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

Abstract: The paper introduces a flexible model for the analysis of multivariate nonlinear time series data. The proposed Functional Coefficients Network Autoregressive (FCNAR) model considers the response of each node in the network to depend in a *nonlinear* fashion to each own past values (autoregressive component), as well as past values of each neighbor (network component). Key issues of model stability/stationarity, together with model parameter identifiability, estimation and inference are addressed for error processes that can be heavier than Gaussian for both fixed and growing number of network nodes. The performance of the estimators for the FCNAR model is assessed on synthetic data and the applicability of the model is illustrated on two data sets; the first on multiple indicators of air pollution data and the second on Covid-19 cases in Florida counties.

Key words and phrases: functional-coefficient regression model; network autoregressive model; ridge penalty; polynomial spline

1. Introduction

Nonlinear time series models gained prominence because of their ability to model, analyze and predict complex patterns in data in a wide range of fields, including economics and finance Franses et al. (2000), climate Donges et al. (2015), cognitive science Ward (2002), geosciences Donner and Barbosa (2008) and engineering Zou et al. (2019). Departure from linearity opens different possibilities of developing nonlinear models. However, a fully nonparametric model that does not impose any constraints on the autoregressive form becomes harder to estimate with limited data or in a multivariate setting (Fan and Yao, 2003). Hence, the literature focused on specific classes of *parametric* nonlinear models for the *conditional mean* for *univariate* time series data, such as the exponential autoregressive (EXPAR) (Haggan and Ozaki, 1981), the threshold autoregressive (TAR) (Tong, 1990) and the smooth transition autoregressive models (Dijk et al., 2002) that added modeling flexibility. The Functional Coeffecient Autoregressive Model (FAR) introduced by Chen and Tsay (1993) encompassed these other classes of models and ergodicity, estimation and inference issues were addressed. Follow-up work by Huang and Shen (2004); Chen and Liu (2001) and Cao et al. (2010) proposed alternative estimation procedures for FAR models, as well as tests of whether the functional form of the FAR

model is constant and hence it reduces to a linear one.

Note that a different line of work focused on nonlinear models for the *conditional variance*, including nonlinear ARCH (Higgins and Bera, 1992), nonlinear thresholded ARCH (Gourieroux and Monfort, 1992) and nonlinear GARCH (Hamadeh and Zakoïan, 2011) models.

In many applications, one is interested in modeling a large number of time series, whose interrelations are reflected through a *network structure*. Motivated by social network related applications, Zhu et al. (2017) developed a network vector autoregression (NAR) that captures the temporal dependence for the time series of each network node through its own history, as well as through that of neighboring nodes. A general form that encompasses various extensions of the basic NAR model for numerical variables was studied in Yin et al. (2023), while Armillotta and Fokianos (2021) considered a basic NAR model for Poisson count data.

Next, we present the general form of the NAR model presented in Yin et al. (2023). Consider a network comprising of N nodes, for which measurements have been collected for T time periods for a variable of interest X; i.e. $X_{it}, i = 1, \cdots, N, \ t = 1, \cdots, T$. Assuming lag-1 temporal dependence for ease of presentation, the model is given by

$$X_{i,t} = a_i X_{i,(t-1)} + b_i w_i^T \mathbb{X}_{t-1} + \gamma_i^T Y_{i,(t-1)} + \epsilon_{i,t}, \qquad (1.1)$$

where $a_i, b_i \in \mathbb{R}$, $\gamma_i \in \mathbb{R}^p$ are regression coefficients, $Y_{i,(t-1)}$ a p-dimensional time-varying covariate vector, and w_i a N-dimensional weight vector comprising of non-negative elements and summing to 1. Hence, in the NAR model, measurements X_{it} for node i are influenced by their own past values, plus past values of combinations of "related" nodes (network lags), plus *time-varying* covariates. Constrained versions of the NAR model (e.g., assuming that all autoregressive/network effects are common across all nodes, i.e. $a_i = a$ and/or $b_i = b$ for all or selected groups of nodes $i = 1, \cdots, N$ are special cases of the model in (1.1); for details see Yin et al. (2023).

However, analogously to the univariate FAR model, both autoregressive and network effect coefficients a_i/b_i can be *time-varying*. Consider the daily PM10 air quality indicator data, discussed in detail in Section 5. As can be seen in Figure 1, the autoregressive and network coefficient functions for selected monitoring stations exhibit strong nonlinear patterns.
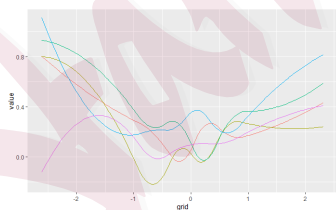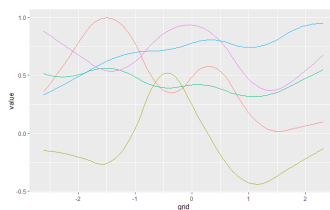
(1.1) $a_i(\cdot)$ (1.2) $b_i(\cdot)$

Figure 1: Plot of $a_i(\cdot)$ and $b_i(\cdot)$ with nonlinear patterns

Airborne particulate matter pollutants smaller in diameter than 10 microns

To accommodate such time-varying behavior, we introduce a Functional Coefficients Network Autoregressive (FCNAR) model described next. The FCNAR(q_1,q_2) model with q_1 autoregressive lags and q_2 network lags takes the form

$$X_{it} = \sum_{j=1}^{q_1} a_{i,j}(U_{it})X_{i,(t-j)} + \sum_{j=1}^{q_2} b_{i,j}(U_{it})w_i^T \mathbb{X}_{t-j} + \epsilon_{it}, \qquad (1.2)$$

where U_{it} denotes a *threshold* variable for node i, and can be either *exogenously* determined or can correspond to lagged values of the outcome variable X_{it}. Hence, for node i and lag j, $a_{i,j}(\cdot)$ and $b_{i,j}(\cdot)$ are unknown functions of the threshold variables U_{it}.

The presence of the threshold variables U_{it} introduces a number of technical challenges that relate to (i) the stability of the FCNAR process, (ii) the identifiability of the autoregressive and network effect parameters and (iii) the asymptotics for the $a_i(\cdot)/b_i(\cdot)$ functions and the FCNAR model parameters that require careful handling and resolved in the sequel.

The remainder of the paper is organized as follows. Section 2 discusses conditions to ensure the stability of the FCNAR process. Section 3 addresses model estimation, inference and testing issues, while Section 4 presents results from synthetic data showcasing the performance of the developed estimators in different simulation settings. Section 5 uses the FCNAR model to analyze the followoing two data sets: (i) air pollution

indicators and (ii) new Covid-19 cases at the county level in the state of Florida. Finally, some concluding remarks are summarized in Section 6.

Notation. Throughout the paper, $\mathbb{R}$ denotes the set of real numbers, respectively. We use $||A||$ and $||A||_F$ to denote the operator norm and Frobenius norm of a matrix A, $\oplus$ the direct sum of two matrices, and $\otimes$ the Kronecker product of two matrices of appropriate dimensions, $\mathrm{diag}(A)$ a diagonal matrix A and a^T/A^T the transpose of a vector/matrix, respectively. For matrices, we use $\to_p$ to denote element-wise convergence in probability, and $\to_d$ to denote convergence in distribution. For a symmetric or Hermitian matrix A, we denote its spectral radius by $\rho(A)$.

2. Stability of the FCNAR process

The first issue addressed is to derive the stability condition of the $\mathrm{FCNAR}(q_1, q_2)$ process for the model defined in (1.2). To proceed, we rewrite equation (1.2) in matrix form:

$$\mathbb{X}_t = \sum_{j=1}^{q_1} A_j(\mathbb{U}_t)\mathbb{X}_{t-j} + \sum_{j=1}^{q_2} B_j(\mathbb{U}_t)W\mathbb{X}_{t-j} + \epsilon_t, \qquad (2.3)$$

where $\mathbb{U}_t$ is the threshold variable, $A_j(\mathbb{U}_t) := \mathrm{diag}\{a_{1,j}(U_{1,(t-1)}) \cdots a_{N,j}(U_{N,(t-1)})\}$, $B_j(\mathbb{U}_t) := \mathrm{diag}\{b_{1,j}(U_{1,(t-1)}) \cdots b_{N,j}(U_{N,(t-1)})\}$ and W denotes the network weight matrix.

Define $G_j(\mathbb{U}_t) := A_j(\mathbb{U}_t) + B_j(\mathbb{U}_t)W$. We can rewrite (2.3) as $\mathbb{X}_t = \sum_{j=1}^{q} G_j(\mathbb{U}_t)\mathbb{X}_{t-j} + \epsilon_t$, wherein $j = 1, 2, \cdots, \max\{q_1, q_2\}$, with the convention that zero matrices are included/padded for the relationship to hold; namely, if $q_1 > q_2$, $B_j(\mathbb{U}_t) = 0$ for $j > q_2$, whereas if $q_1 < q_2$, $A_j(\mathbb{U}_t) = 0$ for $j > q_1$. Let $q = \max\{q_1, q_2\}$. Note that the model (2.3) can also be expressed as a Vector Autoregressive model with a single lag (VAR(1)) (Lütkepohl, 2005) as:

$$\boldsymbol{X}_t = \boldsymbol{G}(\mathbb{U}_t)\boldsymbol{X}_{t-1} + \mathcal{E}_t, \quad \text{with} \tag{2.4}$$

$$\boldsymbol{X}_t := \begin{bmatrix} \mathbb{X}_t^T \\ \mathbb{X}_{t-1}^T \\ \vdots \\ \mathbb{X}_{t-q+1}^T \end{bmatrix}, \ \mathcal{E}_t := \begin{bmatrix} \epsilon_t^T \\ 0^T \\ \vdots \\ 0^T \end{bmatrix}, \boldsymbol{G}(\mathbb{U}_t) := \begin{bmatrix} G_1(\mathbb{U}_t) & \cdots & G_{q-1}(\mathbb{U}_t) & G_q(\mathbb{U}_t) \\ I_N & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_N & 0 \end{bmatrix}. \tag{2.5}$$

Next, we introduce the needed assumptions to establish the stability result.

Assumption 1. *The autorgressive $a_{i,j}(\cdot)$ and network effect $b_{i,j}(\cdot)$ functions are bounded by universal constants $\tilde{a}_{i,j}$ and $\tilde{b}_{i,j}$ for all $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, q$.*

Assumption 2. *The error process $\{\epsilon_t, t \in \mathbb{N}\}$ is a sequence of independent and identically distributed (iid) random vectors with $\mathbb{E}(\epsilon_t) = 0$. Further, the marginal distribution of ϵ_t is absolutely continuous with respect to the Lebesgue measure and its density function is positive everywhere on $\mathbb{R}^N$.*

Next, define $\tilde{A}_j := \text{diag}\{\tilde{a}_{1,j}, \tilde{a}_{2,j}, \cdots, \tilde{a}_{N,j}\}$, $\tilde{B}_j := \text{diag}\{\tilde{b}_{1,j}, \tilde{b}_{2,j}, \cdots, \tilde{b}_{N,j}\}$,

and hence $\tilde{G}_j := \tilde{A}_j + \tilde{B}_j W$, Then, due to Assumption 1, $\boldsymbol{G}(\cdot)$ is elementwise bounded by $\tilde{\boldsymbol{G}}$ where

$$\tilde{\boldsymbol{G}} := \begin{bmatrix} \tilde{G}_1 & \cdots & \tilde{G}_{q-1} & \tilde{G}_q \\ I_N & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_N & 0 \end{bmatrix}.$$

We can then establish the following result.

Theorem 1. *Suppose Assumptions 1-2 hold. Then, the FCNAR(q_1, q_2) process defined in* (1.2) *is geometrically ergodic (and hence stable), if $\rho(\tilde{\boldsymbol{G}}) < 1$.*

Remark 1. The density (number of neighbors) of the weight matrix W impacts the structure of $\tilde{G}$ and thus of the spectral radius $\rho(\tilde{G})$. Numerical evidence (presented in the Supplement, Section 0.2.1) indicates that the value of the spectral radius gradually decreases, once the bandwidth (number of neighbors) of the weight matrix W exceeds a certain threshold.

3. Estimation of the FCNAR model parameters and their Asymptotic Properties

3.1 Estimation Procedure

As mentioned in the introductory section, the functions of univariate FAR models can be estimated by various approaches, including using local polynomial and locally constant methods, local linear regression techniques, as well as polynomial and penalized splines methods.

In this work, for estimating the unknown functions $a_{i,j}(\cdot)$ and $b_{i,j}(\cdot)$ of the FCNAR model, we use splines. Further, for ease of presentation, we assume that $q_1 = q_2 = 1$. The extension to general q_1, q_2 is straightforward.

For the order-M spline constructed by the truncated-power basis set, define the spline basis for $a_{i,j}(\cdot)$ and $b_{i,j}(\cdot)$ to be $\Phi_i(u)$, where:

$$\Phi_i(u) = \begin{bmatrix} \phi_{i1}(u) \ \phi_{i12}(u) \ \cdots \ \phi_{i(M+K)}(u) \end{bmatrix} \tag{3.6}$$

where $\phi_{i1}(u) := 1$, $\phi_{i2}(u) := u$, $\phi_{i3}(u) := u^2, \cdots$, $\phi_{iM}(u) := u^{M-1}$,
$\phi_{i(M+1)}(u) := (u - k_1^{(i)})_+^{M-1}$, $\cdots$, $\phi_{i(M+K)}(u) := (u - k_K^{(i)})_+^{M-1}$.

We can then express the autoregressive and network effect coefficient functions as follows:

$$a_{i,j}(u) = a_{ij1} + \cdots a_{ijM} u^{M-1} + a_{ij(M+1)}(u - k_1^{(i)})_+^{M-1} + \cdots a_{ij(M+K)}(u - k_K^{(i)})_+^{M-1},$$

$$b_{i,j}(u) = b_{ij1} + \cdots b_{ijM} u^{M-1} + b_{ij(M+1)}(u - k_1^{(i)})_+^{M-1} + \cdots b_{ij(M+K)}(u - k_K^{(i)})_+^{M-1},$$

where M denotes the order of the spline and $\{k_j^{(i)}\}_{j=1}^K$ the sequence of spline knots for node i. Hence, estimating the autoregressive $a_{i,j}(\cdot)$ and networr effect $b_{i,j}(\cdot)$ coefficient functions is equivalent to estimating the parameters $a_{ij1}, a_{ij2} \cdots, a_{ij(M+K)}$ and $b_{ij1}, b_{ij2} \cdots, b_{ij(M+K)}$ for the corresponding spline bases.

In the sequel, we consider both ordinary least squares and ridge esti-

mators for the spline base parameters. Defining

$$Z_{i1}(X_{i(t-j)}, U_{it}) := \left[\phi_{i1}(U_{it})X_{i(t-j)} \ \phi_{i2}(U_{it})X_{i(t-j)} \ \cdots \ \phi_{i(M+K)}(U_{it})X_{i(t-j)} \right],$$

$$Z_{i2}(w_i^T \mathbb{X}_{t-j}, U_{it}) := \left[\phi_{i1}(U_{it})w_i^T \mathbb{X}_{t-j} \ \phi_{i2}(U_{it})w_i^T \mathbb{X}_{t-j} \ \cdots \ \phi_{i(M+K)}(U_{it})w_i^T \mathbb{X}_{t-j} \right],$$

$$Z_{i(t-j)} = \left[Z_{i1}(X_{i(t-j)}, U_{it}) \ \ Z_{i2}(w_i^T \mathbb{X}_{t-j}, U_{it}) \right],$$

$$\mathbb{Z}_{i(t-1)} := \left[Z_{i(t-1)} \ Z_{i(t-2)} \ \cdots \ Z_{i(t-q)} \right],$$

the model can then be written in compact form as:

$$\mathbb{X}_t = \mathbb{Z}_{t-1}\beta + \epsilon_t, \tag{3.7}$$

where

$$\mathbb{Z}_{t-1} := \begin{bmatrix} \mathbb{Z}_{1(t-1)} & 0 & \cdots & 0 \\ 0 & \mathbb{Z}_{2(t-1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbb{Z}_{N(t-1)} \end{bmatrix},$$

$$\beta := \left[\beta_1^T \ \beta_2^T \ \cdots \ \beta_N^T \right]^T, \ \ \beta_i := \left[\beta_{i1}^T \ \beta_{i2}^T \ \cdots \ \beta_{iq}^T \right]^T, \ \ and$$

$$\beta_{ij} := \left[a_{ij1} \ a_{ij2} \ \cdots \ a_{ij(M+K)} \ b_{ij1} \ b_{ij2} \ \cdots \ b_{ij(M+K)} \right]^T.$$

Then, the least squares (LS) estimator is given by:

$$\hat{\beta} = (\sum_{t=1}^T \mathbb{Z}_{t-1}^T \mathbb{Z}_{t-1})^{-1} \sum_{t=1}^T \mathbb{Z}_{t-1}^T \mathbb{X}_t.$$

Note that the number of knots K, the placement of knots and the spline order M are critical tuning parameters that impact the quality of the LS estimator in empirical work and are investigated in detail in Section 4.

In the presence of many lags and a large network size, the columns of the design matrix Z may be fairly strongly correlated. Hence, a ridge estimator may be preferable in empirical work. Further, Cao et al. (2010) mention that the number and the location of knots are no longer as critical to be carefully selected (unlike the LS estimator) and can be controlled by the tuning parameter λ of the ridge estimator, the latter defined as:

$$\hat{\beta}^{ridge} = (\sum_{t=1}^{T} \mathbb{Z}_{t-1}^T \mathbb{Z}_{t-1} + \lambda T \Psi)^{-1} \sum_{t=1}^{T} \mathbb{Z}_{t-1}^T \mathbb{X}_t, \tag{3.8}$$

where Ψ is a diagonal matrix with its $(1 + iM+iK)$-th diagonal elements equal to 0 and the rest equal to 1 for $i = 0, 1, \cdots, 2Nq - 1$, which penalizes the higher order coefficients.

3.2 Asymptotic Properties of the Estimators

Next, we introduce the needed assumptions for the main results.

Assumption 3. *Assumptions on the network matrix W:*

(a) The eigenvalues of $E(Y_{it}Y_{it}^T|U_{it} = u)$ are uniformly bounded away from both 0 and infinity for all $u \in \mathbb{R}$ and $i = 1, 2, \cdots, N$, where

$$Y_{it} := \begin{bmatrix} X_{it} & w_i^T \mathbb{X}_t \end{bmatrix}^T.$$

(b) $W \in \mathbb{R}^{N \times N}$ is a row-normalized matrix; i.e., $\sum_{j=1}^{N} w_{ij} = 1$ with $w_{ij} \geq 0$.

Assumption 4. *Assumptions on* $\mathbb{U}_t$:

(a) *The marginal density of* U_{it} *is bounded away from both zero and in-finity uniformly on* $\mathbb{R}$.

(b) *One of the following cases needs to hold:*

- *The threshold process* $\mathbb{U}_t$ *corresponds the lagged values of the outcome process* $\mathbb{X}_t$, *or*

- $\mathbb{U}_t$ *is an exogenous process (independent of* ϵ_t) *and strictly stationary. Further, for some sufficiently large* $m > 0$, $\mathbb{E}(|U_{it}|^m) < \infty$.

Assumption 5. *The error process* $\{\epsilon_t, t \in \mathbb{N}\}$ *satisfies* $\mathbb{E}(\epsilon_t) = 0$, $\Sigma_\epsilon = \sigma^2 I$, *and* $\mathbb{E}(|\epsilon_{it}|^4) < \infty$.

Assumptions 3(a) and 4(a) are needed to ensure the *identifiability* of the autoregressive and network effect coefficient functions (see discussion in Huang and Shen (2004)). Further, assumptions 3(b), 4(b) and 5 are required for establishing the asymptotic properties of the LS and ridge estimators of the model parameters. Specifically, the moment conditions on X_{it} and U_{it} ensure the martingale central limit theorem. Note that 4(b) is also used in univariate FAR models. Assumptions 1, 2 together with 4(b) imply

that $\{\mathbb{U}_t, \mathbb{X}_t\}$ is jointly strictly stationary. Further, note that Assumption 2 ensures identifiability of the autoregressive (q_1) and network effect (q_2) lag-orders. It is also worth noting that this is a standard assumption used also in univariate autoregressive time series models (Hamilton, 2020) to ensure identifiability of the lag-order. Finally, note that since we assume that $\Sigma_\epsilon = \sigma^2 I$, the FCNAR parameter β can be estimated element-wise.

Remark 2. Note that the density (number of neighbors) of the network matrix W impact the maximum and minimum eigenvalue of $E(Y_{it} Y_{it}^T | U_{it} = u)$. Numerical evidence (presented in the Supplement, Section 0.2.1) indicates that the both the maximum and minimum eigenvalues do not change, once the bandwidth (number of neighbors) of the network matrix W is above a certain number. Further, the minimum eigenvalue clearly remains bounded away from zero.

Hence, for both fixed and diverging (with T) network size N, we establish the following three results:

Theorem 2 (Asymptotic Properties of the LS Estimator). *Suppose Assumptions 1-5 hold. Then, the node parameters of the FCNAR model (3.7), i.e., $\mathbb{X}_t = \mathbb{Z}_{t-1}\beta + \epsilon_t$, satisfy as $T \to \infty$*

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

where $P_i := \mathbb{E}(\mathbb{Z}^T_{i(t-1)}\mathbb{Z}_{i(t-1)})$.

Remark 3. In practice, the quantity $\mathbb{E}(\mathbb{Z}^T_{i(t-1)}\mathbb{Z}_{i(t-1)})$ can be estimated by

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{Z}^T_{i(t-1)}\mathbb{Z}_{i(t-1)}.$$

Remark 4. Note that a *uniform convergence result* over the domain of the autoregressive $a_i(\cdot)$ and network effect $b_i(\cdot)$ functions is given in Section 0.2.1 of the Supplement.

Theorem 3 (Asymptotic Properties of the Ridge Estimator). *Suppose Assumptions 1-5 hold and further assume that $\lambda = o(\frac{1}{\sqrt{T}})$. Then, the regression coefficient of the FCNAR model satisfies, as $T \to \infty$*

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

where $P_i := \mathbb{E}(\mathbb{Z}^T_{i(t-1)}\mathbb{Z}_{i(t-1)})$.

The next result provides the asymptotic distribution of any point in the domain of the autoregressive and network effect coefficient functions. Define

$$\beta_i(u) := \begin{bmatrix} a_{i,1}(u) & b_{i,1}(u) & \cdots & a_{i,q}(u) & b_{i,q}(u) \end{bmatrix}^T \in \mathbb{R}^{2q}$$

to be a vector of the unknown autoregressive and network effect coefficient functions, we then have that $\beta_i(u) = (I_{2q} \otimes \Phi_i(u))\beta_i$.

3.2   Asymptotic Properties of the Estimators

Theorem 4. *Suppose Assumptions 1-5 hold. Let $\hat{\beta}_i(u)$ be vectors of either the LS or the ridge estimator. Further, for the ridge estimator, assume that $\lambda = o(\frac{1}{\sqrt{T}})$. Then, for $U_{it} = u$,*

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

Remark 5. In practice, we use the AIC criterion to select the lags of U_{it}, the number of knots K and the spline order M (for a discussion, see Huang and Shen (2004)).

Also, note that since we assume $\Sigma_\epsilon = \sigma^2 I$, for diverging N and T, we can get the joint asymptotic distribution of any sub-vector of the FCNAR parameter β of fixed dimension. Suppose we are interested in a subset of nodes in $\mathbb{S} := \{N_1, N_2, \cdots, N_S\} \subset \{1, 2, \cdots, N\}$, which is of fixed cardinality S. Denote $\beta_{sv} := \begin{bmatrix} \beta_{N_1}^T & \beta_{N_2}^T & \cdots & \beta_{N_S}^T \end{bmatrix}^T \in \mathbb{R}^{2S(M+K)q}$ to be the sub-vector of β. For example, if we are interested in the joint asymptotic distribution of β_1 and β_2, in this case, $\beta_{sv} = \begin{bmatrix} \beta_1^T & \beta_2^T \end{bmatrix}^T$. Similarly, define

$$\beta_{sv}(u) := \begin{bmatrix} \beta_{N_1}(u)^T & \beta_{N_2}(u)^T & \cdots & \beta_{N_S}(u)^T \end{bmatrix}^T \in \mathbb{R}^{2Sq},$$

and $\Phi_{sv}(u) := \oplus_{i \in \mathbb{S}} I_{2q} \otimes \Phi_i(u)$. We then have that $\beta_{sv}(u) = \Phi_{sv}(u)\beta_{sv}$.

Theorem 5. *Suppose assumptions 1-5 hold. Then, for any sub-vector $\beta_{sv} \in \mathbb{R}^{2S(M+K)q}$ of β, the regression coefficient of the FCNAR model satisfies, as*

$T \to \infty$

$$\sqrt{T}(\hat{\beta}_{sv} - \beta_{sv}) \to_d N(0, \sigma^2 P_{sv}^{-1}),$$

where $P_{sv} := \oplus_{i \in \mathbb{S}} P_i.$

Further, assume that $\lambda = o(\frac{1}{\sqrt{T}})$. *Then, the regression coefficient of the ridge satisfies, as* $T \to \infty$

$$\sqrt{T}(\hat{\beta}_{sv}^{ridge} - \beta_{sv}) \to_d N(0, \sigma^2 P_{sv}^{-1}).$$

And for $U_{it} = u,$

$$\sqrt{T}(\hat{\beta}_{sv}(u) - \beta_{sv}(u)) \to_d N(0, \sigma^2 \Phi_{sv}(u) P_{sv}^{-1} \Phi_{sv}(u)^T).$$

3.3   Hypothesis Testing in FCNAR Models

The technical results established in Theorem 2 enable testing hypotheses of interest on a *fixed* subset of network nodes, as discussed next. Recall that a hypothesis involving p linear constraints on β can be written as:

$$H_0 : \mathbb{D}\beta = r \quad \text{vs} \quad H_a : \mathbb{D}\beta \neq r,$$

where $\mathbb{D}$ for the FCNAR model is a $p \times (2NMq + 2NKq)$ matrix and r is a $p \times 1$ column vector. Then, the test statistic takes the form

$$F = ((\mathbb{D}\hat{\beta})'(\hat{\sigma}^2 \mathbb{D}(\sum_{t=1}^{T} \mathbb{Z}_{t-1}^T \mathbb{Z}_{t-1})^{-1} \mathbb{D}')^{-1}(\mathbb{D}\hat{\beta}))/p \sim F(p, N(T - 2Mq - 2Kq))$$

and follows an F-distribution with the provided degrees of freedom, where

$\hat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \mathbb{Z}_{i(t-1)}^T \hat{\beta}_i)^2.$

For the FCNAR model, hypotheses of particular interest are whether the autoregressive effects $a_{ij}(u)$ or the network effects $b_{ij}(u)$ are the same among nodes, and take the following form:

(A) *Homogeneity of the autoregressive effects amongst nodes*:

$H_0 : a_{ijk} = a_{i'jk}, \quad \forall\, k = 1, 2, \cdots, (M+K),\ \forall\, i, i' = 1, 2, \cdots, N$ and $i \neq i'$.

(B) *Homogeneity of network effects amongst the nodes*:

$H_0 : b_{ijk} = b_{i'jk}, \quad \forall\, k = 1, 2, \cdots, (M+K),\ \forall\, i, i' = 1, 2, \cdots, N$ and $i \neq i'$.

Let $\tilde{\mathbf{0}}$ denote the zero matrix of dimension $(M+K) \times (2q-1)(M+K)$ and $\bar{\mathbf{0}}$ the zero matrix of dimension $(M+K) \times (M+K)$. For the homogeneity test of the lag-1 autoregressive effect, the constraint matrix $\mathbb{D}$ takes the form

$$
\mathbb{D} := \begin{bmatrix}
I_{M+K} & \tilde{\mathbf{0}} & -I_{M+K} & \tilde{\mathbf{0}} & \bar{\mathbf{0}} & \tilde{\mathbf{0}} & \cdots & \bar{\mathbf{0}} & \tilde{\mathbf{0}} \\
\bar{\mathbf{0}} & \tilde{\mathbf{0}} & I_{M+K} & \tilde{\mathbf{0}} & -I_{M+K} & \tilde{\mathbf{0}} & \cdots & \bar{\mathbf{0}} & \tilde{\mathbf{0}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
\bar{\mathbf{0}} & \tilde{\mathbf{0}} & \bar{\mathbf{0}} & \tilde{\mathbf{0}} & \bar{\mathbf{0}} & \tilde{\mathbf{0}} & \cdots & -I_{M+K} & \tilde{\mathbf{0}}
\end{bmatrix}.
$$

Another set of hypotheses of particular interest are those pertaining to the functions of the autoregressive effects and the network effects being con-

stant, which translates that the corresponding relationship to the outcome variable is *linear*. These hypotheses take the form:

(C1) *Linearity of the autoregressive effects*:

$$H_0 : a_{ij2} = a_{ij3} = \cdots = a_{ij(M+K)} = 0.$$

(C2) *Linearity of the network effects*:

$$H_0 : b_{ij2} = b_{ij3} = \cdots = b_{ij(M+K)} = 0.$$

If the null hypothesis in any of these two cases holds, it leads to a simplification in the FCNAR model specification and a reduction in the number of model parameters to be estimated. If both hypotheses (A) and (B) hold, the FCNAR model can be reduced to the NAR model.

For hypothesis (A) the constraint matrix $\mathbb{D}$ takes the form

$$\mathbb{D} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & \mathbf{0}^T_{M+K} \\ 0 & 0 & 1 & \cdots & 0 & \mathbf{0}^T_{M+K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \mathbf{0}^T_{M+K} \end{bmatrix}.$$

Then, the F-test statistic is given by:

$$F = ((\mathbb{D}\hat{\beta}_{ij})'(\hat{\sigma}^2 \mathbb{D}(\sum_{t=1}^{T} Z^T_{i(t-j)} Z_{i(t-j)})^{-1} \mathbb{D}')^{-1} (\mathbb{D}\hat{\beta}_{ij}))/(M + K - 1)$$

has a F-distribution with $(M + K - 1, T - 2Mq - 2Kq)$ degrees of freedom, and $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (X_{it} - Z^T_{i(t-1)}\hat{\beta}_i)^2$.

Under the alternative hypothesis, the non-centrality parameter is given by $\delta = (\mathbb{D}\beta_{ij})'(\sigma^2\mathbb{D}(\sum_{t=1}^{T} Z_{i(t-j)}^T Z_{i(t-j)})^{-1})\mathbb{D}')^{-1}(\mathbb{D}\beta_{ij})$. Analogous expressions can be derived for testing hypothesis (B).

4. Performance Evaluation of the LS and Ridge Estimators

We examine the performance of the LS and the ridge estimators for the FCNAR model through numerical experiments based on synthetic data, and the impact that the sample size T, the spline order M, the number of knots K, the size of the neighborhood in W can have.

The data are generated from the following FCNAR$(1, 1)$ model

$$\mathbb{X}_t = A(\mathbb{U}_t)\mathbb{X}_{t-1} + B(\mathbb{U}_t)W\mathbb{X}_{t-1} + \epsilon_t, \tag{4.9}$$

where $a_i(u) := 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$ and $b_i(u) := -0.437 - (0.659 + 1.260u)e^{-3.89u^2}$, $i = 1, 2, \cdots, N$. The network size is fixed as $N = 100$. Further, the K knots $(k_1, k_2, ...k_K)$ in the spline basis functions are evenly placed between the 1% quantile and the 99% quantile of the threshold variable U_{it}.

4.1   Estimation of the Autoregressive and Network Effect Functions

We examine the influence of the sample size T, the number of knots K, the spline order M and the number of neighbors (bandwidth) of the network matrix W on the estimation of the $a(\cdot)$ and $b(\cdot)$ functions. For a grid $(k_1, k_2, ...k_{200})$ of 200 points evenly placed between 5% quantile and 95% quantile of the threshold variable U_{it}, $\hat{a}_i(u)$ and $\hat{b}_i(u)$ are calculated. The experiment is replicated 50 times, and the 95% quantile, the median and the 5% quantile of the values of the functions at the 200 grid of knots are plotted and compared with the true functions.

The following simulation scenarios are considered:

A.1: $\epsilon_{it} \sim N(0,1), U_{it} \sim N(0,1)$ and W is a banded matrix with bandwidth 2 neighbors. The order of the splines is set to 4, and $K = 10$. Finally, the following 3 sample sizes are used: $T = 200, 400, 3200$.

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

4.1 Estimation of the Autoregressive and Network Effect Functions

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 2 and 3 depict the true functions $a_i(\cdot)/b_i(\cdot)$ and their estimated counterparts $\hat{a}_i(\cdot)/\hat{b}_i(\cdot)$. It can be seen that the estimated functions get closer to the true coefficient functions as the sample size T increases. In particular, for sample sizes larger than 800, the approximation is becoming fairly accurate.

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

4.1   Estimation of the Autoregressive and Network Effect 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.
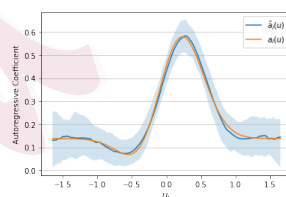
The results confirm that in empirical work wherein the sample size T is given, the spline order M and the number of knots K need to be selected judiciously; for example, Huang and Shen (2004) propose using the AIC criterion for this task.
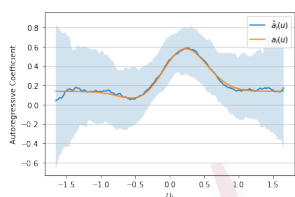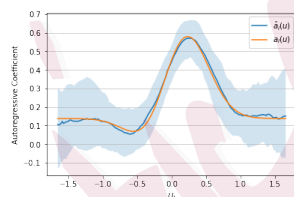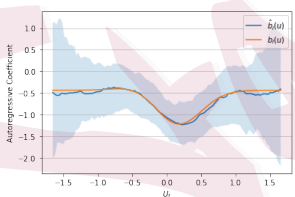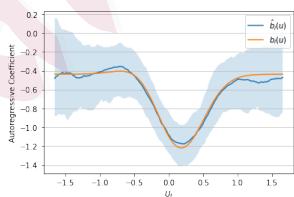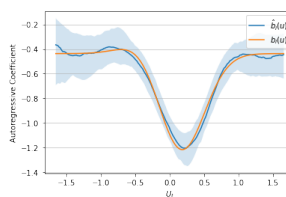

| (2.1) T=200 | (2.2) T=800 | (2.3) T=3200 |

Figure 2: $a_i(\cdot)$ and $\hat{a}_i(\cdot)$ autoregressive functions in scenario A.1.

| (3.1) T=200 | (3.2) T=800 | (3.3) T=3200 |

Figure 3: $b_i(\cdot)$ and $\hat{b}_i(\cdot)$ network effect functions in scenario A.1.

4.2   Confidence Intervals and Coverage Probabilities

Next, we examine the coverage of the confidence intervals based on the results of Theorem 4. The setting considered is as follows: the network size is set to $N = 100$, the error process is distributed according to $\epsilon_{it} \sim N(0,1)$, $U_{it} \sim N(0,1)$ and W is a banded matrix of bandwidth 2. A spline of order $M = 4$ with $K = 20$ knots is used to estimate the functions $a_i(\cdot)$ and $b_i(\cdot)$ while the sample size varies as $T = 200, 400, 800, 1600, 3200$. The experiment is replicated 100 times.

Figure 4 depicts the length of the confidence intervals (CI) and the coverage probabilities (CP) for different thresholds U_{it}. The 95% CI is calculated using $CI_{a_i} = (\hat{a}_i(u) - z_{0.975}SE(\hat{a}_i(u)), \hat{a}_i(u) + z_{0.975}SE(\hat{a}_i(u)))$ and $CI_{b_i} = (\hat{b}_i(u) - z_{0.975}SE(\hat{b}_i(u)), \hat{b}_i(u) + z_{0.975}SE(\hat{b}_i(u)))$, where $SE(\hat{a}_i(u))$ and $SE(\hat{b}_i(u))$ are calculated according to Theorem 4.

It can be seen that the length of the CIs become smaller as the sample size T increases. Further, for any fixed sample size, the length of the CIs is larger for points closer to the boundary of the domain of the threshold variable. On the other hand, with the exception of rather small sample sizes $(T = 200)$, the CP is close to the nominal level, especially for larger values of T. For points closer to the boundary, a larger sample size is required for the CP to be close to the nominal level.

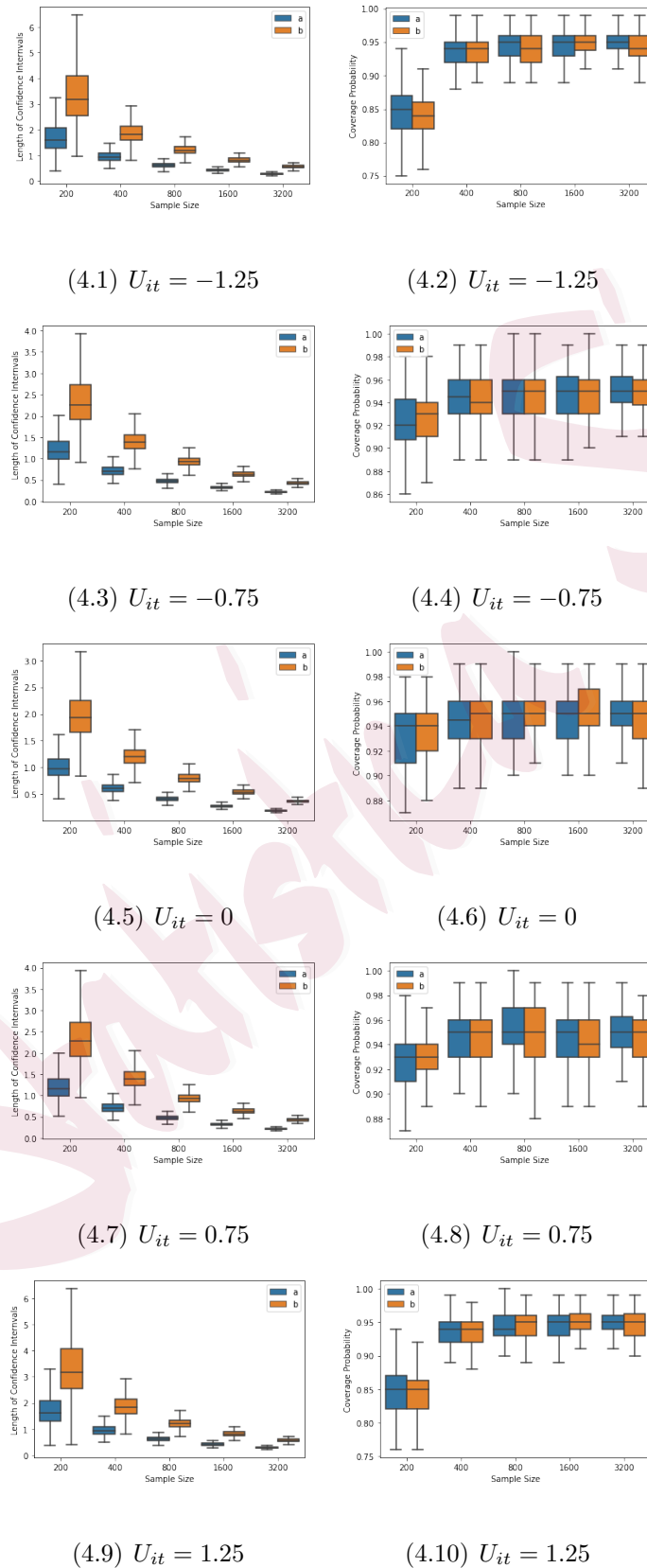## 4.2 Confidence Intervals and Coverage Probabilities

(4.1) $U_{it} = -1.25$

(4.2) $U_{it} = -1.25$

(4.3) $U_{it} = -0.75$

(4.4) $U_{it} = -0.75$

(4.5) $U_{it} = 0$

(4.6) $U_{it} = 0$

(4.7) $U_{it} = 0.75$

(4.8) $U_{it} = 0.75$

(4.9) $U_{it} = 1.25$

(4.10) $U_{it} = 1.25$

Figure 4: CI and CP at different threshold of U_{it}

4.3   Prediction Performance

Next, we compare the predicted RMSE of different models, defined as:

$$\text{predicted RMSE} := \sqrt{\frac{1}{NT} \sum_{t=1}^{T} ||\mathbb{X}_t - \hat{\mathbb{X}}_t||_F^2}.$$

The data is generated by (4.9) with $N = 100$. Two sets of functions $a_i(\cdot)$ and $b_i(\cdot)$ are tested:

B.1 $a_i(u) := 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$ and $b_i(u) := -0.437 - (0.659 + 1.260u)e^{-3.89u^2}$, $i = 1, 2, \cdots, N$.

B.2 $a_i(u) := 0.3I(u \leq 1) - 0.7I(u > 1)$ and $b_i(u) := -0.6I(u \leq 1) + 0.2I(u > 1)$, $i = 1, 2, \cdots, N$.

B.3 $a_i(u) := 0.138 + (0.316 + 0.682u)e^{-0.5u^2}$ and $b_i(u) := -0.437 - (0.259 + 0.560u)e^{-0.5u^2}$, $i = 1, 2, \cdots, N$.

The first $T = 1550$ time points of the data are used for estimating the model parameters, while the last $T = 50$ time points to calibrate the prediction performance. We employ a FCNAR$(1, 1)$, with an order-4 spline with $K = 10$ knots to estimate the autoregressive $a_i(\cdot)$ and network effect $b_i(\cdot)$ parameters. Competing methods include a NAR$(1,1)$ model (i.e, $a_i(\cdot) = a_i, b_i(\cdot) = b_i$, both constants) and a univariate autorgressive time series model with a single lag (as selected by the AIC criterion) applied to

each node i (AR(1)). It can be seen that FCNAR can capture the non-linearity in the data, and thus outperforms the competing models.

Table 1: predicted RMSE for different estimators

	B.1	B.2	B.3
FCNAR(1,1)	1.015	1.022	1.017
NAR(1,1)	1.041	1.104	1.106
AR(1)	1.097	1.135	1.161

5. Applications of the FCNAR model to Real Data Sets

5.1 Application of the FCNAR model to Air Quality Indicators

The proposed FCNAR model is employed to analyze 6 different air quality indicators, namely CO (carbon monoxide), O3 (ozone), SO2 (sulfur dioxide), NO2 (nitrogen dioxide), PM2.5 and PM10 (particulate matter in the air that are less than 2.5 and 10 micrometers in diameter, respectively). The data come from $N = 346$ monitoring stations across China and are collected by the the China National Environmental Monitoring Center. After aggregating the original hourly data for the period 01/02/2015 to 09/05/2020 to a daily cadence and removing missing values, we end up with $T = 1999$ daily measurements (observations). The first 1800 days are used to estimate

5.1 Application of the FCNAR model to Air Quality Indicators

the FCNAR model parameters, while the last 200 days are used for testing purposes. The $AIC = \log |\hat{\Sigma}_\epsilon| + \frac{2(K+M)(q_1+q_2)}{T}$ information criterion is used to select the number of knots, the lag order, the lag of the threshold variable (which corresponds to the response variable) and the spline order. The selected spline order is $M = 3$, the number of knots $K = 5$ and $\mathbb{U}_t := \mathbb{X}_{t-4}$.

The network matrix W is obtained as follows: let D_{ij} be the spatial distance between two monitoring stations, then the ij-th element of W is defined as:

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

$$H_0 : a_{i2} = a_{i3} = \cdots = a_{i8} = 0, \; or$$

$$H_0 : b_{i2} = b_{i3} = \cdots = b_{i8} = 0$$

is used to test, if each node exhibits a non-linear autoregressive or network effect. The location of the monitoring stations that exhibit such non-linear stations are depicted in Figure 5; specifically, red points represent nodes with non-linear autoregressive coefficients, green points represent nodes with non-linear network coefficients, and blue points represent nodes with both non-linear autoregressive and network coefficients. Figure 5 indicates that for most of the stations, there exist either non-linear autoregressive or network coefficients. Hence, the proposed FCNAR model that accounts for

5.1 Application of the FCNAR model to Air Quality Indicators

heterogeneity proves useful for modeling these 6 air quality indicators.

(5.1) CO   (5.2) O3   (5.3) SO2

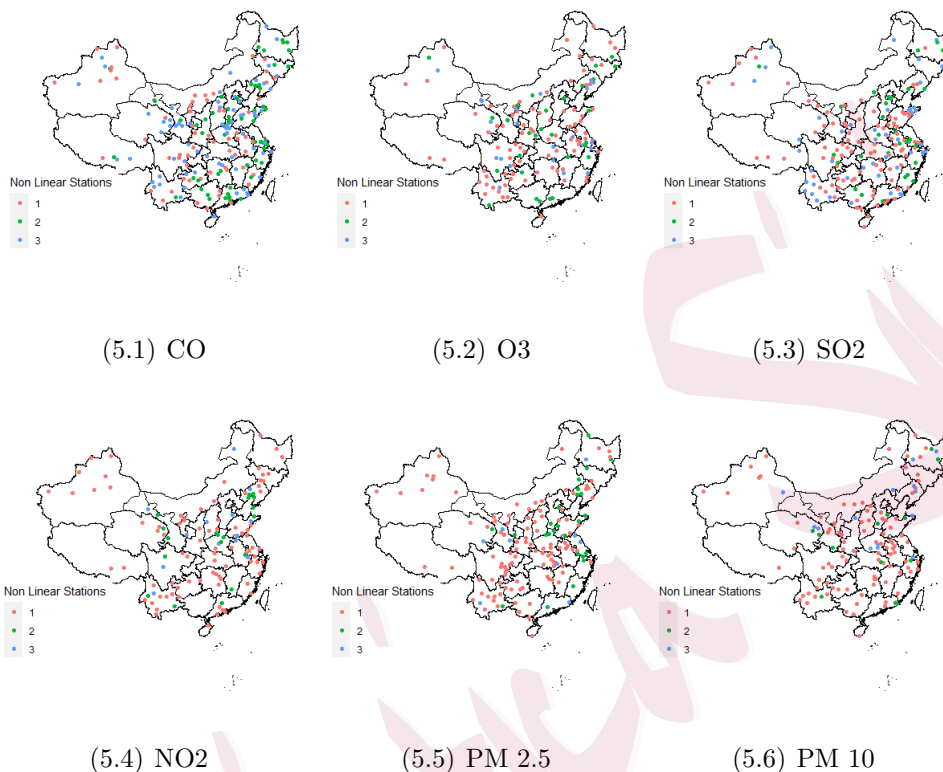(5.4) NO2   (5.5) PM 2.5   (5.6) PM 10

Figure 5: Stations with Non-linear $a(\cdot)$ (red), $b(\cdot)$ (green) and both $a(\cdot)$ and $b(\cdot)$ (blue)

Next, we assess the predictive performance of the FCNAR vis-a-vis selected competitors. Specifically, the predicted RMSE of the ridge based FCNAR(1,1) model is compared to that of a NAR(1,1) model (with linear autoregressive and network effects), a univariate AR(1) model applied to each node, as well as a NAR(1,1) model with homogeneous coefficients (i.e,

5.1 Application of the FCNAR model to Air Quality Indicators

$a_i = a, b_i = b$ for all i nodes). The tuning parameter λ for the ridge based FCNAR(1,1) model is selected by cross-validation. The predicted RMSE of the competing models is given in Table 2. Overall, FCNAR(1,1) performs the best across all air quality indicators, with the linear NAR(1,1) following behind.

Table 2: predicted RMSE for different estimators

	PM10	PM2.5	SO2	NO2	O3	CO
FCNAR(1,1)	0.93425	0.87453	0.71338	0.95353	0.83950	0.66762
NAR(1,1)	0.93575	0.87645	0.71433	0.95380	0.84014	0.66863
AR(1)	0.94832	0.88824	0.72411	0.96565	0.84507	0.67376
NAR with $A = aI$ and $B = bI$	0.94060	0.88210	0.72097	0.96192	0.84420	0.67306

We also compare the relative predicted RMSE improvement of the FCNAR(1,1) model compared to its FAR component (i.e, assuming all $b_i(\cdot) = 0$), and the results are depicted in the next Figure. It can be seen that for the coastal area, the functional network effects are more significant, bringing 4%-6% improvement in predictive performance.

5.1 Application of the FCNAR model to Air Quality Indicators

(6.1) CO   (6.2) O3   (6.3) SO2

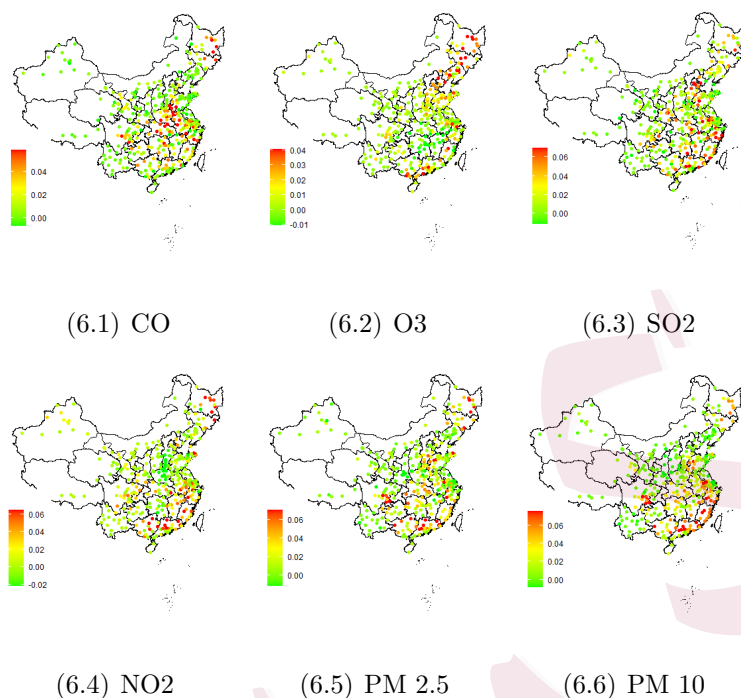(6.4) NO2   (6.5) PM 2.5   (6.6) PM 10

Figure 6: Relative predicted RMSE improvement compared with pure FAR component

Based on the previous finding, we focus on the predictive performance for the stations in Southeast China - the provinces of Fujian and Guangdong. Table 3 compares the predicted RMSE of the FCNAR(1,1) model with the FAR(1) and AR(1) model for the respective stations. Overall, compared to a pure FAR model, there's a 3%-4% improvement in predicted RMSE for the stations in the Southeast by incorporating information from neighboring stations through the FCNAR model.

5.1 Application of the FCNAR model to Air Quality Indicators

Table 3: predicted RMSE for different estimators (southeast)

	PM10	PM2.5	SO2	NO2	O3	CO
FCNAR(1,1)	0.84700	0.87803	0.64093	0.87323	0.82152	0.69485
AR(1)	0.88033	0.90849	0.65938	0.89514	0.82630	0.70224
FAR(1)	0.87931	0.90807	0.65787	0.89576	0.82657	0.70220

For a grid $(k_1, k_2, ... k_{200})$ of 200 points evenly placed between 1% quantile and 99% quantile of the threshold variable U_{it}, $\hat{a}_i(u)$ and $\hat{b}_i(u)$ for $i = 1, 2, \cdots, 346$ are calculated. Based on the 200 $\hat{a}_i(u)$ values and 200 $\hat{b}_i(u)$ values, a total of 400 values for each station, k-means is utilized to partition 346 stations into three clusters. It can be seen that there's a significant geometric pattern of the functional coefficients, from the southeast to the northwest.

(7.1) CO   (7.2) O3   (7.3) SO2

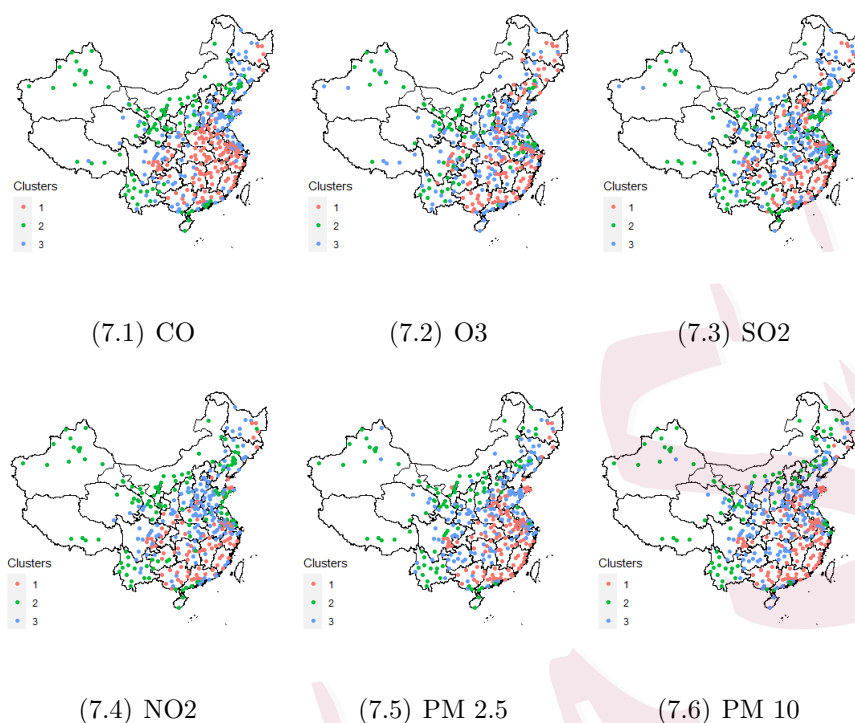(7.4) NO2   (7.5) PM 2.5   (7.6) PM 10

Figure 7: Clusters

5.2   Application to Florida County Level COVID-19 Data

The COVID-19 data used in this study are obtained from The New York
Times (2022) website. The analysis focuses on the 67 counties in the state
of Florida and the raw data include both cases and deaths, as reported by
state and local health departments and compiled by the newspaper. The
network matrix W is calculated from the county adjacency file obtained
from the U.S. Census Bureau (2010). The network matrix is normalized
with each row summing up to 1.

## 5.2 Application to Florida County Level COVID-19 Data

The data we focused on is the daily count of *new* COVID-19 cases for the $N = 67$ Florida counties, covering the period from March 2, 2020 (when the Center for Disease Control reported the first Covid-19 cases in the state of Florida), to Dec 31, 2022, for a total $T = 1036$ observations. The first 900 days are used as the training set, up to 08/17/2022, and the last 136 days are used as the test set.

The $AIC = \log |\hat{\Sigma}_\epsilon| + \frac{2(K+M)(q_1+q_2)}{T}$ information criterion is used to select the number of knots, the lag order, the lag of the threshold variable (which corresponds to the response variable) and the spline order. The model selected is FCNAR$(2, 2)$ with spline order $M = 3$ and $K = 3$ knots, with $\mathbb{U}_t := \mathbb{X}_{t-1}$.

Plots of selected autoregressive $a_i(\cdot)$ and network effect $b_i(\cdot)$ functions that exhibit strong nonlinear patterns are depicted in Figure 8. It can be seen that the larger the lag-value, the larger the autoregressive/network effect. Further, as gleaned from Figure 9 mostly coastal counties exhibit nonlinear network effects

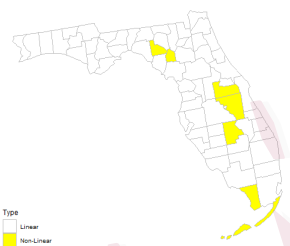## 5.2   Application to Florida County Level COVID-19 Data
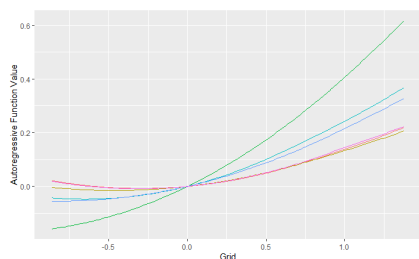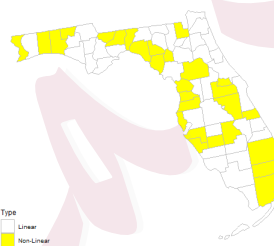

(8.1) $a_i(\cdot)$
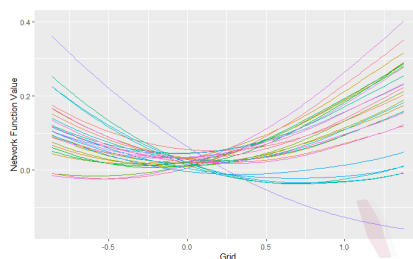
(8.2) $b_i(\cdot)$

Figure 8: Plot of autoregressive $(a_i(\cdot))$ and network $(b_i(\cdot))$ effects with strong nonlinear patterns

(9.1) $a(\cdot)$

(9.2) $b(\cdot)$

Figure 9: Counties with Non-linear autoregressive $(a(\cdot)_i)$ and network $(b_i(\cdot))$ effects.

A closer examination of the network effect functions $(b_i(\cdot))$ shows two distinct patterns, as shown in Figure 10. The first is increasing/decreasing (Figure 10(10.1) and the counties associated with it are shown in Figure 10(10.2), while the other is a quadratic one (Figure 10(10.3)) and the corresponding counties depicted in Figure 10(10.4).

5.2   Application to Florida County Level COVID-19 Data

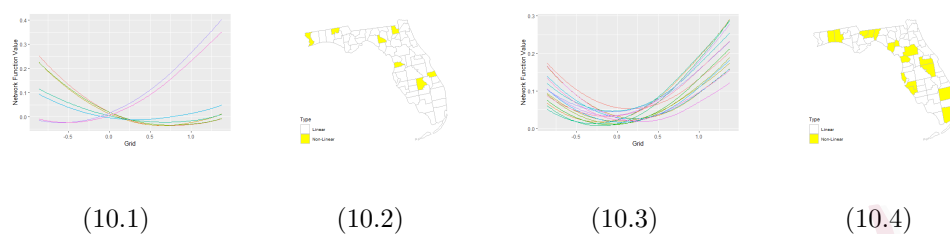| (10.1) | (10.2) | (10.3) | (10.4) |

Figure 10:   The left panels (10(10.1), 10(10.2)) and the right panels (10(10.3), 10(10.4)) depict $b_i(\cdot)$ of the two patterns with their corresponding locations

Next, the predictive performance of the developed FCNAR model is compared with a standard NAR model and a node-wise AR model.

Table 4: predicted RMSE for different estimators

Model	FCNAR(2,2)	FCNAR(2,2) with ridge	NAR(2,2)	AR(2)
predicted RMSE	0.6791	0.6777	0.7521	0.7781

It can be seen that the proposed FCNAR model clearly outperforms the standard NAR one by over 10% in terms of predicted RMSE. The latter also exhibits a small advantage over a univariate AR model fitted to each county separately. For this data set, the added flexibility encompassed in FCNAR contributes to the improved performance.

6. Concluding Remarks

The paper introduces the FCNAR modeling framework that extends both the linear (heterogeneous) NAR model of Yin et al. (2023) by accommodating nonlinear autoregressive and network effects and the univariate FAR model of Chen and Tsay (1993) to a multivariate setting characterized by an underlying network structure. A sufficient condition for the FCNAR process to be stable is provided and we also present least squares and ridge estimators for the functional model parameters that are expressed in spline bases. We establish their asymptotic properties that enable testing of a number of interesting hypotheses, including (i) whether the autoregressive and/or network effects are linear or nonlinear, and (ii) whether the functional network effects components are zero or not, thus reducing the FCNAR to a collection of unrelated FAR models for each node. Experiments on synthetic data assess the performance of the FCNAR model for different sample sizes, and selection of spline basis parameters (order, number of knots). Finally, application to 6 air quality indicators collected across a network of monitoring stations, and on new COVID-19 cases in the state of Florida show that the data exhibit both nonlinear autoregressive and network effects, as well as heterogeneity and hence the proposed FCNAR model has better predictive performance, and also provides better insights

into patterns of the data.

Supplementary Material

The Supplementary material contains statements of auxiliary technical results, the proofs of all Lemmas and Theorems, additional simulation results showcasing the sensitivity of the model parameter estimates to the specification of the network weight matrix W and analysis of an additional data set on wind speeds.

Acknowledgements

References

Armillotta, M. and K. Fokianos (2021). Poisson network autoregression. *arXiv preprint arXiv:2104.06296*.

Cao, Y., H. Lin, T. Z. Wu, and Y. Yu (2010). Penalized spline estimation for functional coefficient regression models. *Computational statistics & data analysis 54*(4), 891–905.

Chen, R. and L.-M. Liu (2001). Functional coefficient autoregressive models: estimation and tests of hypotheses. *Journal of Time Series Analysis 22*(2), 151–173.

REFERENCES

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

Dijk, D. v., T. Teräsvirta, and P. H. Franses (2002). Smooth transition autoregressive models—a survey of recent developments. *Econometric reviews 21*(1), 1–47.

Donges, J. F., J. Heitzig, B. Beronov, M. Wiedermann, J. Runge, Q. Y. Feng, L. Tupikina, V. Stolbova, R. V. Donner, N. Marwan, et al. (2015). Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package. *Chaos: An Interdisciplinary Journal of Nonlinear Science 25*(11), 113101.

Donner, R. V. and S. M. Barbosa (2008). Nonlinear time series analysis in the geosciences. *Lecture Notes in Earth Sciences 112*, 37.

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

Franses, P. H., D. Van Dijk, et al. (2000). *Non-linear time series models in empirical finance*. Cambridge university press.

Gourieroux, C. and A. Monfort (1992). Qualitative threshold arch models. *Journal of econometrics 52*(1-2), 159–199.

Haggan, V. and T. Ozaki (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika 68*(1), 189–196.

Hamadeh, T. and J.-M. Zakoïan (2011). Asymptotic properties of ls and qml estimators for a

REFERENCES

class of nonlinear garch processes. *Journal of Statistical Planning and Inference 141*(1), 488–507.

Hamilton, J. D. (2020). *Time series analysis.* Princeton university press.

Higgins, M. L. and A. K. Bera (1992). A class of nonlinear arch models. *International Economic Review*, 137–158.

Huang, J. Z. and H. Shen (2004). Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scandinavian journal of statistics 31*(4), 515–534.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer Science & Business Media.

The New York Times (2022). Coronavirus (covid-19) data in the united states.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford university press.

U.S. Census Bureau (2010). Coronavirus (covid-19) data in the united states.

Ward, L. M. (2002). *Dynamical cognitive science.* MIT press.

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

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics 45*(3), 1096–1123.

Zou, Y., R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths (2019). Complex network approaches to nonlinear time series analysis. *Physics Reports 787*, 1–97.

REFERENCES

University of Florida, E-mail: hyin@ufl.edu

George Mason University, E-mail: asafikha@gmu.edu

University of California, Los Angeles E-mail: gmichail@ucla.edu