Online Appendix
Synthetic Difference in Differences
Arkhangelsky, Athey, Hirshberg, Imbens, Wager

VI.1 Placebo Study Details

VI.1.1 CPS study

We use the annual CPS data available on the NBER website (`https://data.nber.org/morg/annual`). Following Bertrand et al. (2004) we restrict the sample to 25-50-year-old women in their fourth month of the interview. The complete dataset contains all available years from 1979 to 2018 for 50 states, excluding the District of Columbia. We drop the duplicates on the unique household number, household id, person line number in household, month in the sample, month and year of interview, state, and age. Average log wages and hours are computed using the sample with strictly positive earnings. Unemployment is calculated using the sample of individuals within the labor force.

We use three indicators D_i to estimate the assignment model via logistic regression as described in (13). The first is equal to an indicator that state i has a minimum wage that is higher than the federal minimum wage in the year 2000. This indicator was taken from `http://www.dol.gov/whd/state/stateMinWageHis.htm`; see Barrios et al. (2012) for details. The second indicator comes from a state having an open-carry gun law. This was taken from `https://lawcenter.giffords.org/gun-laws/policy-areas/guns-in-public/open-carry/`. The third indicator comes from the state not having a ban on partial birth abortions. This was taken from `https://www.guttmacher.org/state-policy/explore/overview-abortion-laws`. Table 5 presents the values for these indicators.

VI.1.2 Penn World Table study

We download the data on real annual GDP from the Penn World Table website (`https://www.rug.nl/ggdc/productivity/pwt/`). After removing the countries with missing data we end up with a dataset of 111 countries observed for 48 consecutive years, starting from 1959. To construct the assignment process we use Penn World Table indicators of democracy and education available from the same source.

State	Minimum Wage	Unrestricted Open Carry	Abortion
Alaska	0	1	0
Alabama	0	0	0
Arkansas	0	1	0
Arizona	0	1	0
California	1	0	1
Colorado	0	0	1
Connecticut	0	0	1
Delaware	1	1	1
Florida	0	0	0
Georgia	0	0	0
Hawaii	0	0	1
Idaho	0	1	0
Illinois	0	0	1
Indiana	0	0	0
Iowa	0	0	0
Kansas	0	1	0
Kentucky	0	1	0
Louisiana	0	1	0
Massachusetts	1	0	1
Maine	0	1	1
Maryland	0	0	1
Michigan	0	1	0
Minnesota	0	0	1
Mississippi	0	1	0
Missouri	0	0	0
Montana	0	1	0
Nebraska	0	1	0
Nevada	0	1	1
New Hampshire	0	1	0
New Mexico	0	1	0
North Carolina	0	1	1
North Dakota	0	0	0
New York	0	0	1
New Jersey	0	0	0
Ohio	0	1	0
Oklahoma	0	0	0
Oregon	1	1	1
Pennsylvania	0	0	1
Rhode Island	1	0	0
South Carolina	0	0	0
South Dakota	0	1	0
Tennessee	0	0	0
Texas	0	0	0
Utah	0	0	0
Vermont	1	1	1
Virginia	0	0	0
Washington	1	0	1
West Virginia	0	1	0
Wisconsin	0	1	0
Wyoming	0	1	1

Table 5: State Regulations

	SC	SC (reg)	DIFP	DIFP (reg)
Baseline	0.37	0.78	0.32	0.36
No Correlation	0.38	0.79	0.32	0.36
No $\boldsymbol{M}$	0.18	0.34	0.16	0.14
No $\boldsymbol{F}$	0.23	0.25	0.32	0.36
Only noise	0.14	0.11	0.16	0.14
No noise	0.17	0.34	0.11	0.20
Gun Law	0.27	0.34	0.30	0.40
Abortion	0.31	0.65	0.27	0.35
Random	0.25	0.31	0.27	0.35
Hours	2.03	3.28	1.97	1.91
U-rate	2.31	3.31	2.30	3.32
$T_{post} = 1$	0.59	0.65	0.54	0.50
$N_{tr} = 1$	0.73	0.85	0.83	0.87
$T_{post} = N_{tr} = 1$	1.24	1.23	1.16	1.12
Resample, $N = 200$	0.17	0.16	0.18	0.18
Resample, $N = 400$	0.14	0.11	0.15	0.12
Democracy	0.38	0.35	0.39	0.31
Education	0.53	0.62	0.39	0.29
Random	0.46	0.47	0.45	0.46

Table 6: Comparison of SC and DIFP estimators without regularization and with the regularization parameter used to compute SDID unit weights. Simulation designs correspond to those of Table 2 and 3. All results are based on 1000 simulations and multiplied by 10 for readability.

VI.2 Unit/time weights for California

	DID	SC	SDID
1988	0.053	0.000	0.427
1987	0.053	0.000	0.206
1986	0.053	0.000	0.366
1985	0.053	0.000	0.000
1984	0.053	0.000	0.000
1983	0.053	0.000	0.000
1982	0.053	0.000	0.000
1981	0.053	0.000	0.000
1980	0.053	0.000	0.000
1979	0.053	0.000	0.000
1978	0.053	0.000	0.000
1977	0.053	0.000	0.000
1976	0.053	0.000	0.000
1975	0.053	0.000	0.000
1974	0.053	0.000	0.000
1973	0.053	0.000	0.000
1972	0.053	0.000	0.000
1971	0.053	0.000	0.000
1970	0.053	0.000	0.000

	DID	SC	SDID
Alabama	0.026	0.000	0.000
Arkansas	0.026	0.000	0.003
Colorado	0.026	0.013	0.058
Connecticut	0.026	0.104	0.078
Delaware	0.026	0.004	0.070
Georgia	0.026	0.000	0.002
Idaho	0.026	0.000	0.031
Illinois	0.026	0.000	0.053
Indiana	0.026	0.000	0.010
Iowa	0.026	0.000	0.026
Kansas	0.026	0.000	0.022
Kentucky	0.026	0.000	0.000
Louisiana	0.026	0.000	0.000
Maine	0.026	0.000	0.028
Minnesota	0.026	0.000	0.039
Mississippi	0.026	0.000	0.000
Missouri	0.026	0.000	0.008
Montana	0.026	0.232	0.045
Nebraska	0.026	0.000	0.048
Nevada	0.026	0.204	0.124
New Hampshire	0.026	0.045	0.105
New Mexico	0.026	0.000	0.041
North Carolina	0.026	0.000	0.033
North Dakota	0.026	0.000	0.000
Ohio	0.026	0.000	0.031
Oklahoma	0.026	0.000	0.000
Pennsylvania	0.026	0.000	0.015
Rhode Island	0.026	0.000	0.001
South Carolina	0.026	0.000	0.000
South Dakota	0.026	0.000	0.004
Tennessee	0.026	0.000	0.000
Texas	0.026	0.000	0.010
Utah	0.026	0.396	0.042
Vermont	0.026	0.000	0.000
Virginia	0.026	0.000	0.000
West Virginia	0.026	0.000	0.034
Wisconsin	0.026	0.000	0.037
Wyoming	0.026	0.000	0.001

VII Formal Results

In this section, we will outline the proof of Theorem 1. Recall from Section III.2 the decomposition of the SDID estimator's error into three terms: oracle noise, oracle confounding bias, and the deviation of the SDID estimator from the oracle. Our main task is bounding the deviation term. To do this, we prove an abstract high-probability bound, then derive a more concrete bound using results from a companion paper on penalized high-dimensional least squares with errors in variable (Hirshberg, 2021), and then show that this bound is $o\left((N_{\mathrm{tr}}T_{\mathrm{post}})^{-1/2}\right)$ under the assumptions of Theorem 1. Detailed proofs for each step are included in the next section.

Notation Throughout, each instance of c will denote a potentially different universal constant; $a \lesssim b$, $a \ll b$, and $a \sim b$ will mean $a \leq cb$, $a/b \to 0$, and $c \leq a/b \leq c$ respectively. $\|v\|$ and $\|A\|$ will denote the Euclidean norm $\|v\|_2$ for a vector v and the operator norm $\sup_{\|v\|_2 \leq 1}\|Av\|$ for a matrix A respectively; $\sigma_1(A), \sigma_2(A), \ldots$ will denote the singular values of A; $A_{i\cdot}$ and $A_{\cdot j}$ will denote the ith row and jth column of A; v' and A' will denote the transposes of a vector v and matrix A; and $[v; w] \in \mathbb{R}^{m+n}$ will denote the concatenation of vectors $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$.

VII.1 Abstract Setting

We will begin by describing an abstract setting that arises as a condensed form of the setting considered in our formal results in Section III. We observe an $N \times T$ matrix Y, which we will decompose as the sum $Y_{it} = L_{it} + 1(i = N, j = T)\tau + \varepsilon$ of a deterministic matrix L and a random matrix ε. We will refer to four blocks,

$$Y = \begin{pmatrix} Y_{::} & Y_{:T} \\ Y_{N:} & Y_{NT} \end{pmatrix},$$

where $Y_{::}$ is a submatrix that omits the last row and column, $Y_{N:}$ is the last row omitting its last element, and $Y_{:T}$ is the last column omitting its last element. We will use analogous notation for the parts of L and ε and let $N_0 = N - 1$ and $T_0 = T - 1$.

We assume that rows of ε are independent and subgaussian and that for $i \leq N_0$ they are identically distributed with linear post-on-pretreatment autoregression function $\mathrm{E}[\varepsilon_{iT} \mid \varepsilon_{i:}] = \varepsilon_{i:}\psi$ and covariance $\Sigma = \mathrm{E}\,\varepsilon_{i\cdot}'\varepsilon_{i\cdot}$ and let Σ^N be the covariance matrix of $\varepsilon_{N\cdot}$. We will refer to the covariance of the subvectors $\varepsilon_{i:}$ and $\varepsilon_{N:}$ as $\Sigma_{::}$ and $\Sigma_{::}^N$ respectively.

Our abstract results involve a bound K characterizing the concentration of the rows $\varepsilon_{i\cdot}$.

(34)

$$K \geq \max\left(1,\ \|\varepsilon_{1:}\Sigma_{::}^{-1/2}\|_{\psi_2},\ \|\varepsilon_{N:}(\Sigma_{::}^{N})^{-1/2}\|_{\psi_2}\ \frac{\|\varepsilon_{1T}-\varepsilon_{1:}\psi\|_{\psi_2|\varepsilon_{1:}}}{\|\varepsilon_{1T}-\varepsilon_{1:}\psi\|_{L_2}}\right),$$

$$P\left(\left|\|\varepsilon_{1:}\|^2 - \mathrm{E}\|\varepsilon_{1:}\|^2\right| \geq u\right) \leq c\exp\left(-c\min\left(\frac{u^2}{K^4\,\mathrm{E}\|\varepsilon_{1:}\|^2},\ \frac{u}{K^2\|\Sigma_{::}\|}\right)\right) \quad \text{for all} \quad u \geq 0.$$

Here we follow the convention (e.g., Vershynin, 2018) that the subgaussian norm of a random vector ξ is $\|\xi\|_{\psi_2} := \sup_{\|x\|\leq 1}\|x'\xi\|_{\psi_2}$. The conditional subgaussian norm $\|\cdot\|_{\psi_2|Z}$ is defined like the subgaussian norm the conditional distribution given Z. When the rows of ε are gaussian vectors, these conditions are satisfied for K equal to a sufficiently large universal constant. In the gaussian case, $\varepsilon_{1T}-\varepsilon_{1:}\psi$ is independent of $\varepsilon_{i:}$, the squared subgaussian norm of a gaussian random vector is bounded by a multiple of the operator norm of its covariance, and the concentration of $\|\varepsilon_{1:}\|^2$ as above is implied by the Hanson-Wright inequality (e.g., Vershynin, 2018, Theorem 6.2.1).

VII.2 Concrete Setting

We map from the setting considered in Section III to our condensed form by averaging within blocks as follows.

$$\begin{pmatrix} Y_{::} & Y_{:T} \\ Y_{N:} & Y_{NT} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{\text{co,pre}} & \boldsymbol{Y}_{\text{co,post}}\lambda_{\text{post}} \\ \omega_{\text{tr}}'\boldsymbol{Y}_{\text{tr,pre}} & \omega_{\text{tr}}'\boldsymbol{Y}_{\text{tr,post}}\lambda_{\text{post}} \end{pmatrix}.$$

Here $\lambda_{\text{post}} \in \mathbb{R}^{T_{\text{post}}}$ and $\omega_{\text{tr}} \in \mathbb{R}^{N_{\text{tr}}}$ are vectors with equal weight $1/T_{\text{post}}$ and $1/N_{\text{tr}}$ respectively. When working with this condensed form, we write ω and λ for what is rendered ω_{co} and λ_{tr} in Section III. We will also use Ω and Λ to denote the sets that would be written $\{\omega_{\text{co}} : \omega \in \Omega\}$ and $\{\lambda_{\text{pre}} : \lambda \in \Lambda\}$ in the notation used in Equations 4 and 6. Note that these sets Ω and Λ are the unit simplex in $\mathbb{R}^{N_0} = \mathbb{R}^{N_{\text{co}}}$ and $\mathbb{R}^{T_0} = \mathbb{R}^{T_{\text{pre}}}$ respectively.

In this condensed form, rows $\varepsilon_{i\cdot}$ are independent gaussian vectors with mean zero and covariance matrix Σ for $i \leq N_0$ and $N_{\text{tr}}^{-1}\Sigma$ for $i = N$. This matrix Σ satisfies, with quantities on the right defined as in Section III,

$$\Sigma = \begin{pmatrix} \Sigma_{\text{pre,pre}} & \Sigma_{\text{pre,post}}\lambda_{\text{post}} \\ \lambda_{\text{post}}'\Sigma_{\text{post,pre}} & \lambda_{\text{post}}'\Sigma_{\text{post,post}}\lambda_{\text{post}} \end{pmatrix}.$$

Note that because all rows have the same covariance up to scale, they have the same autoregression vector, $\psi = \arg\min_{v \in \mathbb{R}^{T_0}} \mathrm{E}(\varepsilon_{i:} v - \varepsilon_{iT})^2$. This definition is equivalent to the one given in Section III. And this characterization of $\varepsilon_{i:}\psi$ as a least squares projection implies that $\varepsilon_{i:}\psi - \varepsilon_{iT}$ and $\varepsilon_{i:}$ are uncorrelated and, being jointly normal, therefore independent.

That the eigenvalues of non-condensed-form Σ are bounded and bounded away from zero implies that the eigenvalues of the submatrix $\Sigma_{::} = \Sigma_{\mathrm{pre},\mathrm{pre}}$ are bounded and bounded away from zero. Furthermore, it implies the variance of $\varepsilon_{i:}\psi - \varepsilon_{iT}$ is on the order of $1/T_{\mathrm{post}}$.

To show this, we establish an upper and lower bound of that order. We will write $\sigma_{\min}(\Sigma)$ and $\sigma_{\max}(\Sigma)$ for the smallest and largest eigenvalues of Σ. For the lower bound, we calculate its variance $\mathrm{E}\left(\varepsilon_{i\cdot} \cdot [\psi;\ -\lambda_{\mathrm{post}}]\right)^2 = [\psi;\ -\lambda_{\mathrm{post}}]\ \Sigma\ [\psi;\ -\lambda_{\mathrm{post}}]$, and observe that this is at least $\|[\psi; -\lambda_{\mathrm{post}}]\|^2 \sigma_{\min}(\Sigma)$. This implies an order $1/T_{\mathrm{post}}$ lower bound, as $\|[\psi; -\lambda_{\mathrm{post}}]\|^2 \geq \|\lambda_{\mathrm{post}}\|^2 = 1/T_{\mathrm{post}}$. For the upper bound, observe that because $\varepsilon_{iT} - \varepsilon_{i:}\psi$ is the orthogonal projection of ε_{iT} on a subspace, specifically the subspace orthogonal to $\{\varepsilon_{i:} v : v \in \mathbb{R}^{T_{\mathrm{pre}}}\}$, its variance is bounded by that of ε_{iT}. This is $[0; \lambda_{\mathrm{post}}]\ \Sigma\ [0; \lambda_{\mathrm{post}}] \leq \sigma_{\max}(\Sigma) \|\lambda_{\mathrm{post}}\|^2 = \sigma_{\max}(\Sigma)/T_{\mathrm{post}}$.

VII.3 Theorem 1 in Condensed Form

In the abstract setting we've introduced above, we can write a weighted difference-in-differences treatment effect estimator as the difference between our (aggregate) treated observation Y_{NT} and an estimate $\hat{Y}_{NT}$ of the corresponding (aggregate) control potential outcome. In the concrete setting considered in Section III, this coincides with the estimator defined in (16).

$$\hat{\tau}(\lambda, \omega) = Y_{NT} - \hat{Y}_{NT}(\lambda, \omega) \quad \text{where} \quad \hat{Y}_{NT}(\lambda, \omega) := Y_{N:}\lambda + \omega' Y_{:T} - \omega' Y_{::}\lambda. \tag{35}$$

And the following weights coincide with the definitions used in Section III.

$$\begin{aligned}
\hat{\omega}_0, \hat{\omega} &= \underset{\omega_0, \omega \in \mathbb{R} \times \Omega}{\arg\min} \|\omega_0 + \omega' Y_{::} - Y_{N:}\|^2 + \zeta^2 T_0 \|\omega\|^2, \\
\tilde{\omega}_0, \tilde{\omega} &= \underset{\omega_0, \omega \in \mathbb{R} \times \Omega}{\arg\min} \|\omega_0 + \omega' L_{::} - L_{N:}\|^2 + (\zeta^2 + \sigma^2) T_0 \|\omega\|^2, \\
\hat{\lambda}_0, \hat{\lambda} &= \underset{\lambda_0, \lambda \in \mathbb{R} \times \Lambda}{\arg\min} \|\lambda_0 + Y_{::}\lambda - Y_{:T}\|^2, \\
\tilde{\lambda}_0, \tilde{\lambda} &= \underset{\lambda_0, \lambda \in \mathbb{R} \times \Lambda}{\arg\min} \|\lambda_0 + L_{::}\lambda - L_{:T}\|^2 + N_0 \|\Sigma_{::}^{1/2}(\lambda - \psi)\|^2.
\end{aligned} \tag{36}$$

The following assumptions on the condensed form hold in the setting considered in Theorem 1. The first summarizes our condensed-form model. The second is implied by Assumption 1 for $N_1 = N_{\text{tr}}$ and $T_1 \sim T_{\text{post}}$ as described above in Section VII.2. And the remaining three are condensed-form restatements of Assumptions 2-4, differing only in that we substitute $T_1 \sim T_{\text{post}}$ for T_{post} itself.

Assumption 5 (Model)**.** *We observe $Y_{it} = L_{it} + 1(i = N, t = T)\tau + \varepsilon_{it}$ for deterministic $\tau \in \mathbb{R}$ and $L \in \mathbb{R}^{N\times T}$ and random $\varepsilon \in \mathbb{R}^{N\times T}$. And we define $N_0 = N - 1$ and $T = T_0 - 1$.*

Assumption 6 (Properties of Errors)**.** *The rows $\varepsilon_{i\cdot}$ of the noise matrix are independent gaussian vectors with mean zero and covariance matrix Σ for $i \le N_0$ and $N_1^{-1}\Sigma$ for $i = N$ where the eigenvalues of $\Sigma_{::}$ are bounded and bounded away from zero. Here $N_1 > 0$ can be arbitrary and we define $T_1 = 1/\operatorname{Var}[\varepsilon_{i:}\psi - \varepsilon_{iT}]$ and $\psi = \arg\min_{v\in\mathbb{R}^{T_0}} E(\varepsilon_{i:}v - \varepsilon_{iT})^2$.*

Assumption 7 (Sample Sizes)**.** *We consider a sequence of problems where T_0/N_0 is bounded and bounded away from zero, T_1 and N_1 are bounded away from zero, and $N_0/(N_1T_1\max(N_1, T_1)\log^2(N_0)) \to \infty$.*

Assumption 8 (Properties of L)**.** *For the largest integer $K \le \sqrt{\min(T_0, N_0)}$,*

$$\sigma_K(L_{::})/K \ll \min(N_1^{-1/2}\log^{-1/2}(N_0), T_1^{-1/2}\log^{-1/2}(T_0)).$$

Assumption 9 (Properties of Oracle Weights)**.** *We use weights as in* (36) *for $\zeta \gg (N_1T_1)^{1/4}\log^{1/2}(N_0)$ and the oracle weights satisfy*

$$\begin{aligned}
(i)\ & \max(\|\tilde\omega\|, \|\tilde\lambda - \psi\|) \ll (N_1T_1)^{-1/2}\log^{-1/2}(N_0),\\
(ii.\omega)\ & \|\tilde\omega_0 + \tilde\omega' L_{::} - L_{N:}\| \ll N_0^{1/4}(N_1T_1\max(N_1, T_1))^{-1/4}\log^{-1/2}(N_0),\\
(ii.\lambda)\ & \|\tilde\lambda_0 + L_{::}\tilde\lambda - L_{:T}\| \ll N_0^{1/4}(N_1T_1)^{-1/8},\\
(iii)\ & L_{NT} - \tilde\omega' L_{:T} - L_{N:}\tilde\lambda + \tilde\omega' L_{::}\tilde\lambda \ll (N_1T_1)^{-1/2}.
\end{aligned}$$

The following condensed form asymptotic linearity result implies Theorem 1.

Theorem 3. *If Assumptions 5-9 hold, then $\hat\tau(\hat\lambda, \hat\omega) - \tau = \varepsilon_{NT} - \varepsilon_{N:}\psi + o_p((N_1T_1)^{-1/2})$.*

The following lemma reduces its proof to demonstrating the negligibility of the difference $\Delta_{oracle} := \hat{\tau}(\hat{\omega}, \hat{\lambda}) - \hat{\tau}(\tilde{\omega}, \tilde{\lambda})$ between the SDID estimator and the corresponding oracle estimator. Its proof is a straightforward calculation. Note that the bounds it requires on the oracle weights are looser than what is required by Assumption 9(i); those tighter bounds are used to control Δ_{oracle}.

Lemma 4. *If deterministic $\tilde{\omega}, \tilde{\lambda}$ satisfy $\|\tilde{\omega}\| = o(N_1^{-1/2})$ and $\|\tilde{\lambda} - \psi\| = o(T_1^{-1/2})$ and Assumptions 5, 6, and 9(iii) hold, then $\hat{\tau}(\tilde{\omega}, \tilde{\lambda}) - \tau = \varepsilon_{NT} - \varepsilon_{N:}\psi + o_p((N_1 T_1)^{-1/2})$.*

To show that this difference Δ_{oracle} is small, we use bounds on the difference between the estimated and oracle weights based on Hirshberg (2021, Theorem 1). We summarize these bounds in Lemma 5 below.

Lemma 5. *If Assumptions 5, 6, and 8 hold; T_1 and N_1 are bounded away from zero; $N_0, T_0 \to \infty$ with $N_0 \geq \log^2(T_0)$ and $T_0 \geq \log^2(N_0)$; and we choose weights as in* (36) *for unit simplices $\Omega \subseteq \mathbb{R}^{N_0}$ and $\Lambda \subseteq \mathbb{R}^{T_0}$, then the following bounds hold on an event of probability $1 - c\exp(-c\min(N_0^{1/2}, T_0^{1/2}, N_0/\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|, T_0/\|\tilde{\omega}' L_{::} + \tilde{\omega}_0 - L_{N:}\|))$:*

$$\|\hat{\lambda}_0 - \tilde{\lambda}_0 + L_{::}(\hat{\lambda} - \tilde{\lambda})\| \leq cvr_\lambda, \qquad \|\hat{\lambda} - \tilde{\lambda}\| \leq cvN_0^{-1/2} r_\lambda,$$
$$\|\hat{\omega}_0 - \tilde{\omega}_0 + L'_{::}(\hat{\omega} - \tilde{\omega})\| \leq cvr_\omega, \qquad \|\hat{\omega} - \tilde{\omega}\| \leq cv(\eta^2 T_0)^{-1/2} r_\omega$$

for $\eta^2 = \zeta^2 + 1$, some universal constant c, and

$$r_\lambda^2 = (N_0/T_{eff})^{1/2}\sqrt{\log(T_0)} + \|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|\sqrt{\log(T_0)}, \qquad T_{eff}^{-1/2} = \|\tilde{\lambda} - \psi\| + T_1^{-1/2}$$
$$r_\omega^2 = (T_0/N_{eff})^{1/2}\sqrt{\log(N_0)} + \|L'_{::}\tilde{\omega} + \tilde{\omega}_0 - L'_{N:}\|\sqrt{\log(N_0)}, \qquad N_{eff}^{-1/2} = \|\tilde{\omega}\| + N_1^{-1/2}.$$

When Assumptions 7 and 9(i-ii) hold as well, these bounds hold with probability $1 - c\exp(-cN_0^{1/2})$, as together those assumptions they imply the lemma's conditions on N_0, T_0, N_1, T_1 and that $N_0/\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\| \gg N_0^{3/4}$ and $T_0/\|\tilde{\omega}' L_{::} + \tilde{\omega}_0 - L_{N:}\| \gg N_0^{3/4}$.

We conclude by using bounds of this form, in conjunction with the first order orthogonality of the weighted difference-in-differences estimator $\hat{\tau}(\lambda, \omega)$ to the weights λ and ω, to control Δ_{oracle}. We do this abstractly in Lemma 6, then derive from it a simplified bound from which it will be clear that $\Delta_{oracle} = o_p((N_1 T_1)^{-1/2})$ under our assumptions.

Lemma 6. *In the setting described in Section VII.1, let* $\Lambda \subseteq \mathbb{R}^{T_0}$ *and* $\Omega \subseteq \mathbb{R}^{N_0}$ *be sets with the property that* $\sum_{t \le T_0} \lambda_t = \sum_{i \le N_0} \omega_i = 1$ *for all* $\lambda \in \Lambda$ *and* $\omega \in \Omega$*. Let* $\hat{\lambda}_0, \hat{\lambda} \in \mathbb{R} \times \Lambda$ *and* $\hat{\omega}_0, \hat{\omega} \in \mathbb{R} \times \Omega$ *be random and* $\tilde{\lambda}_0, \tilde{\lambda} \in \mathbb{R} \times \Lambda$ *and* $\tilde{\lambda}_0, \tilde{\lambda} \in \mathbb{R} \times \Omega$ *be deterministic. On the intersection of an event of probability* $1 - c\exp(-u^2)$ *and one on which*

$$
\begin{aligned}
\sigma\|\omega - \tilde{\omega}\| &\le s_\lambda \quad \text{and} \quad \|\hat{\omega}_0 - \tilde{\omega}_0 + (\hat{\omega} - \tilde{\omega})' L_{::}\| \le r_\omega, \\
\|\Sigma_{::}^{1/2}(\hat{\lambda} - \tilde{\lambda})\| &\le s_\omega \quad \text{and} \quad \|\hat{\lambda}_0 - \tilde{\lambda}_0 + L_{::}(\hat{\lambda} - \tilde{\lambda})\| \le r_\lambda,
\end{aligned} \tag{37}
$$

the corresponding treatment effect estimators defined in (35) *are close in the sense that*

$$
\begin{aligned}
|\hat{\tau}(\hat{\lambda}, \hat{\omega}) - \hat{\tau}(\tilde{\lambda}, \tilde{\omega})| \le\ & cuK[N_{eff}^{-1/2} s_\lambda + T_{eff}^{-1/2} s_\omega + \sigma^{-1} s_\omega s_\lambda] \\
&+ cK[(\|\tilde{\omega}\| + \sigma^{-1} s_\omega)\, \mathrm{w}(\Sigma_{::}^{1/2} \Lambda^\star_{s_\lambda}) + (\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\| + s_\lambda)\, \mathrm{w}(\Omega^\star_{s_\omega})] \\
&+ \sigma^{-1} s_\omega \min_{\lambda_0 \in \mathbb{R}} \|S_\lambda^{1/2}(L_{::}\tilde{\lambda} + \lambda_0 - L_{:T})\| + s_\lambda \min_{\omega_0 \in \mathbb{R}} \|S_\omega^{1/2} \Sigma_{::}^{-1/2}(L_{::}'\tilde{\omega} + \omega_0 - L_{N:}')\| \\
&+ \min\left(\|\Sigma_{::}^{-1/2}\| r_\omega s_\lambda,\ \sigma^{-1} s_\omega r_\lambda,\ \min_{k \in \mathbb{N}} \sigma_k(L_{::}^c)^{-1} r_\lambda r_\omega + \sigma^{-1} \|\Sigma_{::}^{-1/2}\| \sigma_{k+1}(L_{::}^c) s_\lambda s_\omega \right)
\end{aligned}
$$

Here c *is a universal constant,* $\mathrm{w}(S)$ *is the gaussian width of the set* S*, and*

$$
\begin{aligned}
T_{eff}^{-1/2} &= \sigma^{-1}(\|\Sigma_{::}^{1/2}(\tilde{\lambda} - \psi)\| + \|\tilde{\varepsilon}_{iT}\|_{L_2}), & N_{eff}^{-1/2} &= \|\tilde{\omega}\| + \|(\Sigma_{::}^N)^{1/2} \Sigma_{::}^{-1/2}\|, \\
\Lambda_s^\star &= \{\lambda - \tilde{\lambda} : \lambda \in \Lambda^\star, \|\Sigma_{::}^{1/2}(\lambda - \tilde{\lambda})\| \le s\}, & \Omega_s^\star &= \{\omega - \tilde{\omega} : \omega \in \Omega^\star, \sigma\|\omega - \tilde{\omega}\| \le s\}, \\
S_\lambda &= I - L_{::}(L_{::}' L_{::} + (\sigma r_\omega / s_\omega)^2 I)^{-1} L_{::}', & S_\omega &= I - \Sigma_{::}^{-1/2} L_{::}'(L_{::} \Sigma_{::}^{-1} L_{::}' + (r_\lambda / s_\lambda)^2 I)^{-1} L_{::} \Sigma_{::}^{-1/2}, \\
L_{::}^c &= L_{::} - N_0^{-1} 1_{N_0} 1_{N_0}' L_{::} - L_{::} T_0^{-1} 1_{T_0} 1_{T_0}'.
\end{aligned}
$$

We simplify this using bounds $s_\omega, s_\lambda, r_\omega, r_\lambda$ from Lemma 5 and bounds $\mathrm{w}(\Omega^\star_{s_\omega}) \lesssim \sqrt{\log(N_0)}$ and $\mathrm{w}(\Lambda^\star_{s_\lambda}) \lesssim \sqrt{\log(T_0)}$ that hold for the specific sets Ω, Λ used in our concrete setting (Hirshberg, 2021, Example 1).

Corollary 7. *Suppose Assumptions 5, 6, and 8 hold with* $T_0 \sim N_0$ *and that* $\log(N_0)$*,* T_1 *and* N_1 *are bounded away from zero. Let* $m_0 = N_0$*,* $m_1 = \sqrt{N_1 T_1}$*, and* $\bar{m}_1 = \max(N_1, T_1)$*. Consider the weights defined in* (36) *with* $\Omega \subseteq \mathbb{R}^{N_0}$ *and* $\Lambda \subseteq \mathbb{R}^{T_0}$ *taken to be the unit simplices and* $\zeta \gg m_1^{1/2} \log^{1/2}(m_0)$*. With probability* $1 - 2\exp(-\min(T_1 \log(T_0), N_1 \log(N_0))) - c\exp(-cN_0^{1/2}))$,

$\hat{\tau}(\hat{\omega}, \hat{\lambda}) - \hat{\tau}(\tilde{\lambda}, \tilde{\omega}) = o_p((N_1 T_1)^{-1/2})$ *if*

$$\max(\|\tilde{\omega}\|, \|\psi - \tilde{\lambda}\|) \ll m_1^{-1} \log^{-1/2}(m_0),$$
$$\|\tilde{\omega}_0 + \tilde{\omega}' L_{::} - L_{N:}\| \ll m_0^{1/4} m_1^{-1/2} \bar{m}_1^{-1/4} \log^{-1/2}(m_0),$$
$$\|\tilde{\lambda}_0 + L_{::}\tilde{\lambda} - L_{:T}\| \ll m_0^{1/4} m_1^{-1/4},$$

and the latter two bounds go to infinity.

These assumptions are implied by Assumptions 5-9. Assumption 7 states our assumptions $T_0 \sim N_0$, $\log(N_0), T_1, N_1 \not\to 0$, and that the (fourth power of) the second bound above goes to infinity; when the second bound does go to infinity, so does the third. As Assumption 7 implies that that $T_0 \sim N_0 \to \infty$, it implies the probability stated in the lemma above goes to one. And Assumption 9(i-ii) states that the bound above hold.

As our assumptions imply the conclusions of Lemma 4 and Corollary 7, and those two results imply the conclusions of Theorem 3, this concludes our proof.

VIII Proof Details

In this section, we complete our proof by proving the lemmas used in the sketch above.

VIII.1 Proof of Lemma 4

First, consider the oracle estimator's bias,

$$\mathrm{E}\,\hat{\tau}(\tilde{\lambda}, \tilde{\omega}) - \tau = (L_{NT} + \tau) - \tilde{\omega}' L_{:T} - L_{N:}\tilde{\lambda} + \tilde{\omega}' L_{::}\tilde{\lambda} - \tau.$$

Assumption 9(iii) is that this is $o_p((N_1 T_1)^{-1/2})$.

Now consider the oracle estimator's variation around its mean,

$$\begin{aligned}
\hat{\tau}(\tilde{\lambda}, \tilde{\omega}) - \mathrm{E}\,\hat{\tau}(\tilde{\lambda}, \tilde{\omega}) &= \varepsilon_{NT} - \varepsilon_{N:}\tilde{\lambda} + \tilde{\omega}'\varepsilon_{:T} + \tilde{\omega}'\varepsilon_{::}\tilde{\lambda} \\
&= (\varepsilon_{NT} - \varepsilon_{N:}\tilde{\lambda}) - \tilde{\omega}'(\varepsilon_{:T} - \varepsilon_{::}\tilde{\lambda}) \\
&= (\varepsilon_{NT} - \varepsilon_{N:}\psi) - \tilde{\omega}'(\varepsilon_{:T} - \varepsilon_{::}\psi) - \varepsilon_{N:}(\tilde{\lambda} - \psi) + \tilde{\omega}'\varepsilon_{::}(\tilde{\lambda} - \psi).
\end{aligned}$$

The conclusion of our lemma holds if all but the first term in the decomposition above are $o_p((N_1T_1)^{-1/2})$. We do this by showing that each term has $o((N_1T_1)^{-1})$ variance.

$$
\begin{aligned}
&\mathrm{E}(\tilde{\omega}'(\varepsilon_{:T}-\varepsilon_{::}\psi))^2 = \|\tilde{\omega}\|^2\,\mathrm{E}(\varepsilon_{1T}-\varepsilon_{i:}\psi)^2 = \|\tilde{\omega}\|^2/T_1,\\
&\mathrm{E}(\varepsilon_{N:}(\tilde{\lambda}-\psi))^2 = (\tilde{\lambda}-\psi)'(\mathrm{E}\,\varepsilon_{N:}'\varepsilon_{N:})(\tilde{\lambda}-\psi) \le \|\tilde{\lambda}-\psi\|^2\|\Sigma_{::}\|/N_1,\\
&\mathrm{E}(\tilde{\omega}'\varepsilon_{::}(\tilde{\lambda}-\psi))^2 = \|\tilde{\omega}\|^2\,\mathrm{E}(\varepsilon_{1:}(\tilde{\lambda}-\psi))^2 \le \|\tilde{\omega}\|^2\|\tilde{\lambda}\|^2\|\Sigma_{::}\|.
\end{aligned}
$$

Our assumption that $\|\Sigma_{::}\|$ is bounded and our assumed bounds on $\|\tilde{\omega}\|$ and $\|\tilde{\lambda}\|$ imply that each of these is $o((N_1T_1)^{-1})$ as required.

VIII.2 Proof of Lemma 5

The bounds involving λ follow from the application of Hirshberg (2021, Theorem 1) with $\eta^2=1$, $A=L_{::}$, $b=L_{:T}$, and $[\varepsilon,\nu]=[\varepsilon_{::},\varepsilon_{:T}]$ with independent rows, using the bound $\mathrm{w}(\Lambda_s^\star)\lesssim\sqrt{\log(T_0)}$ mentioned in its Example 1. The bounds for ω follow from the application of the same theorem with $\eta^2=1+\zeta^2/\sigma^2$ for $\sigma^2=\mathrm{tr}(\Sigma_{::})/T_0$, $A=L_{::}'$, $b=L_{N:}'$, and $[\varepsilon,\nu]=\varepsilon_{::}',\varepsilon_{N:}']$ with independent columns, using the analogous bound $\mathrm{w}(\Omega_s^\star)\lesssim\sqrt{\log(N_0)}$.

In the first case, Hirshberg (2021, Theorem 1) gives bounds of the claimed form for

$$
\begin{aligned}
&r_\lambda^2 = [(N_0/T_{eff})^{1/2} + \|L_{::}\tilde{\lambda}+\tilde{\lambda}_0 - L_{:T}\|]\sqrt{\log(T_0)}+1 \quad \text{holding with probability}\\
&1-c\exp\left(-c\min(N_0\log(T_0)/r_\lambda^2, v^2R, N_0)\right) \text{ if } \sigma_{R+1}(L_{::})/R \le cvT_1^{-1/2}\log^{-1/2}(T_0) \quad \text{and}\\
&R\le\min(v^2(N_0T_{eff})^{1/2}, v^2N_0/\log(T_0), cN_0).
\end{aligned}
$$

To see this, ignore constant order factors of ϕ (≥ 1) and $\|\Sigma\|$ in Hirshberg (2021, Theorem 1) and substitute $s^2=cv^2r_\lambda^2/(\eta^2 n)$ for problem-appropriate parameters $\eta^2=1$, $n=N_0$, $n_{eff}^{-1/2}=T_{eff}^{-1/2}$ $(\ge T_1^{-1/2})$, and $\bar{\mathrm{w}}(\Theta_s)=\sqrt{\log(T_0)}$.

In the second case, Hirshberg (2021, Theorem 1) gives bounds of the claimed form for

$$
\begin{aligned}
&r_\omega^2 = [(T_0/N_{eff})^{1/2} + \|\tilde{\omega}'L_{::}+\tilde{\omega}_0 - L_{N:}\|]\sqrt{\log(N_0)}+\log(N_0) \quad \text{holding with probability}\\
&1-c\exp\left(-c\min(\eta^2T_0\log(N_0)/r_\omega^2, v^2R, T_0)\right) \text{ if } \sigma_{R+1}(L_{::})/R \le cvN_1^{-1/2}\log^{-1/2}(N_0) \quad \text{and}\\
&R\le\min(v^2(T_0N_{eff})^{1/2}, v^2\eta^2T_0/\log(N_0), cT_0).
\end{aligned}
$$

To see this, ignore constant order factors of ϕ (≥ 1) and $\|\Sigma\|$ in Hirshberg (2021, Theorem 1) and substitute $s^2 = cv^2 r_\lambda^2/(\eta^2 n)$ for problem-appropriate parameters $\eta^2 = 1 + \zeta^2/\sigma^2$, $n = T_0$, $n_{eff}^{-1/2} = N_{eff}^{-1/2}$ ($\geq N_1^{-1/2}$), and $\bar{\mathrm{w}}(\Theta_s) = \sqrt{\log(N_0)}$.

We will now simplify our conditions on R. As we have assumed that N_1 and T_1 and therefore N_{eff} and T_{eff} are bounded away from zero, we can choose v of constant order with $v \geq \max(c/T_{eff}, c/N_{eff}, 1)$, so our upper bounds on R simplify to

$$R \leq \min(N_0^{1/2}, N_0/\log(T_0), cN_0) \quad \text{and} \quad R \leq \min(T_0^{1/2}, \eta^2 T_0/\log(N_0), T_0)$$

respectively. Having assumed that that $N_0, T_0 \to \infty$ with $N_0 \geq \log^2(T_0)$ and $T_0 \geq \log^2(N_0)$, these conditions simplify to $R \leq N_0^{1/2}$ and $R \leq T_0^{1/2}$. Thus, it suffices that the largest integer $R \leq \min(N_0, T_0)^{1/2}$ satisfy $\sigma_{R+1}(L_{::})/R \leq c \min(N_1^{-1/2} \log^{-1/2}(N_0), T_1^{-1/2} \log^{-1/2}(T_0))$. This is implied, for any constant c, by Assumption 8.

We conclude by simplifing our probability statements. As noted above, we take $R \sim \min(N_0, T_0)^{1/2}$, so we may make this substitution. Furthermore, again using our assumption that N_{eff} and T_{eff} are bounded away from zero,

$$\begin{aligned}
\frac{N_0 \log(T_0)}{r_\lambda^2} &\gtrsim \min\left(\frac{N_0 \log(T_0)}{(N_0/T_{eff})^{1/2}\sqrt{\log(T_0)}},\ \frac{N_0 \log(T_0)}{\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|\sqrt{\log(T_0)}}, \frac{N_0 \log(T_0)}{1} \right) \\
&\gtrsim \min\left(\sqrt{N_0},\ N_0/\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\| \right), \\
\frac{T_0 \log(N_0)}{r_\omega^2} &\gtrsim \min\left(\frac{T_0 \log(N_0)}{(T_0/N_{eff})^{1/2}\sqrt{\log(N_0)}},\ \frac{T_0 \log(N_0)}{\|\tilde{\omega}' L_{::} + \tilde{\omega}_0 - L_{N:}\|\sqrt{\log(N_0)}},\ \frac{T_0 \log(N_0)}{\log(N_0)} \right) \\
&\gtrsim \min\left(\sqrt{T_0},\ T_0/\|\tilde{\omega}' L_{::} + \tilde{\omega}_0 - L_{N:}\| \right).
\end{aligned}$$

Thus, each bound holds with probability at least $1 - c\exp(-c \min(N_0^{1/2}, T_0^{1/2}, N_0/\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|, T_0/\|\tilde{\omega}' L_{::} + \tilde{\omega}_0 - L_{N:}\|))$. And by the union bound, doubling our leading constant c, both simultaneously with such a probability.

VIII.3 Proof of Lemma 6

We begin with a decomposition of the difference between the SDID estimator and the oracle.

$$
\begin{aligned}
&\tau(\tilde{\lambda}, \tilde{\omega}) - \hat{\tau}(\hat{\lambda}, \hat{\omega}) \\
&= \hat{Y}_{NT}(\hat{\lambda}, \hat{\omega}) - Y_{NT}(\tilde{\lambda}, \tilde{\omega}) \\
&= \Big[Y_{N:}\hat{\lambda} + \hat{\omega}' Y_{:T} - \hat{\omega}' Y_{::}\hat{\lambda}\Big] - \Big[Y_{N:}\tilde{\lambda} + \tilde{\omega}' Y_{:T} - \tilde{\omega}' Y_{::}\tilde{\lambda}\Big] \\
&= Y_{N:}(\hat{\lambda} - \tilde{\lambda}) + (\hat{\omega} - \tilde{\omega})' Y_{:T} - \Big[(\hat{\omega} - \tilde{\omega})' Y_{::}(\hat{\lambda} - \tilde{\lambda}) + \tilde{\omega}' Y_{::}(\hat{\lambda} - \tilde{\lambda}) + (\hat{\omega} - \tilde{\omega})' Y_{::}\tilde{\lambda}\Big] \\
&= (Y_{N:} - \tilde{\omega}' Y_{::})(\hat{\lambda} - \tilde{\lambda}) + (\hat{\omega} - \tilde{\omega})'(Y_{:T} - Y_{::}\tilde{\lambda}) - (\hat{\omega} - \tilde{\omega})' Y_{::}(\hat{\lambda} - \tilde{\lambda}).
\end{aligned}
$$

We bound these terms. As $Y_{it} = L_{it} + 1(i = N, t = T)\tau + \varepsilon$, we can decompose each of these three terms into two parts, one involving L and the other ε. We will begin by treating the parts involving ε.

1. The first term is a sum $\varepsilon_{N:}(\hat{\lambda} - \tilde{\lambda}) - \tilde{\omega}'\varepsilon_{::}(\hat{\lambda} - \tilde{\lambda})$. Because $\hat{\lambda}$ is independent of $\varepsilon_{N:}$, the first of these is subgaussian conditional on $\hat{\lambda}$, with conditional subgaussian norm $\|\varepsilon_{N:}(\hat{\lambda} - \tilde{\lambda})\|_{\psi_2|\hat{\lambda}} \leq \|\varepsilon_{N:}(\Sigma_{::}^{N})^{-1/2}\|_{\psi_2}\|(\Sigma_{::}^{N})^{1/2}\Sigma_{::}^{-1/2}\|\|\Sigma_{::}^{1/2}(\hat{\lambda} - \tilde{\lambda})\|$. It follows that it satisfies a subgaussian tail bound $|\varepsilon_{N:}(\hat{\lambda} - \tilde{\lambda})| \leq cu\|\varepsilon_{N:}(\Sigma_{::}^{N})^{-1/2}\|_{\psi_2}\ \|(\Sigma_{::}^{N})^{1/2}\Sigma_{::}^{-1/2}\|\|\Sigma_{::}^{1/2}(\hat{\lambda} - \tilde{\lambda})\|$ with conditional probability $1 - 2\exp(-u^2)$. This implies that the same bound holds unconditionally on an event of probability $1 - 2\exp(-u^2)$.

   Furthermore, via generic chaining (e.g., Vershynin, 2018, Theorem 8.5.5), on an event of probability $1 - 2\exp(-u^2)$, either $\Sigma_{::}^{1/2}(\hat{\lambda} - \tilde{\lambda}) \notin \Lambda_{s_\lambda}^{\star}$ or $|\tilde{\omega}'\varepsilon_{::}(\hat{\lambda} - \tilde{\lambda})| \leq c\|\tilde{\omega}'\varepsilon_{::}\Sigma_{::}^{-1/2}\|_{\psi_2}(\mathrm{w}(\Sigma_{::}^{1/2}\Lambda_{s_\lambda}^{\star}) + u\,\mathrm{rad}(\Sigma_{::}^{1/2}\Lambda_{s_\lambda}^{\star})) \leq c\|\varepsilon_{i:}\Sigma_{::}^{-1/2}\|_{\psi_2}\|\tilde{\omega}\|(\mathrm{w}(\Sigma_{::}^{1/2}\Lambda_{s_\lambda}^{\star}) + us_\lambda)$. The second comparison here follows from Hoeffding's inequality (e.g., Vershynin, 2018, Theorem 2.6.3). Thus, by the union bound, on the intersection of an event of probability $1 - c\exp(-u^2)$ and one on which (37) holds,

$$
\begin{aligned}
&|(\varepsilon_{N:} - \tilde{\omega}'\varepsilon_{::})(\hat{\lambda} - \tilde{\lambda})| \\
&\leq cu\|\varepsilon_{N:}(\Sigma_{::}^{N})^{-1/2}\|_{\psi_2}\|(\Sigma_{::}^{N})^{1/2}\Sigma_{::}^{-1/2}\|s_\lambda + c\|\varepsilon_{1:}\Sigma^{-1/2}\|_{\psi_2}\|\tilde{\omega}\|(\mathrm{w}(\Sigma_{::}^{1/2}\Lambda_{s_\lambda}^{\star}) + us_\lambda) \\
&\leq cuKN_{eff}^{-1/2}s_\lambda + cK\|\tilde{\omega}\|\,\mathrm{w}(\Sigma_{::}^{1/2}\Lambda_{s_\lambda}^{\star}).
\end{aligned}
$$

2. The second term is similar to the first. It is a sum $(\hat{\omega} - \tilde{\omega})'\tilde{\varepsilon}_{:T} + (\hat{\omega} - \tilde{\omega})'\varepsilon_{::}(\psi - \tilde{\lambda})$ for

$\tilde{\varepsilon}_{:T} = \varepsilon_{:T} - \varepsilon_{::}\psi$. Because $\hat{\omega}$ is a function of $\varepsilon_{::}, \varepsilon_{N:}$ and $\tilde{\varepsilon}_{:T}$ is mean zero conditional on them, the first of these terms is a weighted average of conditionally independent mean-zero subgaussian random variables. Applying Hoeffding's inequality conditionally, it follows that its magnitude is bounded by $cu\|\hat{\omega} - \tilde{\omega}\| \max_{i<N} \|\tilde{\varepsilon}_{iT}\|_{\psi_2|\varepsilon_{::},\varepsilon_{N:}} \leq cuK\|\hat{\omega} - \tilde{\omega}\|\|\tilde{\varepsilon}_{1T}\|_{L_2}$ on an event of probability $1 - 2\exp(-u^2)$. In the second comparison, we've used the independence of rows $\varepsilon_{i\cdot}$, the identical distribution of rows for $i < N$, and the assumption that $\|\tilde{\varepsilon}_{1T}\|_{\psi_2|\varepsilon_{1:}} \leq K\|\tilde{\varepsilon}_{1T}\|_{L_2}$.

Furthermore, via generic chaining, on an event of probability $1 - c\exp(-u^2)$, either $(\hat{\omega} - \tilde{\omega}) \notin \Omega^\star_{s_\omega}$ or $|(\hat{\omega} - \tilde{\omega})\varepsilon_{::}(\psi - \tilde{\lambda})| \leq c\|\varepsilon_{::}(\psi - \tilde{\lambda})\|_{\psi_2}(\mathrm{w}(\Omega^\star_{s_\omega}) + u\,\mathrm{rad}(\Omega^\star_{s_\omega})) \leq cK\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\|(\mathrm{w}(\Omega^\star_{s_\omega}) + u\,\mathrm{rad}(\Omega^\star_{s_\omega}))$. The second comparison here follows from Hoeffding's inequality. Thus, by the union bound, on the intersection of an event of probability $1 - c\exp(-u^2)$ and one on which (37) holds,

$$
\begin{aligned}
&|(\hat{\omega} - \tilde{\omega})'(\varepsilon_{:T} - \varepsilon_{::}\tilde{\lambda})| \\
&\leq cuK\|\tilde{\varepsilon}_{1T}\|_{L_2}\sigma^{-1}s_\omega + cuK\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\|\sigma^{-1}s_\omega + cK\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\|\,\mathrm{w}(\Omega^\star_{s_\omega}) \\
&\leq cuKT_{eff}^{-1/2}s_\omega + cK\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\|\,\mathrm{w}(\Omega^\star_{s_\omega}).
\end{aligned}
$$

3. Via Chevet's inequality (Hirshberg, 2021, Lemma 3), on an event of probability $1 - c\exp(-u^2)$, either $(\hat{\omega}-\tilde{\omega}) \notin \Omega^\star_{s_\omega}$, $(\hat{\lambda}-\tilde{\lambda}) \notin \Lambda^\star_{s_\lambda}$, or $|(\hat{\omega}-\tilde{\omega})'\varepsilon_{::}(\hat{\lambda}-\tilde{\lambda})| \leq cK[\mathrm{w}(\Omega^\star_{s_\omega})\,\mathrm{rad}(\Sigma_{::}^{1/2}\Lambda^\star_{s_\lambda}) + \mathrm{rad}(\Omega^\star_{s_\omega})\,\mathrm{w}(\Sigma_{::}^{1/2}\Lambda^\star_{s_\lambda}) + u\,\mathrm{rad}(\Omega^\star_{s_\omega})\,\mathrm{rad}(\Sigma_{::}^{1/2}\Lambda^\star_{s_\lambda})] \leq cK[\mathrm{w}(\Omega^\star_{s_\omega})s_\lambda + \mathrm{w}(\Sigma_{::}^{1/2}\Lambda^\star_{s_\lambda})\sigma^{-1}s_\omega + u\sigma^{-1}s_\omega s_\lambda]$. On the intersection of this event and one on which (37) holds, the first two possibilities are ruled out and our bound on $|(\hat{\omega} - \tilde{\omega})'\varepsilon_{::}(\hat{\lambda} - \tilde{\lambda})|$ holds.

By the union bound, these three bounds are satisfied on the intersection of one of probability $1 - c\exp(-u^2)$ and one on which (37) holds. And by the triangle inequality, adding our bounds yields a bound on our terms involving ε.

$$
\begin{aligned}
&|(\varepsilon_{N:} - \tilde{\omega}'\varepsilon_{::})(\hat{\lambda} - \tilde{\lambda}) + (\hat{\omega} - \tilde{\omega})'(\varepsilon_{:T} - \varepsilon_{::}\tilde{\lambda}) - (\hat{\omega} - \tilde{\omega})'\varepsilon_{::}(\hat{\lambda} - \tilde{\lambda})| \\
&\leq cuK[N_{eff}^{-1/2}s_\lambda + \phi T_{eff}^{-1/2}s_\omega + \sigma^{-1}s_\omega s_\lambda] \\
&\quad + cK[(\|\tilde{\omega}\| + \sigma^{-1}s_\omega)\,\mathrm{w}(\Sigma_{::}^{1/2}\Lambda^\star_{s_\lambda}) + (\|\Sigma_{::}^{1/2}(\psi - \tilde{\lambda})\| + s_\lambda)\,\mathrm{w}(\Omega^\star_{s_\omega})]
\end{aligned} \tag{38}
$$

We now turn our attention to the terms involving L. For any $\omega_0, \omega \in \mathbb{R} \times \mathbb{R}^{N_0}$, $(L_{N:} -$

$\tilde{\omega}' L_{::})(\hat{\lambda} - \tilde{\lambda}) = (L_{N:} - \omega' L_{::} - \omega_0)(\hat{\lambda} - \tilde{\lambda}) + (\omega - \tilde{\omega})' L_{::}(\hat{\lambda} - \tilde{\lambda})$. The value of the constant ω_0 does not affect the expression because the sum of the elements of $\hat{\lambda} - \tilde{\lambda}$ is zero. By the Cauchy-Schwarz and triangle inequalities, it follows that

$$|(L_{N:} - \tilde{\omega}' L_{::})(\hat{\lambda} - \tilde{\lambda})| \leq \|(L_{N:} - \omega' L_{::} - \omega_0)\Sigma_{::}^{-1/2}\| \|\Sigma_{::}^{1/2}(\hat{\lambda} - \tilde{\lambda})\| + \|\omega - \tilde{\omega}\| \|L_{::}(\hat{\lambda} - \tilde{\lambda})\|$$

Furthermore, substituting bounds implied by (37) and using the elementary bound $x + y \leq 2\sqrt{x^2 + y^2}$, we get a quantity that we can minimize explicitly over ω. The following result; for $A = \Sigma_{::}^{-1/2} L_{::}'$, $b = \Sigma_{::}^{-1/2}(L_{N:}' - \omega_0 1)$, $\alpha = s_\lambda$, and $\beta = r_\lambda$ satisfying $\beta/\alpha = cN_0^{1/2}$; implies the bound

$$\begin{aligned} |(L_{N:} - \tilde{\omega}' L_{::})(\hat{\lambda} - \tilde{\lambda})| &\leq 2 s_\lambda \min_{\omega_0} \|S_\omega^{1/2} \Sigma_{::}^{-1/2}(L_{::}' \tilde{\omega} + \omega_0 - L_{N:}')\| \\ S_\omega &= I - \Sigma_{::}^{-1/2} L_{::}' (L_{::} \Sigma_{::}^{-1} L_{::}' + (r_\lambda / s_\lambda)^2 I)^{-1} L_{::} \Sigma_{::}^{-1/2}. \end{aligned}$$

Lemma 8. *For any real matrix A and appropriately shaped vectors $\tilde{x}$ and b, $\min_x \alpha^2 \|Ax - b\|^2 + \beta^2 \|x - \tilde{x}\|^2 = \alpha^2 \|S^{1/2}(A\tilde{x} - b)\|^2$ for $S = I - A(A'A + (\beta/\alpha)^2 I)^{-1} A'$. If $\beta = 0$, the same holds for $S = I - A(A'A)^\dagger A$.*

Proof. Reparameterizing in terms of $y = x - \tilde{x}$ and defining $v = A\tilde{x} - b$ and $\lambda^2 = \beta^2/\alpha^2$, this is α^2 times $\min_y \|v + Ay\|^2 + \lambda^2 \|y\|^2 = \min_y \|v\|^2 + 2y'A'v + y'(A'A + \lambda^2 I)y$. Setting the derivative of the expression to zero, we solve for the minimizer $y = -(A'A + \lambda^2 I)^{-1} A'v$ and the minimum $v'[I - A(A'A + \lambda^2 I)^{-1} A']v$, then multiply by α^2. □

Analogously, for any $\lambda_0, \lambda \in \mathbb{R} \times \mathbb{R}^{T_0}$,

$$|(\hat{\omega} - \tilde{\omega})'(L_{:T} - L_{::}\tilde{\lambda})| \leq \|L_{:T} - L_{::}\lambda - \lambda_0\| \|\hat{\omega} - \tilde{\omega}\| + \|\lambda - \tilde{\lambda}\| \|(\hat{\omega} - \tilde{\omega})' L_{::}\|.$$

and therefore, when (37) holds,

$$\begin{aligned} |(\hat{\omega} - \tilde{\omega})'(L_{:T} - L_{::}\tilde{\lambda})| &\leq 2\sigma^{-1} s_\omega \min_{\lambda_0} \|S_\lambda^{1/2}(L_{::}\tilde{\lambda} - \lambda_0 - L_{:T})\| \\ S_\lambda &= I - L_{::}(L_{::}' L_{::} + (\sigma r_\omega / s_\omega)^2 I)^{-1} L_{::}'. \end{aligned}$$

Finally, we can take the minimum of two Cauchy-Schwarz bounds on the third term,

$$\begin{aligned}
|(\hat{\omega}-\tilde{\omega})' L_{::}(\hat{\lambda}-\tilde{\lambda})| &= |[(\hat{\omega}_0-\tilde{\omega}_0)+(\hat{\omega}-\tilde{\omega})' L_{::}](\hat{\lambda}-\tilde{\lambda})| \\
&\leq \|(\hat{\omega}_0-\tilde{\omega}_0)+(\hat{\omega}-\tilde{\omega})' L_{::}\| \|\Sigma_{::}^{-1/2}\| \|\Sigma_{::}^{1/2}(\hat{\lambda}-\tilde{\lambda})\|, \\
|(\hat{\omega}-\tilde{\omega})' L_{::}(\hat{\lambda}-\tilde{\lambda})| &= |(\hat{\omega}-\tilde{\omega})'[(\hat{\lambda}_0-\tilde{\lambda}_0)+L_{::}(\hat{\lambda}-\tilde{\lambda})]| \\
&\leq \|\hat{\omega}-\tilde{\omega}\| \|(\hat{\lambda}_0-\tilde{\lambda}_0)+L_{::}(\hat{\lambda}-\tilde{\lambda})\|.
\end{aligned}$$

As above, the inclusion of either intercept does not effect the value of the expression because $\hat{\lambda}-\tilde{\lambda}$ and $\hat{\omega}-\tilde{\omega}$ sum to one. This implies that on an event on which the bounds (37) hold,

$$\begin{aligned}
&|(L_{N:}-\tilde{\omega}' L_{::})(\hat{\lambda}-\tilde{\lambda})+(\hat{\omega}-\tilde{\omega})'(L_{:T}-L\tilde{\lambda})-(\hat{\omega}-\tilde{\omega})' L_{::}(\hat{\lambda}-\tilde{\lambda})| \\
&\leq 2 s_\lambda \min_{\omega_0} \|S_\omega^{1/2}\Sigma_{::}^{-1/2}(L_{::}'\tilde{\omega}+\omega_0-L_{N:}')\| + 2\sigma^{-1} s_\omega \min_{\lambda_0} \|S_\lambda^{1/2}(L_{::}\tilde{\lambda}-\lambda_0-L_{:T})\| \\
&\quad + \min\left(\|\Sigma_{::}^{-1/2}\| r_\omega s_\lambda, \sigma^{-1} s_\omega r_\lambda\right).
\end{aligned} \tag{39}$$

We can include in the minimum in the third term above another bound on $|(\hat{\omega}-\tilde{\omega})' L_{::}(\hat{\lambda}-\tilde{\lambda})|$. We will use one that exploits a potential gap in the spectrum of $L_{::}$, e.g., a bound on the smallest nonzero singular value of $L_{::}$. The abstract bound we will use is one on the inner product $x'Ay$: given bounds $\|x'A\| \leq r_x$, $\|Ay\| \leq r_y$, $\|x\| \leq s_x$, $\|y\| \leq s_y$, it is no larger than $\min_k \sigma_k(A)^{-1} r_x r_y + \sigma_{k+1}(A) s_x s_y$. To show this, we first observe that without loss of generality, we can let A be square, diagonal, and nonnegative with decreasing elemnts on the diagonal: in terms of its singular value decomposition $A = USV'$ and $x_U = U'x$ and $y_V = V'y$, $x'Ay = x_U' S y_V$ where $\|x_U' S\| \leq r_x$, $\|S y_V\| \leq r_y$, $\|x_U\| \leq s_x$, $\|y_V\| \leq s_y$. In this simplified diagonal case, letting

$a_i := A_{ii}$ and $R = \operatorname{rank}(A)$,

$$\begin{aligned}
|x'Ay| &= |\sum_{i=1}^{R} x_i y_i a_i| \\
&\le |\sum_{i=1}^{k} x_i y_i a_i| + |\sum_{i=k+1}^{R} x_i y_i a_i| \\
&\le \sqrt{\sum_{i=1}^{k} x_i^2 a_i^2 \sum_{i=1}^{k} y_i^2} + \sqrt{\sum_{i=k+1}^{R} x_i^2 a_i^2 \sum_{i=k+1}^{R} y_i^2} \\
&\le a_k^{-1} \sqrt{\sum_{i=1}^{k} x_i^2 a_i^2 \sum_{i=1}^{k} y_i^2 a_i^2} + a_{k+1} \sqrt{\sum_{i=k+1}^{R} x_i^2 \sum_{i=k+1}^{R} y_i^2} \\
&\le a_k^{-1} r_x r_y + a_{k+1} s_x s_y.
\end{aligned}$$

We apply this with $x = \hat{\omega} - \tilde{\omega}$, $y = \hat{\lambda} - \tilde{\lambda}$, and $A = L_{::} - N_0^{-1} 1_{N_0} 1'_{N_0} L_{::} - L_{::} T_0^{-1} 1_{T_0} 1'_{T_0}$; because $(\hat{\omega} - \tilde{\omega})' 1_{N_0} = 0$ and $1'_{T_0}(\hat{\lambda} - \tilde{\lambda}) = 0$, $(\hat{\omega} - \tilde{\omega})' L_{::} (\hat{\lambda} - \tilde{\lambda}) = (\hat{\omega} - \tilde{\omega})' A (\hat{\lambda} - \tilde{\lambda}) = x'Ay$. When the bounds in (37) hold, $\|x'A\| \le r_\omega$ and $\|Ay\| \le r_\lambda$, as

$$\|(\hat{\omega} - \tilde{\omega})' A\|^2 = \sum_{t=1}^{T_0} \left[(\hat{\omega} - \tilde{\omega})' L_{:t} - T_0^{-1} \sum_{t=1}^{T_0} (\hat{\omega} - \tilde{\omega})' L_{:t} \right]^2 = \min_{\delta \in \mathbb{R}} \|(\hat{\omega} - \tilde{\omega})' L_{::} - \delta\|^2 \le r_\omega^2.$$

These bounds also imply $\|x\| \le \sigma^{-1} s_\omega$ and $\|y\| \le \|\Sigma_{::}^{-1/2}\| s_\lambda$, so our third term is bounded by

$$|(\hat{\omega} - \tilde{\omega})' L_{::} (\hat{\lambda} - \tilde{\lambda})| \le \min_k \sigma_k(A)^{-1} r_\lambda r_\omega + \sigma^{-1} \|\Sigma_{::}^{-1/2}\| \sigma_{k+1}(A) s_\lambda s_\omega$$

Adding together (38) and (39), including this additional bound in the minimum in the third term of (39), we get the claimed bound on $|\tau(\tilde{\lambda}, \tilde{\omega}) - \hat{\tau}(\hat{\lambda}, \hat{\omega})|$.

VIII.4 Proof of Corollary 7

We begin with the bound from Lemma 6. As the claimed bound is stated up to an unspecified universal constant, we can ignore universal constants throughout. We can ignore K as well; as discussed in Section VII.1, as in the gaussian case we consider, it can be taken to be a universal constant. Furthermore, we can ignore all appearances of powers of σ, $\Sigma_{::}$, and S_θ for $\theta \in \{\lambda, \omega\}$, using bounds $\mathrm{w}(\Sigma_{::}^k \cdot) \le \|\Sigma_{::}^k\| \, \mathrm{w}(\cdot)$, $\|\Sigma_{::}^k \cdot\| \le \|\Sigma^k\| \|\cdot\|$, and $\|S_\theta^{1/2} \cdot\| \le \|S_\theta^{1/2}\| \|\cdot\|$ and observing that $\|S_\theta\| \le 1$ by construction and, under Assumption 6, $\|\Sigma_{::}\|$ and $\|\Sigma_{::}^{-1}\|$ are bounded by universal constants. And we bound minima over ω_0 and $\tilde{\lambda}_0$ by substituting $\tilde{\omega}_0$ and $\tilde{\lambda}_0$. Then, as $\mathrm{w}(\Lambda^\star_{s_\lambda}) \lesssim \sqrt{\log(T_0)}$ and $\mathrm{w}(\Omega^\star_{s_\omega}) \lesssim \sqrt{\log(N_0)}$, Lemma 5 and Lemma 6 together (taking $\sigma = 1$ in the latter), imply that on an event of probability $1 - c\exp(-u^2) - c\exp(-v)$ for v as in Lemma 5, the following bound holds for $\eta^2 = 1 + \zeta^2$.

$$
\begin{aligned}
|\hat{\tau}(\hat{\lambda}, \hat{\omega}) - \hat{\tau}(\tilde{\lambda}, \tilde{\omega})| \lesssim\ & u[N_{eff}^{-1/2} N_0^{-1/2} r_\lambda + T_{eff}^{-1/2} (\eta^2 T_0)^{-1/2} r_\omega + (\eta^2 N_0 T_0)^{1/2} r_\omega r_\lambda] \\
& + (\|\tilde{\omega}\| + (\eta^2 T_0)^{-1/2} r_\omega) \log^{1/2}(T_0) + (\|\psi - \tilde{\lambda}\| + N_0^{-1/2} r_\lambda) \log^{1/2}(N_0) \\
& + (\eta^2 T_0)^{-1/2} r_\omega E_\lambda + N_0^{-1/2} r_\lambda E_\omega + r_\omega r_\lambda M \qquad \text{for any}
\end{aligned}
$$

$$
M \ge \min\left(N_0^{-1/2},\ (\eta^2 T_0)^{-1/2},\ \min_{k \in \mathbb{N}} \sigma_k(L^c_{::})^{-1} + \sigma_{k+1}(L^c_{::})(\eta^2 N_0 T_0)^{-1/2} \right) \quad \text{and}
$$

$$
\begin{aligned}
r_\lambda &= \log^{1/4}(T_0)[(N_0/T_{eff})^{1/4} + E_\lambda^{1/2}], \quad E_\lambda = \|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|, \ T_{eff}^{-1/2} = \|\tilde{\lambda} - \psi\| + T_1^{-1/2}, \\
r_\omega &= \log^{1/4}(N_0)[(T_0/N_{eff})^{1/4} + E_\omega^{1/2}], \quad E_\omega = \|L'_{::}\tilde{\omega} + \tilde{\omega}_0 - L'_{N:}\|, \ N_{eff}^{-1/2} = \|\tilde{\omega}\| + N_1^{-1/2}.
\end{aligned}
$$

Taking $u = \min(T_{eff}^{1/2} \log^{1/2}(T_0), N_{eff}^{1/2} \log^{1/2}(N_0), (\eta^2 N_0 T_0)^{1/2} M)$, we can ignore the first line in the bound above, as its three terms are bounded by the second term in the second line, the first term in the second line, and the final term respectively. Grouping terms with common powers of r_ω, r_λ; redefining $\mathrm{E}_\lambda = \max(E_\lambda, 1)$ and $\mathrm{E}_\omega = \max(E_\omega, 1)$, and expanding r_ω, r_λ yields the following bound.

(40)

$$
\begin{aligned}
& \|\tilde{\omega}\| \log^{1/2}(T_0) + \|\psi - \tilde{\lambda}\| \log^{1/2}(N_0) \\
& + (\eta^2 T_0)^{-1/2} [(T_0/N_{eff})^{1/4} + E_\omega^{1/2}] E_\lambda \log^{1/2}(N_0) \\
& + N_0^{-1/2} [(N_0/T_{eff})^{1/4} + E_\lambda^{1/2}] E_\omega \log^{1/2}(T_0) \\
& + M[(N_0 T_0/N_{eff} T_{eff})^{1/4} + (N_0/T_{eff})^{1/4} E_\omega^{1/2} + (T_0/N_{eff})^{1/4} E_\lambda^{1/2} + (E_\omega E_\lambda)^{1/2}] \log^{1/4}(N_0) \log^{1/4}(T_0).
\end{aligned}
$$

Each term is multiplied by either $\log^{1/2}(T_0)$, $\log^{1/2}(N_0)$, or their geometric mean. For simplicity, we will substitute a common upper bound of $\ell^{1/2}$ for $\ell = \log(\max(N_0, T_0))$. To establish our claim, we must show that each term is $o((N_1T_1)^{-1/2})$.

The first line of our bound is small enough, $N_{eff} \sim N_1$, and $T_{eff} \sim T_1$, if

$$\max(\|\tilde{\omega}\|, \|\tilde{\lambda} - \psi\|) \ll (N_1T_1)^{-1}\ell^{-1/2}, \quad \min(N_1, T_1) \gtrsim 1, \tag{41}$$

If the following bound holds, the remaining terms that do no involve M are small enough.

$$\begin{aligned} E_\omega &\ll N_0^{1/4} N_1^{-1/2} T_1^{-1/4} \ell^{-1/2}, \\ E_\lambda &\ll \eta T_0^{1/4} N_1^{-1/4} T_1^{-1/2} \ell^{-1/2}, \\ (E_\omega E_\lambda)^{1/2} &\ll \min(N_0^{3/8} T_1^{-3/8} N_1^{-1/4},\ \eta^{1/2} T_0^{3/8} N_1^{-3/8} T_1^{-1/4}) \ell^{-1/4}. \end{aligned} \tag{42}$$

To see this, multiply the square root of the first bound by the first part of the third when bounding the term involving $E_\lambda^{1/2} E_\omega$ and the square root of the second by the second part of the third when bounding the term involving $E_\omega^{1/2} E_\lambda$. Note that because our 'redefinition' of E_ω, E_λ requires that they be no smaller than one, these upper bounds must go to infinity, and so long as they do we can interpret them as bounds on $\|L'_{::}\tilde{\omega} + \tilde{\omega}_0 - L'_{N:}\|$, $\|L_{::}\tilde{\lambda} + \tilde{\lambda}_0 - L_{:T}\|$, and their geometric mean respectively.

By substituting the bounds (42) into the term with a factor of M in (40), we can derive a sufficent condition for it to be small enough. To see that it is sufficient, we bound first multiple of M in (40) using the first bound on M below, the second using the second in combination with our bound on E_ω, the third using the third in combination with our bound on E_λ, and the fourth using the second in combination with our first bound on $(E_\omega E_\lambda)^{1/2}$.

$$M \ll \min\left((N_0T_0N_1T_1\ell)^{-1/4},\ N_0^{-3/8} N_1^{-1/4} T_1^{-1/8},\ \eta^{-1/2} T_0^{-3/8} T_1^{-1/4} N_1^{-1/8}\right) \ell^{-1/4}. \tag{43}$$

Equations 41, 42, and 43, so long as the bounds in (42) all go to infinity, are sufficient to imply our claim. Note that because every vector ω in the unit simplex in $\mathbb{R}^{N_0}$ satisfies $\|\omega\| \geq N_0^{-1/2}$, (41) implies an additional constraint on the dimensions of the problem, $N_0 \gg N_1T_1\ell$.

Having established these bounds on E_ω and E_λ, we are now in a position to characterize the probability that our result holds by lower bounding the ratios N_0/E_λ and T_0/E_ω that appear in the probability statement of Lemma 5. As $N_0/E_\lambda \gg N_0^{3/4}$ and $T_0/E_\omega \gg T_0^{3/4}$, the claims of Lemma 5 hold with probability $1 - c\exp(-v)$ for $v = c\min(N_0, T_0)^{1/2}$. Thus, recalling from above that we are working on an event of probability $1 - c\exp(-u^2) - c\exp(-v)$ for $u = \min(T_{eff}^{1/2}\log^{1/2}(T_0), N_{eff}^{1/2}\log^{1/2}(N_0), (\eta^2 N_0 T_0)^{1/2} M)$ and that $N_{eff} \sim N_1$ and $T_{eff} \sim T_1$, this is probability at least $1 - 2\exp(-\min(T_1\log(T_0), N_1\log(N_0), \eta^2 N_0 T_0 M^2)) - c\exp(-c\min(N_0^{1/2}, T_0^{1/2}))$.

We will now derive simplfied sufficient conditions under the assumption that $N_0 \sim T_0$. Let $m_0 = N_0$, $m_1 = (N_1 T_1)^{1/2}$, and $\bar{m}_1 = \max(N_1, T_1)$. Then (43) holds if

$$M \ll \min(m_0^{-1/2} m_1^{-1/2} \ell^{-1/2},\ \eta^{-1/2} m_0^{-3/8} m_1^{-1/4} \bar{m}_1^{-1/4} \ell^{-1/4}).$$

This is not satisfiable with $M = N_0^{-1/2} \sim m_0^{1/2}$. But with $M = (\eta T_0)^{-1/2} \sim \eta^{-1} m_0^{-1/2}$, it is satisfied for $\eta \gg \max(1,\ m_0^{-1/4}\bar{m}_1^{1/2}) m_1^{1/2}\ell^{1/2}$. For such η, (42) hold when

$$\begin{aligned} E_\omega &\ll m_0^{1/4} m_1^{-1/2} \bar{m}_1^{-1/4} \ell^{-1/2}, \\ E_\lambda &\ll \max(m_0^{1/4}\bar{m}_1^{-1/4},\ \bar{m}_1^{1/4}) \\ (E_\omega E_\lambda)^{1/2} &\ll m_0^{3/8} m_1^{-1/2} \bar{m}_1^{-1/8} \ell^{-1/4}. \end{aligned}$$

To keep the statement of our lemma simple, we use the simplified bound $E_\lambda \ll m_0^{1/4}\bar{m}_1^{-1/4}$. Then the geometric mean of our bounds on E_ω and E_λ bounds their geometric mean, and it is $m_0^{1/4} m_1^{-1/4} \bar{m}_1^{-1/4} \ell^{-1/4}$. Thus, our explicit bound on the geometric mean above is redundant as long as the ratio of these two bounds, $m_0^{1/4} m_1^{-1/4} \bar{m}_1^{-1/4} \ell^{-1/4} / m_0^{3/8} m_1^{-1/2} \bar{m}_1^{-1/8} \ell^{-1/4}$, is bounded. As this ratio simplifies to $m_0^{-1/8} m_1^{1/4} \bar{m}_1^{-1/8} \le (m_1/m_0)^{1/8}$ and $m_0 \gg m_1$, it is redundant. And taking $M \sim \eta^{-1} m_0^{-1/2}$ in our probability statement above, our claims hold with probability $1 - 2\exp(-\min(T_1\log(T_0), N_1\log(N_0))) - c\exp(-c m_0^{1/2})$).

To avoid complicating the statement of our result, we will not explore refinements made possible by a nontrivially large gap in the spectrum of $L^c_{::}$, i.e., the case that $M = \min_k \sigma_k(L^c_{::})^{-1} + \sigma_{k+1}(L^c_{::})(\eta^2 N_0 T_0)^{-1/2}$. However, in models with no weak factors, this quantity will be very small, and as a result, Equations 41 and 42 will essentially be sufficient to imply our claim. As we make η large only to control M when it is equal to $(\eta T_0)^{-1/2}$, this provides some justification for the use of weak regularization (ζ small) or no regularization ($\zeta = 0$) when fitting the synthetic control $\hat{\omega}$.

We conclude by observing that the lower bound on ζ above simplifies to $\zeta \gg m_1^{1/2}\ell^{1/2}$ under our stated assumptions. We begin with the assumption that the above upper bound on E_ω goes to infinity. Observing that the other lower bound on ζ as stated above is $\bar{m}_1^{1/4}$ times the reciprocal of the this infinity-tending bound on E_ω, it follows that it must be $o(\bar{m}_1^{1/4})$. As $m_1^{1/2} = \bar{m}_1^{1/4}\min(N_1, T_1)^{1/4}$ and the latter factor and $\ell^{1/2}$ are bounded away from zero by assumption, $\bar{m}_1^{1/4} = O(m_1^{1/2}\ell^{1/2})$, so this other lower bound is indeed smaller than the (other) one that we retain.

IX Proof of Theorem 2

Throughout this proof, we will assume constant treatment effects $\tau_{ij} = \tau$. When treatment effects are not constant, the jackknife variance estimate will include an additional nonnegative term that depends on the amount of treatment heterogeneity, making the inference conservative.

We will write $a \sim_p b$ meaning $a/b \to_p 1$, $a \lesssim_p b$ meaning $a = O_p(b)$, $a \ll_p b$ meaning $a = o_p(b)$, $\sigma_{\min}(\Sigma)$ and $\sigma_{\max}(\Sigma)$ for the smallest and largest eigenvalues of a matrix Σ, and $1_n \in \mathbb{R}^n$ for a vector of ones. And we write $\hat{\lambda}^\star$ to denote the concatenation of $\hat{\lambda}_{\text{pre}}$ and $-\hat{\lambda}_{\text{post}}$.

Now recall that, as discussed in Section III.1,

$$
\begin{aligned}
\hat{\tau} &= \hat{\omega}_{\text{tr}}' Y_{\text{tr,post}} \hat{\lambda}_{\text{post}} - \hat{\omega}_{\text{co}}' Y_{\text{co,post}} \hat{\lambda}_{\text{post}} - \hat{\omega}_{\text{tr}}' Y_{\text{tr,pre}} \hat{\lambda}_{\text{pre}} + \hat{\omega}_{\text{co}}' Y_{\text{co,pre}} \hat{\lambda}_{\text{pre}} \\
&= \hat{\mu}_{\text{tr}} - \hat{\mu}_{\text{co}} \quad \text{where} \\
\hat{\mu}_{\text{co}} &= \sum_{i=1}^{N_{\text{co}}} \hat{\omega}_i \widehat{\Delta}_i, \quad \hat{\mu}_{\text{tr}} = \sum_{i=N_{\text{co}}+1}^{N} \hat{\omega}_i \widehat{\Delta}_i, \quad \widehat{\Delta}_i = Y_{i,\cdot} \hat{\lambda}^\star.
\end{aligned}
\tag{44}
$$

In the jackknife variance estimate defined in Algorithm 3,

$$
\hat{\tau}^{(-i)} = \begin{cases} \hat{\mu}_{\text{tr}} - \frac{\sum_{k \le N_{\text{co}}, k \ne i} \hat{\omega}_k \Delta_k}{1-\hat{\omega}_i} = \hat{\mu}_{\text{tr}} - \left(\hat{\mu}_{\text{co}} - \frac{\hat{\omega}_i(\Delta_i - \hat{\mu}_{\text{co}})}{1-\hat{\omega}_i}\right) & \text{for } \; i \le N_{\text{co}} \\ \frac{\sum_{k \ge N_{\text{co}}, k \ne 1} \hat{\omega}_k \Delta_k}{1-\hat{\omega}_i} - \hat{\mu}_{\text{co}} = \left(\hat{\mu}_{\text{tr}} - \frac{\hat{\omega}_i(\Delta_i - \hat{\mu}_{\text{tr}})}{1-\hat{\omega}_i}\right) - \hat{\mu}_{\text{co}} & \text{for } \; i > N_{\text{co}}. \end{cases}
\tag{45}
$$

Thus, the jackknife variance estimate defined in Algorithm 3 is

$$
\widehat{V}_\tau^{\text{jack}} = \frac{N-1}{N} \left(\sum_{i=1}^{N_{\text{co}}} \left(\frac{\hat{\omega}_i \left(\widehat{\Delta}_i - \hat{\mu}_{\text{co}}\right)}{1-\hat{\omega}_i} \right)^2 + \sum_{i=N_{\text{co}}+1}^{N} \left(\frac{\hat{\omega}_i \left(\widehat{\Delta}_i - \hat{\mu}_{\text{tr}}\right)}{1-\hat{\omega}_i} \right)^2 \right).
\tag{46}
$$

A few simplifications are now in order. We use the bound $\|\hat{\omega}_{\text{co}}\|^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}$ derived in Section IX.0.1 below. This bound implies that the denominators $1 - \hat{\omega}_i$ appearing in the expression above all lie in the interval $[1-\max(\|\hat{\omega}_{\text{co}}\|, N_{\text{tr}}^{-1}),\ 1] = [1-o_p(1),\ 1]$. As each term in that expression is nonnegative, it follows that the ratio between it and the expression below, derived by replacing these denominators with 1, is in this interval and therefore converges to one.

$$\widehat{V}_\tau^{\text{jack}} \sim_p \sum_{i=1}^{N_{\text{co}}} \hat{\omega}_i^2 \left(\widehat{\Delta}_i - \hat{\mu}_{\text{co}}\right)^2 + \sum_{i=N_{\text{co}}+1}^{N} \hat{\omega}_i^2 \left(\widehat{\Delta}_i - \hat{\mu}_{\text{tr}}\right)^2. \tag{47}$$

We will simplify this further by showing that the first term is negligible relative to the second. We'll start by lower bounding the second term. This is straightforward because for $i > N_{\text{co}}$, the unit weights $\hat{\omega}_i$ are equal to the constant $1/N_{\text{tr}}$ and the time weights $\hat{\lambda}$ are independent of $Y_{i,\cdot}$.

$$\begin{aligned}
\mathrm{E} \sum_{i=N_{\text{co}}+1}^{N} \hat{\omega}_i^2 \left(\widehat{\Delta}_i - \hat{\mu}_{\text{tr}}\right)^2 &= N_{\text{tr}}^{-2} \sum_{i=N_{\text{co}}+1}^{N} \mathrm{E}((Y_{i,\cdot} - \hat{\omega}_{\text{tr}}' Y_{\text{tr},\cdot})\hat{\lambda}^\star)^2 \\
&\geq N_{\text{tr}}^{-2} \sum_{i=N_{\text{co}}+1}^{N} \mathrm{E}((\varepsilon_{i,\cdot} - \hat{\omega}_{\text{tr}}' \varepsilon_{\text{tr},\cdot})\hat{\lambda}^\star)^2 \\
&= N_{\text{tr}}^{-1}\, \mathrm{E}\, \hat{\lambda}_\star' \ (1-N_{\text{tr}}^{-1})\Sigma\ \hat{\lambda}^\star && \text{as } \operatorname{Cov}\left[\varepsilon_{i,\cdot} - \hat{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\cdot}\right] = (1-N_{\text{tr}}^{-1})\Sigma \\
&\geq N_{\text{tr}}^{-1} \|\hat{\lambda}^\star\|^2 (1-N_{\text{tr}}^{-1}) \sigma_{\min}(\Sigma) \\
&\geq (N_{\text{tr}} T_{\text{post}})^{-1} (1-N_{\text{tr}}^{-1}) \sigma_{\min}(\Sigma) && \text{as } \|\hat{\lambda}^\star\|^2 \geq \|\hat{\lambda}_{\text{tr}}\|^2 = T_{\text{post}}^{-1}.
\end{aligned}$$

As $\sigma_{\min}(\Sigma)$ is bounded away from zero, it follows that the mean of the second term in (47) is on the order of $(N_{\text{tr}} T_{\text{post}})^{-1}$ or larger. We'll now show that the first term in (47) is $o_p((N_{\text{tr}} T_{\text{post}})^{-1})$, so (47) is equivalent to a variant in which we have dropped its first term.

By Hölder's inequality and the bound $\|\hat{\omega}_{\text{co}}\|^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}$ derived in Section IX.0.1,

$$\sum_{i=1}^{N_{\text{co}}} \hat{\omega}_i^2 \left(\widehat{\Delta}_i - \hat{\mu}_{\text{co}}\right)^2 \leq \|\hat{\omega}_{\text{co}}\|^2 \max_{i \leq N_{\text{co}}} \left(\widehat{\Delta}_i - \hat{\mu}_{\text{co}}\right)^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1} \max_{i \leq N_{\text{co}}} \left(\widehat{\Delta}_i - \hat{\mu}_{\text{co}}\right)^2.$$

Thus, it suffices to show that $\max_{i \leq N_{\text{co}}} (\widehat{\Delta}_i - \hat{\mu}_{\text{co}})^2 \ll \log(N_{\text{co}})$. And it suffices to show that $\max_{i \leq N_{\text{co}}} \widehat{\Delta}_i^2 \ll \log(N_{\text{co}})$, as $(\widehat{\Delta}_i - \hat{\mu}_{\text{co}})^2 \leq 2\widehat{\Delta}_i^2 + 2\hat{\mu}_{\text{co}}^2$ and $\hat{\mu}_{\text{co}}$ is a convex combination of $\widehat{\Delta}_1 \ldots \widehat{\Delta}_{\text{co}}$. This bound holds because, by Hölder's inequality,

$$\max_{i \leq N_{\text{co}}} \left|\widehat{\Delta}_i\right| = \max_{i \leq N_{\text{co}}} \left|Y_{i,\cdot} \hat{\lambda}^\star\right| \leq \left\|\hat{\lambda}^\star\right\|_1 \cdot \max_{i \leq N_{\text{co}},\ j \leq T} |Y_{ij}| \lesssim_p \sqrt{\log(N_{\text{co}})}.$$

In our last comparison above, we use the properties that $\|\hat{\lambda}^{\star}\|_1 = \|\hat{\lambda}_{\text{pre}}\|_1 + \|\hat{\lambda}_{\text{post}}\|_1 = 2$, that the elements of L are bounded, and that the maximum of $K = N_{\text{co}}T$ gaussian random variables ε_{it} is $O_p(\sqrt{\log(K)})$, as well as Assumption 2, which implies that $T \sim N_{\text{co}}$ so $\log(K) \lesssim \log(N_{\text{co}})$. Summarizing,

$$\widehat{V}_{\tau}^{\text{jack}} \sim_p \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i - \hat{\mu}_{\text{tr}}\right)^2. \tag{48}$$

This simplification is as we would hope given that, under the conditions of Theorem 1, we found that all the noise in $\hat{\tau}$ comes from the exposed units. Now, focusing further on (48) we note that, when treatment effects are constant across units, we can verify that they do not contribute to $\widehat{V}_{\tau}^{\text{jack}}$ and so

$$\begin{gathered}\frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i - \hat{\mu}_{\text{tr}}\right)^2 = \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(L) - \hat{\mu}_{\text{tr}}(L) + \widehat{\Delta}_i(\varepsilon) - \hat{\mu}_{\text{tr}}(\varepsilon)\right)^2, \\ \widehat{\Delta}_i(L) = L_{i,\cdot}\hat{\lambda}^{\star} \qquad \widehat{\Delta}_i(\varepsilon) = \varepsilon_{i,\cdot}\hat{\lambda}^{\star},\end{gathered} \tag{49}$$

where $\hat{\mu}_{\text{tr}}(L)$ and $\hat{\mu}_{\text{tr}}(\varepsilon)$ are averages of $\widehat{\Delta}_i(L)$ and $\widehat{\Delta}_i(\varepsilon)$ respectively over the exposed units. Now, by construction, $\hat{\lambda}$ is only a function of the unexposed units and so, given that there is no cross-unit correlation, $\hat{\lambda}$ is independent of $\varepsilon_{i,\cdot}$ for all $i > N_{\text{co}}$. Thus, the cross terms between $\widehat{\Delta}_i(L) - \hat{\mu}_{\text{tr}}(L)$ and $\widehat{\Delta}_i(\varepsilon) - \hat{\mu}_{\text{tr}}(\varepsilon)$ in (49) are mean-zero and concentrate out, and so

$$\widehat{V}_{\tau}^{\text{jack}} \sim_p \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(L) - \hat{\mu}_{\text{tr}}(L)\right)^2 + \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(\varepsilon) - \hat{\mu}_{\text{tr}}(\varepsilon)\right)^2. \tag{50}$$

We will now show that the second term is equivalent to a variant in which $\tilde{\lambda}$ replaces $\hat{\lambda}$. We denote by $\tilde{\Delta}$ and $\tilde{\mu}_{\text{tr}}$ the corresponding variants of $\widehat{\Delta}$ and $\hat{\mu}_{\text{tr}}$. First consider the second term in

(50). $\widehat{\Delta}_i(\varepsilon) = \tilde{\Delta}_i(\varepsilon) + \varepsilon_{i,\text{pre}}(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}})$, so

$$\begin{aligned}
\left(\widehat{\Delta}_i(\varepsilon) - \hat{\mu}_{\text{tr}}(\varepsilon)\right)^2 &= \left([\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon)] + (\varepsilon_{i,\text{pre}} - \hat{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}})\right)^2 \\
&= \left(\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon)\right)^2 \\
&\quad + 2[\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon)](\varepsilon_{i,\text{pre}} - \hat{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}}) \\
&\quad + ((\varepsilon_{i,\text{pre}} - \hat{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}}))^2.
\end{aligned}$$

By the Cauchy-Schwarz inequality, the second and third terms in this decomposition are negligible relative to the first if $\mathrm{E}_{\text{tr}}((\varepsilon_{i,\text{pre}} - \hat{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}}))^2 \ll_p \mathrm{E}_{\text{tr}}(\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon))^2$ where E_{tr} denotes expectation conditional on $\varepsilon_{\text{co},\cdot}$. We calculate both quantities and compare.

$$\begin{aligned}
&\mathrm{E}_{\text{tr}}((\varepsilon_{i,\text{pre}} - \tilde{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}}))^2 = (\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}})'(1 - N_{\text{tr}}^{-1})\Sigma(\hat{\lambda}_{\text{pre}} - \tilde{\lambda}_{\text{pre}}). \\
&\mathrm{E}_{\text{tr}}(\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon))^2 = \mathrm{E}_{\text{tr}}((\varepsilon_{i,\cdot} - \tilde{\omega}_{\text{tr}}'\varepsilon_{\text{tr},\text{pre}})'\tilde{\lambda}^{\star})^2 = \tilde{\lambda}'\ (1 - N_{\text{tr}}^{-1})\Sigma\ \tilde{\lambda}.
\end{aligned}$$

In Section IX.0.2, we show that the first is $\lesssim_p N_{\text{co}}^{-1/2} T_{\text{post}}^{-1/2} \log^{1/2}(N_{\text{co}})$, and the second is $\gtrsim \|\tilde{\lambda}^{\star}\|^2 \geq T_{\text{post}}^{-1}$ because $\sigma_{\min}(\Sigma)$ is bounded away from zero. Thus, because $N_{\text{co}}^{-1/2} \ll T_{\text{post}}^{-1/2} \log^{-1/2}(N_{\text{co}})$ under Assumption 2, the first quantity is negligible relative to the second. As discussed, it follows that

$$\frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(\varepsilon) - \hat{\mu}_{\text{tr}}(\varepsilon)\right)^2 \sim_p \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\tilde{\Delta}_i(\varepsilon) - \tilde{\mu}_{\text{tr}}(\varepsilon)\right)^2. \tag{51}$$

By the law of large numbers, the right side is equivalent ($\sim_p$) to its mean $N_{\text{tr}}^{-1}\tilde{\lambda}'\ (1 - N_{\text{tr}}^{-1})\Sigma\ \tilde{\lambda}$ and therefore to $N_{\text{tr}}^{-1}\tilde{\lambda}'\Sigma\tilde{\lambda}$. It is shown that $N_{\text{tr}}^{-1}\tilde{\lambda}'\Sigma\tilde{\lambda} \sim_p V_\tau$ in the proof of Lemma 4, so

$$\widehat{V}_\tau^{\text{jack}} \sim_p \frac{1}{N_{\text{tr}}^2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(L) - \hat{\mu}_{\text{tr}}(L)\right)^2 + V_\tau. \tag{52}$$

Because the first term is nonnegative, our variance estimate is asymptotically either unbiased or upwardly biased, so our confidence intervals are conservative as claimed. In the remainder, we derive a sufficient condition for the first term to be asymptotically negligible relative to V_τ, so our confidence intervals have asymptotically nominal coverage.

We bound this term using the expansion $\hat{\mu}_{\text{tr}}(L) = N_{\text{tr}}^{-1} 1'_{N_{\text{tr}}}(L_{\text{tr,post}}\hat{\lambda}_{\text{post}} - L_{\text{tr,pre}}\hat{\lambda}_{\text{pre}})$.

$$N_{\text{tr}}^{-2} \sum_{i=N_{\text{co}}+1}^{N} \left(\widehat{\Delta}_i(L) - \hat{\mu}_{\text{tr}}(L)\right)^2 = N_{\text{tr}}^{-2} \|(I - N_{\text{tr}}^{-1} 1_{N_{\text{tr}}} 1'_{N_{\text{tr}}})(L_{\text{tr,pre}}\hat{\lambda}_{\text{pre}} + \hat{\lambda}_0 1_{N_{\text{tr}}} - L_{\text{tr,post}}\hat{\lambda}_{\text{post}})\|^2$$
$$\leq N_{\text{tr}}^{-2} \|L_{\text{tr,pre}}\hat{\lambda}_{\text{pre}} + \hat{\lambda}_0 - L_{\text{tr,post}}\hat{\lambda}_{\text{post}}\|^2.$$

This comparison holds because $\|I - N_{\text{tr}}^{-1} 1_{N_{\text{tr}}} 1_{N_{\text{tr}}}\| \leq 1$. By Assumption (31), this bound is $o_P((N_{\text{tr}} T_{\text{post}})^{-1})$ and therefore negligible relative to V_τ. We conclude by proving our claims about $\|\hat{\omega}_{\text{co}}\|$ and $\|\Sigma_{\text{pre}}^{1/2}(\hat{\lambda}_{\text{co}} - \tilde{\lambda}_{\text{co}})\|$.

IX.0.1 Bounding $\|\hat{\omega}_{\text{co}}\|$

Here we will show that $\|\hat{\omega}_{\text{co}}\|^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}$ under the assumptions of Theorem 1.

$$\begin{aligned}
&\|\hat{\omega}_{\text{co}} - \tilde{\omega}_{\text{co}}\|^2 \lesssim_p \ \zeta^{-2} N_{\text{co}}^{-1} [N_{\text{co}}^{1/2} N_{\text{tr}}^{-1/2} + \|\tilde{\omega}'_{\text{co}} L_{\text{co,pre}} + \tilde{\omega}_0 - \tilde{\omega}'_{\text{tr}} L_{\text{tr,pre}}\|] \log^{1/2}(N_{\text{co}}) \\
&\ll [N_{\text{tr}}^{1/2} T_{\text{post}}^{1/2} \log(N_{\text{co}})]^{-1} N_{\text{co}}^{-1/2} N_{\text{tr}}^{-1/2} \log^{1/2}(N_{\text{co}}) \\
&+ [N_{\text{tr}}^{1/2} T_{\text{post}}^{1/2} \max(N_{\text{tr}}, T_{\text{post}})^{1/2} N_{\text{co}}^{-1/4} \log(N_{\text{co}})]^{-1} N_{\text{co}}^{-3/4} N_{\text{tr}}^{-1/4} T_{\text{post}}^{-1/4} \max(N_{\text{tr}}, T_{\text{post}})^{-1/4} \\
&\ll N_{\text{co}}^{-1/2} N_{\text{tr}}^{-1} T_{\text{post}}^{-1/2} \\
&\ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}.
\end{aligned}$$

Our first bound follows from Lemma 5, in which we can take $N_{eff}^{-1/2} \sim N_{\text{tr}}^{-1/2}$ because $\|\tilde{\omega}_{\text{co}}\| \lesssim N_{\text{tr}}^{-1/2}$ under Assumption 4. To derive our second, we substitute the upper bound $N_{\text{co}}^{1/4} N_{\text{tr}}^{-1/4} T_{\text{post}}^{-1/4} \max(N_{\text{tr}}, T_{\text{post}})^{-1/4} \log^{-1/2}(N_{\text{co}}) \gg \|\tilde{\omega}'_{\text{co}} L_{\text{co,pre}} + \tilde{\omega}_0 - L_{\text{tr,pre}}\|$ from Assumption 4 and substitute (in brackets) two lower bounds on ζ^2 chosen as in Theorem 1: the first is implied by squaring the lower bound $\zeta \gg (N_{\text{tr}} T_{\text{post}})^{1/4} \log^{1/2}(N_{\text{co}})$ and the second by multiplying this lower bound by an alternative lower bound, $\zeta \gg (N_{\text{tr}} T_{\text{post}})^{1/4} \max(N_{\text{tr}}, T_{\text{post}})^{1/2} N_0^{-1/4} \log^{1/2}(N_{\text{co}})$. The third is a simplification, and the fourth follows because $T_{\text{post}} \log^2(N_{\text{co}}) \ll N_{\text{co}}$ under Assumption 2. Furthermore, as $\|\tilde{\omega}_{\text{co}}\|^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}$ under Assumption 4, by the triangle inequality, $\|\hat{\omega}_{\text{co}}\|^2 \ll (N_{\text{tr}} T_{\text{post}} \log(N_{\text{co}}))^{-1}$ as claimed.

IX.0.2 Bounding $\|\Sigma_{\text{pre,pre}}(\hat{\lambda}_{\text{co}} - \tilde{\lambda}_{\text{co}})\|$

Here we will show that $\|\Sigma_{\text{pre,pre}}(\hat{\lambda}_{\text{co}} - \tilde{\lambda}_{\text{co}})\|^2 \lesssim_p N_{\text{co}}^{-1/2} T_{\text{post}}^{-1/2} \log^{1/2}(N_{\text{co}})$. Because Assumption 1 implies that $\|\Sigma_{\text{pre},pre}\|$ is bounded, it suffices to bound $\|\hat{\lambda}_{\text{co}} - \tilde{\lambda}_{\text{co}}\|$.

$$\begin{aligned}
\|\hat{\lambda}_{\text{co}} - \tilde{\lambda}_{\text{co}}\|^2 &\lesssim_p N_{\text{co}}^{-1}[N_{\text{co}}^{1/2} T_{\text{post}}^{-1/2} + \|L_{\text{co,pre}}\tilde{\lambda}_{\text{pre}} + \tilde{\lambda}_0 - L_{\text{co,post}}\tilde{\lambda}_{\text{post}}\|] \log^{1/2}(N_{\text{co}}) \\
&\lesssim N_{\text{co}}^{-1/2} T_{\text{post}}^{-1/2} \log^{1/2}(N_{\text{co}}) + N_{\text{co}}^{-3/4} N_{\text{tr}}^{-1/8} T_{\text{post}}^{-1/8} \log^{1/2}(N_{\text{co}}) \\
&\lesssim N_{\text{co}}^{-1/2} T_{\text{post}}^{-1/2} \log^{1/2}(N_{\text{co}}).
\end{aligned}$$

Our first bound follows from Lemma 5, in which we can take $T_{eff}^{-1/2} \sim T_{\text{post}}^{-1/2}$ because $\|\tilde{\lambda}_{\text{pre}} - \psi\| \lesssim T_{\text{post}}^{-1/2}$ under Assumption 4. To derive our second, we substitute the upper bound $N_{\text{co}}^{1/4} N_{\text{tr}}^{-1/8} T_{\text{post}}^{-1/8} \gg \|L_{\text{co,pre}}\tilde{\lambda}_{\text{pre}} + \tilde{\lambda}_0 - L_{\text{co,post}}\lambda_{\text{post}}\|$ from Assumption 4. The third follows because $N_{\text{co}}^{-1/4} \ll N_{\text{tr}}^{-1/4} T_{\text{post}}^{-1/4} \max(N_{\text{tr}}, T_{\text{post}})^{-1/4} \leq N_{\text{tr}}^{-3/8} T_{\text{post}}^{-3/8}$ under Assumption 2.