Supplement of "A Functional Coefficients Network Autoregressive Model"

Hang Yin, Abolfazl Safikhani and George Michailidis

University of Florida, George Mason University and University of California, Los Angeles

0.1 Lemmas and Proofs

0.1.1 Proof of Stability Result (Section 2)

Following along the lines in Fan and Yao (2003) and Chan and Tong (1985), we first introduce the following lemmas:

Lemma 1. *Let $\{\boldsymbol{X}_t\}$ be a $\phi-$irreducible Markov chain on a normed topological space. If the transition probability $P(x,\cdot)$ is strongly continuous, namely, the transition probability $P(x,A)$ from x to a measurable set A is continuous in x, then a sufficient condition for the geometric ergodicity is that there exists a compact set $\boldsymbol{K}$ and a positive constant $\rho<1$ such that*

$$\mathbb{E}(||\boldsymbol{X}_{t+1}|||\boldsymbol{X}_t=x)<\begin{cases}\infty, & for\ x\in\boldsymbol{K}\\ \rho||x||, & for\ x\notin\boldsymbol{K}\end{cases}$$

Lemma 2. *Let $\{\boldsymbol{X}_t\}$ be an aperiodic Markov chain, and let m be a positive integer. Then, the geometric ergodicity of the subsequence $\{\boldsymbol{X}_{mt}\}$ entails the geometric ergodicity of the original series $\{\boldsymbol{X}_t\}$.*

Lemma 3. *The Markov chain $\{\boldsymbol{X}_t\}$ is aperiodic and $\phi-$irreducible with ϕ being the Lebesgue measure if ϵ_t has an absolutely continuous component with a positive density everywhere and $A(\cdot)$ and $B(\cdot)$ are bounded over bounded sets.*

Proof of Theorem 1. The proof follows along similar arguments as in Chen and Tsay (1993) and Theorem 8.1 in Fan and Yao (2003). Let $||\cdot||$ be the Euclidean norm. Let λ_{max} be the maximum eigenvalue of $\tilde{\boldsymbol{G}}$. Then $||\tilde{\boldsymbol{G}}^n||^{1/n} \to |\lambda_{max}| < 1$, so there exists a positive integer $\delta < 1$ and an integer m such that $||\tilde{\boldsymbol{G}}^m|| < \delta$.

For the subchain $\{\boldsymbol{X}_{mt}, t = 1, 2, \cdots\}$,

$$\boldsymbol{X}_{m(t+1)} = \prod_{i=0}^{m-1} \boldsymbol{G}(\mathbb{U}_{mt+i+1})\boldsymbol{X}_{mt} + \sum_{i=1}^{m}[\prod_{j=1}^{m-1} \boldsymbol{G}(\mathbb{U}_{mt+j+1})]\mathcal{E}_{mt+i}. \tag{1}$$

For any vector $c = (c_1, \cdots, c_{Nq})^T$, let $(d_1, \cdots, d_{Nq})^T = \boldsymbol{G}(\mathbb{U})c$. Then

$$|d_i| = |\sum_{j=1}^{q} a_{i,j}(U)c_{1+(j-1)N} + \sum_{j=1}^{q}\sum_{k=1,k\neq i}^{N} b_{ij}(U)w_{ik}c_{k+(j-1)N}| \leq \sum_{j=1}^{q} \tilde{a}_{i,j}|c_{1+(j-1)N}| + \sum_{j=1}^{q}\sum_{k=1,k\neq i}^{N} \tilde{b}_{ij}w_{ik}|c_{k+(j-1)N}|,$$

for $i = 1, 2, \cdots, N$ and $|d_i| = |c_i|$ for $i = N+1, \cdots, Nq$. So

$$||\boldsymbol{G}(\mathbb{U})c|| \leq ||\tilde{\boldsymbol{G}}|c|||.$$

Repeatedly applying this to 1, we obtain

$$||\boldsymbol{X}_{(m+1)t}|| = ||\tilde{\boldsymbol{G}}^m|\boldsymbol{X}_{mt}||| + ||\sum_{i=1}^{m} \tilde{\boldsymbol{G}}^{m-i}|\mathcal{E}_{mt+i}|||.$$

The first term is bounded by

$$||\tilde{\boldsymbol{G}}^m||||\boldsymbol{X}_{mt}|| \leq \delta||\boldsymbol{X}_{mt}||.$$

Hence

$$\mathbb{E}(||\boldsymbol{X}_{m(t+1)}|||\boldsymbol{X}_{mt} = \boldsymbol{x}) \leq \delta||\boldsymbol{x}|| + \mathbb{E}||\sum_{i=1}^{m} \tilde{\boldsymbol{G}}^{m-i}|\mathcal{E}_{mt+i}|||.$$

The second term is bounded and is independent of $\boldsymbol{x}$. Let D denote the bound. Then

$$\mathbb{E}(||\boldsymbol{X}_{m(t+1)}|||\boldsymbol{X}_{mt} = \boldsymbol{x}) \leq \delta||\boldsymbol{x}|| + D.$$

Let $\rho \in (\delta, 1)$ and set $M = D(\rho - \delta)^{-1}$. Then for all $||\boldsymbol{x}|| > M$,

$$\mathbb{E}(||\boldsymbol{X}_{m(t+1)}|||\boldsymbol{X}_{mt}) \leq \rho||\boldsymbol{x}||.$$

Hence, by Lemma 1, with $K = \{\boldsymbol{x} : ||\boldsymbol{x}|| \leq M\}$, the sequence $\{\boldsymbol{X}_{mt}\}$ is geometrically ergodic. As a result, the original sequence $\{\boldsymbol{X}_t\}$ is geometrically ergodic by Lemma 2. □

0.1.2 Proofs of Asymptotic Properties

For completeness of the exposition, we provide the central limit theorem for martingale differences (Theorem 5.3.4 in Fuller (2009)) next:

Lemma 4. *Let $\{Z_{tn} : 1 \leq t \leq n, n \geq 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathcal{A}, P)$, and let $\{\mathcal{A}_{tn} : 0 \leq t \leq n, n \geq 1\}$ be any triangular array of sub-sigma-fields of $\mathcal{A}$ such that for each n and $1 \leq t \leq n$, Z_{tn} is $\mathcal{A}_{tn}$-measurable and $\mathcal{A}_{t-1,n}$ is contained in $\mathcal{A}_{tn}$. For $1 \leq k \leq n, 1 \leq j \leq n$, and $n \geq 1$, let*

$$S_{kn} = \sum_{t=1}^{k} Z_{tn},$$

$$\delta_{tn}^2 = E(Z_{tn}^2 | \mathcal{A}_{t-1,n}),$$

$$V_{jn}^2 = \sum_{t=1}^{j} \delta_{tn}^2,$$

and

$$s_{nn}^2 = E(V_{nn}^2).$$

Assume

- $E(Z_{tn} | \mathcal{A}_{t-1,n}) = 0$ *a.s. for* $1 \leq t \leq n$,
- $V_{nn}^2 s_{nn}^{-2} \to_p 1$,
- $\lim_{n\to\infty} s_{nn}^{-2} \sum_{j=1}^{n} E(Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn})) = 0$ *for all* $\epsilon > 0$,

where $I(A)$ denotes the indicator function of a set A. Then, as $n \to \infty$,

$$s_{nn}^{-1} S_{nn} \to_d N(0, 1).$$

Proof of Theorem 2

Note that

$$\begin{aligned}\sqrt{T}(\hat{\beta}_i - \beta_i) &= (\frac{1}{T}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)})^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T(\mathbb{Z}_{i(t-1)}\beta_i + \epsilon_{it}) - \sqrt{T}\beta_i \\ &= (\frac{1}{T}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)})^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it}.\end{aligned}$$

Hence, it suffices to show the following:

(a) $\frac{1}{T}\sum\limits_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} \to_p P_i$, which follows from ergodicity.

Specifically, following the proof of Proposition 11.1 in Hamilton (1994), $\rho(\tilde{\boldsymbol{G}}) < 1$ ensures that the MA(∞) representation:

$$\boldsymbol{X}_t = \sum_{j=0}^{\infty}\boldsymbol{G}^j(\mathbb{U}_{t-j})\mathcal{E}_{t-j} \tag{2}$$

is absolutely summable. Hence, $\boldsymbol{X}_t$ is ergodic for the first moments from Proposition 10.2(b), 10.5(a) of Hamilton (1994), and is also ergodic for the second moments from Proposition 10.2(d) of Hamilton (1994).

Hence, $\boldsymbol{X}_t$ is a geometrically ergodic process which implies that $\frac{1}{T}\sum\limits_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t^T \to_p E(\boldsymbol{X}_t\boldsymbol{X}_t^T)$. We also get similar results for $\mathbb{U}_t$ because of strict stationarity and (2). Note that by assumption 3, the row sum of W is equal to 1, so $w_i^T\mathbb{X}_{t-j}$ can be seen as the weighted average of $\mathbb{X}_{t-j}$, and elements in $\mathbb{Z}_{i(t-1)}$ include $Z_{i1}(X_{i(t-j)}, U_{it})$, $Z_{i2}(w_i^T\mathbb{X}_{t-j}, U_{it})$, where $Z_{i1}(\cdot)$ and $Z_{i2}(\cdot)$ are continuous functions. Thus, $\frac{1}{T}\sum\limits_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} \to_p P_i$.

(b) $\frac{1}{\sqrt{T}}\sum\limits_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} \to_d \sigma^2 P_i$.

For simplicity, denote $Y_{ij,l}$ to be the l-th element of $Z_{i(t-j)}$. So

$$Z_{i(t-j)} := \begin{bmatrix} Z_{i1}(X_{i(t-j)}, U_{it}) & Z_{i2}(w_i^T\mathbb{X}_{t-j}, U_{it})\end{bmatrix} = \begin{bmatrix} Y_{ij,1} & \cdots & Y_{ij,M+K} & Y_{ij,M+K+1} & \cdots & Y_{ij,2M+2K}\end{bmatrix}.$$

By assumption 5 and Lemma 3 of Yin et al. (2023), $E(|X_{it}|^4) < \infty$, $E(|U_{it}|^{4(M-1)}) < \infty$, and by assumption 3, it is obvious $E(|w_i^T\mathbb{X}_t|^4) < \infty$, so $E(|Y_{ij,k}|^4) < \infty$.

Let $\eta_{2q(M+K)\times 1}$ be a column vector of arbitrary real numbers such that $\eta^T\eta \neq 0$. Let

$$\frac{1}{\sqrt{T}}\eta^T\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} = \sum_{t=1}^{T}Z_{tT} = S_{TT},$$

where $Z_{tT} = \frac{1}{\sqrt{T}}\eta^T\mathbb{Z}_{i(t-1)}^T\epsilon_{it}$. Since $E(\epsilon_t|\mathcal{A}_{t-1}) = 0$ where $\mathcal{A}_{t-1}$ is the sigma-field generated by $\{\epsilon_j : j \leq t-1\}$, we have $E(Z_{tT}|\mathcal{A}_{t-1}) = 0$, so condition 1 of Lemma 4 (in the supplement) is satisfied.

Define $\delta_{tT}^2 := E(Z_{tT}^2|\mathcal{A}_{t-1}) = E(\frac{1}{T}\eta^T\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)}\eta\epsilon_{it}^2|\mathcal{A}_{t-1}) = \frac{1}{T}\eta^T\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)}\eta\sigma^2$ and $V_{TT}^2 := \frac{1}{T}\sum_{t=1}^{T}\eta^T\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)}\eta\sigma^2$.

Hence, we have $V_{TT}^2 \rightarrow_p \eta^T P_i\eta$ by ergodicity of $\mathbb{X}_t$. Further,

$$s_{TT}^2 := E(V_{TT}^2) = \sum_{t=1}^{T}\frac{1}{T}\eta^T P_i\eta\sigma^2 = \eta^T P_i\eta\sigma^2,$$

thus $V_{TT}^2 s_{TT}^{-2} \rightarrow_p 1$. Therefore, condition 2 of Lemma 4 is satisfied.

Condition 3 of Lemma 4 gives

$$\begin{aligned}
&s_{TT}^{-2}\sum_{t=1}^{T}E(Z_{tT}^2 I(|Z_{tT}| \geq \epsilon s_{TT}))\\
&\leq s_{TT}^{-2}\sum_{t=1}^{T}E((\epsilon s_{TT})^{-2}Z_{tT}^4 I(|Z_{tT}| \geq \epsilon s_{TT}))\\
&\leq s_{TT}^{-4}\epsilon^{-2}\sum_{t=1}^{T}E((\frac{1}{\sqrt{T}}\eta^T\mathbb{Z}_{i(t-1)}^T\epsilon_{it})^4)\\
&= s_{TT}^{-4}\epsilon^{-2}\sum_{t=1}^{T}E((\frac{1}{\sqrt{T}}\eta^T\mathbb{Z}_{i(t-1)}^T\epsilon_{it})^4).
\end{aligned}$$

Define $a_i := \frac{1}{\sqrt{T}}\sum_{j=1}^{q}\sum_{k=1}^{2M+2K}\eta_{(iNq-Nq+j-1)(2M+2K)+k}Y_{ij,k}$, then for $E((\eta^T\frac{1}{\sqrt{T}}\mathbb{Z}_{i(t-1)}^T\epsilon_{it})^4)$, we have:

$$\begin{aligned}
&E((\eta^T\frac{1}{\sqrt{T}}\mathbb{Z}_{i(t-1)}^T\epsilon_{it})^4) = E((\frac{1}{\sqrt{T}}a_i\epsilon_{it})^4)\\
&= E(E(\frac{1}{T^2}a_i^4\epsilon_{it}^4|\mathcal{A}_{t-1})) \overset{(a)}{\leq} c_1 E((\frac{1}{T^2}a_i^4))
\end{aligned}$$

where (a) follows from $E(|\epsilon_{it}^4|\mathcal{A}_{t-1}) \leq c_1 < \infty$.

Since $E(|Y_{ij,k}|^4) \leq c_2 < \infty$, $E(\frac{1}{T^2}a_i^4) = O(\frac{1}{T^2})$. Therefore, $s_{TT}^{-2}\sum_{t=1}^{T} E(Z_{tT}^2 I(|Z_{tT}| \geq \epsilon s_{TT})) = O(\frac{1}{T}) \to 0$. Thus, we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} \to_d N(0, \sigma^2 P_i).$$

Combining parts (a) and (b) implies that $\sqrt{T}(\hat{\beta}_i - \beta_i) \to_d N(0, \sigma^2 P_i^{-1})$.

Proof of Theorem 3

Proof of Theorem 3. By the definition of the ridge estimator in Equation 3.8, we get

$$\begin{aligned}
\hat{\beta}_i^{ridge} &= argmin\frac{1}{T}\sum_{t=1}^{T}(X_{it} - Z_{i(t-1)}\beta_i)^2 + ||\lambda\Psi\beta_i||^2 \\
&= (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T X_{it} \\
&= (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T X_{it} \\
&= (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}\sum_{t=1}^{T}(\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)}\beta_i + \mathbb{Z}_{i(t-1)}^T\epsilon_{it}) \\
&= (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}(\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda TM)\beta_i + (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} \\
&- (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}T\Psi\beta_i \\
&= \beta_i + (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} - (\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + TM)^{-1}\lambda T\Psi\beta_i
\end{aligned}$$

Since,

$$\frac{1}{T}(\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi) \to_p P_i + \lambda\Psi,$$

we thus obtain:

$$\sqrt{T}(\hat{\beta}_i^{ridge} - \beta_i) = \sqrt{T}(\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\mathbb{Z}_{i(t-1)} + \lambda T\Psi)^{-1}(\sum_{t=1}^{T}\mathbb{Z}_{i(t-1)}^T\epsilon_{it} - \lambda T\Psi\beta_i).$$

If $\lambda = o(\frac{1}{\sqrt{T}})$, using the above and part (b) in the proof of Theorem 2,

$$\sqrt{T}(\hat{\beta}_i^{ridge} - \beta_i) \to_d N(0, \sigma^2 P_i^{-1}).$$

□

Proof of Theorem 4

Proof of Theorem 4. Since $\beta_i(u) = (I_{2q} \otimes \Phi_i(u))\beta_i$, and

$$\sqrt{T}(\hat{\beta}_i - \beta_i) \to_d N(0, \sigma^2 P_i^{-1}).$$

So by the delta method,

$$\sqrt{T}(\hat{\beta}_i(u) - \beta_i(u)) \to_d N(0, \sigma^2 (I_{2q} \otimes \Phi_i(u)) P_i^{-1} (I_{2q} \otimes \Phi_i(u))^T).$$

□

Proof of Theorem 5

Proof of Theorem 5. Note that we assume $\Sigma_\epsilon = \sigma^2 I$, β_i of each node is estimated separately. The proof of 5 is parallel to that of Theorems 2, 3, and 4. □

Uniform Convergence of FNAR Model Parameters

Corollary 1. *Suppose assumptions 1-5 hold. Further, assume u satisfies $P(u \notin [-L, L]) \to 0$ as $L \to \infty$ where $L \in \mathbb{R}$.*

- *If $\frac{L^{2M}}{T} = o(1)$,*

$$sup_u |\beta^i(u) - \beta^i(u)| \to_p 0 \textit{ for } i = 1, 2, \cdots, N.$$

- *If $\frac{L^{2M}N}{T} = o(1)$,*

$$sup_u ||\beta^i(u) - \beta^i(u)||_2 \to_p 0.$$

Proof. For $U_{it} = u$,

$$\sqrt{T}(\hat{\beta}_i(u) - \beta_i(u)) \to_d N(0, \sigma^2 (I_{2q} \otimes \Phi_i(u)) P_i^{-1} (I_{2q} \otimes \Phi_i(u))^T).$$

Denote $\beta_{ij}(u)$ to be the j-th element of $\beta_i(u)$, and denote the corresponding asymptotic variance as $\sigma_{ij}^2(u)$, which is the j-th diagonal element of

$$\sigma^2(I_{2q} \otimes \Phi_i(u))P_i^{-1}(I_{2q} \otimes \Phi_i(u))^T.$$

Thus, $\sigma_{ij}^2(u) := O(u^{2M})$.

$$\begin{aligned} & P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon) \\ = & P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon, u \in [-L, L]) + P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon, u \notin [-L, L]). \end{aligned} \tag{3}$$

For $u \in [-L, L]$,

$$\begin{aligned} & P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon, u \in [-L, L]) \\ &= P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon \mid u \in [-L, L]) \cdot P(u \in [-L, L]) \\ &= P(|\hat{\beta}_{ij}(u_0) - \beta_{ij}(u_0)| \geq \epsilon \mid u_0 \in [-L, L]) \cdot P(u \in [-L, L]) \\ &= O(\frac{L^{2M}}{T\epsilon^2}) \cdot P(u \in [-L, L]) \\ &\to 0 \text{ as } T \to \infty, \end{aligned} \tag{4}$$

if $\frac{L^{2M}}{T} = o(1)$, since $\beta_{ij}(u)$ is a continuous function of u.

For $u \notin L$,

$$\begin{aligned} & P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon, u \notin [-L, L]) \\ &= P(\sup_u |\hat{\beta}_{ij}(u) - \beta_{ij}(u)| \geq \epsilon \mid u \notin [-L, L]) \cdot P(u \notin [-L, L]) \\ &\to 0, \end{aligned} \tag{5}$$

if $P(u \notin [-L, L]) \to 0$ as $L \to \infty$.

Thus, the condition $P(u \notin [-L, L]) \to 0$ as $L \to \infty$ and $\frac{L^{2M}}{T} = o(1)$ needs to be satisfied.

Similarly,

$$\begin{aligned}
&P(\sup_u ||\hat{\beta}_i(u) - \beta_i(u)||_2 \geq \epsilon) \\
&= P(\sup_u \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{2q}(\hat{\beta}_{ij}(u) - \beta_{ij}(u))^2} \geq \epsilon) \\
&= P(\sup_u \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{2q}(\hat{\beta}_{ij}(u) - \beta_{ij}(u))^2} \geq \epsilon, u \in [-L, L]) \qquad (6) \\
&+ P(\sup_u \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{2q}(\hat{\beta}_{ij}(u) - \beta_{ij}(u))^2} \geq \epsilon, u \notin [-L, L]) \\
&\to 0
\end{aligned}$$

if $\frac{L^{2M}N}{T} = o(1)$, and $P(u \notin [-L, L]) \to 0$ as $L \to \infty$. □

0.2 Additional Simulation Results

The following simulation scenarios are considered:

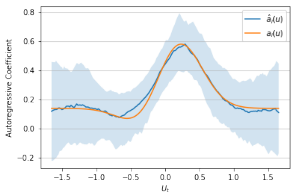
A.2: The sample size is set to $T = 400$ and 2400, and the spline order varies between $M = 2, 4, 6$. Further, the error and threshold processes and W are set as in scenario A.1.

A.3: The order of the spline basis is set to $M = 4$, while the number of knots varies according to $K = 5, 10, 15$. The sample size, the error and threshold processes and the network matrix W are as in scenario A.2.

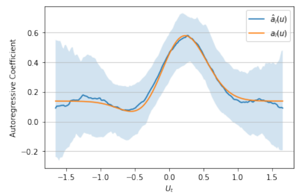
A.4: The order of the spline basis is set to $M = 4$ with $K = 10$ knots. The sample size and the error and threshold processes are set as in scenario A.2. The network matrix W is banded and the bandwidth is set to $2, 10, 50$.

Figures 1 and 2 in the supplementary material suggest that for exponential autoregressive and network effect functions, splines of low order ($M = 2$) provide good estimates compared to higher order splines for $T = 400$. This may be related to the limited sample size. As T increases to 2400, the performance of splines with $M = 4$ and 6 improve.

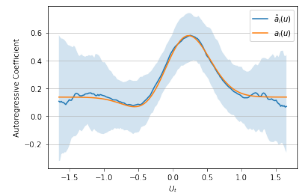
Figures 3 and 4 in the supplementary material show that the number of knots K needs to be selected in accordance with the sample size T. Specifically, $K = 10$ gives the best result, while for smaller K, there might be bias for the estimated functions, and for larger K the estimated functions become unstable near the boundary. Finally, Figures 5 and 6 in the supplementary material confirm the robustness of the autoregressive function estimates over network matrices W with different bandwidths (number of neighboring nodes included), while the network effect estimates become less accurate for more connected W matrices.



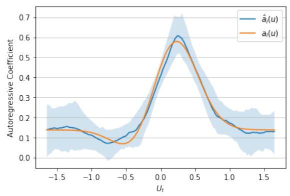
(a) M=2, T=400



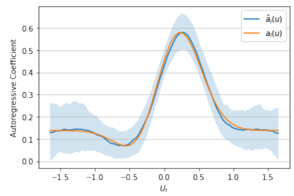
(b) M=4, T=400



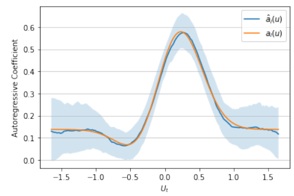
(c) M=6, T=400



(d) M=2, T=2400

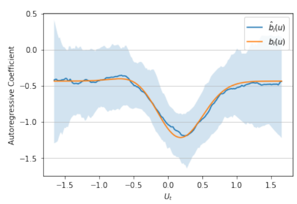


(e) M=4, T=2400

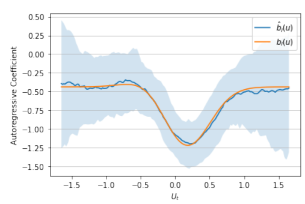


(f) M=6, T=2400

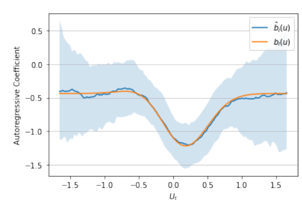
Figure 1: $a_i(\cdot)$ of A.2



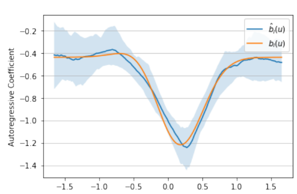
(a) M=2, T=400



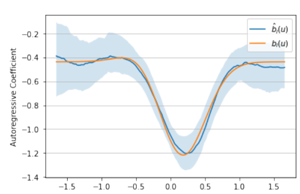
(b) M=4, T=400



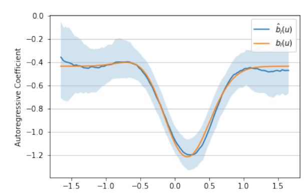
(c) M=6, T=400



(d) M=2, T=2400

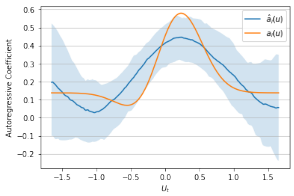


(e) M=4, T=2400

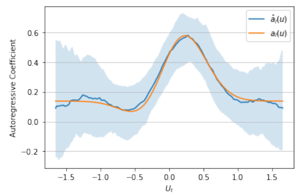


(f) M=6, T=2400

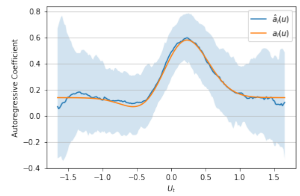
Figure 2: $b_i(\cdot)$ of A.2



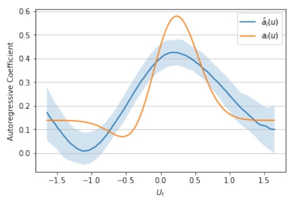
(a) K=5, T=400



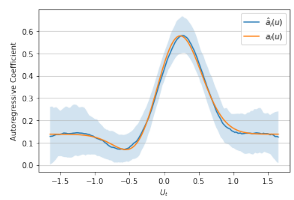
(b) K=10, T=400



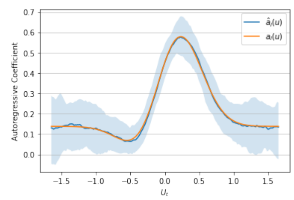
(c) K=15, T=400



(d) K=5, T=2400

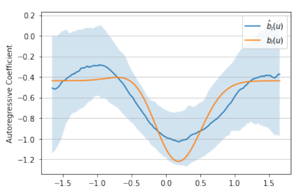


(e) K=10, T=2400

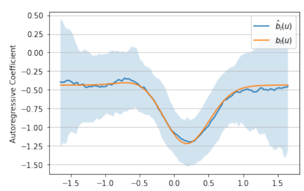


(f) K=15, T=2400

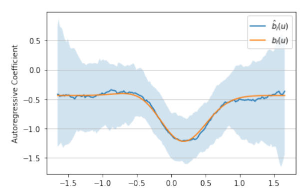
Figure 3: $a_i(\cdot)$ of A.3



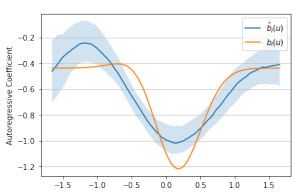
(a) K=5, T=400



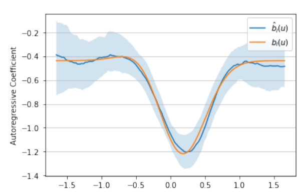
(b) K=10, T=400



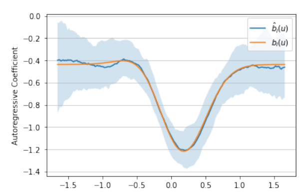
(c) K=15, T=400



(d) K=5, T=2400

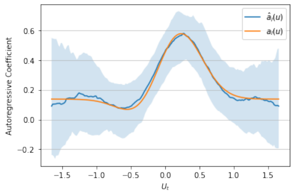


(e) K=10, T=2400

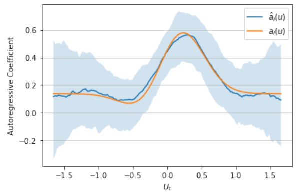


(f) K=15, T=2400

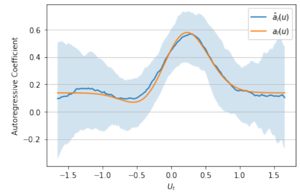
Figure 4: $b_i(\cdot)$ of A.3



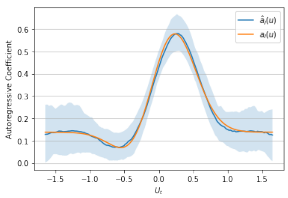
(a) BW=2, T=400



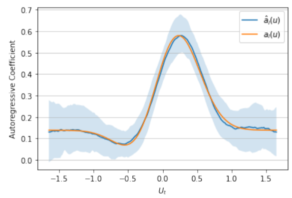
(b) BW=10, T=400



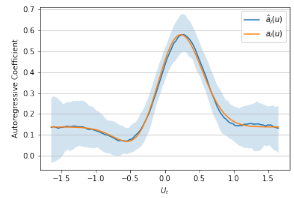
(c) BW=50, T=400



(d) BW=2, T=2400

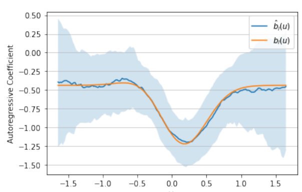


(e) BW=10, T=2400

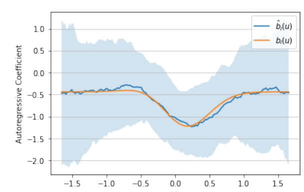


(f) BW=50, T=2400

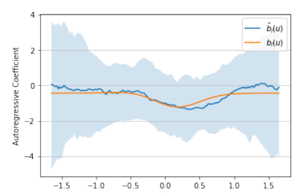
Figure 5: $a_i(\cdot)$ of A.4



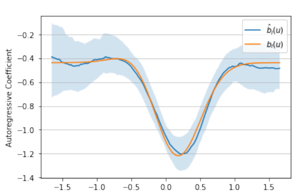
(a) BW=2, T=400



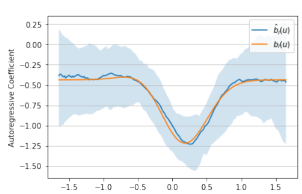
(b) BW=10, T=400



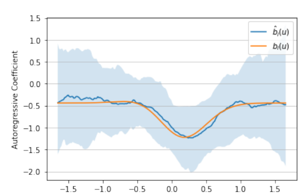
(c) BW=50, T=400



(d) BW=2, T=2400



(e) BW=10, T=2400



(f) BW=50, T=2400

Figure 6: $b_i(\cdot)$ of A.4

0.2.1 Impact of the network matrix W on the Spectral Radius and Assumption 3(a) of the FCNAR model

Theorem 1 shows that a sufficient condition for the stability of the FNCAR process is that $\rho(\tilde{G}) < 1$. Further, Assumption 3(a) is used to establish the asymptotic properties of the estimators of the FCNAR model parameters. The network matrix W impacts both the spectral radius and the maximum and minimum eigenvalues of $E(Y_{it}Y_{it}^T|U_{it} = u$; however, analytical results on its impact are difficult to obtain. In the sequel, we provide numerical evidence of how these important quantities behave as a function of the bandwidth of W.

(I) Impact of W on the Spectral Radius

Consider FCNAR(1,1) with $N = 100$, $a_i(u) := 0.3I(u \leq 0) - 0.4I(u > 0)$ and $b_i(u) := -0.5I(u \leq 0) + 0.4I(u > 0)$, $i = 1, 2, \cdots, N$. Further, both the autorgressive $a_i(\cdot)$ and network effect $b_i(\cdot)$ functions are bounded by universal constants $\tilde{a}_i = 0.4$ and $\tilde{b}_i = 0.5$ for all $i = 1, 2, \cdots, N$. We consider equally weighted network matrices W of bandwidths 1 to 100.

Figure 7 depicts the spectral radius of $\tilde{\boldsymbol{G}}$ as the bandwidth of W increases.

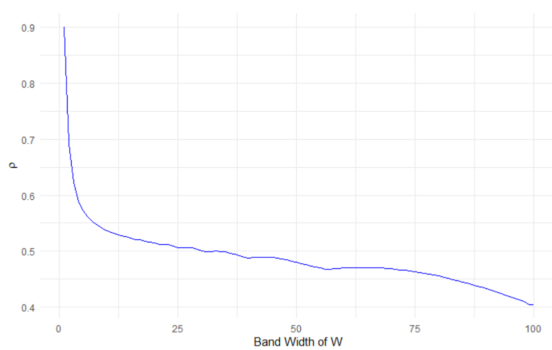


Figure 7: Spectral Radius $\rho(\tilde{\boldsymbol{G}})$ as a function of an increasing bandwidth of W

It can be seen that the spectral radius $\rho(\cdot)$ is a decreasing function of the bandwidth of W. Specifically, as the bandwidth (number of neighbors) increases from 1 to ~ 5, the spectral radius decreases rapidly to a value around 0.55. A further increase in the bandwidth of W (from 10 to the maximum value of 100, wherein every node in the network impacts every other node), gradually decreases the spectral radius to a value of 0.4.

Hence, after fixing all the other parameters that determine the FCNAR process, the impact of an increasing bandwidth of the network matrix W on the spectral radius,beyond a certain number of neighbors, is gradual and small.

We also plotted $\rho(\boldsymbol{G}(u))$ as a function of u. Consider FCNAR(1,1) with $N = 100$, $a_i(u) := 0.1 - 0.03u + 0.05u^2 - 0.16(u+0.5)_+^2 + 0.12(u-0.5)_+^2$ for $u \in [-2, 2]$ and constant otherwise, and $b_i(u) := -0.12 + 0.08u - 0.05u^2 + 0.18(u + 0.5)_+^2 - 0.12(u - 0.5)_+^2$ for $u \in [-2, 2]$ and constant otherwise, $i = 1, 2, \cdots, N$. Fiture 8 depicts the spectral radius as a function of u.

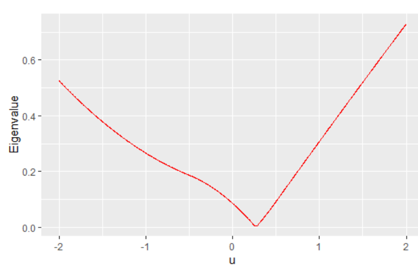


Figure 8: Spectral Radius $\rho(\boldsymbol{G}(u))$ as a function of u

For FCNAR(1,1) with $N = 100$, $a_i(u) := 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$ and $b_i(u) := -0.437 - (0.659 + 1.260u)e^{-3.89u^2}$, $i = 1, 2, \cdots, N$. Fiture 9 depicts the spectral radius as a function of u.

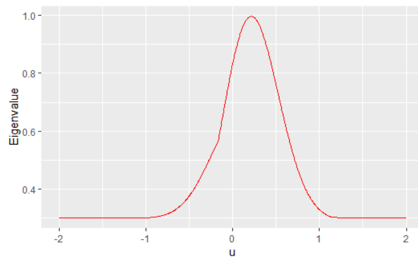


Figure 9: Spectral Radius $\rho(\boldsymbol{G}(u))$ as a function of u

(II) Impact of W on Assumption 3(a)

Consider again the same setting for FCNAR(1,1) with $N = 100$, $a_i(u) := 0.3I(u \leq 0) - 0.4I(u > 0)$ and $b_i(u) := -0.5I(u \leq 0) + 0.4I(u > 0)$, $i = 1, 2, \cdots, N$. The network matrix W is constructed in an analogous fashion as in (I) above.

Figure 10 depicts the maximum and minimum eigenvalue of $E(Y_{it}Y_{it}^T|U_{it} = -0.1)$, as the bandwidth of W increases.

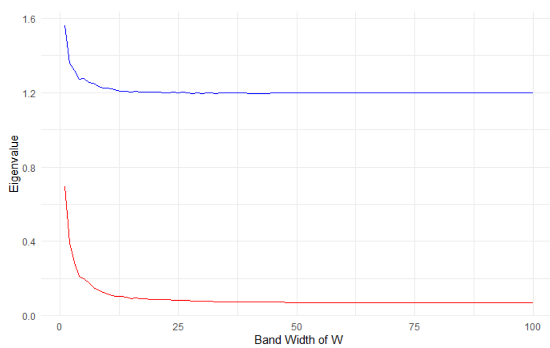


Figure 10: Maximum and minimum eigenvalue of $E(Y_{it}Y_{it}^T|U_{it} = -0.1)$ as a function of the bandwidth of W

It can be seen that for a bandwidth larger than 15, the maximum eigenvalue stabilizes around 1.2. The same happens to the minimum eigenvalue, which more importantly remains clearly bounded away from 0.

Hence, after fixing all the other parameters that determine the FCNAR process, the impact of an increasing bandwidth of the network matrix W on Assumption 3(a), beyond a certain number of neighbors, is marginal.

0.3 Results from a Data Set on Wind Speeds

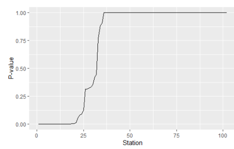
We also tested the posited FCNAR model based on a wind speed data set available in the `GNAR` `R` package. The data set contains wind speed measurements for 102 weather stations in England and Wales, whose location coordinates are available. The length of the wind speed series is 721 observations. W correspond to a row-normalized adjacent matrix obtained as follows: let D_{ij} be the Euclidean distance between two stations, then the ij-th element of W is defined as: $\phi_{ij} = w_{ij} := \frac{D_{ij}^{-1}}{\sum\limits_{j} D_{ij}^{-1}}$ if $i \neq j$ and $D_{ij} \leq 100$ and 0 otherwise.

The F-test (for details see Section 3.3) for the following two null hypotheses

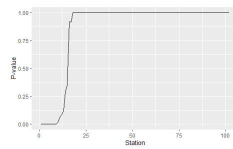
$$H_0 : a_{i2} = a_{i3} = \cdots = a_{i(M+K)} = 0,$$

$$H_0 : b_{i2} = b_{i3} = \cdots = b_{i(M+K)} = 0.$$

is used to test, if each node exhibits a non-linear autoregressive or network effect. However, since only a small number (around 21 for $a_i(\cdot)$ and 10 for $b_i(\cdot)$) of the stations have non-linear autoregressive and network coefficients according to the hypothesis testing, the model can be simplified to the linear models. We plot the adjusted p-values in ascending order in Figure 11:



(a) $a_i(\cdot)$



(b) $b_i(\cdot)$

Figure 11: Adjusted p-values sorted in ascending order

Bibliography

Chan, K. S. and H. Tong (1985). On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Advances in applied probability 17*(3), 666–678.

Chen, R. and R. S. Tsay (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association 88*(421), 298–308.

Fan, J. and Q. Yao (2003). *Nonlinear time series: nonparametric and parametric methods*, Volume 20. Springer.

Fuller, W. A. (2009). *Introduction to statistical time series*, Volume 428. John Wiley & Sons.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Yin, H., A. Safikhani, and G. Michailidis (2023). A general modeling framework for network autoregressive processes. *Technometrics 65*(4), 579–589.