

8-2017

Parsimonious Space-Time Temperature Series Modeling

Yanbo Xia

Clemson University, yanbox@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation

Xia, Yanbo, "Parsimonious Space-Time Temperature Series Modeling" (2017). *All Dissertations*. 2009.
https://tigerprints.clemson.edu/all_dissertations/2009

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

PARSIMONIOUS SPACE-TIME TEMPERATURE SERIES MODELING

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

by
Yanbo Xia
August 2017

Accepted by:
Dr. Robert Lund, Committee Chair
Dr. Peter Kiessler
Dr. Colin Gallagher
Dr. Xiaoqian Sun

Abstract

Climatological time series are often periodically and spatially correlated. High dimensionality issues arise when modeling periodically and spatially correlated time series data – often, even simple multivariate models have more parameters than data points. This dissertation develops parsimonious methods for fitting periodically and spatially correlated multivariate time series data. Parsimonious VAR (vector autoregressive) and PVAR (periodic VAR) models are pursued here. The layered procedure introduced by Lund et al. (1995) is adopted as a basic scheme, which removes periodic correlation from the data in the first layer, and fits a stationary VAR model in the second layer. The method is applied to a daily maximum temperature data set of seven cities in southeastern U.S.. In addition, a portmanteau test is proposed for diagnosing serial correlations in periodic multivariate residuals. The performance of the test is examined in simulated data.

Table of Contents

Title Page	i
Abstract	ii
List of Figures	iv
1 Introduction	1
1.1 Literature Review	2
1.2 Periodic Vector Series	8
1.3 Data	9
2 Methods	14
2.1 A Layered Method	14
2.2 Parsimonious PVAR models	16
2.3 Model Evaluation	24
3 Results	31
3.1 Simulations: Periodic Multivariate Portmanteau Test	31
3.2 Simulations: Parsimonious PVAR Model	34
3.3 Fitting the seven-city Data	36
4 Conclusions and Discussion	39
References	41

List of Figures

1.1	Sample summaries by day of year: Atlanta	10
1.2	Average squared coherences for the Atlanta series	11
1.3	Sample cross correlations between Atlanta and Charlotte	11
1.4	Sample auto-correlations and cross-correlations	12
1.5	Sample cross correlations vs distance between stations	13
2.1	Components of $\hat{\Gamma}_\nu(0)^{-\frac{1}{2}}$ by day of year	16
2.2	Time warping function and its effect	23
3.1	Empirical Type I errors vs series lengths, $d = 3, T = 4$	33
3.2	Empirical Type I error plots for models with different dimensions	33
3.3	Power of periodic multivariate portmanteau test on PVMA series	34
3.4	Empirical Type I error rates testing PVAR residuals	35
3.5	Empirical Type I error rates testing residuals of parsimonious PVAR model	35
3.6	Coherence test of standardization procedures	36
3.7	VAR fit	37
3.8	"AIC, BIC, log likelihood, and average p -values for sparse VAR fits"	38

Chapter 1

Introduction

Time series modeling in the environmental settings often involves two aspects. On one hand, climatological measurements such as air temperature, humidity, precipitation, and wind speed naturally show periodic variations over time. On the other hand, the continuity of geological characteristics and fluidity of the atmosphere often injects cross-correlation between climatological measurements in spatially adjacent areas, making it beneficial to include multiple time series in the same model. Hence, it is often necessary to consider both periodicity and multivariate aspects.

When studying such series, many methods from periodic time series theory and multivariate time series can be used. However, one should also expect classical difficulties. Two of the most prominent difficulties are the high dimensionality of multivariate models and the lack of data. Multivariate periodic time series usually have an enormous number of parameters, which in return causes problems such as poor asymptotic behavior and poor interpretability of parameters. As explained by Franses and Paap (2004), periodic time series data typically only contain a few dozen data points in each season, which is problematic given the large number of model parameters.

In this chapter, we first study the literature on periodic and multivariate time series. Then we briefly introduce periodic multivariate time series models and a multivariate climatological data set with periodic features.

1.1 Literature Review

Several types of theories are worth studying before considering methods for periodic vector time series. The theories of periodic time series and multivariate time series are prominent in our topic. Parsimonious periodic time series models and sparse vector time series models are useful when dealing with the high dimensionality of periodic multivariate time series data. Additionally, some methods from spatial statistics will be used in our methods.

1.1.1 Periodic Time Series

To date, periodic time series has been studied in detail (Lund et al. (1995), Lund and Basawa (2000), Lund et al. (2006), Franses and Paap (2002), Franses and Paap (2004), Fuller (2009)). A univariate time series $\{X_t; t \in \mathbb{Z}\}$ is called periodically correlated (PC) with period T if $\mathbb{E}(X_{t+T}) = \mathbb{E}(X_t)$ and $\text{cov}(X_{t+T}, X_{s+T}) = \text{cov}(X_t, X_s)$ for all $t, s \in \mathbb{Z}$. An alternative seasonal notation $\{X_{nT+\nu}; n \in \mathbb{Z}, \nu = 1, 2, \dots, T\}$ is often used to emphasize periodic features of a PC series, with $nT + \nu$ denoting season ν of the n th period. We further define the seasonal mean, seasonal autocovariances, and seasonal autocorrelations respectively as follows:

$$\begin{aligned}\mu_\nu &= \mathbb{E}[X_{nT+\nu}], \\ \gamma_\nu(h) &= \mathbb{E}[(X_{nT+\nu} - \mu_\nu)(X_{nT+\nu-h} - \mu_{\nu-h})], \quad h \in \mathbb{Z}, \\ \rho_\nu(h) &= \frac{\gamma_\nu(h)}{\sqrt{\gamma_\nu(0)\gamma_{\nu-h}(0)}} \quad h \in \mathbb{Z},\end{aligned}$$

where $\nu \in \{1, 2, \dots, T\}$ and all quantities are interpreted periodically with period T . In general, three types of periodic changes are considered in such series, including changes in the mean ($\{\mu_\nu\}_{\nu=1}^T$), variance ($\{\gamma_\nu(0)\}_{\nu=1}^T$), and autocorrelations ($\{\rho_\nu(h)\}_{\nu=1}^T$ for $h \neq 0$). When the seasonal autocorrelation $\rho_\nu(h)$ does not depend on ν , the time series is merely a stationary time series linearly transformed with deterministic periodic parameters. Such series are often referred to as seasonally adjusted series (Lund et al., 1995) and can be written in the form

$$X_{nT+\nu} = \mu_\nu + \sigma_\nu Y_{nT+\nu}$$

where $\{Y_{nT+\nu}\}$ is a stationary time series. Because a seasonally adjusted series is easier to handle and usually requires a model with a much smaller number of parameters, it is worth checking this model reduction before considering models featuring periodic autocorrelations.

A hypothesis test is proposed by Lund et al. (1995) to detect periodicities in autocorrelations. This test uses an average squared coherence statistic for each lag h constructed from the discrete Fourier transform of the series at a set of Fourier frequencies. If periodic changes in autocorrelations are confirmed present, a periodic time series model should be considered for the data. Periodic autoregressive moving-average (PARMA) models have been studied in the literature (Vecchia, 1985; Lund and Basawa, 2000; Franses and Paap, 2004; Anderson et al., 2007). Methods for parameter estimation and recursive predictions have been developed. Maximum likelihood estimates can be computed via numerical methods, and PAR models can be estimated by solving the periodic Yule-Walker moment equations. PARMA models can also be concatenated into stationary vector ARMA models (Vecchia, 1985), thus allowing vector time series methods to be applied. For example, the following PARMA model with AR order $p < T$ and MA order $q < T$,

$$X_{nT+\nu} = \sum_{i=1}^p \phi_\nu(i) X_{nT+\nu-i} + Z_{nT+\nu} + \sum_{j=1}^q \theta_\nu(j) Z_{nT+\nu-j},$$

can be concatenated into

$$\mathbf{\Phi}_0 \mathbf{X}_n = \mathbf{\Phi}_1 \mathbf{X}_{n-1} + \mathbf{\Theta}_0 \mathbf{Z}_n + \mathbf{\Theta}_1 \mathbf{Z}_{n-1}, \quad (1.1)$$

where $\mathbf{X}'_n = \begin{pmatrix} X_{(n+1)T} & X_{(n+1)T-1} & \cdots & X_{nT+1} \end{pmatrix}$, $\mathbf{Z}'_n = \begin{pmatrix} Z_{(n+1)T} & Z_{(n+1)T-1} & \cdots & Z_{nT+1} \end{pmatrix}$, and

$$\mathbf{\Phi}_0 = \begin{pmatrix} 1 & -\phi_T(1) & -\phi_T(2) & \cdots & 0 \\ 0 & 1 & -\phi_{T-1}(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_2(2) & \phi_2(3) & \cdots & 0 \\ \phi_1(1) & \phi_1(2) & \cdots & 0 \end{pmatrix},$$

$$\mathbf{\Theta}_0 = \begin{pmatrix} 1 & \theta_T(1) & \theta_T(2) & \cdots & 0 \\ 0 & 1 & \theta_{T-1}(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Theta}_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_2(2) & \theta_2(3) & \cdots & 0 \\ \theta_1(1) & \theta_1(2) & \cdots & 0 \end{pmatrix}.$$

Since Φ_0 is invertible, (1.1) is equivalent to the vector autoregressive moving-average VARMA(1,1) model

$$\mathbf{X}_n = \Phi_0^{-1} \Phi_1 \mathbf{X}_{n-1} + \Phi_0^{-1} \Theta_0 \mathbf{Z}_n + \Phi_0^{-1} \Theta_1 \mathbf{Z}_{n-1}.$$

However, general VARMA methods usually have a large number of parameters. For PAR model parsimony, Lund et al. (1995) first introduced a layered method that uses a PAR model of low order to eliminate periodic changes in correlations, avoiding over-parameterization. Another technique by Lund et al. (1995) and Lund et al. (2006) uses a short Fourier expansion of parameters to parsimonize the model.

1.1.2 Multivariate Time Series

On the other side, an extensive theory for multivariate (or vector) time series has been compiled (Lütkepohl, 1993, 2005; Reinsel, 2003; Porcu et al., 2012; Tsay, 2013). Most of the emphasis has been placed on stationary vector series, introducing models with autoregressive (AR) and/or moving-average (MA) structures; non-stationary vector series and periodic vector series are discussed in Lütkepohl (1993).

Consider a d -dimensional vector series $\{\mathbf{X}_t \in \mathbb{R}^d; t \in \mathbb{Z}\}$. Let $\boldsymbol{\mu}_t = \mathbb{E}(\mathbf{X}_t)$ be the mean and $\boldsymbol{\Gamma}(t, s) = \mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu}_t)(\mathbf{X}_s - \boldsymbol{\mu}_s)']$ the matrix-valued autocovariance of the series. The definition of (weak) stationarity is that the first and the second moments (mean and variance) are constants, i.e., there exists $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Gamma}_0(\cdot) : \mathbb{Z} \rightarrow \mathbb{R}^d$, such that $\boldsymbol{\mu}_t = \boldsymbol{\mu}$ and $\boldsymbol{\Gamma}(t, t-h) = \boldsymbol{\Gamma}_0(h)$, $\forall t, h \in \mathbb{Z}$. For simplicity, we assume $\{\mathbf{X}_t\}$ to be mean-zero in this paper unless otherwise stated, i.e. $\boldsymbol{\mu}_t = \mathbf{0}$, $\forall t \in \mathbb{Z}$.

The series $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$ is said to be a vector autoregressive (VAR) series if it is a solution to the linear difference equation

$$\mathbf{X}_t - \sum_{k=1}^p \boldsymbol{\Phi}(k) \mathbf{X}_{t-k} = \mathbf{Z}_t, \quad \nu = 1, 2, \dots, T, \quad (1.2)$$

where $\{\mathbf{Z}_t\}_{t=-\infty}^{\infty}$ is a white noise (WN) series with mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}$. Classic methods for estimating univariate AR parameters have been extended to VAR models, including the Yule-Walker method, least squares, and maximum likelihood (Lütkepohl, 1993).

1.1.3 Sparse VAR models

Davis et al. (2016) introduced a two-stage method for fitting VAR models with some parameters constrained as zero. Consider the following d -dimensional VAR model

$$\mathbf{Y}_t = \sum_{k=1}^p \tilde{\Phi}_k \mathbf{Y}_{t-k} + \tilde{\mathbf{Z}}_t, \quad (1.3)$$

with parameter constraint $\text{vec}(\tilde{\Phi}_1 | \cdots | \tilde{\Phi}_p) = \mathbf{R}\beta$, where β is an m -dimensional vector of free parameters, and \mathbf{R} is a full rank $d^2 p \times m$ matrix. Since the only purpose of this constraint is to set some of the parameters to zero, \mathbf{R} should only have entries equal to 0 or 1. For example, for a 2-dimensional VAR(2) model with VAR parameters of the form

$$\tilde{\Phi}_1 = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}, \quad \tilde{\Phi}_2 = \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$$

the constraint matrix \mathbf{R} and the parameter β should be

$$\mathbf{R}' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} a \\ b \end{pmatrix}.$$

To derive parameter estimates of the constrained VAR model, the two stages of the method in Davis et al. (2016) use a GLS method introduced by Lütkepohl (1993). The difference between the two stages is in determining which parameters are set to zero. In the first stage, the Partial Spectral Coherence (PSC) of each pair (i, j) of components of the vector series is computed, and the pairs are sorted by the supremum of their squared PSC. The M pairs with the largest supremum squared PSC's are chosen, and VAR coefficients corresponding to all other pairs are set to zero. The constrained model is estimated for each $M \in \{1, 2, \dots, d(d-1)/2\}$, and an optimal number of pairs, M^* , is determined according to the BIC criterion. In the second stage, more parameters are picked from the M^* non-zero pairs of the last stage, and set to zero. The parameters are prioritized according to the magnitude of their corresponding t -statistics from the last stage. The remaining non-zero parameters are estimated using the method from stage one.

1.1.4 Spatial Statistics

A major part of our method relies on constructing a feasible autocovariance function of a vector time series. Consider a stationary mean-zero vector time series $\{\mathbf{X}_t\}_{t=-\infty}^{+\infty}$. The autocovariance of this series is denoted by $\mathbf{\Gamma}(h) = \mathbb{E}[\mathbf{X}_t \mathbf{X}'_{t-h}] = [\gamma_{i,j}(h)]_{i,j=1}^d$. By stationarity, $\mathbf{\Gamma}(\cdot)$ has the following properties:

- (i) $\mathbf{\Gamma}(h) = \mathbf{\Gamma}(-h)'$,
- (ii) $|\gamma_{ij}(h)| \leq (\gamma_{ii}(0)\gamma_{jj}(0))^{1/2}$, $i, j \in \{1, \dots, d\}$,
- (iii) $\gamma_{ii}(\cdot)$ is an autocovariance function, $i \in \{1, \dots, m\}$, and
- (iv) $\sum_{j,k=1}^n \mathbf{a}'_j \mathbf{\Gamma}(j-k) \mathbf{a}_k \geq 0$ for all $n \in \{1, 2, \dots\}$ and $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$.

The non-negative definiteness condition (iv) is equivalent to

- (iv') The following matrix is non-negative definite for all $n \in \{1, 2, \dots\}$:

$$\mathbf{G}_n = \begin{pmatrix} \mathbf{\Gamma}(0) & \mathbf{\Gamma}(1) & \mathbf{\Gamma}(2) & \cdots & \mathbf{\Gamma}(n-1) \\ \mathbf{\Gamma}(-1) & \mathbf{\Gamma}(0) & \mathbf{\Gamma}(1) & \cdots & \mathbf{\Gamma}(n-1) \\ \mathbf{\Gamma}(-2) & \mathbf{\Gamma}(-1) & \mathbf{\Gamma}(0) & \cdots & \mathbf{\Gamma}(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}(1-n) & \mathbf{\Gamma}(2-n) & \mathbf{\Gamma}(3-n) & \cdots & \mathbf{\Gamma}(0) \end{pmatrix}. \quad (1.4)$$

Conditions (i)–(iv) are sufficient to guarantee the existence of a stationary Gaussian series with covariance function $\mathbf{\Gamma}(\cdot)$ (but they are not a necessary condition for stationarity, see Lütkepohl 1993). The key property in building a parsimonious model in space is that these conditions are naturally satisfied if $\{\gamma_{ij}(h); h \in \mathbb{Z}\}$ are from a covariogram of a stationary spatial series.

Consider a univariate stationary zero-mean spatial series (also known as a random field) $\{\tilde{X}(\mathbf{w}); \mathbf{w} \in \mathbb{R}^k\}$ on a k -dimensional space. The covariogram of the spatial series is defined by

$$C(\mathbf{h}) = \mathbb{E}[\tilde{X}(\mathbf{w}) \tilde{X}(\mathbf{w} - \mathbf{h})], \quad (1.5)$$

where $\mathbf{h} \in \mathbb{R}^k$ is called the spatial lag. The covariogram is said to be *separable* if there exists covariograms $C_1(\cdot), \dots, C_s(\cdot)$ on the complementary subspaces $\mathbb{R}^{k_1}, \dots, \mathbb{R}^{k_s}$ of \mathbb{R}^k , such that

$$C(\mathbf{h}) \equiv C_1(\mathbf{h}_1)C_2(\mathbf{h}_2) \dots C_s(\mathbf{h}_s)$$

for $\mathbf{h}_1 \in \mathbb{R}^{k_1}, \dots, \mathbf{h}_s \in \mathbb{R}^{k_s}$, and $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_s)$. The covariogram is said to be *isotropic* if it depends only on the magnitude of \mathbf{h} , i.e. if there exists a univariate function $\tilde{C}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$, such that $C(\mathbf{h}) \equiv \tilde{C}(\|\mathbf{h}\|)$. Here $\|\cdot\|$ can be any valid norm on \mathbb{R}^k .

Much statistical research involves finding valid covariograms (Matérn, 2013; Cressie, 1993; Cressie and Wikle, 2015). Several examples (see Cressie, 1993) of isotropic covariograms on \mathbb{R}^k are

$$C(\mathbf{h}) = \sigma^2 \exp(-a^2 \|\mathbf{h}\|^2), \quad \mathbf{h} \in \mathbb{R}^k, \quad (1.6)$$

$$C(\mathbf{h}) = \sigma^2 [1 + \|\mathbf{h}\|^2/b^2]^{-\beta}, \quad \mathbf{h} \in \mathbb{R}^k, \beta > 0, \quad (1.7)$$

$$C(\mathbf{h}) = \sigma^2 (a^2 \|\mathbf{h}\|/2)^\nu 2K_\nu(a^2 \|\mathbf{h}\|)/\Gamma(\nu), \quad \mathbf{h} \in \mathbb{R}^k, \nu > 0, \quad (1.8)$$

where K_ν is the modified Bessel function of the second kind.

By using different norms, it is possible to define a larger family of feasible covariograms. Applying linear transformations on the coordinates can also help construct different covariograms. This is supported by the following lemma from Porcu et al. (2012).

Lemma 1. *If $\{C(\mathbf{h})\}$ is a covariogram of a k -dimensional stationary spatial series, and \mathbf{A} is an $k \times k$ invertible matrix, then $\{C(\mathbf{A}\mathbf{h})\}$ is also a legitimate covariogram.*

Apart from using isotropic covariograms, one dimension reduction tactic involves a separable covariogram of form

$$C(\mathbf{v}, t) = C_1(\mathbf{v})C_2(t), \quad (1.9)$$

for some covariograms $C_1(\cdot)$ on \mathbb{R}^{k-1} and $C_2(\cdot)$ on \mathbb{R}^1 . The separable form for time and space possesses more flexibility than isotropic covariograms, as it allows modeling the effect of time and space differently. However, in section 2.2.2, we will show that this construction imparts a covariance with all components independent of all other component's historical past. As this does not seem to hold in our temperature data's cross correlations, this construction does not satisfy our needs.

1.2 Periodic Vector Series

The emphasis in this dissertation is on periodic multivariate (vector) time series. A d -dimensional series $\{\mathbf{X}_t\}$ is called periodically correlated (PC) with period $T \in \mathbb{Z}^+$ if for any $t, s \in \mathbb{Z}$

$$\begin{aligned}\mathbb{E}(\mathbf{X}_t) &= \mathbb{E}(\mathbf{X}_{t+T}), \quad \text{and} \\ \mathbb{E}(\mathbf{X}_t - \mathbb{E}(\mathbf{X}_t))(\mathbf{X}_s - \mathbb{E}(\mathbf{X}_s))' &= \mathbb{E}(\mathbf{X}_{t+T} - \mathbb{E}(\mathbf{X}_{t+T}))(\mathbf{X}_{s+T} - \mathbb{E}(\mathbf{X}_{s+T}))'.\end{aligned}$$

We denote its seasonal mean and seasonal autocovariance function of season $\nu = 1, 2, \dots, T$ by

$$\begin{aligned}\boldsymbol{\mu}_\nu &= \mathbb{E}(\mathbf{X}_\nu), \quad \text{and} \\ \boldsymbol{\Gamma}_\nu(h) &= \mathbb{E}(\mathbf{X}_\nu - \mathbb{E}(\mathbf{X}_\nu))(\mathbf{X}_{\nu-h} - \mathbb{E}(\mathbf{X}_{\nu-h})) \quad h \in \mathbb{Z}\end{aligned}$$

Again, we assume that $\boldsymbol{\mu}_\nu = \mathbf{0}$ for all $\nu = 1, 2, \dots, T$ unless otherwise noted. One simply subtracts periodic sample means otherwise.

The series is called a p th order periodic vector autoregressive (PVAR) if it obeys

$$\mathbf{X}_{nT+\nu} - \sum_{k=1}^p \boldsymbol{\Phi}_\nu(k) \mathbf{X}_{nT+\nu-k} = \mathbf{Z}_{nT+\nu}, \quad \nu = 1, 2, \dots, T, \quad (1.10)$$

where $\{\mathbf{Z}_{nT+\nu}; n \in \mathbb{Z}, \nu = 1, 2, \dots, T\}$ is periodic white noise: $\mathbb{E}[\mathbf{Z}_t] = \mathbf{0}$, $\mathbb{E}[\mathbf{Z}_t \mathbf{Z}_s'] = 0$ when $t \neq s$, and $\mathbb{E}[\mathbf{Z}_{nT+\nu} \mathbf{Z}_{nT+\nu}'] = \boldsymbol{\Sigma}_\nu$ for each $\nu = 1, 2, \dots, T$ and for all $n \in \mathbb{Z}$. The marginal distribution of $\mathbf{Z}_{nT+\nu}$ may change with ν .

Periodic vector series share many properties with univariate PC series. The concatenating tactic described in section 1.1.1 can also be applied on periodic vector series (see Tsay, 2013). For example, define

$$\vec{\mathbf{X}}_n = (\mathbf{X}'_{nT+T}, \mathbf{X}'_{nT+T-1}, \dots, \mathbf{X}'_{nT+1})'; \quad (1.11)$$

a periodic VAR model can be concatenated into a VAR model:

$$\vec{\mathbf{X}}_n = \tilde{\boldsymbol{\Phi}}_1 \vec{\mathbf{X}}_{n-1} + \tilde{\mathbf{Z}}_n. \quad (1.12)$$

This blocking allows methods for stationary vector series to be applied on periodic vector series, but also has noteworthy disadvantages. First, as the dimension of the vector is multiplied by T ,

this method heavily increases the dimension of the problem, and hence relies on the estimation of a large number of parameters, which could waste computational resources. In fact, a VAR model of dimension dT requires pTd^2 autoregressive parameters, $Td(d+1)/2$ white noise variance parameters, and solving linear systems of dimension pdT . As climate data is often recorded monthly or daily, periods of $T = 12$ or $T = 365$ are often encountered. A first order periodic VAR model for daily data with $d = 7$ stations has 28105 free parameters to estimate, far more than can be easily handled with standard gradient search techniques to optimize a model likelihood, and can even exceed the number of data points. Second, it isolates parameters in different seasons, ignoring the fact that the parameters may vary smoothly in the seasons. Besides, several components in the concatenated VAR parameters (such as $\tilde{\Phi}_1$ in Equation (1.12)) may need to be set to zero, thus requiring sparse VAR methods to be applied, and further complicating the model.

Although techniques for univariate periodic time series as introduced in Lund et al. (2006) and Anderson et al. (2007) are helpful for reducing the number of parameters in the model arising due to time, how to reduce the number of parameters arising in the series due to the spatial component is less clear. Also, there is no guarantee that any parsimonious form is from a legitimate causal model replete with a non-negative definite autocovariance structure. This dissertation will show how to build a periodic vector time series model with parsimonious use of parameters in both spatial and temporal aspects, while guaranteeing a legitimate autocovariance structure.

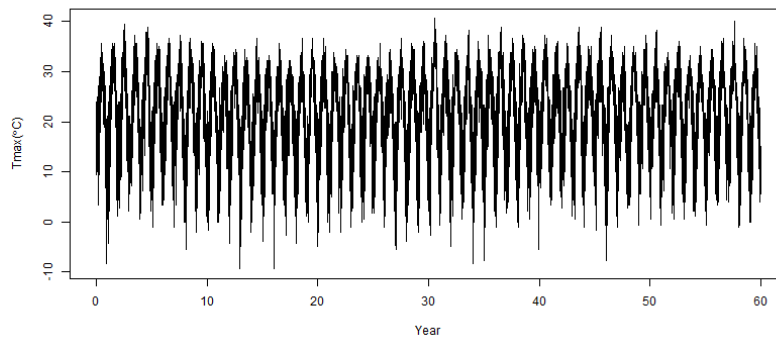
1.3 Data

Weather data is often periodic. Multivariate time series sampled spatially often display interdependence between different components. Motivating our methods are seven daily temperature series from the Southeastern United States: Athens GA, Atlanta GA, Macon GA, Charlotte NC, Greensboro NC, Raleigh NC, and Columbia SC. The data were downloaded from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI) website ¹ and contain daily maximum (high) temperatures from 1/1/1950 to 12/31/2009. To enforce a period of $T = 365$ days, we omit any February 29th observations. This results in a minimal loss of precision.

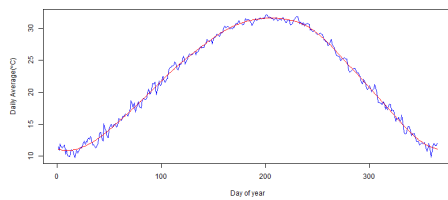
Figure 1.1a shows a plot of the Atlanta series. The daily sample mean, variance, and

¹<http://www.ncdc.noaa.gov/>

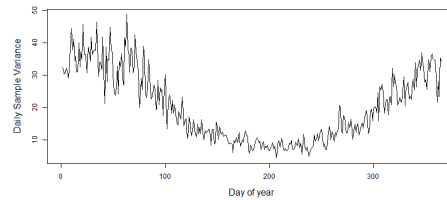
autocorrelations are also shown. Significant structure is evident in these plots. First, there is a clear seasonal mean cycle in the series. Second, temperatures appear less variable in the summer than in the winter. Third, autocorrelation during the Spring appears less than that in the Fall, at both lags one and three. This is reasonable because cold fronts rapidly pass through the area in the Spring, changing the day-to-day weather greatly, while in the Fall, long stretches of dry weather and accompanying high pressure dominate, imparting more persistence (higher autocorrelation) into the series.



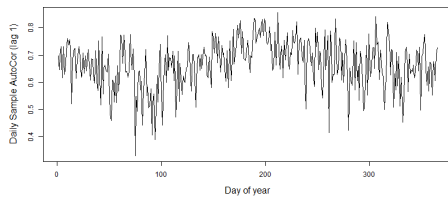
(a) Daily maximum temperatures of Atlanta



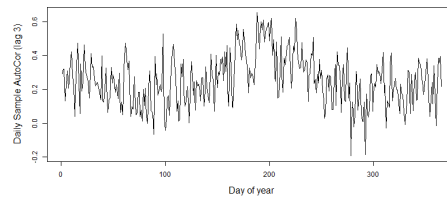
(b) Fitted daily mean: Atlanta



(c) Daily sample variance: Atlanta



(d) Daily sample autocorrelation at lag 1



(e) Daily sample autocorrelation at lag 3

Figure 1.1: Sample summaries by day of year: Atlanta

As mentioned in the last section, it is important to confirm if the autocorrelation of the series is varying periodically. We adopt the coherence test introduced by Lund et al. (1995) to test

each of the seven series. Figure 1.2 is a plot of the average squared coherences of the Atlanta series. A coherence spike is evident at $h = 60$, which indicates existence of periodic changes in the autocorrelations of the Atlanta series according to Lund et al. (1995). Coherence plots of the other six series are similar to this.

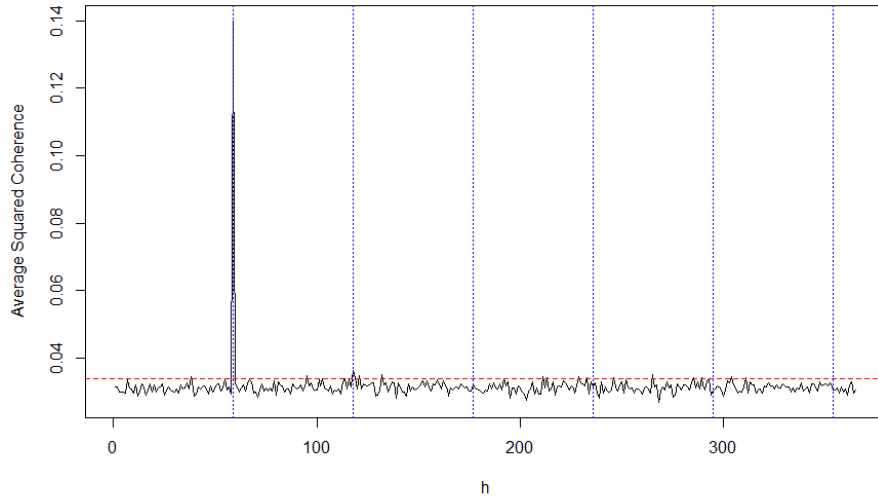


Figure 1.2: Average squared coherences for the Atlanta series

Since we are studying multiple time series, it is also worthwhile to learn if seasonally varying cross-correlations exist between components. Figure 1.3 displays cross autocorrelations between the Atlanta, GA series and the Charlotte, NC series at lag two (the Atlanta series two days ahead of the Charlotte series) and lag minus one. Both of these quantities appear to be seasonally varying, and similar features are seen in other cross-correlations.

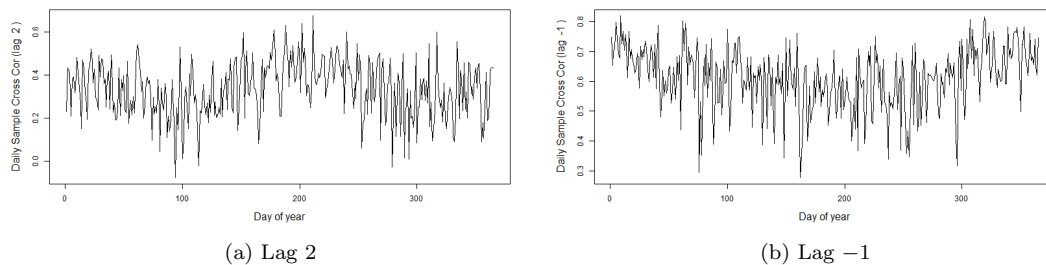


Figure 1.3: Sample cross correlations between Atlanta and Charlotte

In Figure 1.4, we study how cross-correlations between temperatures at two stations are related to the distance between the stations. Recall the introduction of spatial statistics in section 1.1.4. Here, we treat temperature as a spatial series $\{\tilde{X}(\mathbf{v}, t)\}$, where \mathbf{v} is a vector of geological coordinates: (longitude, latitude). Figure 1.4a plots the sample cross-correlations between $\{X(\mathbf{v}_1, t)\}_{t=1}^N$ and $\{X(\mathbf{v}_2, t)\}_{t=1}^N$ against the distance $\|\mathbf{v}_1 - \mathbf{v}_2\|_{Euc}$. Figure 1.4b plots the sample cross-correlations between $\{X(\mathbf{v}_1, t)\}_{t=1}^{N-1}$ and $\{X(\mathbf{v}_2, t+1)\}_{t=1}^{N-1}$ against the distances. Since the longitude and latitude are in degrees, the distance should also be in degrees. The cross-correlations at lag 0 show a roughly linear relation with the spatial lag, while at lag 1, they also decrease roughly linearly as the spatial lag increases. This feature is similar to the autocorrelations decrease as time lag increases, which inspired us to treat time and space dimensions similarly in one of our future models.

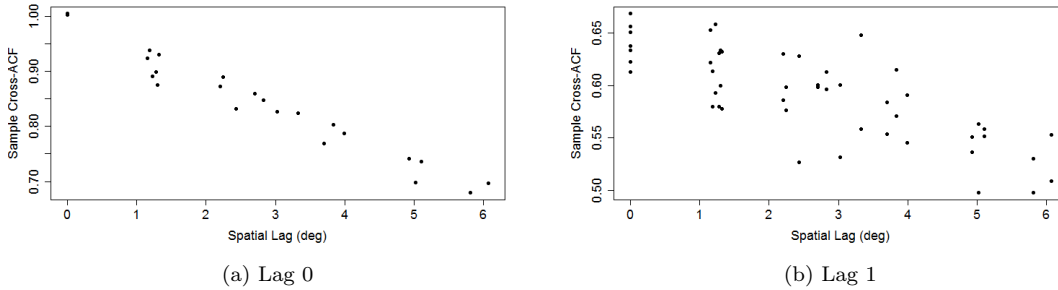


Figure 1.4: Sample auto-correlations and cross-correlations

Our parsimonious PVAR model fits the correlations and cross-correlations with one function. It is interesting to observe the structure of the sample correlations and sample cross-correlations. Figure 1.5 plots all seven sample autocorrelations between lags zero and ten, along with all 42 pairwise cross correlation functions. Note that the 42 pairs include each pair of cities in different orders, so that all 49 curves can be compared in the same graph. These curves have very similar shapes, thus rendering possible to fit them with one model.

The rest of this dissertation proceeds as follows. In chapter 2, we introduce methods to parsimoniously fit periodic vector time series. We also propose a residual test for periodic vector residuals, and establish its asymptotic properties. The residual test and the methods are applied to simulated data in sections 3.1 and 3.2, and on the temperature data introduced in section 3.3. Section 4 concludes with several comments.

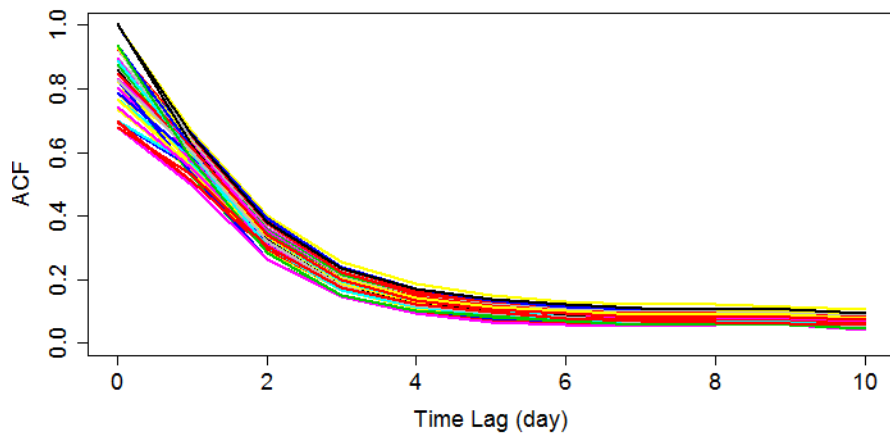


Figure 1.5: Sample cross correlations vs distance between stations

Chapter 2

Methods

In this chapter, we propose two methods to parsimoniously fit a periodic vector time series, and a portmanteau test for periodic vector series. Section 2.1 introduces a layered method similar to that in Lund et al. (1995). By using spline fits of seasonal means and other statistics, it significantly reduces the number of parameters needed to eliminate the periodic means and variances of the series. It then fits a sparse VAR model to the standardized series. Section 2.2 explain how to consolidate parameters in a PVAR model. The final method we adopt will be a combination of these two methods.

2.1 A Layered Method

As mentioned in Section 1.1, if a multivariate periodic time series $\{\mathbf{X}_{nT+\nu}\}$ can be linearly transformed into a stationary series,

$$\mathbf{Y}_{nT+\nu} = \mathbf{S}_\nu(\mathbf{X}_{nT+\nu} - \boldsymbol{\mu}_\nu) \tag{2.1}$$

then it would be parametrically more efficient to model the corresponding stationary series $\{\mathbf{Y}_t\}$. A layered method for periodic vector series analogous to that in Lund et al. (1995) for univariate series can be used.

The first layer is aimed at eliminating the periodic mean and variance in the series. Define

the seasonal sample mean and variance of the series by

$$\hat{\boldsymbol{\mu}}_\nu = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{X}_{nT+\nu},$$

$$\hat{\boldsymbol{\Gamma}}_\nu(0) = \frac{1}{N} \sum_{n=0}^{N-1} (\mathbf{X}_{nT+\nu} - \hat{\boldsymbol{\mu}}_\nu)(\mathbf{X}_{nT+\nu} - \hat{\boldsymbol{\mu}}_\nu)',$$

for $\nu = 1, 2, \dots, T$, where N is the number of cycles in the data. In practice, it is possible that a set of data does not contain a full number of cycles, i.e., the record length divided by period T is not an integer, a case we dismiss as trite work. The series $\{\mathbf{X}_{nT+\nu}\}$ is then standardized by the following linear transformation for each season ν ,

$$\mathbf{Y}_{nT+\nu} = \hat{\boldsymbol{\Gamma}}_\nu(0)^{-\frac{1}{2}}(\mathbf{X}_{nT+\nu} - \hat{\boldsymbol{\mu}}_\nu), \quad (2.2)$$

where $\hat{\boldsymbol{\Gamma}}_\nu(0)^{-\frac{1}{2}}$ denotes the matrix square root of the inverse of $\hat{\boldsymbol{\Gamma}}_\nu(0)$, i.e., a matrix \mathbf{S}_ν such that $\mathbf{S}_\nu^2 = \hat{\boldsymbol{\Gamma}}_\nu(0)^{-1}$.

As simple as this procedure is, it takes many parameters to standardize $\{\mathbf{X}_{nT+\nu}\}$. A total of dT parameters arise in $\hat{\boldsymbol{\mu}}_\nu$, and $dT + d(d+1)T/2$ parameters in \mathbf{S}_ν , which exceeds 10000 in the 7-city daily temperature data. Therefore, it is worth modifying this procedure for parsimony. As is shown in Section 1.3, each of the seven components in the seasonal means $\hat{\boldsymbol{\mu}}_\nu$ (as a function of ν) can be fit with a spline with as few as six parameters, expending only 42 parameters in seasonal mean estimation. For the variance standardizing, we fit each component of the matrix \mathbf{S}_ν with a spline, because these are the coefficients directly applied on the series. More specifically, the i th component of $\mathbf{Y}_{nT+\nu}$ is a linear combination of components of $\mathbf{X}_{nT+\nu} - \hat{\boldsymbol{\mu}}_\nu$, with coefficients arising in the i th row of \mathbf{S}_ν . For example, Figure 2.1 plots the 1st, 2nd, and 7th coefficients in the first row. All components plotted show a relatively smooth change over the days of year. After this modification, we are using only $6(d + d(d+1)/2) = 210$ parameters in the first layer. Note that this number is proportional to d^2 ; essentially, we have only parsimonized the time component of these parameters. This count can still increase rapidly with d . With a large number of series in the model, it would be beneficial to also parsimonize the spatial component. This will be discussed in the next section.

According to Lund et al. (1995), we should check the coherence of the autocorrelations of $\{\mathbf{Y}_{nT+\nu}\}$. If there is still evidence of periodic autocorrelation, we can try other methods such as fitting a PVAR(1) model to $\{\mathbf{X}_{nT+\nu}\}$ and defining $\{\mathbf{Y}_{nT+\nu}\}$ as the residuals. These methods could

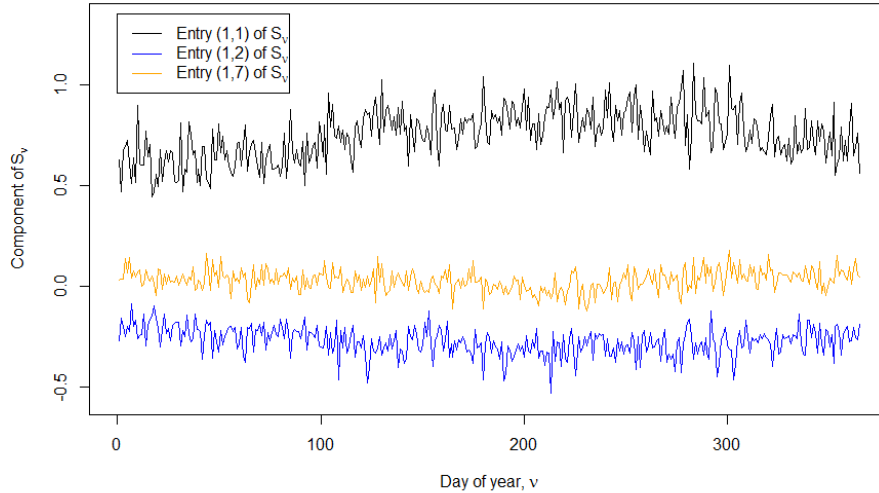


Figure 2.1: Components of $\hat{\Gamma}_\nu(0)^{-\frac{1}{2}}$ by day of year

use even more parameters and hence be harder to parsimonize. As such, we propose parsimonious PVAR models.

In the second layer, a stationary time series model is fitted to $\{\mathbf{Y}_t\}$. Univariate ARMA models or AR models can be fit to each component of $\{\mathbf{Y}_t\}$ to suggest model orders for the eventual VARMA/VAR model. In practice, although VARMA models tend to have less parameters than VAR models, their fit is more computationally demanding than that for VAR models. However, a VAR model for a daily series may need an AR order of about 30, meaning that there are well over 1,000 parameters arising in this layer. So if a VAR model is to be fitted, certain parsimonious method needs to be adopted. In this work, we fit a sparse VAR model to the series $\{\mathbf{Y}_t\}$ using the method from Davis et al. (2016).

2.2 Parsimonious PVAR models

When modeling periodic vector series, one frequently deals with models with many parameters. Several attempts to reduce the number of parameters were made in the last section, but all of them only reduce time parameters. When d is big, the models above can still have a large number of parameters, and this number can increase very rapidly due to the term d^2 arising in the model parameter counts. The purpose of this section is to study ways of modeling a periodic vector series

using a smaller set of parameters, and eventually develop a model that is parsimonious in parameters in both time and space.

2.2.1 Time Parameter Consolidation

Lund et al. (2006) propose parsimonious periodic ARMA modeling strategies for univariate series. A natural analogy is to apply the same methods to periodic vector series. Consider the zero-mean periodic d -dimensional VARMA model governed by

$$\mathbf{X}_{nT+\nu} - \sum_{i=1}^p \Phi_{\nu}(i) \mathbf{X}_{nT+\nu-i} = \mathbf{Z}_{nT+\nu} + \sum_{j=1}^q \Theta_{\nu}(j) \mathbf{Z}_{nT+\nu-j}, \quad \nu = 1, 2, \dots, T, \quad (2.3)$$

where $\{\mathbf{Z}_{nT+\nu}\}$ is a d -dimensional white noise with mean $\mathbf{0}$ and variance Σ_{ν} during season ν . Unconstrained, this model has $[d^2(p+q) + d(d+1)/2]T$ parameters. An example of parametric consolidation constraints on the autoregressive, moving-average, and white noise variance matrices are:

$$\begin{aligned} \Phi_{\nu}(i) &= \sum_{k=0}^K A_k^{\Phi}(i) b_k(\nu), \quad i = 1, 2, \dots, p; \\ \Theta_{\nu}(j) &= \sum_{k=0}^K A_k^{\Theta}(j) b_k(\nu), \quad j = 1, 2, \dots, q; \\ \Sigma_{\nu} &= \sum_{k=0}^K A_k^{\Sigma} b_k(\nu), \end{aligned} \quad (2.4)$$

for $\nu = 1, \dots, T$, where $A_k^{\Phi}(i)$, $A_k^{\Theta}(j)$, and A_k^{Σ} are $d \times d$ matrix coefficients, and $\{b_k(\cdot); k = 1, \dots, K\}$ is a set of basis functions with period T . For example, a 1st order Fourier representation of $\Phi_{\nu}(i)$ is

$$\Phi_{\nu}(i) = A_0^{\Phi}(i) + A_1^{\Phi}(i) \cos\left(\frac{2\pi\nu}{T}\right) + A_2^{\Phi}(i) \sin\left(\frac{2\pi\nu}{T}\right).$$

Other basis functions such as splines and wavelets may also be considered when appropriate.

It is necessary to check causality of the model in 2.3 with parameters constrained by equations 2.4. However, since each equation in 2.3 involves random vectors of different seasons, the VAR causality conditions cannot be applied on these equations separately. One needs to examine the causality condition of the concatenated model as described in section 1.2. For example, the concatenated VAR model 1.12 is causal if and only if the solution to the equation $\det(\mathbf{I} - \tilde{\Phi}_1 z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

The number of free parameters under this constraint is $[d^2(p+q) + d(d+1)/2]K$, which can be considerably smaller than that of the original PVARMA model. However, the factor $[d^2(p+q) + d(d+1)/2]$, which is related to the spatial aspect of the model, has not changed. In fact, under this framework, it is difficult to associate the space-related components of matrices $A_k^\Phi(i)$, $A_k^\Theta(j)$, and A_k^Σ to some other factor, for example, distance, because the matrices lack spatial connotations. To reduce parameter counts in this first factor, we need to find a circumstance where the components of a matrix are closely related to the model, so that we can effectively “predict” all components from a smaller number of parameters.

2.2.2 Spatial Parameter Consolidation

Spatial correlation analysis often associates correlations to spatial parameters, and can help consolidate parameters. We start with some parameter consolidations on stationary vector series, and then proceed to periodic vector series.

Consider a zero-mean stationary vector time series $\{\mathbf{X}_t = (X_t^{(1)}, \dots, X_t^{(d)})'; t \in \mathbb{Z}\}$. Let $\mathbf{\Gamma}(h) = \mathbb{E}(\mathbf{X}_t \mathbf{X}_{t-h}')$ be its lag- h autocovariance. Denote the entries of the autocovariance matrix by $\gamma_{ij}(h) = \text{ent}_{ij} \mathbf{\Gamma}(h)$. Note that $\gamma_{ij}(h)$ is a covariance:

$$\gamma_{ij}(h) = \mathbb{E}(X_t^{(i)} X_{t-h}^{(j)}) = \text{cov}(X_t^{(i)}, X_{t-h}^{(j)}),$$

where $X_t^{(i)}$ and $X_{t-h}^{(j)}$ are respectively associated with the space and time indices (i, t) and $(j, t-h)$. Taking the temperature data as an example, $X_t^{(i)}$ is the temperature of the i th station on day t , and $X_{t-h}^{(j)}$ is the temperature of the j th station on day $t-h$, each of them respectively associated with coordinates (x_i, y_i, t) and $(x_j, y_j, t-h)$. Note that this coordinate system has combined temporal and spatial dimensions. It is possible to relate the covariance to these coordinates or to the distance between the coordinates.

Suppose that there exist d row vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ of dimension $k-1$, such that

$$X_{i,t} \equiv \tilde{X}(\mathbf{v}_i, t).$$

Then the components $\{(X_{1,t}, \dots, X_{d,t})\}_{t=-\infty}^{+\infty}$ of vector time series $\{\mathbf{X}_t\}$ are viewed as a sample from the spatial series $\{\tilde{X}(\mathbf{w}); \mathbf{w} = (\mathbf{v}, t) \in \mathbb{R}^k\}$. Thus, the covariance of the vector time series $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$

satisfies

$$\begin{aligned}
\gamma_{ij}(h) &= \mathbb{E}[X_{i,t}X_{j,t-h}] \\
&= \mathbb{E}[\tilde{X}(\mathbf{v}_i, t) \tilde{X}(\mathbf{v}_j, t-h)] \\
&= C((\mathbf{v}_i - \mathbf{v}_j, h)).
\end{aligned} \tag{2.5}$$

Since $\mathbf{\Gamma}(h) = \mathbb{E}[\mathbf{X}_t \mathbf{X}'_{t-h}]$, the matrix in (1.4) is equivalent to

$$\mathbf{G}_n = \text{Var} \left[(\mathbf{X}'_n | \mathbf{X}'_{n-1} | \cdots | \mathbf{X}'_1)' \right],$$

where the random vector

$$(\mathbf{X}'_n | \mathbf{X}'_{n-1} | \cdots | \mathbf{X}'_1) = \left(\tilde{X}(\mathbf{v}_1, n) \dots \tilde{X}(\mathbf{v}_d, n) \mid \cdots \mid \tilde{X}(\mathbf{v}_1, 1) \dots \tilde{X}(\mathbf{v}_d, 1) \right).$$

Based on the non-negative definiteness of a valid covariogram, \mathbf{G}_n must be non-negative definite. In short, as long as $C(\cdot)$ is a legitimate covariogram of a stationary spatial series, $\mathbf{\Gamma}(\cdot)$ defined by (2.5) is a legitimate covariance of a stationary vector time series.

In the context of our temperature data, we can assume that $k = 3$ and let $\mathbf{v} = (x, y)$ be the coordinates (longitude, latitude) of the stations. Thus, the random variable $\tilde{X}(\mathbf{v}, t)$ denotes the temperature at position $\mathbf{v} = (x, y)$ at time t . If the covariogram of $\{\tilde{X}(\mathbf{w})\}$ is isotropic, then the covariance of $\tilde{X}(\mathbf{w}_1)$ and $\tilde{X}(\mathbf{w}_2)$ is a function of the “distance” between their indices, $\|(\mathbf{w}_1 - \mathbf{w}_2)\|$. For example, the distance based on the Euclidean norm is

$$\|(\mathbf{w}_1 - \mathbf{w}_2)\|_{Euc} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (t_1 - t_2)^2}.$$

Based on lemma 1, we can apply a linear transformation on the indices to enlarge the family of feasible covariance functions. For example, a model we will use fits the covariances of the seven-city temperature data by covariogram 1.7, with norm $\|\cdot\|$ defined by

$$\|(x, y, t)\| = \sqrt{(b_1x + b_2t)^2 + (b_3x + b_4y + b_5t)^2 + t^2}.$$

Now, suppose that $\{\mathbf{X}_t\}$ is a VAR(p) series defined by equation (1.2). With the covariance function form specified, a VAR model can then be fitted by solving the Yule-Walker moment

equations, which are

$$\begin{pmatrix} \Gamma(0) & \Gamma(1) & \cdots & \Gamma(p-1) \\ \Gamma(-1) & \Gamma(0) & \cdots & \Gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(1-p) & \Gamma(2-p) & \cdots & \Gamma(0) \end{pmatrix} \begin{pmatrix} \Phi(1)' \\ \Phi(2)' \\ \vdots \\ \Phi(p)' \end{pmatrix} = \begin{pmatrix} \Gamma(-1) \\ \Gamma(-2) \\ \vdots \\ \Gamma(-p) \end{pmatrix}, \quad (2.6)$$

and

$$\Gamma(0) = \sum_{i=1}^p \sum_{j=1}^p \Phi(i)\Gamma(j-i)\Phi(j)' + \Sigma. \quad (2.7)$$

Note that all autoregressive parameters are determined by the free parameters in the covariogram. The optimal values of these parameters can be obtained by maximizing the model likelihood using numerical methods. We summarize parameter estimation for this method as follows:

1. Construct the likelihood function of $\{\mathbf{X}_t\}$ given the vector of free parameters β :
 - (a) Given β , construct the covariance function $\Gamma(\cdot|\beta)$;
 - (b) Solve the Yule-Walker equations for the VAR parameters $\Phi^{(\beta)}(k)$, $k = 1, \dots, p$, and $\Sigma^{(\beta)}$;
 - (c) Compute residuals given the series and the parameters defined above;
 - (d) Compute and return the model likelihood $L(\{\mathbf{X}_t\}|\beta)$.
2. Numerically maximize the likelihood function to obtain the optimal parameter estimator $\hat{\beta}$;
3. Go through steps (1a) and (1b) to estimate the VAR model.

Before proceeding, we provide a short digression on space-time separable covariograms satisfying

$$C(\mathbf{v}, t) = C_1(\mathbf{v})C_2(t), \quad (2.8)$$

for some covariograms $C_1(\cdot)$ on \mathbb{R}^{k-1} and $C_2(\cdot)$ on \mathbb{R}^1 . The following proposition shows that taking a space-time separable covariogram in our method results in a VAR model whose coefficient matrices Φ_1, \dots, Φ_p are all diagonal (not what we need).

Proposition 1. *Suppose that $\{X_{i,t}\}_{t=-\infty}^{\infty}$ and $i \in \{1, \dots, d\}$ are the components of a VAR series $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$. Then the following are equivalent.*

i) $\{X_{i,t}; i = 1, \dots, d\}_{t=-\infty}^{\infty}$ is a collection of univariate AR series, satisfying

$$X_{i,t} = \sum_{k=1}^p \phi_{ii}(k) X_{i,t-k} + Z_{i,t}, \quad i = 1, \dots, d.$$

ii) $\Phi(k)$, $k = 1, 2, \dots, p$ are all diagonal;

iii) The autocovariance of $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$ is separable in that

$$\mathbf{\Gamma}(h) = \mathbf{\Gamma}(0)C_2(h)$$

for some univariate autocovariance function $C_2(\cdot)$.

Proof. Obviously, i) and ii) are equivalent and sufficient for iii). If iii) holds, then the Yule-Walker equations become

$$\tilde{\mathbf{C}}_0 \otimes \mathbf{\Gamma}(0) \begin{pmatrix} \Phi(1)' \\ \vdots \\ \Phi(p)' \end{pmatrix} = \tilde{\mathbf{C}}_1 \otimes \mathbf{\Gamma}(0), \quad (2.9)$$

where \otimes is the Kronecker product, and

$$\tilde{\mathbf{C}}_0 = \begin{pmatrix} C_2(0) & C_2(1) & \cdots & C_2(p-1) \\ C_2(-1) & C_2(0) & \cdots & C_2(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_2(1-p) & C_2(2-p) & \cdots & C_2(0) \end{pmatrix}, \quad \text{and} \quad \tilde{\mathbf{C}}_1 = \begin{pmatrix} C_2(-1) \\ C_2(-2) \\ \vdots \\ C_2(-p) \end{pmatrix}$$

Applying the property $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ yields

$$\left(\mathbf{I}_p \otimes \mathbf{\Gamma}(0) \right) \left(\tilde{\mathbf{C}}_0 \otimes \mathbf{I}_d \right) \begin{pmatrix} \Phi(1)' \\ \vdots \\ \Phi(p)' \end{pmatrix} = \left(\mathbf{I}_p \otimes \mathbf{\Gamma}(0) \right) \left(\tilde{\mathbf{C}}_1 \otimes \mathbf{I}_d \right).$$

where \mathbf{I}_k is the $k \times k$ identity matrix. Since $\mathbf{\Gamma}(0)$ is positive definite, $\mathbf{I}_p \otimes \mathbf{\Gamma}(0)$ is invertible. Hence,

$$\left(\tilde{\mathbf{C}}_0 \otimes \mathbf{I}_d\right) \begin{pmatrix} \mathbf{\Phi}(1)' \\ \vdots \\ \mathbf{\Phi}(p)' \end{pmatrix} = \left(\tilde{\mathbf{C}}_1 \otimes \mathbf{I}_d\right).$$

Now left multiply by $\left(\tilde{\mathbf{C}}_0 \otimes \mathbf{I}_d\right)^{-1}$ and apply the property $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, producing

$$\begin{pmatrix} \mathbf{\Phi}(1)' \\ \vdots \\ \mathbf{\Phi}(p)' \end{pmatrix} = \left(\tilde{\mathbf{C}}_0^{-1} \otimes \mathbf{I}_d\right) \left(\tilde{\mathbf{C}}_1 \otimes \mathbf{I}_d\right) = \left(\tilde{\mathbf{C}}_0^{-1} \tilde{\mathbf{C}}_1\right) \otimes \mathbf{I}_d.$$

Therefore, $\mathbf{\Phi}(\ell)$ is diagonal for each $\ell \in \{1, \dots, p\}$. □

Proposition 1 shows one way that a parsimonious VAR model can be constructed. Specifically, if $C_2(\cdot)$ uses ℓ parameters, then $\mathbf{\Gamma}(0)C_2(h)$ only uses $\ell + d(d+1)/2$ free parameters. Unfortunately, our data does not appear space-time separable.

2.2.3 Space-Time Parametric Consolidation

A parsimonious PVAR model can be constructed using a similar tactic. In this section, a covariogram model for a periodically stationary VAR series will be constructed, and the periodic versions of the Yule-Walker equations and likelihood will be given.

Suppose $f(t) = t + f_0(t)$ is a time warping function where $f_0(\cdot)$ is a function with period T . Define $X_{i,t} = \tilde{X}(\mathbf{v}_i, f(t))$, where \mathbf{v}_i and $\{\tilde{X}(\mathbf{v}, t)\}$ are defined as before. Then $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$ with $\mathbf{X}_t = (X_{1,t}, \dots, X_{d,t})'$ is a periodically stationary vector time series. To identify the periodically varying covariance $\mathbf{\Gamma}_\nu(h) = \mathbb{E}[\mathbf{X}_{nT+\nu} \mathbf{X}'_{nT+\nu-h}]$, note that

$$\begin{aligned} \gamma_{ij}^\nu(h) &:= \text{ent}_{ij}[\mathbf{\Gamma}_\nu(h)] \\ &= \mathbb{E}[X_{i,nT+\nu} X_{j,nT+\nu-h}] \\ &= \mathbb{E}[\tilde{X}(\mathbf{v}_i, f(nT + \nu)) \tilde{X}(\mathbf{v}_j, f(nT + \nu - h))] \\ &= C(\mathbf{v}_i - \mathbf{v}_j, f(nT + \nu) - f(nT + \nu - h)) \\ &= C(\mathbf{v}_i - \mathbf{v}_j, h + f_0(\nu) - f_0(\nu - h)). \end{aligned}$$

To illustrate its role, the $f_0(\cdot)$ obtained in our model fit is plotted in Figure 2.2a. The function increases faster in Spring, causing a greater distance between $f(nT + \nu + h)$ and $f(nT + \nu)$. Thus, a time lag of h days in Spring is like a lag of about $3h$ days in Fall, resulting in a lower correlation between observations at these two time points. Correspondingly, the fitted autocorrelation at lag 1 (for a single station) is plotted in Figure 2.2b, which is close to the shape of the sample autocorrelation at lag 1 in Figure 1.1d and supports our analysis.

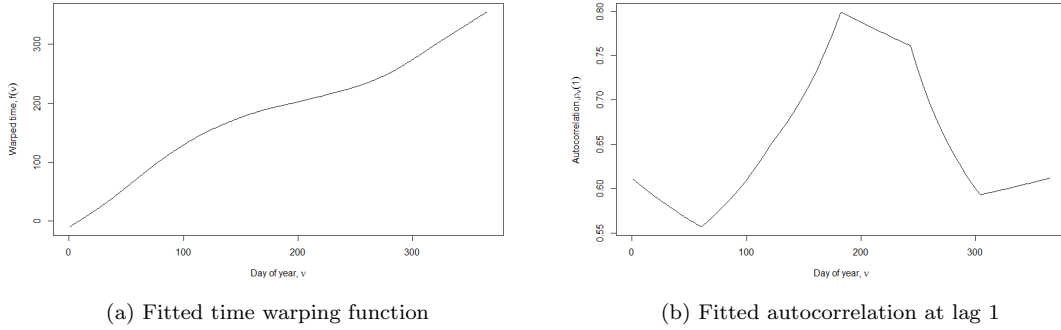


Figure 2.2: Time warping function and its effect

If $\{\mathbf{X}_{nT+\nu}\}$ is the PVAR series in (1.10), then the PVAR coefficients can be computed from the covariances via the Yule-Walker equations:

$$\begin{pmatrix} \Gamma_{\nu-1}(0) & \Gamma_{\nu-1}(1) & \cdots & \Gamma_{\nu-1}(p-1) \\ \Gamma_{\nu-2}(-1) & \Gamma_{\nu-2}(0) & \cdots & \Gamma_{\nu-2}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{\nu-p}(1-p) & \Gamma_{\nu-p}(2-p) & \cdots & \Gamma_{\nu-p}(0) \end{pmatrix} \begin{pmatrix} \Phi_{\nu}(1)' \\ \Phi_{\nu}(2)' \\ \vdots \\ \Phi_{\nu}(p)' \end{pmatrix} = \begin{pmatrix} \Gamma_{\nu-1}(-1) \\ \Gamma_{\nu-2}(-2) \\ \vdots \\ \Gamma_{\nu-p}(-p) \end{pmatrix},$$

and Σ_{ν} via

$$\Gamma_{\nu}(0) = \sum_{i=1}^p \sum_{j=1}^p \Phi_{\nu}(i) \Gamma_{\nu-i}(j-i) \Phi_{\nu}(j)' + \Sigma_{\nu}.$$

Note that the parameters used to construct the PVAR coefficients $\Phi_{\nu}(i)$ and Σ_{ν} , $i = 1, \dots, p$, $\nu \in \{1, \dots, T\}$ are those used in $C(\cdot)$ and in $f(\cdot)$, which can be much less than the number of original PVAR parameters.

2.3 Model Evaluation

2.3.1 Likelihood

The methods introduced in sections 2.2.2 and 2.2.3 require evaluation of a Gaussian likelihood. Given data $\{\mathbf{X}_t\}_{t=1}^N$, the likelihood of a PVAR model of order p is

$$L_X(\boldsymbol{\theta}) = (2\pi)^{-Nd/2} (\det \mathbf{G}_p)^{-1/2} \prod_{t=p+1}^N (\det \boldsymbol{\Sigma}_{\nu(t)}) \times \exp \left\{ -\frac{1}{2} \vec{\mathbf{X}}_p' \mathbf{G}_p^{-1} \vec{\mathbf{X}}_p - \frac{1}{2} \sum_{t=p+1}^N (\mathbf{X}_t - \hat{\mathbf{X}}_t)' \boldsymbol{\Sigma}_{\nu}^{-1} (\mathbf{X}_t - \hat{\mathbf{X}}_t) \right\},$$

where $\nu(t)$ is the season of time t , i.e., $\nu(t) = ((t-1) \bmod T) + 1$; $\vec{\mathbf{X}}_p$ is the concatenated vector $\vec{\mathbf{X}}_p' = (\mathbf{X}'_p, \mathbf{X}'_{p-1}, \dots, \mathbf{X}'_1)$; $\hat{\mathbf{X}}_{nT+\nu} = \sum_{k=1}^p \boldsymbol{\Phi}_{\nu}(k) \mathbf{X}_{nT+\nu-k}$ for $t > p$; and

$$\mathbf{G}_p = \text{var}(\vec{\mathbf{X}}_p) = \begin{pmatrix} \boldsymbol{\Gamma}_{\nu}(0) & \boldsymbol{\Gamma}_{\nu}(1) & \cdots & \boldsymbol{\Gamma}_{\nu}(p-1) \\ \boldsymbol{\Gamma}_{\nu-1}(-1) & \boldsymbol{\Gamma}_{\nu-1}(0) & \cdots & \boldsymbol{\Gamma}_{\nu-1}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{\nu-p+1}(1-p) & \boldsymbol{\Gamma}_{\nu-p+1}(2-p) & \cdots & \boldsymbol{\Gamma}_{\nu-p+1}(0) \end{pmatrix}.$$

One may also use the likelihood conditional on the first p observations,

$$L_X(\boldsymbol{\theta} | \mathbf{X}_1, \dots, \mathbf{X}_p) = (2\pi)^{-\frac{(N-p)d}{2}} \prod_{t=p+1}^N (\det \boldsymbol{\Sigma}_{\nu(t)})^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=p+1}^N (\mathbf{X}_t - \hat{\mathbf{X}}_t)' \boldsymbol{\Sigma}_{\nu}^{-1} (\mathbf{X}_t - \hat{\mathbf{X}}_t) \right\}, \quad (2.10)$$

What needs further discussion is the likelihood of models in the layered method in section 2.1. Recall that in the second layer, we fit the following sparse VAR model

$$\mathbf{Y}_t = \sum_{k=1}^p \tilde{\boldsymbol{\Phi}}_k \mathbf{Y}_{t-k} + \tilde{\mathbf{Z}}_t, \quad (2.11)$$

with the parameter constraint $\text{vec}(\tilde{\boldsymbol{\Phi}}_1 | \cdots | \tilde{\boldsymbol{\Phi}}_p) = \mathbf{R}\boldsymbol{\beta}$ and $\{\tilde{\mathbf{Z}}_t\}$ a vector white noise sequence with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}_{(\mathbf{Y})}$. For simplicity, we assume that $\{\mathbf{X}_{nT+\nu}\}$ has zero mean. Then definition

(2.2) becomes $\mathbf{Y}_{nT+\nu} = \hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} \mathbf{X}_{nT+\nu}$. Substituting this into (2.11) yields

$$\begin{aligned} \hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} \mathbf{X}_{nT+\nu} &= \sum_{k=1}^p \tilde{\mathbf{\Phi}}_k \hat{\mathbf{\Gamma}}_{\nu-k}(0)^{-\frac{1}{2}} \mathbf{X}_{nT+\nu-k} + \tilde{\mathbf{Z}}_t, \\ \iff \mathbf{X}_{nT+\nu} &= \sum_{k=1}^p \hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{\Phi}}_k \hat{\mathbf{\Gamma}}_{\nu-k}(0)^{-\frac{1}{2}} \mathbf{X}_{nT+\nu-k} + \hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{Z}}_t; \end{aligned}$$

thus $\{\mathbf{X}_{nT+\nu}\}$ is a PVAR(p) series with

$$\begin{aligned} \mathbf{\Phi}_\nu(k) &= \hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{\Phi}}_k \hat{\mathbf{\Gamma}}_{\nu-k}(0)^{-\frac{1}{2}}, \\ \mathbf{Z}_{nT+\nu} &= \hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{Z}}_t. \end{aligned}$$

It follows that the covariance matrix $\mathbf{\Sigma}_\nu$ for the PVAR(p) series is

$$\begin{aligned} \mathbf{\Sigma}_\nu &= \mathbb{E}(\mathbf{Z}_{nT+\nu} \mathbf{Z}'_{nT+\nu}) \\ &= \mathbb{E}((\hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{Z}}_t)(\hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \tilde{\mathbf{Z}}_t)') \\ &= \hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}} \mathbf{\Sigma}_{(\mathbf{Y})} (\hat{\mathbf{\Gamma}}_\nu(0)^{\frac{1}{2}})'. \end{aligned}$$

From the derivations above, we know that $\det[\mathbf{\Sigma}_{\nu(t)}] = \det[\mathbf{\Sigma}_{(\mathbf{Y})}] \det[\hat{\mathbf{\Gamma}}_\nu(0)]$ and

$$\begin{aligned} (\mathbf{X}_t - \hat{\mathbf{X}}_t)' \mathbf{\Sigma}_\nu^{-1} (\mathbf{X}_t - \hat{\mathbf{X}}_t) &= (\hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t))' \mathbf{\Sigma}_\nu^{-1} (\hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} (\mathbf{X}_t - \hat{\mathbf{X}}_t)) \\ &= (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)' \hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} \mathbf{\Sigma}_\nu^{-1} \hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t) \\ &= (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)' \mathbf{\Sigma}_{(\mathbf{Y})} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t). \end{aligned}$$

Hence, the conditional likelihood in (2.10) is

$$\begin{aligned} L_X(\boldsymbol{\theta} | \mathbf{X}_1, \dots, \mathbf{X}_p) &= (2\pi)^{-\frac{(N-p)d}{2}} \prod_{t=p+1}^N (\det \mathbf{\Sigma}_{(\mathbf{Y})} \det \mathbf{\Gamma}_\nu(0))^{-\frac{1}{2}} \times \\ &\quad \exp \left\{ -\frac{1}{2} \sum_{t=p+1}^N (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)' \mathbf{\Sigma}_{(\mathbf{Y})} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t) \right\} \\ &= (2\pi)^{-\frac{(N-p)d}{2}} (\det \mathbf{\Sigma}_{(\mathbf{Y})})^{-\frac{N-p}{2}} \prod_{t=p+1}^N (\det \mathbf{\Gamma}_\nu(0))^{-\frac{1}{2}} \times \\ &\quad \exp \left\{ -\frac{1}{2} \sum_{t=p+1}^N (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)' \mathbf{\Sigma}_{(\mathbf{Y})} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t) \right\}. \end{aligned} \tag{2.12}$$

Note that the conditional likelihood of the VAR model (2.11) is

$$L_Y(\boldsymbol{\beta}|\mathbf{Y}_1, \dots, \mathbf{Y}_p) = (2\pi)^{-\frac{(N-p)d}{2}} (\det \boldsymbol{\Sigma}_{(\mathbf{Y})})^{-\frac{N-p}{2}} \times \exp\left\{-\frac{1}{2} \sum_{t=p+1}^N (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)' \boldsymbol{\Sigma}_{(\mathbf{Y})}^{-1} (\mathbf{Y}_t - \hat{\mathbf{Y}}_t)\right\}. \quad (2.13)$$

Comparing (2.12) with (2.13) yields

$$L_X(\boldsymbol{\theta}|\mathbf{X}_1, \dots, \mathbf{X}_p) = L_Y(\boldsymbol{\beta}|\mathbf{Y}_1, \dots, \mathbf{Y}_p) \prod_{t=p+1}^N (\det \boldsymbol{\Gamma}_\nu(0))^{-\frac{1}{2}}.$$

Since our model evaluation will be based on the full model for the periodic vector series $\{\mathbf{X}_{nT+\nu}\}$, this equation will be helpful as it relates the likelihood of $\{\mathbf{X}_{nT+\nu}\}$ to that of $\{\mathbf{Y}_t\}$.

2.3.2 Residual Diagnostics

Let $\{\hat{\mathbf{Z}}_t\}_{t=p+1}^N$ be the residual series of a fitted VAR(p) or PVAR(p) model. Note that this series is based on $\{\mathbf{X}_t\}_{t=1}^N$; hence, the time index runs from $p+1$ to N . For a VAR(p) series, the residual is defined as $\hat{\mathbf{Z}}_t = \mathbf{X}_t - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}(j)\mathbf{X}_{t-j}$, while the residual of a PVAR(p) series is defined as $\hat{\mathbf{Z}}_{kT+\nu} = \mathbf{X}_t - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_\nu(j)\mathbf{X}_{kT+\nu-j}$. If the model is adequate, then $\{\hat{\mathbf{Z}}_t\}_{t=p+1}^N$ should have no significant serial or cross-sectional correlations. To test this property, let \mathbf{R}_ℓ be the theoretical lag ℓ cross-correlation matrix of the residual series. Our hypothesis of interest is

$$H_0 : \mathbf{R}_1 = \dots = \mathbf{R}_m = \mathbf{0} \quad \text{vs} \quad \mathbf{R}_\ell \neq \mathbf{0} \text{ for some } \ell \in \{1, \dots, m\}$$

To test this hypothesis in the VAR case, a multivariate portmanteau statistic introduced by Hosking (1980) is widely used. Among its variants, the original definition of the statistic is

$$Q(m) = N \sum_{\ell=1}^m \text{tr}(\hat{\mathbf{C}}(\ell)' \hat{\mathbf{C}}(0)^{-1} \hat{\mathbf{C}}(\ell) \hat{\mathbf{C}}(0)^{-1}), \quad (2.14)$$

where $\text{tr}(\cdot)$ is the trace function, and $\hat{\mathbf{C}}(\ell) = \frac{1}{N-p} \sum_{t=p+\ell+1}^N \hat{\mathbf{Z}}_t \hat{\mathbf{Z}}_{t-\ell}'$ is the lag ℓ cross-covariance matrix of the residual series. If $\{\mathbf{X}_t\}$ is a stationary process and was adequately fitted by a VAR(p) series, then $Q_k(m)$ follows a χ^2 distribution asymptotically with $d^2(m-p)$ degrees of freedom.

The above statistic, however, does not apply directly to assess the residual correlations of

a periodic time series. We follow the method of Hosking (1980) to construct portmanteau statistics for testing residual correlations of a PVAR series. Define the seasonal autocovariance of residuals by

$$\hat{\mathbf{C}}_\nu(\ell) = \frac{1}{n} \sum_{p+\ell < kT+\nu \leq N} \hat{\mathbf{Z}}_{kT+\nu} \hat{\mathbf{Z}}'_{kT+\nu-\ell},$$

where $\nu \in \{1, 2, \dots, T\}$ denotes the season, and $n = \lceil (N-p)/T \rceil$ the number of cycles of the data. The test statistic for season ν is defined by

$$Q_\nu(m) = n \sum_{\ell=1}^m \text{tr}(\hat{\mathbf{C}}_\nu(\ell)' \hat{\mathbf{C}}_\nu(0)^{-1} \hat{\mathbf{C}}_\nu(\ell) \hat{\mathbf{C}}_{\nu-\ell}(0)^{-1}). \quad (2.15)$$

We will show that, under certain conditions, $Q_\nu(m)$ has an asymptotic $\chi^2_{(m-p)d^2}$ distribution. It can also be shown that the statistic for each season is asymptotically independent of that in other seasons, based on which a statistic combining all seasons can be introduced.

For a causal PVAR series,

$$\hat{\mathbf{Z}}_{kT+\nu} = \mathbf{Z}_{kT+\nu} + \sum_{j=1}^p (\hat{\Phi}_\nu(j) - \hat{\Phi}_\nu(j)) \mathbf{X}_{kT+\nu-j}.$$

Let $\mathbf{X}_{kT+\nu} = \sum_{w=0}^{\infty} \Psi_\nu(w) \mathbf{Z}_{kT+\nu-w}$ be the moving average representation of the PVAR model.

Then

$$\hat{\mathbf{Z}}_{kT+\nu} = \mathbf{Z}_{kT+\nu} + \sum_{j=1}^p \sum_{w=0}^{\infty} (\hat{\Phi}_\nu(j) - \hat{\Phi}_\nu(j)) \Psi_{\nu-j}(w) \mathbf{Z}_{kT+\nu-j-w}.$$

To formulate the sample autocorrelation of the residual series, note that

$$\begin{aligned}
& \hat{\mathbf{Z}}_{kT+\nu} \hat{\mathbf{Z}}'_{kT+\nu-\ell} \\
&= \left[\mathbf{Z}_{kT+\nu} + \sum_{j=1}^p \sum_{w=0}^{\infty} (\Phi_{\nu}(j) - \hat{\Phi}_{\nu}(j)) \Psi_{\nu-j}(w) \mathbf{Z}_{kT+\nu-j-w} \right] \\
&\quad \times \left[\mathbf{Z}'_{kT+\nu-\ell} + \sum_{j=1}^p \sum_{w=0}^{\infty} \mathbf{Z}'_{kT+\nu-\ell-j-w} \Psi_{\nu-\ell-j}(w)' (\Phi_{\nu-\ell}(j)' - \hat{\Phi}_{\nu-\ell}(j)') \right] \\
&= \mathbf{Z}_{kT+\nu} \mathbf{Z}'_{kT+\nu-\ell} \\
&\quad + \sum_{j=1}^p \sum_{w=0}^{\infty} (\Phi_{\nu}(j) - \hat{\Phi}_{\nu}(j)) \Psi_{\nu-j}(w) \mathbf{Z}_{kT+\nu-j-w} \mathbf{Z}'_{kT+\nu-\ell} \\
&\quad + \sum_{j=1}^p \sum_{w=0}^{\infty} \mathbf{Z}_{kT+\nu} \mathbf{Z}'_{kT+\nu-\ell-j-w} \Psi_{\nu-\ell-j}(w)' (\Phi_{\nu-\ell}(j)' - \hat{\Phi}_{\nu-\ell}(j)') \\
&\quad + O_p(n^{-1}).
\end{aligned}$$

Hence,

$$\begin{aligned}
\hat{\mathbf{C}}_{\nu}(\ell) &= \mathbf{C}_{\nu}(\ell) + \sum_{j=1}^p \sum_{w=0}^{\infty} (\Phi_{\nu}(j) - \hat{\Phi}_{\nu}(j)) \Psi_{\nu-j}(w) \mathbf{C}_{\nu-j-w}(\ell - j - w) \\
&\quad + \sum_{j=1}^p \sum_{w=0}^{\infty} \mathbf{C}_{\nu}(\ell + j + w) \Psi_{\nu-\ell-j}(w)' (\Phi_{\nu-\ell}(j)' - \hat{\Phi}_{\nu-\ell}(j)') \\
&\quad + O_p(n^{-1}).
\end{aligned}$$

Note that for any $\ell \neq 0$,

$$\mathbf{C}_{\nu}(\ell) = \frac{1}{n} \sum_{p+\ell < kT+\nu \leq N} \mathbf{Z}_{kT+\nu} \mathbf{Z}'_{kT+\nu-\ell} = O_p(n^{-1/2}),$$

and since $\Psi_{\nu}(k) = \mathbf{0}$ for any $k < 0$,

$$\begin{aligned}
\hat{\mathbf{C}}_{\nu}(\ell) &= \mathbf{C}_{\nu}(\ell) + \sum_{j=1}^p (\Phi_{\nu}(j) - \hat{\Phi}_{\nu}(j)) \Psi_{\nu-j}(\ell - j) \mathbf{C}_{\nu-\ell}(0) \\
&\quad + \sum_{j=1}^p \mathbf{C}_{\nu}(0) \Psi_{\nu-\ell-j}(-\ell - j)' (\Phi_{\nu-\ell}(j)' - \hat{\Phi}_{\nu-\ell}(j)') + O_p(n^{-1}) \\
&= \mathbf{C}_{\nu}(\ell) + \sum_{j=1}^p (\Phi_{\nu}(j) - \hat{\Phi}_{\nu}(j)) \Psi_{\nu-j}(\ell - j) \mathbf{C}_{\nu-\ell}(0) + O_p(n^{-1}).
\end{aligned}$$

This is equivalent to

$$\text{vec}[\hat{\mathbf{C}}_\nu(\ell)] = \text{vec}[\mathbf{C}_\nu(\ell)] + \sum_{j=1}^p \left[(\mathbf{C}_{\nu-\ell}(0)' \boldsymbol{\Psi}_{\nu-j}(\ell-j)') \otimes \mathbf{I}_d \right] \text{vec}[\boldsymbol{\Phi}_\nu(j) - \hat{\boldsymbol{\Phi}}_\nu(j)],$$

or

$$\text{vec}[\hat{\mathbf{C}}_\nu(\ell)] = \text{vec}[\mathbf{C}_\nu(\ell)] + \mathbf{V}_\nu(\ell) \text{vec}[\boldsymbol{\Phi}_\nu - \hat{\boldsymbol{\Phi}}_\nu] \quad (2.16)$$

where

$$\mathbf{V}_\nu(\ell) = \left((\mathbf{C}_{\nu-\ell}(0)' \boldsymbol{\Psi}_{\nu-1}(\ell-1)') \otimes \mathbf{I}_d \quad \dots \quad (\mathbf{C}_{\nu-\ell}(0)' \boldsymbol{\Psi}_{\nu-p}(\ell-p)') \otimes \mathbf{I}_d \right)$$

and $\boldsymbol{\Phi}_\nu = \begin{pmatrix} \boldsymbol{\Phi}_\nu(1) & \boldsymbol{\Phi}_\nu(2) & \dots & \boldsymbol{\Phi}_\nu(p) \end{pmatrix}$ (and similarly for $\hat{\boldsymbol{\Phi}}_\nu$). Let $\mathbf{C}_\nu = \begin{pmatrix} \mathbf{C}_\nu(1) & \dots & \mathbf{C}_\nu(m) \end{pmatrix}$ (and similarly $\hat{\mathbf{C}}_\nu$), and $\mathbf{V}'_\nu = (\mathbf{V}_\nu(1)', \dots, \mathbf{V}_\nu(m)')$. Then we can write (2.16) as

$$\text{vec}[\hat{\mathbf{C}}_\nu] = \text{vec}[\mathbf{C}_\nu] + \mathbf{V}_\nu \text{vec}[\boldsymbol{\Phi}_\nu - \hat{\boldsymbol{\Phi}}_\nu]. \quad (2.17)$$

Lemma 2.

$$\mathbf{W}_\nu := \text{cov}(\text{vec}[\mathbf{C}_\nu]) = \text{diag}(\boldsymbol{\Sigma}_{\nu-\ell} \otimes \boldsymbol{\Sigma}_\nu, \ell = 1, 2, \dots, m).$$

Lemma 3.

$$\mathbf{V}_\nu \mathbf{W}_\nu^{-1} \text{vec}[\mathbf{C}_\nu] = O(n^{-1/2}).$$

Then, in a few steps, we can show that $\text{vec}[\hat{\mathbf{C}}_\nu] = (\mathbf{I}_{d^2} - \mathbf{Q}_\nu) \text{vec}[\mathbf{C}_\nu] + O_p(n^{-1})$, where $\mathbf{Q}_\nu = \mathbf{V}_\nu (\mathbf{V}'_\nu \mathbf{W}_\nu^{-1} \mathbf{V}_\nu)^{-1} \mathbf{V}'_\nu \mathbf{W}_\nu^{-1}$. Left multiplying (2.16) by \mathbf{Q}_ν and applying Lemma 3 yields

$$\begin{aligned} \mathbf{Q}_\nu \mathbf{V}_\nu \text{vec}[\boldsymbol{\Phi}_\nu - \hat{\boldsymbol{\Phi}}_\nu] &= \mathbf{Q}_\nu \text{vec}[\hat{\mathbf{C}}_\nu]_\nu - \mathbf{Q}_\nu \text{vec}[\mathbf{C}_\nu] \\ &= -\mathbf{Q}_\nu \text{vec}[\mathbf{C}_\nu]. \end{aligned}$$

Note that $\mathbf{Q}_\nu \mathbf{V}_\nu = \mathbf{V}_\nu$; hence,

$$\mathbf{V}_\nu \text{vec}[\boldsymbol{\Phi}_\nu - \hat{\boldsymbol{\Phi}}_\nu] + \mathbf{Q}_\nu \text{vec}[\mathbf{C}_\nu] = 0.$$

Comparing with (2.16) yields $\text{vec}[\hat{\mathbf{C}}_\nu] = (\mathbf{I}_{d^2} - \mathbf{Q}_\nu) \text{vec}[\mathbf{C}_\nu] + O_p(n^{-1})$.

Theorem 1. *Asymptotically,*

$$\text{vec}[\hat{\mathbf{C}}_\nu] \sim N_{d^2} \left(\mathbf{0}, n^{-1}(\mathbf{I}_{md^2} - \mathbf{Q}_\nu)\mathbf{W}_\nu \right). \quad (2.18)$$

Theorem 2. *Asymptotically,*

$$n \left(\text{vec}[\hat{\mathbf{C}}_\nu] \right)' \text{diag}(\hat{\mathbf{C}}_{\nu-\ell} \otimes \hat{\mathbf{C}}_\nu, \ell = 1, 2, \dots, m) \text{vec}[\hat{\mathbf{C}}_\nu] \sim \chi_{(m-p)d^2}^2. \quad (2.19)$$

And the above statistic is equivalent to

$$P_\nu(m) = n \sum_{\ell=1}^m \text{tr}(\hat{\mathbf{C}}_\nu(\ell)' \hat{\mathbf{C}}_\nu(0)^{-1} \hat{\mathbf{C}}_\nu(\ell) \hat{\mathbf{C}}_{\nu-\ell}(0)). \quad (2.20)$$

Chapter 3

Results

3.1 Simulations: Periodic Multivariate Portmanteau Test

In this section, we study the behavior of the periodic vector portmanteau test statistics via simulation. The study is performed first on white noise series, and then on residual series from model estimations. If the test statistics have the desired asymptotic distribution under the null hypothesis, the corresponding p -values are supposed to follow uniform distribution on the interval $[0, 1]$. We observe the sample quantiles of p -values, which are equivalent to the empirical type I error rates of the simulations.

3.1.1 Testing Periodic Vector White Noise

The first set of periodic vector white noise series $\{\mathbf{Z}_{nT+\nu}\}$ is generated as follows. Set vector dimension $d = 3$, period $T = 4$, series length of 60 cycles (hence series length is 240), and seasonal

variances

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \begin{pmatrix} 1.08 & 0.08 & -0.1 \\ 0.08 & 1.49 & 0.24 \\ -0.1 & 0.24 & 1.72 \end{pmatrix}, & \boldsymbol{\Sigma}_2 &= \begin{pmatrix} 1.29 & -0.18 & 0.31 \\ -0.18 & 0.87 & -0.23 \\ 0.31 & -0.23 & 1.46 \end{pmatrix}, \\ \boldsymbol{\Sigma}_3 &= \begin{pmatrix} 1.05 & 0 & 0.01 \\ 0 & 0.36 & 0.37 \\ 0.01 & 0.37 & 0.85 \end{pmatrix}, & \boldsymbol{\Sigma}_4 &= \begin{pmatrix} 0.88 & 0.45 & 0.36 \\ 0.45 & 1.1 & -0.28 \\ 0.36 & -0.28 & 1.11 \end{pmatrix}. \end{aligned}$$

The size 60 sample from each season ν , $\mathbf{Z}_\nu, \mathbf{Z}_{T+\nu}, \dots, \mathbf{Z}_{59T+\nu}$, is generated from a multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}_\nu$. The periodic multivariate test statistics $Q(\ell)$, $\ell = 1, 2, \dots, 12$ of the white noise series $\{\mathbf{Z}_{nT+\nu}\}$ are computed. According to Theorem 2, $Q(\ell) \sim \chi_{\ell d^2 T}^2$; here the p -values are defined by $p_\ell = \mathbb{P}(\chi_{\ell d^2 T}^2 > Q(\ell))$. We generate $n = 10^5$ such series and denote the p -values of each trial by $p_\ell^{(i)}$, $i = 1, 2, \dots, n$, and the empirical Type I error rates by $A_\ell(\alpha) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{p_\ell^{(i)} < \alpha\}}$, where $\mathbf{1}_E$ is the indicator function equal to 1 when E occurs, and 0 otherwise. The empirical Type I error rates for $\alpha = 0.05, 0.1$, and 0.2 are as follows.

lag ℓ	1	2	3	4	5	6	7	8	9	10
$\alpha = 0.05$	0.047	0.047	0.05	0.048	0.044	0.044	0.042	0.04	0.036	0.034
$\alpha = 0.1$	0.098	0.101	0.102	0.096	0.092	0.09	0.087	0.083	0.079	0.073
$\alpha = 0.2$	0.214	0.212	0.207	0.197	0.197	0.192	0.18	0.171	0.163	0.159

While the empirical Type I error rates at lower lags are reasonably close to the desired values, the values for higher lags are noticeably lower. Further simulations show that, as the series lengths increase, the p -values for any lag eventually converge to the uniform distribution $Unif(0, 1)$. More series of the same dimensions are generated, and the empirical Type I error rates are plotted against the series lengths (in number of cycles) in Figure 3.1. Only the test results for lags 1, 4, 7, and 10 are plotted. Clearly empirical Type I errors at all lags converge appropriately as the series length increases.

Empirical Type I error behaviors under different dimensions (d, T) are shown in Figure 3.2. While the convergence rate is significantly slower when d is larger, a larger T does not seem to have a heavy effect on the convergence rate in terms of cycles; however, it should be noted that the same number of cycles now correspond to a longer series as a period contains more seasons.

Next, periodic vector moving-average (PVMA) series are simulated to illustrate the Type

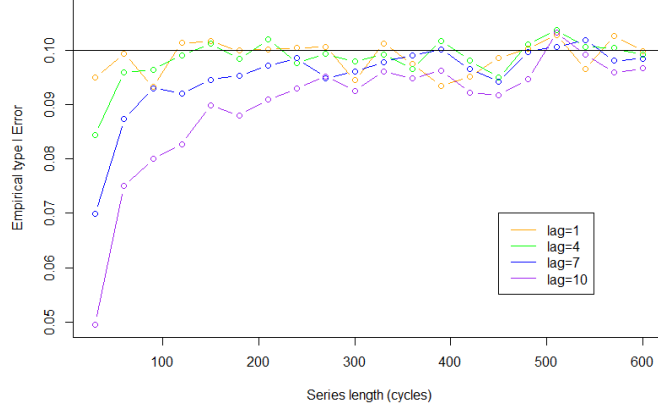


Figure 3.1: Empirical Type I errors vs series lengths, $d = 3, T = 4$

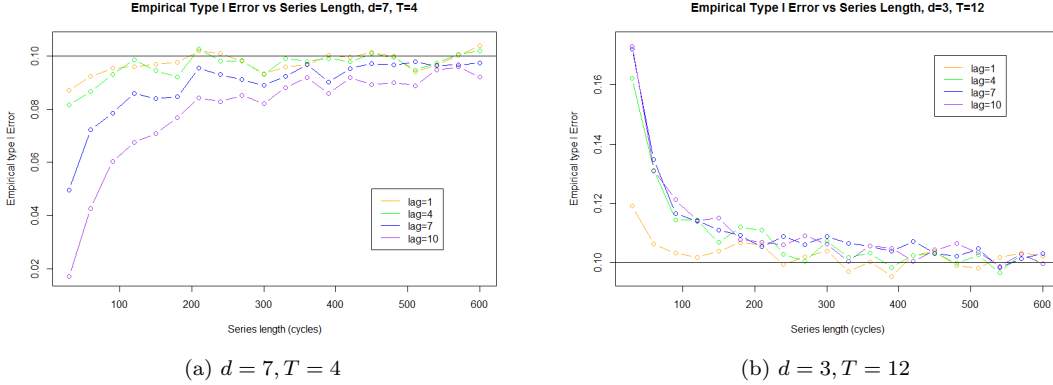


Figure 3.2: Empirical Type I error plots for models with different dimensions

II error rate of the periodic multivariate portmanteau test. The series is defined by $\mathbf{X}_{nT+\nu} = \mathbf{Z}_{nT+\nu} + \Theta \mathbf{Z}_{nT+\nu}$ where $\{\mathbf{Z}_{nT+\nu}\}$ is the same vector white noise defined above, and

$$\Theta = \begin{pmatrix} 0.5 & 0.1 & -0.2 \\ -0.2 & -0.4 & 0.1 \\ -0.1 & 0 & 0.45 \end{pmatrix}.$$

Periodic multivariate portmanteau test statistics of the series $\{\mathbf{X}_{nT+\nu}\}$ and their corresponding p -values are computed, and empirical powers of the test $A_\ell(\alpha) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{p_\ell^{(i)} < \alpha\}}$ are plotted against series lengths in Figure 3.3. All tests successfully confirm that $\{\mathbf{X}_{nT+\nu}\}$ is not a white noise series with a sample series length of 100 or more.

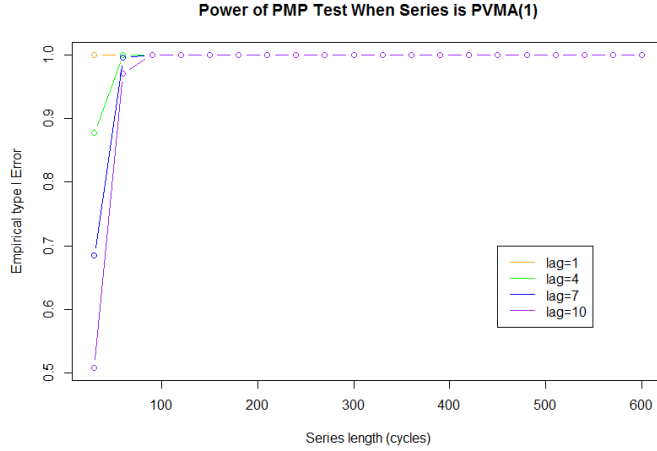


Figure 3.3: Power of periodic multivariate portmanteau test on PVMA series

3.1.2 Testing PVAR Residuals

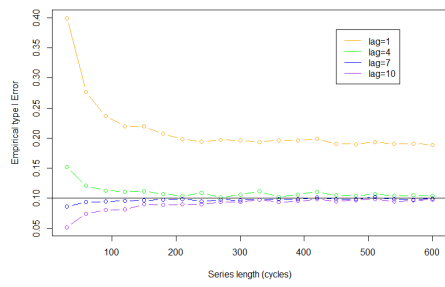
Next, we generate series from the PVAR(2) model

$$\mathbf{X}_{nT+\nu} = \Phi_{\nu}(1)\mathbf{X}_{nT+\nu-1} + \Phi_{\nu}(2)\mathbf{X}_{nT+\nu-2} + \mathbf{Z}_{nT+\nu}, \quad \nu = 1, 2, \dots, T.$$

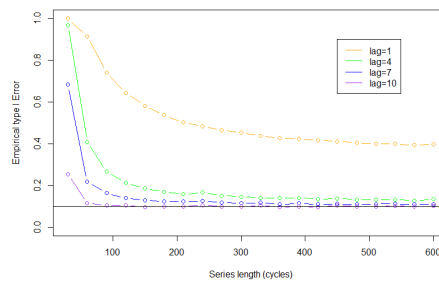
The parameters are estimated by solving the periodic Yule-Walker moment equations, and periodic multivariate portmanteau test statistics $Q(\ell), \ell = 1, 2, \dots, 12$ of the residuals are computed. According to Theorem 2, $Q(\ell) \sim \chi_{(\ell-2)d^2T}^2$, so the p -values are defined by $p_{\ell} = \mathbb{P}(\chi_{(\ell-2)d^2T}^2 > Q(\ell))$. $n = 10^5$ series are generated for each combination of dimensions, and the empirical Type I error rates are plotted in Figure 3.4.

3.2 Simulations: Parsimonious PVAR Model

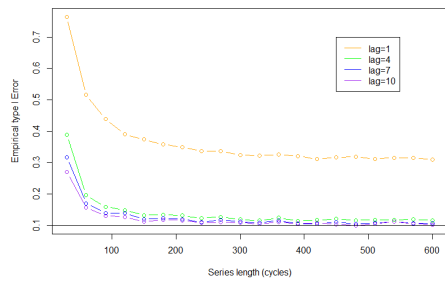
The same series simulated in section 3.1.2 are fitted with a parsimonious PVAR(2) model. Periodic multivariate portmanteau test statistics $Q(\ell), \ell = 1, 2, \dots, 12$ of the residuals are computed. Here we define the p -values by $p_{\ell} = \mathbb{P}(\chi_{\ell d^2 T - 12}^2 > Q(\ell))$, where 12 is the number of parameters in the parsimonious PVAR(2) model. Note that we are proposing this degree of freedom without a proof. Since the parsimonious PVAR fit is based on numerically maximizing the likelihood, it runs much slower, so we have only generated 1000 series for each given series length. As series length increases, the empirical Type I error rates converges to the expected value, 0.1.



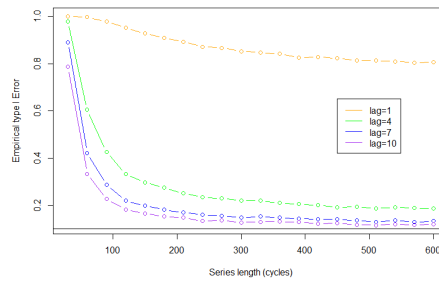
(a) $d = 3, T = 4$



(b) $d = 7, T = 4$



(c) $d = 3, T = 12$



(d) $d = 7, T = 12$

Figure 3.4: Empirical Type I error rates testing PVAR residuals

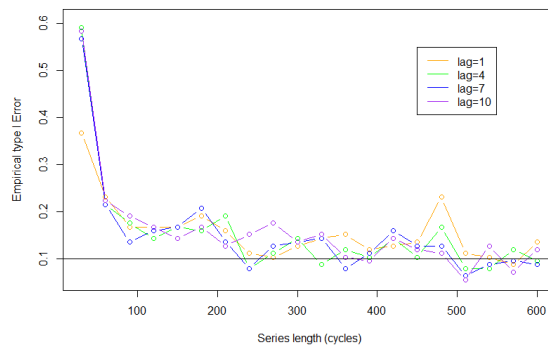


Figure 3.5: Empirical Type I error rates testing residuals of parsimonious PVAR model

3.3 Fitting the seven-city Data

The layered procedure described in section 2.1 is adopted as a basic scheme to fit the seven-city daily temperature data. The periodic means are removed prior to fitting any model. In the first layer, we compare different methods to eliminate the periodic variances. The methods include:

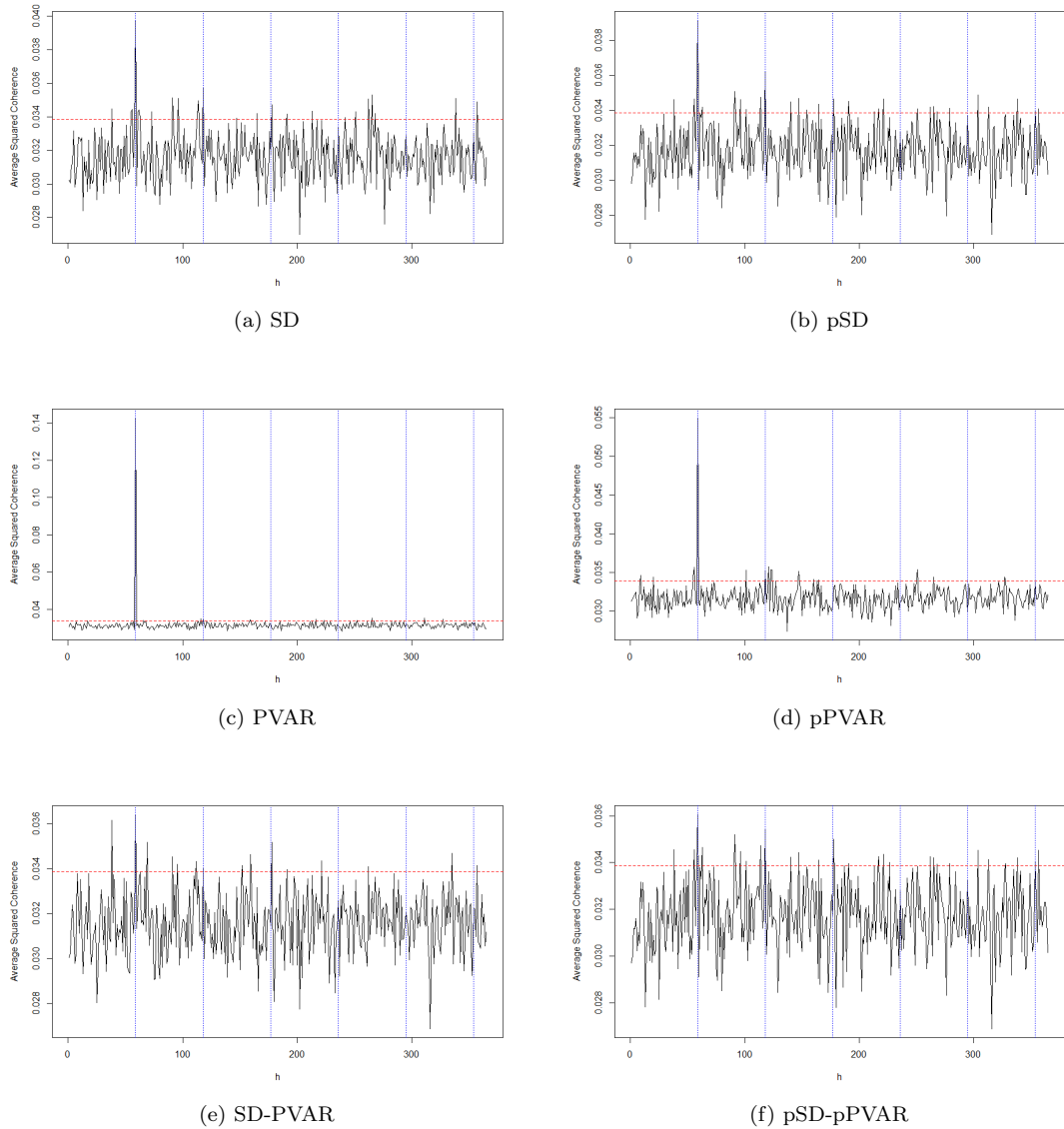


Figure 3.6: Coherence test of standardization procedures

SD Standardize the series according to equation 2.1.

pSD Standardize the series using spline fit of \mathbf{S}_ν in equation 2.1.

PVAR Fit a PVAR model to the series, and define the new series as the residuals from the PVAR model.

pPVAR Fit a parsimonious PVAR model to the series, and defined the new series as the residuals from the parsimonious PVAR model.

Note that only pSD and pPVAR are parsimonious methods. Coherence tests are performed on each component of the resulting series. The test results for the first component for each method is plotted in Figures 3.6a – 3.6d.

Here, the standardization methods SD and pSD failed to eliminate the periodicity, as the coherence plots for both methods show a significant spike at $h = 59$. The PVAR method and the pPVAR method also produce residual series with periodic correlations. Methods combining standardization and fitting PVAR models are then used. Method SD-PVAR fits a PVAR(1) to the standardized series defined by equation (2.2). Method standardizes the series using spline fits of components of $\hat{\mathbf{\Gamma}}_\nu(0)^{-\frac{1}{2}}$ in equation (2.2), and fit the standardized series with a parsimonious PVAR(3) model. Coherence tests are performed on each component of the residual series, and the test plot of the first component from each method is plotted in Figure 3.6e – 3.6f. Both of these two methods roughly eliminated periodicity in the series. Note that in method SD-PVAR, a total of $\frac{1}{2}d(d+1)T + d^2T = 28105$ parameters arise, while method pSD-pPVAR uses $\frac{6}{2}d(d+1) + 12 = 120$ parameters.

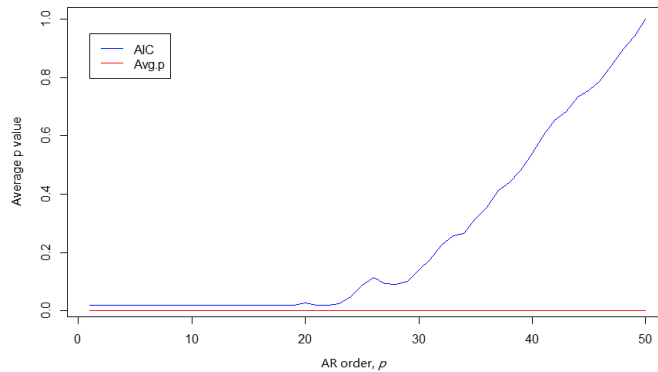


Figure 3.7: VAR fit

In the second layer, we fit the residual series from method pSD-pPVAR. A VAR model is first fitted to the residuals to provide a suggestion of autoregressive order p . The AIC values and average p -values from the multivariate Ljung-Box test are plotted in Figure 3.7. All VAR fits resulted in very small p -values, indicating significant correlation in residuals of the VAR fits. The AIC values suggest an AR order of $p \leq 22$ be used.

Next, a sparse VAR model with $p = 21$ is fitted to the residuals from the first layer, with choice of m , the number of unconstrained parameters, ranging from $d^2 = 49$ to $pd^2 = 1029$. Log likelihood, AIC and BIC values are computed, and residuals from the sparse VAR fits are tested by multivariate Ljung-Box test, with average p -values from the first 12 lags computed. These values are plotted in Figure 3.8. Although BIC was suggested as criterion for selection of m value, neither

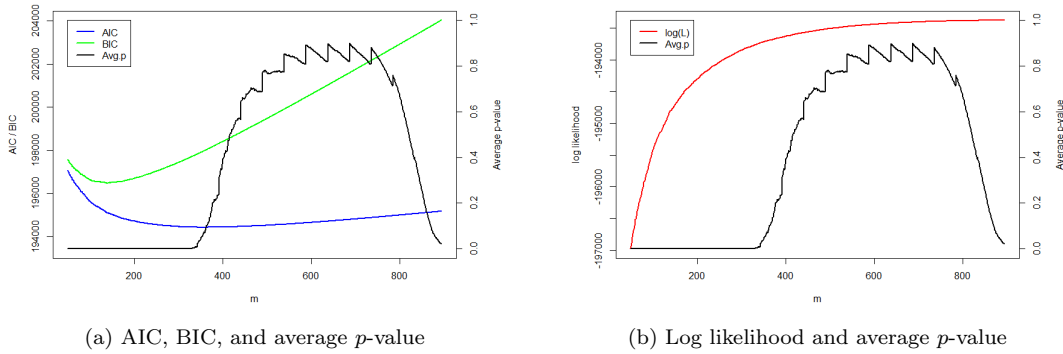


Figure 3.8: "AIC, BIC, log likelihood, and average p -values for sparse VAR fits"

BIC nor AIC suggests a value of m such that the sparse VAR fit would pass the residual test. Thus, we propose a composite criterion which guarantees that the sparse VAR fit passes residual tests. The m value is chosen to first guarantee that the average p -value is no less than 0.5, and yield the smallest AIC/BIC value. This criterion suggests 417 as the best value for m , with log likelihood -193603.7 , AIC= 194437.7, and BIC=198575.3.

Chapter 4

Conclusions and Discussion

In this dissertation, two parsimonious methods are proposed for fitting periodic vector time series. The layered method is proposed as a general tactic for fitting periodic vector time series. The parsimonious PVAR method is based on the construction of a feasible periodic autocovariance function. The number of parameters it uses depends only on those arising in the model of the periodic autocovariance function, and hence remains small even when the PVAR model dimensions are big. However, it also suffers from several limitations:

- It relies heavily on a good fit of the periodic autocovariance function.
- It is limited to multivariate data with spatial relations. Consistent autocorrelation and cross-correlation behaviors are expected for the method to work, in which case a larger dimension could result in a better fit of the ACF. On the contrary, a vector series with a small dimension may see failure to fit the ACF.
- The parameter estimation is based on numerical maximization of the likelihood function, which requires constructing and solving the Yule-Walker equations of all seasons. Hence the estimation is slow, especially when d and T are both large.

As a standalone method, the parsimonious PVAR method cannot produce a good fit on the seven-city temperature data. However, as part of the layered procedure, it successfully removes the periodic correlations from the original series, and worked as effective as the non-parsimonious method (SD-PVAR) as suggested by the coherence test.

A sparse VAR model was fitted to the seasonally adjusted series from layer one. Several adjustments are made to the sparse VAR method:

- In the estimation of VAR models with linear constraints, the estimated generalized least squares (EGLS) method from Lütkepohl (1993) was used, where the white noise variance Σ was approximated by estimators either from a least squares method or from an asymptotic theorem. In practice, both these approximations are computationally demanding due to the large dimensions and the long record length. Instead of adopting these tactics of estimating Σ , we use an iterative method, which turns out to converge within three iterations, much faster and more accurate than the approximation method from Lütkepohl (1993).
- In the original sparse VAR method from Davis et al. (2016), the model fit is performed in two stages, the first of which determines to constrain all parameters to zero which correspond to some pairs of components. For example, the $(1, 3)$ entry of $\Phi(1), \dots, \Phi(p)$ of a VAR(p) model are all constrained to zero. When fitting the seven-city temperature data, this stage of model fitting always suggest no parameter set to zero.
- Davis et al. (2016) suggests to depend the choice of m (the number of non-zero parameters) solely on BIC. We suggest that the residual test results should also be considered, and a cross-validation be performed when possible.

We have shown simulation results for the periodic multivariate portmanteau test, and have indicated that the test should have bad asymptotic behavior when the period T is much larger. Thus, the test is not performed on residuals of the seven-city data since the test statistic does not converge to the proposed distribution at a series length of 60 years. However, the test is suggested to be used when the period of the series is smaller.

Bibliography

- Anderson, P. L., Y. G. Tesfaye, and M. M. Meerschaert (2007). Fourier-parma models and their application to river flows. *Journal of Hydrologic Engineering* 12(5), 462–472.
- Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N. and C. K. Wikle (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.
- Davis, R. A., P. Zang, and T. Zheng (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* 25(4), 1077–1096.
- Franses, P. H. and R. Paap (2002). *Forecasting with periodic autoregressive time-series models*. Oxford, UK, Blackwell Publishers Ltd.
- Franses, P. H. and R. Paap (2004). *Periodic time series models*. OUP Oxford.
- Fuller, W. A. (2009). *Introduction to statistical time series*, Volume 428. John Wiley & Sons.
- Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association* 75(371), 602–608.
- Lund, R. and I. Basawa (2000). Recursive prediction and likelihood evaluation for periodic arma models. *Journal of Time Series Analysis* 21(1), 75–93.
- Lund, R., H. Hurd, P. Bloomfield, and R. Smith (1995). Climatological time series with periodic correlation. *Journal of Climate* 8(11), 2787–2809.
- Lund, R., Q. Shao, and I. Basawa (2006). Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics* 48(1), 33–47.
- Lütkepohl, H. (1993). *Introduction to multiple time series*. Springer Verlag, Berlin.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Matérn, B. (2013). *Spatial variation*, Volume 36. Springer Science & Business Media.
- Porcu, E., J.-M. Montero, and M. Schlather (2012). *Advances and challenges in space-time modelling of natural events*, Volume 207. Springer Science & Business Media.
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media.
- Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley & Sons.
- Vecchia, A. (1985). Periodic autoregressive-moving average (parma) modeling with applications to water resources. *JAWRA Journal of the American Water Resources Association* 21(5), 721–730.