Improved Inference for Interactive Fixed Effects Model with Cross Sectional Dependence

Zhenhao Gong
University of Connecticut

Outline

- Consider

$$Y_{it} = X_{it}'\beta_0 + u_{it},\ u_{it} = \lambda_i' F_t + \varepsilon_{it}.$$

  - X_{it} is potentially correlated with λ_i or F_t alone, or both.
  - ε_{it} is allowed to be weakly correlated and heteroskedastic.
  - Let $\lambda_i = (\alpha_i, 1)'$ and $F_t = (1, \delta_t)'$, then $\lambda_i' F_t = \alpha_i + \delta_t$;
  - Model the unobservable common shocks to impact the cross-sectional units heterogeneously.
    $\Rightarrow$ More flexible and general than the fixed effects model.
  - Incidental parameters problem in estimation.

- $(\hat{\beta}, \hat{F}, \hat{\Lambda})$ minimizes

$$\mathrm{SSR}(\beta_0, F, \Lambda) = \sum_{i=1}^{N} (Y_i - X_i\beta_0 - F\lambda_i)' (Y_i - X_i\beta_0 - F\lambda_i),$$

  subject to $F'F/T = I_r$ and $\Lambda'\Lambda$ being diagonal.

- Concentrating out Λ, the LS estimator for β_0 given F is:

$$\hat{\beta}(F) = \left(\sum_{i=1}^{N} X_i' M_F X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_F Y_i,$$

  where $M_F = I_T - F(F'F)^{-1}F'$.

- Given β_0, the model reduces to a pure factor model, so we can estimate F using PCA:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}\right)\left(Y_i - X_i\hat{\beta}\right)'\right]\hat{F} = \hat{F}V_{NT}.$$

- The solution $(\hat{\beta}, \hat{F})$ can be obtained by iteration until convergence. Given $(\hat{\beta}, \hat{F})$, $\hat{\Lambda} = T^{-1}(Y - X\hat{\beta})'\hat{F}$ by least square.

Asymptotic Bias

- Bai (2009) shows that as $N, T \to \infty$, under some regularity assumptions and if $T/N \to \rho > 0$,

$$\sqrt{NT}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(\rho^{1/2} B_0 + \rho^{-1/2} C_0, H_0^{-1} H_Z H_0^{-1}\right).$$

  where B_0 and C_0 arise from cross-sectional and serial correlations and heteroskedasticities in ε_{it} $\Rightarrow$ invalid inference.

- In the presence of serial correlation, we can control the bias by the truncated kernel method of Newey and West (1987).
- **Goal:** Developing a valid inference procedure under cross-sectional correlation and heteroskedasticity, assuming no serial correlation ($C_0 = 0$).
  - Bias correction (Estimate B_0).
  - Robust estimation of H_Z.

- The asymptotic bias is the probability limit of B_{NT} with

$$B_{NT} = -H\left(F\right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} w_i \lambda_k \left(\frac{1}{T} \sum_{t=1}^{T} E \varepsilon_{it} \varepsilon_{kt} \right),$$

where

$$H(F) = \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' M_F X_k a_{ik} \right],$$

$$w_i = \operatorname{plim} \left[\frac{\left(X_i - V_i\right)' F^0}{T} \right] \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda' \Lambda}{N} \right)^{-1},$$

$$V_i = \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k, \text{ and } a_{ik} = \lambda_i' (\Lambda' \Lambda / N)^{-1} \lambda_k.$$

- Panel data models with interactive fixed effects:
  - Ahn et al.(2001); Pesaran (2006); Bai, (2009); Kao et al. (2012); Moon and Weidner (2015); Su et al. (2015); Harding and Lamarche (2014); etc.
  - Empirical studies: Kim and Oka (2013); Gobillon and Magnac (2016); Totty (2017); etc.
- Methods for cross-sectional correlated bias:
  - Bai (2009): CS-HAC method.
  - Bai and Liao (2017): GLS method.
- The spatial HAC method:
  - Conley (1996, 1999); Conley and Molinari (2007); Kelejian and Prucha (2007); Kim and Sun (2011, 2013); Bester et al.,(2017); Mueller and Watson (2021); etc.

1. CS-HAC estimator (Bai, 2009)

$$\hat{B}_{NT} = -\hat{H}_0^{-1} \frac{1}{n_{sub}} \sum_{i=1}^{n_{sub}} \sum_{k=1}^{n_{sub}} \hat{w}_i \hat{\lambda}_k \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{kt} \right),$$

where $H_0 = \text{plim} H(F)$; $\hat{H}_0$ and $\hat{w}_i$ are the estimators of H_0 and w_i with F, λ_i, and Λ replaced by $\hat{F}$, $\hat{\lambda}_i$, and $\hat{\Lambda}$.

- Consistent as $n_{sub}/N \to 0$ and $n_{sub}/T \to 0$.
- Hard to implement properly. Performance highly depends on the sub-sample selection and there is no practical guidance to select.

2. GLS method (Bai and Liao, 2017)

$$\hat{\beta}(\Sigma_{\varepsilon}^{-1}) = arg\min_{\beta} \sum_{t=1}^{T} (Y_t - X_t\beta - \Lambda F_t)\, \Sigma_{\varepsilon}^{-1} \,(Y_t - X_t\beta - \Lambda F_t)\,,$$

where $\Sigma_{\varepsilon} = \underbrace{cov(\varepsilon_t)}_{N\times N}$.

- They assume
  - Σ_{ε} is sparse and apply the thresholding method.
  - serial independence in $\{\varepsilon_t : t \geq 1\}$.

- The major advantages of $\hat{\beta}(\Sigma_{\varepsilon}^{-1})$ are:
  - Faster rate of convergence when $N = o(T)$.
  - Smaller asymptotic variance.
  - No need to correct the bias.
  - No need to use the robust covariance matrix.

- Estimating Procedures:
  1. Obtain $\hat{\beta}$, $\hat{\Lambda}$, $\hat{F}_t$, and $\hat{\varepsilon}_t = Y_t - X_t\hat{\beta} - \hat{\Lambda}\hat{F}_t$. (Bai, 2009)
  2. Compute
  $$\hat{R} = (\hat{r}_{ij})_{N\times N} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'.$$
  3. Apply adaptive thresholding
  $$\hat{\Sigma}_\varepsilon = (s_{ij}(\hat{r}_{ij}))_{N\times N}, \quad s_{ij}(\cdot) : \text{hard, soft} \cdots .$$
  4. Using $\hat{\beta}$ as the initial value and $\hat{\Sigma}_\varepsilon^{-1}$ as the weight matrix, obtain $\hat{\beta}(\hat{\Sigma}_\varepsilon^{-1})$.

- Thresholds
  - Hard-thresholding: $s(t) = t1\{|t| > \tau\}$.
  - Soft-thresholding: $s(t) = (t - \tau)1\{|t| > \tau\}$ when $t > 0$; symmetric when $t < 0$.
  - Threshold τ_{ij}:

$$\tau_{ij} = C\sqrt{\hat{R}_{ii}\hat{R}_{jj}}\left(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}\right),$$

where the tuning parameter $C > 0$ can be chosen through multifold cross-validation in practice.

- Consider the following DGP:

$$
\begin{aligned}
&Y_{it} = X_{it}\beta + \lambda_i' F_t + \varepsilon_{it},\\
&X_{it} = \mu + c\lambda_i' F_t + \iota'\lambda_i + \iota' F_t + \eta_{it}, \iota' = (1,1);\\
&F_{rt} = \rho F_{r,t-1} + \sqrt{1-\rho^2}u_{rt}, r = 1,2;\\
&\lambda_{ir}, \eta_{it}, u_{rt} \overset{iid}{\sim} N(0,1).
\end{aligned}
$$

- We set $\beta = \mu = c = 1$ and $\rho = 0.3$. The number of common factors is two, and is assumed to be known.

- We generates cross-sectional dependent data using a popular spatial MA model:

$$\varepsilon_t = \left(I_N + \theta M_1 + \theta^2 M_2\right) v_t,$$
$$v_t = (v_{1t}, \ldots, v_{Nt})', v_{it} \overset{iid}{\sim} N(0,1).$$

- $M_1 = [m_{1,ik}]_{i,k=1}^N$ and $M_2 = [m_{2,ik}]_{i,k=1}^N$ are $(N \times N)$ spatial weight matrices such that

$$m_{1,ik} = \begin{cases} 1 & \text{if } d_{ik} = 1 \\ 0 & \text{if } d_{ik} \neq 1 \end{cases} \text{ and } m_{2,ik} = \begin{cases} 1 & \text{if } d_{ik} = \sqrt{2} \\ 0 & \text{if } d_{ik} \neq \sqrt{2} \end{cases}.$$

Table: $\sqrt{NT}$ scaled Bias and RMSE of the interactive estimator $\hat{\beta}$ and GLS estimator $\hat{\beta}_{gls}$ comparison. Null rejection probabilities, 5% level.

T	N	$B(\hat{\beta})$	RMSE	Rej	$B(\hat{\beta}_{gls})$	RMSE	Rej
				$\theta = 0$			
50	144	0.851	1.077	0.073	0.878	1.103	0.089
100	144	0.833	1.052	0.071	0.846	1.072	0.076
150	144	0.840	1.057	0.066	0.852	1.074	0.075
200	144	0.788	0.978	0.052	0.798	0.992	0.042
				$\theta = .4$			
50	144	1.513	1.975	0.199	0.801	1.021	0.051
100	144	1.553	1.991	0.202	0.542	0.679	0.166
150	144	1.563	2.004	0.226	0.447	0.563	0.278
200	144	1.720	2.148	0.261	0.421	0.523	0.349

TA-SHAC Estimator

- **We propose TA-SHAC estimators to improve the inference of $\hat{\beta}$ by correcting the bias and estimating the covariance matrix.**
- Recall

$$
\begin{aligned}
B_{NT} &= -H\left(F\right)^{-1} \underbrace{\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} w_i \lambda_k \left(\frac{1}{T}\sum_{t=1}^{T} E\varepsilon_{it}\varepsilon_{kt}\right)}_{=J_{NT}}, \\
&= -H\left(F\right)^{-1} J_{NT}.
\end{aligned}
$$

- $H\left(F\right)$ is easy to estimate, our focus is on consistent estimation of J_{NT}.

- The TA-SHAC estimator of J_{NT} is given by

$$
\begin{aligned}
\hat{J}_{NT} &= \frac{1}{T}\sum_{t=1}^{T}\underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} K\left(\frac{d_{ik}}{d_n^1}\right)\hat{w}_i\hat{\varepsilon}_{it}\hat{\varepsilon}_{kt}\hat{\lambda}_k\right]}_{=\hat{J}_t} \\
&= \frac{1}{T}\sum_{t=1}^{T}\hat{J}_t.
\end{aligned}
$$

- $\hat{J}_t$ is a standard spatial HAC estimator in the literature and $\hat{J}_{NT}$ can be viewed as a time average of $\hat{J}_t, t = 1, \cdots, T$.
- $K(\cdot)$ is a real-valued kernel function. d_{ik} is the distance measure between i and k and d_n^1 is a bandwidth parameter.

- To establish the consistency of $\hat{J}_{NT}$, we introduce an infeasible estimator $\tilde{J}_{NT}$:

$$\tilde{J}_{NT} = \frac{1}{T}\sum_{t=1}^{T} \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} K\left(\frac{d_{ik}}{d_n^1}\right) w_i \varepsilon_{it}\varepsilon_{kt}\lambda_k\right]}_{=\tilde{J}_t}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\tilde{J}_t.$$

- We have

$$\hat{J}_{NT} - J_{NT} = \underbrace{(\hat{J}_{NT} - \tilde{J}_{NT})}_{\text{Estimation error}} + \underbrace{(\tilde{J}_{NT} - E\tilde{J}_{NT})}_{\text{Variation}} + \underbrace{(E\tilde{J}_{NT} - J_{NT})}_{\text{Bias}}.$$

- We assume that ε_{it} has a linear representation:

$$\varepsilon_{it} = \sum_{\ell=1}^{\infty} c_{it,\ell} e_{\ell},$$

where $\{c_{it,\ell}\}$ are unknown constants and $e_\ell \overset{iid}{\sim} (0,1)$.

- We can rely on Andrews (1991), Kim and Sun (2011, 2013) to show that

$$\tilde{J}_{NT} - E\tilde{J}_{NT} = O\left(\sqrt{\frac{\ell_N}{NT}}\right) \text{ and } E\tilde{J}_{NT} - J_{NT} = O\left(\frac{1}{d_n^q}\right),$$

where

$$\ell_i = \sum_{k=1}^{N} 1\left\{d_{ik} \leq d_n\right\} \text{ and } \ell_N = \frac{1}{N}\sum_{i=1}^{N} \ell_i.$$

- Based on the arguments in Bai (2009), we can show that

$$\hat{J}_{NT} - \tilde{J}_{NT} = o\left(1\right).$$

Robust Estimation of H_Z

- Recall under cross-sectional dependence

$$\sqrt{NT}\left(\hat{\beta}-\beta_0\right)\xrightarrow{d}N\left(\rho^{1/2}B_0, H_0^{-1}H_ZH_0^{-1}\right),$$

where $H_Z = \text{plim}\,\frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{t=1}^{T}E(\varepsilon_{it}\varepsilon_{kt})Z_{it}Z_{kt}'$, and $Z_i = M_{F_0}X_i - \frac{1}{N}\sum_{k=1}^{N}a_{ik}M_{F_0}X_k$.

- H_Z is conventional estimated as

$$\hat{H}_Z = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2\left(\frac{1}{T}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\right),$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$. Not valid with cross-sectional dependence.

- Bai (2009) suggests CS-HAC estimator

$$\hat{H}_{CS} = \frac{1}{n_{sub}} \sum_{i=1}^{n_{sub}} \sum_{k=1}^{n_{sub}} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{Z}_{it} \hat{Z}'_{kt} \hat{\varepsilon}_{it} \hat{\varepsilon}_{kt} \right).$$

  - Hard to implement in practice.
  - Sensitive to the sub-sample selection.
  - Still consistent if the whole sample size is used,

$$\tilde{H}_{CS} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{Z}_{it} \hat{Z}'_{kt} \hat{\varepsilon}_{it} \hat{\varepsilon}_{kt} \right).$$

- We propose TA-SHAC estimator given by,

$$\hat{H}_{NT} = \frac{1}{T}\sum_{t=1}^{T}\underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{Z}_{it}\hat{Z}'_{kt}\hat{\varepsilon}_{it}\hat{\varepsilon}_{kt}\mathfrak{K}_F\left(\frac{d_{ik}}{d_n^2}\right)\right]}_{\hat{H}_t} = \frac{1}{T}\sum_{t=1}^{T}\hat{H}_t.$$

- We use the flat-top kernels $\mathfrak{K}_F$ to include the estimator $\tilde{H}_{CS}$ as a special case:

$$\mathfrak{K}_F = \left(\mathfrak{K}(\cdot) : \mathfrak{K}(x) = \begin{cases} 1 & \text{if } |x| \le c_F \\ \mathcal{G}(x) & \text{otherwise} \end{cases}\right),$$

where $c_F \le 1$ and $\mathcal{G} : |x| \in (c_F, 1] \to [0, 1]$. e.g. Trapezoidal kernel: $\mathcal{G}(x) = \max\{(|x| - 1)/(c_F - 1), 0\}$.

There are two major challenges in implementing our method:

- **How to choose the distance measure?**
  - Auxiliary variable: transportation cost (Conley and Ligon, 2000), geographic distance (Pinkse et al., 2002), etc.
    $\Rightarrow$ May not be available and appropriate.
- **How to select the bandwidths jointly?**
  - Fixed bandwidth (Kelejian and Prucha, 2007), bandwidth selection by asymptotic MSE (Kim and Sun, 2011)
    $\Rightarrow$ May not applicable to our estimators.

- We define the distance that reflects the dependence structure directly,

$$d_{ik} = |1/\rho_{ik}| - 1.$$

where $\rho_{ik} = corr(\varepsilon_{it}, \varepsilon_{kt})$. d_{ik} is unobservable but we can use the sample counter part,

$$\tilde{d}_{ik} = \min\{1/|\tilde{\rho}_{ik}|, 100\} - 1,$$

where $\tilde{\rho}_{ik} = \sum_{t=1}^{T} \tilde{\varepsilon}_{it}\tilde{\varepsilon}_{kt} / \sqrt{\sum_{t=1}^{T} \tilde{\varepsilon}_{it}^2 \sum_{t=1}^{T} \tilde{\varepsilon}_{kt}^2}$.

  - Mantegna (1998), Fernandez (2011), Kim (2020), etc.
  - No need prior information for implementation.
  - Does not satisfy triangle inequality.

- Kim and Sun (2017) use a simulation-based choice in time series kernel method to select two smoothing parameters in their test procedure.
- How to replicate cross-sectional dependence?
  - Silvia Goncalves (2011)
  - Timothy Vogelsang (2012)
  - Javier Hidalgo and Marcia Schafgans (2017)
- Bootstrap-based procedure: avoid to use the parameter model in time series case.

- We consider a bootstrap-based bandwidth selection procedure. Let $D_{nM}^1 = \{d_{n1}^1, \cdots, d_{nM}^1\}$ and $D_{nS}^2 = \{d_{n1}^2, \cdots, d_{nS}^2\}$ be the sets of reasonable d_n^1 and d_n^2.
  1. Estimate $\hat{\beta}$, $\hat{F}_t$, $\hat{\Lambda}$, and $\hat{\varepsilon}_t = Y_t - X_t\hat{\beta} - \hat{\Lambda}\hat{F}_t$. (Bai, 2009)
  2. Generate bootstrap sample Y_t^* based on $Y_t^* = X_t\hat{\beta} + \hat{\Lambda}\hat{F}_t + \varepsilon_t^*$ with $\varepsilon_t^* = \hat{\varepsilon}_t\xi_t$, and $\xi_t \overset{iid}{\sim} (0,1)$ (e.g. Rademacher distribution).
  3. Estimate $\hat{\beta}^*$, $\hat{F}_t^*$, $\hat{\Lambda}^*$, and $\hat{\varepsilon}_t^*$. Construct the bootstrap version of the bias term $\hat{B}_{NT}^*\left(d_{nm}^1\right)$ by TA-SHAC with $d_{nm}^1 \in D_{nM}^1$.
  4. Estimate the bootstrap version of covariance matrix $\hat{H}_{NT}^*\left(d_{ns}^2\right)$ by TA-SHAC with $d_{ns}^2 \in D_{nS}^2$.

5. Generate B bootstrap samples. Compute the bootstrap-based t-test statistics:

$$t_k^*(d_{nm}^1, d_{ns}^2) = \hat{\beta}_{cor}^* / se(\hat{\beta}^*), \text{ for } k = 1, 2, \cdots, B;$$

$$\text{with} \quad \hat{\beta}_{cor}^* = \hat{\beta}^* - \frac{1}{N} B_{NT}^* \left(d_{nm}^1\right) \quad \text{and}$$

$$se(\hat{\beta}^*) = \sqrt{\frac{H(\hat{F}^*)^{-1} \hat{H}_{NT}^* \left(d_{ns}^2\right) H(\hat{F}^*)^{-1}}{NT}}.$$

6. For each pair of $(d_{nm}^1, d_{ns}^2) \in (D_{nM}^1, D_{nS}^2)$, compute the corresponding 5% rejection ratio in step 5, and select $\left(d_{nm}^{1*}, d_{ns}^{2*}\right)$ that gives the rejection ratio closest to 0.05.

- Consider the following DGP:

$$
\begin{aligned}
&Y_{it} = X_{it}\beta + \lambda_i' F_t + \varepsilon_{it}, \\
&X_{it} = \mu + c\lambda_i' F_t + \iota'\lambda_i + \iota' F_t + \eta_{it}, \iota' = (1,1); \\
&F_{rt} = \rho F_{r,t-1} + \sqrt{1-\rho^2} u_{rt}, r = 1,2; \\
&\lambda_{ir}, \eta_{it}, u_{rt} \overset{iid}{\sim} N(0,1).
\end{aligned}
$$

- We set $\beta = \mu = c = 1$ and $\rho = 0.3$. The number of common factors is two, and is assumed to be known.

- We generates cross-sectional dependent data using a popular spatial MA model.

$$\varepsilon_t = \left(I_N + \theta M_1 + \theta^2 M_2\right) v_t,$$
$$v_t = (v_{1t}, \ldots, v_{Nt})', v_{it} \overset{iid}{\sim} N(0,1).$$

- $M_1 = [m_{1,ik}]_{i,k=1}^N$ and $M_2 = [m_{2,ik}]_{i,k=1}^N$ are $(N \times N)$ spatial weight matrices such that

$$m_{1,ik} = \begin{cases} 1 & \text{if } d_{ik} = 1 \\ 0 & \text{if } d_{ik} \neq 1 \end{cases} \text{ and } m_{2,ik} = \begin{cases} 1 & \text{if } d_{ik} = \sqrt{2} \\ 0 & \text{if } d_{ik} \neq \sqrt{2} \end{cases}.$$

Table: Scaled Bias and RMSE of different estimators

						TA-SHAC (d_{ik}^T)		TA-SHAC (d_{ik}^D)	
T	N	$B(\hat{\beta})$	RMSE	$B(\hat{\beta}_{gls})$	RMSE	$B(\tilde{\beta}^*_{hac})$	RMSE	$B(\hat{\beta}^*_{hac})$	RMSE
					$\theta = .4$				
50	144	1.597	2.019	0.897	1.137	1.426	1.807	1.492	1.892
100		1.584	1.956	0.601	0.756	1.308	1.704	1.393	1.728
150		1.642	2.072	0.491	0.617	1.367	1.734	1.383	1.764
200		1.660	2.087	0.453	0.577	1.426	1.816	1.346	1.697
50	196	1.442	1.851	0.837	1.069	1.336	1.703	1.361	1.742
100		1.368	1.708	0.550	0.686	1.260	1.624	1.261	1.568
150		1.387	1.766	0.454	0.566	1.235	1.560	1.220	1.560
200		1.475	1.861	0.428	0.535	1.228	1.525	1.264	1.584

Note: Scaled bias equals the difference between each estimator and its true value scaled by $\sqrt{NT}$. RMSE is the corresponding root mean square error scaled by $\sqrt{NT}$. d_{ik}^T denotes the true distance. d_{ik}^D denotes the data driven distance measure.

Table: 95% empirical coverage rates of different estimators

				TA-SHAC (d_{ik}^T)			TA-SHAC (d_{ik}^D)		
T	N	$\hat{\beta}$	$\hat{\beta}_{gls}$	$\tilde{\beta}_{hac1}$	$\tilde{\beta}_{hac2}$	$\tilde{\beta}^*_{hac}$	$\hat{\beta}_{hac1}$	$\hat{\beta}_{hac2}$	$\hat{\beta}^*_{hac}$
					$\theta = .4$				
50	144	0.771	0.969	0.829	0.824	0.864	0.817	0.796	0.849
100		0.800	0.902	0.834	0.851	0.867	0.821	0.843	0.879
150		0.777	0.797	0.806	0.854	0.878	0.796	0.841	0.860
200		0.754	0.734	0.772	0.821	0.854	0.786	0.853	0.879
50	196	0.809	0.972	0.846	0.837	0.877	0.843	0.835	0.868
100		0.855	0.911	0.866	0.881	0.898	0.874	0.885	0.902
150		0.842	0.784	0.876	0.890	0.908	0.857	0.880	0.894
200		0.823	0.678	0.872	0.902	0.921	0.847	0.896	0.911

Note: For $\hat{\beta}_{hac1}$, we estimate covariance matrix only by TA-SHAC without bias correction. For $\hat{\beta}_{hac2}$, we correct the bias only by TA-SHAC. We correct the bias and estimate the covariance matrix by TA-SHAC for $\hat{\beta}^*_{hac}$.

1. **Effects of divorce law reforms**

- **Background:** During and after 1970s, most of states in U.S. shifted from a consent divorce regime to no-fault unilateral divorce laws. The new laws allowed people to seek a divorce without the consent of their spouse.
- **Research question:** the causal relationships between divorce law reforms and divorce rates.

- **Literature:**
  - Allen (1992), Peters (1982,1992) suggested that divorce rates were unaffected by the change of the laws.
  - After controlling for fixed state and year effects, Firedberg (1998) found that states' law reforms have contributed to one-sixth of the rise in state-level divorce rates in the first eight years following reforms, but longer effects is unclear.
  - By using a standard fixed effects panel data model, Wolfers (2006) identified negatives effects of law reforms on the divorce rates after nine years to fourteen years in most states.

- Specifically, Wolfers (2006) studied the model as

$$
\begin{aligned}
y_{st} &= T_{st} + f(v_{st}, t) + u_{st}, \\
u_{st} &= \delta_s + \alpha_t + \varepsilon_{st},
\end{aligned}
$$

  where y_{st} is the annual divorce rates; $f(v_{st}, t)$ is the time trend; δ_s and α_t are the state and the time fixed effects.

- The treatment effects T_{st} is

$$
\begin{aligned}
T_{st} =& \mathbf{1}_{T_s \le t \le T_s+1}\beta_1 + \mathbf{1}_{T_s+2 \le t \le T_s+3}\beta_2 \\
& + \cdots + \mathbf{1}_{T_s+12 \le t \le T_s+13}\beta_7 + \mathbf{1}_{T_s+14 \le t}\beta_8,
\end{aligned}
$$

  where T_s is the law reform year of state s.

- The robustness of Wolfers (2006) has been doubted, due to the additive structure in u_{st} is not flexible to capture factors varying across time and state, and cross-sectional correlation in error terms.
- Kim and Oka (2014) apply the interactive fix effects model proposed by Bai (2009) for the study, in which model u_{st} as

$$u_{st} = \lambda_s' F_t + \varepsilon_{st},$$

where F_t is principle components of u_{it}, which dominant the portion of divorce rates not explained by the included regressors. λ_s stands for the heterogeneous effect of F_t to each state.

- To correct the cross-sectional correlation bias and make efficient estimation, Bai and Liao (2017) re-estimate the model of Kim and Oka (2014) using the GLS method.
- We apply the proposed method to correct the cross-sectional correlation bias and provide valid inference.

Table: Methods comparison in effects of divorce law reform

	$\hat{\beta}$		$\hat{\beta}^*_{hac}$		$\hat{\beta}_{gls}$	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
First 2 years	0.0183**	[0.003, 0.034]	0.0175*	[-0.002, 0.037]	0.0138**	[0.000, 0.027]
3–4 years	0.0418***	[0.020, 0.064]	0.0402***	[0.016, 0.065]	0.0340***	[0.014, 0.054]
5–6 years	0.0322**	[0.004, 0.060]	0.0297**	[0.004, 0.057]	0.0249**	[0.000, 0.050]
7–8 years	0.0293*	[-0.005, 0.063]	0.0266	[-0.005, 0.061]	0.0152	[-0.015, 0.045]
9–10 years	0.0073	[-0.032, 0.047]	0.0039	[-0.034, 0.046]	-0.0061	[-0.040, 0.028]
11–12 years	0.0092	[-0.037, 0.051]	0.0055	[-0.038, 0.051]	-0.0078	[-0.044, 0.028]
13–14 years	0.0050	[-0.041, 0.051]	0.0008	[-0.048, 0.052]	-0.0092	[-0.048, 0.029]
15 years+	0.0306	[-0.020, 0.081]	0.0264	[-0.027, 0.084]	0.0093	[-0.033, 0.052]

Note: 95 % confidence intervals are reported. The number of factors $r = 10$.
$^{*}\ p < .1$. $^{**}\ p < .05$. $^{***}\ p < .01$.

2. **Effects of water and sewerage interventions**

- **Background:** From 1880 to 1920, when Boston authorities developed a sewerage and water district, infant mortality plummeted from around 1/5 to 1/16 white infants, and deaths of noninfants under 5 years decreased by a factor of seven in Massachusetts.
- **Research question:** the causal relationships between water and sewerage interventions and child mortality.

- **Literature:**
  - Cutler and Miller (2015) studied the impact of water chlorination and filtration on the death rate from waterborne diseases across 13 US cities. Their results suggest that improved water quality decreases 47 percent in log infant mortality from 1900 to 1936.
  - Alsan and Goldin (2019) exploited the independent and combined effects of clean water and effective sewerage systems on under-5 mortality in Massachusetts, 1880-1920. They identified the two interventions together account for approximately one-third of the decline in log child mortality during the 41 years.

- Specifically, Alsan and Goldin (2019) estimate

$$
\begin{aligned}
y_{it} &= \mu + \beta_1 W_{it} + \beta_2 S_{it} + \beta_3 (W * S)_{it} + \Omega X_{it} + u_{it}, \\
u_{it} &= \delta_i + \alpha_t + \delta_i t + \varepsilon_{it},
\end{aligned}
$$

  - i is municipality and t is year; y_{it} is the log under-5 mortality rate. X_{it} is a vector of time- and municipality-varying demographic controls.
  - u_{it} captures the unobserved heterogeneity, which includes municipality and time fixed effects, municipality-specific time trends.
  - The standard errors are clustered at the municipality level with 60 clusters in their analysis.

- Since they used the municipality level data, the potential unobserved heterogeneity and cross-sectional correlation in the errors may affect the results.
- To check the robustness of their results, we first apply the interactive fixed effects model with u_{it} as

$$u_{it} = \lambda_i' F_t + \varepsilon_{it},$$

where F_t dominant the portion of child mortality rates not explained by the included regressors. λ_i stands for the heterogeneous effect of F_t to each municipality.

- Then, we apply the proposed method to correct the bias and provide valid inference for the interactive fixed effects model.
- Note that if we let $\lambda_i = (\delta_i, 1, \delta_i)'$ and $F_t = (1, \alpha_t, t)'$, then u_{it} in above equations are the same. Hence, We choice three factors in our model to include the original model as a special case.
- Finally, we apply the GLS method for the study to compare with our method.

Table: Estimated effects of clean water and sewerage

	Panel A. Standard Fixed Effects				
	(1)	(2)	(3)	(4)	(5)
Safe water	-0.127		-0.102		0.108
	[-0.280, 0.026]		[-0.252, 0.047]		[-0.043, 0.258]
Sewerage		-0.124***	-0.106**		-0.068
		[-0.214, -0.033]	[-0.194, -0.018]		[-0.156, 0.021]
Interaction				-0.239***	-0.307***
				[-0.395 -0.084]	[-0.509, -0.106]
	Panel B. Interactive Fixed Effects				
Safe water	-0.060***		-0.051**		0.126***
	[-0.103, -0.017]		[-0.096, -0.006]		[0.055, 0.197]
Sewerage		-0.052***	-0.042**		-0.003
		[-0.092, -0.013]	[-0.085, 0.001]		[-0.045, 0.044]
Interaction				-0.151***	-0.262***
				[-0.198, -0.104]	[-0.346, -0.177]

Note: 95 % confidence intervals are reported.
* $p < .1$. ** $p < .05$. *** $p < .01$.

Table: Estimated effects of clean water and sewerage

	(1)	(2)	(3)	(4)	(5)
	Panel C. TA-SHAC Estimation				
Safe water	-0.056		-0.048		0.119**
	[-0.126, 0.012]		[-0.120, 0.022]		[0.013, 0.225]
Sewerage		-0.049*	-0.039		-0.003
		[-0.107, 0.009]	[-0.100, 0.022]		[-0.068, 0.062]
Interaction				-0.147***	-0.252***
				[-0.218, -0.076]	[-0.376, -0.128]
	Panel D. GLS Estimation				
Safe water	-0.021		-0.020		0.116***
	[-0.074, 0.033]		[-0.075, 0.034]		[0.028, 0.205]
Sewerage		-0.024	-0.023		0.006
		[-0.071, 0.023]	[-0.072, 0.025]		[-0.044, 0.058]
Interaction				-0.100***	-0.205***
				[-0.159, -0.040]	[-0.310, -0.101]

Note: 95 % confidence intervals are reported.
* $p < .1$. ** $p < .05$. *** $p < .01$.

Conclusion

1. We propose TA-SHAC approach to improve the inference for interactive fixed effects model with cross-sectional dependence and heteroskedasticity.
2. We establish the consistency and rate of convergence of the TA-SHAC estimators.
3. To implement our approach, we develop a data driven distance that does not rely on prior information and bandwidth selection procedure based on bootstrap method.
4. We show that our approach performs well in simulation with finite samples.