mathrm{op}}\\ &\le C\end{aligned}$$

using Assumption 3.2(i) for $\|\gamma_1\|=O(1)$ and Assumption 3.1(v) for $\|\Sigma_{\mathbf{u}_{XM,0}}\|_{\mathrm{op}}\le C$.

Step 4: The single influence function and the vanishing cross term.
Substituting (C.43), (C.49), and (C.51) into the fold decomposition (C.14), on the common fold scale $\sqrt{m}$,

$$\sqrt{m}\,(\hat\theta_{\mathcal{D}}-\theta_0)=\sqrt{\tfrac{m}{n_t}}\Big[\tfrac{1}{\sqrt{n_t}}\sum_{i:A_i=1}\xi_{1,i}\Big]+\sqrt{\tfrac{m}{n_c}}\Big[\tfrac{1}{\sqrt{n_c}}\sum_{i:A_i=0}\xi_{2,i}\Big]+o_p(1).$$

On the treated group, regrouping the constants gives us

$$\begin{aligned}\sqrt{\tfrac{m}{n_t}}\cdot\tfrac{1}{\sqrt{n_t}}\sum_{i:A_i=1}\xi_{1,i} &= \tfrac{1}{\sqrt{m}}\cdot\tfrac{m}{n_t}\sum_{i:A_i=1}\xi_{1,i}\\ &= \tfrac{1}{\sqrt{m}}\sum_{i\in\mathcal{D}}\mathbf{1}(A_i=1)\,\tfrac{m}{n_t}\,\xi_{1,i}.\end{aligned}$$

Since $\frac{m}{n_t} \to_p \frac{1}{\pi_1}$ by Assumption 5.2 and $\frac{1}{\sqrt{m}} \sum_{i \in \mathcal{D}} \mathbf{1}(A_i = 1)\xi_{1,i} = O_p(1)$, the replacement of $\frac{m}{n_t}$ by $\frac{1}{\pi_1}$ costs $\left(\frac{m}{n_t} - \frac{1}{\pi_1}\right) \cdot O_p(1) = o_p(1)$ by Slutsky's theorem. The control group is analogous with $\frac{m}{n_c} \to_p \frac{1}{\pi_0}$. Hence

$$\sqrt{m}\,(\hat{\theta}_{\mathcal{D}} - \theta_0) = \frac{1}{\sqrt{m}} \sum_{i \in \mathcal{D}} \psi_i + o_p(1), \tag{C.53}$$

where

$$\psi_i = \frac{\mathbf{1}(A_i = 1)}{\pi_1} \big(\Omega_1^{-1} a_1\big)^\top \mathbf{f}_{1,i} \epsilon_{1,i} + \frac{\mathbf{1}(A_i = 0)}{\pi_0} \big(\Omega_{X0}^{-1} e_{\mathbf{1}}\big)^\top \mathbf{f}_{X1,i} \eta_i. \tag{C.54}$$

By the conditional mean-zero properties of Steps 2—3, $\mathbb{E}[\psi_i] = 0$.

The cross term vanishes exactly. Write $\psi_i = \psi_i^{(1)} + \psi_i^{(0)}$ for the treated and control parts. Each product $\psi_i^{(1)} \psi_i^{(0)}$ contains the factor $\mathbf{1}(A_i = 1)\,\mathbf{1}(A_i = 0) = 0$, since A_i takes exactly one value, so

$$\psi_i^{(1)} \psi_i^{(0)} = 0 \quad \text{for every } i,$$

pointwise, not merely in expectation. Therefore

$$\begin{aligned} \sigma_n^2 :&= \operatorname{Var}(\psi_i) \\ &= \mathbb{E}\big[(\psi_i^{(1)})^2\big] + 2\,\mathbb{E}\big[\psi_i^{(1)} \psi_i^{(0)}\big] + \mathbb{E}\big[(\psi_i^{(0)})^2\big] \\ &= \mathbb{E}\big[(\psi_i^{(1)})^2\big] + \mathbb{E}\big[(\psi_i^{(0)})^2\big]. \end{aligned}$$

Evaluating each square with $\mathbb{E}[\mathbf{1}(A_i = 1)] = \pi_1$ and the conditional variances (C.50) and (C.52),

$$\begin{aligned} \mathbb{E}\big[(\psi_i^{(1)})^2\big] &= \frac{1}{\pi_1^2}\,\mathbb{E}\big[\mathbf{1}(A_i = 1)\,\{(\Omega_1^{-1} a_1)^\top \mathbf{f}_{1,i}\,\epsilon_{1,i}\}^2\big] \\ &= \frac{1}{\pi_1^2}\,\mathbb{E}\big[\mathbf{1}(A_i = 1)\big]\,\mathbb{E}\big[\{(\Omega_1^{-1} a_1)^\top \mathbf{f}_{1,i}\,\epsilon_{1,i}\}^2 \,\big|\, A_i = 1\big] \\ &= \frac{\sigma_1^2}{\pi_1}, \\ \mathbb{E}\big[(\psi_i^{(0)})^2\big] &= \frac{\sigma_2^2}{\pi_0}, \end{aligned}$$

so that

$$\sigma_n^2 = \frac{\sigma_1^2}{\pi_1} + \frac{\sigma_2^2}{\pi_0}. \tag{C.55}$$

There is no 2ρ cross term, as the mutually exclusive treatment indicators force the same-observation product of the two groups to be exactly zero.

Step 5: Lyapunov central limit theorem.
The $\{\psi_i\}_{i \in \mathcal{D}}$ are i.i.d. and mean zero. Fix $\omega \in (0, 1]$. Each part of ψ_i is a fixed scalar ($\frac{1}{\pi_1}$ or $\frac{1}{\pi_0}$) times the product of a bounded linear form in a sub-Gaussian factor vector and a sub-Gaussian noise ($\epsilon_{1,i}$ or $\eta_i = \gamma_1^\top \mathbf{u}_{XM,0,i}$). By the elementary bound $|x + y|^{2+\omega} \le 2^{1+\omega}(|x|^{2+\omega} + |y|^{2+\omega})$, the Cauchy–Schwarz inequality, and the finiteness of all moments of sub-Gaussian variables (Assumption 5.1), we get that

$$\begin{aligned} \mathbb{E}|\psi_i|^{2+\omega} &\le 2^{1+\omega}\Big(\tfrac{1}{\pi_1^{2+\omega}}\mathbb{E}\big|(\Omega_1^{-1} a_1)^\top \mathbf{f}_{1,i}\,\epsilon_{1,i}\big|^{2+\omega} + \tfrac{1}{\pi_0^{2+\omega}}\mathbb{E}\big|(\Omega_{X0}^{-1} e_{\mathbf{1}})^\top \mathbf{f}_{X1,i}\,\eta_i\big|^{2+\omega}\Big) \\ &\le C < \infty, \end{aligned}$$

uniformly in i. With the lower bound $\sigma_n^2 \ge c > 0$ established below, the i.i.d. Lyapunov ratio satisfies

$$\begin{aligned} \frac{\sum_{i \in \mathcal{D}} \mathbb{E}|\psi_i|^{2+\omega}}{\big(\sum_{i \in \mathcal{D}} \operatorname{Var} \psi_i\big)^{1+\omega/2}} &= \frac{m\,\mathbb{E}|\psi_1|^{2+\omega}}{(m\sigma_n^2)^{1+\omega/2}} \\ &\le \frac{C\,m}{(mc)^{1+\omega/2}} \\ &= O\big(m^{-\omega/2}\big) \to 0, \end{aligned}$$

verifying the Lyapunov condition. By the Lyapunov central limit theorem for i.i.d. arrays,

$$\frac{1}{\sqrt{m}\,\sigma_n}\sum_{i\in\mathcal{D}}\psi_i \xrightarrow{d} N(0,1),$$

and combining with (C.53) and Slutsky's theorem gives $\sqrt{m}\,(\hat{\theta}_{\mathcal{D}}-\theta_0)/\sigma_n \xrightarrow{d} N(0,1)$.

Step 6: Cross-fitting.
The expansion above is for one regression fold. Let $\mathcal{D}_1, \mathcal{D}_2$ be the two folds with $|\mathcal{D}_1| = |\mathcal{D}_2| = n/2$, the whitening for fold $\mathcal{D}_k$ built on its complement, and we get that $\hat{\theta}_0^{\mathrm{cf}} = \frac{1}{2}(\hat{\theta}_{\mathcal{D}_1} + \hat{\theta}_{\mathcal{D}_2})$. Steps 1–5 applied within fold $\mathcal{D}_k$ give

$$\hat{\theta}_{\mathcal{D}_k} - \theta_0 = \frac{1}{|\mathcal{D}_k|}\sum_{i\in\mathcal{D}_k}\psi_i + o_p\big(|\mathcal{D}_k|^{-1/2}\big), \tag{C.56}$$

with the same influence function ψ_i of (C.53). The training-fold whitening differs across folds, but by Step 1 both training folds produce sandwich matrices converging to the same population limits Ω_1, Ω_{X0}, so the leading summand in (C.56) is the same fixed influence function on both folds and the difference is absorbed into the same o_p remainder.

The two fold sums are over disjoint index sets $\mathcal{D}_1, \mathcal{D}_2$, so $\{\psi_i\}_{i\in\mathcal{D}_1}$ and $\{\psi_i\}_{i\in\mathcal{D}_2}$ are independent collections of the same i.i.d., the cross-fold estimation dependence enters only through the training-fold whitening, whose effect is o_p. Hence, the two centered fold sums are asymptotically uncorrelated and, using $2|\mathcal{D}_k| = n$, we have that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_0^{\mathrm{cf}} - \theta_0) &= \frac{\sqrt{n}}{2}\sum_{k=1}^{2}(\hat{\theta}_{\mathcal{D}_k} - \theta_0) \\
&= \frac{\sqrt{n}}{2}\sum_{k=1}^{2}\frac{1}{|\mathcal{D}_k|}\sum_{i\in\mathcal{D}_k}\psi_i + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i + o_p(1).
\end{aligned}$$

The $\{\psi_i\}_{i=1}^n$ are i.i.d. mean zero with variance σ_n^2 over the full sample, so the fold-level Lyapunov argument of Step 5 applies verbatim with m replaced by n yields

$$\frac{\sqrt{n}(\hat{\theta}_0^{\mathrm{cf}} - \theta_0)}{\sigma_n} \xrightarrow{d} N(0,1).$$

Boundedness and nondegeneracy of σ_n^2. It remains to show $0 < c \le \sigma_n^2 \le C < \infty$. The upper bound is immediate from (C.55), since $\sigma_1^2, \sigma_2^2 = O(1)$ (Steps 2-3) and π_1, π_0 are bounded away from 0. For the lower bound, σ_n^2 is a sum of two nonnegative terms, so it suffices to bound one of them below. Conditioning on $\mathbf{f}_{1,i}$ inside (C.50) and using the conditional residual variance bound $\mathrm{Var}(\epsilon_{1,i} \mid \mathbf{f}_{1,i}, A_i = 1) \ge \underline{\sigma}^2 > 0$ (Assumption 3.1(i) and 5.1(ii)),

$$\begin{aligned}
\sigma_1^2 &= \mathbb{E}\big[\,\mathrm{Var}(\epsilon_{1,i} \mid \mathbf{f}_{1,i}, A_i = 1)\,\{(\Omega_1^{-1}a_1)^\top \mathbf{f}_{1,i}\}^2 \,\big|\, A_i = 1\big] \\
&\ge \underline{\sigma}^2\,\mathbb{E}\big[\{(\Omega_1^{-1}a_1)^\top \mathbf{f}_{1,i}\}^2 \,\big|\, A_i = 1\big].
\end{aligned}$$

The inner expectation is the Ω_1-quadratic form,

$$\begin{aligned}
\mathbb{E}\big[\{(\Omega_1^{-1}a_1)^\top \mathbf{f}_{1,i}\}^2 \,\big|\, A_i = 1\big] &= (\Omega_1^{-1}a_1)^\top \Omega_1 (\Omega_1^{-1}a_1) \\
&= a_1^\top \Omega_1^{-1} a_1 \\
&\ge \lambda_{\min}(\Omega_1^{-1})\,\|a_1\|^2.
\end{aligned}$$

Since $\|a_1\| \geq 1$ by (C.47) and $\lambda_{\min}(\Omega_1^{-1}) = \lambda_{\max}(\Omega_1)^{-1} > 0$,

$$\begin{aligned}\sigma_n^2 &\geq \frac{\sigma_1^2}{\pi_1} \\ &\geq \frac{\underline{\sigma}^2\, \lambda_{\max}(\Omega_1)^{-1}}{\pi_1} \\ &= c > 0.\end{aligned}$$

Now, combining the two bounds gives $0 < c \leq \sigma_n^2 \leq C < \infty$, which completes the proof of asymptotic normality of the cross-fitted estimator $\hat{\theta}_0^{\text{cf}}$. □

D Auxiliary Material

D.1 Technical Lemmas

Lemma D.1 (Inverse of an augmented second-moment matrix)**.** *Suppose Assumptions 3.1 and 3.2 hold. Let* $\mathbf{H} = diag(\mathbf{H}_X, \mathbf{H}_M)$ *where* $\mathbf{H}_X$ *and* $\mathbf{H}_M$ *be defined as in Definition 3.1, and* $\mathbf{e} = (\mathbf{e}_X^\top \quad \mathbf{e}_M^\top)^\top$ *be a vector with mean zero, covariance* $\Sigma_{\mathbf{e}} \succ 0$ *as defined in Section A.1. Set* $L = (1, (\mathbf{Hf} + \mathbf{e})^\top)^\top$*, then* $M := \mathbb{E}[LL^\top \mid A = 1]$ *is invertible, with*

$$M^{-1} = \begin{pmatrix} 1 + \mu_{\mathbf{f}|1}^\top \mathbf{H}^\top (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}\mathbf{H}\mu_{\mathbf{f}|1} & -\mu_{\mathbf{f}|1}^\top \mathbf{H}^\top (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1} \\ -(\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}\mathbf{H}\mu_{\mathbf{f}|1} & (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1} \end{pmatrix}.$$

Proof. Using Assumption 3.1(ii), we know that $\mathbf{e}$ do not depend on A, so $\mathbb{E}[\mathbf{e} \mid A = 1] = \mathbb{E}[\mathbf{e}] = 0$, $\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid A = 1] = \mathbb{E}[\mathbf{e}\mathbf{e}^\top] = \Sigma_{\mathbf{e}}$ and $\mathbb{E}[\mathbf{f}\mathbf{e}^\top \mid A = 1] = 0$. From $L = (1, (\mathbf{Hf} + \mathbf{e})^\top)^\top$, we have

$$LL^\top = \begin{pmatrix} 1 & (\mathbf{Hf} + \mathbf{e})^\top \\ \mathbf{Hf} + \mathbf{e} & (\mathbf{Hf} + \mathbf{e})(\mathbf{Hf} + \mathbf{e})^\top \end{pmatrix}.$$

Now, expanding the bottom-right block,

$$(\mathbf{Hf} + \mathbf{e})(\mathbf{Hf} + \mathbf{e})^\top = \mathbf{Hf}\ \mathbf{f}^\top\mathbf{H}^\top + \mathbf{Hf}\mathbf{e}^\top + \mathbf{e}\mathbf{f}^\top\mathbf{H}^\top + \mathbf{e}\mathbf{e}^\top.$$

We take the conditional expectation given $A = 1$ entrywise. In the linear blocks, we have that

$$\mathbb{E}[\mathbf{Hf} + \mathbf{e} \mid A = 1] = \mathbf{H}\,\mathbb{E}[\mathbf{f} \mid A = 1] + \mathbb{E}[\mathbf{e}] = \mathbf{H}\mu_{\mathbf{f}|1},$$

using $\mathbb{E}[\mathbf{e}] = 0$. Also, the two cross terms satisfy

$$\begin{aligned}\mathbb{E}[\mathbf{Hf}\mathbf{e}^\top \mid A = 1] &= \mathbf{H}\,\mathbb{E}[\mathbf{f}\mathbf{e}^\top \mid A = 1] = 0, \\ \mathbb{E}[\mathbf{e}\mathbf{f}^\top\mathbf{H}^\top \mid A = 1] &= \mathbb{E}[\mathbf{f}\mathbf{e}^\top \mid A = 1]^\top\mathbf{H}^\top = 0,\end{aligned}$$

as $\mathbb{E}[\mathbf{f}\mathbf{e}^\top \mid A = 1] = 0$, and the remaining noise term is $\mathbb{E}[\mathbf{e}\mathbf{e}^\top \mid A = 1] = \mathbb{E}[\mathbf{e}\mathbf{e}^\top] = \Sigma_{\mathbf{e}}$. So, collecting these gives us

$$M = \begin{pmatrix} 1 & \mu_{\mathbf{f}|1}^\top\mathbf{H}^\top \\ \mathbf{H}\mu_{\mathbf{f}|1} & \mathbf{H}\,\mathbb{E}[\mathbf{f}\mathbf{f}^\top \mid A = 1]\,\mathbf{H}^\top + \Sigma_{\mathbf{e}} \end{pmatrix}.$$

We write $A_{11} = 1$, $A_{12} = \mu_{\mathbf{f}|1}^\top\mathbf{H}^\top$, and $A_{22} = \mathbf{H}\,\mathbb{E}[\mathbf{f}\ \mathbf{f}^\top \mid A = 1]\,\mathbf{H}^\top + \Sigma_{\mathbf{e}}$. The Schur complement of A_{11} is

$$\begin{aligned}&A_{22} - A_{12}^\top A_{11}^{-1} A_{12} = \mathbf{H}\,\mathbb{E}[\mathbf{f}\ \mathbf{f}^\top \mid A = 1]\,\mathbf{H}^\top + \Sigma_{\mathbf{e}} - \mathbf{H}\mu_{\mathbf{f}|1}\mu_{\mathbf{f}|1}^\top\mathbf{H}^\top \\ &= \mathbf{H}\big(\mathbb{E}[\mathbf{f}\ \mathbf{f}^\top \mid A = 1] - \mu_{\mathbf{f}|1}\mu_{\mathbf{f}|1}^\top\big)\mathbf{H}^\top + \Sigma_{\mathbf{e}} = \mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}},\end{aligned}$$

using $\Sigma_{\mathbf{f}|1} = \mathbb{E}[\mathbf{f}\ \mathbf{f}^\top \mid A=1] - \mu_{\mathbf{f}|1}\mu_{\mathbf{f}|1}^\top$ in the last line. This is positive definite, so M is invertible. Applying block-inverse formula with these blocks, we get

$$\begin{aligned}
B_{22} &= (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}, \\
B_{12} &= -\mu_{\mathbf{f}|1}^\top \mathbf{H}^\top (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}, \\
B_{11} &= 1 + \mu_{\mathbf{f}|1}^\top \mathbf{H}^\top (\mathbf{H}\Sigma_{\mathbf{f}|1}\mathbf{H}^\top + \Sigma_{\mathbf{e}})^{-1}\mathbf{H}\mu_{\mathbf{f}|1},
\end{aligned}$$

which is the formula. □

Finally, we record the Woodbury identity, used repeatedly to invert signal plus noise matrices.

Lemma D.2 (Woodbury matrix identity)**.** *Let $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{k\times k}$ be invertible, and let $U \in \mathbb{R}^{n\times k}$, $V \in \mathbb{R}^{k\times n}$. If $C^{-1} + VA^{-1}U$ is invertible, then $A + UCV$ is invertible and*

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\big(C^{-1} + VA^{-1}U\big)^{-1}VA^{-1}.$$

Proof. We verify the right-hand side is a right inverse of $A + UCV$. Since both matrices are square, a right inverse is a two-sided inverse. We have

$$\begin{aligned}
&(A+UCV)\big[A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}\big] \\
&= I + UCVA^{-1} - U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} - UCVA^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
&= I + UCVA^{-1} - U\big[I + CVA^{-1}U\big](C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
&\overset{(1)}{=} I + UCVA^{-1} - UC\big[C^{-1}+VA^{-1}U\big](C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
&= I + UCVA^{-1} - UCVA^{-1} = I,
\end{aligned}$$

Here, (1) follows simplyy using $C[C^{-1}+VA^{-1}U](C^{-1}+VA^{-1}U)^{-1} = C$. □

D.2 Calculation of Effects

Throughout this subsection, we center the latent factors so that $\mathbb{E}[\mathbf{f}_X] = \mathbb{E}[\mathbf{f}_M] = 0$. The intercepts α_a and δ_a therefore account for the corresponding mean levels. Note that the arm-specific conditional means $\mu_{\mathbf{f}|1}$ and $\mu_{\mathbf{f}_X|0}$ used elsewhere in the Appendix can still be nonzero.

By the regression model in (2.1), for $a \in \{0,1\}$,

$$\mathbb{E}[Y \mid \mathbf{f}_X, \mathbf{f}_M, A=a] = \alpha_a + \beta_a^\top \mathbf{f}_X + \gamma_a^\top \mathbf{f}_M, \qquad \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=a] = \delta_a + B_a \mathbf{f}_X.$$

For $Y^{(a,M^{(a')})}$, we first average over the mediator distribution under $A=a'$ conditional on $\mathbf{f}_X$, and then over the marginal distribution of $\mathbf{f}_X$. Therefore,

$$\begin{aligned}
\mathbb{E}\big[Y^{(a,M^{(a')})}\big] &= \alpha_a + \mathbb{E}[\mathbf{f}_X]^\top \beta_a + \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=a']^\top \gamma_a \\
&= (\alpha_a + \delta_{a'}^\top \gamma_a) + \mathbb{E}[\mathbf{f}_X]^\top (\beta_a + B_{a'}^\top \gamma_a) \\
&= \alpha_a + \delta_{a'}^\top \gamma_a,
\end{aligned}$$

where the last equality follows from $\mathbb{E}[\mathbf{f}_X] = 0$.

For the four choices of (a, a'), we have

$$\begin{aligned}
\mathbb{E}[Y^{(1,M^{(1)})}] &= \alpha_1 + \mathbb{E}[\mathbf{f}_X]^\top \beta_1 + \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=1]^\top \gamma_1 \\
&= (\alpha_1 + \delta_1^\top \gamma_1) + \mathbb{E}[\mathbf{f}_X]^\top (\beta_1 + B_1^\top \gamma_1) \\
&= \alpha_1 + \delta_1^\top \gamma_1 \,, \\
\mathbb{E}[Y^{(0,M^{(0)})}] &= \alpha_0 + \mathbb{E}[\mathbf{f}_X]^\top \beta_0 + \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=0]^\top \gamma_0 \\
&= (\alpha_0 + \delta_0^\top \gamma_0) + \mathbb{E}[\mathbf{f}_X]^\top (\beta_0 + B_0^\top \gamma_0) \\
&= \alpha_0 + \delta_0^\top \gamma_0 \,, \\
\mathbb{E}[Y^{(1,M^{(0)})}] &= \alpha_1 + \mathbb{E}[\mathbf{f}_X]^\top \beta_1 + \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=0]^\top \gamma_1 \\
&= (\alpha_1 + \delta_0^\top \gamma_1) + \mathbb{E}[\mathbf{f}_X]^\top (\beta_1 + B_0^\top \gamma_1) \\
&= \alpha_1 + \delta_0^\top \gamma_1 \,, \\
\mathbb{E}[Y^{(0,M^{(1)})}] &= \alpha_0 + \mathbb{E}[\mathbf{f}_X]^\top \beta_0 + \mathbb{E}[\mathbf{f}_M \mid \mathbf{f}_X, A=1]^\top \gamma_0 \\
&= (\alpha_0 + \delta_1^\top \gamma_0) + \mathbb{E}[\mathbf{f}_X]^\top (\beta_0 + B_1^\top \gamma_0) \\
&= \alpha_0 + \delta_1^\top \gamma_0 \,.
\end{aligned}$$

Note that $\mathbb{E}[Y^{(1,M^{(0)})}] = \alpha_1 + \delta_0^\top \gamma_1$ is exactly the estimand $\theta_0 = \alpha_1 + \gamma_1^\top \delta_0$ studied in Theorem 3.1 and throughout the rest of this Appendix.

The natural indirect effect holds exposure fixed at treatment and varies the mediator between its two natural values

$$\begin{aligned}
\text{NIE} &= \mathbb{E}[Y^{(1,M^{(1)})}] - \mathbb{E}[Y^{(1,M^{(0)})}] \\
&= (\alpha_1 + \delta_1^\top \gamma_1) - (\alpha_1 + \delta_0^\top \gamma_1) \\
&= (\delta_1 - \delta_0)^\top \gamma_1 \,.
\end{aligned}$$

The natural direct effect holds the mediator fixed at its control-arm natural value and varies exposure

$$\begin{aligned}
\text{NDE} &= \mathbb{E}[Y^{(1,M^{(0)})}] - \mathbb{E}[Y^{(0,M^{(0)})}] \\
&= (\alpha_1 + \delta_0^\top \gamma_1) - (\alpha_0 + \delta_0^\top \gamma_0) \\
&= \alpha_1 - \alpha_0 + \delta_0^\top (\gamma_1 - \gamma_0) \,.
\end{aligned}$$

The remaining potential outcome, $\mathbb{E}[Y^{(0,M^{(1)})}] = \alpha_0 + \delta_1^\top \gamma_0$, is not needed for NIE/NDE above but gives the complementary decomposition, $\text{TDE} = \mathbb{E}[Y^{(1,M^{(1)})}] - \mathbb{E}[Y^{(0,M^{(1)})}]$ and $\text{TIE} = \mathbb{E}[Y^{(0,M^{(1)})}] - \mathbb{E}[Y^{(0,M^{(0)})}]$, which satisfies the same identity $\text{TDE} + \text{TIE} = \text{NDE} + \text{NIE}$.

D.3 More on Simulation Study and Sensitivity Analysis

		$\sigma = 0.1$				$\sigma = 1$			
$p = q$	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
500	500	0.1988	0.1990	0.7799	0.956	0.2209	0.2211	0.8665	0.944
	1000	0.1493	0.1494	0.5855	0.942	0.1496	0.1497	0.5867	0.954
	1500	0.1156	0.1157	0.4534	0.948	0.1207	0.1206	0.4729	0.956
	2000	0.1046	0.1046	0.4098	0.936	0.1060	0.1043	0.4089	0.948
	2500	0.0851	0.0849	0.3330	0.954	0.0939	0.0937	0.3674	0.954
1000	500	0.1645	0.1646	0.6452	0.950	0.1355	0.1352	0.5299	0.950
	1000	0.1078	0.1075	0.4214	0.952	0.0937	0.0938	0.3678	0.952
	1500	0.0923	0.0924	0.3621	0.948	0.0747	0.0745	0.2920	0.944
	2000	0.0813	0.0814	0.3189	0.950	0.0651	0.0652	0.2556	0.942
	2500	0.0746	0.0746	0.2923	0.954	0.0557	0.0557	0.2185	0.948
2500	500	0.1737	0.1738	0.6814	0.954	0.2758	0.2760	1.0818	0.964
	1000	0.1266	0.1267	0.4968	0.948	0.1900	0.1897	0.7436	0.956
	1500	0.0951	0.0952	0.3731	0.930	0.1575	0.1574	0.6169	0.936
	2000	0.0887	0.0888	0.3480	0.948	0.1265	0.1266	0.4961	0.956
	2500	0.0763	0.0763	0.2993	0.956	0.1205	0.1206	0.4726	0.952
5000	500	0.1754	0.1755	0.6880	0.952	0.2106	0.2108	0.8264	0.952
	1000	0.1190	0.1191	0.4668	0.958	0.1323	0.1324	0.5188	0.940
	1500	0.1012	0.1013	0.3970	0.948	0.1077	0.1078	0.4227	0.942
	2000	0.0880	0.0872	0.3417	0.942	0.0976	0.0977	0.3830	0.956
	2500	0.0791	0.0792	0.3103	0.950	0.0852	0.0850	0.3332	0.968

Table 3: Performance of $\hat{\theta}_0$ under variance-matched t_6 errors, over 500 replications.

Table 4: Performance of $\hat{\theta}_0$ as the strength of the mediated pathway varies, $\gamma_1 = \gamma_0 = c\,g$, over 500 replications.

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	c	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
1000	0	500	0.0085	0.0085	0.0333	0.948	0.0859	0.0859	0.3365	0.950
		1000	0.0056	0.0056	0.0219	0.956	0.0585	0.0585	0.2292	0.950
		1500	0.0047	0.0047	0.0185	0.950	0.0469	0.0468	0.1836	0.952
		2000	0.0039	0.0039	0.0152	0.956	0.0393	0.0384	0.1506	0.952
		2500	0.0036	0.0036	0.0142	0.950	0.0371	0.0372	0.1456	0.954
1000	0.1	500	0.0204	0.0203	0.0797	0.958	0.0840	0.0840	0.3294	0.958
		1000	0.0145	0.0145	0.0570	0.938	0.0590	0.0590	0.2311	0.948
		1500	0.0112	0.0112	0.0438	0.946	0.0465	0.0465	0.1823	0.946
		2000	0.0101	0.0101	0.0395	0.942	0.0395	0.0394	0.1544	0.950
		2500	0.0085	0.0085	0.0333	0.948	0.0366	0.0366	0.1433	0.958
1000	0.25	500	0.0433	0.0432	0.1692	0.950	0.0938	0.0939	0.3681	0.948
		1000	0.0317	0.0317	0.1242	0.946	0.0635	0.0635	0.2488	0.958
		1500	0.0250	0.0251	0.0982	0.948	0.0569	0.0570	0.2234	0.952
		2000	0.0225	0.0225	0.0881	0.948	0.0452	0.0453	0.1774	0.944
		2500	0.0200	0.0200	0.0784	0.954	0.0399	0.0398	0.1561	0.940

Continued on next page

Table 4 — continued

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	c	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.0908	0.0908	0.3561	0.952	0.1164	0.1165	0.4567	0.944
		1000	0.0632	0.0632	0.2479	0.948	0.0879	0.0879	0.3447	0.944
1000	0.5	1500	0.0531	0.0531	0.2083	0.942	0.0699	0.0700	0.2743	0.954
		2000	0.0463	0.0463	0.1817	0.952	0.0611	0.0610	0.2392	0.940
		2500	0.0382	0.0382	0.1499	0.946	0.0535	0.0533	0.2089	0.960
		500	0.1885	0.1886	0.7394	0.958	0.1942	0.1941	0.7610	0.952
		1000	0.1263	0.1263	0.4950	0.950	0.1442	0.1443	0.5658	0.950
1000	1	1500	0.0999	0.1000	0.3920	0.944	0.1120	0.1121	0.4393	0.940
		2000	0.0886	0.0884	0.3464	0.950	0.0966	0.0963	0.3776	0.954
		2500	0.0858	0.0857	0.3361	0.960	0.0835	0.0835	0.3275	0.956
		500	0.0093	0.0093	0.0364	0.950	0.0965	0.0966	0.3785	0.954
		1000	0.0066	0.0066	0.0257	0.942	0.0658	0.0659	0.2583	0.956
2500	0	1500	0.0050	0.0050	0.0195	0.962	0.0508	0.0504	0.1977	0.948
		2000	0.0044	0.0044	0.0173	0.946	0.0429	0.0429	0.1681	0.958
		2500	0.0040	0.0039	0.0154	0.952	0.0390	0.0390	0.1529	0.944
		500	0.0185	0.0185	0.0726	0.948	0.0936	0.0937	0.3672	0.944
		1000	0.0132	0.0132	0.0518	0.948	0.0655	0.0655	0.2568	0.942
2500	0.1	1500	0.0103	0.0103	0.0405	0.962	0.0528	0.0527	0.2066	0.948
		2000	0.0092	0.0092	0.0361	0.950	0.0467	0.0467	0.1832	0.960
		2500	0.0085	0.0085	0.0334	0.960	0.0408	0.0408	0.1599	0.946
		500	0.0410	0.0410	0.1606	0.940	0.1009	0.1010	0.3959	0.948
		1000	0.0321	0.0321	0.1258	0.958	0.0741	0.0740	0.2900	0.938
2500	0.25	1500	0.0236	0.0237	0.0927	0.952	0.0579	0.0580	0.2272	0.952
		2000	0.0195	0.0196	0.0767	0.952	0.0468	0.0468	0.1835	0.952
		2500	0.0186	0.0187	0.0731	0.942	0.0403	0.0403	0.1579	0.952
		500	0.0831	0.0830	0.3254	0.948	0.1314	0.1315	0.5155	0.952
		1000	0.0597	0.0598	0.2342	0.950	0.0808	0.0808	0.3169	0.942
2500	0.5	1500	0.0467	0.0467	0.1832	0.950	0.0684	0.0684	0.2682	0.958
		2000	0.0384	0.0384	0.1505	0.948	0.0606	0.0603	0.2363	0.950
		2500	0.0350	0.0350	0.1371	0.956	0.0514	0.0514	0.2014	0.952
		500	0.1615	0.1615	0.6331	0.948	0.1838	0.1838	0.7205	0.948
		1000	0.1153	0.1154	0.4522	0.946	0.1348	0.1349	0.5287	0.958
2500	1	1500	0.0884	0.0885	0.3467	0.954	0.1045	0.1045	0.4095	0.944
		2000	0.0826	0.0825	0.3232	0.952	0.0891	0.0887	0.3478	0.950
		2500	0.0734	0.0733	0.2874	0.952	0.0811	0.0812	0.3183	0.946
		500	0.0099	0.0099	0.0387	0.948	0.1024	0.1024	0.4014	0.952
		1000	0.0065	0.0065	0.0255	0.948	0.0650	0.0651	0.2551	0.940
5000	0	1500	0.0055	0.0055	0.0216	0.952	0.0495	0.0495	0.1940	0.944
		2000	0.0045	0.0045	0.0176	0.956	0.0452	0.0453	0.1775	0.958
		2500	0.0042	0.0042	0.0163	0.956	0.0420	0.0420	0.1645	0.934
		500	0.0210	0.0210	0.0823	0.956	0.0998	0.0999	0.3914	0.942
		1000	0.0159	0.0159	0.0625	0.950	0.0647	0.0647	0.2536	0.954
5000	0.1	1500	0.0119	0.0119	0.0465	0.948	0.0551	0.0550	0.2155	0.946
		2000	0.0100	0.0100	0.0391	0.952	0.0480	0.0480	0.1883	0.948
		2500	0.0094	0.0094	0.0367	0.946	0.0434	0.0435	0.1704	0.952

Continued on next page

Table 4 — continued

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	c	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.0495	0.0495	0.1940	0.960	0.1117	0.1116	0.4376	0.946
		1000	0.0364	0.0364	0.1426	0.960	0.0730	0.0730	0.2862	0.960
5000	0.25	1500	0.0274	0.0275	0.1076	0.948	0.0593	0.0593	0.2325	0.954
		2000	0.0259	0.0259	0.1015	0.942	0.0527	0.0528	0.2069	0.952
		2500	0.0220	0.0221	0.0865	0.956	0.0462	0.0463	0.1814	0.954
		500	0.0960	0.0960	0.3764	0.950	0.1483	0.1483	0.5815	0.942
		1000	0.0660	0.0661	0.2590	0.950	0.0942	0.0943	0.3696	0.962
5000	0.5	1500	0.0559	0.0560	0.2194	0.934	0.0797	0.0797	0.3126	0.946
		2000	0.0477	0.0477	0.1871	0.958	0.0684	0.0685	0.2683	0.964
		2500	0.0435	0.0435	0.1707	0.952	0.0623	0.0623	0.2444	0.948
		500	0.1888	0.1890	0.7409	0.944	0.2238	0.2240	0.8781	0.950
		1000	0.1433	0.1432	0.5613	0.958	0.1534	0.1535	0.6017	0.940
5000	1	1500	0.1126	0.1126	0.4415	0.940	0.1216	0.1216	0.4766	0.946
		2000	0.0983	0.0983	0.3854	0.954	0.1069	0.1070	0.4194	0.946
		2500	0.0875	0.0876	0.3432	0.946	0.0916	0.0916	0.3592	0.956

Table 5: Sensitivity of $\hat{\theta}_0$ to the common working dimension $\kappa = \tilde{p} = \tilde{q}$, with $\bar{p} = \bar{q} = 3$, over 500 replications.

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	κ	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.2470	0.1902	0.7454	0.880	0.2609	0.2072	0.8122	0.894
		1000	0.2044	0.1355	0.5311	0.796	0.2140	0.1518	0.5952	0.840
1000	1	1500	0.1925	0.1076	0.4218	0.684	0.1975	0.1200	0.4704	0.770
		2000	0.1846	0.0994	0.3895	0.654	0.1916	0.1000	0.3919	0.630
		2500	0.1726	0.0875	0.3432	0.610	0.1764	0.0909	0.3564	0.610
		500	0.1986	0.1983	0.7772	0.960	0.2207	0.2209	0.8659	0.944
		1000	0.1461	0.1451	0.5690	0.944	0.1609	0.1610	0.6311	0.954
1000	2	1500	0.1183	0.1179	0.4620	0.944	0.1321	0.1302	0.5102	0.944
		2000	0.1067	0.1060	0.4154	0.950	0.1122	0.1104	0.4328	0.942
		2500	0.0878	0.0870	0.3411	0.942	0.1008	0.0996	0.3903	0.946
		500	0.1956	0.1951	0.7647	0.952	0.2240	0.2242	0.8790	0.954
		1000	0.1484	0.1484	0.5818	0.946	0.1485	0.1483	0.5815	0.944
1000	3	1500	0.1151	0.1152	0.4515	0.944	0.1269	0.1269	0.4975	0.952
		2000	0.1053	0.1054	0.4130	0.944	0.1106	0.1107	0.4339	0.952
		2500	0.0931	0.0932	0.3653	0.954	0.1016	0.1015	0.3980	0.952
		500	0.2138	0.2140	0.8390	0.948	0.2285	0.2287	0.8965	0.950
		1000	0.1492	0.1493	0.5852	0.946	0.1700	0.1697	0.6653	0.948
1000	5	1500	0.1252	0.1253	0.4913	0.936	0.1306	0.1307	0.5123	0.948
		2000	0.1098	0.1098	0.4306	0.954	0.1104	0.1104	0.4329	0.938
		2500	0.0900	0.0901	0.3531	0.944	0.1035	0.1032	0.4047	0.942
		500	0.2235	0.2236	0.8767	0.960	0.2349	0.2351	0.9216	0.950
		1000	0.1500	0.1500	0.5881	0.948	0.1608	0.1609	0.6308	0.962
1000	7	1500	0.1190	0.1191	0.4670	0.944	0.1297	0.1297	0.5084	0.944
		2000	0.1052	0.1049	0.4112	0.950	0.1110	0.1111	0.4353	0.956
		2500	0.1012	0.1012	0.3966	0.960	0.1019	0.1019	0.3996	0.948

Continued on next page

Table 5 — continued

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	κ	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.2242	0.2244	0.8795	0.956	0.2392	0.2394	0.9384	0.942
		1000	0.1536	0.1535	0.6017	0.942	0.1643	0.1643	0.6440	0.942
1000	10	1500	0.1211	0.1212	0.4752	0.956	0.1269	0.1270	0.4979	0.950
		2000	0.1035	0.1035	0.4058	0.948	0.1210	0.1210	0.4744	0.944
		2500	0.0925	0.0926	0.3630	0.950	0.0965	0.0962	0.3772	0.948
		500	0.2176	0.2175	0.8525	0.954	0.2529	0.2531	0.9922	0.948
		1000	0.1501	0.1503	0.5891	0.948	0.1674	0.1675	0.6565	0.942
1000	15	1500	0.1233	0.1234	0.4837	0.960	0.1265	0.1265	0.4961	0.950
		2000	0.1067	0.1067	0.4182	0.954	0.1111	0.1112	0.4361	0.954
		2500	0.0949	0.0950	0.3723	0.960	0.1047	0.1048	0.4109	0.962
		500	0.2290	0.2293	0.8987	0.948	0.2499	0.2493	0.9774	0.958
		1000	0.1577	0.1577	0.6180	0.952	0.1678	0.1677	0.6575	0.948
1000	20	1500	0.1244	0.1243	0.4873	0.948	0.1350	0.1350	0.5291	0.948
		2000	0.1036	0.1037	0.4065	0.950	0.1176	0.1176	0.4611	0.944
		2500	0.0970	0.0971	0.3807	0.950	0.1026	0.1027	0.4027	0.950
		500	0.2370	0.2371	0.9293	0.950	0.2665	0.2665	1.0445	0.952
		1000	0.1634	0.1635	0.6409	0.948	0.1680	0.1682	0.6592	0.964
1000	30	1500	0.1222	0.1222	0.4792	0.954	0.1342	0.1342	0.5262	0.946
		2000	0.1080	0.1078	0.4225	0.934	0.1099	0.1099	0.4309	0.946
		2500	0.0961	0.0959	0.3758	0.954	0.1094	0.1095	0.4291	0.944
		500	0.3131	0.1519	0.5956	0.566	0.3236	0.1722	0.6749	0.656
		1000	0.2947	0.1080	0.4233	0.274	0.2878	0.1262	0.4947	0.464
2500	1	1500	0.2747	0.0889	0.3485	0.154	0.2923	0.1004	0.3935	0.234
		2000	0.2747	0.0743	0.2912	0.052	0.2855	0.0861	0.3375	0.118
		2500	0.2692	0.0660	0.2586	0.028	0.2704	0.0759	0.2976	0.066
		500	0.3131	0.1993	0.7814	0.770	0.3550	0.2184	0.8563	0.744
		1000	0.3093	0.1662	0.6516	0.626	0.3039	0.1807	0.7083	0.728
2500	2	1500	0.3041	0.1402	0.5496	0.532	0.3036	0.1581	0.6196	0.616
		2000	0.2897	0.1437	0.5632	0.590	0.2920	0.1383	0.5421	0.564
		2500	0.2938	0.1370	0.5370	0.536	0.3044	0.1463	0.5734	0.530
		500	0.1532	0.1533	0.6010	0.944	0.1749	0.1749	0.6858	0.946
		1000	0.1159	0.1159	0.4542	0.952	0.1249	0.1251	0.4903	0.956
2500	3	1500	0.0859	0.0860	0.3372	0.948	0.1016	0.1017	0.3987	0.964
		2000	0.0738	0.0739	0.2897	0.962	0.0905	0.0903	0.3540	0.958
		2500	0.0692	0.0692	0.2713	0.934	0.0766	0.0767	0.3006	0.962
		500	0.1584	0.1582	0.6202	0.944	0.1870	0.1869	0.7328	0.956
		1000	0.1127	0.1128	0.4421	0.948	0.1254	0.1254	0.4917	0.958
2500	5	1500	0.0899	0.0899	0.3523	0.944	0.1037	0.1038	0.4067	0.950
		2000	0.0745	0.0745	0.2922	0.944	0.0891	0.0891	0.3493	0.960
		2500	0.0672	0.0673	0.2637	0.958	0.0787	0.0788	0.3089	0.958
		500	0.1592	0.1590	0.6234	0.944	0.1846	0.1845	0.7233	0.944
		1000	0.1108	0.1108	0.4345	0.942	0.1239	0.1235	0.4842	0.946
2500	7	1500	0.0843	0.0843	0.3306	0.950	0.0983	0.0984	0.3857	0.946
		2000	0.0794	0.0793	0.3108	0.950	0.0877	0.0877	0.3438	0.946
		2500	0.0701	0.0701	0.2748	0.958	0.0783	0.0784	0.3073	0.950

Continued on next page

Table 5 — continued

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	κ	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.1494	0.1496	0.5863	0.946	0.1835	0.1834	0.7189	0.954
		1000	0.1073	0.1074	0.4210	0.944	0.1272	0.1273	0.4991	0.954
2500	10	1500	0.0939	0.0938	0.3677	0.950	0.0986	0.0987	0.3868	0.940
		2000	0.0776	0.0776	0.3042	0.946	0.0869	0.0870	0.3410	0.948
		2500	0.0681	0.0681	0.2671	0.950	0.0793	0.0790	0.3098	0.936
		500	0.1678	0.1679	0.6580	0.944	0.1960	0.1962	0.7689	0.946
		1000	0.1054	0.1054	0.4131	0.948	0.1305	0.1306	0.5120	0.952
2500	15	1500	0.0847	0.0846	0.3314	0.960	0.1096	0.1096	0.4298	0.952
		2000	0.0704	0.0703	0.2755	0.934	0.0862	0.0862	0.3380	0.940
		2500	0.0648	0.0647	0.2537	0.954	0.0764	0.0765	0.2998	0.950
		500	0.1651	0.1653	0.6479	0.954	0.2014	0.2015	0.7900	0.948
		1000	0.1108	0.1109	0.4346	0.948	0.1374	0.1375	0.5390	0.948
2500	20	1500	0.0853	0.0854	0.3346	0.940	0.1043	0.1043	0.4089	0.938
		2000	0.0779	0.0777	0.3046	0.954	0.0874	0.0875	0.3428	0.948
		2500	0.0685	0.0685	0.2685	0.942	0.0747	0.0745	0.2919	0.954
		500	0.1851	0.1851	0.7254	0.942	0.2226	0.2225	0.8724	0.958
		1000	0.1134	0.1135	0.4448	0.954	0.1374	0.1373	0.5382	0.950
2500	30	1500	0.0913	0.0914	0.3581	0.960	0.1072	0.1073	0.4206	0.948
		2000	0.0771	0.0772	0.3024	0.952	0.0923	0.0924	0.3621	0.956
		2500	0.0675	0.0675	0.2648	0.948	0.0809	0.0810	0.3174	0.944
		500	0.3050	0.2377	0.9318	0.872	0.3197	0.2629	1.0304	0.908
		1000	0.2479	0.1642	0.6437	0.790	0.2586	0.1761	0.6903	0.818
5000	1	1500	0.2335	0.1285	0.5038	0.670	0.2356	0.1440	0.5645	0.750
		2000	0.2201	0.1151	0.4513	0.630	0.2325	0.1264	0.4956	0.632
		2500	0.2148	0.1017	0.3988	0.540	0.2205	0.1063	0.4167	0.556
		500	0.3087	0.2478	0.9712	0.888	0.3139	0.2486	0.9745	0.884
		1000	0.2540	0.1662	0.6515	0.794	0.2758	0.1974	0.7738	0.840
5000	2	1500	0.2442	0.1371	0.5375	0.696	0.2420	0.1427	0.5593	0.726
		2000	0.2297	0.1208	0.4736	0.640	0.2282	0.1220	0.4782	0.646
		2500	0.2235	0.1031	0.4042	0.528	0.2349	0.1136	0.4451	0.560
		500	0.2370	0.2370	0.9292	0.958	0.2530	0.2519	0.9872	0.950
		1000	0.1759	0.1760	0.6900	0.962	0.1818	0.1815	0.7116	0.946
5000	3	1500	0.1357	0.1358	0.5325	0.958	0.1457	0.1457	0.5713	0.948
		2000	0.1261	0.1262	0.4946	0.944	0.1384	0.1383	0.5420	0.948
		2500	0.1079	0.1080	0.4235	0.956	0.1098	0.1099	0.4307	0.954
		500	0.2346	0.2347	0.9201	0.942	0.2498	0.2500	0.9801	0.950
		1000	0.1640	0.1642	0.6436	0.946	0.1834	0.1835	0.7195	0.948
5000	5	1500	0.1378	0.1379	0.5405	0.936	0.1538	0.1539	0.6034	0.954
		2000	0.1180	0.1180	0.4626	0.950	0.1223	0.1225	0.4800	0.946
		2500	0.1077	0.1078	0.4226	0.948	0.1091	0.1092	0.4279	0.948
		500	0.2342	0.2345	0.9191	0.944	0.2677	0.2678	1.0497	0.952
		1000	0.1777	0.1776	0.6963	0.958	0.1751	0.1750	0.6861	0.960
5000	7	1500	0.1399	0.1400	0.5489	0.942	0.1463	0.1460	0.5724	0.952
		2000	0.1226	0.1227	0.4810	0.956	0.1236	0.1231	0.4823	0.946
		2500	0.1097	0.1098	0.4304	0.942	0.1094	0.1090	0.4273	0.948

Continued on next page

Table 5 — continued

			$\sigma = 0.1$				$\sigma = 1$			
$p = q$	κ	n	RMSE	SD	CI Length	Coverage	RMSE	SD	CI Length	Coverage
		500	0.2441	0.2444	0.9579	0.954	0.2755	0.2751	1.0785	0.956
		1000	0.1640	0.1641	0.6434	0.954	0.1807	0.1809	0.7090	0.946
5000	10	1500	0.1423	0.1425	0.5585	0.936	0.1546	0.1545	0.6058	0.942
		2000	0.1246	0.1247	0.4887	0.956	0.1346	0.1347	0.5281	0.950
		2500	0.1061	0.1057	0.4142	0.938	0.1269	0.1268	0.4970	0.956
		500	0.2487	0.2489	0.9758	0.954	0.2773	0.2775	1.0879	0.950
		1000	0.1818	0.1820	0.7134	0.952	0.1875	0.1870	0.7331	0.956
5000	15	1500	0.1401	0.1398	0.5481	0.944	0.1476	0.1476	0.5787	0.958
		2000	0.1174	0.1174	0.4601	0.958	0.1242	0.1244	0.4875	0.936
		2500	0.1073	0.1074	0.4212	0.946	0.1143	0.1145	0.4487	0.954
		500	0.2577	0.2576	1.0099	0.942	0.2903	0.2905	1.1389	0.960
		1000	0.1664	0.1665	0.6525	0.954	0.1900	0.1901	0.7453	0.962
5000	20	1500	0.1416	0.1417	0.5556	0.954	0.1469	0.1470	0.5764	0.942
		2000	0.1189	0.1190	0.4663	0.956	0.1227	0.1226	0.4805	0.944
		2500	0.1118	0.1120	0.4389	0.948	0.1157	0.1158	0.4540	0.944
		500	0.2774	0.2775	1.0878	0.950	0.3107	0.3110	1.2193	0.950
		1000	0.1789	0.1791	0.7021	0.940	0.1963	0.1965	0.7702	0.944
5000	30	1500	0.1445	0.1446	0.5668	0.954	0.1548	0.1549	0.6071	0.948
		2000	0.1202	0.1203	0.4717	0.950	0.1252	0.1253	0.4912	0.946
		2500	0.1117	0.1118	0.4381	0.948	0.1167	0.1168	0.4577	0.948

All 500 replications were usable in every cell.

		$\sigma = 0.1$			$\sigma = 1$		
$p = q$	$\tilde{p}$	$\tilde{q} = 3$	$\tilde{q} = 7$	$\tilde{q} = 20$	$\tilde{q} = 3$	$\tilde{q} = 7$	$\tilde{q} = 20$
	3	0.0739 (0.940)	0.0702 (0.968)	0.0722 (0.956)	0.0699 (0.946)	0.0690 (0.946)	0.0720 (0.940)
1000	7	0.0690 (0.946)	0.0688 (0.934)	0.0706 (0.954)	0.0733 (0.944)	0.0749 (0.952)	0.0711 (0.938)
	20	0.0726 (0.956)	0.0750 (0.942)	0.0714 (0.944)	0.0724 (0.946)	0.0746 (0.942)	0.0764 (0.946)
	3	0.0581 (0.948)	0.0550 (0.952)	0.0568 (0.950)	0.0725 (0.954)	0.0687 (0.954)	0.0694 (0.948)
2500	7	0.0536 (0.948)	0.0581 (0.962)	0.0578 (0.956)	0.0721 (0.948)	0.0714 (0.952)	0.0669 (0.962)
	20	0.0594 (0.952)	0.0571 (0.950)	0.0534 (0.940)	0.0704 (0.956)	0.0707 (0.946)	0.0701 (0.944)
	3	0.0932 (0.942)	0.0972 (0.946)	0.0938 (0.952)	0.0829 (0.944)	0.0816 (0.946)	0.0820 (0.954)
5000	7	0.0966 (0.938)	0.0933 (0.956)	0.1036 (0.954)	0.0832 (0.944)	0.0829 (0.964)	0.0827 (0.954)
	20	0.1002 (0.940)	0.0973 (0.944)	0.1014 (0.956)	0.0828 (0.946)	0.0862 (0.948)	0.0842 (0.956)

Table 6: RMSE of $\hat{\theta}_0$ (coverage in parentheses) for $\tilde{p}$ and $\tilde{q}$ varied separately, with $\bar{p} = \bar{q} = 3$ and $n = 2500$, over 500 replications.

D.4 Covariates

Variable Name	Description
PTDOBYY	Date of Birth (in year)
PTGENDER	Gender of the subject (e.g., Male or Female)
PTMARRY	Marital status (e.g., Married, Single, Divorced)
PTEDUCAT	Years of formal education completed
PTNOTRT	Indicator for treatment group or non-treatment
VSWEIGHT	Body weight (in pounds)
VSPULSE	Pulse rate (beats per minute)
VSRESP	Respiratory rate (breaths per minute)
VSTEMP	Body temperature (in Fahrenheit)
VSHEIGHT	Height (in inches)

Table 7: Descriptions of patient background variables used as covariates in the model.

E Comparison to Related Literature

E.1 Comparison to Zhang et al. (2026a)

The method of Zhang et al. (2026a) is related to FARMA because both use latent factor models in high-dimensional mediation settings. However, the high-dimensional variables and latent factors play different roles in the two methods, and the causal targets are also different.

Using the notation of Zhang et al. (2026a), let $\boldsymbol{x} \in \mathbb{R}^q$ be a high-dimensional exposure vector, $\boldsymbol{m} \in \mathbb{R}^p$ a high-dimensional mediator vector, and $y \in \mathbb{R}$ the outcome. Their model is

$$y = \boldsymbol{\alpha}_0^\top \boldsymbol{m} + \boldsymbol{\alpha}_1^\top \boldsymbol{x} + \varepsilon_1,$$
$$\boldsymbol{m} = \boldsymbol{\Gamma}^\top \boldsymbol{x} + \boldsymbol{\varepsilon}.$$

The direct, indirect, and total effects are

$$\boldsymbol{\alpha}_1, \quad \boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\alpha}_0, \quad \boldsymbol{\gamma} = \boldsymbol{\alpha}_1 + \boldsymbol{\beta}.$$

In order to account for strong dependence among the high-dimensional variables, PRIME uses the factor models

$$\boldsymbol{x} = B_1 \boldsymbol{f}_1 + \boldsymbol{v}_1, \qquad \begin{pmatrix} \boldsymbol{m} \\ \boldsymbol{x} \end{pmatrix} = B_2 \boldsymbol{f}_2 + \boldsymbol{v}_2.$$

For each prespecified exposure group g, PRIME estimates $\boldsymbol{\alpha}_{1,g}$ and $\boldsymbol{\gamma}_g$, and gives us the indirect effect from

$$\widehat{\boldsymbol{\beta}}_g = \widehat{\boldsymbol{\gamma}}_g - \widehat{\boldsymbol{\alpha}}_{1,g}.$$

The main differences from FARMA are given below.

1. **Observed variables and surrogate measurements.**
   In Zhang et al. (2026a), $\boldsymbol{x}$ and $\boldsymbol{m}$ are observed exposures and mediators, and the direct and indirect effects are defined for their original coordinates. The factors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are introduced to account for the dependence among these observed variables.
   In FARMA, X and M are not the covariates and mediators themselves. They are high-dimensional surrogate measurements of the latent factors $\boldsymbol{f}_X$ and $\boldsymbol{f}_M$:
   $$X = \Gamma_X \boldsymbol{f}_X + \boldsymbol{u}_X, \quad M = \Gamma_M \boldsymbol{f}_M + \boldsymbol{u}_M.$$
   Therefore, the individual coordinates of M are not treated as separate mediators. Instead, they provide information about the underlying latent mediator $\boldsymbol{f}_M$.

2. **Exposure and causal target.**
   Zhang et al. (2026a) considers a high-dimensional exposure vector and studies direct and indirect effects for individual exposures or prespecified groups:
   $$H_{0,\mathrm{dir}}^{(g)} : \boldsymbol{\alpha}_{1,g} = 0, \qquad H_{0,\mathrm{ind}}^{(g)} : \boldsymbol{\beta}_g = 0.$$
   The corresponding inference problem is therefore high-dimensional and involves multiple testing.
   FARMA considers a binary treatment $A \in \{0, 1\}$. The main target is the population-level counterfactual mean
   $$\theta_0 = E\left\{ Y^{(1, \boldsymbol{f}_M^{(0)})} \right\},$$
   which is then used together with the treatment-specific counterfactual means to obtain the natural direct, natural indirect, and total effects. The goal is therefore different from identifying which coordinate of a high-dimensional exposure has a direct or indirect effect.
3. **Role of the factor structure.**
   In PRIME, the exposures and mediators are observed, and the factors are used to handle the strong dependence among the high-dimensional variables.
   In FARMA, the factors are themselves the latent covariates and mediators of interest. As a result, the factor structure also plays an important role in identification. With fixed p and q, the latent mediation functional is generally not point identified from (X, M) alone. We show that the observed-data functional approaches the latent target as the numbers of surrogate measurements increase. We also allow the working dimensions to be larger than the true factor dimensions:
   $$\widetilde{p} \geq \bar{p}, \qquad \widetilde{q} \geq \bar{q}.$$
   Thus, the exact latent factor dimensions do not need to be known.
4. **Model assumptions.**
   PRIME uses a high-dimensional linear structural equation model with sparsity assumptions on the regression coefficients. The main model does not include unmeasured confounding, although observed confounders can be added. The mediator-outcome coefficient is also common across exposure levels.
   FARMA includes latent pre-treatment covariates $\boldsymbol{f}_X$, which adjust for confounding among the treatment, mediator, and outcome. Since the regressions involving the latent factors are low-dimensional, our theory does not require sparsity. We also allow the coefficients
   $$(B_0, B_1), \qquad (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1), \qquad (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$$
   to vary across treatment levels.
5. **Estimation and inference.**
   PRIME estimates the factor and idiosyncratic components and then uses group-wise partially penalized least squares. Inference is based on tests for direct and indirect effects together with false discovery rate control.
   FARMA uses projected surrogate variables followed by two cross-fitted low-dimensional regression steps. The estimator targets a scalar mediation functional. We show that the estimator is $\sqrt{n}$-consistent and asymptotically normal, which allows inference for the population natural direct and indirect effects.

The difference can also be seen in the ADNI applications. Zhang et al. (2026a) use SNPs as high-dimensional exposures and regional PET measurements as observed mediators. In our application, depressive status is the treatment, while high-dimensional DNA methylation measurements are used as surrogates for an underlying latent mediating process.

Summary of differences:

	Zhang et al. (2026a): PRIME	**FARMA**
High-dimensional variables	Observed exposures and mediators	Surrogate measurements X and M
Role of factors	Account for dependence	Latent covariates and mediators
Exposure	High-dimensional $\boldsymbol{x}$	Binary treatment A
Target	Group-wise direct and indirect effects	Population NDE, NIE, and TE
Covariates	Not included in the core model	Latent $\boldsymbol{f}_X$ measured through X
Sparsity	Required for high-dimensional coefficients	Not required
Treatment-specific coefficients	No	Yes
Estimation	Factor-adjusted penalized regression	Cross-fitted projected regressions
Inference	Group-wise tests and FDR control	$\sqrt{n}$-based inference