

Corrections to “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models”

Political Analysis 25(1): 57–76 (2017)

Yiqing Xu*

October 18, 2026

Abstract

This note corrects and clarifies several statements in Xu (2017) and its Supplementary Materials (SM) and provides complete proofs of the corrected results. Under assumptions following Bai (2009), each treated unit’s estimation error equals its post-treatment shock up to a term that vanishes in probability. The corrections address an error in the proof of a lemma in the SM and omitted or misstated conditions, primarily those concerning serial dependence and the scaling of an identification condition. The note also qualifies claims about the bootstrap (Algorithm 2), cross-validation (Algorithm 1), and dynamic models; adds an integrability condition sufficient for a statement about the mean of the estimation error; and fixes typos. The method and estimator remain unchanged. Software issues in earlier versions of the gsynth and fect packages are discussed in Xu (2026), and I will provide a more comprehensive review separately.

*Yiqing Xu, Associate Professor, Department of Political Science, Stanford University. Email: yiqingxu@stanford.edu. I used AI tools (Anthropic’s Claude and OpenAI’s Codex) to help me understand better the underlying issues, draft the corrected proofs and accompanying text, and proofread. I checked their findings and take full responsibility for the content of this note.

1. Summary

Table 1 lists the problems. Panel A has the one error. Several steps in the proof of SM Lemma 3(i) are incorrect as printed, and this note gives a new proof under the corrected assumptions. Panel B has imprecisions, which are claims made without proof or results that need a condition the article or the SM does not state, and typos that change a formula or a limit. For Algorithm 2 and SM Assumption 3, an example shows that the printed conditions do not suffice. Smaller slips are listed in Section 10. “Article” is the published paper and “SM” its Online Appendix. Starred assumptions, lemmas and propositions are the corrected versions in this note.

TABLE 1. PROBLEMS AND CORRECTIONS

Where	Problem	Correction
A. Error		
SM Lemma 3(i)	The rotation is transposed ($H^{-1}\hat{f}_t$ for $(H')^{-1}\hat{f}_t$); the bound on a_6 does not give its claimed rate, and the bounds on a_7 and b_1 use inequalities that are false as written.	Lemma 2*, with a complete proof.
B. Imprecisions and typos		
Article, introduction; Algorithm 1	“With sufficient data, the cross-validation procedure can pick up the correct number of factors with high probability.” The support is Monte Carlo evidence with errors independent over time. With serially correlated errors the procedure can choose too many factors in finite samples.	Restated as simulation evidence (Sections 5 and 9).
Article, Remark 1	“The proposed method can accommodate dynamic models as long as the error terms are not serially correlated.” This holds for $\hat{\beta}$ (Moon and Weidner, 2015), but the treatment-effect results assume strict exogeneity, which a lagged outcome violates.	Qualified (Section 6).
Article, Remark 2	Typo in the limit (“ $N_{co}, T_0 \rightarrow 0$ ” for $\rightarrow \infty$). The remark attributes bias at small T_0 to imprecise loadings, but that term has mean zero given the control data.	Restated (Section 6).
Article, inference paragraph; Algorithm 2	“Under Assumptions 1–5, this procedure provides valid uncertainty estimates.” Equal error covariances (Assumption 5) do not make the treated and control error distributions equal, so coverage is not ensured (Section 4).	Identically distributed errors, a sufficient condition (Assumption 5*); claim restated as Monte Carlo evidence.
Article, Algorithm 2, Step 4	The percentile interval agrees with the basic interval when the error distribution is symmetric. With asymmetric errors it need not have nominal coverage (Section 4).	Qualified, with the basic interval for asymmetric errors.

Where	Problem	Correction
SM Assumption 3	Omits the summability and eigenvalue conditions of Bai (2009, Assumption C(ii)) on serial dependence, so the bound on $R_{3,it}$ in SM Lemma 3 does not follow (example in Section 3).	Bai’s Assumption C, with independence across units (Assumption 2*).
SM Assumption 4.2	A typo drops $1/N_{co}$ inside the inverse in a_{ik} , which SM Lemma 2 includes. As printed, the condition can hold when β is not identified (Section 7).	$a_{ik} = \lambda'_i(\Lambda'_{co}\Lambda_{co}/N_{co})^{-1}\lambda_k$ (Assumption 3*).
SM A.3	States that a unique solution exists and that the iteration “can lead to” it. The theory concerns the global minimizer, which the iteration is not guaranteed to reach.	Restated (Section 8).
SM Lemma 1(ii)	Treats the number of post-treatment periods q as fixed without saying so.	q fixed (Assumption 4*); Lemma 3* .
SM Lemmas 3(ii) and 4	Lemma 4 states $o_p(T_0^{-1})$ but its proof shows $O_p(T^{-1})$.	Replaced by one identity (Lemma 4*); Lemma 4 is not needed.
SM Proposition 1	Passes from convergence in probability to convergence of the conditional mean by asserting uniform integrability, which the printed moment conditions do not establish.	Proposition 1* , with the conditional-mean statement under an added condition.

2. Setting and corrected assumptions

The notation is that of the article. There are N_{co} control units in $\mathcal{C}$ and N_{tr} treated units in $\mathcal{T}$, observed for $T = T_0 + q$ periods; the treated units are untreated in periods $1, \dots, T_0$ and treated afterwards. For a T -row matrix or vector, the superscript 0 keeps the first T_0 rows. Outcomes follow [Assumption 1](#) of the article,

$$Y_{it} = \delta_{it}D_{it} + x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it},$$

with a fixed, known number r of factors, $F = (f_1, \dots, f_T)'$ and Λ_{co} the $N_{co} \times r$ matrix of control loadings. Write $\mathcal{G} = (D, X, \Lambda, F)$, $N = N_{co}$ inside the control-group results, $\delta_{NT} = \min(\sqrt{N_{co}}, \sqrt{T})$, $\|\cdot\|$ for the Frobenius norm and $\|\cdot\|_{op}$ for the spectral norm. The estimator is the article’s: Step 1 fits the interactive fixed effects model to the controls with $\hat{F}'\hat{F}/T = I_r$, giving $(\hat{\beta}, \hat{F})$; Step 2 sets $\hat{\lambda}_i = (\hat{F}^{0'}\hat{F}^0)^{-1}\hat{F}^{0'}(Y_i^0 - X_i^0\hat{\beta})$ for $i \in \mathcal{T}$; Step 3 predicts $\hat{Y}_{it}(0) = x'_{it}\hat{\beta} + \hat{\lambda}'_i\hat{f}_t$ for $t > T_0$. When $\hat{F}^{0'}\hat{F}^0$ is singular, which can happen with positive probability in a finite sample, Step 2 is defined with the Moore–Penrose inverse; the identities below are stated on the event that the matrix is invertible, whose probability tends to one ([Lemma 3*](#)).

The corrected assumptions replace [Assumptions 2–4](#) of the article and SM. [Assumption 1](#) is unchanged; [Assumption 5](#) is needed only for inference and is replaced in [Section 4](#).

Assumption 1* (Strict exogeneity). The array of errors $\{\varepsilon_{it}\}$ is independent of $\mathcal{G}$.

Assumption 2* (The error term). The error vectors $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$, $i \in \mathcal{C} \cup \mathcal{T}$, are independent across units. For a constant $M < \infty$ not depending on N_{co} or T :

- (a) $\mathbb{E}(\varepsilon_{it}) = 0$ and $\mathbb{E}|\varepsilon_{it}|^8 \leq M$;
- (b) with $\Omega_i = \mathbb{E}(\varepsilon_i \varepsilon_i')$, the largest eigenvalue of Ω_i is at most M for every unit and every T , and $|\mathbb{E}(\varepsilon_{it} \varepsilon_{is})| \leq \tau_{ts}$ for all i , with $T^{-1} \sum_{t,s} \tau_{ts} \leq M$;
- (c) for every (t, s) , $\mathbb{E} \left| N_{co}^{-1/2} \sum_{i \in \mathcal{C}} [\varepsilon_{is} \varepsilon_{it} - \mathbb{E}(\varepsilon_{is} \varepsilon_{it})] \right|^4 \leq M$;
- (d) the two fourth-moment conditions of SM Assumption 3.3 hold for the controls.

Assumption 1* writes the article's strict exogeneity jointly, as the proofs use it; Bai (2009, Assumption D) uses the article's notation. Assumption 2* strengthens the SM's uncorrelatedness across units to independence. For the controls, it implies Bai (2009, Assumption C). Part (b) is the condition SM Assumption 3 omitted. It rules out an error that is constant over time and allows stationary autoregressive errors.

Assumption 3* (Regularity). (a) $\mathbb{E} \|x_{it}\|^4 \leq M$ for all units and periods. Let $\mathcal{F} = \{F : F'F/T = I_r\}$; then, for some constant $c > 0$, $\Pr\{\inf_{F \in \mathcal{F}} \lambda_{\min}(D(F)) \geq c\} \rightarrow 1$, where

$$D(F) = \frac{1}{N_{co}T} \sum_{i \in \mathcal{C}} Z_i' Z_i, \quad Z_i = M_F X_i - \frac{1}{N_{co}} \sum_{k \in \mathcal{C}} M_F X_k a_{ik}, \quad a_{ik} = \lambda_i' \left(\frac{\Lambda_{co}' \Lambda_{co}}{N_{co}} \right)^{-1} \lambda_k.$$

- (b) $\mathbb{E} \|f_t\|^4 \leq M$ and $F'F/T \rightarrow_p \Sigma_F$, positive definite, as $T \rightarrow \infty$.
- (c) $\mathbb{E} \|\lambda_i\|^4 \leq M$ and $\Lambda_{co}' \Lambda_{co}/N_{co} \rightarrow_p \Sigma_\Lambda$, positive definite, as $N_{co} \rightarrow \infty$.

Part (a) is Bai (2009, Assumption A), with positive definiteness stated as a lower bound that holds with probability approaching one, as his proofs use it. Parts (b) and (c) are his Assumption B.

Assumption 4* (Asymptotic regime). $N_{co} \rightarrow \infty$ and $T_0 \rightarrow \infty$, with $q = T - T_0$, N_{tr} , the treated units' loadings, covariates and effects held fixed.

Remark 1. The treated units may have different adoption times; every result below holds for each treated unit with its own $T_{0,i}$, provided each $q_i = T - T_{0,i}$ is fixed.

3. Corrected results and proofs

3.1. Results used from Bai (2009)

Let V_{NT} be the $r \times r$ diagonal matrix of the r largest eigenvalues of $(N_{co}T)^{-1} \sum_{i \in \mathcal{C}} (Y_i - X_i \hat{\beta})(Y_i - X_i \hat{\beta})'$, so that this matrix times $\hat{F}$ equals $\hat{F}V_{NT}$, and let

$$H = \left(\frac{\Lambda'_{co} \Lambda_{co}}{N_{co}} \right) \left(\frac{F' \hat{F}}{T} \right) V_{NT}^{-1}, \quad E = \hat{F} - FH, \quad e_t = \hat{f}_t - H' f_t,$$

where e'_t is row t of E .

Lemma 1* (**Bai, 2009**). Under Assumptions 1 and **1*-3***, as $N_{co}, T \rightarrow \infty$: (i) $\hat{\beta} - \beta \rightarrow_p 0$; (ii) $V_{NT} \rightarrow_p V$, a diagonal matrix of the eigenvalues of $\Sigma_\Lambda \Sigma_F$, which is positive definite; (iii) $T^{-1} \|E\|^2 = O_p(\|\hat{\beta} - \beta\|^2) + O_p(\delta_{NT}^{-2})$; (iv) $\|H\| = O_p(1)$ and $\|H^{-1}\| = O_p(1)$.

Proof. Parts (i)–(iii) are **Bai (2009, Proposition 1(i) and Proposition A.1)**, whose Assumptions A–D are implied by Assumptions **1*-3*** for the control units. For (iv),

$$\left\| F' \hat{F} / T \right\| \leq (\|F\| / \sqrt{T}) (\|\hat{F}\| / \sqrt{T}) = O_p(1),$$

so $\|H\| = O_p(1)$ by (ii) and Assumption **3*(c)**. Next, Proposition 1(ii) of **Bai (2009)** gives $\|P_{\hat{F}} - P_F\| \rightarrow_p 0$. Since $\hat{F}' P_{\hat{F}} \hat{F} / T = I_r$,

$$\left(\frac{\hat{F}' F}{T} \right) \left(\frac{F' F}{T} \right)^{-1} \left(\frac{F' \hat{F}}{T} \right) = \frac{\hat{F}' P_F \hat{F}}{T} = I_r + \frac{\hat{F}' (P_F - P_{\hat{F}}) \hat{F}}{T} \rightarrow_p I_r,$$

because $\left\| \hat{F}' (P_F - P_{\hat{F}}) \hat{F} / T \right\| \leq r \|P_F - P_{\hat{F}}\|$. With $F' F / T \rightarrow_p \Sigma_F$ positive definite, the smallest singular value of $F' \hat{F} / T$ is bounded away from zero with probability approaching one, so $\left\| (F' \hat{F} / T)^{-1} \right\| = O_p(1)$. Then $\|H^{-1}\| \leq \|V_{NT}\| \left\| (F' \hat{F} / T)^{-1} \right\| \left\| (\Lambda'_{co} \Lambda_{co} / N_{co})^{-1} \right\| = O_p(1)$. $\square$

For the coefficient estimate, the results below use only Lemma **1*(i)**, not the expansion quoted in SM Lemma 2 (**Bai's** Corollary A.1, which holds under $T/N^2 \rightarrow 0$).

3.2. The factor estimate in one period

This lemma corrects SM Lemma 3(i).

Lemma 2*. Under Assumptions 1 and **1*-3***, for any deterministic sequence of dates $t = t_{NT} \in \{1, \dots, T\}$,

$$\left\| \hat{f}_t - H' f_t \right\| = O_p(\|\hat{\beta} - \beta\|) + O_p(N_{co}^{-1/2}) + O_p(T^{-1/2}),$$

and therefore $\left\| \hat{f}_t \right\| = O_p(1)$.

Proof. Let $d = \beta - \hat{\beta}$ and $W_i = Y_i - X_i \hat{\beta} = F \lambda_i + \varepsilon_i + X_i d$ for $i \in \mathcal{C}$. Expanding $W_i W_i'$ in the eigen-equation $(N_{co} T)^{-1} \sum_{i \in \mathcal{C}} W_i W_i' \hat{F} = \hat{F} V_{NT}$ gives

$$\hat{F} V_{NT} = F \left(\frac{\Lambda'_{co} \Lambda_{co}}{N_{co}} \right) \left(\frac{F' \hat{F}}{T} \right) + \sum_{k=1}^8 J_k,$$

where, with all sums over $i \in \mathcal{C}$ and $N = N_{co}$,

$$\begin{aligned} J_1 &= \frac{1}{NT} \sum X_i d d' X_i' \hat{F}, & J_2 &= \frac{1}{NT} \sum X_i d \lambda_i' F' \hat{F}, & J_3 &= \frac{1}{NT} \sum X_i d \varepsilon_i' \hat{F}, & J_4 &= \frac{1}{NT} \sum F \lambda_i d' X_i' \hat{F}, \\ J_5 &= \frac{1}{NT} \sum \varepsilon_i d' X_i' \hat{F}, & J_6 &= \frac{1}{NT} \sum F \lambda_i \varepsilon_i' \hat{F}, & J_7 &= \frac{1}{NT} \sum \varepsilon_i \lambda_i' F' \hat{F}, & J_8 &= \frac{1}{NT} \sum \varepsilon_i \varepsilon_i' \hat{F}. \end{aligned}$$

Multiplying on the right by V_{NT}^{-1} gives $\hat{F} = FH + \sum_k J_k V_{NT}^{-1}$. Row t , transposed, is

$$\hat{f}_t - H' f_t = V_{NT}^{-1} \sum_{k=1}^8 j_{k,t}, \quad j_{k,t} = J_k' \iota_t, \quad (1)$$

where ι_t is the t -th unit vector in $\mathbb{R}^T$. (SM Lemma 3 writes the left-hand side as $H^{-1} \hat{f}_t - f_t$, which transposes H .) Since $\|V_{NT}^{-1}\| = O_p(1)$ by Lemma 1*(ii), it suffices to bound each $j_{k,t}$. Throughout, $\|\hat{F}\|/\sqrt{T} = \sqrt{r}$, $\|F\|/\sqrt{T} = O_p(1)$, $\|f_t\| = O_p(1)$, and averages over controls such as $N^{-1} \sum_i (\|X_i\|/\sqrt{T}) \|x_{it}\|$, $N^{-1} \sum_i \|\lambda_i\| \|x_{it}\|$ and $N^{-1} \sum_i (\|\varepsilon_i\|/\sqrt{T}) \|x_{it}\|$ are $O_p(1)$ by the Cauchy–Schwarz inequality and the fourth-moment bounds of Assumptions 2* and 3*.

Terms with d . Using $j_{1,t} = (NT)^{-1} \sum \hat{F}' X_i d d' x_{it}$ and the analogous expressions,

$$\begin{aligned} \|j_{1,t}\| &\leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{1}{N} \sum \frac{\|X_i\|}{\sqrt{T}} \|x_{it}\| \|d\|^2, & \|j_{2,t}\| &\leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{\|F\|}{\sqrt{T}} \frac{1}{N} \sum \|\lambda_i\| \|x_{it}\| \|d\|, \\ \|j_{3,t}\| &\leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{1}{N} \sum \frac{\|\varepsilon_i\|}{\sqrt{T}} \|x_{it}\| \|d\|, & \|j_{4,t}\| &\leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{1}{N} \sum \frac{\|X_i\|}{\sqrt{T}} \|\lambda_i\| \|f_t\| \|d\|, \\ \|j_{5,t}\| &\leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{1}{N} \sum \frac{\|X_i\|}{\sqrt{T}} |\varepsilon_{it}| \|d\|. \end{aligned}$$

Each is $O_p(\|d\|)$, since $\|d\| = o_p(1)$ by Lemma 1*(i).

Terms $j_{6,t}$ and $j_{7,t}$. Let $u_s = N^{-1} \sum_{i \in \mathcal{C}} \lambda_i \varepsilon_{is}$. By Assumptions 1* and 2*, the ε_{is} are independent across i , mean zero and independent of Λ , so $\mathbb{E} \|u_s\|^2 = N^{-2} \sum_i \mathbb{E} \|\lambda_i\|^2 \mathbb{E} \varepsilon_{is}^2 \leq C/N$ for every s and a constant C . Hence $T^{-1} \sum_s \|u_s\|^2 = O_p(N^{-1})$ by Markov's inequality, and $\|u_t\| = O_p(N^{-1/2})$. Now $j_{6,t} = T^{-1} \sum_s \hat{F}' u_s' f_t$ and $j_{7,t} = (\hat{F}' F/T) u_t$, so by the Cauchy–Schwarz inequality

$$\|j_{6,t}\| \leq \frac{\|\hat{F}\|}{\sqrt{T}} \left(\frac{1}{T} \sum_s \|u_s\|^2 \right)^{1/2} \|f_t\| = O_p(N^{-1/2}), \quad \|j_{7,t}\| \leq \frac{\|\hat{F}\|}{\sqrt{T}} \frac{\|F\|}{\sqrt{T}} \|u_t\| = O_p(N^{-1/2}).$$

(SM Lemma 3 bounds these terms by a product containing $\|\varepsilon\|/\sqrt{T}$, which is $O_p(\sqrt{N})$, and by the inequality $\|\sum_i v_i\| \leq (\sum_i \|v_i\|^2)^{1/2}$, which is false in general.)

Term $j_{8,t}$. Write $j_{8,t} = T^{-1} \sum_s \hat{f}_s N^{-1} \sum_i \varepsilon_{is} \varepsilon_{it} = b_1 + b_2$, with

$$b_1 = \frac{1}{T} \sum_s \hat{f}_s \zeta_{st}, \quad \zeta_{st} = \frac{1}{N} \sum_{i \in \mathcal{C}} [\varepsilon_{is} \varepsilon_{it} - \mathbb{E}(\varepsilon_{is} \varepsilon_{it})], \quad b_2 = \frac{1}{T} \sum_s \hat{f}_s \bar{\omega}_{st}, \quad \bar{\omega}_{st} = \frac{1}{N} \sum_{i \in \mathcal{C}} \mathbb{E}(\varepsilon_{is} \varepsilon_{it}).$$

By Assumption 2*(c), $\mathbb{E} \zeta_{st}^2 \leq M^{1/2}/N$ for every (s, t) , so $\|b_1\| \leq (\|\hat{F}\|/\sqrt{T})(T^{-1} \sum_s \zeta_{st}^2)^{1/2} = O_p(N^{-1/2})$. For b_2 , let $\bar{\Omega} = N^{-1} \sum_{i \in \mathcal{C}} \Omega_i$, whose (s, t) entry is $\bar{\omega}_{st}$. Then $b_2 = T^{-1} \hat{F}' \bar{\Omega} \iota_t$ and, by Assumption 2*(b),

$$\|b_2\| \leq \frac{1}{T} \|\hat{F}\| \|\bar{\Omega}\|_{\text{op}} \|\iota_t\| \leq \frac{\sqrt{rT}}{T} \frac{1}{N} \sum_{i \in \mathcal{C}} \|\Omega_i\|_{\text{op}} \leq \sqrt{r} M T^{-1/2}.$$

(SM Lemma 3 obtains $O_p(T^{-1/2})$ from a bound on each covariance alone; that step needs the eigenvalue or summability condition.)

Every constant in these bounds is uniform in the date, so they hold along any sequence t_{NT} . Combining the eight bounds in (1) proves the rate. Finally, $\|\hat{f}_t\| \leq \|H\| \|f_t\| + \|e_t\| = O_p(1)$ by Lemma 1*(iv). $\square$

3.3. The pretreatment Gram matrix

This lemma makes explicit the fixed- q regime used in the proof of SM Lemma 1(ii).

Lemma 3*. Under Assumptions 1 and 1*-4*, $\hat{F}^{0'} \hat{F}^0$ is invertible with probability approaching one, and $T \left\| (\hat{F}^{0'} \hat{F}^0)^{-1} \right\|_{\text{op}} = O_p(1)$.

Proof. Let $Q = \sum_{s=T_0+1}^T \hat{f}_s \hat{f}_s'$. It has q terms, q is fixed, and each $\|\hat{f}_{T_0+h}\| = O_p(1)$, $h = 1, \dots, q$, by Lemma 2* applied to the sequence $t = T_0 + h$; a finite union bound gives $\|Q\|_{\text{op}} = O_p(1)$. Since $\hat{F}' \hat{F} = T I_r$, $\hat{F}^{0'} \hat{F}^0 = T I_r - Q = T(I_r - Q/T)$, and the smallest eigenvalue of $I_r - Q/T$ is at least $1 - \|Q\|_{\text{op}}/T$, which exceeds $1/2$ with probability approaching one. On that event $\left\| (\hat{F}^{0'} \hat{F}^0)^{-1} \right\|_{\text{op}} \leq 2/T$. $\square$

Without a bound on q the lemma can fail. For example, if $T_0 = \lfloor \sqrt{T} \rfloor$ and f_t is a linear trend normalized so that $F' F/T = 1$, then $F^{0'} F^0$ is of order $T^{-1/2}$, and $T \left\| (\hat{F}^{0'} \hat{F}^0)^{-1} \right\|_{\text{op}}$ grows like $T^{3/2}$ even when the factor is known.

3.4. The projection onto the pretreatment factors

This lemma replaces SM Lemma 3(ii) and Lemma 4. It bounds directly the quantity that enters the prediction error.

Lemma 4*. Under Assumptions 1 and **1*-4***, for $t = T_0 + h$ with $h \in \{1, \dots, q\}$ fixed,

$$\left\| f_t - F^{0'} \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t \right\| = O_p(\left\| \hat{\beta} - \beta \right\|) + O_p(\delta_{NT}^{-1}).$$

Proof. From $\hat{F} = FH + E$ and the invertibility of H (Lemma **1***), the first T_0 rows give $F^0 = (\hat{F}^0 - E^0)H^{-1}$, and row t gives $f_t = (H')^{-1}(\hat{f}_t - e_t)$. On the event of Lemma **3***,

$$F^{0'} \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t = (H')^{-1}(\hat{F}^0 - E^0)' \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t = (H')^{-1}[\hat{f}_t - E^{0'} \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t].$$

Subtracting from f_t ,

$$f_t - F^{0'} \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t = (H')^{-1}[E^{0'} \hat{F}^0 (\hat{F}^{0'} \hat{F}^0)^{-1} \hat{f}_t - e_t].$$

Hence the norm is at most

$$\|H^{-1}\| \left(\|E\| \left\| \hat{F}^0 \right\| \left\| (\hat{F}^{0'} \hat{F}^0)^{-1} \right\|_{\text{op}} \left\| \hat{f}_t \right\| + \|e_t\| \right).$$

Here $\|E\| = \sqrt{T} [O_p(\left\| \hat{\beta} - \beta \right\|) + O_p(\delta_{NT}^{-1})]$ by Lemma **1***(iii), $\left\| \hat{F}^0 \right\| \leq \left\| \hat{F} \right\| = \sqrt{rT}$, $\left\| (\hat{F}^{0'} \hat{F}^0)^{-1} \right\|_{\text{op}} = O_p(T^{-1})$ by Lemma **3***, and $\left\| \hat{f}_t \right\| = O_p(1)$ and $\|e_t\| = O_p(\left\| \hat{\beta} - \beta \right\|) + O_p(\delta_{NT}^{-1})$ by Lemma **2***. With $\|H^{-1}\| = O_p(1)$, the bound follows. $\square$

3.5. The limit of the estimation error

Proposition 1*. Under Assumptions 1 and **1*-4***, for each $i \in \mathcal{T}$ and $t = T_0 + h$ with $h \in \{1, \dots, q\}$ fixed,

$$\hat{\delta}_{it} - \delta_{it} = \varepsilon_{it} + R_{1,it} + R_{2,it} + R_{3,it}, \quad (2)$$

where

- (i) $R_{1,it} + R_{2,it} = O_p(\left\| \hat{\beta} - \beta \right\|) + O_p(\delta_{NT}^{-1})$, and $R_{1,it}$ and $R_{2,it}$ are functions of the controls' data and $\mathcal{G}$ only;
- (ii) $\mathbb{E}(R_{3,it} \mid \varepsilon_{\mathcal{C}}, \mathcal{G}) = 0$ and $\mathbb{E}(R_{3,it}^2 \mid \varepsilon_{\mathcal{C}}, \mathcal{G}) \leq M \|w_{it}\|^2$ on the event of Lemma **3***, with w_{it} defined below and $\|w_{it}\|^2 = O_p(T_0^{-1})$; hence $R_{3,it} = O_p(T_0^{-1/2})$.

Consequently $\hat{\delta}_{it} - \delta_{it} - \varepsilon_{it} = O_p(\left\| \hat{\beta} - \beta \right\|) + O_p(N_{co}^{-1/2}) + O_p(T_0^{-1/2}) = o_p(1)$, and

$$\widehat{ATT}_t - ATT_t - \frac{1}{N_{tr}} \sum_{i \in \mathcal{T}} \varepsilon_{it} \rightarrow_p 0.$$

If, in addition, $Z_{it} = \hat{\delta}_{it} - \delta_{it} - \varepsilon_{it}$ satisfies $\mathbb{E}(|Z_{it}| \mid \mathcal{G}) < \infty$ almost surely at each sample size and is uniformly integrable conditional on $\mathcal{G}$, in the sense that $\lim_{K \rightarrow \infty} \limsup \Pr\{\mathbb{E}(|Z_{it}| \mathbf{1}\{|Z_{it}| > K\} \mid \mathcal{G}) > \eta\} = 0$ for every $\eta > 0$, then $\mathbb{E}_\varepsilon(\widehat{ATT}_t \mid \mathcal{G}) - ATT_t \rightarrow_p 0$.

Proof. Decomposition. Substituting $Y_i^0 = X_i^0\beta + F^0\lambda_i + \varepsilon_i^0$ into Step 2 and Step 3, as in the SM,

$$\begin{aligned} R_{1,it} &= [x'_{it} - \hat{f}'_t(\hat{F}^{0'}\hat{F}^0)^{-1}\hat{F}^{0'}X_i^0](\beta - \hat{\beta}), \\ R_{2,it} &= \lambda'_i[f_t - F^{0'}\hat{F}^0(\hat{F}^{0'}\hat{F}^0)^{-1}\hat{f}_t], \\ R_{3,it} &= -w'_{it}\varepsilon_i^0, \quad w_{it} = \hat{F}^0(\hat{F}^{0'}\hat{F}^0)^{-1}\hat{f}_t. \end{aligned}$$

The estimates $\hat{\beta}$ and $\hat{F}$ use the controls only, so $R_{1,it}$, $R_{2,it}$ and w_{it} are functions of $(\varepsilon_{\mathcal{C}}, \mathcal{G})$.

Part (i). By Lemma 3* and Lemma 2*,

$$\begin{aligned} \|\hat{f}'_t(\hat{F}^{0'}\hat{F}^0)^{-1}\hat{F}^{0'}X_i^0\| &\leq \|\hat{f}_t\| \left\| (\hat{F}^{0'}\hat{F}^0)^{-1} \right\|_{\text{op}} \|\hat{F}^0\| \|X_i^0\| \\ &= O_p(1) \cdot O_p(T^{-1}) \cdot \sqrt{rT} \cdot O_p(\sqrt{T_0}) = O_p(1), \end{aligned}$$

using $\mathbb{E}\|x_{it}\|^4 \leq M$ for the treated unit. Hence $|R_{1,it}| = O_p(\|\hat{\beta} - \beta\|)$. By Lemma 4*, $|R_{2,it}| \leq \|\lambda_i\| [O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})]$.

Part (ii). By Assumptions 1* and 2*, ε_i is independent of $(\varepsilon_{\mathcal{C}}, \mathcal{G})$ for $i \in \mathcal{T}$. Since w_{it} is a function of $(\varepsilon_{\mathcal{C}}, \mathcal{G})$,

$$\mathbb{E}(R_{3,it} \mid \varepsilon_{\mathcal{C}}, \mathcal{G}) = -w'_{it}\mathbb{E}(\varepsilon_i^0) = 0, \quad \mathbb{E}(R_{3,it}^2 \mid \varepsilon_{\mathcal{C}}, \mathcal{G}) = w'_{it}\Omega_i^0 w_{it} \leq \|\Omega_i\|_{\text{op}} \|w_{it}\|^2 \leq M \|w_{it}\|^2,$$

where Ω_i^0 is the leading $T_0 \times T_0$ block of Ω_i , whose largest eigenvalue is at most that of Ω_i . Moreover $\|w_{it}\|^2 = \hat{f}'_t(\hat{F}^{0'}\hat{F}^0)^{-1}\hat{f}_t \leq \|\hat{f}_t\|^2 \left\| (\hat{F}^{0'}\hat{F}^0)^{-1} \right\|_{\text{op}} = O_p(T^{-1})$. For any $c > 0$ and $K > 0$, the conditional Chebyshev inequality gives

$$P(|R_{3,it}| > cT^{-1/2}) \leq P(T\|w_{it}\|^2 > K) + \frac{MK}{c^2},$$

plus the probability that $\hat{F}^{0'}\hat{F}^0$ is singular, which tends to zero; choosing K and then c shows $R_{3,it} = O_p(T^{-1/2})$, which is $O_p(T_0^{-1/2})$ since q is fixed. (The SM bounds $\mathbb{E}(R_{3,it}^2)$ with Rosenthal's inequality, which needs independence over time, and with a bound on each covariance. Under SM Assumption 3, take a known constant factor, $f_t = 1$, so that $w_{it} = T_0^{-1}\mathbf{1}$, and a treated error $\varepsilon_{is} = U_i$ for all s with $U_i = \pm 1$ with equal probability; then $R_{3,it} = -U_i$ and $\mathbb{E}(R_{3,it}^2) = 1$ for every T_0 .)

Conclusion. Parts (i) and (ii), Lemma 1*(i) and $\delta_{NT}^{-1} \leq N_{co}^{-1/2} + T^{-1/2}$ give the rate for $\hat{\delta}_{it} - \delta_{it} - \varepsilon_{it}$; averaging over the fixed set $\mathcal{T}$ gives the statement for $\widehat{ATT}_t$. For the last claim, $\mathbb{E}(\varepsilon_{it} \mid \mathcal{G}) = 0$ by Assumption 1*; for $K > 0$, $|\mathbb{E}(Z_{it} \mid \mathcal{G})| \leq \mathbb{E}(\min(|Z_{it}|, K) \mid \mathcal{G}) + \mathbb{E}(|Z_{it}|\mathbf{1}\{|Z_{it}| > K\} \mid \mathcal{G})$, where the first term tends to zero in probability because $\mathbb{E}\min(|Z_{it}|, K) \rightarrow 0$ by bounded convergence and Markov's inequality, and the second is small in probability for large K by the condition. $\square$

Remark 2. The uniform integrability in the last claim is an added condition. The SM asserted it

“by the moment conditions specified in Assumptions 3 and 4,” but those conditions bound moments of the data, while the terms in question contain the random inverses $(\hat{F}^{0'}\hat{F}^0)^{-1}$ and $D(\hat{F})^{-1}$, whose moments need not exist. The method uses only the statement in probability, which needs no such condition. Likewise, part (ii) bounds the variance of $R_{3,it}$ conditional on the control data; it does not bound $\mathbb{E}(R_{3,it}^2)$.

4. Inference: Algorithm 2

The article states: “Under Assumptions 1–5, this procedure provides valid uncertainty estimates for the proposed method without making particular distributional assumptions of the error terms.” The assumptions do not ensure general confidence-interval coverage, for two reasons.

Equal covariances do not give equal distributions. By Proposition 1*, the estimation error of $\widehat{ATT}_t$ converges to $N_{tr}^{-1} \sum_{i \in \mathcal{T}} \varepsilon_{it}$, whose distribution is set by the treated units’ errors. With N_{tr} fixed, the estimator need not be consistent, and an interval for ATT_t is a prediction interval whose width does not shrink. Step 1 of Algorithm 2 approximates it with prediction errors of held-out control units, which converge to the controls’ errors. Assumption 5 equates second moments only. Suppose, for example, that the control errors are standard normal and that one treated unit’s error equals 0 with probability 0.9 and $\pm\sqrt{10}$ with probability 0.05 each, independently over time. The variances agree and Assumptions 1–5 hold, but a nominal 95% interval built from the control errors covers with probability $\Pr(|\varepsilon_{it}| \leq 1.96) = 0.90$. Bai and Ng (2021, p. 1756) make the same point for their estimator, and they estimate the shock’s distribution from the residuals by assuming that the shocks are identically distributed across units and periods. A sufficient condition for matching the treated and control error distributions is:

Assumption 5* (Exchangeable errors). The error vectors ε_i , $i \in \mathcal{C} \cup \mathcal{T}$, are identically distributed, in addition to Assumptions 1* and 2*.

The interval’s orientation. Steps 3 and 4 add $\widehat{ATT}_t$ to bootstrap estimates computed with the effect set to zero and take percentiles of the result, giving $[\widehat{ATT}_t + \hat{q}_{\alpha/2}, \widehat{ATT}_t + \hat{q}_{1-\alpha/2}]$, where $\hat{q}_p$ are quantiles of the simulated estimation error U^* . If U^* reproduces the distribution of the estimation error $U = \widehat{ATT}_t - ATT_t$ exactly, this interval covers ATT_t with probability $\Pr(-q_{1-\alpha/2} \leq U \leq -q_{\alpha/2})$, which is not $1 - \alpha$ in general; symmetry of the distribution of U , with continuous quantiles, is sufficient. For $U = Z - 1$ with Z standard exponential, the coverage of a nominal 95% interval is $1 - e^{-1.975} \approx 0.86$. The basic interval inverts the distribution of U (with exact quantiles of a continuous distribution its coverage is $1 - \alpha$, and at least $1 - \alpha$ with atoms):

$$[\widehat{ATT}_t - \hat{q}_{1-\alpha/2}, \widehat{ATT}_t - \hat{q}_{\alpha/2}].$$

The variance in Step 4 is unaffected, so intervals of the form $\widehat{ATT}_t \pm z_{1-\alpha/2} \widehat{se}$ are unaffected as well; they rely on approximate normality of U .

Corrected statement. The quoted sentence should say that the article’s Monte Carlo evidence (independent normal errors) shows close to nominal coverage; that under Assumption 5*, Step 1 targets the right distribution; and that Step 4 should use the basic interval, or a normal interval when a normal approximation is justified. A proof that the full algorithm, with its refitting steps, gives asymptotically valid intervals is not given here.

5. Cross-validation: Algorithm 1

The article supports its claim about Algorithm 1 (Table 1) with Monte Carlo evidence (SM Table A5, Figure A4) in which errors are independent over time. With serially correlated errors, which Assumption 2* allows, the held-out error is correlated with the errors used to refit the loadings, so the validation score can differ from the prediction error it estimates. In a simulation with $N_{co} = 40$, $N_{tr} = 5$, $T_0 = 20$, two true factors, two covariates and AR(1) errors with coefficient ρ (60 replications per value, ranks 1 to 4), the procedure chose two factors in 82% of replications at $\rho = 0$, 23% at $\rho = 0.5$ and 2% at $\rho = 0.8$, and more than two in 18%, 75% and 97%. This finite-sample experiment does not show that the procedure fails as the sample grows. The results in Section 3 take the number of factors as known and do not cover the number chosen by Algorithm 1.

6. Remarks 1 and 2

Remark 1. Remark 1 follows Bai (2009, pp. 1243, 1263), who notes that strict exogeneity rules out lagged dependent variables only to simplify the proofs, requires serially uncorrelated errors in that case, and says that the consistency argument carries over. Moon and Weidner (2015, Theorem 4.1, p. 1554) prove that $\hat{\beta}$ is consistent with predetermined regressors such as a lagged outcome. Remark 1 is therefore supported for $\hat{\beta}$. The treatment-effect results here assume strict exogeneity and do not cover this case. The counterfactual raises a further issue: after treatment, a treated unit’s lagged outcome is a treated outcome, so predicting $Y_{it}(0)$ requires imputing the untreated lag recursively. The sentence should read: “Under suitable identification and regularity conditions, the coefficients remain consistent in dynamic models with lagged outcomes when the errors are not serially correlated (Moon and Weidner, 2015); the treatment-effect results in this paper assume strict exogeneity and do not cover such models.”

Remark 2. The limit should read “as $N_{co}, T_0 \rightarrow \infty$ ”. The explanation of bias should read: “Under the corrected assumptions, the term from estimating each treated unit’s loadings has mean zero given the control data, and its conditional variance is of order T_0^{-1} . Bias can arise from estimating β and the factors; Proposition 1* gives the conditions under which it vanishes.”

7. Identification of β

As printed, SM Assumption 4.2 defines $a_{ik} = \lambda'_i(\Lambda'_{co}\Lambda_{co})^{-1}\lambda_k$ while keeping the factor $1/N_{co}$ in front of the sum in S_i . The second term of S_i is then of order $1/N_{co}$, and the condition reduces to $\inf_F N_{co}^{-1}T^{-1}\sum_i X'_i M_F X_i > 0$, which does not ensure identification. The following example satisfies the printed condition. Let T be even, $r = 1$, one covariate, no errors, $h_t = 1$, $g_t = (-1)^t$, and control loadings and covariates $\lambda_i = 1$, $X_i = h - \mu_i g$ with $\mu_i \in \{-1, 1\}$ and $\sum_{i \in C} \mu_i = 0$, so that $Y_i = h$ when $\beta = 0$ and $F = h$. The parameters $(\beta, F, \lambda_i) = (0, h, 1)$ and $(1, g, \mu_i)$ fit the controls exactly. With the printed a_{ik} , $\inf_F D(F)$ stays bounded away from zero; a treated unit with loading 1, $X_i^0 = h^0$ and $x_{iT} = 0$ then has a prediction error of 1 at the second solution, for every sample size. With the corrected a_{ik} of Assumption 3*(a), $Z_i = 0$ for $F = g$, so $D(g) = 0$ and the condition correctly fails.

8. Estimation (SM A.3)

The SM's claim that a unique solution exists (Table 1) concerns the global minimizer of the least-squares problem in Step 1, as does the theory here and in Bai (2009). That problem is not convex, and the alternating iteration is not guaranteed to reach its global minimum. Under the normalization $\hat{F}'\hat{F}/T = I_r$ with $\hat{\Lambda}'_{co}\hat{\Lambda}_{co}$ diagonal, $\hat{F}$ is unique for a given β up to the signs of its columns when the r largest eigenvalues of the residual matrix are distinct and exceed the $(r+1)$ -th, with columns ordered by eigenvalue. The claim should say instead that a solution exists and that the iteration should be started from several values.

9. Implications for gsynth and fecl

The methods of the article are implemented in the R packages gsynth and fecl. gsynth had its own implementation through version 1.2.1 (2021) and calls fecl from version 1.2.2 (2025). Almost none of the corrections change what the packages compute. The estimator is unchanged, and both packages report normal intervals by default, so the orientation issue in Section 4 does not affect their default output, although this alone does not establish the intervals' coverage. The two paragraphs below concern how the output is read, and the last one points to a separate discussion of software issues.

Bootstrap intervals. With the standard error computed from the bootstrap draws, gsynth through version 1.2.1 reported the normal interval $\widehat{ATT}_t \pm z_{1-\alpha/2} \widehat{se}$. fecl's default, `ci.method = "normal"`, is the same interval, and its alternative, `ci.method = "basic"`, added in version 2.4.2, has the orientation of the basic interval. The normal interval relies on equal treated and control error variances (the article's Assumption 5), the basic interval on identically distributed errors (Assumption 5*, a sufficient condition), and neither package checks them. Comparing the spread of the controls' residuals with that of the treated units' pretreatment residuals is a simple diagnostic.

With few treated units and non-normal shocks, the normal interval can also under-cover even when the standard error is right.

Cross-validation. The caution in Section 5 applies to gsynth through version 1.2.1, which implemented Algorithm 1 as described. fecl’s cross-validation, used by gsynth from version 1.2.2, holds out cells of the control units instead, and the simulation in Section 5 does not concern it.

Software issues. Some earlier versions of the packages worked in ways their documentation did not describe. Before gsynth 1.3.1 (December 2025), for example, the parametric bootstrap for the IFE-EM estimator (`EM = TRUE`) omitted Step 1 of Algorithm 2 and drew errors from in-sample residuals (Green and Aronow, 2026). Xu (2026) discusses this routine, and changes to the optional quantile-based intervals and p-values in fecl are recorded in the package’s release notes. I will provide a more comprehensive review of these issues separately.

10. Errata

This section lists slips in the article and the SM beyond the items in Table 1. Most concern notation, indexing and citations, and none changes a result.

Where	Correction
Article, footnote 10	Iterating $Y_{it} = \rho Y_{i,t-1} + x'_{it}\beta + \varepsilon_{it}$ gives $Y_{it} = \rho^t Y_{i0} + \sum_{s=1}^t \rho^{t-s} x'_{is}\beta + \nu_{it}$ with $\nu_{it} = \sum_{s=1}^t \rho^{t-s} \varepsilon_{is}$. The article omits the lagged covariate terms.
Article, footnote 11	With $\tilde{F} = FA$, the loadings are $\tilde{\lambda}_i = A^{-1}\lambda_i$ and $\tilde{\Lambda}_{co} = \Lambda_{co}(A^{-1})'$; the diagonal matrix is $A^{-1}\Lambda'_{co}\Lambda_{co}(A^{-1})'$.
Article, after Assumption 3	“The space spanned by F (or $F'F/T$)” should read “(or the projection $F(F'F)^{-1}F'$)”; $F'F/T$ is the identity by normalization.
Article, dimension of X	The number of covariates is k throughout (the article also writes p).
Article, Algorithm 1	F_{-s}^0 should be $\hat{F}_{-s}^0$, and $X_{i,-s}^{0'}\hat{\beta}$ should be $X_{i,-s}^0\hat{\beta}$.
Article, Algorithm 2	Step 3(a): the treated units are indexed by j on both sides; Step 4: B is B_2 .
SM Assumption 3.1	Implied by Assumption 4.1; the dependence restriction used here is Assumption 2*(b).
SM Assumption 4.2	“ $\mathbb{E} \ x_{it}\ ^4 \leq M$ ” is a separate condition; the sum in S_i runs over $k \in \mathcal{C}$.
SM A.3	The matrix whose eigenvectors give $\hat{F}$ is $T \times T$, not $N_{co} \times N_{co}$.
SM A.4	$\sum_i \xi_i = 0$ should be $\sum_t \xi_t = 0$; scalars subtracted from T_0 -vectors should be multiplied by $\mathbf{1}_{T_0}$.
SM Lemma 1(ii)	Q is positive semidefinite (rank at most q).

Where	Correction
SM Lemma 2	The leading term is $O_p((N_{co}T)^{-1/2})$, which is dominated by $O_p(N_{co}^{-1}) + O_p(T^{-1})$; $1/(NT)$ in ζ should be $1/(N_{co}T)$; the assumptions cited should include the standing Assumption 2 (strict exogeneity).
SM Lemma 3(i)	$(Y_i - X_i)$ should be $(Y_i - X_i\hat{\beta})$; ε' should be ε'_i ; in a_1 and a_4 , X'_i should be X_i ; with G as defined in the SM, the corrected expansion uses G' , and $\hat{f}_t = Gf_t$ should be $\tilde{f}_s = G'\hat{f}_s$ inside the sum; the bound on $\mathbb{E}(\varepsilon_{is}\varepsilon_{it})$ should cite Assumption 3 (the SM cites 4.2).
SM Lemma 3(ii)	The statement’s superscripts 0 should be dropped (the proof treats the full-sample quantity); in the closing claim for errors without serial correlation, the condition $\mathbb{E}(\varepsilon_{is}\varepsilon_{it}) = 0$, printed for all (s, t) , should be stated for $s \neq t$.
SM Lemma 4	The sum defining B starts at $s = T_0 + 1$.
SM Proposition 1	The bound $ \hat{f}'_t(\sum_l \hat{f}_l \hat{f}'_l)^{-1} \hat{f}_s \leq O(T_0^{-1})$ is a stochastic bound; the second moment of ε_{it} is from Assumption 4.1; $N_{tr}^{-1} \cdot o(1)$ should be $N_{tr}^{-1} \sum_{i \in \mathcal{T}} o(1) = o(1)$.
SM B.4	In the displayed IFE specification, $\sum_{t=T_0+1}^T \delta_t D_{it}$ should be $\delta_t D_{it}$, with one coefficient for each post-treatment period.
SM Table A3	The column labels N_{co} and N_{tr} are swapped: $N_{tr} = 20$ throughout and N_{co} varies from 40 to 200. The entries are unchanged.

References

- Bai, Jushan. 2009. “Panel Data Models with Interactive Fixed Effects.” *Econometrica* 77(4):1229–1279.
- Bai, Jushan and Serena Ng. 2021. “Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data.” *Journal of the American Statistical Association* 116(536):1746–1763.
- Green, Beniamino and P. M. Aronow. 2026. “Undocumented Behavior in the gsynth R Package and Its Consequences for Three Published Studies.” arXiv:2606.18113.
- Moon, Hyungsik Roger and Martin Weidner. 2015. “Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects.” *Econometrica* 83(4):1543–1579.
- Xu, Yiqing. 2017. “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models.” *Political Analysis* 25(1):57–76.
- Xu, Yiqing. 2026. “A Response to Green and Aronow (2026).” Working paper, April 26, 2026. <https://yiqingxu.org/public/gsynth-note.pdf>.