A. Supplementary Materials for

A Bayesian Alternative to Synthetic Control for Comparative Case Studies

Xun Pang (Tsinghua) Licheng Liu (M.I.T.) Yiqing Xu (Stanford)

Table of Contents

A.1. A Formal Proof of The Predictive Posterior Distribution

Under the assumption of exchangeability and by de Finetti's theorem (de Finetti 1963), given some parameters $\boldsymbol{\theta}$ and their prior distributions, we can write $\Pr(\mathbf{Y}(\mathbf{0})^{mis}, \mathbf{Y}(\mathbf{0})^{obs}, \mathbf{X}')$, $\mathbf{X}' = (\mathbf{X}, \mathbf{U})$ as $i.i.d.$:

$$\begin{aligned} \Pr(\mathbf{X}', \mathbf{Y}(\mathbf{0})) &= \int \left[\prod_{it \in S} f(y_{it}(c), \mathbf{X}'_{it} | \boldsymbol{\theta}) \right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int \left[\prod_{it \in S} f(y_{it}(c) | \mathbf{X}'_{it}, \boldsymbol{\theta}_{y.x'}) f(\mathbf{X}'_{it} | \boldsymbol{\theta}_{x'}) \right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \tag{A1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{y.x}, \boldsymbol{\theta}_{x'})$ and $\boldsymbol{\theta}_{x'} = (\boldsymbol{\theta}_x, \boldsymbol{\theta}_u)$, in which $\boldsymbol{\theta}_x$ are the parameters that govern the data-generating process (DGP) of $\mathbf{X}$, $\boldsymbol{\theta}_u$ are the parameters that govern the DGP of $\mathbf{U}$, and $\boldsymbol{\theta}_{y.x'}$ are the parameters that govern the DGP of $\mathbf{Y}(\mathbf{0})$ given $\mathbf{X}' = (\mathbf{X}, \mathbf{U})$. We also assume that $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_{y.x'})\pi(\boldsymbol{\theta}_{x'})$, i.e., the parameters that govern the DGPs of $\mathbf{X}'$ and $\mathbf{Y}$, are a prior independent. Equation (A1) can be written as follows:

$$\begin{aligned} &\int \prod_{it \in S_1} f(y_{it}(c)^{mis} | \mathbf{X}'_{it}, \boldsymbol{\theta}_{y.x'}) \prod_{it \in S_0} f(y_{it}(c)^{obs} | \mathbf{X}'_{it}, \boldsymbol{\theta}_{y.x'}) \pi(\boldsymbol{\theta}_{y.x'}) d\boldsymbol{\theta}_{y.x'} \times \int \prod_{it \in S} f(\mathbf{X}'_{it} | \boldsymbol{\theta}_{x'}) \pi(\boldsymbol{\theta}_{x'}) d\boldsymbol{\theta}_{x'} \\ &\propto \int \underbrace{\left(\prod_{it \in S_1} f(y_{it}(c)^{mis} | \mathbf{X}_{it}, \boldsymbol{\theta}'_{y.x'}) \right)}_{\text{posterior predictive distribution}} \times \underbrace{\left(\prod_{it \in S_0} f(y_{it}(c)^{obs} | \mathbf{X}_{it}, \boldsymbol{\theta}'_{y.x'}) \right)}_{\text{likelihood}} \pi(\boldsymbol{\theta}_{y.x'}) d\boldsymbol{\theta}_{y.x'} \end{aligned}$$

In the last step, we regard the unobserved covariates $\mathbf{U}$ as unknown parameters and denote $\boldsymbol{\theta}'_{y.x'} = (\boldsymbol{\theta}_{y.x'}, \mathbf{U})$. Returning to Equation (**??**) in the main text, we can write the posterior predictive distribution as

$$\Pr(\mathbf{Y}(\mathbf{0})^{mis} | \mathbf{X}', \mathbf{Y}(\mathbf{0})^{obs}, \mathcal{A}) \propto \int \Pr(\mathbf{Y}(\mathbf{0})^{mis} | \mathbf{X}, \boldsymbol{\theta}'_{y.x'}) \Pr(\boldsymbol{\theta}'_{y.x'} | \mathbf{X}, \mathbf{Y}(\mathbf{0})^{obs}) \pi(\boldsymbol{\theta}_{y.x'}) d\boldsymbol{\theta}_{y.x'},$$

where the parameterized posterior predictive distribution of the counterfactual is

$$\Pr(\mathbf{Y}(\mathbf{0})^{mis} | \mathbf{X}, \boldsymbol{\theta}'_{y.x'}) = \prod_{it \in S_1} f(y_{it}(c)^{mis} | \mathbf{X}_{it}, \boldsymbol{\theta}'_{y.x'}), \tag{A2}$$

and the likelihood function is:

$$\Pr(\boldsymbol{\theta}'_{y.x'} | \mathbf{X}, \mathbf{Y}^{obs}) = \prod_{it \in S_0} f(y_{it}(c)^{obs} | \mathbf{X}_{it}, \boldsymbol{\theta}'_{y.x'}). \tag{A3}$$

Recall that our objective is to impute the untreated potential outcomes for treated observations. Hence, we first estimate parameters based on Equation (A3) and then predict missing values $y_{it}(c)^{mis}$ for treated units at $a_i \leqslant t \leqslant T$ based on Equation (A2). If we can correctly estimate $\pi(\mathbf{U} | \mathbf{X}, \mathbf{Y}(\mathbf{0})^{obs})$ – the posterior distributions of $\mathbf{U}$ – using a factor analysis, we can draw samples of treated counterfactuals $y_{it}(c)^{mis}$ from its posterior predictive distribution by integrating out the parameters, including $\mathbf{U} = \mathbf{\Gamma}'\mathbf{F}$:

$$\Pr(\mathbf{Y}(\mathbf{0})^{mis}|\mathbf{X}, \mathbf{Y}(\mathbf{0})^{obs}, \mathcal{A}) \propto \int \Pr(\mathbf{Y}(\mathbf{0})^{mis}|\mathbf{X}, \mathbf{U}, \mathbf{Y}(\mathbf{0})^{obs})\pi(\mathbf{U}|\mathbf{X}, \mathbf{Y}(\mathbf{0})^{obs})d\mathbf{U}.$$

Note that, because $\mathbf{Y}(\mathbf{0})^{obs}$ implies $\mathcal{A}$, the posterior of $\mathbf{U}$ is unconditional on $\mathcal{A}$ given $\mathbf{Y}(\mathbf{0})^{obs}$. In other words, if we can build a flexible model and estimate its parameters using observed data, we can predict the counterfactuals using the posterior predictive distribution for treated observations in the post-treatment period ($a_i \leqslant t \leqslant T, \ \forall i$).

A.2. The MCMC Algorithm

Model searching and parameter estimation are based on the reduced form model after re-parameterization, which can be written as the following:

$$y(c)_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + \mathbf{Z}'_{it}(\boldsymbol{\omega_\alpha} \cdot \tilde{\boldsymbol{\alpha}}_i) + \mathbf{A}'_{it}(\boldsymbol{\omega_\xi} \cdot \tilde{\boldsymbol{\xi}}_t) + (\boldsymbol{\omega_\gamma} \cdot \tilde{\gamma}_i)'\mathbf{f}_t + \epsilon_{it}, \tag{A4}$$

and $\tilde{\boldsymbol{\xi}}_t = \Phi_\xi \tilde{\boldsymbol{\xi}}_{t-1} + \tilde{\boldsymbol{e}}_t$, and $\tilde{\boldsymbol{\alpha}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p_2})$, $\tilde{\gamma}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$, $\tilde{\boldsymbol{e}}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p_3})$. We assign Bayesian Lasso priors to the following parameters:

$$\begin{aligned}
&\beta_k|\tau^2_{\beta_k} \sim \mathcal{N}(0, \tau^2_{\beta_k}), \ \tau^2_{\beta_k}|\lambda_\beta \sim Exp(\frac{\lambda^2_\beta}{2}), \ \lambda^2_\beta \sim \mathcal{G}(a_1, a_2), \ \forall 1 \leq k \leq p_1; \\
&\omega_{\alpha_j}|\tau^2_{\alpha_j} \sim \mathcal{N}(0, \tau^2_{\alpha_j}), \ \tau^2_{\alpha_j}|\lambda_\alpha \sim Exp(\frac{\lambda^2_\alpha}{2}), \ \lambda^2_\alpha \sim \mathcal{G}(b_1, b_2), \ \forall 1 \leq j \leq p_2; \\
&\omega_{\xi_j}|\tau^2_{\xi_j} \sim \mathcal{N}(0, \tau^2_{\xi_j}), \ \tau^2_{\xi_j}|\lambda_\xi \sim Exp(\frac{\lambda^2_\xi}{2}), \ \lambda^2_\xi \sim \mathcal{G}(c_1, c_2), \ \forall 1 \leq j \leq p_3; \\
&\omega_{\gamma_j}|\tau^2_{\gamma_j} \sim \mathcal{N}(0, \omega^2_{\gamma_j}), \ \omega^2_{\gamma_j}|\lambda_\gamma \sim Exp(\frac{\lambda^2_\gamma}{2}), \ \lambda^2_\gamma \sim \mathcal{G}(k_1, k_2), \ \forall 1 \leq j \leq r.
\end{aligned} \tag{A5}$$

Note that the shrinkage on the factor term is imposed on the factor loadings, and the latent factors are not re-parameterized and the state-space equation remains as $\mathbf{f}_t = \Phi_f \mathbf{f}_{t-1} + \boldsymbol{\nu}_t$, with $\boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$. We assume $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2_\epsilon)$ and $\sigma^{-2}_\epsilon \sim \mathcal{G}(e_1, e_2)$. Because the number of factors r is unknown, we presume a reasonably large positive integer for r (the initial number of factors) and let the algorithm determine its value, based on the posterior distributions of $\boldsymbol{\omega}_\gamma$.

The MCMC algorithm for the DM-LFM, which simulates parameter posteriors and predicts $Y(\mathbf{0})^{miss}$ using data $\mathcal{D}$, takes the following recursive steps:

1. Start with the initial values of the parameters $\boldsymbol{\theta}^{(0)}$;
2. in the gth iteration, sample from the following conditional distributions based on the most updated values $\boldsymbol{\theta}^{(g-1)}$;

(a) update $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \sim \mathcal{N}(\bar{\boldsymbol{\beta}}, \boldsymbol{B}_1)$$
$$\boldsymbol{B}_1 = \left(\sigma_\epsilon^{-2} \sum_{w_{it}=0} \boldsymbol{X}_{it}\boldsymbol{X}_{it}' + \boldsymbol{B}_0^{-1} \right)^{-1}$$
$$\bar{\boldsymbol{\beta}} = \boldsymbol{B}_1 \left(\sigma_\epsilon^{-2} \sum_{w_{it}=0} \boldsymbol{X}_{it} u_{it} \right)$$
$$\boldsymbol{B}_0^{-1} = \mathrm{Diag}(\tau_{\beta_1}^{-2}, \ldots, \tau_{\beta_{p_1}}^{-2})$$
$$u_{it} = y_{it}(c)^{obs} - \mathbf{Z}_{it}'(\boldsymbol{\omega_\alpha} \cdot \tilde{\boldsymbol{\alpha}}_i) - \mathbf{A}_{it}'(\boldsymbol{\omega_\xi} \cdot \tilde{\boldsymbol{\xi}}_t) - (\boldsymbol{\omega_\gamma} \cdot \tilde{\boldsymbol{\gamma}}_i)'\mathbf{f}_t;$$

(b) update $(\tilde{\boldsymbol{\alpha}}_i', \tilde{\boldsymbol{\gamma}}_i')'$ [A1] :

Denote $\tilde{\boldsymbol{Z}}_{it} = (\boldsymbol{Z}_{it}' \cdot \boldsymbol{\omega}_\alpha',\ \boldsymbol{\omega}_\gamma' \cdot \mathbf{f}_t')'$

$$(\tilde{\boldsymbol{\alpha}}_i', \tilde{\boldsymbol{\gamma}}_i')' \sim \mathcal{N}(\bar{\boldsymbol{\alpha}}, \mathbf{H}_1)$$
$$\mathbf{H}_1 = \left(\sigma_\epsilon^{-2} \sum_{i, w_{it}=0} \tilde{\mathbf{Z}}_{it}\tilde{\mathbf{Z}}_{it}' + \mathbf{I}_{(p_2+r)} \right)^{-1}$$
$$\bar{\boldsymbol{\alpha}} = \mathbf{H}_1 \left(\sigma_\epsilon^{-2} \sum_{i, w_{it}=0} \tilde{\mathbf{Z}}_{it} u_{it} \right)$$
$$u_{it} = y_{it}(c)^{obs} - \mathbf{X}_{it}'\boldsymbol{\beta} - \mathbf{A}_{it}'(\boldsymbol{\omega_\xi} \cdot \tilde{\boldsymbol{\xi}}_t);$$

(c) update $(\tilde{\boldsymbol{\xi}}_t', \mathbf{f}_t')'$ [A2] :

Denote $\boldsymbol{\Psi}_t = (\tilde{\boldsymbol{\xi}}_t', \mathbf{f}_t')'$, $\tilde{\mathbf{A}}_{it} = (\mathbf{A}_{it}' \cdot \boldsymbol{\omega}_\xi', \boldsymbol{\omega}_\gamma' \cdot \tilde{\boldsymbol{\gamma}}_i')'$ and $\Phi = \mathrm{diag}(\phi_{\xi_1}, \ldots, \phi_{\xi_{p_3}}, \phi_{f_1}, \ldots, \phi_{f_r})$.

[A1] When some covariates appear in both $\mathbf{X}_{it}$ and $\mathbf{Z}_{it}$ or $\mathbf{A}_{it}$, an interweave from the non-centered parameterization to the centered parameterization using the algorithm developed by Yu and Meng (2011) can improve mixing.

[A2] When the number of units is relatively small, we jointly update $\tilde{\boldsymbol{\Xi}} = (\tilde{\boldsymbol{\xi}}_1, \ldots, \tilde{\boldsymbol{\xi}}_T)$ and $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)$ using the algorithm developed by Carter and Kohn (1994).

Assume $\boldsymbol{\Psi}_0 = \mathbf{0}$ as initial state.

$$\boldsymbol{\Psi}_t \sim \mathcal{N}(\boldsymbol{\Omega}_t^{-1}\boldsymbol{\mu}_t, \boldsymbol{\Omega}_t^{-1})$$

$$\boldsymbol{\mu}_t = \begin{cases} \Phi\boldsymbol{\Psi}_{t-1} + \sigma_\epsilon^{-2} \sum\limits_{t,w_{it}=0} \tilde{\mathbf{A}}_{it} u_{it} + \Phi\Psi_{t+1}, & 1 \le t \le T-1 \\ \Phi\Psi_{t-1} + \sigma_\epsilon^{-2} \sum\limits_{t,w_{it}=0} \tilde{\mathbf{A}}_{it} u_{it}, & t = T \end{cases}$$

$$\boldsymbol{\Omega}_t = \begin{cases} \mathbf{I}_{(p_3+r)} + \sigma_\epsilon^{-2} \sum\limits_{t,w_{it}=0} \tilde{\mathbf{A}}_{it}\tilde{\mathbf{A}}_{it}' + \Phi\Phi, & 1 \le t \le T-1 \\ \mathbf{I}_{(p_3+r)} + \sigma_\epsilon^{-2} \sum\limits_{t,w_{it}=0} \tilde{\mathbf{A}}_{it}\tilde{\mathbf{A}}_{it}', & t = T \end{cases}$$

$$u_{it} = y_{it}(c)^{obs} - \mathbf{X}_{it}'\boldsymbol{\beta} - \mathbf{Z}_{it}'(\boldsymbol{\omega}_\alpha \cdot \tilde{\boldsymbol{\alpha}}_i);$$

(d) update $(\boldsymbol{\omega}_\alpha', \boldsymbol{\omega}_\xi', \boldsymbol{\omega}_\gamma')'$: [A3]

Denote $\boldsymbol{\omega} = (\boldsymbol{\omega}_\alpha', \boldsymbol{\omega}_\xi', \boldsymbol{\omega}_\gamma')'$ and $\tilde{\mathbf{Z}}_{it} = (\mathbf{Z}_{it}' \cdot \tilde{\boldsymbol{\alpha}}_i', \mathbf{A}_{it}' \cdot \tilde{\boldsymbol{\xi}}_t', \tilde{\boldsymbol{\gamma}}_i' \cdot \mathbf{f}_t')'$.

$$\boldsymbol{\omega} \sim \mathcal{N}(\bar{\boldsymbol{\omega}}, \boldsymbol{\Omega}_1)$$

$$\boldsymbol{\Omega}_1 = \left(\sigma_\epsilon^{-2} \sum_{w_{it}=0} \tilde{\mathbf{Z}}_{it}\tilde{\mathbf{Z}}_{it}' + \boldsymbol{\Omega}_0^{-1}\right)^{-1}$$

$$\bar{\boldsymbol{\omega}} = \boldsymbol{\Omega}_1 \left(\sigma_\epsilon^{-2} \sum_{w_{it}=0} \tilde{\mathbf{Z}}_{it} u_{it}\right)$$

$$\boldsymbol{\Omega}_0^{-1} = \mathrm{Diag}(\tau_{\alpha_1}^{-2}, \ldots, \tau_{\alpha_{p_2}}^{-2}, \tau_{\xi_1}^{-2}, \ldots, \tau_{\xi_{p_3}}^{-2}, \tau_{f_1}^{-2}, \ldots, \tau_{f_r}^{-2})$$

$$u_{it} = y_{it}(c)^{obs} - \mathbf{X}_{it}'\boldsymbol{\beta};$$

(e) update autoregressive coefficients:

i. ϕ_{ξ_j} in Φ_ξ for $j = 1, ..., p_3$

$$\phi_{\xi_j} \sim \mathcal{N}(\bar{\phi}, \Phi_1)$$

$$\Phi_1 = \left(\sum_{t=1}^{T} \tilde{\xi}_{j,t-1}^2 + \sigma_\phi^{-2}\right)^{-1}$$

$$\bar{\phi} = \Phi_1 \left(\sum_{t=1}^{T} \tilde{\xi}_{j,t}\tilde{\xi}_{j,t-1}\right)$$

[A3] When the number of covariates and the pre-specified number of factors are relatively small, we jointly update $\boldsymbol{\beta}$, $\boldsymbol{\omega}_\alpha$, $\boldsymbol{\omega}_\xi$ and $\boldsymbol{\omega}_\gamma$ to improve mixing.

ii. ϕ_{f_j} in Φ_f for $j = 1, ..., r$

$$\phi_{f_j} \sim \mathcal{N}(\bar{\phi}, \Phi_1)$$
$$\Phi_1 = \left(\sum_{t=1}^{T} f_{j,t-1}^2 + \sigma_\phi^{-2} \right)^{-1}$$
$$\bar{\phi} = \Phi_1 \left(\sum_{t=1}^{T} f_{j,t} f_{j,t-1} \right);$$

(f) update $\tau_{\beta_j}^2$: [A4]

$$\tau_{\beta_j}^{-2} \sim IG(\sqrt{\frac{\lambda_\beta^2}{\beta_j^2}}, \lambda_\beta^2), \quad \forall 1 \leq j \leq p_1;$$

(g) update $\tau_{\alpha_j}^2$:

$$\tau_{\alpha_j}^{-2} \sim IG(\sqrt{\frac{\lambda_\alpha^2}{\omega_{\alpha_j}^2}}, \lambda_\alpha^2), \quad \forall 1 \leq j \leq p_2;$$

(h) update $\tau_{\xi_j}^2$:

$$\tau_{\xi_j}^{-2} \sim IG(\sqrt{\frac{\lambda_\xi^2}{\omega_{\xi_j}^2}}, \lambda_\xi^2), \quad \forall 1 \leq j \leq p_3;$$

(i) update $\tau_{\gamma_j}^2$:

$$\tau_{\gamma_j}^{-2} \sim IG(\sqrt{\frac{\lambda_\gamma^2}{\omega_{\gamma_j}^2}}, \lambda_\gamma^2), \quad \forall 1 \leq j \leq r;$$

(j) update λ_β^2:

$$\lambda_\beta^2 \sim \mathcal{G}(p_1 + a_1, \frac{1}{2} \sum_{j=1}^{p_1} \tau_{\beta_j}^2 + a_2);$$

(k) update λ_α^2:

$$\lambda_\alpha^2 \sim \mathcal{G}(p_2 + b_1, \frac{1}{2} \sum_{j=1}^{p_2} \tau_{\alpha_j}^2 + b_2);$$

(l) update λ_ξ^2:

$$\lambda_\xi^2 \sim \mathcal{G}(p_3 + c_1, \frac{1}{2} \sum_{j=1}^{p_3} \tau_{\xi_j}^2 + c_2);$$

[A4] Here IG stands for inverse-Gaussian distribution with parameters a and b, and its probability density function is $f(x) = \sqrt{\frac{b}{2\pi} x^{-3/2}} \exp\left(\frac{-b(x-a)^2}{2a^2 x}\right)$ for $x > 0$. Park and Casella (2008) state that conditioning on σ_ϵ can form an efficient Gibbs sampler for updating parameters. When heteroskedasticity exits, it's impossible to condition on σ_ϵ. Here we don't condition on σ_ϵ following Belmonte et al. (2014).

(m) update λ_γ^2:

$$\lambda_\gamma^2 \sim \mathcal{G}(r + k_1, \frac{1}{2}\sum_{j=1}^{r} \tau_{\gamma_j}^2 + k_2);$$

(n) update σ_ϵ^2:

$$\sigma_\epsilon^{-2} \sim \mathcal{G}(N_{obs} + e_1, \frac{1}{2}\sum_{D_{it}=0} (y_{it} - u_{it})^2 + e_2)$$
$$N_{obs} = N \times T - N_{tr} \times (T - T_0)$$
$$u_{it} = y_{it}(c)^{obs} - \mathbf{X}'_{it}\boldsymbol{\beta} - \mathbf{Z}'_{it}(\boldsymbol{\omega}_{\boldsymbol{\alpha}} \cdot \tilde{\boldsymbol{\alpha}}_i) - \mathbf{A}'_{it}(\boldsymbol{\omega}_{\boldsymbol{\xi}} \cdot \tilde{\boldsymbol{\xi}}_t) - (\boldsymbol{\omega}_{\boldsymbol{\gamma}} \cdot \tilde{\boldsymbol{\gamma}}_i)'\mathbf{f}_t;$$

(o) update predicted $y_{it}(c)^{mis}$ for observations under treatment:[A5]

$$y_{it}(c)^{mis} \sim \mathcal{N}(\mu_{it}, \sigma_\epsilon^2), \text{ for } it \in S_1$$
$$\mu_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + \mathbf{Z}'_{it}(\boldsymbol{\omega}_{\boldsymbol{\alpha}} \cdot \tilde{\boldsymbol{\alpha}}_i) + \mathbf{A}'_{it}(\boldsymbol{\omega}_{\boldsymbol{\xi}} \cdot \tilde{\boldsymbol{\xi}}_t) + (\boldsymbol{\omega}_{\boldsymbol{\gamma}} \cdot \tilde{\boldsymbol{\gamma}}_i)'\mathbf{f}_t;$$

(p) obtain an estimate for δ_{it}: $\delta_{it} = y_{it} - y_{it}(c)^{mis}, \text{ for } it \in S_1$.

3. Repeat (a)-(p) until convergence, and obtain G draws for each parameter, counterfactual ($y_{it}(c)^{mis}, \ it \in S_1$), and the individual causal effect ($\delta_{it}, \text{ for } it \in S_1$).

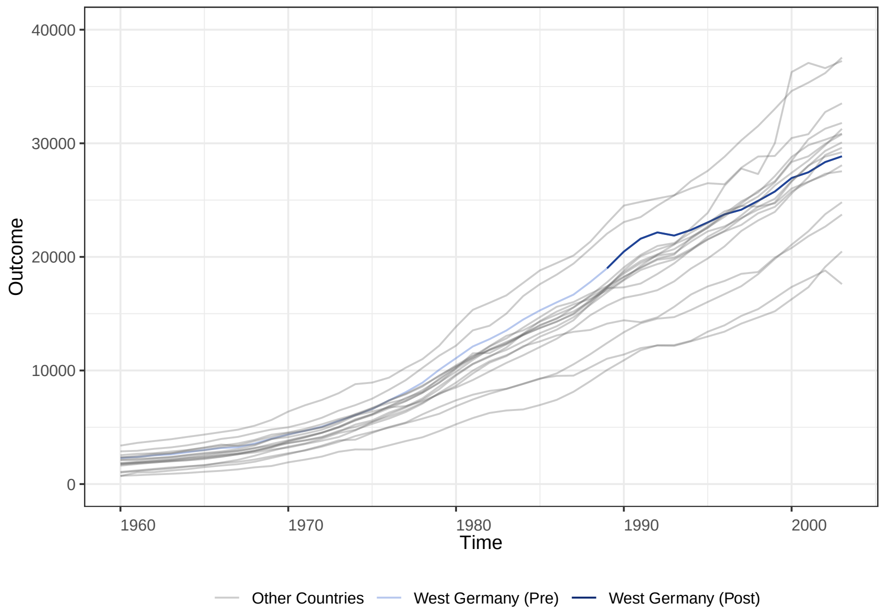
[A5]The draws of $y_{it}(c)^{mis}$ are not used to update the parameters.

A.3. Additional Information on the Empirical Examples

A.3.1. ADH (2015): German Reunification

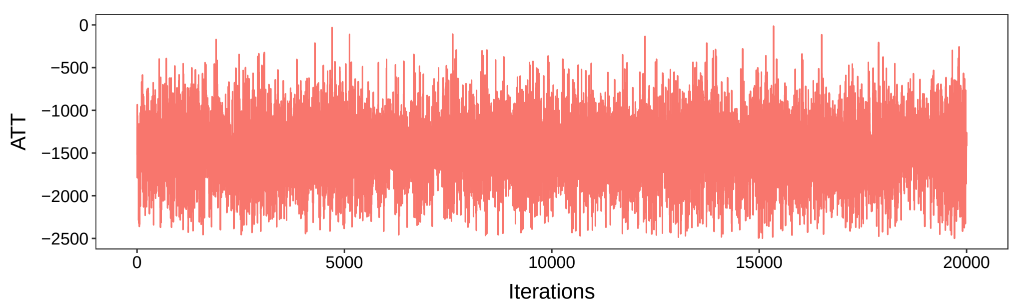
Figure A1 shows the outcome trajectories in ADH (2015)'s data, in which West Germany is shown in blue. Figure A2 shows the traceplot corresponding to the Markov Chain of the ATT (averaged over time).

FIGURE A1. RAW DATA: OUTCOME



Note: The above figure shows the outcome trajectories in ADH (2015)'s data.

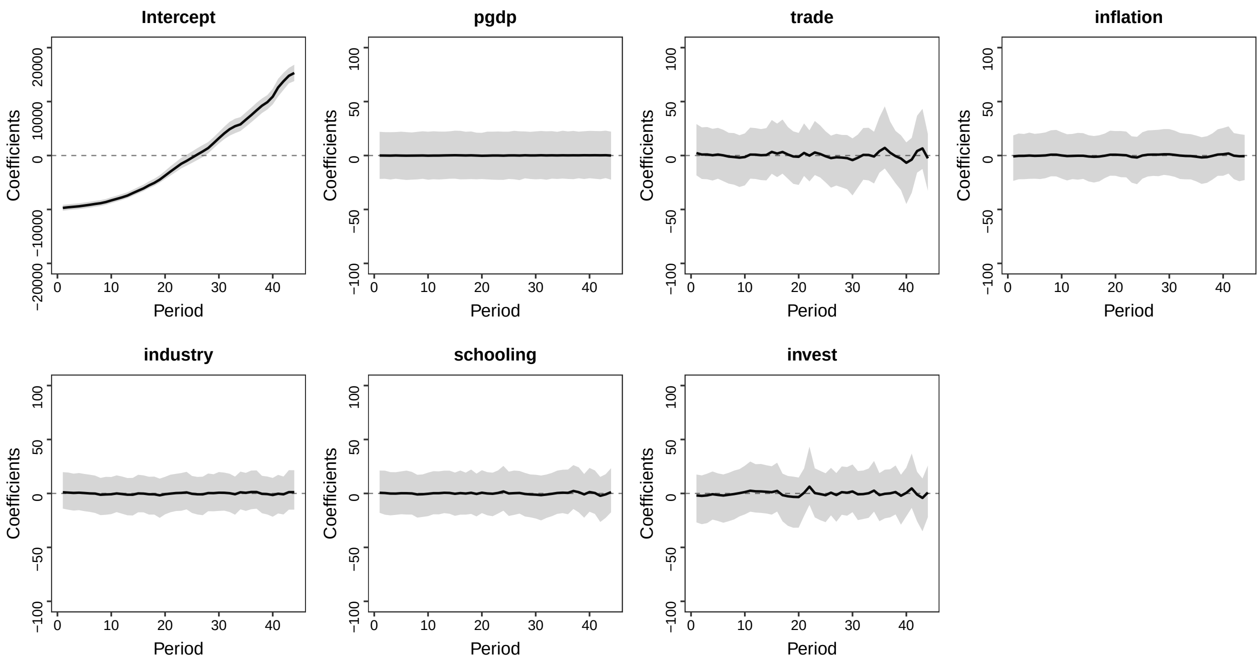
FIGURE A2. THE TRACEPLOT OF THE ATT MARKOV CHAIN



Note: The above figure shows the traceplot corresponding to the Markov Chain of the estimated treatment effect of German reunification on West Germany's economy. The first 5,000 draws (the burn-in period) are dropped.

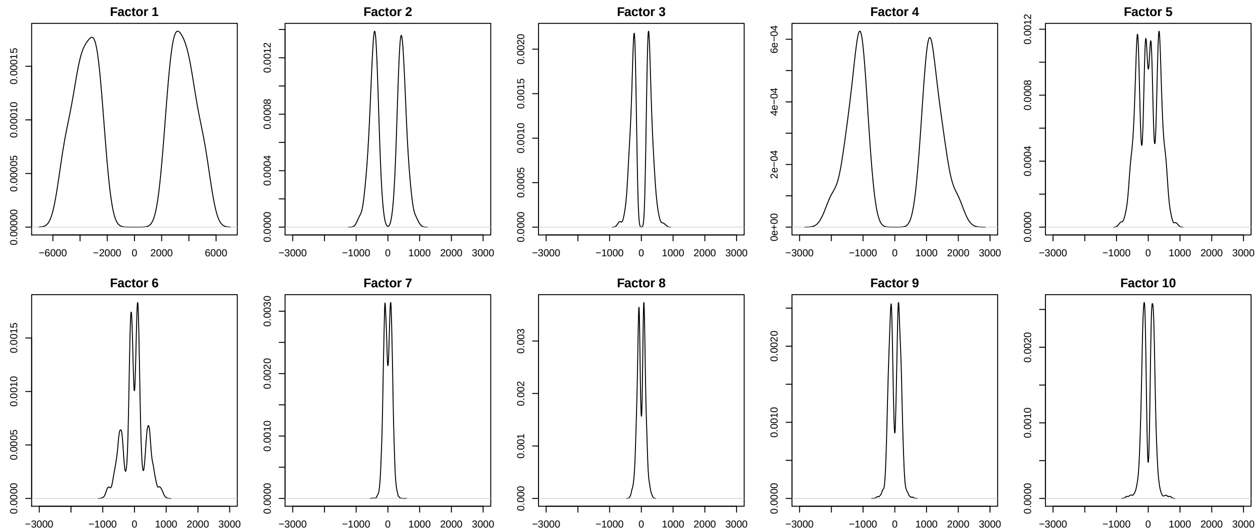
Figure A3 shows the posterior means and 95% credible intervals of the time-varying component of covariate coefficients $\boldsymbol{\xi}_t$. Figure A4 shows $\boldsymbol{\omega}_\gamma$ for the 10 factors included in the model, which captures the relative importance of the factors in explaining the outcome. Figure A5 shows the influence of the first latent factor on each country, i.e., $\gamma_{i1}' f_{1t}$.

Figure A3. Time-varying Coefficients $\boldsymbol{\xi}_t$



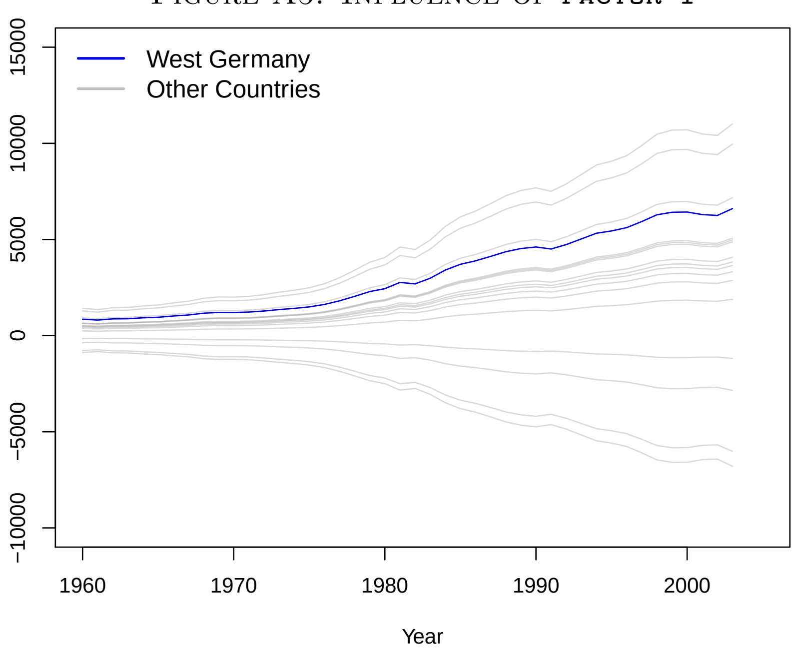
Note: The above figures show the posterior mean and 95% credibility intervals of $\boldsymbol{\xi}_t$.

Figure A4. $\boldsymbol{\omega}_\gamma$ for Factor Selection



Note: The above figures show the posterior distribution of $\boldsymbol{\omega}_\gamma$ for each of the 10 factors. Each $\boldsymbol{\omega}_\gamma$ (a scaling parameter) captures the importance of the corresponding factor.

Figure A5. Influence of Factor 1



Note: The above figure shows the influence of the first latent factor on each country $\gamma_{i1}' f_{1t}$.

A.3.2. Xu (2017): EDR on Voter Turnout

Figures A6 and A7 show the treatment status and outcome trajectories in Xu (2017)'s data.

FIGURE A6. TREATMENT STATUS

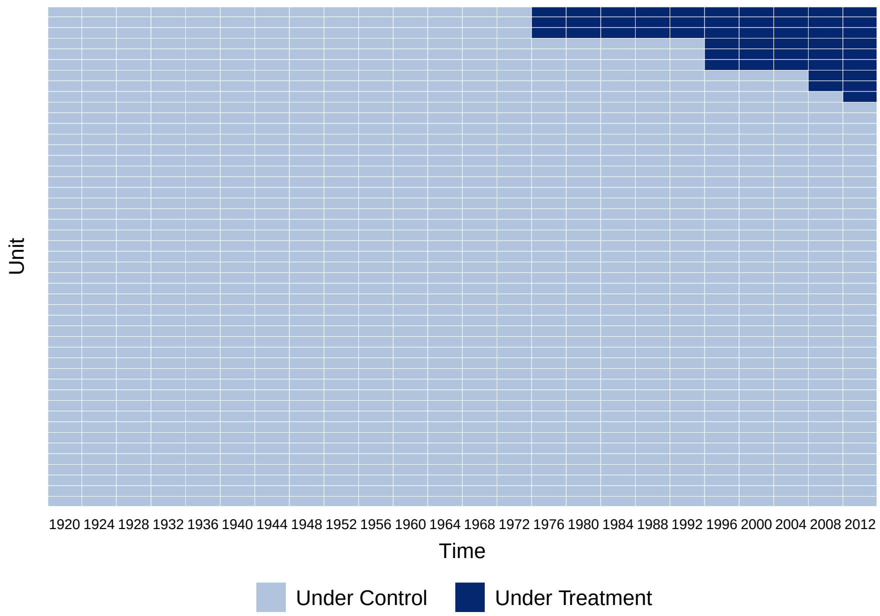


FIGURE A7. RAW DATA: OUTCOME

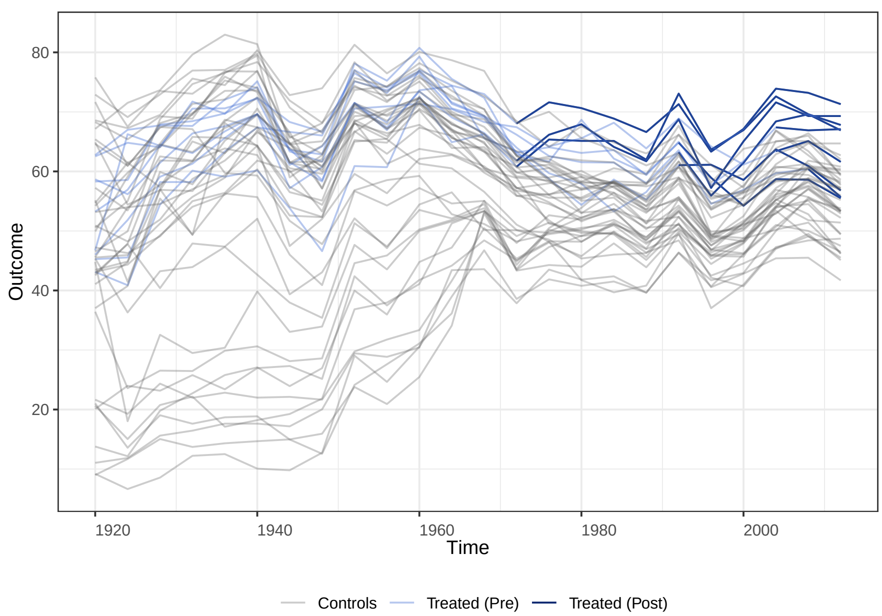
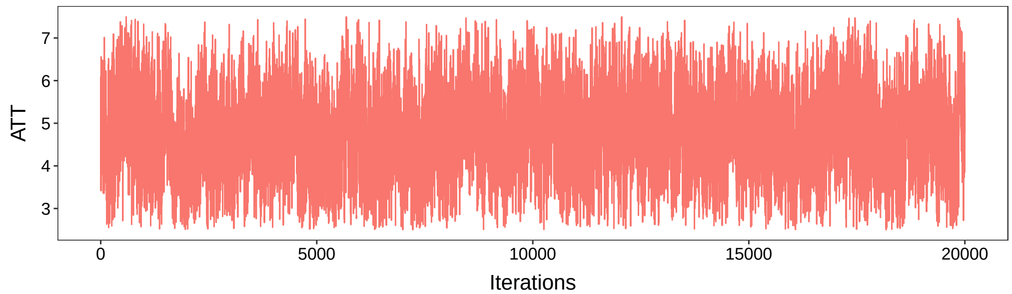


Figure A8 is a traceplot corresponding to the Markov Chain of the ATT (averaged over time and treated units).

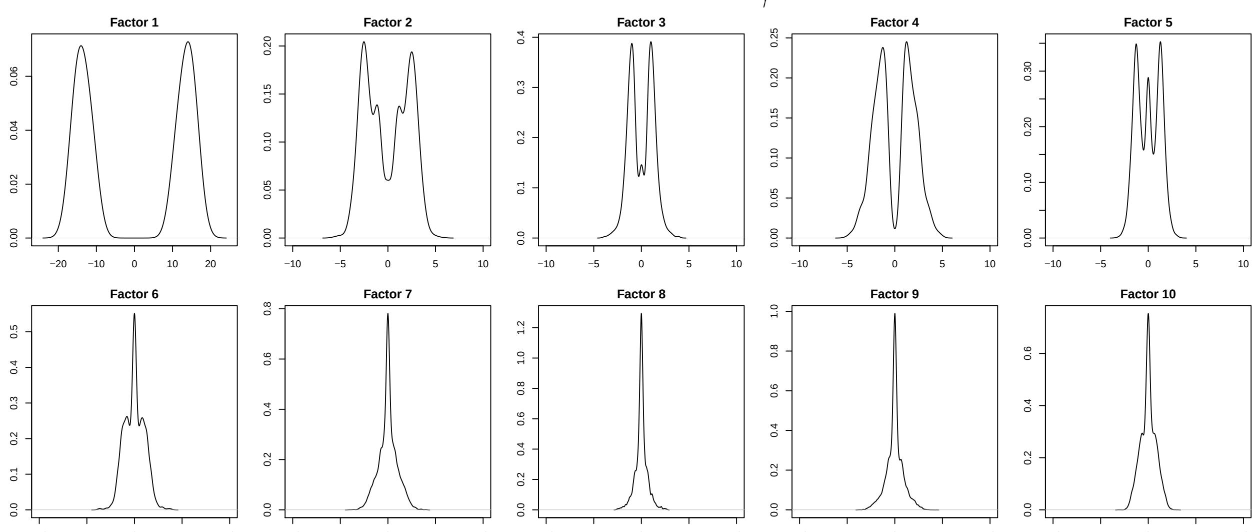
Figure A9 shows $\boldsymbol{\omega}_\gamma$ for the 10 factors included in the model, which capture the relative importance of the factors in explaining the outcome.

Figure A8. The Traceplot of the ATT Markov Chain



Note: The above figure shows the traceplot corresponding to the Markov Chain of the estimated ATT of EDR on voter turnout. The first 5,000 draws (the burn-in period) are dropped.

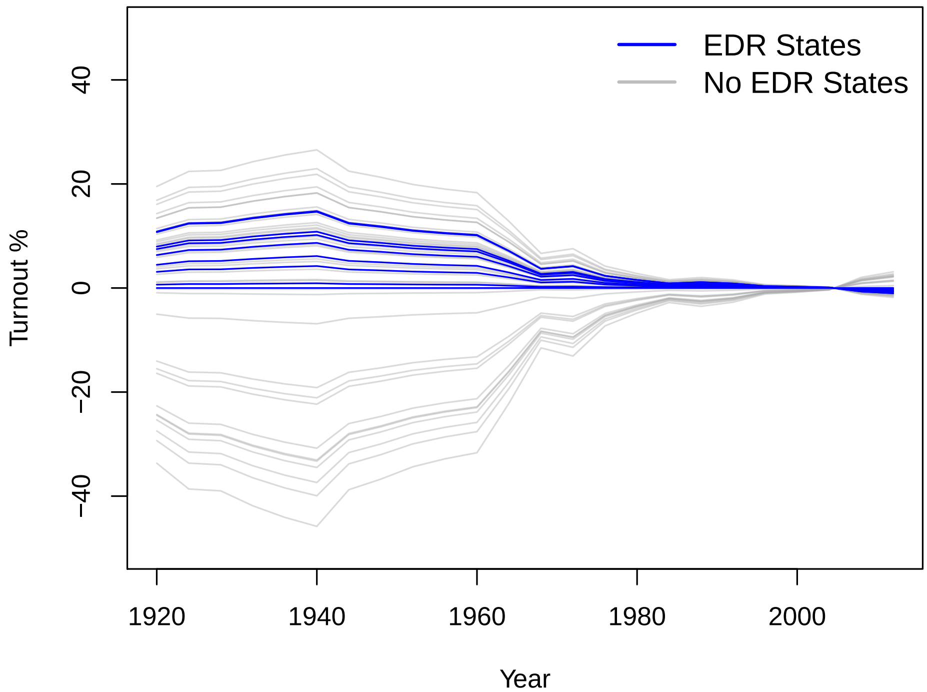
Figure A9. EDR on Voter Turnout: $\boldsymbol{\omega}_\gamma$ for Factor Selection



Note: The above figures show the posterior distribution of $\boldsymbol{\omega}_\gamma$ for each of the 10 factors. Each $\boldsymbol{\omega}_\gamma$ (a scaling parameter) captures the influence of the corresponding factor. These figures suggest that at least 4 factors can explain significant proportions of the variation in the non-treatment outcome.

Figure A10 displays the influence of the first latent factor on each state, i.e., $\gamma'_{i1} f_{1t}$.

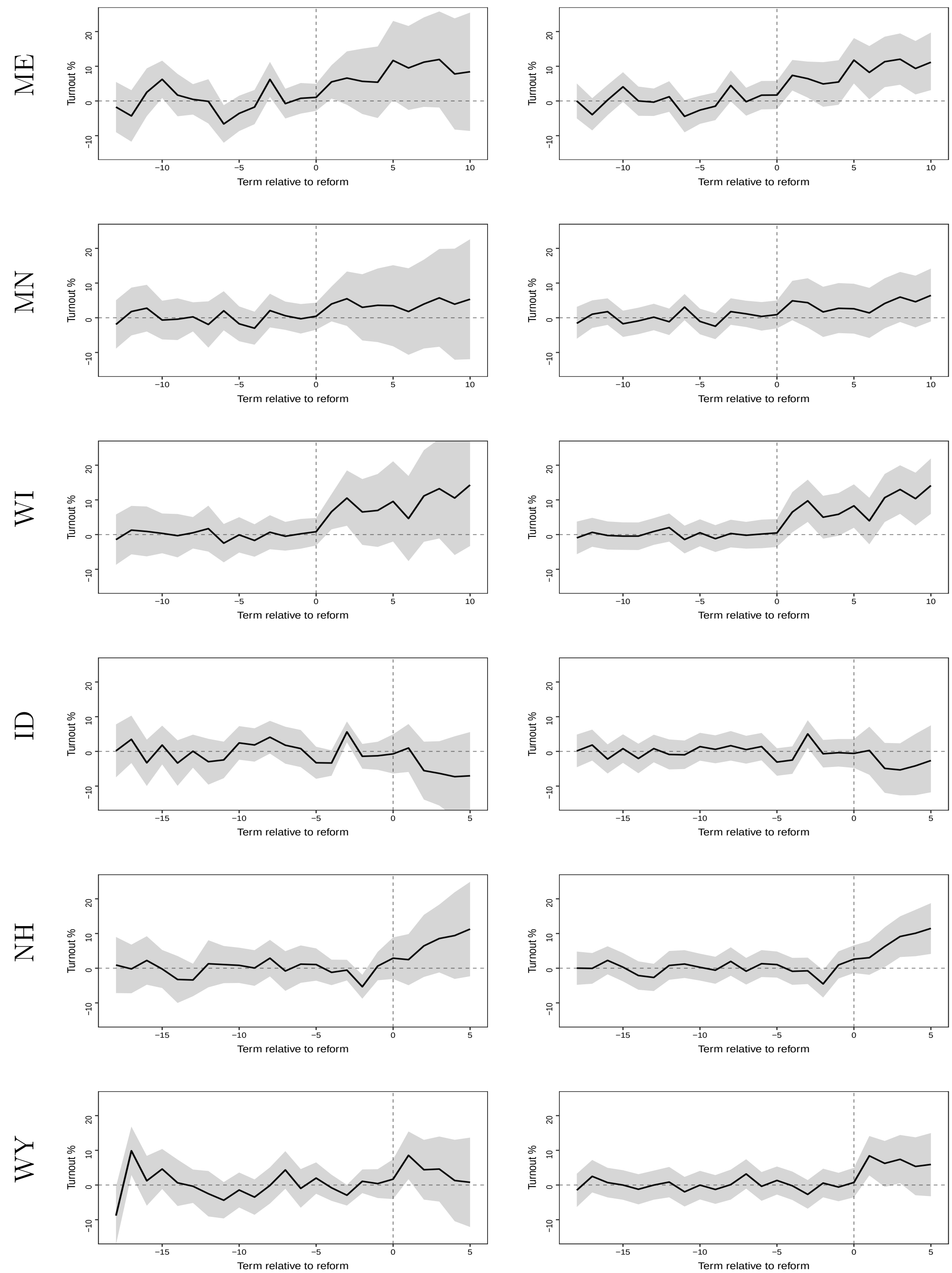
FIGURE A10. INFLUENCE OF FACTOR 1



Note: The above figure shows the influence of the first latent factor on each state $\gamma'_{i1} f_{1t}$. [Please check the PDF version if the gray lines do not show up in a printed-out copy.]

Figure A11 (next page) shows the estimated individual treatment effects of EDR on turnout using both Gsynth (left) and DM-LFM (right). It demonstrates that DM-LFM produces better model fit and narrower uncertainty estimates.

Figure A11. Estimated Individual Treatment Effects of EDR on Turnout



(a) Gsynth

(b) DM-LFM

Note: The above figures show the estimated treatment effect (with corresponding 95% confidence intervals or 95% credibility intervals) of EDR on voter turnout in 6 treated states using Gsynth (a) and DM-LFM (b), respectively.

A.4. Monte Carlo Evidence

We present Monte Carlo evidence in this section. First, we illustrate a basic data generating process (DGP) using a simulated example (A.4.1). Based on this DGP (and its close variants), we conduct additional Monte Carlo exercises, which aim at

1. evaluating the importance of the key components in DM-LFMs and studying their finite sample properties (Table A1 in A.4.2);
2. testing the model's robustness to non-normal error terms (Table A2 in A.4.2);
3. testing the model's robustness to non-AR1 factors (Table A3 in A.4.2);
4. showing the model's performance when T_0 or N_{co} is small (Table A4 in A.4.2);
5. comparing DM-LFMs with the synthetic control method and the generalized synthetic control method in cases with only one treated unit (A.4.3).

Overall, we find that the point and uncertainty estimates produced by DM-LFMs have desirable finite sample properties (from a frequentist perspective).

A.4.1. A Simulated Example

As stated in the main text, the simulated example is produced by the following DGP:

$$y_{it} = \delta_{it} w_{it} + \mathbf{X}'_{it}(\boldsymbol{\beta} + \boldsymbol{\alpha}_i + \boldsymbol{\xi}_t) + \boldsymbol{\gamma}'_i \mathbf{f}_t + \epsilon_{it} \tag{A6}$$

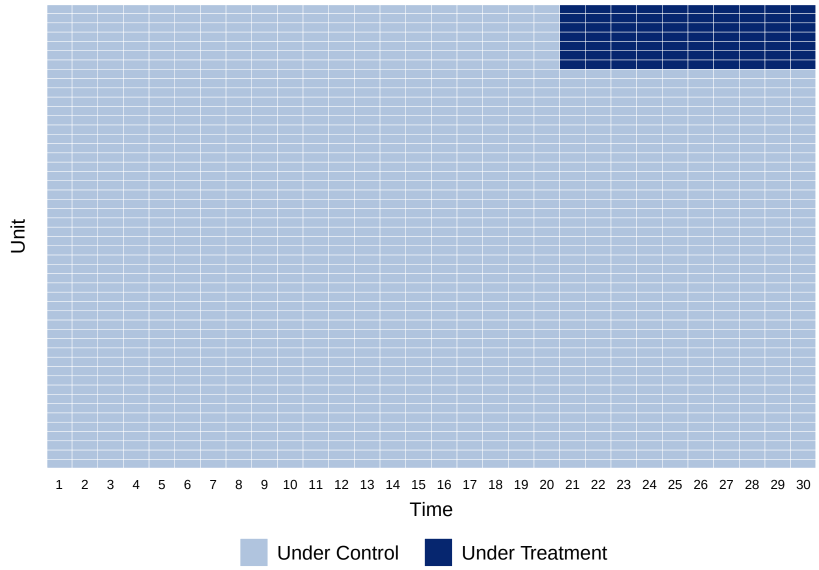
with $N = 50$ and $T = 30$. The treatment assignment follows a DiD setup: a unit is assigned either to the treatment group and receives the treatment in Period 21 ($T_0 = 20$), or to the control group and never receives the treatment. The treatment assignment is determined by a latent variable $tr_i^* = 0.7\gamma_{i1} + 0.3\gamma_{i2} + \pi_i$, in which $\pi_i \overset{i.i.d.}{\sim} N(0, 0.25)$. When tr_i^* is bigger than the 90 percentile of the population distribution of tr_i^*, unit i belongs to the treatment group; otherwise, it belongs to the control group—as a result, the number of treated units may slightly vary from sample to sample. This means that units with bigger values of γ_{i1} and γ_{i2} are more likely to be assigned to the treatment group. The selection on the factor loadings will lead to biases in the causal estimates if these factors are not accounted for in the estimation. The heterogeneous treatment effects are governed by $\delta_{it,t>20} = t - 20 + \tau_{it}$, in which $\tau_{it} \overset{i.i.d.}{\sim} N(0, 0.25)$.

$\mathbf{X}_{it}$ is a vector of control variables, including an intercept and nine pretreatment covariates, each of which is $i.i.d.$ $N(0, 1)$. $\boldsymbol{\beta} = (3\ 6\ 4\ 2\ 0\ 0\ 0\ 0\ 0)$ is the invariant part of the covariate coefficients. For the intercept and the first three covariates, their corresponding ξ_{jt} each follows an AR(1) process: $\xi_{jt} = 0.6\xi_{j,t-1} + e_{jt}$ in which $e_{jt} \overset{i.i.d.}{\sim} N(0, 1)$, $j = 0, 1, 2, 3$; their corresponding $\boldsymbol{\alpha}_i$ each follows a normal distribution: $\alpha_{ij} \overset{i.i.d.}{\sim} N(0, 0.25\beta_j^2)$, $j = 0, 1, 2, 3$. However, $\xi_{jt} = 0$, $\alpha_{ij} = 0, \forall i, \forall t$ for the remaining six covariates.

The factor vector $\mathbf{f}_t = (f_{1t},\ f_{2t})$ is two-dimensional. Both factors follow an AR(1) process: $f_{1t} = 0.7 f_{1,t-1} + \nu_{1t}$; $f_{2t} = 0.7 f_{2,t-1} + \nu_{2t}$. The two factor loadings γ_{i1}, $\gamma_{i2} \overset{i.i.d.}{\sim} N(0, 1)$. The error term $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, 1)$.

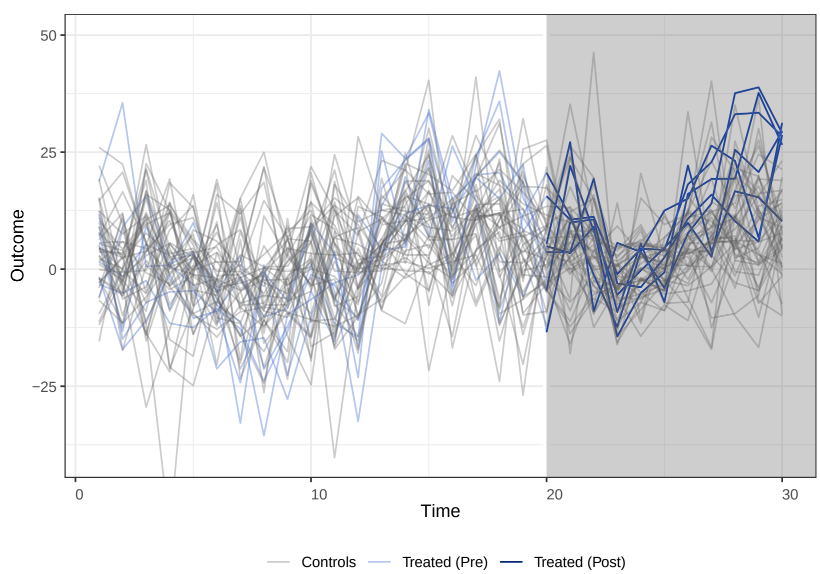
Figure A12 shows the treatment status of the simulated data and Figure A13 shows the outcome trajectories, among which the treated units are shown in blue.

Figure A12. Treatment Status



Note: The above figure shows the treatment status of the simulated data.

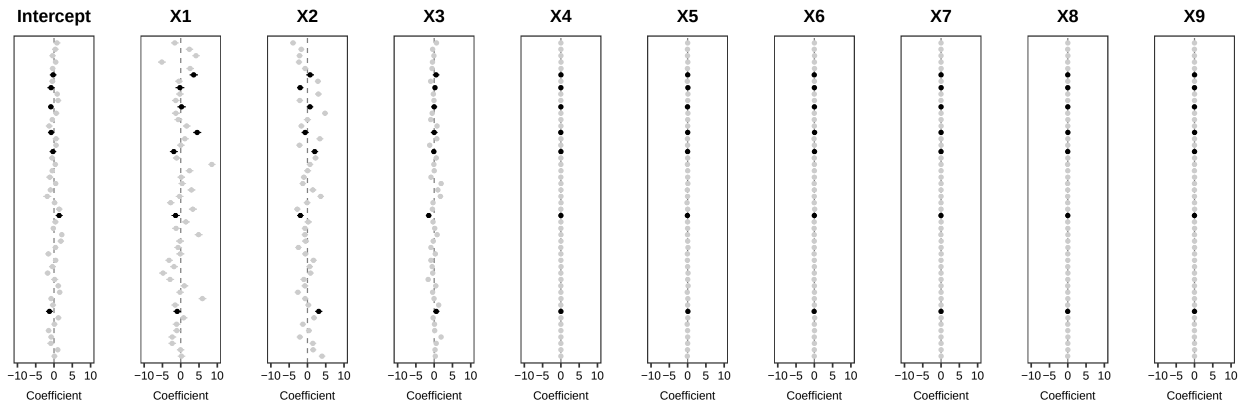
Figure A13. Raw Data: Outcome



Note: The above figure shows the outcome trajectories in the simulated data.

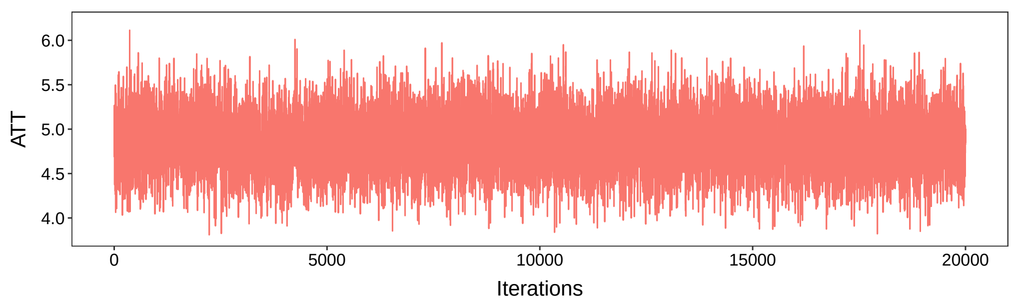
Figure A14 shows the unit-varying part of the covariate coefficients $\boldsymbol{\beta}_{it}$. Figure A15 shows the traceplots corresponding to the two Markov Chains of the ATT and the individual treatment effect on Unit 1 in Period 25 (i.e., $\hat{\delta}_{1,25}$).

FIGURE A14. POSTERIORS OF $\boldsymbol{\alpha}_i$

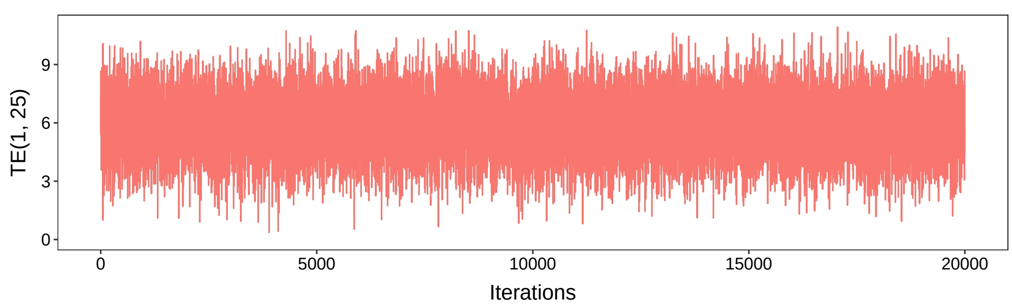


Note: The above figure shows the estimated posteriors of the unit-varying part of the covariate coefficients.

FIGURE A15. TRACEPLOTS OF THE MARKOV CHAINS OF TREATMENT EFFECTS



(a) ATT (average over time and units)



(b) Treatment effect on Unit 1 in Period 25

Note: The above figures show the traceplots corresponding to the two Markov Chains of the ATT and individual treatment effect (for Unit 1 in Period 25) estimates, respectively, using the simulated data. The first 5,000 draws (the burn-in period) are dropped.

A.4.2. Varying Model Specifications and Sample Sizes

In this subsection, we study the role of each of the main components of a DM-LFM in estimation and investigate how the sample size affects the model performance. We simulate samples using the DGP as in Equation (**??**) while varying the sample size (both the total number of units n and the number of pretreatment periods T_0). We estimate three models all by using MCMC with the same priors for common parameters:

(a) A model with covariates but without factors: $y_{it} = \delta_{it} w_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + \epsilon_{it}$, which is analogous to a DiD model including covariates with fixed coefficients. In other words, we shut off the factor component in the model.

(b) A model without covariates but with 10 latent factors: $y_{it} = \delta_{it} w_{it} + \boldsymbol{\gamma}'_i \mathbf{f}_t + \epsilon_{it}$, which is analogous to Gsynth without time-varying covariates. In other words, we shut off the covariate component in the model.

(c) A model with both covariates and factors, but the coefficients of the covariates are set to be fixed: $y_{it} = \delta_{it} w_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + \boldsymbol{\gamma}'_i \mathbf{f}_t + \epsilon_{it}$.

(d) The full model with both covariates with time- and unit-specific coefficients and factors (DM-LFM): $y_{it} = \delta_{it} w_{it} + \mathbf{X}'_{it}\boldsymbol{\beta}_{it} + \boldsymbol{\gamma}'_i \mathbf{f}_t + \epsilon_{it}$.

For each of the six combinations of $N = 40, 80$ and $T_0 = 20, 40, 80$, we simulate 500 samples and apply all four models to each dataset. Each estimation is based on 10,000 MCMC iterations with the first 2,000 runs as the burn-in period.

Figure **??** in the main text shows the RMSE for the posterior mean estimates of the ATT (left panel) and coverage rate of 95% credibility intervals of the ATT estimates (right panel) of the four models when $N = 40$.[A6] Table A1 shows the full results, including biases and standard deviation (SD) of the posterior estimates.

Non-normal errors. We re-do the above exercises with Model (d) using an asymmetric error term: $\epsilon_{it} \sim Gamma(2, 2)$, whose distribution is left skewed but has a fat right tail. Table A2 shows the results.

[A6] The coverage rate cannot be interpreted in the conventional way because the Bayesian 95% credibility interval does not rely on any repeated sampling but is the 0.95 probability that a parameter varies in an interval given a particular set of data. Nonetheless, we follow the convention in the literature to calculate the coverage rates to investigate the frequentist property of the Bayesian estimator.

Normal (non-AR1) factors. We re-do the above exercises with a variant of Model (d) using normally distributed factors: $f_{1t}, f_{2t} \sim N(0, 1)$. In the DGP as well as in Model (d), we assume $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}$ to have a fair comparison with Gsynth and the matrix completion methods, both of which do not allow time- and unit-specific coefficients. Table A3 shows the results.

Key takeaways:

1. Not surprisingly, Model (d) (DM-LFM) outperforms the other three models in terms of bias, SD, RMSE, and coverage. Each of the key components of the model, including factors and heterogeneity in the coefficients of covariates over time and across units, contributes to improving performance in causal effect estimation (Table A1).

2. The DF-LFM approaches the correct coverage rates as the number of pretreatment periods (T_0) grows despite relatively small N. The RMSE decreases as T_0 or N increases.

3. In our simulations, the DF-LFM is not sensitive to non-normal error terms (Table A2).

4. The DF-LFM performs well when the factors are drawn from a normal distribution; in fact, it works almost as well as Gsynth. Both methods perform better than the matrix completion method (Table A3). This is likely because the latter mis-specifies the model—it uses soft-impute instead of best subset to penalize the factors (see Liu et al. (2020) for a more detailed comparison).

5. However, the DF-LFM does not perform very well in terms of both RMSE and the coverage rate when the number of pretreatment periods (T_0) or the number of control units (N_{co}) is small (Table A4). This is likely because in such scenarios, the method lacks overlapped data to correctly adjust for the influence of factor term.

Table A1. Comparison of Model Performance

		w/o factors, w/ covariates (fixed coefficients)				w/ factors, w/o covariates				w/ factors & covariates (fixed coefficients)				w/ factors & covariates (varying coefficients)			
N	T_0	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover
40	20	0.40	3.23	3.26	0.37	0.30	2.77	2.79	0.65	0.15	1.31	1.31	0.79	0.01	0.31	0.30	0.93
40	40	0.04	3.11	3.11	0.32	0.01	2.38	2.37	0.74	0.00	0.89	0.89	0.89	0.01	0.23	0.23	0.93
40	80	0.10	2.89	2.89	0.39	0.01	1.58	1.58	0.90	0.01	0.86	0.86	0.91	0.00	0.19	0.18	0.94
80	20	0.00	3.09	3.08	0.24	0.06	2.13	2.13	0.63	0.00	1.03	1.03	0.72	0.01	0.21	0.20	0.92
80	40	0.13	2.89	2.89	0.25	0.05	1.60	1.60	0.73	0.05	0.66	0.66	0.86	0.00	0.14	0.14	0.96
80	80	0.03	2.90	2.90	0.24	0.07	1.29	1.29	0.84	0.01	0.55	0.55	0.89	0.00	0.13	0.12	0.95

Note: Each set of results (one row) is based on 500 simulated samples. For each sample, we run 10,000 MCMC iterations (among which 2,000 are burn-in) for each of the three models. SD, RMSE, and Cover represent standard deviation, root mean squared error, and coverage rate, respectively. The coverage rate is based on 95% credibility intervals of the ATT posteriors.

Table A2. Robust to Non-Normal Errors: $\epsilon_{it} \sim Gamma(2,2)$

		DM-LFM (full model)			
n	T_0	Bias	SD	RMSE	Cover
40	20	0.00	0.32	0.31	0.92
40	40	0.00	0.22	0.22	0.94
40	80	0.02	0.18	0.18	0.96
80	20	0.01	0.21	0.21	0.92
80	40	0.02	0.15	0.15	0.92
80	80	0.00	0.13	0.12	0.95

Note: Each set of results (one row) is based on 500 simulated samples. For each sample, we run 10,000 MCMC iterations (among which 2,000 are burn-in). SD, RMSE, and Cover represent standard deviation, root mean squared error, and coverage rate, respectively. The coverage rate is based on 95% credibility intervals of the ATT posteriors.

Table A3. Robustness to Non-AR1 Factors

		DM-LFM (full model)				Gsynth				Matrix Completion			
n	T_0	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover	Bias	SD	RMSE	Cover
50	20	0.00	0.22	0.21	0.93	0.00	0.22	0.21	0.93	0.00	0.44	0.44	0.70
50	40	0.01	0.18	0.17	0.95	0.01	0.18	0.17	0.95	0.00	0.34	0.33	0.76
50	80	0.00	0.17	0.16	0.95	0.00	0.17	0.17	0.93	0.00	0.30	0.29	0.76
100	20	0.01	0.14	0.14	0.94	0.01	0.14	0.14	0.95	0.03	0.40	0.40	0.57
100	40	0.00	0.12	0.11	0.96	0.00	0.12	0.12	0.95	0.01	0.27	0.26	0.71
100	80	0.01	0.12	0.11	0.95	0.01	0.12	0.11	0.93	0.00	0.23	0.23	0.72

Note: For the DM-LFM, we assume covariate parameters do not vary by unit or time; 10 factors are included as usual; shrinkage is imposed on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. For both Gsynth and matrix completion methods, we assume the number of factors/tuning parameters are unknown and use a cross-validation procedure to choose them. Each set of results (one row) is based on 500 simulated samples. For each sample, we run 10,000 MCMC iterations (among which 2,000 are burn-in). SD, RMSE, and Cover represent standard deviation, root mean squared error, and coverage rate, respectively. The coverage rate is based on 95% credibility intervals of the ATT posteriors.

Table A4. Poor Performance with Small T_0 or Small N_{co}

		DM-LFM (full model)			
$N_{tr}\%$	T_0	Bias	SD	RMSE	Cover
0.20	5	0.10	1.30	1.31	0.67
0.20	10	0.01	0.59	0.58	0.87
0.20	15	0.00	0.38	0.38	0.90
0.20	20	0.00	0.33	0.33	0.90
0.40	5	0.02	1.38	1.38	0.60
0.40	10	0.00	0.62	0.61	0.81
0.40	15	0.02	0.40	0.40	0.89
0.40	20	0.01	0.33	0.33	0.89
0.60	5	0.00	1.59	1.59	0.55
0.60	10	0.05	0.78	0.78	0.77
0.60	15	0.03	0.67	0.67	0.84
0.60	20	0.01	0.48	0.48	0.86

Note: The sample size is fixed at 50. Each set of results (one row) is based on 500 simulated samples. For each sample, we run 10,000 MCMC iterations (among which 2,000 are burn-in). SD, RMSE, and Cover represent standard deviation, root mean squared error, and coverage rate, respectively. The coverage rate is based on 95% credibility intervals of the ATT posteriors.

A.4.3. Comparisons of Methods with a Single Treated Unit

In this section, we compare DM-LFM with two existing methods for comparative case studies: the synthetic control method (SCM) and the generalized synthetic control method (Gsynth). We simulate samples based on Equation (A6) with two modifications: (1) we set the in-sample percentile threshold of being treated to $(N-1)/N$ such that there is a single treated unit in each sample;[A7] (2) we set unit-specific heterogeneity ($\boldsymbol{\alpha}_i$, including unit fixed effects) to be 0 to reduce the chances that the treated unit trajectory lies outside the convex hull of those of the control units, which is a required assumption for the SCM.

Setups. We consider the following three setups:

(a) A model with three relatively strong factors but no covariates: $y_{it} = \delta_{it}w_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \epsilon_{it}$, in which $\gamma_{i1},\ \gamma_{i2},\ \gamma_{i3} \overset{i.i.d.}{\sim} N(0, 4^2)$. *Treatment assignment*: $tr_i^* = 0.1\gamma_{i1} + 0.1\gamma_{i2} + \pi_i$, $\pi_i \overset{i.i.d.}{\sim} N(0,1)$ (selection on the loadings of the first two factors; the unit with the biggest tr_i^* is assigned the treatment).

(b) A model with eight relatively weak factors but no covariates: $y_{it} = \delta_{it}w_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \epsilon_{it}$, in which $\gamma_{i1},\ \gamma_{i2}, \cdots, \gamma_{i8} \overset{i.i.d.}{\sim} N(0, 2^2)$. Treatment assignment is the same as above.

(c) A model with three relatively strong factors and four covariates: $y_{it} = \delta_{it}w_{it} + \mathbf{Z}_i'(\boldsymbol{\beta} + \boldsymbol{\xi}_t) + \boldsymbol{\gamma}_i'\mathbf{f}_t + \epsilon_{it}$, in which $\gamma_{i1},\ \gamma_{i2},\ \gamma_{i3} \overset{i.i.d.}{\sim} N(0, 4^2)$ and $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4})$ is a vector of covariates following $N(0,1)$; $\boldsymbol{\beta} = (4\ 3\ 2\ 1)$ and $\boldsymbol{\xi}_t$ is the same as in the simulated example. *Treatment assignment*: $tr_i^* = 0.1\gamma_{i1} + 0.1\gamma_{i2} + 0.1Z_{i1} + 0.1Z_{i2} + \pi_i$, $\pi_i \overset{i.i.d.}{\sim} N(0,1)$ (selection on the loadings of the first two factors and the first two covariates; the unit with the biggest tr_i^* is assigned the treatment). Note that here we do not include unit-specific coefficients $\boldsymbol{\alpha}_i$ because they are incompatible with the SCM.

We set the treatment effect $\delta_{it} = 0$. Once again, for each setup, we change the sample size by varying N and T_0.

Estimators. With the SCM, we supply the algorithm with all four relevant covariates as well as two irrelevant ones, Z_{i5} and Z_{i6}. We also vary the number of pretreatment outcomes serving as "special predictors." SCM$^{(1)}$ uses three pretreatment outcomes for matching in Setups (a) and (c), and eight in Setup (b). The rest of the pretreatment outcomes are taken as the validation set. SCM$^{(2)}$ uses $T_0 - 5$ pretreatment outcomes for matching and five periods for validation.

[A7] This creates small correlations in treatment status across units, which in theory violates our modeling assumptions.

With Gsynth, we do not include covariates (Gsynth does not support time-invariant covariates). Gsynth$^{(1)}$ assumes the correct number of factors: three in Setup (a), eight in Setup (b), and seven in Setup (c) (the influence of covariates $\mathbf{Z}_i'\boldsymbol{\beta}_t$ can be potentially captured by additional latent factors), while Gsynth$^{(2)}$ employs the cross-validation procedure to select the number of factors.

With the Bayesian DM-LFM, we include all six covariates, including the irrelevant ones, and set the maximal number of factors at 10. For each estimation, we run 10,000 MCMC iterations and discard the first 2,000 runs as the burn-in period.

Results. Tables A5, A6, and A7 present the full results for each of the three setups. Each table has four panels, showing the biases of the estimates, RMSE, coverage rates, and average run time using each method. Figure A16 visualizes the results from Setup (3). A few key takeaways:

1. The fact that Gsynth$^{(1)}$ performs well is not surprising because it assumes the true model.
2. In both Setups (a) and (b), DM-LFM performs almost as well as Gsynth$^{(1)}$, DM-LFM often out-performs Gsynth$^{(2)}$; the relative performance of DM-LFM improves when there are more factors or the factors are weaker.
3. In Setup (c), DM-LFM almost always performs better than Gsynth$^{(2)}$ in terms of RMSE, and works better than Gsynth$^{(1)}$ when T_0 is small.
4. The uncertainty estimates produced by DM-LFM give relatively good coverage rates in all three setups. When T_0 is small, Gsynth slightly over-covers the truth while DM-LFM slightly under-covers it.
5. Gsynth is computationally the most efficient, while DM-LFM takes the longest time and its run-time grows rapidly as the sample size grows.

Table A5. Monte Carlo: Single Treated Unit w/ Three Strong Factors

		Bias					RMSE				
N_{co}	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20	-0.03	-0.05	-0.10	-0.14	-0.02	4.38	4.05	2.39	2.65	2.52
30	40	-0.08	0.11	0.05	0.03	0.08	4.16	3.91	1.83	1.93	1.84
30	60	-0.19	-0.03	-0.03	-0.05	-0.02	3.89	3.57	1.84	1.86	1.87
50	20	-0.19	-0.16	0.03	0.14	0.03	4.18	3.44	2.15	2.43	2.14
50	40	0.24	0.13	0.06	0.10	0.09	4.03	3.63	1.82	1.86	1.85
50	60	0.15	-0.00	0.04	0.03	0.06	4.00	3.38	1.79	1.83	1.84
		Coverage					Run Time (sec)				
N	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20			0.93	0.95	0.92	2.31	13.26	0.32	0.33	11.00
30	40			0.96	0.96	0.96	2.34	11.96	0.37	0.42	16.18
30	60			0.94	0.94	0.95	2.31	17.32	0.42	0.49	20.65
50	20			0.96	0.96	0.94	3.56	20.19	0.38	0.39	15.28
50	40			0.93	0.93	0.94	3.43	36.85	0.58	0.62	22.43
50	60			0.94	0.95	0.93	3.36	25.19	0.66	0.73	30.09

Note: The above table compares the performance of the synthetic control method (SCM), the generalized synthetic control method (Gsynth), and the Bayesian DM-LFM in estimating the treatment effects on a single treated unit using Monte Carlo exercises. N_{co} and T_0 represent the number of control units and the number of pretreatment periods in each setting. SCM$^{(1)}$ and SCM$^{(2)}$ stand for the SCM using 5 and (T_0-5) pretreatment outcomes for matching, respectively. Gsynth$^{(1)}$ and Gsynth$^{(2)}$ represent Gsynth using the correct number of factors ($r = 3$) and Gsynth that cross-validates the number of factors. For Synth, we obtain the point estimates. Each cell is produced based on results from 500 simulations.

Table A6. Monte Carlo: Single Treated Unit w/ Eight Weak Factors

		Bias					RMSE				
N	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20	0.03	0.10	-0.03	0.10	0.07	3.54	3.62	3.45	3.45	3.09
30	40	-0.21	-0.28	-0.02	-0.12	-0.04	3.50	3.72	2.78	2.89	2.80
30	60	-0.01	0.11	-0.05	-0.04	-0.02	3.61	4.01	2.42	2.48	2.53
50	20	-0.02	-0.07	0.10	0.20	0.07	3.74	3.80	3.55	3.73	2.96
50	40	-0.10	-0.26	0.17	0.18	0.09	3.62	3.61	2.51	2.57	2.50
50	60	-0.13	-0.16	-0.08	-0.10	-0.06	3.41	3.53	2.16	2.21	2.17
		Coverage					Run Time (sec)				
N	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20			0.97	0.97	0.92	16.96	24.43	0.80	0.81	26.49
30	40			0.95	0.97	0.91	18.06	28.75	1.01	1.12	42.10
30	60			0.91	0.94	0.91	17.47	42.65	1.10	1.28	52.33
50	20			0.97	0.97	0.92	26.11	40.30	1.04	1.03	40.65
50	40			0.96	0.96	0.92	25.03	89.98	1.60	1.70	58.53
50	60			0.96	0.96	0.94	24.37	64.79	1.83	2.03	79.88

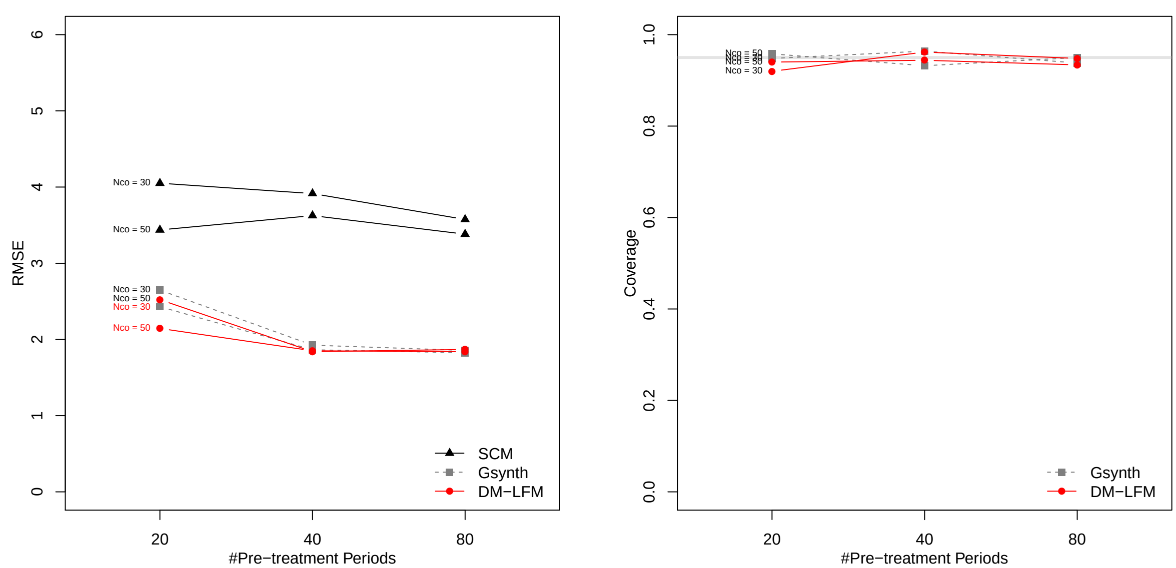
Note: The setup is the same as that in Table A5 except that the DGP includes eight relatively weak factors instead of three relatively strong factors. Different from Table A5, SCM$^{(1)}$ stands for the SCM using eight pretreatment outcomes for matching, while SCM$^{(2)}$ still uses (T_0-5) pretreatment outcomes.

Table A7. Monte Carlo: Single Treated Unit w/ Three Strong Factors and Time-Invariant Covariates

		Bias					RMSE				
N	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20	0.25	0.55	0.23	0.23	0.34	5.45	6.17	3.71	4.01	3.69
30	40	0.22	0.39	0.01	0.03	0.10	4.85	5.59	2.60	2.63	2.64
30	60	0.74	0.73	0.11	0.14	0.18	5.04	5.84	2.37	2.44	2.45
50	20	-0.18	0.04	-0.30	-0.23	-0.14	5.23	5.74	3.63	3.68	3.30
50	40	0.09	0.19	-0.04	-0.08	0.02	4.92	5.91	2.45	2.48	2.50
50	60	0.73	0.77	0.09	0.08	0.16	4.80	5.73	2.23	2.27	2.30
		Coverage					Run Time (sec)				
N	T_0	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM	SCM$^{(1)}$	SCM$^{(2)}$	Gsynth$^{(1)}$	Gsynth$^{(2)}$	DM-LFM
30	20			0.96	0.97	0.91	17.81	36.08	0.76	0.79	24.36
30	40			0.96	0.97	0.94	16.73	30.61	0.84	0.94	34.03
30	60			0.95	0.95	0.93	16.25	43.24	0.93	1.08	42.84
50	20			0.97	0.97	0.93	28.67	61.49	1.03	1.06	39.25
50	40			0.97	0.96	0.95	27.80	109.09	1.59	1.70	57.39
50	60			0.97	0.97	0.94	28.07	82.17	1.84	2.04	79.73

Note: The setup is the same as that in Table A5 except that the simulated data now include six time-invariant covariates, four of which enter the DGP with time-varying coefficients. Because Gsynth cannot accommodate time-invariant covariates with time-varying coefficients, Gsynth$^{(1)}$ estimates a seven factor model—three account for the unobserved factors and four for the influence of the covariates. Gsynth$^{(2)}$ uses a cross-validation scheme to set the number of factors.

Figure A16. Single Treated Unit: Comparing Different Methods (Setup 1)



(a) RMSE

(b) Coverage

Note: The left panel shows the RMSE of ATT estimates using SCM$^{(2)}$, Gsynth$^{(2)}$, and DM-LFM. The right panel shows the coverage rates of 95 confidence/credibility intervals using Gsynth$^{(2)}$ and DM-LFM.

References

Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science 59*(2), 495–510.

Belmonte, M. A., G. Koop, and D. Korobilis (2014). Hierarchical Shrinkage in Time-Varying Parameter Models. *Journal of Forecasting 33*(1), 80–94.

Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika 81*(3), 541–553.

de Finetti, B. (1963). Foresight: Its Logical Laws, Its Subjective Sources. In jr. Henry E. Kyburg and H. E. Smokler (Eds.), *Studies in Subjective Probability*. New York: John Wiley.

Liu, L., Y. Wang, and Y. Xu (2020). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. Mimeo, Stanford University.

Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Eects Models. *Political Analysis 25*(1), 57–76.

Yu, Y. and X.-L. Meng (2011). To Center or Not to Center: That is Not the Question—an Ancillarity—Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.