



A General Modeling Framework for Network Autoregressive Processes

Hang Yin, Abolfazl Safikhani & George Michailidis

To cite this article: Hang Yin, Abolfazl Safikhani & George Michailidis (2023): A General Modeling Framework for Network Autoregressive Processes, *Technometrics*, DOI: [10.1080/00401706.2023.2203184](https://doi.org/10.1080/00401706.2023.2203184)

To link to this article: <https://doi.org/10.1080/00401706.2023.2203184>



View supplementary material [↗](#)



Published online: 30 May 2023.



Submit your article to this journal [↗](#)



Article views: 130



View related articles [↗](#)



View Crossmark data [↗](#)



A General Modeling Framework for Network Autoregressive Processes

Hang Yin^a, Abolfazl Safikhani^b, and George Michailidis^c

^aDepartment of Statistics, University of Florida, Gainesville, FL; ^bDepartment of Statistics, George Mason University, Fairfax, VA; ^cDepartment of Statistics & Informatics Institute, University of Florida, Gainesville, FL

ABSTRACT

A general flexible framework for Network Autoregressive Processes (NAR) is developed, wherein the response of each node in the network linearly depends on its past values, a prespecified linear combination of neighboring nodes and a set of node-specific covariates. The corresponding coefficients are node-specific, and the framework can accommodate heavier than Gaussian errors with spatial-autoregressive, factor-based, or in certain settings general covariance structures. We provide a sufficient condition that ensures the stability (stationarity) of the underlying NAR that is significantly weaker than its counterparts in previous work in the literature. Further, we develop ordinary and (estimated) generalized least squares estimators for both fixed, as well as diverging numbers of network nodes, and also provide their ridge regularized counterparts that exhibit better performance in large network settings, together with their asymptotic distributions. We derive their asymptotic distributions that can be used for testing various hypotheses of interest to practitioners. We also address the issue of misspecifying the network connectivity and its impact on the aforementioned asymptotic distributions of the various NAR parameter estimators. The framework is illustrated on both synthetic and real air pollution data.

ARTICLE HISTORY

Received October 2022
Accepted April 2023

KEYWORDS

Inference; Misspecified network connectivity; Network autoregressive process; Ridge penalty; Stability

1. Introduction

Consider a network comprising of N nodes, for which we collect measurements over T time periods for a variable X ; that is, $X_{it}, i = 1, \dots, N, t = 1, \dots, T$. Depending on the application of interest, these nodes may correspond to agents/actors in a social network, companies in an economic network, sensors in an environmental network and even physical sites or devices in an engineering network. Further, for each node i we also observe p covariates $Y_{i,t} \in \mathbb{R}^p$ that are also time-varying. The model posited next assumes that the measurements X_{it} for node i are influenced by their past values (self-lags), plus past values of “related” nodes (network lags), after adjusting for the effect of covariates. Henceforth, we refer to this model as the Network Autoregressive (NAR) model. The corresponding NAR(q_1, q_2) process with q_1 self-lags and q_2 network lags takes the form:

$$X_{it} = \sum_{j=1}^{q_1} a_i^{(j)} X_{i(t-j)} + \sum_{j=1}^{q_2} b_i^{(j)} \sum_{k=1}^N w_{ik} X_{k(t-j)} + \gamma_i^T Y_{i,t-1} + \epsilon_{it},$$

$$i = 1, \dots, N, \quad (1)$$

where $a_i^{(j)} \in \mathbb{R}, b_i^{(j)} \in \mathbb{R}, \gamma_i \in \mathbb{R}^p$ are regression coefficients for the self-lags, the network lags and the covariates, respectively; further, $w_{ik} \in [0, 1]$ are weights capturing the degree of dependence among node i and other nodes $k \neq i$. We impose further constraints on these weights in the sequel (see Assumption 2).

Finally, ϵ_{it} is an error term with $E(\epsilon_{it}) = 0$ and $E(\epsilon_{it}^4) < \infty$, which is assumed to be independent of the covariates $Y_{i,t}$.

An example of an NAR(1,1) model with three nodes is given in Figure 1. Every node i is influenced by its past values and a linear combination of its neighbors' past values through the i th row of weight matrix w_i . Additional conditions on the error processes are discussed in the sequel.

The posited model encompasses as special cases a number of models that appeared in recent literature, and also extends other related models, as discussed next. Specifically, Zhu et al. (2017) consider an NAR model with $a_i^{(j)} = a^{(j)}, b_i^{(j)} = b^{(j)}$ for all $i = 1, \dots, N$, while Zhu and Pan (2018) assume that the nodes belong to K groups $G_k, k = 1, \dots, K$ and thus all nodes in group G_k share the same coefficients; that is, $a_i^{(j)} = a_k^{(j)}, b_i^{(j)} = b_k^{(j)}$, for all $i \in G_k$. The assignment of nodes into groups is obtained from the data, by assuming a mixture model. Further, in both cases the error term is homoscedastic, that is, $\epsilon_k \sim N(0, \sigma_k I), k = 1, \dots, K$. A variation of the model in Zhu and Pan (2018) is presented in Chen, Fan, and Zhu (2020), wherein the adjacency matrix of the network W is assumed to be generated by a Stochastic Block model with K communities, which allows interactions between nodes belonging to the same community, as well as belonging to different communities. Further, the covariance matrix of the error term can exhibit factor structure, while the community structure is estimated from the data through spectral clustering. Knight et al. (2020) allow for different autoregressive coefficients for the nodes, and different

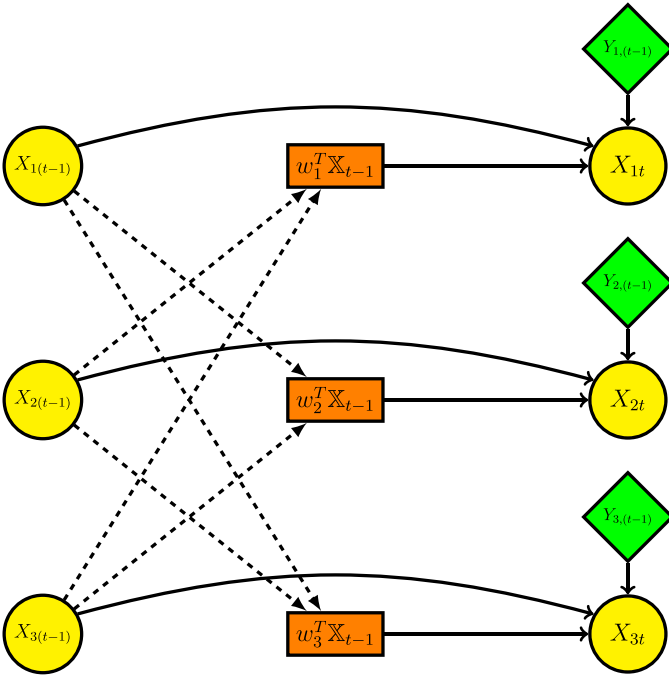


Figure 1. An example of an NAR(1,1) model instance with three network nodes.

specifications of the network effects with a single regression coefficient for each network specification, but do not consider exogenous covariates. A paper focusing on the economic impact of COVID-19 mitigation strategies and fatality rates (Nason and Wei 2021) enhances the model in Knight et al. (2020) by incorporating exogenous covariates. Further, a variant of the popular in the econometrics literature Seemingly Unrelated Regressions model in Zellner (1962) is also encompassed by the NAR one; specifically, by letting $y_{it} = \beta_i^T x_{it} + \epsilon_{it}$, where $\epsilon_{it} \sim F(0, \Sigma)$, and $y_{it} := X_{it}$, $x_{it} := [X_{i(t-1)} \ w_i^T X_{t-1} \ Y_{i(t-1)}^T]^T$, where $Y_{i(t-1)}$ are defined to be exogenous covariates.

Specific variants of the NAR model have been employed in diverse application areas, including social media analysis (Zhu et al. 2017), pollution (Zhu and Pan 2018), environmental monitoring and economic growth studies (Knight et al. 2020) and predicting stock market returns (Chen, Fan, and Zhu 2020).

In the study of the NAR multivariate $\{X_{it}\}$ process, the following two issues need to be addressed at the technical level: (i) conditions that ensure its stability/stationarity, and (ii) estimation of the model parameters and inference in different regimes; namely, (a) wherein the number of network nodes is fixed, while the number of time points grows and (b) wherein both the number of network nodes and time points grow at appropriate rates. For the first issue, the work in the literature has adopted a rather stringent sufficient condition, that this article substantially relaxes (see Theorem 2.1 and ensuing discussion in Remark 3). For the second issue, the nature of the posited model dictates the estimation procedure and associated inference results. Specifically, Zhu et al. (2017) use ordinary least squares and establish asymptotic normality for the *fixed number* of estimated parameters present in their model specification. Zhu and Pan (2018) use the EM algorithm to identify the underlying group structure, and then apply the NAR model defined in Zhu et al. (2017) to each group. Hence, a fixed number of model

parameters is also considered. Knight et al. (2020) use a least squares criterion to fit the model and establish the asymptotic normality of the model parameters assuming that the *network size* N is fixed. The same setting of a fixed network size is used in the extension by Nason and Wei (2021) that incorporates exogenous covariates. Chen, Fan, and Zhu (2020) use a multi-step estimation procedure to first identify the community structure, then the factor structure of the error term and finally through generalized least squares obtain estimates of the model parameters. Further, asymptotic distributions for the parameters are also derived. Finally, Armillotta and Fokianos (2021) consider a network autoregressive model for count data, with *common parameters* for all nodes and use quasi-likelihood methods for inference for the model parameters.

In contrast, the posited model in (1) can accommodate *heterogeneity* across network nodes, by having a different network effect parameter for each node. Further, the network size is allowed to grow as a function of the time observations, which together with the model formulation leads to increasing parameter space. Hence, estimation and inference issues require technical care, as seen in Propositions C.1 and D.1. In addition, a general structure of the covariance matrix is assumed for the error term, which is also flexible, but further contributes to the technical challenges. Hence, the key contributions of this work are: (i) building a general flexible modeling framework for network autoregressive data (Section 1), (ii) developing a relaxed sufficient condition for stability/stationarity of the underlying NAR process (Section 2), (iii) establishing inference procedures for the growing number of model parameters, including regularized variants of the (empirical) generalized least squares estimates (Section 3), and (iv) addressing model misspecification issues regarding the network matrix W (Appendix E).

Notation. Throughout the article, we use $\|A\|_\infty$ to denote the matrix induced infinity norm of matrix $A \in \mathbb{R}^{m \times n}$, that is, $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$. We use $\|A\|_{\max}$, $\|A\|$ and $\|A\|_F$ to

denote the element-wise max norm, the operator norm and Frobenius norm of A , respectively. We use e_i to denote the i th unit vector in $\mathbb{R}^p$. For matrices, we use $\rightarrow_p$ to denote element-wise convergence in probability, and $\rightarrow_d$ to denote convergence in distribution. For a symmetric or Hermitian matrix A , we denote its spectral radius by $\rho(A)$, where the spectral radius of a square matrix is the maximum of the absolute values of its eigenvalues.

2. Stability of the NAR Process

The first issue addressed is to derive conditions that ensure the stability/stationarity of the NAR(q_1, q_2) process for the model posited in (1). To proceed, some additional notation is required.

Let $X_t := [X_{1t} \ X_{2t} \ \dots \ X_{Nt}]^T$, $\epsilon_t := [\epsilon_{1t} \ \epsilon_{2t} \ \dots \ \epsilon_{Nt}]^T$, $A_i := \text{diag}\{a_1^{(i)}, a_2^{(i)}, \dots, a_N^{(i)}\} \in \mathbb{R}^{N \times N}$ for $i = 1, 2, \dots, q_1$, $B_j := \text{diag}\{b_1^{(j)}, b_2^{(j)}, \dots, b_N^{(j)}\} \in \mathbb{R}^{N \times N}$ for $j = 1, 2, \dots, q_2$, $C_k := \text{diag}\{\gamma_{1k}, \gamma_{2k}, \dots, \gamma_{Nk}\} \in \mathbb{R}^{N \times N}$ for $k = 1, 2, \dots, p$ where γ_{ik} is the k th covariate for node i , and $G_\ell := A_\ell + B_\ell W$, wherein $\ell = 1, 2, \dots, \max\{q_1, q_2\}$, with the convention that zero matrices are included/padded for the relationship to hold; namely, if $q_1 > q_2$,

$B_j = 0$ for $j > q_2$, whereas if $q_1 < q_2$, $A_j = 0$ for $j > q_1$. Let $q = \max\{q_1, q_2\}$, then NAR(q_1, q_2) posited in matrix form can be written as

$$\begin{aligned}\mathbb{X}_t &= \sum_{i=1}^{q_1} A_i \mathbb{X}_{t-i} + \sum_{j=1}^{q_2} B_j W \mathbb{X}_{t-j} + \sum_{k=1}^p C_k \mathbb{Y}_{k,(t-1)} + \epsilon_t \\ &= \sum_{\ell=1}^q G_\ell \mathbb{X}_{t-\ell} + \sum_{k=1}^p C_k \mathbb{Y}_{k,(t-1)} + \epsilon_t,\end{aligned}\quad (2)$$

wherein $\mathbb{X}_t$ and $\mathbb{X}_{t-\ell} \in \mathbb{R}^N$ and $\mathbb{Y}_{k,(t-1)} = [Y_{1k,(t-1)} \cdots Y_{Nk,(t-1)}]^T \in \mathbb{R}^N$ and $Y_{ik,(t-1)}$ being the k th element of $Y_{i,(t-1)}$. We impose $Y_{ik,(t-1)}$ further constraints on these weights in the sequel (see [Assumption 2](#)). Finally, ϵ_{it} is an error process with $E(\epsilon_{it}) = 0$ and $E(\epsilon_{it})^4 < \infty$, which is assumed to be independent of the covariates $Y_{i,t}$. For technical developments, it is convenient to also express (2) in the following form:

$$\mathbb{X}_t = \mathbb{Z}_{t-1} \beta + \epsilon_t, \quad (3)$$

where $\mathbb{Z}_{t-1} := [Z_{t-1} \cdots Z_{t-q} \text{diag}\{\mathbb{Y}_{1,(t-1)}\} \cdots \text{diag}\{\mathbb{Y}_{p,(t-1)}\}]$, $Z_{t-\ell} := [\text{diag}\{\mathbb{X}_{t-\ell}\} \text{diag}\{W \mathbb{X}_{t-\ell}\}]$, $\beta := [\beta_1^T \beta_2^T \cdots \beta_q^T \gamma_1^T \gamma_2^T \cdots \gamma_p^T]^T$, $\beta_\ell := [a_1^{(\ell)} a_2^{(\ell)} \cdots a_N^{(\ell)} b_1^{(\ell)} b_2^{(\ell)} \cdots b_N^{(\ell)}]^T$, and $\gamma_k := [\gamma_{1k} \gamma_{2k} \cdots \gamma_{Nk}]^T$, for $\ell = 1, \dots, q$ and $k = 1, \dots, p$.

Let $\tilde{\epsilon}_t = \epsilon_t + \sum_{k=1}^p C_k \mathbb{Y}_{k,(t-1)}$ and rewrite (2) as $\mathbb{X}_t = \sum_{\ell=1}^q G_\ell \mathbb{X}_{t-\ell} + \tilde{\epsilon}_t$, which can be considered as a vector autoregressive model (VAR) with transition matrix G and error term $\tilde{\epsilon}_t$. The latter model can also be expressed as a VAR(1) model (see (Lütkepohl 2005)):

$$\mathbf{X}_t = \mathbf{G} \mathbf{X}_{t-1} + \mathcal{E}_t, \quad \text{with} \quad (4)$$

$$\mathbf{X}_t := \begin{bmatrix} \mathbb{X}_t^T \\ \mathbb{X}_{t-1}^T \\ \vdots \\ \mathbb{X}_{t-q+1}^T \end{bmatrix}, \quad \mathcal{E}_t := \begin{bmatrix} \tilde{\epsilon}_t^T \\ 0^T \\ \vdots \\ 0^T \end{bmatrix}, \quad \mathbf{G} := \begin{bmatrix} G_1 & \cdots & G_{q-1} & G_q \\ I_N & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_N & 0 \end{bmatrix}. \quad (5)$$

Before stating the main result, we introduce the following assumptions:

Assumption 1. Moment conditions on ϵ_t and $\mathbb{Y}_t$:

- (i) $\{\epsilon_t, t \in \mathbb{N}\}$ is an iid sequence over the time index t of random vectors satisfying $E(\epsilon_t) = 0$, and $\Sigma_\epsilon = E(\epsilon_t \epsilon_t^T)$ is non-singular. Further, for some finite constant c_1 , the following relationship holds

$$E|\epsilon_{it} \epsilon_{jt} \epsilon_{kt} \epsilon_{mt}| \leq c_1 \text{ for } i, j, k, m = 1, \dots, N, \text{ and all } t.$$

- (ii) $\{\mathbb{Y}_t, t \in \mathbb{N}\}$, where $\mathbb{Y}_t := [\mathbb{Y}_{1,t}^T \mathbb{Y}_{2,t}^T \cdots \mathbb{Y}_{p,t}^T]^T$, is a stationary process satisfying $E(\mathbb{Y}_t) = 0$, $E(\mathbb{Y}_t \mathbb{Y}_t^T) = \Gamma_Y(0) = \Sigma_Y$, $E(\mathbb{Y}_t \mathbb{Y}_{t-j}^T) = \Gamma_Y(j)$, and with absolutely summable autocovariance function. For $i_1, i_2, i_3, i_4 = 1, \dots, N$, $j_1, j_2, j_3, j_4 = 1, \dots, p$ and $t_1, t_2, t_3, t_4 = 1, \dots, T$ and some finite constant c_2 , the following holds:

$$E|Y_{i_1 j_1, t_1} Y_{i_2 j_2, t_2} Y_{i_3 j_3, t_3} Y_{i_4 j_4, t_4}| \leq c_2$$

- (iii) The error process $\{\epsilon_t, t \in \mathbb{N}\}$ is independent of the covariate process $\{\mathbb{Y}_t, t \in \mathbb{N}\}$.

Assumption 2. $W \in \mathbb{R}^{N \times N}$ is a row-normalized matrix; that is, $\sum_{j=1}^N w_{ij} = 1$ with $w_{ii} = 0$ and $w_{ij} \geq 0$ for $i \neq j$, $\forall i, j = 1, 2, \dots, N$.

Assumption 3. For diverging network size N as a function of time T , the error $\{\epsilon_t, t \in \mathbb{N}\}$ and covariate $\{\mathbb{Y}_{t-1}, t \in \mathbb{N}\}$ processes are assumed to be sub-Weibull (see definition in Appendix A of the supplementary materials.)

Assumptions 1 (i)–(ii) require existence of fourth moments for the error term as well as the covariate process. While finiteness of second moments is sufficient to ensure the existence of a unique stationary solution to the recursive equations (2), finiteness of fourth moments is required to establish the asymptotic normality of the various estimators presented in [Section 3](#). In addition, [Assumption 1](#) (ii) is satisfied by a wide class of processes, such as stable linear processes that include vector autoregressive and moving average processes as special cases. [Assumption 1](#) (iii) requires independence between the error and the covariate process, which makes the latter process *exogenous*. [Assumption 2](#) is required for the identifiability of the network coefficients and is needed both for establishing the stability/stationarity of the NAR process and for the asymptotic properties of the estimators of the model parameters. Finally, [Assumption 3](#) imposes a mild condition on the tail behavior of the distribution of the error and the covariate processes that encompass a wide range of possibilities, including sub-Gaussian and sub-exponential random variables.

Remark 1. Note that all prior work in the literature Zhu et al. (2017), Zhu and Pan (2018), Chen, Fan, and Zhu (2020), Knight et al. (2020), and Nason and Wei (2021) assumes that both the exogenous variables Y_t , as well as the error terms ϵ_t are normally distributed. [Assumption 3](#) relaxes significantly this requirement. Further, Zhu et al. (2017) and Zhu and Pan (2018) assume that the exogenous process Y_t is an iid sequence over the time index t , whereas [Assumption 1](#) (ii) allows for temporal dependence that proves useful in applications (see [Section 5](#)).

Remark 2 (Missing Data in the NAR Model). [Assumption 2](#) requires row normalization of the weights of the network matrix W . A small adjustment on this assumption suffices for the technical framework for the NAR process to hold, in the presence of missing data either for the process $\{X_{it}\}$, $t = 1, \dots, T$, $i = 1, \dots, N$, or whenever network edges drop-out and then reappear. The key idea is that the influence of nodes with missing values (or missing network edges) will receive zero weights in their corresponding columns and the remaining weight values will be adjusted to ensure that each row sums up to one and hence satisfy [Assumption 2](#). Thus, data missing from a specific node or in the presence of a missing network edge do not enter the network effect estimation calculations, while the identifiability assumption on weight matrices still holds due to other nodes' weight adjustments. Note that this adjustment obviates the need for using imputation techniques. This issue is handled in analogous manner for the GNAR model (see details in section 2.6 in Knight et al. (2020)).

Theorem 2.1. Consider the NAR(q_1, q_2) process defined recursively by $\mathbb{X}_t = \sum_{\ell=1}^q G_\ell \mathbb{X}_{t-\ell} + \tilde{\epsilon}_t$ where $G_\ell = A_\ell + B_\ell W$ and

$\tilde{\epsilon}_t = \epsilon_t + \sum_{k=1}^p C_k \mathbb{Y}_{k,(t-1)}$. Assume [Assumptions 1–2](#) hold. Then, $\mathbb{X}_t$ is a stationary process with a finite first-order moment that can be expressed as

$$\mathbb{X}_t = \begin{bmatrix} I_N & \mathbf{0}_{N \times N(q-1)} \end{bmatrix} \sum_{j=0}^{\infty} \mathbf{G}^j \mathcal{E}_{t-j},$$

if $\rho(\mathbf{G}) < 1$. Equivalently, $\mathbb{X}_t$ is a stationary process if

$$\det(I_{Nq} - \mathbf{G}z) = \det(I_N - G_1 z - \dots - G_q z^q) \neq 0 \text{ for } |z| \leq 1.$$

Remark 3. (i) For row-normalized W , $\max_{1 \leq i \leq N} \{ \sum_{l=1}^q (|a_i^{(l)}| + |b_i^{(l)}|) \} < 1$ is only a sufficient condition for an NAR(q_1, q_2) model to be stationary.

(ii) The stability/stationarity condition in [Theorem 2.1](#) is significantly weaker than those in the literature for even special cases of the posited model as illustrated in [Appendix A.1](#) in the supplementary materials.

3. Estimation Procedures for the NAR Model and their Asymptotic Properties

We consider the following estimators for the NAR model: (a) generalized least squares (GLS) (with ordinary least squares (OLS) being a special case), and (b) the empirical counterpart of (a): empirical generalized least squares (EGLS), in the following two regimes: (I) $N \leq T$ and (II) $N > T$. In both regimes, both the network size N and the number of time points T grow to infinity at appropriately defined rates that impact the asymptotics. The case of a fixed size network size can be subsumed in case (I).

Note that the GLS estimator is primarily of theoretical interest, but serves as a building step for its empirical counterpart used in practice, by identifying the assumptions required for its consistency and asymptotic normality.

3.1. GLS Estimator

The GLS estimator is defined next, and uses the covariance structure of the error term:

$$\begin{aligned} \hat{\beta}_{\text{GLS}} &= \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1} \right)^{-1} \sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{X}_t \\ &= \beta + \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1} \right)^{-1} \sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \epsilon_t, \end{aligned} \quad (6)$$

where $\Sigma_\epsilon := E(\epsilon_t \epsilon_t^T)$.

Note that if one ignores the structure of the covariance matrix Σ_ϵ , then the OLS estimator is obtained by setting $\Sigma_\epsilon = I$.

(I) Growing network size with $N \leq T$:

Theorem 3.1 (Asymptotic Properties of the GLS Estimator). Suppose [Assumptions 1–3](#) hold. Let $\mathbb{X}_t$ be a stationary process

generated by the NAR(q_1, q_2) model (3), that is, $\mathbb{X}_t = \mathbb{Z}_{t-1} \beta + \epsilon_t$ with growing network size $N \leq T$. Define $D \in \mathbb{R}^{k \times (2Nq+Np)}$ for any finite k . Further, assume:

- D has bounded row sums; that is, for $i = 1, \dots, k$, there exists a finite constant c such that $\sum_{j=1}^{2Nq+Np} d_{ij} \leq c$ where d_{ij} is the ij th element of D .
- Σ_ϵ^{-1} has bounded row sums.

Then,

$$\frac{1}{\sqrt{T}} D \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1} \right) (\hat{\beta}_{\text{GLS}} - \beta) \rightarrow_d N(0, D Q D^T) \quad (7)$$

where $Q := E(\mathbb{Z}_t^T \Sigma_\epsilon^{-1} \mathbb{Z}_t)$.

Remark 4. Note that if the network size N is fixed, the expression in (7) simplifies to $\sqrt{T}(\hat{\beta}_{\text{GLS}} - \beta) \rightarrow_d N(0, Q^{-1})$; the detailed result and its derivation is presented in [Appendix C.1](#) (supplementary materials).

(II) Growing network size with $N > T$:

In this case, *regularization* needs to be introduced. We consider a ridge GLS estimator that uses the covariance structure of the error term:

$$\hat{\beta}_{\text{ridge}} = \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1} + TM \right)^{-1} \sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{X}_t, \quad (8)$$

where $\Sigma_\epsilon := E(\epsilon_t \epsilon_t^T)$ and $M := \text{diag}\{\lambda_1 I_N, \lambda_2 I_N, \dots, \lambda_1 I_N, \lambda_2 I_N, \lambda_3 I_{Np}\}$. Specifically, λ_1, λ_2 , and λ_3 are the tuning parameters for the autoregressive, the network and the exogenous covariates coefficients, respectively.

Theorem 3.2 (Asymptotic Properties of the ridge regularized GLS Estimator). Suppose [Assumptions 1–3](#) hold. Let $\mathbb{X}_t$ be a stationary process generated by the NAR(q_1, q_2) model (3), that is, $\mathbb{X}_t = \mathbb{Z}_{t-1} \beta + \epsilon_t$ with growing network size N . Define $D \in \mathbb{R}^{k \times (2Nq+Np)}$ for any finite k . Further, assume:

- D has bounded row sums; that is, for $i = 1, \dots, k$, there exists a finite constant c such that $\sum_{j=1}^{2Nq+Np} d_{ij} \leq c$ where d_{ij} is the ij th element of D .
- Σ_ϵ^{-1} has bounded row sums.
- $N > T$, and $\lambda_i = o(\frac{1}{\sqrt{T}})$ for $i = 1, 2, 3$. Then,

$$\frac{1}{\sqrt{T}} D \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1} + TM \right) (\hat{\beta}_{\text{ridge}} - \beta) \rightarrow_d N(0, D Q D^T) \quad (9)$$

where $Q := E(\mathbb{Z}_t^T \Sigma_\epsilon^{-1} \mathbb{Z}_t)$.

3.2. EGLS Estimator

As previously mentioned, the EGLS estimator defined as

$$\hat{\beta}_{\text{EGLS}} = \left(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \hat{\Sigma}_\epsilon^{-1} \mathbb{Z}_{t-1} \right)^{-1} \sum_{t=1}^T \mathbb{Z}_{t-1}^T \hat{\Sigma}_\epsilon^{-1} \mathbb{X}_t \quad (10)$$

is of primary interest in practice. The asymptotic properties established for the GLS estimator in [Theorem 3.1](#) hold for EGLS, as long as a *consistent estimator* $\hat{\Sigma}_\epsilon$ for Σ_ϵ is available (full details provided in Proposition D.1 in Appendix D).

Such an estimator is readily available for the case $N < T$, by using the residuals from the OLS estimator to obtain $\hat{\Sigma}_\epsilon$ (see Proposition D.2 in Appendix D). However, this is not the case for $N > T$. To address this issue, we consider two popular models for the structure of the covariance matrix of the error term for the NAR(q_1, q_2) model. We consider both a factor model and a spatial autoregressive structure one for Σ_ϵ , described in detail (due to space considerations) in Appendices D.2, D.3 in the supplement and used in the application presented in [Section 5](#).

3.3. Testing Hypotheses of Interest

The technical results established for the (E)GLS estimator enable testing a number of hypotheses of interest. Next, we present the null hypotheses of the test statistic for selected ones:

- (A) *Homogeneity of the autoregressive effects:* $H_0 : a_1^{(\ell)} = \dots = a_N^{(\ell)}, \ell = 1, \dots, q$.
- (B) *Homogeneity of the network effects:* $H_0 : b_1^{(\ell)} = \dots = b_N^{(\ell)}, \ell = 1, \dots, q$.
- (C) *Homogeneity of both the autoregressive and the network effects:* $H_0 : \beta_1 = \dots = \beta_q$.

If the null hypothesis in any of these three cases holds, it leads to a simplification in the model specification and a reduction in the number of model parameters to be estimated.

Let K denote a matrix of size $(N-1) \times (Nq_1 + Nq_2 + Np)$ that encodes the null hypothesis of interest. For example, for $a_1^{(1)} = \dots = a_N^{(1)}$, K takes the form

$$K := \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{0}_{N(q_1-1+q_2+p)}^T \\ \mathbf{0}_{N(q_1-1+q_2+p)}^T \\ \mathbf{0}_{N(q_1-1+q_2+p)}^T \end{bmatrix}.$$

Then, the test statistic has the following form:

$$F = ((K\hat{\beta})'(K(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \hat{\Sigma}_\epsilon^{-1} \mathbb{Z}_{t-1})K')^{-1}(K\hat{\beta}))/((N-1) \sim F_{(N-1), NT-N \times (q_1+q_2+p)}.$$

Under the alternative hypothesis, the noncentrality parameter is given by

$$\delta = (K\beta)'(K(\sum_{t=1}^T \mathbb{Z}_{t-1}^T \Sigma_\epsilon^{-1} \mathbb{Z}_{t-1})K')^{-1}(K\beta).$$

A brief evaluation of the F -test statistic is given in Section G.3 of the Appendix.

4. Performance Evaluation

Several factors influence the performance of the various estimators proposed, including the sample size T , the number of network nodes N , the structure of the weight matrix W , the lag orders (q_1, q_2) and the parameterization of the error covariance

matrix Σ_ϵ (spatial autoregressive structure vs. factor structure). The performance metrics considered include the root-mean-square error (RMSE) of the model parameters, together with the coverage probability of the constructed confidence intervals and their average length.

Next, we describe the data-generating mechanism and the settings considered. Each experiment is based on 500 replications of data generated from the NAR(q_1, q_2) model $\mathbb{X}_t = A\mathbb{X}_{t-1} + B\mathbb{W}\mathbb{X}_{t-1} + \mathbb{Y}_{t-1}\gamma + \epsilon_t$ where $\mathbb{Y}_{t-1} := [\mathbb{Y}_{1,(t-1)} \ \mathbb{Y}_{2,(t-1)} \ \dots \ \mathbb{Y}_{p,(t-1)}] \in \mathbb{R}^{N \times p}$ and $\gamma := [\gamma_1 \ \gamma_2 \ \dots \ \gamma_p]^T$. We fix the network size to $N = 100$. Further, the error terms are generated either through a spatial autoregressive (SAR) structure $[\epsilon_t \sim N(0, (I - \rho\Phi)^{-1}(I - \rho\Phi)^{-T})]$ or through a factor structure $\epsilon_t = \Lambda F_t + u_t$ with $u_t \sim N(0, I)$ for factor structure. Finally, the exogenous covariates are generated according to $Y_{ik,(t-1)} \sim N(0, 1)$, for all i, k, t . The specific form of the primary model parameters (A, B, γ) together with W, Φ and Λ are specified in subsequent sections.

4.1. Estimation Accuracy

We focus on the performance of the EGLS estimator. We examine the influence of T, W , the structure of Σ_ϵ , and the lag order $q = q_1 = q_2$ on the RMSE metric. The latter is defined as follows for the three sets of model parameters: self-lags $\|\hat{A} - A\|_F / \|A\|_F$, network lags $\|\hat{B} - B\|_F / \|B\|_F$ and exogenous covariates $\|\hat{\gamma} - \gamma\|_F / \|\gamma\|_F$. We set $A = \text{diag}\{0.1I_{25}, 0.2I_{25}, 0.3I_{25}, 0.4I_{25}\}$, $B = \text{diag}\{0.4I_{25}, 0.3I_{25}, 0.2I_{25}, 0.1I_{25}\}$ while we set $\gamma = (-0.81I_3^T, -0.41I_2^T, 0.41I_2^T, 0.81I_3^T)^T$ in Tables 4–8 in Appendix G, supplementary materials. In Table 9 of Appendix G, we fix lag order $q_1 = q_2 = 2, A_1 = 0.3I, A_2 = 0.3I, B_1 = 0.15I, B_2 = 0.15I$ and $\gamma = 0.5\mathbf{1}_{10}$. The experiment is replicated 500 times, and we calculate the average over 500 estimates. For convenience, we set the weight matrix W to be a row-normalized banded matrix with different bandwidths. Due to space considerations, we show in [Figure 2](#) the results for the influence of the sample size T , and present all Tables assessing the influence of the factors (e.g., $T, N, W, (q_1, q_2)$), the structure of Σ_ϵ in Appendix G, but provide a summary of the results in the sequel.

It can be seen that the accuracy of all model parameters -self-lags, network lags and regression coefficients of the exogenous covariates- increases as the sample size increases.

Next, a summary of the results of the impact of the other factors corresponding to Tables 5–9 (in Appendix G) is presented. The results in Table 5 (Table 8) confirm the robustness of the EGLS estimates over weight matrices W (Φ) with different bandwidths (number of neighboring nodes included), while the RMSE exhibits a slight increase for network effect parameters when the bandwidth increases. Table 6 summarizes the performance of the EGLS estimates for different values of ρ (for the SAR model based Σ_ϵ). It can be seen that the estimates themselves are stable, while the RMSE slightly decreases when ρ increases (from close to zero to close to one). Further, Table 7 shows that the number of factors k (for the factor model based Σ_ϵ) has no significant effect on EGLS estimation, while the RMSE for the network parameters decreases slightly when k increases. Finally, Table 9 shows that EGLS performance improves for larger T , for NAR processes with higher temporal

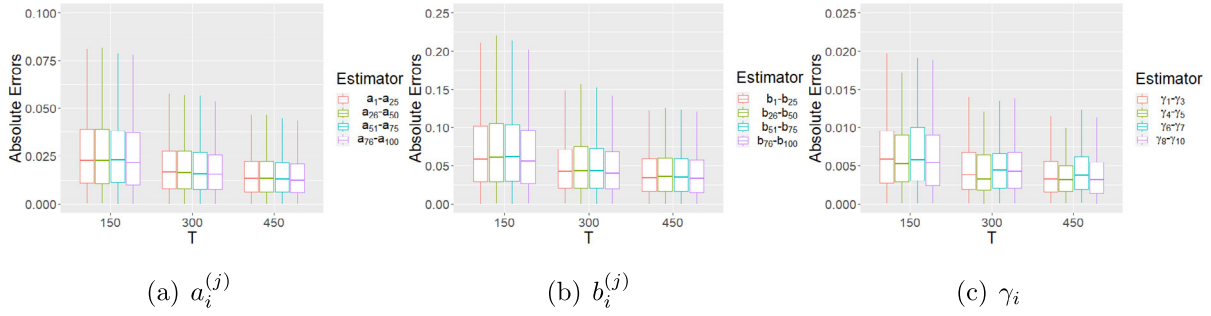


Figure 2. Absolute errors of different coefficients for increasing T under normal errors.

Table 1. $N = 100, q_1 = q_2 = 1, \rho = 0.5$ and W and Φ being banded matrices of width 5.

Estimator	a_i			b_i			γ_i		
	OLS	GLS	EGLS	OLS	GLS	EGLS	OLS	GLS	EGLS
T=150	0.950	0.949	0.949	0.950	0.951	0.950	0.959	0.955	0.955
T=300	0.948	0.948	0.948	0.948	0.949	0.949	0.952	0.951	0.951
T=450	0.948	0.950	0.950	0.950	0.950	0.950	0.951	0.954	0.954

NOTE: Coverage probability of confidence intervals for $T = 150, 300, 450$.

lags. An overall conclusion of the various simulation scenarios is that the performance of the EGLS estimator for the network parameters (B) is more sensitive to changes in W , Φ , and ρ than the autoregressive parameters (A).

4.2. Coverage Probability and Length of Confidence Intervals

We consider $\text{NAR}(1, 1)$ and $\text{NAR}(2, 2)$ processes. For the former, we fix $A = B = 0.4I_N$ and $\gamma = 0.4 \times \mathbf{1}_{10}$ and for the latter $A_1 = B_1 = 0.3I_N$, $A_2 = B_2 = 0.15I_N$ and $\gamma = 0.5 \times \mathbf{1}_{10}$. We explore how different factors influence the coverage probability (CP) and length of confidence intervals (CI) of the various estimators. The results are based on 500 replicates. We set the network size $N = 100$, and ϵ_t follows either a spatial autoregressive model with parameter ρ , or a k factor model. For each model parameter (100 α_i 's, 100 β_i 's, 10 γ_i 's), we calculate its CI and the corresponding CP and length. The 95% CI is calculated using $CI_i = (\hat{\beta}_i - z_{0.975}SE(\hat{\beta}), \hat{\beta}_i + z_{0.975}SE(\hat{\beta}))$, where $SE(\hat{\beta}) = (\sum_t Z_t^T Z_t)^{-1} (\sum_t Z_t^T \Sigma_\epsilon Z_t) (\sum_t Z_t^T Z_t)^{-1}$ for the OLS estimator and $SE(\hat{\beta}) = (\sum_t Z_t^T \Sigma_\epsilon^{-1} Z_t)^{-1}$ for the GLS and EGLS estimators.

In Table 1, the influence of the sample size T is investigated. We fix $\epsilon_t = \rho W \epsilon_t + u_t$ with $\rho = 0.5$, W as a banded matrix of width 5 and explore how the CP and length of the CI are influenced by T .

Figure 3 depicts the length of the confidence intervals that decrease for larger sample sizes T for all estimators. Further, the OLS one has longer CIs, since it does not incorporate information about the error covariance matrix.

Table 1 shows the coverage probability of the confidence intervals for the three estimators, for the different model parameters and varying sample sizes T . It can be seen that the coverage probability is basically at the nominal level.

Due to space considerations, the results of the influence of the other factors (e.g., N , W , (q_1, q_2)), are given in Appendix G in Tables 10–14, supplementary materials, but a summary of the main findings follows. Table 10 shows that with lag 2, the length of the confidence intervals decreases for larger sample sizes T , while the coverage probabilities improve (get closer to the nominal 95% level). Table 11 shows that as the bandwidth of W increases, the length of the 95% confidence intervals of the network effect parameters B increases, while the corresponding coverage probabilities are robust. From Table 12, it can be seen that both the length of CIs and the coverage probabilities are robust for both the GLS and EGLS estimates with respect to changes in ρ (except for the network effect parameters B for which the CI length decreases slightly), while the length of CIs for the OLS estimates increases for all model parameters. Table 13 shows that as the number of factors k increases, the length of the 95% confidence intervals for A , B , and γ increases for the OLS estimator, while for the GLS and EGLS estimators, the corresponding length of the CIs decreases. Finally, results summarized in Table 14 confirm that the length of CIs and coverage probabilities are robust with respect to changes in the bandwidth of matrix Φ (for the SAR model for Σ_ϵ). Overall, the coverage probabilities in all simulation scenarios are close to the nominal level (95%), which implies that the estimators are unbiased and the estimated variances are close to the true variances.

4.3. Comparison of Competing Modeling Approaches Based on Simulated Data

The Predictive MSE (PMSE) defined as

$$\text{PMSE} := \frac{1}{N|T_{\text{test}}|} \sum_{t \in T_{\text{test}}} \|\mathbb{X}_t - Z_{t-1} \hat{\beta}\|_F^2. \quad (11)$$

is used as the performance metric to compare the following models: (a) $\text{NAR}(1,1)$, (b) homogeneous $\text{NAR}(1,1)$ with global a and b ($A = aI$ and $B = bI$), (c) GNAR Knight et al. (2020) with global autoregressive coefficient α and (d) GNAR with local autoregressive coefficients. We set $N = 100$, $T = 400$, and the weight matrix W to be a row-normalized banded matrix with bandwidth 1. We consider two mechanisms for data generation. The experiment is replicated 500 times, and we provide the boxplots of the PMSE for different models. The following two data generation mechanisms are considered:

- (A) The data are generated through a $\text{NAR}(1,1)$ model with $(a_1, a_2, \dots, a_{100})$ being a sequence of 100 points evenly

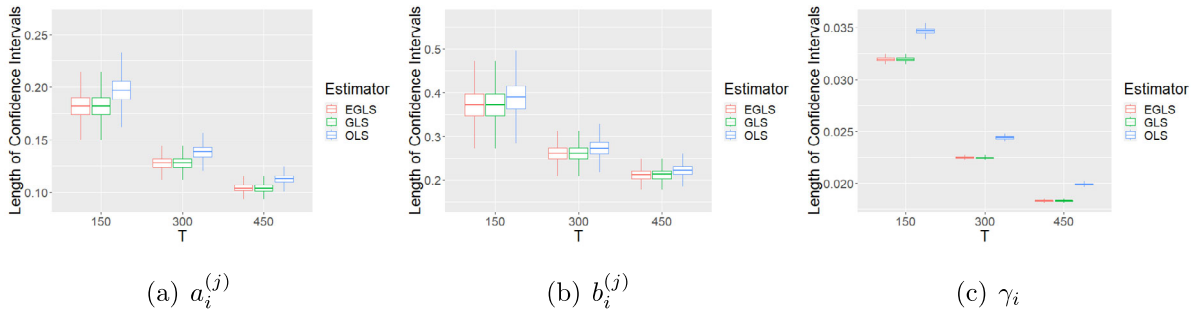


Figure 3. Length of confidence Intervals for different estimators for various values of T .

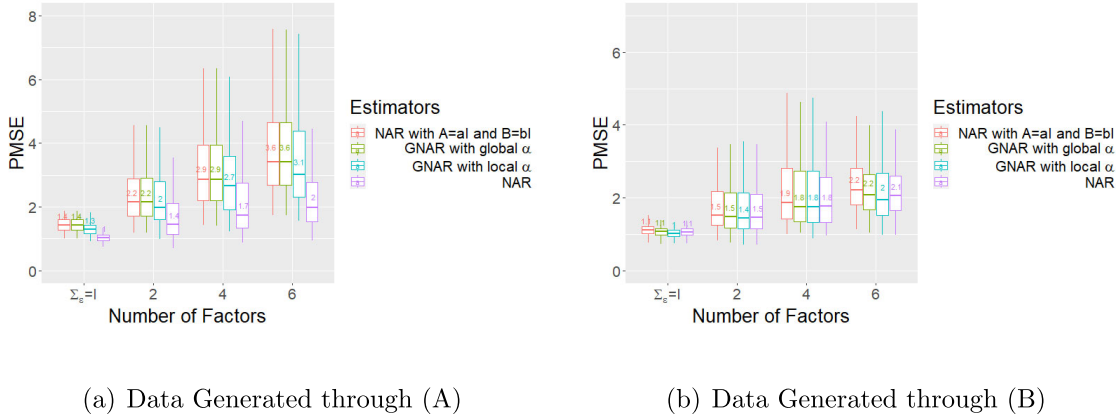


Figure 4. Comparisons of PMSE among competing network autoregressive models ($T=400$).

placed between -0.5 and 0.5 , $(b_1, b_2, \dots, b_{100})$ satisfies that $b_i = (-1)^i(0.8 - |a_i|)$, $i = 1, 2, \dots, 100$. Further, the error terms are generated through (a) $\Sigma_\epsilon = I$, and (b) a factor structure $\epsilon_t = \Lambda F_t + u_t$ with $u_t \sim N(0, I)$ with a different number of factors.

- (B) The data are generated through a GNAR(1,[5]) model with local α autoregressive coefficients, where $(\alpha_1, \alpha_2, \dots, \alpha_{100})$ is a sequence of 100 points evenly placed between -0.3 and 0.3 , and $\beta = [0.3, -0.2, 0.1, -0.05, 0.01]$. Further, the error terms are generated through (a) $\Sigma_\epsilon = I$, and (b) a factor structure $\epsilon_t = \Lambda F_t + u_t$ with $u_t \sim N(0, I)$ with a different number of factors.

It can be seen that when the data are generated from an heterogeneous NAR model (mechanism (A)), the corresponding NAR specification outperforms in terms of PMSE competing specifications for both sample sizes, as seen in the (a) panels of Figures 4 and 5. For data generated through mechanism (B), the heterogeneous NAR model essentially matches the performance of the GNAR model that was used to generate the data. The result suggests that the added flexibility of node specific network effects allows the NAR model to perform well in terms of prediction even in settings wherein the data are generated from a different mechanism.

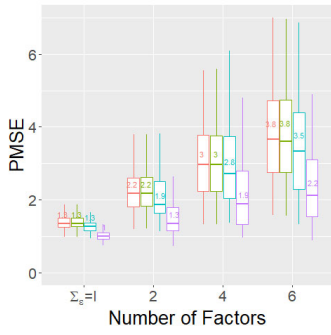
4.4. Impact of the Error Covariance Structure Estimation Method on EGLS

Next, we assess the performance of the EGLS estimator based on different methods to estimate the error covariance matrix. We

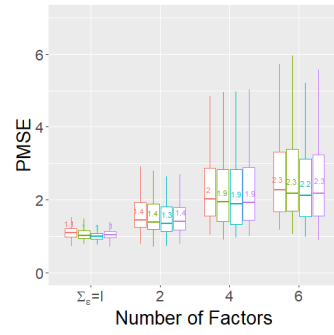
select a setting with $N = 95$ and $T = 100$, that highlights the issue and is also relevant in applications (see data description in Section 5.) The data are generated through an NAR(1,1) model with different network autoregressive effects $a_i^{(1)}$ and a single network effect $b^{(1)}$, so that the model specification is the same as that of the GNAR model. Specifically, $(a_1^{(1)}, a_2^{(1)}, \dots, a_{95}^{(1)})$ is a sequence of 95 values evenly placed between -0.5 and 0.5 , and $b^{(1)} = -0.45$. Further, the error terms are generated through a factor structure $\epsilon_t = \Lambda F_t + u_t$ with $u_t \sim N(0, I)$ with five factors.

The EGLS estimator is employed with the following choices for estimating the error covariance matrix: (i) estimate $\hat{\Sigma}_\epsilon$ from the factor model; (ii) estimate $\hat{\Sigma}_\epsilon$ from the least squares residuals; and (iii) estimate only the *diagonal* elements of $\hat{\Sigma}_\epsilon$ from the least squares residuals.

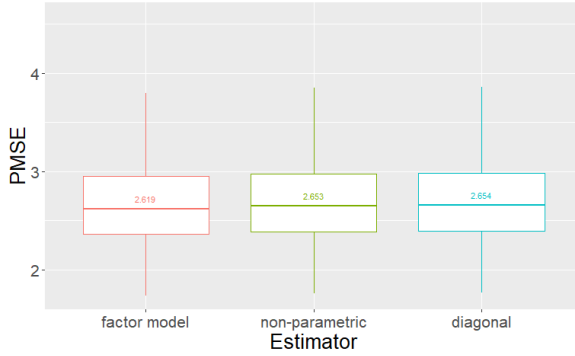
Figure 6 depicts boxplots for the PMSE of each node (left panel), as well as for the length of the confidence intervals (right panel) for the three estimation methods of the covariance matrix. As expected, the PMSEs are not impacted, since the EGLS estimator is consistent for all three methods. On the other hand, the diagonal-based covariance estimator exhibits the largest CI lengths, while its coverage probability is 94.7% (very close to the nominal level), while the nonparametric-based estimator (option (ii)) the smallest ones by a large margin, with a coverage probability 31% (thus, severing underestimating the uncertainty of the model parameters). Finally, the factor-based estimated covariance matrix has a coverage probability of 95%, as expected. These discrepancies become smaller when T increases; for example, for the same setting with $N = 95$ and



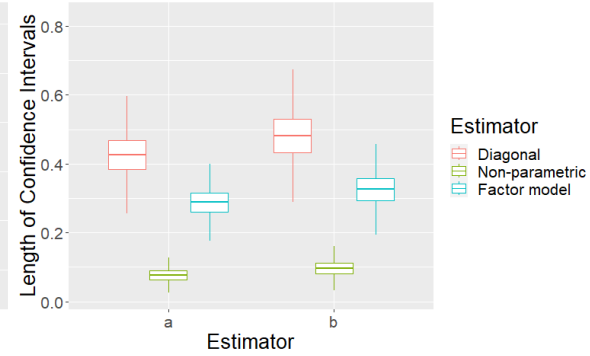
(a) Data Generated through (A)



(b) Data Generated through (B)

Figure 5. Comparisons of PMSE for competing network autoregressive models ($T = 150$).

(a) PMSE



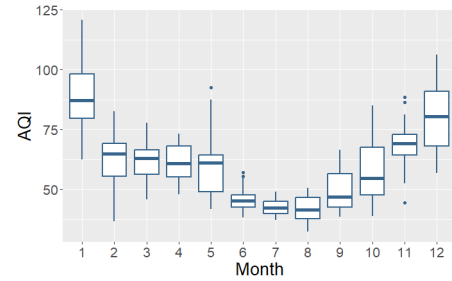
(b) CI

Figure 6. The impact of the error covariance structure estimate on the EGLS estimator.

$T = 400$, the coverage probability of the diagonal-based EGLS is 94.9%, while that of the nonparametric-based estimate improves to 88%, but still falls behind the nominal coverage level.

5. Application to Air Quality Index Data

We employ the proposed $\text{NAR}(q_1, q_2)$ model to analyze Air Quality Index (AQI) data together with relevant weather condition covariates, collected from $N = 319$ stations across China for the period from March 20th, 2019 to March 19th, 2020, for a total of $T = 366$ observations. The AQI data are obtained from the China National Environmental Monitoring Centre, while the weather covariates are from the National Centers for Environmental Information.¹ Boxplots of the AQI for each month across all stations are depicted in Figure 7, and the locations of the stations (left panel), and the average AQI for each station across all observations (right panel) are depicted in Figure 8. We consider the log-transformed AQI as the response variable. It can be seen from the middle panel of Figure 8 that the average AQI reaches its peak in winter, while the pollution level is relatively low in summer; hence, we fit *separate* models for each season. Exogenous covariates $\mathbb{Y}_t$ included in the $\text{NAR}(q_1, q_2)$ model include air temperature, relative humidity, wind speed rate and sky condition total coverage.

**Figure 7.** Boxplot of Monthly AQI values.

Note also that the right panel of Figure 8 indicates substantial heterogeneity, with the north-northwest regions of the country exhibiting higher AQI levels.

Given a large number of stations (319) and limited sample size for each season, the autoregressive a_i and network lag b_i coefficients, together with those of the external covariates were obtained based on regularized estimators. The tuning parameters λ_1 and λ_2 for the ridge regression is determined by generalized cross-validation Golub, Heath, and Wahba

$$(1979) \text{ by minimizing: } \text{GCV}(\lambda) := \frac{\frac{1}{T} \sum_{t=1}^T \|\mathbb{X}_t - \hat{\mathbb{X}}_t\|_F^2}{\left\{ \frac{1}{T} \sum_{t=1}^T \text{Trace}(I - H_t(\lambda)) \right\}^2},$$

where $H_t(\lambda) := \mathbb{Z}_{t-1}(\sum_{j=1}^T \mathbb{Z}_{j-1}^T \mathbb{Z}_{j-1} + TM)^{-1} \mathbb{Z}_{t-1}^T$ and $M := \text{diag}\{\lambda_1 I_{319}, \lambda_2 I_{319}, \mathbf{0}_4\}$.

¹ <https://www.ncdc.noaa.gov/>

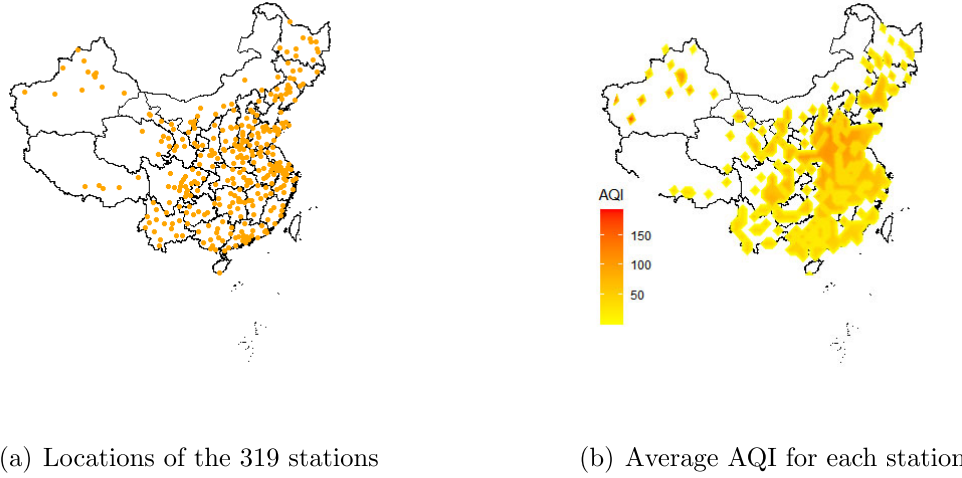


Figure 8. Spatial distribution of AQI monitoring stations and their average values.

The model is fitted with both a SAR covariance structure and a factor model one. To construct the network for the SAR version, both W and Φ correspond to a row-normalized adjacent matrix obtained as follows: let D_{ij} be the spatial distance between two stations and σ^2 be the variance of all distances, then the ij th element of Φ is defined as $\phi_{ij} := \frac{D_{ij}^{-1}}{\sum_j D_{ij}^{-1}}$ if $i \neq j$ and 0 otherwise. Recall that Φ aims to capture any additional spatial dependence not reflected in the structure of the NAR model. Further, the ij th element of W is defined as $w_{ij} := \frac{D_{ij}^{-1}}{\sum_j D_{ij}^{-1}}$, if $i \neq j$ and $D_{ij} \leq 500$ km and 0 otherwise.

Based on the following BIC criterion

$$\text{BIC}(q) = \log |\hat{\Sigma}_\epsilon| + \frac{(Nq_1 + Nq_2 + p) \log T}{T}, \quad (12)$$

an NAR(1,1) model was selected for each season. A plot of the partial autocorrelation function for the AQI variable (not shown) corroborates this choice for the temporal autoregressive and network lags. To select the number of factors in the corresponding model, we employed the following information criterion: $\text{IC}(k) = \log(S(k)) + \frac{k(N+T-k)}{NT} \log(NT)$. It resulted in selecting a single factor ($k = 1$) for each season's NAR model.

The results are depicted in Figures 9 and 10 and tabulated in Tables 18–21 (in Appendix H), respectively. The NAR model estimates show great variation amongst different regions and different seasons. Tables 18–21 show that all covariates employed are statistically significant. For relative humidity, its magnitude remains constant across the four seasons and its impact is positive in reducing air pollution (the negative sign of the regression coefficient). Analogously, the impact of the wind speed is fairly similar across the four seasons and positive for air quality. The impact of the air temperature is positive and similar during the Summer and Fall seasons; further, it exhibits a bigger positive impact in Winter and a small negative impact in Spring.

To aid interpretation, Table 22 (in the supplementary materials) presents the average (over monitoring stations) autoregressive and network lag coefficients for all the provinces and selected big cities. Air quality in China is impacted by multiple

factors and exhibits large variability over regions and seasons. It can be seen that the regions with the largest autoregressive coefficients are in southwest China, while the regions with the highest network coefficients are in southeast China, along the coastal area. The topography of the regions (island or plateau) may be related to the presence of such large autoregressive coefficients (e.g., Yunnan–Guizhou Plateau, Tibetan Plateau and Hainan Island). In contrast to regions with large autoregressive coefficients, regions with the largest network coefficients are coastal areas. During Winter months, northern regions tend to have larger autoregressive coefficients and smaller network coefficients compared to other seasons, and temperature inversion may be the cause. During an inversion, warmer air is held above cooler air, so air pollution is trapped by it, which makes air pollution hard to diffuse.

The results are broadly in accordance with findings in recent studies that have investigated spatial and temporal variations of air pollutants in China. (Wang et al. 2014). In North East China, coal-based industries such as iron and steel manufacturing and coal-fired power plants are key drivers for increased AQI levels the main causes of air pollution. In the Northern China Plain, the network effect is high compared to other regions. Emissions from fossil fuel combustion and biomass burning for home heating in the winter months result in a high concentration of air pollutants. Surrounded by mountains, particulates brought by south easterly winds may accumulate in the region, whereas cold fronts from the north together with their winds are weakened by the mountains and hence result in increased pollution levels. Further, sandstorms from the deserts in the north also contribute to the network effects observed for Northern China Plain stations. A number of studies have discussed air pollution patterns in this region and potential drivers (Wang et al. 2017; Xiao et al. 2020). For the Yangtze River Delta in the east coastal area, the network coefficient is large compared to the autoregressive coefficient. Particulates brought by cold fronts from the Mongolian Plateau in the north also contribute to the network effect. Finally, it is worth noting that eastern regions exhibit on average larger network coefficients, while western regions have higher autoregressive coefficients.

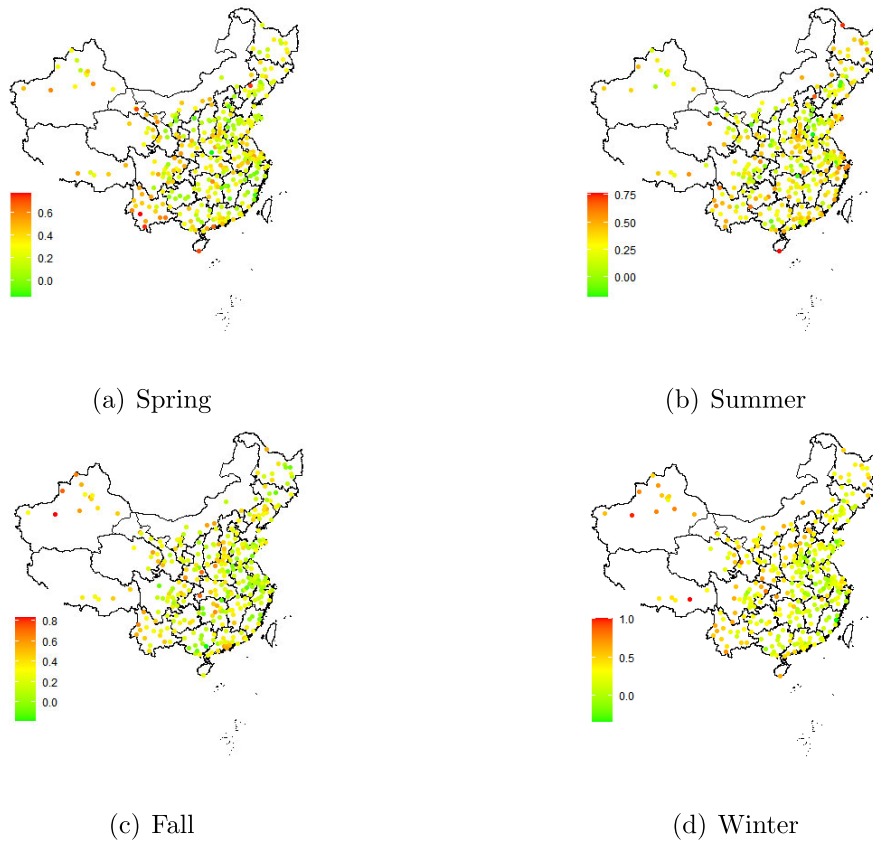


Figure 9. Autoregressive coefficients a_i .

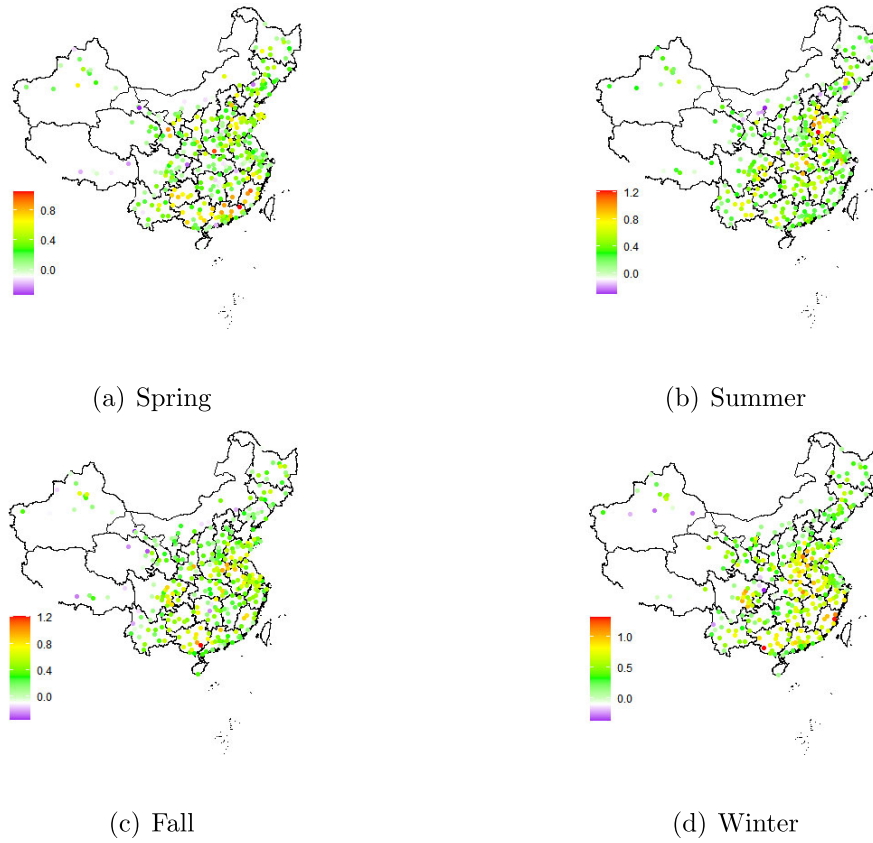


Figure 10. Network lag coefficients b_l .

Table 2. PMSE for different estimators across different seasons.

	Spring	Summer	Fall	Winter
OLS	0.0656	0.0665	0.0999	0.0889
EGLS w/ spatial covariance	0.0654	0.0656	0.0990	0.0880
EGLS w/ factor structure	0.0614	0.0660	0.1030	0.0884
NAR with $A = aI$ and $B = bI$	0.0658	0.0661	0.1019	0.0925
VAR(1) with ridge	0.0927	0.0835	0.1466	0.1180
VAR(1) with lasso	0.0727	0.0791	0.1239	0.1068
AR(1)	0.0684	0.0693	0.1136	0.1054
GNAR(1, s)	0.0733	0.0730	0.1131	0.0990
GNARX(1, s)	0.0687	0.0673	0.1038	0.0932

Next, we consider how the estimated NAR(1,1) model performs in terms of forecasts together with a number of competing models. The parameters of the NAR(1,1) model are estimated by both OLS and EGLS, with SAR and factor covariance structures and a ridge penalty (since $N > T$). For comparison purposes, we also consider a NAR(1,1) model with $A = aI$ and $B = bI$, a GNARX(1, s) model where the number of stages s is determined by BIC, a regularized VAR(1) model with ridge and lasso penalties and finally a simple AR(1) model, applied to each station's data. The evaluation is based on the last 20 days (test data) of each season, which are used to calculate PMSEs for the different models, defined as $PMSE := \frac{1}{N|T_{\text{test}}|} \sum_{t \in T_{\text{test}}} ||\mathbb{X}_t - \mathbb{Z}_{t-1}\hat{\beta}||_F^2$. The results are given in Table 2.

It can be seen that the NAR-based predictions clearly outperform the VAR and AR(1) ones, across all seasons. Further, EGLS for the posited NAR model exhibits better performance than its OLS counterpart and also the predictions of the homogeneous NAR model. Differences are minuscule for Summer, but around 5% in magnitude for the other seasons. Finally, the GNAR/GNAR-X models perform well, but fall behind compared to the NAR specification, probably due to the heterogeneity of the network effects that the latter model captures more effectively.

6. Conclusion and Discussion

The article presented a general flexible framework for NAR processes that can accommodate node-specific network effects in a growing size network (number of nodes *exceeding* the number of available time observations), exogenous covariates, errors that can exhibit heavier than Gaussian tails and a variety of error covariance matrices. It can also be regarded as a VAR model with a specific structure in the transition matrices that reduces the number of parameters, and also aids in interpretability. The latter connection enables us to provide a significantly weaker stability condition compared to those available in the literature for significantly simpler models, thus, expanding the applicability of the framework, as also illustrated in the real data applications. The parameter reduction requires a priori knowledge of the weight matrices. However, the results established show that the model parameters estimates are robust to a certain degree of misspecification of these matrices (see Appendix E in the supplementary materials). The flexibility of the proposed NAR model, which allows for a *node specific* heterogeneous network effect proves useful in applications as the results for both the air quality data and the wind speed data (see Appendix F.1 in the supplementary materials) show.

To that end, how to “design” the weight matrices W to optimize performance is a topic of future research. Further, in very high-dimensional settings, the use of sparsity-inducing penalties (such as the lasso and its variants) is of interest, together with inference procedures based on ideas of debiasing the resulting parameter estimates.

Supplementary Materials

The PDF file contains proofs of all Theorems, additional lemmas and other technical material not included in the main file due to space considerations, as well as additional simulation results, tables and figures. The HTML files contain R code to reproduce Tables 4, 1 and part of Table 2, as well as the comparisons in Section F.1.

Acknowledgments

The authors would like to thank the Editor, Associate Editor, and anonymous reviewers for many constructive comments and suggestions.

Funding

The work of GM and AS was supported in part by NSF grant DMS 2124507.

ORCID

George Michailidis  <https://orcid.org/0000-0002-3676-1739>

References

- Armiliotta, M., and Fokianos, K. (2021), “Poisson Network Autoregression,” arXiv preprint arXiv:2104.06296. [2]
- Chen, E. Y., Fan, J., and Zhu, X. (2020), “Community Network Autoregression for High-Dimensional Time Series.” Available at <https://doi.org/10.1016/j.jeconom.2022.10.005>. [1,2,3]
- Golub, G. H., Heath, M., and Wahba, G. (1979), “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” *Technometrics*, 21, 215–223. [8]
- Knight, M., Leeming, K., Nason, G., and Nunes, M. (2020), “Generalized Network Autoregressive Processes and the GNAR Package,” *Journal of Statistical Software*, 96, 1–36. Available at <https://www.jstatsoft.org/index.php/jss/article/view/v096i05>. [1,2,3,6]
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Berlin: Springer. [3]
- Nason, G. P., and Wei, J. L. (2021), “Quantifying the Economic Response to COVID-19 Mitigations and Death Rates via Forecasting Purchasing Managers’ Indices using Generalised Network Autoregressive Models with Exogenous Variables,” arXiv preprint arXiv:2107.07605. [2,3]
- Wang, Y., Liu, H., Mao, G., Zuo, J., and Ma, J. (2017), “Inter-regional and Sectoral Linkage Analysis of Air Pollution in Beijing–Tianjin–Hebei (Jing-Jin-Ji) Urban Agglomeration of China,” *Journal of Cleaner Production*, 165, 1436–1444. [9]
- Wang, Y., Ying, Q., Hu, J., and Zhang, H. (2014), “Spatial and Temporal Variations of Six Criteria Air Pollutants in 31 Provincial Capital Cities in China during 2013–2014,” *Environment international*, 73, 413–422. [9]
- Xiao, C., Chang, M., Guo, P., Gu, M., and Li, Y. (2020), “Analysis of Air Quality Characteristics of Beijing–Tianjin–Hebei and Its Surrounding Air Pollution Transport Channel Cities in China,” *Journal of Environmental Sciences*, 87, 213–227. [9]
- Zellner, A. (1962), “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,” *Journal of the American statistical Association*, 57, 348–368. [2]
- Zhu, X., and Pan, R. (2018), “Grouped Network Vector Autoregression,” *Statistica Sinica*, 30, 1437–1462. [1,2,3]
- Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017), “Network Vector Autoregression,” *The Annals of Statistics*, 45, 1096–1123. [1,2,3]