ANALYSIS OF POISSON COUNT TIME SERIES WITH UNKNOWN PERIODICITY

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2011

Sarah Jervis
School of Mathematics

Contents

Word count xxxxx

List of Tables

List of Figures

The University of Manchester

Sarah Jervis
Doctor of Philosophy
Analysis of Poisson count time series with unknown periodicity
November 14, 2011

There are well-established methods for estimating and removing seasonality from a stationary Gaussian time series and in recent years, models for integer-valued, non-stationary time series, such as count data, have been developed. However, little research has been done into the analysis of time series which are both non-stationary integer-valued data sets and display strong periodic behaviour. Sunspot numbers and case counts of endemic diseases are just two examples of data sets dependent upon one or more periods of unknown length, illustrating the need for models which can simultaneously capture non-stationarity and unknown periodicity.

Conditional on a latent process ε_t, let $\mathbf{Y}$ be a T-element time series of counts with distribution given by

$$y_t|\varepsilon_t \sim Po\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\varepsilon_t\right)$$

where $\{\varepsilon_t\}$ is an unobserved stationary process. This is known as a serially correlated error model or latent variable model.

In this thesis, we shall consider the properties of this model in the specific case where $\{\varepsilon_t\}$ is a series of non-negative trigonometric functions

$$\varepsilon_t = \sum_{i=1}^{k} a_i \cos^2\left(\omega_i t + \phi_i\right)$$

where $\{\omega_i\}$ are unknown frequencies and $\{\phi_i\}$ are uniform, $(0, 2\pi)$, random variables. In Chapter 1, we give examples of count data time series with strong periodicity and evaluate the unconditional mean and covariance structure of $\{y_t\}$ in the case where $k = 1$. In Chapter 2, we study several previously suggested methods for estimating a single unknown frequency from the data, including least-squares parameter estimation for non-linear models and an autocovariance estimator for the latent process as mentioned in Zeger (1988). In Chapter 3, we introduce a GLM-based mean parameter estimator and a Discrete Fourier transform (DFT) based estimator for ω, then establish consistency of both and asymptotic normality of the former. In Chapter 4, extensions of the single-period model to cover data sets with extra short-term dependence or several periods are introduced and the parameter estimators and their properties are extended to cover these models. Finally, in Chapter 5, we study a data set of twenty-three years of measles case counts using results established in previous chapters and consider the wide applications for extensions of the latent variable model to analyse longitudinal, multivariate or spatial-temporal data.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

I would like to thank both my supervisors, Professor T. Subba Rao, for suggesting a highly interesting and original topic and getting me started, and Dr P. Neal for his guidance and encouragement throughout my research. I would also like to thank the staff and other PhD students in the probability and statistics department for many interesting discussions and comments and their helpful suggestions over the years. I would especially like to thank my family for all the support they have given me. This work was funded by the Engineering and Physical Sciences Research Council, (EPSRC).

Chapter 1

Introduction

1.1 Periodic Poisson count data

The focus of this thesis is the estimation of periodicity as part of a parametric model for time series of Poisson counts. Many examples of time series count data, such as the hourly number of hospital admissions, will only be influenced by predictable periods such as a 24-hour, seasonal or yearly cycle. Emergency department admissions measured hourly might be low at some time during the day when many people are in a relatively safe work environment or late at night when most are sleeping, and high in the evening or early morning due to factors such as poor outdoor visibility and increased travel. Over longer periods of time, one might expect hospital admissions to vary according to the day of the week and the time of the year. However there are other events, such as annual cases of an endemic disease, where the frequency of any underlying oscillatory pattern might be much harder to predict but of significant importance. To illustrate some of the wide variety of situations which produce periodic Poisson time series, three rather different sets of count data are displayed graphically below.

Figure 1.1 shows the monthly numbers of serious car crashes in the UK, which is a classic example of a discrete time series with predictable periodicity and step-changes due to outside factors. Although quite "jagged", the data plot displays a clear twelve-month period, a significant decrease from 1983 when wearing seatbelts in the front seat was made compulsory and a steady increase until 1974 when the movement of petrol prices changed from a slow decrease to a rapid rise. The annual cycle peaks around December every year, as one would predict after considering factors such as visibility, road grip and changes in car travel rates due to weather and socialising patterns. In contrast to car crashes, annual lynx trappings or averaged sunspot numbers are not expected to be dependent on any highly seasonal factors like weather. Thus, although the graphs in Figures 1.2 and 1.3 all suggest that both data sets have a regular cycle, one can not predict the periodicity of either series in advance. Astronomical events within the solar system are far more likely to be influenced by the orbits of Jupiter and other gas giants than little rocky planets, so the periodicity of the sunspots is unlikely to be an integer or even a rational number of years, although it is easy to infer from the data plots that it is close to eleven years. The nine to ten-year periodicity apparent in the yearly lynx counts is likely to be one part of a long-term "predator-prey" cycle, where populations of one or more primary organisms and their consumers rise and fall almost in synchrony apart from a short time-lag. Without more information, one does not know whether the lynx are the predators or the prey, with the trappers acting as predators in the latter case.

Both the sunspot and the lynx counts are good examples of data sets where precise estimation of the periodicity would be useful. Both are very periodic, yet the non-integer length of the sunspot cycle and the varying lengths of time between consecutive peaks in the lynx counts would make estimates of either period directly from the data rather imprecise. The relative magnitudes of peak observations in both series also suggest that there could be secondary, longer-term periods which might also be of interest.

Figure 1.1: Plot of the number of drivers killed or seriously injured in car accidents, 1969-1984

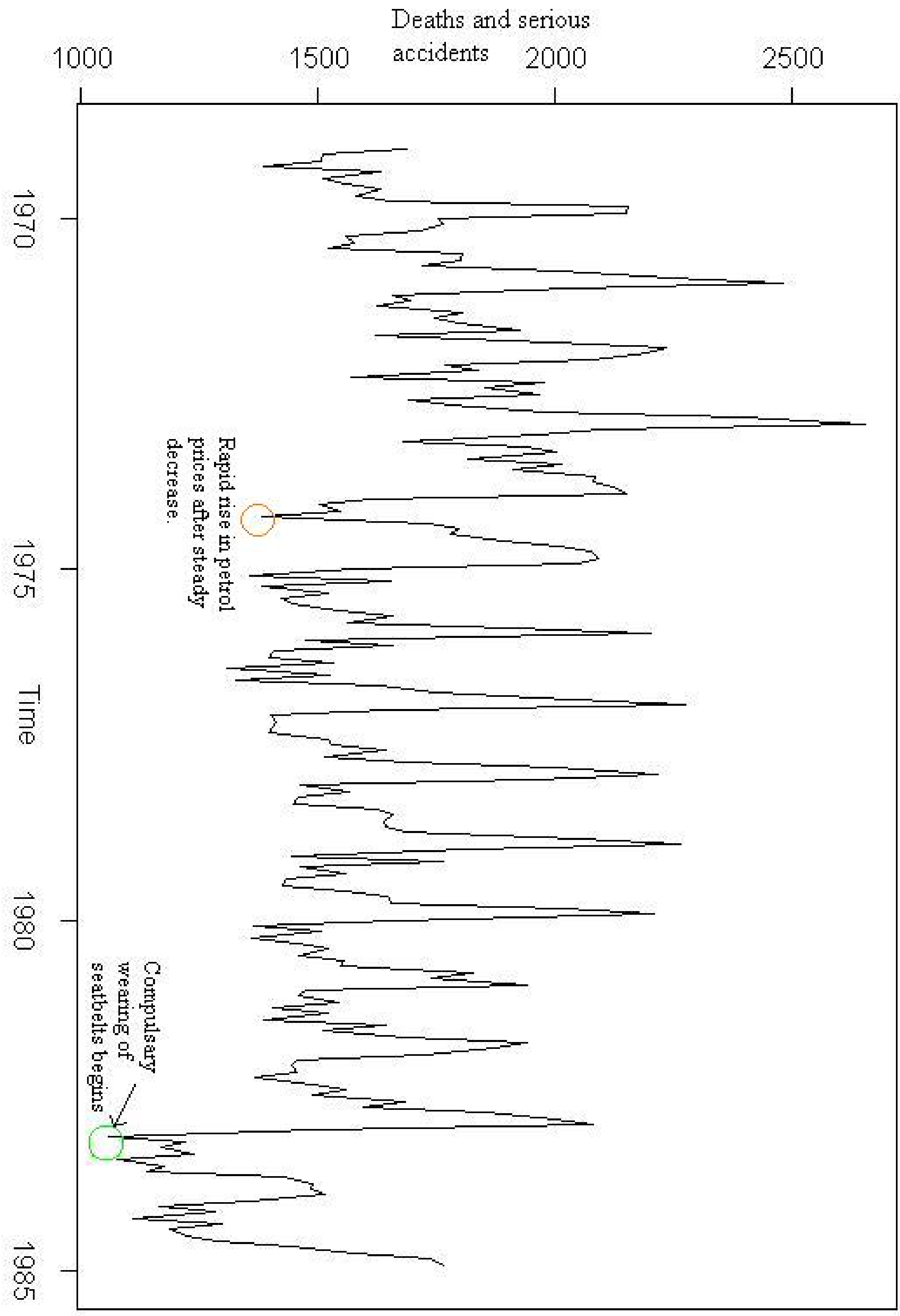


Figure 1.2: Averaged sunspot numbers, measured yearly 1700-1988 and monthly 1749-1997

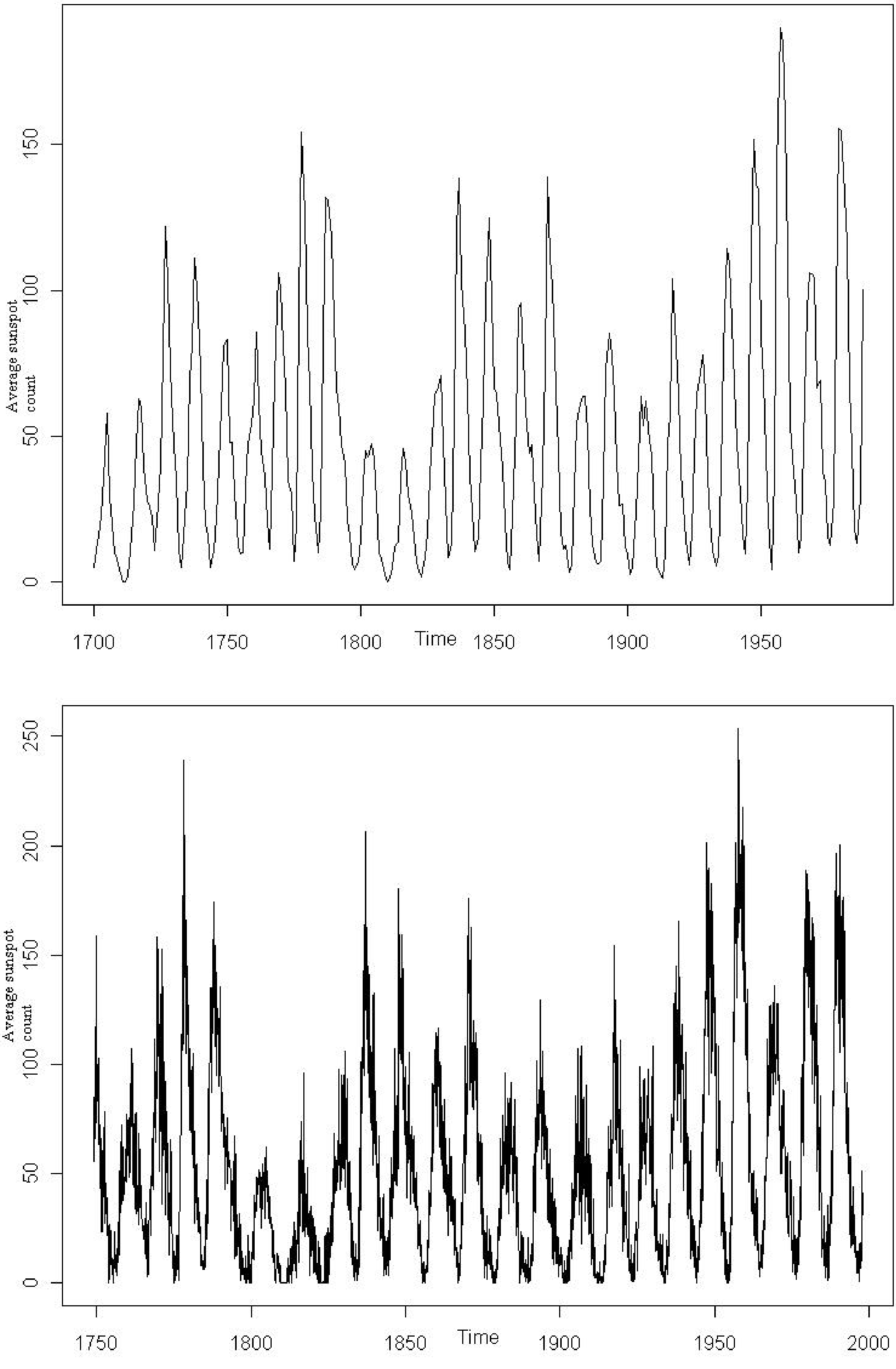
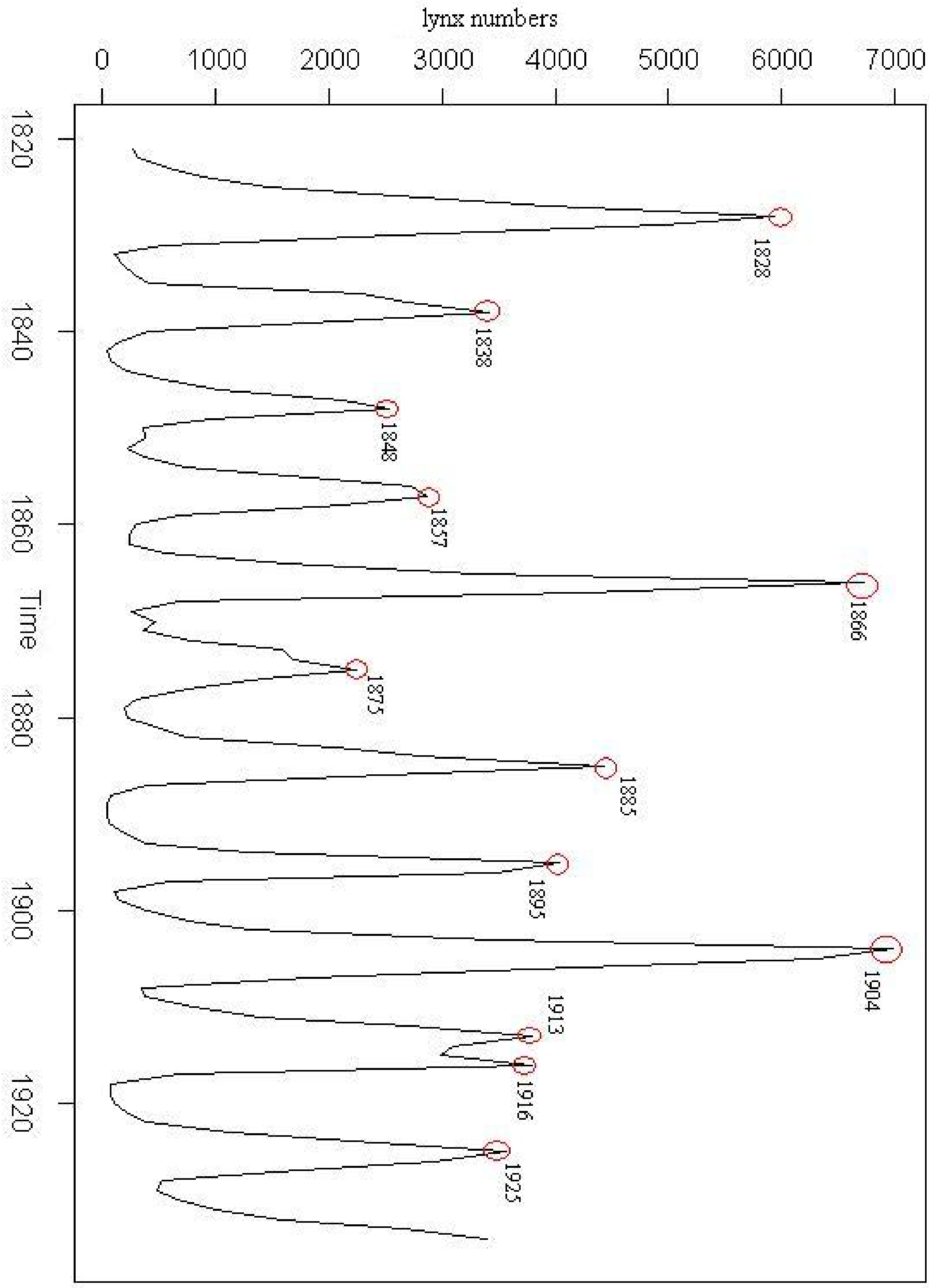


Figure 1.3: Annual Canadian lynx trappings, 1821 to 1934



1.2 A review of time series analysis

In this section we shall give a summary of methods for time series modelling. First there shall be a brief introduction to the very widely-used ARMA(p,q) model, as described in Priestley (1981) and Brockwell and Davis (1991), then a discussion of the analyses of non-homogeneous Poisson processes developed recently.

1.2.1 The ARMA model

The best known type of model for time series is the autoregressive moving-average model of order p and q, denoted the ARMA(p,q) model. A time series $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ is said to be an ARMA(p,q) process if for every t,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} \qquad (1.2.1)$$

where $\{e_t, t - 0, \pm 1, \pm 2, \ldots\}$ is a sequence of independent and identically distributed (iid) random variables, often Gaussian, with mean μ and variance σ^2. Let B denote the backward shift operator, defined as $B^j X_t = X_{t-j}$ for $j = 0, \pm 1, \pm 2, \ldots$. The ARMA(p,q) process can then be re-written as

$$\phi(B) X_t = \theta(B) e_t \qquad (1.2.2)$$

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$ are referred to as the autoregressive and the moving average polynomials of the difference equation (1.2.1) respectively. The ARMA(p, q) process is said to be causal if all the roots of the characteristic equation $\phi(B) = 0$ lie outside the unit circle. Similarly, the model is said to be invertible if the roots of $\theta(B) = 0$ lie outside the unit circle.

The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be weakly stationary if

(i) $E\left(X_t^2\right) < \infty$ for all $t \in \mathbb{Z}$

(ii) $E\left(X_t\right) = m$ for all $t \in \mathbb{Z}$

(iii) $Cov\left(X_t, X_s\right) = Cov\left(X_{r+t}, X_{r+s}\right) = \gamma_X\left(|t - s|\right)$ for all $r, s, t \in \mathbb{Z}$.

Any time series $\{X_t\}$ which can be expressed as a causal and invertible ARMA process can be written as

$$X_t = e_t + \sum_{i=1}^{\infty} \psi_i e_{t-i} \tag{1.2.3}$$

where $\sum_{i=1}^{\infty} \psi_i < \infty$. As $\{e_t\}$ is stationary, it follows that $\{X_t\}$ is weakly stationary. ARMA processes are usually assumed to be weakly stationary, as this facilitates model selection, parameter estimation and forecasting. However, in many cases, observed time series do not appear to have constant mean or variance. It is therefore common practice to first remove any trend or seasonality from a data set. These are usually expressed as components in the representation

$$X_t = m_t + s_t + Y_t \tag{1.2.4}$$

where m_t is a slowly increasing or decreasing function and s_t is a function of a known period d, known as the "trend component" and the "seasonal component" respectively. and Y_t is a stationary process referred to as a "random noise component". If seasonal or random noise fluctuations are increasing over time, a prior transformation such as logarithm or square-root is often used to improve the fit of the data to a model of the form (1.2.4). Similarly, an exponential or quadratic transform can be used for data with decreasing fluctuations. When there appears to be no seasonal component, the trend component can be estimated by fitting a n^{th}-order time polynomial $a_0 + a_1 t + \ldots + a_n t^n$ by least-squares estimation, isolated as a moving-average function $\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j}$ or eliminated by repeatedly differencing the data until the observations resemble a stationary process $\{W_t\}$.

Differencing, using the single lag-d difference operator ∇_d defined as

$$\nabla_d X_t = X_t - X_{t-d} \tag{1.2.5}$$

can similarly be used to remove the seasonal component. The trend of the resulting seasonless process

$$X_t^* = \nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d} \tag{1.2.6}$$

can then be eliminated using one of the three methods described previously. An alternative method is to estimate the seasonal component s_t. Suppose the data set $\{x_1, \ldots, x_N\}$ can be divided into j subsets $\{x_1, \ldots, x_d\}, \{x_{d+1}, \ldots, x_{2d}\}, \ldots, \{x_{(j-1)d+1}, \ldots, x_{jd}\}$. First, a preliminary moving-average estimate of the trend is calculated.

$$\hat{m}_t = \begin{cases} \frac{1}{d}\left(0.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + 0.5x_{t+q}\right) & d = 2q \\ \frac{1}{d}\left(x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + x_{t+q}\right) & d = 2q+1 \end{cases}. \quad (1.2.7)$$

This estimate is considered optimal in terms of eliminating the effects of both the seasonal and the random components. For each $1 \leq k \leq d$, s_k is then estimated as

$$\hat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i \quad (1.2.8)$$

where

$$w_k = \frac{1}{j-1}\sum_{i:q<k+id<dj-q} \left(x_{k+id} - \hat{m}_{k+id}\right). \quad (1.2.9)$$

The final step is to re-estimate or eliminate the trend from the deseasonalised data set

$$d_t = x_t - \hat{s}_t \quad (1.2.10)$$

by one of the methods described for non-seasonal data.

The ARMA model is well developed, flexible and easy to implement. However, its feasibility depends on the data being weakly stationary as well as continuous and analysis is less straightforward for non-Gaussian random variables. The Poisson counts we are interested in will always be non-stationary due to the underlying harmonic behaviour as well as any linear regressors such as time or location. The methods suggested for detrending are clearly not applicable to discrete data sets, as neither x_t/q or the time polynomial parameters are necessarily integer-valued. The de-seasonalised data set $x_t - \hat{s}_t$ would also contain non-integer values and a reasonable idea of the underlying periodicity is needed to estimate the seasonal component s_t in the first place. The white noise terms $\{e_t\}$ would not follow a Gaussian distribution, so overall the standard ARMA model will have limited usefulness.

1.2.2 Models for non-homogeneous Poisson processes

A time-inhomogeneous Poisson Process is a counting process $\{N(t) : t \geq 0\}$, characterised by a time-varying intensity function $\lambda(t)$ such that the number of occurrences $N(b) - N(a)$, in time $(a, b]$, follows a Poisson distribution

$$P\left((N(b) - N(a)) = k\right) = e^{-\lambda_{a,b}} \frac{\lambda_{a,b}^k}{k!} \tag{1.2.11}$$

where $\lambda_{a,b} = \int_a^b \lambda(t).dt$ Much work has been done recently in modeling time-inhomogeneous Poisson Processes, usually via estimation of the intensity function (see Willis (1964), Kuhl et al. (1997),(1998), Helmers et al. (2003),(2005) and others). These developments might, at first, appear to be of use in our area of interest, the analysis of Poisson count time series, but have severe limitations. One such handicap is the importance of extra measurements in the Poisson process models, namely the occurrence or arrival times of the events. Such data would not be possible to gather for many Poisson counts, due to the counted events being continuous changes in state rather than discrete occurrences. At what precise time does an individual catch a disease or another sunspot appear on a continuous time scale? The developments in processes have other limitations besides their dependence on arrival times. Many of the models are nonparametric and among those which have been developed for periodic data, most are specified only for data with a known periodicity or for cyclic data. Thus the Poisson process models are not readily extendable to a parametric model for a sample of Poisson counts influenced by unknown frequencies or any non-cyclic time-varying regressors such as a trend function. The few papers written on estimation of frequency, such as Hannan(1974), Vere-Jones(1982) and Bebbington et.al.(2004), are also restricted to series with purely cyclic behaviour. The methods used could consequently be very sensitive to nonlinear link functions or the changes in variance arising from underlying monotonic behaviour. However, as a series of Poisson counts can often be thought of as a Poisson process with the number of occurrences only recorded at discrete time points, some of the methods and models studied in this thesis could be applicable to Poisson processes. For example, periodicity estimators developed for Poisson counts might be applicable to some Poisson processes.

1.3 The basic model with trigonometric latent process

We have discussed the limitations of the popular ARMA model and methods developed for Poisson processes for the analysis of periodic Poisson count series. Most other recent developments are extensions of the ARMA model and include the ARCH, GARCH, GLARMA and INAR models. These are known as observation- driven models as all observations are specified as functions of previous observations. We shall initially be considering a purely parametric model, similar to that first proposed in Zeger(1988), with correlation and deviations from a log-linear mean described by an underlying latent process.

Suppose that we have a series of count data random variables $Y = (y_1, y_2, \ldots, y_T)$. The hypothesised model for the data is $\{y_t|\phi\} \sim Po\left(\exp\left(\mathbf{x}_t^T \boldsymbol{\vartheta}\right) e_t^2\right)$ where ϕ is a random variable with a $U(0, 2\pi)$ distribution, $\boldsymbol{\vartheta}$ is a parameter vector, $\mathbf{x}_t^T$ is the design vector at timepoint t and $e_t^2 = \cos^2(\omega t + \phi)$ for an unknown frequency ω. Then

$$\begin{aligned} E(y_t \mid \phi) &= E(y_t \mid \phi) = \exp\left(\mathbf{x}_t^T \boldsymbol{\vartheta}\right) \cos^2(\omega t + \phi) = V(y_t \mid \phi) \\ Cov(y_t, y_{t+s} \mid \phi) &= 0 \quad \text{for all } s \neq 0 \\ E(y_t) &= E(E(y_t \mid \phi)) = 0.5 \exp\left(\mathbf{x}_t^T \boldsymbol{\vartheta}\right) E\left(2\cos^2(\omega t + \phi)\right) \end{aligned}$$

where

$$\begin{aligned} E\left(2\cos^2(\omega t + \phi)\right) &= \frac{1}{2\pi} \int_0^{2\pi} 1 + \cos(2\omega t + 2\phi)\, d\phi = \frac{1}{2\pi} \left[1 + \frac{1}{2} \sin(2\omega t + 2\phi)\right]_0^{2\pi} \\ &= \frac{2\pi + \frac{1}{2}\sin(2\omega t) - \frac{1}{2}\sin(2\omega t)}{2\pi} = 1 \\ \text{Thus } E(y_t) &= \frac{\exp\left(\mathbf{x}_t^T \boldsymbol{\vartheta}\right)}{2} = \mu_t, \text{ say.} \end{aligned} \tag{1.3.1}$$

The model can then be rewritten as $\{y_t|\phi\} \sim Po(\mu_t \varepsilon_t)$ where $\mu_t = \frac{1}{2}\exp\left(\mathbf{x}_t^T \boldsymbol{\vartheta}\right)$ and $\varepsilon_t = 2e_t^2$.

By the theorem of total variance,

$$\begin{aligned} V(y_t) &= V(E(y_t \mid \phi)) + E(V(y_t|\phi)) = V(\mu_t \varepsilon_t) + E(y_t) \\ &= \mu_t^2 E\left((\varepsilon_t - 1)^2\right) + \mu_t \end{aligned}$$

where

$$
\begin{aligned}
E\left((\varepsilon_t - 1)^2\right) &= E\left(\cos^2(2\omega t + 2\phi)\right) = \frac{1}{2}E\left(1 + \cos\left(4\left(\omega t + \phi\right)\right)\right) \\
&= \frac{1}{4\pi}\int_0^{2\pi}\left[1 + \cos\left(4\left(\omega t + \phi\right)\right)\right] d\phi = \frac{1}{4\pi}\left[1 + \frac{1}{4}\sin\left(4\left(\omega t + \phi\right)\right)\right]_0^{2\pi} \\
&= \frac{2\pi + \frac{1}{4}\sin(4\omega t) - \frac{1}{4}\sin(4\omega t)}{4\pi} = \frac{1}{2} \\
\text{Thus } V(y_t) &= \mu_t + \frac{\mu_t^2}{2}. \qquad (1.3.2)
\end{aligned}
$$

This unconditional variance is always larger than the expectation of y_t, a useful property when many real-world incidences of count data appear overdispersed. For all $s \neq 0$,

$$
\begin{aligned}
Cov(y_t, y_{t+s}) &= Cov\left(E(y_t \mid \phi), E(y_t \mid \phi)\right) + E\left(Cov(y_t, y_{t+s} | \phi)\right) \\
&= E\left((\mu_t\varepsilon_t - \mu_t)(\mu_{t+s}\varepsilon_{t+s} - \mu_{t+s})\right) + E(0) \\
&= \mu_t\mu_{t+s}E\left((\varepsilon_t - 1)(\varepsilon_{t+s} - 1)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
E\left((\varepsilon_t - 1)(\varepsilon_{t+s} - 1)\right) &= E\left(\cos\left(2\left(\omega t + \phi\right)\right)\cos\left(2\left(\omega t + \omega s + \phi\right)\right)\right) \\
&= \frac{1}{2}E\left(\cos(4\omega t + 2\omega s + 4\phi) + \cos(2s\omega)\right) \\
&= \frac{1}{4\pi}\int_0^{2\pi}\left[\cos(4\omega t + 2\omega s + 4\phi)\cos(2s\omega)\right] d\phi \\
&= \frac{1}{4\pi}\left[\cos(2s\omega) + \frac{1}{4}\sin(4\omega t + 2\omega s + 4\phi)\right]_0^{2\pi} = \frac{\cos(2s\omega)}{2} \\
\text{giving } Cov(y_t, y_{t+s}) &= \frac{\mu_t\mu_{t+s}\cos(2s\omega)}{2} \quad \text{for all } s \neq 0. \qquad (1.3.3)
\end{aligned}
$$

Looking at the model $y_t \sim Po(\mu_t\epsilon_t)$ which is discussed in Davis et al.(2000), we can see that our model is actually identical to that used there, prior to any assumptions on the distribution or autocorrelation structure of ε_t. Our variable term ε_t, in other words, is a weakly stationary process with mean 1, variance σ^2 $(= 1/2)$ and correlation function $\rho(s)$ $(= \cos(2\omega s))$. The unconditional mean μ_t is of exponential type, equal to $\exp\left(\mathbf{x}_t^T\boldsymbol{\vartheta} - \log[2]\right) = \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)$. These similarities suggest that many of the methods used in Davis et al.(2000) could be still be applicable to the analysis of periodic count data when the underlying frequencies are unknown.

1.4 Outline of the thesis

Now that we have our basic model with its conditional Poisson distribution and unconditional trigonometric covariance structure, the rest of the thesis will be divided between studying possible methods for parameter estimation, extensions to the basic model, and real data analysis. In Chapter 2, we shall trial several methods for estimating the periodicity directly from a data set and Chapter 3 will consist of proofs of the asymptotic properties of empirically promising parameter estimators for the basic model. In Chapter 4, we shall study extensions to the basic models obtained by adding a secondary latent process, in particular an ARMA process. Finally in Chapter 5 we analyse UK measles case count data 1944-1966.

Chapter 2

Methods for estimating ω directly from the data

2.1 Introduction

There are multiple methods for estimating the parameters of any linear model for uncorrelated Poisson counts and, more recently, some for correlated count data. Thus our area of interest is the estimation of the frequency parameter ω. Three different approaches will be examined in this chapter - estimating ω as one of several parameters of a nonlinear model, identifying an unknown period from graphical plots of the data and transformations of the data, and the potential for estimating the frequency or period from the estimator of the latent process autocovariance function suggested in Zeger (1988). The poor empirical evidence for the first two approaches will be revealed and possible improvements upon Zeger's covariance estimator are discussed.

2.2 The nonlinear model

Theoretically ω can be estimated as one of $p+2$ parameters of a nonlinear model for $\mathbf{Y}$. Following the methods proposed in Gallant (1987), the sum of squares $S(\mathbf{\Theta}) = \sum_{t=1}^{T} (y_t - \mu_t \varepsilon_t)^2$ (conditional on ϕ) is first replaced by a second-order Taylor expansion. The vector $\hat{\mathbf{\Theta}}$ which minimises this Taylor series approximation is taken as the LSE of $\mathbf{\Theta}_0 = \left(\boldsymbol{\theta}^T, \omega, \phi\right)$ although this may be a local rather than a global minimising vector and accuracy can depend on the closeness of the approximation. Denoting $E(\mathbf{Y}|\phi)$ by $\mathbf{f}(\mathbf{\Theta})$ and $\frac{\partial}{\partial \mathbf{\Theta}}(\mathbf{f}(\mathbf{\Theta}))$ by $\mathbf{F}(\mathbf{\Theta})$, $\hat{\mathbf{\Theta}}$ is the vector satisfying $\mathbf{F}^T\left(\hat{\mathbf{\Theta}}\right)\left[\mathbf{Y} - \mathbf{f}\left(\hat{\mathbf{\Theta}}\right)\right] = 0$. Substituting in a first-order Taylor approximation of $\mathbf{f}\left(\hat{\mathbf{\Theta}}, t\right)$ and rearranging terms then gives the iterative solution

$$\hat{\mathbf{\Theta}}_n = \hat{\mathbf{\Theta}}_{n-1} + \left[\mathbf{F}^T\left(\hat{\mathbf{\Theta}}_{n-1}\right)\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right]^{-1}\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\left[\mathbf{Y} - \mathbf{f}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right]. \tag{2.2.1}$$

This is usually referred to as the *Gauss-Newton* method. A second-order approximation of $S(\mathbf{\Theta})$ gives the iterative solution

$$\hat{\mathbf{\Theta}}_n = \hat{\mathbf{\Theta}}_{n-1} + \left[-\mathbf{H}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right]^{-1}\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\left[\mathbf{Y} - \mathbf{f}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right] \tag{2.2.2}$$

where $\mathbf{H} = \dfrac{\partial^2 S(\mathbf{\Theta})}{\partial \mathbf{\Theta} \partial \mathbf{\Theta}^T}$

This is the *Newton-Raphson* method. To ensure that the successive estimates are converging to a minimising value in the sense that $S(\mathbf{\Theta}_n) \leq S(\mathbf{\Theta}_{n-1})$, it is common to use a rescaled shift rather than the whole shift. If we express either of the above methods as $\hat{\mathbf{\Theta}}_n = \hat{\mathbf{\Theta}}_{n-1} + \hat{\mathbf{D}}_{n-1}$, it is often preferable to use the iterative step $\tilde{\mathbf{\Theta}}_n = \tilde{\mathbf{\Theta}}_{n-1} + \lambda_n \hat{\mathbf{D}}_{n-1}$ where λ_n is the largest number in $(0, 1]$ such that $S\left(\tilde{\mathbf{\Theta}}_n\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n-1}\right)$. In general, trialing successive numbers in the sequence $\left\{1, 0.9, 0.8, 0.7, 0.6, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^n}, \ldots\right\}$, until a parameter vector with a smaller sum-of-squares is found, is thought to be nearly always adequate. We shall refer to this as "λ-adjustment". There is also an alternative rescaling method explained in Marquardt (1963). The matrix $\hat{\mathbf{A}}_{n-1} = \mathbf{F}^T\left(\hat{\mathbf{\Theta}}_{n-1}\right)\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)$ or $-\mathbf{H}\left(\hat{\mathbf{\Theta}}_{n-1}\right)$ is adjusted by the addition of $\delta \mathbf{I}_p$ with δ chosen at each iteration as follows: Take a small starting value such as 0.01 for δ_0 and a value close to 1 such as 1.1 for v. Compute

$$\begin{aligned}
\tilde{\mathbf{\Theta}}_{n1} &= \hat{\mathbf{\Theta}}_{n-1} + [\mathbf{A}_{n-1} + \delta_1 \mathbf{I}_p]^{-1}\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\left[\mathbf{Y} - \mathbf{f}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right] \\
\tilde{\mathbf{\Theta}}_{n2} &= \hat{\mathbf{\Theta}}_{n-1} + [\mathbf{A}_{n-1} + \delta_2 \mathbf{I}_p]^{-1}\mathbf{F}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\left[\mathbf{Y} - \mathbf{f}\left(\hat{\mathbf{\Theta}}_{n-1}\right)\right]
\end{aligned} \tag{2.2.3}$$

where $\delta_1 = \delta_0$ and $\delta_2 = \frac{\delta_0}{v}$. Calculate $S\left(\tilde{\mathbf{\Theta}}_{n1}\right)$ and $S\left(\tilde{\mathbf{\Theta}}_{n2}\right)$

If $S\left(\tilde{\mathbf{\Theta}}_{n-1}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n1}\right)$ and $S\left(\tilde{\mathbf{\Theta}}_{n-1}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n2}\right)$, then compute successive $\tilde{\mathbf{\Theta}}_{nr}$ using $\delta_r = \delta_0 v^r \ (r = 1, 2, \ldots)$ until an estimate $\tilde{\mathbf{\Theta}}_{np}$ with a smaller sum of squares than $\tilde{\mathbf{\Theta}}_{n-1}$ is found. $\tilde{\mathbf{\Theta}}_{np} = \hat{\mathbf{\Theta}}_n$

If $S\left(\tilde{\mathbf{\Theta}}_{n2}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n1}\right)$ and $S\left(\tilde{\mathbf{\Theta}}_{n2}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n-1}\right)$, then $\hat{\mathbf{\Theta}}_n = \tilde{\mathbf{\Theta}}_{n2}$

If $S\left(\tilde{\mathbf{\Theta}}_{n1}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n-1}\right)$ and $S\left(\tilde{\mathbf{\Theta}}_{n1}\right) \leq S\left(\tilde{\mathbf{\Theta}}_{n2}\right)$, then $\hat{\mathbf{\Theta}}_n = \tilde{\mathbf{\Theta}}_{n1}$

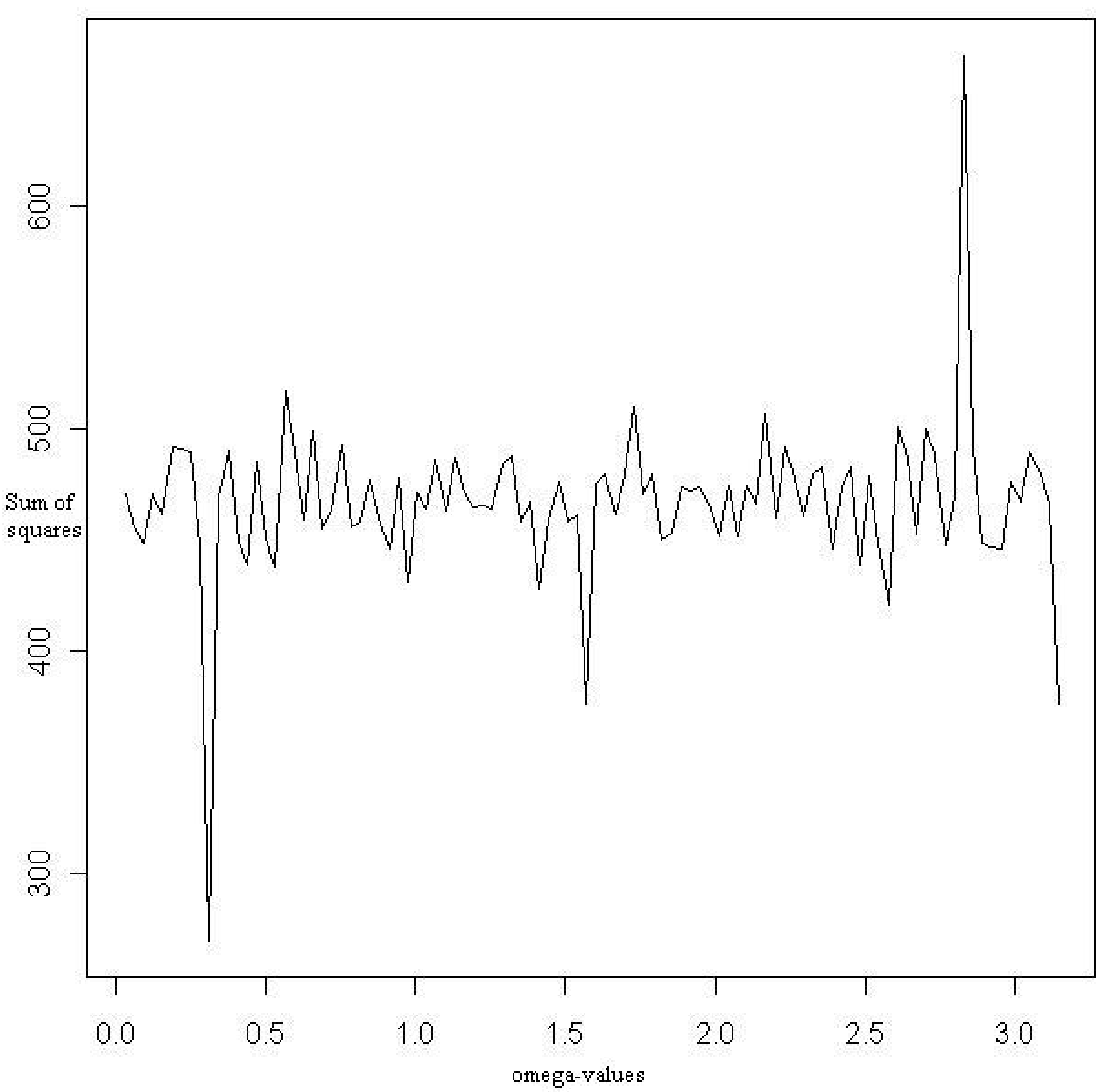
The ordinary least-squares (OLS) approach, as with linear models, is optimal for Normal data, not Poisson counts. An improved algorithm could be generated by using weighted least squares instead, with weight matrix $\mathbf{W} = diag\left(\frac{1}{\mu_t \varepsilon_t}\right)$. Weighting the squares by the inverse of the conditional variance gives a consistent estimator $\hat{\mathbf{\Theta}}$ of $\mathbf{\Theta}$ with asymptotic distribution $N_p\left(\mathbf{\Theta}, \left[\mathbf{F}^T\mathbf{W}\mathbf{F}\right]^{-1}\right)$ (Gallant 1987, Chapter 4). The ordinary least squares estimator may however be inconsistent. Programs for performing the OLS Gauss-Newton algorithm with λ-adjustment (G-N lambda), the OLS Newton-Raphson algorithm with λ-adjustment (N-R lambda), the WLS Gauss-Newton algorithm with λ-adjustment (Weighted G-N lambda) and the OLS Gauss-Newton algorithm adjusted using Marquardt's method (G-N Marquadt) have been written for the statistical package $\mathbb{R}$ and tested on several simulated data sets. All appear to produce similar results, with the Newton-Raphson algorithm being the fastest but least robust and weighted Gauss-Newton is rather slower. For the simple model $\{y_t \mid \phi\} \sim Po\left(\exp\left(\alpha + \beta t\right)\cos^2\left(\omega t + \phi\right)\right)$ all four algorithms appear to estimate β quite accurately and robustly, α very poorly (with iterations often diverging from rather than converging to the true value) and estimates of ω and ϕ are accurate only when the starting values are very close to the true values. The results, all from starting values $(\hat{\alpha}, \hat{\beta}, \hat{\omega}, \hat{\phi})$=(1,0,0.5,0.5), for four simulations of size 500 are displayed in Table 2.1. The accurate estimates of the trend compared with the other parameters is due to β being the parameter with most influence over the moments of the data. Small changes in the value of β have greater effect on the sum of squares than the other parameters. Plots of the sum of squares for various values for α, ω and ϕ are displayed below. Together they suggest another explanation for the comparatively poor estimation of the other parameters. The SS-minimisers of α and ϕ differ from their true values when ω does, while the SS-plot when ω is varying has multiple local minima. Both the Gauss-Newton and the Newton-Raphson method are algorithms designed to find the vector $\mathbf{\Theta}_0$ which satisfies $\frac{\partial}{\partial \mathbf{\Theta}}\left(S\left(\mathbf{\Theta}\right)\right) = 0$ but

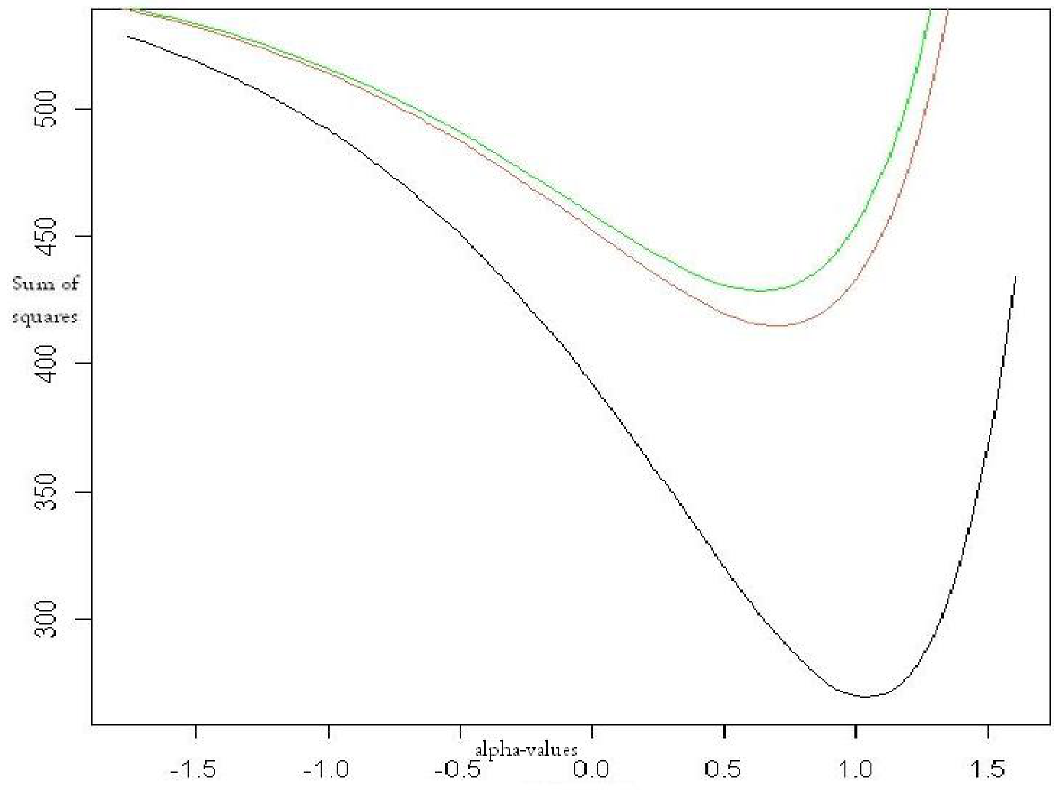
Table 2.1: Parameter estimates from four non-linear least-square algorithms

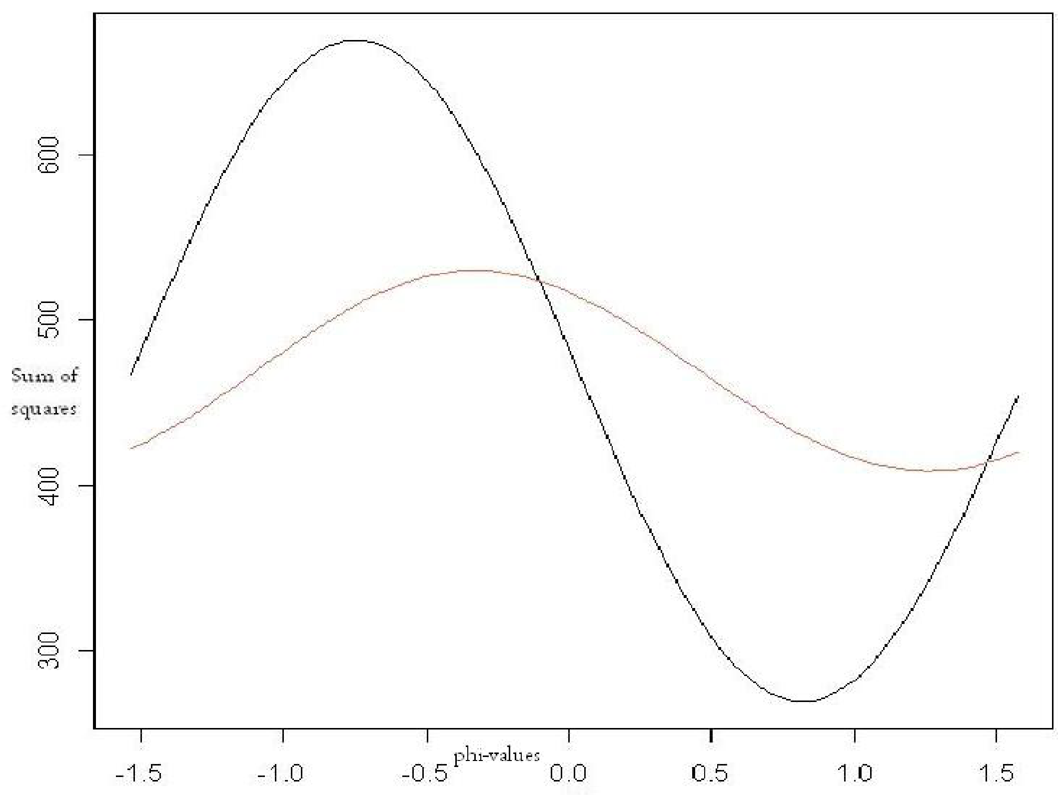
True	G-N lambda	N-R lambda	G-N Marquad	Weighted G-N
1	-0.32235892	-0.32235892	0.999994272	-4.2933452
0.005	0.006351296	0.006351296	0.004575719	0.0156511
0.314159265	0.498948248	0.498948248	0.499878807	0.4883015
0.785398163	1.26738712	1.26738712	1.570796327	1.5707963
1	0.8985067	0.8985067	0.500680567	0.676854059
-0.005	-0.004409429	-0.004409429	-0.004332558	-0.000610715
0.314159265	0.503375423	0.503375423	0.509408731	0.499421269
0.785398163	1.08654907	1.08654907	1.570796327	1.570796327
1	-0.31676643	-0.31676643	0.999997824	-0.935974695
0.005	0.00624283	0.00624283	0.004541366	0.007231628
0.314159265	0.49624497	0.49624497	0.499799555	0.446885157
0	0.43770272	0.43770272	1.570796327	1.570796327
1	0.995311833	0.995311833	0.611020959	0.61029996
-0.005	-0.004944706	-0.004944706	-0.004872163	-0.001052844
0.314159265	0.510306002	0.510306002	0.499053744	0.498547234
0	0.180400024	0.180400024	1.570796327	1.570796327

there is no way of distinguishing whether such a vector is of the true values or a set of locally minimising or maximising (LMM) values. Consequently ω is quickly fixed at a point very close to its starting value due to the multitude of minima. Then the intercept and phase shift converge to values that minimise the sum of squares for the local minima of ω. It is probably this which is also accountable for the ϕ estimated by the ordinary Gauss-Newton algorithm with Marquardt adjustment and the weighted Gauss-Newton algorithm with λ-adjustment converging to identical values to each other in all four simulations - by chance these two algorithms estimate ω as the same LMM in each simulation, which then causes the same mis-estimation of ϕ. This local minimisation problem is the main drawback in estimating ω as one of several parameters in a nonlinear model. None of the algorithms tried so far appear robust enough to give reliable estimates of the frequency when the starting value is not itself a good estimate.

Figure 2.1: Plots of the unweighted sum-of-squares for a simulation from the model $y_t \sim Po\left(e^{1+0.005*t}\cos^2\left(\pi t/10+\pi/4\right)\right)$ over a range of values of frequency, intercept and phase shift. The black line in all three shows the SS when all other parameters are at their true values, the red lines when ω is fixed at $\pi/11$ and the green line when ϕ is fixed at $\pi/2$




Sum of squares
500
450
400
350
300
alpha-values
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5


Sum of squares
600
500
400
300
phi-values
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

2.3 ω as a pattern in data plots

An alternative approach to estimating ω is to examine transforms of the data, which may be chosen to exaggerate the effects of the trigonometric elements in the conditional means. For various simulated data sets, the raw data, their logarithms and their differenced logarithms were the first such transformed data sets to be examined. Some of the corresponding sample plots did appear to have a regular periodic structure, being similar in shape to a sine or cosine function with or without a linear or exponential trend. This degree of periodicity is apparent in the data plot for the simulation with conditional mean $\exp(1 + 0.02)\cos^2(2\pi t/40)$ and is pronounced in the plots of the logarithm and differenced logarithm. However all these plots were among those of data based on models with positive trends large enough to cause unrealistic "explosive" behavior in the simulations. In the first graph below, the counts rapidly grow from between 0 and 10 to around 50000. It is highly unlikely that any of the real-life situations we might be interested in, such as astronomical events or pandemic outbreaks, would actually show such rapid growth. Although the "spikiness" of the second and third sets of plots, of 500 and 150-element simulations with very small negative trends, do suggest that the data has some periodicity, the patterns are too irregular to infer what the period is with any accuracy.

The sample periodogram was the next transform to be investigated. Variations of the periodogram, which is the modulus of a discrete Fourier transform of the data, are the method of choice in frequency estimation papers such as Hannan(1973), mentioned briefly in the introduction. This method appeared to be slightly better suited to our purpose in some respects than logarithms or difference operators of the data, in that the positive results were not confined to simulations of data with unrealistic mean growth. Distinguishable peaks near 0.05, indicating a periodicity of $0.05^{-1} = 20$, can be seen in the periodograms for all three simulated data sets. The high incidence of insignificant peaks generated by random variation, particularly in the third periodogram (for a relatively small simulated data set), suggests that the periodogram is far from ideal as a sample statistic. It is unlikely that one could infer from a periodogram with two distinct peaks which peak corresponded to the true periodicity of the data set or indeed whether the data set is dependent on more than one frequency. To summarise, the sample periodogram has several pros and cons as a sample statistic, namely

- The periodogram does usually display noticeable peaks corresponding to the true frequency and sometimes multiples of it, particularly for large data sets with small trend functions.

- The periodogram is actually only defined as a statistic for frequency estimation on samples of stationary data. Even then it is not a consistent estimator, as proven in detail in Priestley (1981). The type of data we are studying here is nonstationary either conditionally or unconditionally. Thus even when the periodogram of a sample does have one or more noticeable peaks, there is no theoretical evidence that the frequencies corresponding to those peaks are of significance. Ideally we would have theoretical as well as empirical evidence for a sample statistic.

- There is empirical evidence that random variation alone can cause noticeable peaks, which might be indistinguishable from true frequencies, particularly when sample size is quite small.

Figure 2.2: Plots of three simulated data sets, on the left, and their logarithms, on the right. Simulations Y1, Y2 and Y3 have conditional means $\exp(1 + 0.02t)\cos^2(2\pi t/40)$, $\exp(1 - 0.001t)\cos^2(2\pi t/40)$ and $\exp(1 - 0.001t)\cos^2(2\pi t/40)$. Sample sizes are 500, 500 and 150 respectively

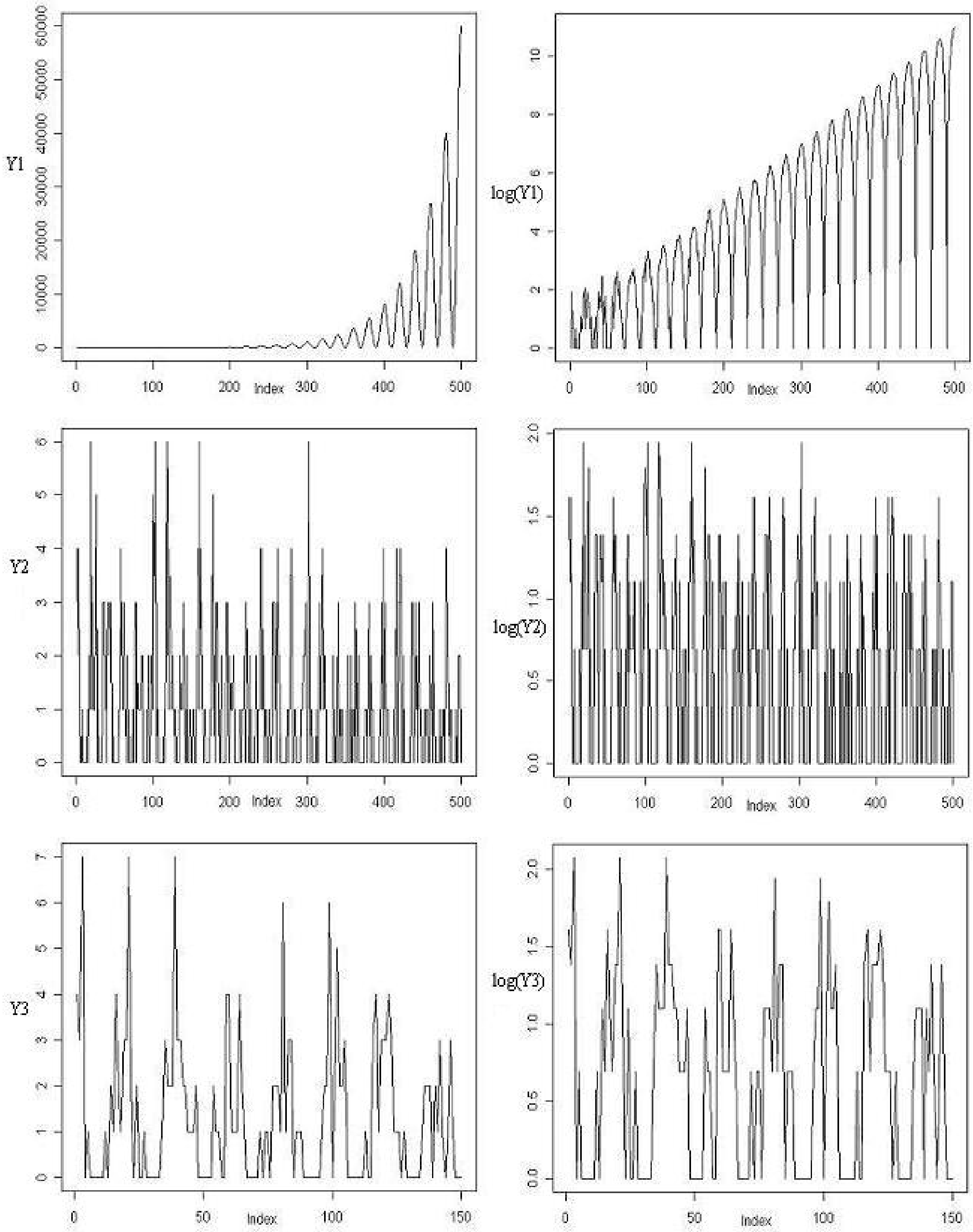
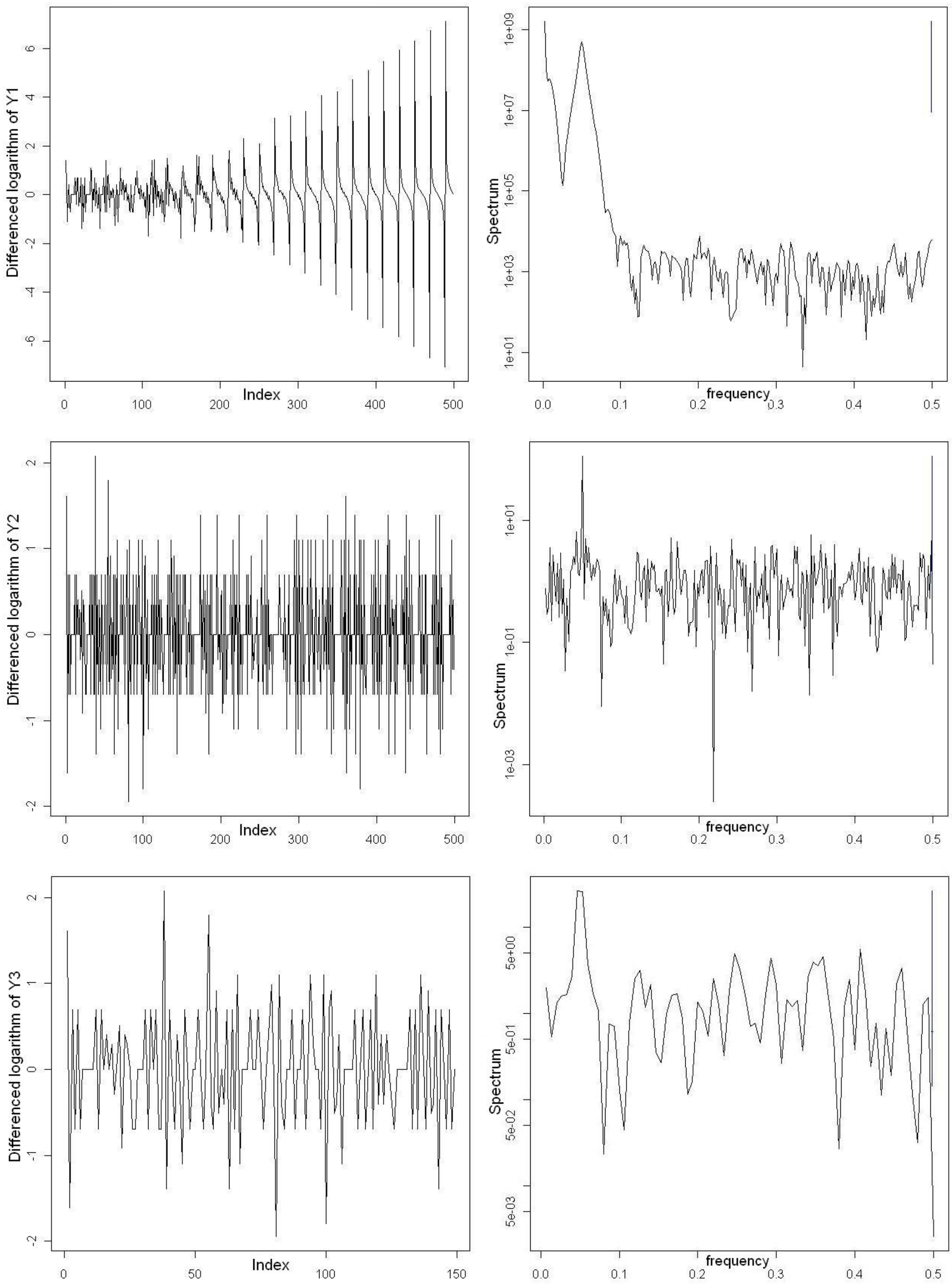


Figure 2.3: Plots of the periodograms of the three simulated data sets, as estimates of their spectrums, on the right, and their differenced logarithms, on the left.



2.4 Estimating ω from Zeger's latent ACF estimator

In Zeger (1988), a method-of-moments estimator $\hat{\gamma}(s)$ for the latent process autocovariance, similar to the ratio of the sample covariance functions for the data and for a unconditional mean estimator, is proposed. There is strong empirical evidence that these covariance estimates, as a data set, are smoothly periodic enough to estimate ω from graphically, either directly from the data plot or via a Fourier transform. We will show examples of empirical evidence and establish theoretical evidence for Zeger's estimator when the mean estimator used in $\hat{\gamma}(s)$ is unbiased, then examine the bias and other problems which arise when using a natural choice of mean estimator such as $e^{\mathbf{x}_t^T\hat{\boldsymbol{\theta}}}$ where $\hat{\boldsymbol{\theta}}$ is a consistent estimator of the mean parameter vector.

2.4.1 The periodicity of an unbiased estimator

A more successful approach than the previous two, on the basis of empirical evidence, is the estimation of ω via estimation of the covariances of ε_t. Recall that

$$Cov\left(y_t, y_{t+s}\right) = \mu_t\mu_{t+s}Cov\left(\varepsilon_t, \varepsilon_{t+s}\right) = \frac{\mu_t\mu_{t+s}\cos\left(2\omega s\right)}{2}$$

which gives for all $t \geq 1$

$$Cov\left(\varepsilon_t, \varepsilon_{t+s}\right) = \frac{\cos\left(2\omega s\right)}{2}$$

Zeger (1988) suggests the method-of-moments estimator

$$\hat{\gamma}\left(s\right) = \frac{\sum_{t=1}^{T-s}\left(y_t - \hat{\mu}_t\right)\left(y_{t+s} - \hat{\mu}_{t+s}\right)}{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}$$

for $Cov\left(\varepsilon_t, \varepsilon_{t+s}\right)$, using some estimate $\hat{\mu}_t$ of $E\left(y_t\right)$. If we assume that $\{\hat{\mu}_t\}$ is a series of unbiased and mutually uncorrelated estimators and approximate the mean of the ratio by the ratio of the means (Davis et al. (2000),p10)

$$\begin{aligned} E\left(\hat{\gamma}\left(s\right)\right) &\simeq \frac{\sum_{t=1}^{T-s} E\left(\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)\right)}{\sum_{t=1}^{T-s} E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)} = \frac{\sum_{t=1}^{T-s} Cov\left(y_t, y_{t+s}\right)}{\sum_{t=1}^{T-s} \mu_t\mu_{t+s}} \\ &= \frac{\frac{\cos\left(2\omega s\right)}{2}\sum_{t=1}^{T-s}\mu_t\mu_{t+s}}{\sum_{t=1}^{T-s}\mu_t\mu_{t+s}} = \frac{\cos\left(2\omega s\right)}{2}. \end{aligned} \tag{2.4.1}$$

Thus $\hat{\gamma}\left(s\right)$ is an asymptotically unbiased estimate of $Cov\left(\varepsilon_t, \varepsilon_{t+s}\right)$. Having accurate estimates of $\{\cos\left(2t\omega\right)\}$ for $1 \leq t \leq \tau < T$, we should be able to gain a fairly precise estimate of ω from either the basic plot of $\{\hat{\gamma}\left(s\right)\}$ or from examining the modulus of a Discrete Fourier Transform (DFT) of $\{\hat{\gamma}\left(s\right)\}$. The maxima of the former, which we will refer to as the direct plot, would be expected to occur at lags corresponding to multiples of the true period $\frac{\pi}{\omega}$, while it is well established that the DFT of an N-element harmonic process with frequency θ will be maximised at $\frac{N}{\theta}$ if θ is a multiple of $\frac{2\pi}{N}$ and at k for $\frac{2\pi k}{N}$ closest to θ otherwise.

Several data sets, of various sizes, have been simulated from the single-period model with various trend and frequency parameters and the distinctness of their periods/frequencies compared for the $\{\hat{\gamma}\left(s\right)\}$ DFT and direct plot of each. From these graphs, there appears to be good empirical evidence for both methods, a suggestion that both produce slightly less clear results for data with negative trends, and also some indication of their applicability with respect to each other. The direct-plot approach appears easier to estimate the period from than the DFT in the case of a model with a single, integer-valued period. This is due to the direct plot having multiple maxima opposed to a single maximum on the DFT modulus. Maxima at all multiples of the period could allow one to distinguish which of two consecutive values near the first maximum is the period, by studying later extrema, all of which are expected to be located at multiples of the first maximising value. The graph of the estimates has an extra advantage when the period p is coprime to τ as well as integer-valued. In this case $p \neq \tau/k$ for any integer, so the true frequency $\frac{2\pi}{p} \neq \frac{2\pi k}{\tau}$.

Thus whilst the direct plot will have repeated maxima at $p, 2p, \ldots$, the DFT will have a rough peak close to but unequal to the true frequency. However it would be easier to estimate the frequency from the DFT when ω is a multiple of $\frac{2\pi}{T}$ and period $p = \frac{2\pi}{\omega}$ is not integer-valued. Due to the expectation of the direct plot being a harmonic curve and that of the DFT being close to a delta function, the DFT also has more potential to be applied to models with multiple periods.

2.4.2 Zeger's covariance estimator under a GLM mean

The hypothesis used throughout the previous section was that the mean estimator $\hat{\mu}_t$ is unbiased and mutually uncorrelated, but the actual estimates used to gather empirical evidence for the two methods based on $\hat{\gamma}(s)$ were all calculated as $\exp\left(\mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{GLM}\right)$. This parameter estimator is the maximum likelihood estimate from the conditional quasi-likelihood function

$$l_T(\boldsymbol{\theta}, \alpha) = \frac{1}{T}\sum_{t=1}^{T}\left[y_t\left(\mathbf{x}_t^T\boldsymbol{\theta} + f(y_{t-1}, \alpha)\right) - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + f(y_{t-1}, \alpha)\right)\right]$$

maximised using a Fisher scoring algorithm. This is very similar to the methodology used to estimate the parameters in a generalised linear model with canonical link function (see McCullagh and Nelder (1989)) so the parameter estimator and the functions $f(\hat{\boldsymbol{\theta}}$ which are estimators of $f(\boldsymbol{\theta}$ are consequently referred to as *generalised linear model* GLM estimators. This is the parameter estimate used throughout Davis et al. (2000), where it is proven, for a single parameter model with log-linear latent process, to be consistent with asymptotic distribution $\hat{\boldsymbol{\theta}}_{GLM} \sim N_p\left(\boldsymbol{\theta}, \boldsymbol{\Omega}_{1T}^{-1} + \boldsymbol{\Omega}_{1T}^{-1}\boldsymbol{\Omega}_{11T}\boldsymbol{\Omega}_{1T}^{-1}\right)$.

$$\boldsymbol{\Omega}_{1T} = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^T \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)$$

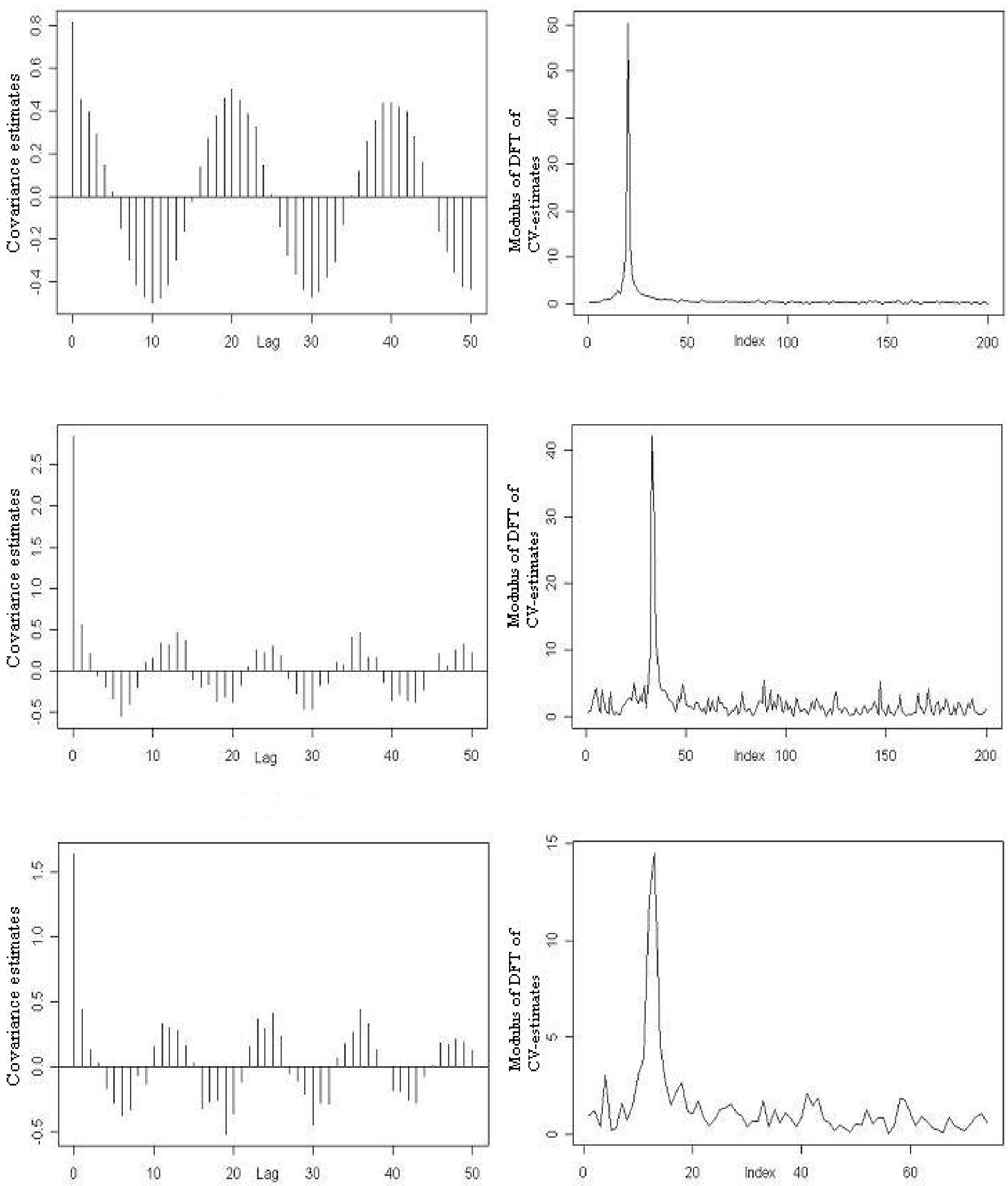
and

$$\boldsymbol{\Omega}_{11T} = \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_t\mathbf{x}_s^T \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}\right)\gamma(t-s). \tag{2.4.2}$$

This result is conjectured in Davis et al. (2000) to be extendable to models with bounded regressors and certain non-linear latent processes. Although the former is easy to ensure in our model, there is little information in Davis et al. (2000)

on what criteria the latent process must satisfy for the parameter estimate to be consistent. Assuming that the class of convergent models does include our model with its trigonometric process, it is easily shown that $\hat{\mu}_{t,GLM} = \exp\left(\mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{GLM}\right)$ is not in fact an unbiased estimate of μ_t.

Figure 2.4: Line plots and DFT-moduli of covariance estimates from simulations with conditional means $\exp(1+0.005)\cos^2(2\pi t/40)$,$\exp(1-0.004)\cos^2(2\pi t/24)$ and $\exp(1-0.004)\cos^2(2\pi t/24)$. Sample sizes are 500, 500 and 200 respectively



For $Z \sim N(\mu, \sigma^2)$

$$\begin{aligned}
E(\exp(Z)) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2}\right) \exp(z)\, \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(z^2 - 2z(\mu+\sigma^2) + \mu^2)}{2\sigma^2}\right) \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(z-(\mu+2\sigma^2))^2 + 2\sigma^2\left(\mu+\frac{\sigma^2}{2}\right)}{2\sigma^2}\right) \mathrm{d}z \\
&= \exp\left(\mu + \frac{\sigma^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(z-(\mu+2\sigma^2))^2}{2\sigma^2}\right) \mathrm{d}z \\
&= \exp\left(\mu + \frac{\sigma^2}{2}\right). \qquad (2.4.3)
\end{aligned}$$

From multivariate statistics,

$\mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{GLM} \sim N_p\left(\mathbf{x}_t^T \boldsymbol{\theta}, \mathbf{x}_t^T \left[\boldsymbol{\Omega}_{1T}^{-1} + \boldsymbol{\Omega}_{1T}^{-1}\boldsymbol{\Omega}_{11T}\boldsymbol{\Omega}_{1T}^{-1}\right]\mathbf{x}_t\right)$

so

$$\begin{aligned}
E(\hat{\mu}_t) &= E\left(\exp\left(\mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{GLM}\right)\right) = \exp\left(\mathbf{x}_t^T \boldsymbol{\theta} + \mathbf{x}_t^T \frac{\mathbf{G}_T}{2}\mathbf{x}_t\right) \\
&= \mu_t \exp\left(\mathbf{x}_t^T \frac{\mathbf{G}_T}{2}\mathbf{x}_t\right) \qquad (2.4.4)
\end{aligned}$$

where $\mathbf{G}_T = \left[\boldsymbol{\Omega}_{1T}^{-1} + \boldsymbol{\Omega}_{1T}^{-1}\boldsymbol{\Omega}_{11T}\boldsymbol{\Omega}_{1T}^{-1}\right]$.

Similarly,

$(\mathbf{x}_s + \mathbf{x}_t)^T \hat{\boldsymbol{\theta}}_{GLM} \sim N_p\left((\mathbf{x}_s + \mathbf{x}_t)^T \boldsymbol{\theta}, (\mathbf{x}_s + \mathbf{x}_t)^T \mathbf{G}_T (\mathbf{x}_s + \mathbf{x}_t)\right)$

so

$$\begin{aligned}
E(\hat{\mu}_s\hat{\mu}_t) &= E\left(\exp\left((\mathbf{x}_s + \mathbf{x}_t)^T \hat{\boldsymbol{\theta}}_{GLM}\right)\right) \\
&= \exp\left((\mathbf{x}_s + \mathbf{x}_t)^T \boldsymbol{\theta} + (\mathbf{x}_s + \mathbf{x}_t)^T \frac{\mathbf{G}_T}{2}(\mathbf{x}_s + \mathbf{x}_t)\right) \\
&= \mu_s\mu_t \exp\left((\mathbf{x}_s + \mathbf{x}_t)^T \frac{\mathbf{G}_T}{2}(\mathbf{x}_s + \mathbf{x}_t)\right). \qquad (2.4.5)
\end{aligned}$$

Note that

$$E\left(\frac{\sum_{t=1}^{T-s}(y_t - \hat{\mu}_t)(y_{t+s} - \hat{\mu}_{t+s})}{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}\right) = E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}(y_t - \hat{\mu}_t)(y_{t+s} - \hat{\mu}_{t+s})}{\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}\right)$$

and

$$\frac{E\left(\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)\right)}{\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}=E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}\right).$$

Therefore

$$\begin{aligned}
& E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}\right)-\frac{E\left(\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)} \\
= \;& E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}-\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}\right) \\
= \;& E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}\left(\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)-\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}\right)\right).
\end{aligned} \tag{2.4.6}$$

By Chebychev's inequality

$$\begin{aligned}
& P\left(\left|\hat{\mu}_{t+s}\hat{\mu}_t-E\left(\hat{\mu}_{t+s}\hat{\mu}_t\right)\right|>\epsilon\right)\leq\frac{V\left(\hat{\mu}_{t+s}\hat{\mu}_t\right)}{\epsilon^2} \\
= \;& \epsilon^{-2}\mu_t^2\mu_{t+s}^2\exp\left(\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)^T V\left(\hat{\boldsymbol{\theta}}_{GLM}\right)\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)\right) \\
\times \;& \left(\exp\left(\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)^T V\left(\hat{\boldsymbol{\theta}}_{GLM}\right)\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)\right)-1\right)\longrightarrow 0 \\
& \text{as } T\longrightarrow\infty.
\end{aligned} \tag{2.4.7}$$

Thus $\hat{\mu}_{t+s}\hat{\mu}_t \underset{P}{\longrightarrow} E\left(\hat{\mu}_{t+s}\hat{\mu}_t\right)$ so

$$\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}-\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)\underset{P}{\longrightarrow}0.$$

Since it is easy to verify that

$$V\left(\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}\right)\longrightarrow 0 \text{ as } T\longrightarrow\infty,$$

we have that

$$\left\{\left|\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}-\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)\right|\right\}$$

is uniformly integrable. Hence

$$E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T^2}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}\left(\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)-\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}\right)\right)\longrightarrow 0$$

$$\Longrightarrow\quad E\left(\frac{\frac{1}{T}\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\frac{1}{T}\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}\right)\longrightarrow\frac{E\left(\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}$$

$$\text{as } T\longrightarrow\infty. \tag{2.4.8}$$

Thus we can approximate

$$E\left(\frac{\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)}{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}}\right)\quad\text{by}\quad\frac{E\left(\sum_{t=1}^{T-s}\left(y_t-\hat{\mu}_t\right)\left(y_{t+s}-\hat{\mu}_{t+s}\right)\right)}{\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}.$$

Inserting the first two moments of $\hat{\mu}_t$ into the second expression,

$$\begin{aligned}E\left(\hat{\gamma}\left(s\right)\right) &\approx \frac{\sum_{t=1}^{T-s}E\left(y_ty_{t+s}\right)-E\left(y_{t+s}\right)E\left(\hat{\mu}_t\right)-E\left(\hat{\mu}_{t+s}\right)E\left(y_t\right)+E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}{\sum_{t=1}^{T-s}E\left(\hat{\mu}_t\hat{\mu}_{t+s}\right)}\\ &= 1+\frac{\sum_{t=1}^{T-s}\mu_t\mu_{t+s}\left(1+\frac{\cos\left(2\omega s\right)}{2}-\exp\left(\mathbf{x}_t^T\frac{\mathbf{G}_T}{2}\mathbf{x}_t\right)-\exp\left(\mathbf{x}_{t+s}^T\frac{\mathbf{G}_T}{2}\mathbf{x}_{t+s}\right)\right)}{\sum_{t=1}^{T-s}\mu_{t+s}\mu_t\exp\left(\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)^T\frac{\mathbf{G}_T}{2}\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)\right)}\end{aligned} \tag{2.4.9}$$

$$\begin{aligned}\exp\left(\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)^T\frac{\mathbf{G}_T}{2}\left(\mathbf{x}_{t+s}+\mathbf{x}_t\right)\right) &\geq 1\\ \text{and } 1-\exp\left(\mathbf{x}_t^T\frac{\mathbf{G}_T}{2}\mathbf{x}_t\right)-\exp\left(\mathbf{x}_{t+s}^T\frac{\mathbf{G}_T}{2}\mathbf{x}_{t+s}\right) &\leq -1.\end{aligned} \tag{2.4.10}$$

Consequently the denominator of the Zeger estimator is an overestimate of $\sum_{t=1}^{T-s}\mu_t\mu_{t+s}$

while the numerator is an underestimate of $\sum_{t=1}^{T-s}(y_t-\mu_t)(y_{t+s}-\mu_{t+s})$. Taken together, these inequalities clearly suggest that $\hat{\gamma}(s)$ is a significantly biased estimator of $\gamma(s)$. There are several different methods of reducing this bias, most of which replace Zeger's estimator by a sequence modified using prior covariance estimates. One such method is the direct modification of the Zeger estimator developed in Davis et al.(2000). Under the assumption of asymptotic normality,

$$\hat{\mu}_t\hat{\mu}_{t+s}\exp\left(-(\mathbf{x}_s+\mathbf{x}_t)^T\frac{\hat{\mathbf{G}}_T}{2}(\mathbf{x}_s+\mathbf{x}_t)\right)$$

is an asymptotically unbiased estimate of $\mu_t\mu_{t+s}$, with asymptotic parameter covariance matrix estimated by $\hat{\mathbf{G}}_T=\left[\hat{\mathbf{\Omega}}_{1T}^{-1}+\hat{\mathbf{\Omega}}_{1T}^{-1}\hat{\mathbf{\Omega}}_{11T}\hat{\mathbf{\Omega}}_{1T}^{-1}\right]$ where $\mathbf{\Omega}_{1T}=\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^T\hat{\mu}_t$ and $\mathbf{\Omega}_{11T}=\sum_{t=1}^{\mathcal{T}}\sum_{s=1}^{\mathcal{T}}\mathbf{x}_t\mathbf{x}_s^T\hat{\mu}_t\hat{\mu}_s\hat{\gamma}(t-s)$. Using this and similar approximations gives us a sequence of adjusted estimators

$$\begin{aligned}\hat{\gamma}_{UB}(s) &= \frac{\sum_{t=1}^{T-s}(y_t-\hat{\mu}_t)(y_{t+s}-\hat{\mu}_{t+s})}{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}g_{t,s}} \\ &- \frac{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}g_{t,s}\left(1-e^{\mathbf{x}_t^T\frac{\hat{\mathbf{G}}_T}{2}\mathbf{x}_t}-e^{\mathbf{x}_{t+s}^T\frac{\hat{\mathbf{G}}_T}{2}\mathbf{x}_{t+s}}+e^{(\mathbf{x}_s+\mathbf{x}_t)^T\frac{\hat{\mathbf{G}}_T}{2}(\mathbf{x}_s+\mathbf{x}_t)}\right)}{\sum_{t=1}^{T-s}\hat{\mu}_t\hat{\mu}_{t+s}g_{t,s}}\end{aligned}$$

where $g_{t,s}=\exp\left(-(\mathbf{x}_s+\mathbf{x}_t)^T\frac{\hat{\mathbf{G}}_T}{2}(\mathbf{x}_s+\mathbf{x}_t)\right)$ and $\hat{\mathbf{G}}_T$ is the parameter covariance matrix estimate computed using the Zeger estimates $\hat{\gamma}(s)$ and used to adjust the current estimates.

Another alternative to the Zeger estimators, which are rescalings of the sample covariance function, is the approximated sample covariance function $\tilde{\gamma}(s)=\sum_{t=1}^{T-s}\widehat{\varepsilon_t\varepsilon_{t+s}}-1$ of the latent process, calculated using some estimator. $\widehat{\varepsilon_t\varepsilon_{t+s}}$ of $\varepsilon_t\varepsilon_{t+s}$. As such a covariance estimator is a sum of quotients rather than a quotient function of two sums, its moments can theoretically be calculated much more easily than those of Zeger's estimator. The standard moments estimate $\hat{\varepsilon}_t\hat{\varepsilon}_{t+s}=\frac{y_ty_{t+s}}{\hat{\mu}_t\hat{\mu}_{t+s}}$ could be used, or if empirical evidence suggests significant bias, replaced by the approximately unbiased estimate $\widetilde{\varepsilon_t\varepsilon_{t+s}} = \dfrac{y_t y_{t+s} \exp\left(-\left(\mathbf{x}_t^T + \mathbf{x}_{t+s}\right)\frac{\hat{\mathbf{G}}_T}{2}\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\right)}{\hat{\mu}_t\hat{\mu}_{t+s}}$, with $\hat{\mathbf{G}}_T$ as defined in the adjustments to Zeger's estimator. As

$$E\left(\frac{1}{\hat{\mu}_t\hat{\mu}_{t+s}}\right) = \frac{1}{\mu_t\mu_{t+s}} \exp\left(\left(\mathbf{x}_t^T + \mathbf{x}_{t+s}\right)\frac{\hat{\mathbf{G}}_T}{2}\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\right),$$

this would remove nearly all of the bias from the above estimator (as the variance matrix used is an estimate itself, we can not be certain that $\widetilde{\varepsilon_t\varepsilon_{t+s}}$ is completely unbiased).

Chapter 3

Asymptotic results for the GLM and DFT estimators

3.1 Introduction

The main topic of this chapter will be the properties of the mean-parameter and frequency estimators. Taking the simple model with trend and intercept regressors only, the consistency of a GLM estimator of $\boldsymbol{\theta}$ which maximises a pseudo-likelihood function will be established and then the asymptotic normality of the estimator will be established via characteristic functions.

In Section 3.4, an estimator of ω will be introduced and its convergence to the true frequency value verified. This estimator is derived from the estimated covariance function of the latent process and we will utilise the asymptotic distribution of the mean estimator to prove pointwise convergence. Sections 3.2, 3.3 and 3.4 will also supply us with the key results, using Chebychev's inequality and Abel's lemma, which will significantly simplify the analysis of other, more complicated models.

The extension of the properties of the mean-parameter estimator to models with regressors other than trend and intercept will be studied in the final section. Covariates such as population size and weather statistics will often be very informative when modelling time series of counts, but the randomness of the values involved, opposed to a deterministic covariate like time, means that the results do not automatically follow.

3.2 Convergence of $\hat{\boldsymbol{\theta}}_{GLM}$

The first asymptotic result which we shall establish is the pointwise convergence of $\hat{\boldsymbol{\theta}}_{GLM}$ to the true parameter vector $\boldsymbol{\theta}_0$ where $\hat{\boldsymbol{\theta}}_{GLM}$ is the vector which maximises the pseudo-log-likelihood function

$$l_T(\boldsymbol{\theta}, \alpha) = \frac{1}{T}\sum_{t=1}^{T}\left[y_t\left(\mathbf{x}_t^T\boldsymbol{\theta} + f(y_{t-1}, \alpha)\right) - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + f(y_{t-1}, \alpha)\right)\right]$$

Maximum likelihood estimates are found using a Fisher scoring algorithm. This is very similar to the methodology used to estimate the parameters in a generalised linear model with canonical link function (see McCullagh and Nelder (1989)) so the parameter estimators and the functions $f\left(\hat{\boldsymbol{\theta}}\right)$ which are used as estimators of $f(\boldsymbol{\theta})$ will be referred to as GLM estimators. After establishing that $\boldsymbol{\theta}_0$ is the unique minimiser of $E[l_T(\boldsymbol{\theta})]$, it is sufficient to prove that $\sup\left|l_T\left(\hat{\boldsymbol{\theta}}\right) - E[l_T(\boldsymbol{\theta}_0)]\right| \xrightarrow{\mathcal{P}} 0$, known as uniform (rather than pointwise) convergence in probability
This shall be achieved by proving that $l_T(\boldsymbol{\theta})$ is both pointwise convergent in probability to $E[l_T(\boldsymbol{\theta})]$ and is equicontinuous in probability.

Let $l_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left[y_t\mathbf{x}_t^T\boldsymbol{\theta} - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right]$. Then

$$\begin{aligned} E(l_T(\boldsymbol{\theta})) = \tilde{l}(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=1}^{T}\left[E(y_t)\,\mathbf{x}_t^T\boldsymbol{\theta} - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\left[\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\mathbf{x}_t^T\boldsymbol{\theta} - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right]. \qquad (3.2.1)\end{aligned}$$

This is maximised by $\boldsymbol{\theta}$ solving $\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^T\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right) - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right) = \mathbf{0}$.
This is clearly the case when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ Let $\hat{\boldsymbol{\theta}}_T$ be the value which maximises $l_T(\boldsymbol{\theta})$.

$$
\begin{aligned}
V\left(l_T\left(\boldsymbol{\theta}\right)\right) &= \frac{1}{T^2}V\left(\sum_{t=1}^{T} y_t\mathbf{x}_t^T\boldsymbol{\theta} - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right) \\
&= \frac{1}{T^2}E\left(\sum_{t=1}^{T}\left(y_t - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\right)\mathbf{x}_t^T\boldsymbol{\theta}\right)^2 \\
&= \frac{1}{T^2}E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left(y_s - \exp\left(\mathbf{x}_s^T\boldsymbol{\theta}_0\right)\right)\left(y_t - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\right)\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(y_t, y_s)\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}_0\right)\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\frac{\cos(2(t-s)\omega)}{2} \\
&+ \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right) \\
&= \frac{1}{2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}_0\right)\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right) \\
&\quad \times\Big(\cos(2\omega t)\cos(2\omega s) + \sin(2\omega t)\sin(2\omega s)\Big) + \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}_0\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right) + \left(\frac{1}{\sqrt{2}T}\sum_{t=1}^{T}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\sin(2\omega t)\right)^2 \\
&+ \left(\frac{1}{\sqrt{2}T}\sum_{t=1}^{T}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right)\cos(2\omega t)\right)^2. \qquad (3.2.2)
\end{aligned}
$$

For $\mathbf{x}_t^T\boldsymbol{\theta} = \alpha + \beta t/T$, let $a = |\alpha|$ and $b = |\beta|$. Then

$$
\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_0\right) \leq \frac{1}{T^2}\sum_{t=1}^{T}(a+b)^2\exp\left(a_0+b_0\right) \\
&= \frac{(a+b)^2\exp\left(a_0+b_0\right)}{T} \longrightarrow 0 \text{ as } T \longrightarrow \infty. \qquad (3.2.3)
\end{aligned}
$$

Using Abel's Lemma:

$$
\sum_{t=1}^{T} f(t)g(t) = f(T)\sum_{t=0}^{T} g(t) - f_0g_0 - \sum_{t=0}^{T-1}\left(\left(f(t+1)-f(t)\right)\sum_{s=0}^{t} g(s)\right) \qquad (3.2.4)
$$

with $f_t = \mathbf{x}_t^T \boldsymbol{\theta} \exp\left(\mathbf{x}_t^T \boldsymbol{\theta}_0\right)$ and $g_t = \cos(2\omega t)$

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbf{x}_t^T \boldsymbol{\theta} \exp\left(\mathbf{x}_t^T \boldsymbol{\theta}_0\right) \cos(2\omega t) \\
= \;& (\alpha + \beta) \exp(\alpha_0 + \beta_0) \sum_{t=0}^{T} \cos(2\omega t) - \alpha \exp(\alpha_0) \\
- \;& \sum_{t=0}^{T-1} \left[\left(e^{\alpha_0 + \beta_0 (t+1)/T} \left(\alpha + \beta (t+1)/T\right) - e^{\alpha_0 + \beta_0 t/T} \left(\alpha + \beta t/T\right) \right) \sum_{s=0}^{t} \cos(2\omega s) \right]
\end{aligned} \tag{3.2.5}$$

$$(\alpha + \beta) \exp(\alpha_0 + \beta_0) \sum_{t=0}^{T} \cos(2\omega t) = (\alpha + \beta) \exp(\alpha_0 + \beta_0) \frac{\sin\left((T+1)\,\omega\right) \cos(T\omega)}{\sin(\omega)}$$

$$\begin{aligned}
&\sum_{t=0}^{T-1} \left[\left(e^{\alpha_0 + \beta_0 (t+1)/T} \left(\alpha + \beta (t+1)/T\right) - e^{\alpha_0 + \beta_0 t/T} \left(\alpha + \beta t/T\right) \right) \sum_{s=0}^{t} \cos(2\omega s) \right] \\
= \;& \sum_{t=0}^{T-1} \exp(\alpha_0 + \beta_0 t/T) \left((\alpha + \beta t/T) \left(\exp(\beta_0/T) - 1 \right) + \beta/T \exp(\beta_0/T) \right) \\
& \times \frac{\sin\left((t+1)\,\omega\right) \cos(\omega t)}{\sin(\omega)} \\
\leq \;& \frac{1}{\sin(\omega)} \sum_{t=0}^{T-1} \exp(\alpha_0 + \beta_0 t/T) \left((\alpha + \beta t/T) \left(\exp(\beta_0/T) - 1 \right) + \beta/T \exp(\beta_0/T) \right) \\
\propto \;& \sum_{t=0}^{T-1} \exp(\alpha_0 + \beta_0 t/T) \left((\alpha + \beta t/T) \left(\exp(\beta_0/T) - 1 \right) + \beta/T \exp(\beta_0/T) \right) \\
= \;& \left(e^{\beta_0/T} - 1\right) \sum_{t=0}^{T-1} (\alpha + \beta t/T)\, e^{\alpha_0 + \beta_0 t/T} + \beta/T e^{\beta_0/T} \sum_{t=0}^{T-1} e^{\alpha_0 + \beta_0 t/T} \\
= \;& \left(e^{\beta_0/T} - 1\right) e^{\alpha_0} \left(\alpha \frac{e^{\beta_0} - 1}{e^{\beta_0/T} - 1} + \frac{e^{\beta_0}}{e^{\beta_0/T} - 1} + \frac{e^{\beta_0/T}\left(e^{\beta_0} - 1\right)}{T\left(e^{\beta_0/T} - 1\right)^2} \right) \\
+ \;& \frac{\beta e^{\alpha_0 + \beta_0/T}}{T} \left(\frac{e^{\beta_0} - 1}{e^{\beta_0/T} - 1} \right) \\
= \;& e^{\alpha_0} \left(\alpha \left(e^{\beta_0} - 1\right) - e^{\beta_0} \right) + \frac{(\beta + 1)\, e^{\alpha_0 + \beta_0/T}}{T} \left(\frac{e^{\beta_0} - 1}{e^{\beta_0/T} - 1} \right) \\
\longrightarrow \;& e^{\alpha_0} \left(\alpha \left(e^{\beta_0} - 1\right) - e^{\beta_0} \right) + \frac{(\beta + 1)\, e^{\alpha_0}}{\beta} \left(e^{\beta_0} - 1\right) \\
\text{as} \quad & T \longrightarrow \infty.
\end{aligned} \tag{3.2.6}$$

Thus

$$\begin{aligned}
& \frac{1}{\sqrt{2}T}\sum_{t=1}^{T}\left(\alpha+\beta t/T\right)e^{\alpha_0+\beta_0 t/T}\cos\left(2\omega t\right)\\
\propto\quad & -\frac{\alpha e^{\alpha_0}}{\sqrt{2}T}+\frac{1}{\sqrt{2}T}\sum_{t=0}^{T-1}e^{\alpha_0+\beta_0 t/T}\left(\left(\alpha+\beta t/T\right)\left(e^{\beta_0/T}-1\right)+\beta/Te^{\beta_0/T}\right)\\
+\quad & \left(\alpha+\beta\right)e^{\alpha_0+\beta_0}\frac{\sin\left(\left(T+1\right)\omega\right)\cos\left(T\omega\right)}{\sqrt{2}T\sin\left(\omega\right)}\\
\propto\quad & \frac{1}{\sqrt{2}T}\left[-\alpha\exp\left(\alpha_0\right)+\left(\alpha+\beta\right)\exp\left(\alpha_0+\beta_0\right)\frac{\sin\left(\left(T+1\right)\omega\right)\cos\left(T\omega\right)}{\sin\left(\omega\right)}\right.\\
+\quad & \left.e^{\alpha_0}\left(\alpha\left(e^{\beta_0}-1\right)-e^{\beta_0}\right)+\frac{\left(\beta+1\right)e^{\alpha_0+\beta_0/T}}{\beta_0}\left(e^{\beta_0}-1\right)\right]\\
\longrightarrow\quad & 0\text{ as }T\longrightarrow\infty
\end{aligned}\tag{3.2.7}$$

as all terms inside the square brackets are fixed and finite. By the same reasoning, $\frac{1}{\sqrt{2}T}\sum_{t=1}^{T}\left(\alpha+\beta t/T\right)\exp\left(\alpha_0+\beta_0 t/T\right)\sin\left(2\omega t\right)\longrightarrow 0$

Thus for $\mathbf{x}_t^T\boldsymbol{\theta}=\alpha+\beta t/T$, $V\left(l_T\left(\boldsymbol{\theta}\right)\right)\longrightarrow 0$ as $T\longrightarrow\infty$

Using Chebychev's Inequality, $P\left(\left|X-E\left(X\right)\right|>\epsilon\right)\leq\frac{V\left(X\right)}{\epsilon^2}$, $l_T\left(\boldsymbol{\theta}\right)$ is pointwise convergent in probability to $\tilde{l}\left(\boldsymbol{\theta}\right)$ or in other words, $\left|l_T\left(\boldsymbol{\theta}\right)-\tilde{l}\left(\boldsymbol{\theta}\right)\right|\underset{P}{\longrightarrow}0\;\forall\boldsymbol{\theta}=\left(\alpha,\beta\right)\in\mathbb{R}^2$.

Similarly, for the score function $\Delta l\left(\boldsymbol{\theta}\right)=\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\left(y_t-\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\right)$,

$$\begin{aligned}
\mathbf{V}\left(\Delta l\left(\boldsymbol{\theta}\right)\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_t\mathbf{x}_s^T Cov\left(y_t,y_s\right)\\
&= \frac{1}{2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\begin{pmatrix}1 & s/T\\ t/T & st/T^2\end{pmatrix}\exp\left(\alpha_0+\beta_0\right)\exp\left(\alpha_0+\beta_0\right)\cos\left(2\left(t-s\right)\omega\right)\\
&+ \frac{1}{T^2}\sum_{t=1}^{T}\begin{pmatrix}1 & t/T\\ t/T & t^2/T^2\end{pmatrix}\exp\left(2\left(\alpha_0+\beta_0\right)\right)\\
&= \frac{1}{2T^2}\left[\left(\sum_{t=1}^{T}\begin{pmatrix}1\\ t/T\end{pmatrix}e^{\alpha_0+\beta_0}\cos\left(2t\omega\right)\right)\left(\sum_{s=1}^{T}\begin{pmatrix}1\\ s/T\end{pmatrix}e^{\alpha_0+\beta_0}\cos\left(2s\omega\right)\right)^T\right.\\
&+ \left.\sum_{t=1}^{T}\begin{pmatrix}1 & t/T\\ t/T & t^2/T^2\end{pmatrix}\exp\left(2\left(\alpha_0+\beta_0\right)\right)\right].
\end{aligned}\tag{3.2.8}$$

Evaluating each term as slight adjustments to those of $V\left(l_T\left(\boldsymbol{\theta}\right)\right)$, it is clear that $\mathbf{V}\left(\Delta l_T\left(\boldsymbol{\theta}\right)\right)\longrightarrow\mathbf{0}$ as $T\longrightarrow\infty$.

Let $\mathbf{J}_T = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}1\\ t/T\end{pmatrix} y_t = \begin{pmatrix}J_{1T}\\ J_{2T}\end{pmatrix}$ and $E(\mathbf{J}_T) = \bar{\mathbf{J}}_T = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}1\\ t/T\end{pmatrix}\mu_t = \begin{pmatrix}\bar{J}_{1T}\\ \bar{J}_{2T}\end{pmatrix}$. $\mathbf{J}_T \geq \Delta l_T(\boldsymbol{\theta}), V(\mathbf{J}_T) = V(\Delta l_T(\boldsymbol{\theta}))$ and $\mathbf{J}_T \geq 0\, \forall T$. Using basic axioms of probability, Chebychev's Inequality and the triangle inequality,

$$\begin{aligned}
\forall k > 0,\ P\left(|\mathbf{J}_T| > \left|\bar{\mathbf{J}}_T\right| + k\right) &= P\left(|\mathbf{J}_T| - \left|\bar{\mathbf{J}}_T\right| > k\right) < P\left(\left|\mathbf{J}_T - \bar{\mathbf{J}}_T\right| > k\right) \\
= P\left(\left|\mathbf{J}_T - \bar{\mathbf{J}}_T\right|^2 > k^2\right) &\leq P\left(\left(J_{1T} - \bar{J}_{1T}\right)^2 > \frac{k^2}{2}\right) + P\left(\left(J_{2T} - \bar{J}_{2T}\right)^2 > \frac{k^2}{2}\right) \\
&\leq \frac{2V(J_{1T})}{k^2} + \frac{2V(J_{2T})}{k^2} \longrightarrow 0 \\
&\Longrightarrow P(|\mathbf{J}_T| > M) < \frac{2V(J_{1T})}{\left(M - \left|\bar{J}_{1T}\right|\right)^2} + \frac{2V(J_{2T})}{\left(M - \left|\bar{J}_{2T}\right|\right)^2}
\end{aligned} \tag{3.2.9}$$

for all $M > \left|\bar{J}_T\right|$.

$$\begin{aligned}
P(|l_T(\boldsymbol{\theta}_1) - l_T(\boldsymbol{\theta}_2)| > \epsilon) &\leq P(\sup|\Delta l_T(\boldsymbol{\theta})| \cdot |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| > \epsilon) \text{ (by MVT)} \\
&\leq P(\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \sup|\Delta l_T(\boldsymbol{\theta})| > \epsilon) \\
&\leq P(\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|\, |\mathbf{J}_T| > \epsilon) \\
&= P\left(|\mathbf{J}_T| > \frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|}\right) \\
&\leq \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} - \left|\bar{J}_{1T}\right|\right)^2} + \frac{2V(J_{2T})}{\left(\frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} - \left|\bar{J}_{2T}\right|\right)^2}.
\end{aligned} \tag{3.2.10}$$

A function f defined over a range of values $x \in \mathfrak{X}$ is said to be equicontinuous in probability if the following condition is satisfied:

For any $\epsilon > 0$, there exists $\delta > 0$ such that if $|x_1 - x_2| \leq \delta$,

then $P(|f(x_1) - f(x_2)| > \epsilon) \longrightarrow 0$.

Let $\begin{pmatrix}A\\ B\end{pmatrix} = \sup\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ and $\tilde{\mathbf{J}}_T = \begin{pmatrix}\tilde{J}_{T1}\\ \tilde{J}_{T2}\end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}1\\ t/T\end{pmatrix} e^{A+Bt/T}$. For any $\epsilon > 0$, let $\delta = \frac{\epsilon}{2\sup|\tilde{J}_{T1}|} = \frac{\epsilon}{2\lim_{T\to\infty}(\tilde{J}_{T1})}$ as $\tilde{J}_{T1}$ is monotonically increasing. Denote $\lim_{T\to\infty}(\tilde{J}_{T1})$ by $J_\infty = \frac{e^A\left(e^B - 1\right)}{B}$.

For all $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ such that $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| < \delta$

$$\begin{aligned}
\frac{\epsilon}{\delta} < \frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} \Longrightarrow & \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} - |\bar{J}_{1T}|\right)^2} < \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\delta} - |\bar{J}_{1T}|\right)^2} \\
= & \frac{2V(J_{1T})}{\left(2J_\infty - |\bar{J}_{1T}|\right)^2}. \qquad (3.2.11)
\end{aligned}$$

For all $0 < 1 \leq T$, $t/T \leq 1$ and $\mu_t > 0$, so $\bar{J}_{T2} < \bar{J}_{T1}$. Thus $\left(2J_\infty - |\bar{J}_{1T}|\right)^2 < \left(2J_\infty - |\bar{J}_{1T}|\right)^2 < \left(2J_\infty - |\bar{J}_{2T}|\right)^2$

Putting all these inequalities together, we have

$$\begin{aligned}
P\left(|l_T(\boldsymbol{\theta}_1) - l_T(\boldsymbol{\theta}_2)| > \epsilon\right) \leq & \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} - |\bar{J}_{1T}|\right)^2} + \frac{2V(J_{2T})}{\left(\frac{\epsilon}{\sup|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|} - |\bar{J}_{2T}|\right)^2} \\
\leq & \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\delta} - \bar{J}_{1T}\right)^2} + \frac{2V(J_{1T})}{\left(\frac{\epsilon}{\delta} - \bar{J}_{1T}\right)^2} \\
= & \frac{2V(J_{1T})}{\left(2J_\infty - \bar{J}_{1T}\right)^2} + \frac{2V(J_{1T})}{\left(2J_\infty - \bar{J}_{1T}\right)^2} \\
\leq & \frac{2\left(V(J_{1T}) + V(J_{2T})\right)}{\left(2J_\infty - \bar{J}_{1T}\right)^2} \leq \frac{2\left(V(J_{1T}) + V(J_{2T})\right)}{J_\infty^2} \\
= & 2\left(V(J_{1T}) + V(J_{2T})\right) \frac{B^2}{e^{2A}\left(e^B - 1\right)^2} \\
\longrightarrow & 0 \text{ as } T \longrightarrow \infty. \qquad (3.2.12)
\end{aligned}$$

Thus for any $\epsilon > 0$, there exists $\delta = \dfrac{\epsilon}{2J_\infty} > 0$ such that if $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \leq \delta$, then $P\left(|l(\boldsymbol{\theta}_1) - l(\boldsymbol{\theta}_2)| > \epsilon\right) \longrightarrow 0$.

Therefore the sequence $l_T(\boldsymbol{\theta})$ is equicontinuous in probability, it is pointwise convergent in probability to $\tilde{l}_T(\boldsymbol{\theta})$ and the parameter space $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ is compact. This is sufficient to establish uniform convergence of $l_T(\boldsymbol{\theta})$, as stated in Berkes et. al. (2003), Lemma 5.4. Thus $\sup\left|l_T(\boldsymbol{\theta}) - \tilde{l}_T(\boldsymbol{\theta})\right| \xrightarrow{\mathcal{P}} 0$. Since

$$\begin{aligned}
& l_T(\boldsymbol{\theta}_0) \leq l_T\left(\hat{\boldsymbol{\theta}}\right) \xrightarrow{\mathcal{P}} \tilde{l}_T\left(\hat{\boldsymbol{\theta}}\right) \leq \tilde{l}_T(\boldsymbol{\theta}_0) \\
\Longrightarrow & l_T(\boldsymbol{\theta}_0) - \tilde{l}_T(\boldsymbol{\theta}_0) \leq l_T\left(\hat{\boldsymbol{\theta}}\right) - \tilde{l}_T(\boldsymbol{\theta}_0) \leq l_T\left(\hat{\boldsymbol{\theta}}\right) - \tilde{l}_T\left(\hat{\boldsymbol{\theta}}\right) \\
\Longrightarrow & \left|l_T\left(\hat{\boldsymbol{\theta}}\right) - \tilde{l}_T(\boldsymbol{\theta}_0)\right| \leq \max\left\{\left|l_T\left(\hat{\boldsymbol{\theta}}\right) - \tilde{l}_T\left(\hat{\boldsymbol{\theta}}\right)\right|, \left|l_T(\boldsymbol{\theta}_0) - \tilde{l}_T(\boldsymbol{\theta}_0)\right|\right\} \\
\leq & \sup\left|l_T(\boldsymbol{\theta}) - \tilde{l}_T(\boldsymbol{\theta})\right| \xrightarrow{\mathcal{P}} 0 \\
\Longrightarrow & \left|l_T\left(\hat{\boldsymbol{\theta}}\right) - \tilde{l}_T(\boldsymbol{\theta}_0)\right| \xrightarrow{\mathcal{P}} 0. \qquad (3.2.13)
\end{aligned}$$

As the minimum of $\tilde{l}_T(\boldsymbol{\theta})$ is unique, it is clear from this that $\hat{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$. Therefore $\hat{\boldsymbol{\theta}}$ is a consistent estimator of the intercept and trend parameters.

3.3 Normality of $\hat{\boldsymbol{\theta}}_{GLM}$

We shall now establish that $\sqrt{T}\hat{\boldsymbol{\theta}}_{GLM}$ converges in distribution to a Gaussian random variable. As $\hat{\boldsymbol{\theta}}_{GLM}$ is both a vector and the outcome of an iterative algorithm, stages will be taken to simplify the task. Using the Mean Value theorem, we will first show that asymptotic normality of a rescaled score function is a sufficient condition for asymptotic normality of $\sqrt{T}\hat{\boldsymbol{\theta}}_{GLM}$. Asymptotic normality of any scalar projection of this score function will be proven by computing the limit of the characteristic function, after verifying that this in turn is sufficient to prove that the score function is asymptotically Gaussian. We abbreviate $\hat{\boldsymbol{\theta}}_{GLM}$ by $\hat{\boldsymbol{\theta}}$ throughout

Consider the score function $S_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(y_t - \mu_t)$, where $\mathbf{x}_t = \begin{pmatrix} 1 \\ t/T \end{pmatrix}$. By the mean value theorem,

$$\begin{aligned} \sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\nabla S_T(\boldsymbol{\theta}_1) &= \sqrt{T}\left(S_T\left(\hat{\boldsymbol{\theta}}\right) - S_T(\boldsymbol{\theta}_0)\right) = -\sqrt{T}S_T(\boldsymbol{\theta}_0) \\ \Longrightarrow \quad \sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) &= \left(-\nabla S_T(\boldsymbol{\theta}_1)\right)^{-1}\sqrt{T}S_T(\boldsymbol{\theta}_0) \end{aligned} \tag{3.3.1}$$

where $\nabla S_T(\boldsymbol{\theta}_1) = -\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^T \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}_1\right)$ and $\boldsymbol{\theta}_1$ lies between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Since $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$, by the sandwich theorem $\boldsymbol{\theta}_1 \xrightarrow{P} \boldsymbol{\theta}_0$. Also, for all $\boldsymbol{\theta}$, using the integral approximation theorem, we have that

$$\nabla S_T(\boldsymbol{\theta}) \longrightarrow M(\boldsymbol{\theta}) = \begin{pmatrix} \int_0^1 e^{\alpha+\beta x}\mathrm{d}x & \int_0^1 xe^{\alpha+\beta x}\mathrm{d}x \\ \int_0^1 e^{\alpha+\beta x}\mathrm{d}x & \int_0^1 xe^{\alpha+\beta x}\mathrm{d}x \end{pmatrix} \quad \text{as } T \longrightarrow \infty. \tag{3.3.2}$$

Therefore by the continuous mapping theorem,

$$\nabla S_T(\boldsymbol{\theta}_1) \xrightarrow{P} M(\boldsymbol{\theta}_0) \quad \text{as } T \longrightarrow \infty. \tag{3.3.3}$$

This result is easily proven using Chapter 5 of Billingsley (1968).

We proceed by proving that $\sqrt{T}S_T(\boldsymbol{\theta}_0)$ converges in distribution to a Gaussian random variable $S_T^*(\boldsymbol{\theta}_0)$ as $T \longrightarrow 0$. Then by Billingsley (1968), Theorem 4.4,

$$\begin{aligned}\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) &= \left(-\nabla S_T(\boldsymbol{\theta}_1)\right)^{-1}\sqrt{T}S_T(\boldsymbol{\theta}_0) \\ &\xrightarrow{D} -M(\boldsymbol{\theta}_0)^{-1} S_T^*(\boldsymbol{\theta}_0). \qquad (3.3.4)\end{aligned}$$

By the Cramer-Wold theorem, $\sqrt{T}S_T(\boldsymbol{\theta}_0)$ is asymptotically Gaussian if and only if any linear projection of $\sqrt{T}S_T(\boldsymbol{\theta})$, $P_T(\boldsymbol{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\gamma,\delta)\,\mathbf{x}_t(y_t-\mu_t)$, is too. To prove asymptotic normality, we shall examine the characteristic function $\varphi_{S_T}(\lambda) = E(\exp(i\lambda P_T(\boldsymbol{\theta})))$.

$$\begin{aligned} & E(\exp(i\lambda P_T(\boldsymbol{\theta}))) = E(E(\exp(i\lambda P_T(\boldsymbol{\theta})) \mid \phi)) \\ = \; & E\left(\frac{1}{\sqrt{T}}\prod_{t=1}^{T} E(\exp((\gamma+\delta t/T)\,i\lambda(y_t-\mu_t)) \mid \phi)\right). \qquad (3.3.5)\end{aligned}$$

Studying the inner expectation,

$$\begin{aligned} & \prod_{t=1}^{T} E(\exp((\gamma+\delta t/T)\,i\lambda(y_t-\mu_t)) \mid \phi) \\ = \; & \prod_{t=1}^{T}\exp\left[\mu_t\varepsilon_t\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1\right) - i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\mu_t\right] \\ = \; & \exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1-i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)\mu_t\right] \\ \times \; & \exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1\right)\mu_t\cos(2(\omega t+\phi))\right]. \qquad (3.3.6)\end{aligned}$$

It follows that

$$\begin{aligned} & E(\exp[i\lambda P_T(\boldsymbol{\theta})]) \\ = \; & \exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1-i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)\exp(\alpha+\beta t/T)\right]\times \\ & E\left\{\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1\right)\exp(\alpha+\beta t/T)\cos(2(\omega t+\phi))\right]\right\}. \\ & \qquad (3.3.7)\end{aligned}$$

Starting with the latter term on the right-hand side of (3.3.7), we have that

$$\begin{aligned}
&\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1\right)\exp\left(\alpha+\beta t/T\right)\cos\left(2\left(\omega t+\phi\right)\right)\\
= &\left[\left(\frac{e^{\frac{\lambda}{\sqrt{T}}(\gamma+\delta+\delta/T)}}{1+e^{2\beta/T}e^{2i\lambda\delta/T\sqrt{T}}-2e^{\beta/T}e^{i\lambda\delta/T\sqrt{T}}\cos\left(2\omega\right)}\right.\right.\\
&\left.-\frac{1}{1+e^{2\beta/T}-2e^{\beta/T}\cos\left(2\omega\right)}\right)e^{\beta+\beta/T}\cos\left(2\left(\phi+T\omega\right)\right)\\
- &\left(\frac{e^{\frac{\lambda}{\sqrt{T}}(\gamma+\delta)}}{1+e^{2\beta/T}e^{2i\lambda\delta/T\sqrt{T}}-2e^{\beta/T}e^{i\lambda\delta/T\sqrt{T}}\cos\left(2\omega\right)}\right.\\
&\left.-\frac{1}{1+e^{2\beta/T}-2e^{\beta/T}\cos\left(2\omega\right)}\right)e^{\beta}\cos\left(2\left(\phi+\left(T+1\right)\omega\right)\right)\\
- &\left(\frac{e^{\frac{\lambda}{\sqrt{T}}(\gamma+\delta/T)}}{1+e^{2\beta/T}e^{2i\lambda\delta/T\sqrt{T}}-2e^{\beta/T}e^{i\lambda\delta/T\sqrt{T}}\cos\left(2\omega\right)}\right.\\
&\left.-\frac{1}{1+e^{2\beta/T}-2e^{\beta/T}\cos\left(2\omega\right)}\right)e^{\beta/T}\cos\left(2\phi\right)\\
+ &\left(\frac{1}{1+e^{2\beta/T}e^{2i\lambda\delta/T\sqrt{T}}-2e^{\beta/T}e^{i\lambda\delta/T\sqrt{T}}\cos\left(2\omega\right)}\right.\\
&\left.\left.-\frac{1}{1+e^{2\beta/T}-2e^{\beta/T}\cos\left(2\omega\right)}\right)\cos\left(2\left(\phi+\omega\right)\right)\right]. \qquad (3.3.8)
\end{aligned}$$

The above expression converges to zero for all values of $\alpha, \beta, \gamma, \delta, \lambda$ and ϕ, so we can conclude that

$$\begin{aligned}
&E\left\{\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\delta t/T}{\sqrt{T}}\right)\right)-1\right)\exp\left(\alpha+\beta t/T\right)\cos\left(2\left(\omega t+\phi\right)\right)\right]\right\}\\
\longrightarrow\quad &\exp\left(0\right)=1 \text{ as } T\longrightarrow\infty. \qquad (3.3.9)
\end{aligned}$$

For all sufficiently large T,

$$
\begin{aligned}
&\sum_{t=1}^{T}\left[\exp\left(\frac{\lambda i\left(\gamma+\delta t/T\right)}{\sqrt{T}}\right)-1-\frac{\lambda i\left(\gamma+\delta t/T\right)}{\sqrt{T}}-\frac{-\lambda^{2}\left(\gamma+\delta t/T\right)^{2}}{2T}\right]\\
\times\quad &\exp\left(\alpha+\beta t/T\right)\\
=\quad &\sum_{t=1}^{T}\left(\sum_{k=3}^{\infty}\frac{\left(\lambda i\left(\gamma+\delta t/T\right)\right)^{k}}{T^{\frac{k}{2}}k!}\right)\exp\left(\alpha+\beta t/T\right)\\
\leq\quad &\sum_{t=1}^{T}\left|\frac{\lambda}{\sqrt{T}}\left(\gamma+\frac{\delta t}{T}\right)\right|^{3}\exp\left(\alpha+\beta t/T\right)\\
=\quad &\sum_{t=1}^{T}\frac{1}{T\sqrt{T}}\left|\lambda^{3}\left(\gamma+\frac{\delta t}{T}\right)^{3}\right|\exp\left(\alpha+\beta t/T\right)\\
\leq\quad &\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\left|\lambda\left(\gamma+\delta\right)\right|^{3}\exp\left(\alpha+\beta t/T\right)\\
\leq\quad &\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\left|\lambda\left(\gamma+\delta\right)\right|^{3}\exp\left|\alpha+\beta\right|\\
=\quad &\frac{\left|\lambda\left(\gamma+\delta\right)\right|^{3}\exp\left|\alpha+\beta\right|}{\sqrt{T}}\longrightarrow 0\ \text{ as }\ T\longrightarrow\infty. \qquad (3.3.10)
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\Bigg|\sum_{t=1}^{T}\left[\exp\left(\frac{\lambda i\left(\gamma+\delta t/T\right)}{\sqrt{T}}\right)-1-\frac{\lambda i\left(\gamma+\delta t/T\right)}{\sqrt{T}}\right]\exp\left(\alpha+\beta t/T\right)\\
&-\sum_{t=1}^{T}\frac{-\lambda^{2}\left(\gamma+\delta t/T\right)^{2}}{2T}\exp\left(\alpha+\beta t/T\right)\Bigg|\\
\longrightarrow\quad &0\ \text{ as } T\longrightarrow\infty. \qquad (3.3.11)
\end{aligned}
$$

By the Integral Approximation Rule for series, the last expression

$$
\sum_{t=1}^{T}\frac{-\lambda^{2}\left(\gamma+\delta t/T\right)^{2}}{2T}\exp\left(\alpha+\beta t/T\right)
$$

converges to

$$\int_0^1 \frac{-\lambda^2(\gamma+\delta x)^2}{2}\exp(\alpha+\beta x)\,\mathrm{d}x = \int_0^1 \frac{\lambda^2(\gamma+\delta x)^2}{2}\exp(\alpha+\beta x)\,\mathrm{d}x$$

$$= -\frac{\lambda^2\exp(\alpha)}{2}\left(\left[\frac{(\gamma+\delta x)^2}{\beta}\exp(\beta x)\right]_0^1 - \int_0^1 \frac{2\delta(\gamma+\delta x)}{\beta}\exp(\beta x)\,\mathrm{d}x\right)$$

$$= -\frac{\lambda^2\exp(\alpha)}{2}\left(\left[\left(\frac{(\gamma+\delta x)^2}{\beta}-\frac{2\delta(\gamma+\delta x)}{\beta^2}\right)\exp(\beta x)\right]_0^1 + \int_0^1 \frac{2\delta^2}{\beta^2}\exp(\beta x)\,\mathrm{d}x\right)$$

$$= -\frac{\lambda^2\exp(\alpha)}{2}\left[\left(\frac{(\gamma+\delta x)^2}{\beta}-\frac{2\delta(\gamma+\delta x)}{\beta^2}+\frac{2\delta^2}{\beta^3}\right)\exp(\beta x)\right]_0^1$$

$$= \frac{\lambda^2 e^\alpha}{2}\left(\frac{2\delta(\gamma+\delta)-\beta(\gamma+\delta)^2}{\beta^2}e^\beta + \frac{2\delta^2}{\beta^3}\left(1-e^\beta\right) + \frac{\beta\gamma^2-2\delta\gamma}{\beta^2}\right). \quad (3.3.12)$$

The exponential of this is the characteristic function of a Normal random variable with zero mean and variance

$$\frac{e^\alpha}{\beta^3}\left(\left(2\delta\beta(\gamma+\delta)-\beta(\gamma+\delta)^2\right)e^\beta + 2\delta^2\left(1-e^\beta\right)+\beta\left(\beta\gamma^2-2\delta\gamma\right)\right).$$

3.4 Consistency of the latent ACF estimator and its DFT

In Section 2.4, we investigated the method-of-moments estimator of the latent-process covariance function proposed in Zeger (1988), after obtaining good empirical evidence for the maximiser of a DFT of these estimates as an estimator for ω. An alternative to Zeger's covariance estimator is

$$\bar{\gamma}(s) = \frac{1}{T-s}\sum_{t=1}^{T-s}\left(\frac{y_t}{\hat{\mu}_t}-1\right)\left(\frac{y_{t+s}}{\hat{\mu}_{t+s}}-1\right).$$

The real part of the DFT of the resulting covariance estimates, which we shall denote the "real Fourier transform" (RFT) appears very similar to that of Zeger's covariance estimates, while the structure of $\bar{\gamma}(s)$ as the sample covariance of variable ratios, rather than the ratio of two sample covariances estimator, makes mean and variance computation much more straightforward.

Although, due to the log-normal distribution of $\{\hat{\mu}_t\}$, the covariance estimates will not be unbiased, we shall show that the maximiser $\hat{\omega}$ of the RFT of $\{\hat{\gamma}(s)\}$ still converges in probability to the true frequency ω.

Let

$$\begin{aligned}\tilde{R}_T(\theta) &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\right)\cos(\theta s),\\ \bar{R}_T(\theta) &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\bar{\varepsilon}_t\bar{\varepsilon}_{t+s}\right)\cos(\theta s)\end{aligned}$$

and

$$R_T(\theta) = \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\cos(2\omega s)\cos(\theta s)$$

where $\tilde{\varepsilon}_t = y_t\exp\left(-\mathbf{x}_t^T\boldsymbol{\theta}_0\right) - 1$ and $\bar{\varepsilon}_t = y_t\exp\left(-\mathbf{x}_t^T\hat{\boldsymbol{\theta}}\right) - 1$.

These three summations can be thought of as the RFTs of the covariance estimates using the true and estimated means respectively and the RFT of the actual latent-process covariances.

After showing that $R_T(\theta)$ is uniquely maximised at ω, it is sufficient to prove that $\bar{R}_T(\theta)$ converges in probability to $R_T(\theta)$. This in turn will be established by first proving that $\left|\bar{R}_T(\theta) - R_T(\theta)\right| \underset{P}{\longrightarrow} 0$ using Chebychev's inequality, then proving that $\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| \underset{P}{\longrightarrow} 0$ via a step-by-step process using the asymptotic normality of $\hat{\boldsymbol{\theta}}_{GLM}$.

We shall first prove that $|\tilde{R}_T(\theta) - R(\theta)| \underset{P}{\longrightarrow} 0$ and hence that

$$\tilde{R}_T(\theta) \underset{P}{\longrightarrow} r(\theta) = \lim_{\tau\to\infty}\left(R_T(\theta)\right) = \begin{cases} 0 & \theta \neq 2\omega \\ \frac{1}{2} & \theta = 2\omega\end{cases}$$

$$\begin{aligned}E\left(\tilde{R}_T(\theta)\right) &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}E\left(\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\right)\right)\cos(\theta s)\\ &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\frac{E\left[\left(y_t-\mu_t\right)\left(y_{t+s}-\mu_{t+s}\right)\right]}{\mu_t\mu_{t+s}}\right)\cos(\theta s) \qquad (3.4.1)\\ &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\frac{\cos(2\omega s)}{2}\right)\cos(\theta s) = \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\cos(2\omega s)\cos(\theta s).\end{aligned}$$

Thus $\tilde{R}_T(\theta)$ is an unbiased estimator of $R(\theta)$

$$\begin{aligned}\frac{1}{2\tau}\sum_{s=0}^{\tau-1}\cos(2\omega s)\cos(\theta s) &= \frac{1}{4\tau}\sum_{s=0}^{\tau-1}\cos\left((2\omega-\theta)s\right)+\cos\left((2\omega+\theta)s\right)\\ &= \frac{1}{4\tau}\left[\frac{\sin\left(\frac{\tau}{2}(2\omega-\theta)\right)\cos\left(\frac{\tau-1}{2}(2\omega-\theta)\right)}{\sin\left(\frac{2\omega-\theta}{2}\right)}+\frac{\sin\left(\frac{\tau}{2}(2\omega+\theta)\right)\cos\left(\frac{\tau+1}{2}(2\omega+\theta)\right)}{\sin\left(\frac{2\omega+\theta}{2}\right)}\right]\\ &\approx \begin{cases}0 & \text{if } \theta\neq 2\omega\\ \frac{1}{2} & \text{if } \theta=2\omega\end{cases} \qquad (3.4.2)\end{aligned}$$

This approximation is exact if θ and 2ω are both multiples of $\dfrac{2\pi}{\tau}$ and a limit as $\tau \longrightarrow \infty$ otherwise Hence $R(\theta)=E\left(\tilde{R}_T(\theta)\right)$ is maximised at $\theta=2\omega$

$$\begin{aligned}V\left(\tilde{R}_T(\theta)\right) &= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\left\{\sum_{t=1}^{T-s}\sum_{p=1}^{T-r}\left(\frac{E\left(\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\tilde{\varepsilon}_p\tilde{\varepsilon}_{p+r}\right)}{(T-s)(T-r)}-\frac{\cos(2\omega s)\cos(2\omega r)}{4}\right)\right\}\\ &\quad\times\cos(\theta s)\cos(\theta r). \qquad (3.4.3)\end{aligned}$$

Given that $t, t+s, p$ and $p+r$ are all different,

$$\begin{aligned}&E\left(\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\tilde{\varepsilon}_p\tilde{\varepsilon}_{p+r}\right)\\ =\ & E\Big(E\left(\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\tilde{\varepsilon}_p\tilde{\varepsilon}_{p+r}|\,\phi\right)\Big)\\ =\ & \frac{1}{\mu_t\mu_{t+s}\mu_p\mu_{p+r}}E\Big[E\left((y_t-\mu_t)\left(y_{t+s}-\mu_{t-s}\right)\left(y_p-\mu_p\right)\left(y_{p+r}-\mu_{p+r}\right)|\,\phi\right)\Big]\\ =\ & E\Big(\cos(2\omega t+2\phi)\cos(2\omega(t+s)+2\phi)\cos(2\omega p+2\phi)\cos(2\omega(p+r)+2\phi)\Big)\\ =\ & \frac{1}{4}E\Big[(\cos(4\omega t+2\omega s+4\phi)+\cos(2\omega s))(\cos(4\omega p+2\omega s+4\phi)+\cos(2\omega r))\Big]\\ =\ & \frac{1}{8}E\Big[2\cos(2\omega s)\cos(2\omega r)+\cos(4\omega(t-p)+2\omega(s-r))\Big]\\ =\ & \frac{\cos(2\omega s)\cos(2\omega r)}{4}+\frac{\cos(4\omega(t-p)+2\omega(s-r))}{8}. \qquad (3.4.4)\end{aligned}$$

Similar calculations for the cases where $t=p, t+s=p+r, t=p+r$ or $t+s=p$, yields

$$
\begin{aligned}
V\left(\tilde{R}_T(\theta)\right) &= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\left(\sum_{t=1}^{T-s}\sum_{p=1}^{T-r}\frac{\cos 2(2(t-p)+s-r)\omega}{8} + \sum_{t=1}^{T-s-r}\frac{\cos 2(s+r)}{2\mu_{t+s}}\right. \\
&\quad \left. + \sum_{p=1}^{T-s-r}\frac{\cos 2(s+r)\omega}{2\mu_{p+r}} + \sum_{t=1}^{T-\max(s,r)}\frac{\cos 2(s-r)\omega}{2}\left(\frac{1}{\mu_t}+\frac{1}{\mu_{t+s}}\right)\right) \\
&\quad \times\frac{\cos(\theta s)\cos(\theta r)}{(T-s)(T-r)} \\
&+ \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{t=1}^{T-s}\frac{(2+\cos(2\omega s))\cos^2(\theta s)}{2\mu_t\mu_{t+s}(T-s)^2}. \qquad (3.4.5)
\end{aligned}
$$

Considering the last term first,

$$
\begin{aligned}
&\frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{t=1}^{T-s}\frac{(2+\cos(2s\omega))\cos^2(s\theta)}{2\mu_t\mu_{t+s}(T-s)^2} \\
&= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{t=1}^{T-s}\frac{(2+\cos(2s\omega))\cos^2(s\theta)}{(T-s)^2}e^{-\alpha-\beta t/T}e^{-\alpha-\beta(t+s)/T} \\
&\leq \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{t=1}^{T-s}\frac{3e^{2(|\alpha|+|\beta|)}}{2(T-s)^2} \\
&= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\frac{3e^{2(|\alpha|+|\beta|)}}{2(T-s)} \\
&\leq \frac{3e^{2(|\alpha|+|\beta|)}}{2\tau(T-\tau)} \longrightarrow 0 \text{ as } T\longrightarrow\infty. \qquad (3.4.6)
\end{aligned}
$$

Using the fact that for any $\psi \in R$ and $s = 1, 2, \ldots, \tau$, $\left|\frac{\cos\psi}{\mu_t(T-s)}\right| \leq \frac{e^{|\alpha|+|\beta|}}{T-\tau}$, we have that

$$\begin{aligned}
&\frac{1}{(T-s)(T-r)}\left(\sum_{t=1}^{T-\max(s,r)} \frac{\cos 2(s-r)}{2}\left(\frac{1}{\mu_t}+\frac{1}{\mu_{t+s}}\right)\right. \\
&\left.+\sum_{t=1}^{T-s-r} \frac{\cos 2(s+r)}{2\mu_{t+s}}+\sum_{p=1}^{T-s-r} \frac{\cos 2(s+r)}{2\mu_{p+r}}\right) \\
\leq\ & \frac{(T-\max(s,r))e^{|\alpha|+|\beta|}+(T-s-r)e^{|\alpha|+|\beta|}}{(T-s)(T-r)} \\
\leq\ & \frac{2e^{|\alpha|+|\beta|}}{T-\tau} \longrightarrow 0 \\
&\frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\left(\sum_{t=1}^{T-\max(s,r)} \frac{\cos 2(s-r)}{2}\left(\frac{1}{\mu_t}+\frac{1}{\mu_{t+s}}\right)\right. \\
&\left.+\sum_{t=1}^{T-s-r} \frac{\cos 2(s+r)}{2\mu_{t+s}}+\sum_{p=1}^{T-s-r} \frac{\cos 2(s+r)}{2\mu_{p+r}}\right)\frac{\cos(\theta s)\cos(\theta r)}{(T-s)(T-r)} \\
\leq\ & \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\frac{2e^{|\alpha|+|\beta|}}{T-\tau} \\
=\ & \frac{2e^{|\alpha|+|\beta|}}{T-\tau} \longrightarrow 0 \text{ as } T \longrightarrow \infty\ . \qquad (3.4.7)
\end{aligned}$$

Finally,

$$\sum_{t=1}^{T-s}\sum_{p=1}^{T-r}\frac{\cos 2(2(t-p)+s-r)\omega}{8} = \frac{\sin 2(T-r)\omega \sin 2(T-s)\omega}{8\sin^2(2\omega)} \leq \frac{1}{8\sin^2(2\omega)}.$$

This implies that

$$\begin{aligned}
&\frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\left(\sum_{t=1}^{T-s}\sum_{p=1}^{T-r}\frac{\cos 2(2(t-p)+s-r)}{8}\right)\frac{\cos(\theta s)\cos(\theta r)}{(T-s)(T-r)} \\
\leq\ & \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\frac{1}{8(T-s)(T-r)\sin^2(2\omega)} \\
\leq\ & \frac{1}{8(T-\tau)^2\sin^2(2\omega)} \longrightarrow 0 \\
&\text{as } T \longrightarrow \infty. \qquad (3.4.8)
\end{aligned}$$

To establish the convergence in probability of $|\bar{R}_T(\theta)-\tilde{R}_T(\theta)|$ to 0, it is sufficient to prove that for any $\epsilon > 0$, $P\left(\left|\bar{R}_T(\theta)-\tilde{R}_T(\theta)\right| > \epsilon\right) \longrightarrow 0$ as $T \longrightarrow \infty$. This is best accomplished via a series of short steps.

Step 1: Let $M_T = \sup\{y_1, y_2, \ldots, y_T\}$. For any $\kappa > 0$,

$$\begin{aligned}
P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon\right) &= P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T \leq T^\kappa\right) \\
&+ P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T > T^\kappa\right) \\
&\leq P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T \leq T^\kappa\right) + P\left(M_T > T^\kappa\right) \\
&= P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \mid M_T \leq T^\kappa\right) P\left(M_T \leq T^\kappa\right) \\
&+ P\left(M_T > T^\kappa\right). \qquad (3.4.9)
\end{aligned}$$

Step 2: Let $\tilde{Z} \sim Po(\lambda)$ where $\lambda = 2e^{\alpha+|\beta|}$.

Then for all $0 < t \leq T$, $2\exp(\alpha + \beta t/T) \leq \lambda$.

Therefore for all $k > 0$ and $0 < \phi < 2\pi$, $P\left(y_t > k \mid \phi\right) \leq P\left(\tilde{Z} > k\right)$.

By Markov's inequality,

$$\begin{aligned}
P\left(y_t > T^\kappa \mid \phi\right) &\leq P\left(\tilde{Z} > T^\kappa\right) \\
&\leq \frac{E\left(s^{\tilde{Z}}\right)}{s^{T^\kappa}} = \frac{\exp\left((s-1)\lambda\right)}{s^{T^\kappa}}. \qquad (3.4.10)
\end{aligned}$$

Taking $s = 2$, $P\left(y_t > T^\kappa \mid \phi\right) \leq \dfrac{\exp(\lambda)}{2^{T^\kappa}}$

Let $\tilde{M}_T = \sup\left\{\tilde{Z}_t\right\}$

$$\begin{aligned}
P\left(\tilde{M}_T > x\right) &= P\left(\bigcup_{t=1}^{T}\left\{\tilde{Z}_t > x\right\}\right) \leq \sum_{t=1}^{T-s} P\left(\tilde{Z}_t > x\right) \\
&= TP\left(\tilde{Z} > x\right)\left(\leq \frac{T}{2^x}e^\lambda\right). \qquad (3.4.11)
\end{aligned}$$

Thus for all ϕ, $P\left(M_T > T^\kappa|\phi\right) \leq P\left(\tilde{M}_T > T^\kappa\right)$ which gives

$$\begin{aligned}
P\left(M_T > T^\kappa\right) &= \int_0^{2\pi} P\left(M_T > T^\kappa|\phi\right) f(\phi).\mathrm{d}\phi \\
&\leq \int_0^{2\pi} P\left(\tilde{M}_T > T^\kappa\right) f(\phi).\mathrm{d}\phi \\
&= P\left(\tilde{M}_T > T^\kappa\right) \int_0^{2\pi} f(\phi).\mathrm{d}\phi \\
&= P\left(\tilde{M}_T > T^\kappa\right) \\
&\longrightarrow 0 \text{ as } T \longrightarrow \infty. \qquad (3.4.12)
\end{aligned}$$

Thus $P\left(M_T > T^\kappa\right) \le P\left(\tilde{M}_T > T^\kappa\right)$, so $P\left(M_T > T^\kappa\right) \le \dfrac{T}{2^{T^\kappa}} e^\lambda \longrightarrow 0$ as $T \longrightarrow \infty$.

Step 3: Conditioning on $M_T \le T^\kappa$

$$\begin{aligned}
\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| &\le \frac{1}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s}\left|\bar{\varepsilon}_t\bar{\varepsilon}_{t+s} - \tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\right| \\
&\le \frac{1}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s} y_t y_{t+s}\left|e^{-(\mathbf{x}_t+\mathbf{x}_{t+s})\hat{\boldsymbol{\theta}}} - e^{-(\mathbf{x}_t+\mathbf{x}_{t+s})\boldsymbol{\theta}_0}\right| \\
&+ \frac{1}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s}\left\{y_t\left|e^{-\mathbf{x}_t\hat{\boldsymbol{\theta}}} - e^{-\mathbf{x}_t\boldsymbol{\theta}_0}\right| + y_{t+s}\left|-e^{\mathbf{x}_{t+s}\hat{\boldsymbol{\theta}}} - e^{\mathbf{x}_{t+s}\boldsymbol{\theta}_0}\right|\right\} \\
(\text{if } M_T \le T^\kappa) &\le \frac{1}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s} T^{2\kappa}\left|e^{-(\mathbf{x}_t+\mathbf{x}_{t+s})\hat{\boldsymbol{\theta}}} - e^{-(\mathbf{x}_t+\mathbf{x}_{t+s})\boldsymbol{\theta}_0}\right| \\
&+ \frac{1}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s} T^{\kappa}\left\{\left|e^{-\mathbf{x}_t\hat{\boldsymbol{\theta}}} - e^{-\mathbf{x}_t\boldsymbol{\theta}_0}\right| + \left|-e^{\mathbf{x}_{t+s}\hat{\boldsymbol{\theta}}} - e^{\mathbf{x}_{t+s}\boldsymbol{\theta}_0}\right|\right\} \\
&= N_{T1} + N_{T2} + N_{T3} \\
&= N_T. \qquad (3.4.13)
\end{aligned}$$

Note that

$$\begin{aligned}
&P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \,\middle|\, M_T \le T^\kappa\right) P\left(M_T \le T^\kappa\right) \\
&\quad \le P\left(N_T > \epsilon\right) P\left(M_T \le T^\kappa\right) \le P\left(N_T > \epsilon\right). \qquad (3.4.14)
\end{aligned}$$

Step 4: Proof of $P\left(N_T > \epsilon\right) \longrightarrow 0$.

For all $0 < \xi < \frac{1}{4}$,

$$T^{2\xi}\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \underset{P}{\longrightarrow} 0. \qquad (3.4.15)$$

Thus

$$\begin{aligned}
P\left(N_{T1} > \epsilon\right) &\le P\left(N_{T1} > \epsilon | T^\xi\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \le T^{-\xi}\right) + P\left(T^\xi\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| > T^{-\xi}\right) \\
& P\left(T^\xi\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \ge T^{-\xi}\right) \longrightarrow 0 \text{ as } T \longrightarrow \infty \qquad (3.4.16)
\end{aligned}$$

by the central limit theorem.

Conditioning upon $T^\xi\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \le T^{-\xi}$

$$\begin{aligned}
&\left|\exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\hat{\boldsymbol{\theta}}\right) - \exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\boldsymbol{\theta}_0\right)\right| \\
&= \exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\boldsymbol{\theta}_0\right)\left|\exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right) - 1\right|. \qquad (3.4.17)
\end{aligned}$$

Since $\mathbf{x}_t = \binom{1}{t/T}$,

$$\begin{aligned}\left|(\mathbf{x}_t + \mathbf{x}_{t+s})\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right| &= \left|2\left(\hat{\alpha} - \alpha_0\right) + \frac{2t+s}{T}\left(\hat{\beta} - \beta_0\right)\right| \\ &\leq 2\left\{\left|\hat{\alpha} - \alpha_0\right| + \left|\hat{\beta} - \beta_0\right|\right\} \\ &\leq 4\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|. \qquad (3.4.18)\end{aligned}$$

For all sufficiently large T,

$$\begin{aligned}\left|\exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right) - 1\right| &\leq 2\left|(\mathbf{x}_t + \mathbf{x}_{t+s})\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right| \\ &\leq 2\left\{4\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|\right\} \\ &\leq 8T^{-2\xi}. \qquad (3.4.19)\end{aligned}$$

Also, $\exp\left(-\left(\mathbf{x}_t + \mathbf{x}_{t+s}\right)\boldsymbol{\theta}_0\right) \leq \exp\left(2|\alpha_0| + 2|\beta_0|\right)$.

Therefore conditioning upon $\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \leq T^{-2\xi}$, for any $0 < \kappa < \xi$

$$\begin{aligned}N_{T1} &\leq \frac{8}{\tau}\sum_{s=0}^{\tau-1}\frac{1}{T-s}\sum_{t=1}^{T-s}T^{2\kappa}\exp\left(2|\alpha_0| + 2|\beta_0|\right)T^{-2\xi} \\ &= T^{2(\kappa-\xi)}\exp\left(2|\alpha_0| + 2|\beta_0|\right) \\ &\longrightarrow 0 \text{ as } T \longrightarrow \infty \text{ for any } \kappa < \xi. \qquad (3.4.20)\end{aligned}$$

Thus $P\left(N_{T1} > \epsilon \middle| T^{\xi}\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| \leq T^{-\xi}\right) \longrightarrow 0$ as $T \longrightarrow \infty$.

The same argument applies for N_{T2} and N_{T3}.

Thus $P\left(|N_T| > \epsilon\right) \longrightarrow 0$ as $T \longrightarrow \infty$.

As

$$\begin{aligned}\left|\hat{R}_T - r_T\right| &\leq \left|\hat{R}_T - \tilde{R}_T\right| + \left|\tilde{R}_T - R_T\right| + \left|R_T - r_T\right| \\ P\left(\left|\hat{R}_T - \tilde{R}_T\right| > \epsilon/2\right) &\leq P(|N_T| > \epsilon/2) \longrightarrow 0 \text{ as } T \longrightarrow \infty \\ P\left(\left|\tilde{R}_T - R_T\right| > \epsilon\right) &\longrightarrow 0 \qquad (3.4.21)\end{aligned}$$

it is now clear that $P\left(\left|\bar{R}_T\right| > \epsilon/2\right) \longrightarrow 0$. Our estimator for the frequency of our model is convergent in probability to the true frequency. This result, although thoroughly verified only for the intercept-and-trend case, holds for any log-linear mean function provided that the regressors are all bounded. This can be seen by assuming all terms in $\mathbf{x}_t$ are in $|-1,1|$, replacing all $e^{\alpha+\beta t/T}$ terms by $e^{\alpha+\beta t/T+\gamma u_t}$, bounding the latter by $e^{\alpha+|\gamma|+\beta t/T}$ and proceeding.

3.5 Extension of results on $\hat{\boldsymbol{\theta}}_{GLM}$ to include random-valued regressors

Although trend and intercept parameters are needed in most cases to obtain an overall picture of the data, the inclusion of other regressors will often be very useful. The incidence of almost anything related to accidents (car insurance, traffic accidents or hospital admissions) will have some dependence on the weather and the number of cases seen of most diseases are related to population size. We will call such realisations of random variables *random-valued regressors*

The asymptotic results established previously are all proven by studying the limiting behaviour of functions of summations, an approach that is easily extendable to models with random-valued covariates. As covariates like population sizes are stochastic rather than deterministic like time covariates, one however can not directly compute summations of such regressors or their exponentials. This problem will be bypassed using assumptions on the distribution and deviance of any random-valued regressors, as well as further applications of Chebychev's inequality and Abel's Lemma, to construct bounds for these summations.

3.5.1 Consistency of $\hat{\boldsymbol{\theta}}_{GLM}$ for the model with random-valued regressors

The convergence of the variance of the log-likelihood $V\left(l_T(\boldsymbol{\theta})\right)$ to zero as $T \longrightarrow \infty$ is a sufficient condition for $\{\hat{\boldsymbol{\theta}}_{GLM} - \boldsymbol{\theta}_0\}$ to be pointwise convergent to 0 and equicontinuous in probability, which together establish uniform convergence of $\hat{\boldsymbol{\theta}}_{GLM}$ to $\boldsymbol{\theta}_0$. Looking at the methods used to establish the consistency of trend and intercept parameters, it is clear in turn that

$$f(T) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^T\boldsymbol{\theta}\exp(\mathbf{x}_t^T\boldsymbol{\theta})\cos(2\omega t) \propto \sqrt{V(l(\boldsymbol{\theta}))} \longrightarrow 0 \text{ as } T \longrightarrow \infty$$

is a sufficient condition for the convergence of $V\left(l_T(\boldsymbol{\theta})\right)$ to zero, provided that $|\mathbf{x}_t^T\boldsymbol{\theta}|$ is bounded

Thus for a model with a single random-valued regressor, we shall examine the limiting behaviour of its likelihood variance via that of $f(T)$.

We will be working under the assumption that for any random-valued regressor x_t, $u_t = x_t - E(x_t)$ has a Normal distribution and can be expressed as a causal and invertible ARMA process with $E(x_t) = P_t + S_t$ where P_t is a finite-degree polynomial of time and S_t is a seasonality function consisting either of a step process (for the daily, weekly or monthly mean rainfall in a country with a "wet" and a "dry" season) or a first-degree trigonometric polynomial with known period (for hours of sunshine or temperature in a temperate country). Both parts can be estimated via least squares.

First consider the case when x_t is zero-meaned, with autocorrelation function $R(t, t \pm s) = \sigma^2 \rho(s)$.

Note that

$$
\begin{aligned}
f(T) &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^T \boldsymbol{\theta} \exp\left(\mathbf{x}_t^T \boldsymbol{\theta}\right) \cos(2\omega t) \\
&= \frac{1}{T}\sum_{t=1}^{T} (\alpha + \beta t/T + \gamma u_t) e^{\alpha + \beta t/T + \gamma u_t} \cos(2\omega t)
\end{aligned}
$$

and that

$$
\begin{aligned}
\tilde{f}(T) &= E\left(f(T)\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\alpha + \beta t/T + \gamma^2\sigma^2\right) e^{\alpha + \beta t/T + \gamma^2\sigma^2/2} \cos(2\omega t) \longrightarrow 0. \qquad (3.5.1)
\end{aligned}
$$

By Chebychev's inequality,

$$
\begin{aligned}
& P\left(|f(T) - \tilde{f}(T)| > \epsilon\right) \leq \frac{V(f(T))}{\epsilon^2} \\
&= \frac{1}{T^2\epsilon^2}\sum_{s=1}^{T}\sum_{t=1}^{T} e^{2\alpha + \beta(t+s)/T} e^{\gamma^2\sigma^2} \Big[\gamma^2 R(t-s) - \left(\alpha + \beta t/T + \gamma^2\sigma^2\right)\left(\alpha + \beta s/T + \gamma^2\sigma^2\right) \\
&\quad + e^{\gamma^2 R(t-s)} \Big(\alpha + \beta t/T + \gamma^2(\sigma^2 + R(t-s))\Big)\Big(\alpha + \beta s/T + \gamma^2(\sigma^2 + R(t-s))\Big)\Big] \\
&\quad \times \cos(2(t-s))\omega \\
&= \sum_{s=1}^{T}\sum_{t=s}^{T} \frac{2e^{\alpha+\beta t/T} e^{\alpha + \beta s/T}}{T^2\epsilon^2} e^{\gamma^2\sigma^2} \Big[\gamma^2 R(t-s) e^{\gamma^2 R(t-s)} \left(2\alpha + \beta(t+s)/T + 2\sigma^2\right) \\
&\quad + \left(e^{\gamma^2 R(t-s)} - 1\right)\left(\alpha + \beta t/T + \gamma^2\sigma^2\right)\left(\alpha + \beta s/T + \gamma^2\sigma^2\right) + \gamma^4 R(t-s)^2 \\
&\quad + \gamma^2 R(t-s)\Big] \cos(2(t-s)\omega, \qquad (3.5.2)
\end{aligned}
$$

where $R(t-s) = cov(u_t, u_s)$ for all $t \geq s$.

For any causal ARMA process, there exists $r \in (0,1)$ and a constant $0 < K < \infty$ such that for all $h = 0, 1, \ldots$, $|R(h)| \leq K r^h$. Thus

$$\begin{aligned}
& \frac{2}{T} \sum_{s=1}^{T}\sum_{t=s}^{T} e^{2\alpha+\frac{\beta}{T}(t+s)T} e^{\gamma^2\sigma^2}\Big[\left(e^{\gamma^2 R(t-s)}-1\right)\left(\alpha+\beta t/T+\gamma^2\sigma^2\right)\left(\alpha+\beta s/T+\gamma^2\sigma^2\right) \\
& \quad +\gamma^2 R(t-s)+\gamma^2 R(t-s) e^{\gamma^2 R(t-s)}\left(2\alpha+\beta(t+s)/T+2\sigma^2\right)+\gamma^4 R(t-s)^2\Big] \\
& \quad \times \cos(2(t-s)\omega) \\
\leq & \frac{2}{T} \sum_{s=1}^{T}\sum_{t=s}^{T} e^{2\alpha+\frac{\beta}{T}(t+s)T} e^{\gamma^2\sigma^2}\Big[\left(e^{\gamma^2 R(t-s)}-1\right)\left(\alpha+\beta t/T+\gamma^2\sigma^2\right)\left(\alpha+\beta s/T+\gamma^2\sigma^2\right) \\
& \quad +\gamma^2 R(t-s)+\gamma^2 R(t-s) e^{\gamma^2 R(t-s)}\left(2\alpha+\beta(t+s)/T+\sigma^2\right)+\gamma^4 R(t-s)^2\Big] \\
\leq & \frac{2}{T} \sum_{s=1}^{T}\sum_{t=s}^{T} e^{2\alpha+\frac{\beta}{T}(t+s)T} e^{\gamma^2\sigma^2}\Big[\left(e^{K\gamma^2 r^{t-s}}-1\right)\left(\alpha+\beta t/T+\gamma^2\sigma^2\right)\left(\alpha+\beta s/T+\gamma^2\sigma^2\right) \\
& \quad +K\gamma^2 r^{t-s}+K\gamma^2 r^{t-s} e^{K\gamma^2 r^{t-s}}\left(2\alpha+\beta(t+s)/T+\sigma^2\right)+K^2\gamma^4 r^{2(t-s)}\Big]. \qquad (3.5.3)
\end{aligned}$$

All parts inside the square brackets are proportional to r^{t-s}, $r^{2(t-s)}e^{K\gamma^2 r^{t-s}}$, $\left(e^{K\gamma^2 r^{t-s}}-1\right)(\alpha+\beta t/T)(\alpha+\beta/T)$ or $(\alpha+\beta t/T)\, r^{t-s}e^{K\gamma^2 r^{t-s}}$.

Considering these four terms in turn, we have that

$$\begin{aligned}
& \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T} r^{t-s} e^{2\alpha+\frac{\beta}{T}(t+s)} \leq \frac{e^{2(|\alpha|+|\beta|)}}{T^2}\sum_{s=1}^{T} r^{-s}\sum_{t=s}^{T} r^t \\
= & \frac{e^{2(|\alpha|+|\beta|)}}{T^2}\sum_{s=1}^{T} r^{-s}\frac{r^{T+1}-r^s}{r-1} \\
\propto & \frac{1}{T^2}\sum_{s=1}^{T}\frac{r^{T+1-s}-1}{r-1} = \frac{1}{T(1-r)} - \frac{r^{T+1}(r^{-T}-1)}{T^2(1-r)^2} \\
= & \frac{1}{T(1-r)} - \frac{r(1-r^T)}{T^2(1-r)^2} \longrightarrow \frac{1}{0} \text{ as } T \longrightarrow \infty. \qquad (3.5.4)
\end{aligned}$$

Similarly we can show

$$\begin{aligned}
& \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T} r^{2(t-s)} e^{K\gamma^2 r^{2(t-s)}} e^{2\alpha+\frac{\beta}{T}(t+s)} \\
\leq & \frac{e^{K\gamma^2}}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T} (r^2)^{t-s} e^{2\alpha+\frac{\beta}{T}(t+s)T} \longrightarrow 0 \text{ as } T \longrightarrow \infty, \qquad (3.5.5)
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T}(\alpha+\beta t/T)\, r^{t-s} e^{K\gamma^2 r^{t-s}} e^{2\alpha+\frac{\beta}{T}(t+s)} \\
\leq\; & \frac{(|\alpha|+|\beta|)e^{K\gamma^2}}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T} r^{t-s} e^{2\alpha+\frac{\beta}{T}(t+s)} \longrightarrow 0 \text{ as } T\longrightarrow\infty. \qquad (3.5.6)
\end{aligned}$$

Using the identity that for any $x \leq K$ $|e^x - 1| \leq e^K x$

$$\begin{aligned}
& \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T}\left(\alpha+\frac{\beta t}{T}\right)\left(\alpha+\frac{\beta s}{T}\right) e^{2\alpha+\frac{\beta}{T}(t+s)}\left(e^{K\gamma^2 r^{t-s}}-1\right) \\
\leq\; & (|\alpha|+|\beta|)^2\, e^{|\alpha|+|\beta|}\sum_{s=1}^{T}\sum_{t=s}^{T}\left(e^{K\gamma^2 r^{t-s}}-1\right) \\
\leq\; & (|\alpha|+|\beta|)^2\, e^{|\alpha|+|\beta|}\sum_{s=1}^{T}\sum_{t=s}^{T} K\gamma^2 r^{t-s} e^{K\gamma^2} \longrightarrow 0 \text{ as } T\longrightarrow\infty. \qquad (3.5.7)
\end{aligned}$$

Thus $|f(T)-\tilde{f}(T)| \longrightarrow 0$ as $T \longrightarrow \infty$ when x_t is a stationary ARMA process, which combined with $\tilde{f}(T) \longrightarrow 0$ implies that $f(T) \longrightarrow 0$ as $T \longrightarrow \infty$. This is sufficient to establish consistency of $\hat{\boldsymbol{\theta}}$.

To extend this to the case when the random regressor has a non-zero mean, it is sufficient to establish the convergence to zero of $f(T)$ when $\mathbf{x}_t^T\boldsymbol{\theta} = \alpha+\beta t/T + P_t + S_t$. The trend component P_t can be incorporated into the trend function of the mean and S_t can be incorporated into the intercept when it is a step function. Therefore the problem can be reduced to proving the convergence to zero of $f(T)$ when $\mathbf{x}_t^T\boldsymbol{\theta} = \alpha + \beta t/T + \psi\cos(\theta t+\delta)$, where $\theta = 2\pi/P$ is a known frequency corresponding to an integer period P.

$$\begin{aligned}
f(T) &= \frac{1}{T}\sum_{t=1}^{T}(\alpha+\beta t/T+\psi\cos(\theta t+\delta))\, e^{\alpha+\beta t/T+\psi\cos(\theta t+\delta)}\cos(2\omega t) \\
&= \frac{1}{T}\sum_{t=1}^{T}(\alpha+\beta t/T)\, e^{\alpha+\beta t/T} e^{\psi\cos(\theta t+\delta)}\cos(2\omega t) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\psi\cos(\theta t+\delta) e^{\alpha+\beta t/T} e^{\psi\cos(\theta t+\delta)}\cos(2\omega t). \qquad (3.5.8)
\end{aligned}$$

Considering that $\cos\left(\frac{2\pi}{P}(kP+t)+\delta\right) = \cos\left(\frac{2\pi}{P}t+\delta\right)$ for any values of t and k and expressing T as $MP+N$ for some $N \leq P$, the above expression can be rewritten as

$$\begin{aligned} f(t) &= \frac{1}{T}\sum_{k=1}^{P}\left[\sum_{t=0}^{M-1}\left(\alpha+\frac{\beta(Pt+k)}{T}\right)e^{\alpha+\beta(Pt+k)/T}\cos(2\omega(Pt+k))\right. \\ &\quad \left. +\psi\cos(\theta k+\delta)\sum_{t=0}^{M-1}e^{\alpha+\beta(Pt+k)/T}\cos(2\omega(Pt+k))\right]e^{\psi\cos(\theta k+\delta)} \\ &+ \frac{1}{T}\sum_{t=1}^{N}\left(\alpha+\frac{\beta(MP+t)}{T}+\psi\cos(\theta t+\delta)\right)e^{\alpha+\beta(MP+t)/T+\psi\cos(\theta t+\delta)} \\ &\times \cos(2\omega(MP+t)) \end{aligned} \tag{3.5.9}$$

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{N}\left(\alpha+\frac{\beta(MP+t)}{T}+\psi\cos(\theta t+\delta)\right)e^{\alpha+\beta(MP+t)/T+\psi\cos(\theta t+\delta)}\cos(2\omega(MP+t)) \\ \leq\ &\frac{N}{T}\left(|\alpha|+|\beta|+|\psi|\right)e^{|\alpha|+|\beta|+|\psi|} \longrightarrow 0. \end{aligned} \tag{3.5.10}$$

$$\begin{aligned} &\sum_{t=0}^{M-1}e^{\alpha+\beta(Pt+k)/T}\cos(2\omega(Pt+k)) = \frac{e^{\alpha+\beta k/T}}{1+e^{2\beta P/T}-e^{\beta P/T}\cos(2\omega P)} \\ \times\ &\left(e^{\beta P(M+1)/T}\cos\left(2((M-1)P+k)\omega\right)-e^{\beta PM/T}\cos\left(2(MP+k)\omega\right)-e^{\beta P/T}+\cos\left(2\omega k\right)\right) \\ \leq\ &\frac{e^{\alpha+|\beta|}}{1+e^{2\beta P/T}-e^{\beta P/T}\cos(2\omega P)}\left(e^{|\beta|P(M+1)/T}+e^{|\beta|PM/T}+e^{|\beta|P/T}+1\right). \end{aligned} \tag{3.5.11}$$

This implies that

$$\begin{aligned} &\frac{1}{T}\sum_{k=1}^{P}\psi\cos(\theta k+\delta)e^{\psi\cos(\theta k+\delta)}\sum_{t=0}^{M-1}e^{\alpha+\beta(Pt+k)/T}\cos(2\omega(Pt+k)) \\ \leq\ &\frac{P|\psi|e^{\alpha+|\beta|+|\psi|}}{T\left(1+e^{2\beta P/T}-e^{\beta P/T}\cos(2\omega P)\right)}\left(e^{|\beta|P(M+1)/T}+e^{|\beta|PM/T}+e^{|\beta|P/T}+1\right) \\ \propto\ &\frac{P}{T}\longrightarrow 0. \end{aligned} \tag{3.5.12}$$

Similarly,

$$\begin{aligned} &\frac{1}{T}\sum_{k=1}^{P}\sum_{t=0}^{M-1}\left(\alpha+\frac{\beta(Pt+k)}{T}\right)e^{\alpha+\beta(Pt+k)/T+\psi\cos(\theta k+\delta)}\cos(2\omega(Pt+k)) \\ \longrightarrow\ &0 \text{ as } T\longrightarrow\infty. \end{aligned} \tag{3.5.13}$$

Therefore $f(T) \longrightarrow 0$ as $T \longrightarrow \infty$. This is sufficient proof that $\hat{\boldsymbol{\theta}}_{GLM}$ is a consistent estimator of the mean parameters of a model including trigonometric regressors. Together with the previous proof, it is a sufficient condition for the consistency of $\hat{\boldsymbol{\theta}}_{GLM}$ for a model including random-valued regressors with a seasonal mean function.

3.5.2 Asymptotic normality of $\hat{\boldsymbol{\theta}}_{GLM}$ with random-valued regressors

Let $P_T(\boldsymbol{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{A}\mathbf{x}_t^T(y_t-\mu_t)$ be any linear projection of the rescaled score function $\sqrt{T}S_T(\boldsymbol{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t^T(y_t-\mu_t)$.

As established earlier in Section 3.2, the convergence of the characteristic function $\varphi_X(\lambda) = E[\exp(i\lambda X)]$ of $P_T(\boldsymbol{\theta})$ to that of a Gaussian random variable is a sufficient condition for $\sqrt{T}S_T(\boldsymbol{\theta})$ to be asymptotically normal. By the Mean Value theorem, this in turn is a sufficient condition for $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{GLM}-\boldsymbol{\theta}_0\right)$ to converge in distribution to a mean zero normal random variable.

Thus to extend the asymptotic normality of $\hat{\boldsymbol{\theta}}_{GLM}$ from a trend and intercept parameter vector to one containing random-valued regressors satisfying certain conditions, we shall compute the characteristic function of $P_T(\boldsymbol{\theta})$ and examine its limiting behaviour in the case where $\mathbf{x}_t^T = (1, \frac{t}{T}, u_t)$ for a random-valued regressor $\{u_t\}$ in $[-1,1]$. As $\{u_t\}$ is a series of known values, it can always be bounded by rescaling.

$$\begin{aligned} E\big[\exp\left(i\lambda P_T(\boldsymbol{\theta})\right)\big] &= E\big[E\left(\exp\left(i\lambda P_T(\boldsymbol{\theta})\right)\right) \mid \phi\big] \\ &= E\left(\frac{1}{\sqrt{T}}\prod_{t=1}^{T}E\big[\exp\left((a+bt/T+cu_t)\,i\lambda\left(y_t-\mu_t\right)\right) \mid \phi\big]\right). \end{aligned} \tag{3.5.14}$$

Since $\{y_t|\phi\} \sim Po\left(\mu_t\varepsilon_t\right)$ where $\mu_t = \exp(\alpha+\beta t/T+\gamma u_t)$ and $\epsilon_t = 1+\cos\left(2(\omega t+\phi)\right)$

$$\begin{aligned} &\prod_{t=1}^{T}E\big[\exp\left((a+bt/T+cu_t)\,i\lambda\left(y_t-\mu_t\right)\right) \mid \phi\big] \\ &= \prod_{t=1}^{T}\exp\left[\mu_t\varepsilon_t\left(\exp\left(\frac{i\lambda}{\sqrt{T}}\left(a+bt/T+cu_t\right)\right)-1\right)-\frac{i\lambda}{\sqrt{T}}\left(a+bt/T+cu_t\right)\mu_t\right] \\ &= \exp\left[\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}\left(a+bt/T+cu_t\right)\right)-1-\frac{i\lambda}{\sqrt{T}}\left(a+bt/T+cu_t\right)\right)\right. \\ &\qquad \left.\times\ \exp\left(\alpha+\beta t/T+\gamma u_t\right)\right] \\ &\times\ \exp\left[\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}\left(a+bt/T+cu_t\right)\right)-1\right)\exp\left(\alpha+\beta t/T+\gamma u_t\right)\cos(2(\omega t+\phi))\right]. \end{aligned} \tag{3.5.15}$$

Thus

$$\begin{aligned}
&E\big(\exp[i\lambda P_T(\boldsymbol{\theta})]\big)\\
=\;& \exp\left[\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1-\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)\right.\\
&\left.\times\exp(\alpha+\beta t/T+\gamma u_t)\right]\times E\left\{\exp\left[\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1\right)\right.\right.\\
&\left.\left.\times\exp(\alpha+\beta t/T+\gamma u_t)\cos(2(\omega t+\phi))\right]\right\}. \qquad (3.5.16)
\end{aligned}$$

By (3.3.10) ,as for trend and intercept only,

$$\begin{aligned}
&\sum_{t=1}^{T}\left[\exp\left(\frac{\lambda i(a+bt/T+cu_t)}{\sqrt{T}}\right)-1-i\lambda\left(\frac{a+bt/T+cu_t}{\sqrt{T}}\right)\right]\exp(\alpha+\beta t/T+\gamma u_t)\\
-\;&\frac{-\lambda^2}{2}\sum_{t=1}^{T}\left[\frac{(a+bt/T+cu_t)^2}{T}\right]\exp(\alpha+\beta t/T+\gamma u_t)\\
\longrightarrow\;& 0 \text{ as } T\longrightarrow 0. \qquad (3.5.17)
\end{aligned}$$

Therefore a random variable X_T with characteristic function

$$\sum_{t=1}^{T}\left[\exp\left(\frac{\lambda i(a+bt/T+cu_t)}{\sqrt{T}}\right)-1-i\lambda\left(\frac{a+bt/T+cu_t}{\sqrt{T}}\right)\right]\exp(\alpha+\beta t/T+\gamma u_t) \qquad (3.5.18)$$

has the same limiting distribution as $T\longrightarrow 0$ as

$$Y_T \sim \sum_{t=1}^{T}\left[\frac{(a+bt/T+cu_t)^2}{T}\right]\exp(\alpha+\beta t/T+\gamma u_t). \qquad (3.5.19)$$

Using Abel's lemma ,

$$\begin{aligned}
&\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1\right)\exp(\alpha+\beta t/T+\gamma u_t)\cos(2(\omega t+\phi))\\
=\;&\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+b+cu_T)\right)-1\right)\exp(\alpha+\beta+\gamma u_T)\sum_{t=0}^{T}\cos(2(\omega t+\phi))\\
-\;&\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+cu_0)\right)-1\right)\exp(\alpha+\gamma u_0)\\
-\;&\sum_{t=0}^{T-1}\Delta\left[\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1\right)\exp(\alpha+\beta t/T+\gamma u_t)\right]\sum_{s=0}^{t}\cos(2(\omega s+\phi)), \qquad (3.5.20)
\end{aligned}$$

where $\Delta f(t)$ denotes the difference function $f(t+1)-f(t)$

Note that

$$
\begin{aligned}
& \Delta\left[\left(\exp\left(i\lambda\left(a+bt/T+cu_t\right)/\sqrt{T}\right)-1\right)\exp\left(\alpha+\beta t/T+\gamma u_t\right)\right] \\
= & \left[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\left(\exp\left(\beta/T\right)\exp\left(i\lambda b/T\sqrt{T}\right)\exp\left((\gamma+i\lambda c/T)u_{t+1}\right)\right.\right. \\
& \left.\left.-\exp\left((\gamma+i\lambda c/T)u_t\right)\right)-\left(\exp\left(\beta/T\right)\exp\left(\gamma u_{t+1}\right)-\exp\left(\gamma u_t\right)\right)\right]\exp\left(\alpha+\beta t/T\right) \\
= & \left[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\left\{\left(\exp\left((\gamma+i\lambda c/T)u_{t+1}\right)-\exp\left((\gamma+i\lambda c/T)u_t\right)\right)\right.\right. \\
& \left.+\left(\exp\left(\beta/T\right)\exp\left(i\lambda b/T\sqrt{T}\right)-1\right)\exp\left((\gamma+i\lambda c/T)u_{t+1}\right)\right\} \\
& \left.-\left\{\left(\exp\left(\beta/T\right)-1\right)\exp\left(\gamma u_{t+1}\right)+\left(\exp\left(\gamma u_{t+1}\right)-\exp\left(\gamma u_t\right)\right)\right\}\right]\exp\left(\alpha+\beta t/T\right).
\end{aligned}
\tag{3.5.21}
$$

$$
\begin{aligned}
& \sum_{t=0}^{T-1}\left[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\left(\exp\left(\beta/T\right)\exp\left(i\lambda b/T\sqrt{T}\right)-1\right)\right. \\
& \left.\times\exp\left((\gamma+i\lambda c/T)u_{t+1}\right)-\left(\exp\left(\beta/T\right)-1\right)\exp\left(\gamma u_{t+1}\right)\right]\exp\left(\alpha+\beta t/T\right) \\
= & \sum_{t=0}^{T-1}\left[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\left(\exp\left(\beta/T\right)\exp\left(i\lambda b/T\sqrt{T}\right)-1\right)\exp\left(i\lambda c/Tu_{t+1}\right)\right. \\
& \left.-\left(\exp\left(\beta/T\right)-1\right)\right]\exp\left(\gamma u_{t+1}\right)\exp\left(\alpha+\beta t/T\right).
\end{aligned}
\tag{3.5.22}
$$

Expressing each exponential term as a Taylor series,

$$
\begin{aligned}
& \left[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\left(\exp\left(\beta/T\right)\exp\left(i\lambda b/T\sqrt{T}\right)-1\right)\exp\left(i\lambda c/Tu_{t+1}\right)\right. \\
- & \left.\left(\exp\left(\beta/T\right)-1\right)\right] \\
= & \left\{1+\frac{i\lambda(a+bt/T)}{\sqrt{T}}-\frac{\lambda^2(a+bt/T)^2}{2T}+O\left(T^{-3/2}\right)\right\}\left\{\frac{\beta}{T}+\frac{i\lambda b}{T\sqrt{T}}+O\left(T^{-2}\right)\right\} \\
& \times\left\{1+\frac{i\lambda c u_{t+1}}{\sqrt{T}}-\frac{\lambda c^2u_{t+1}^2}{2T}+O\left(T^{-3/2}\right)\right\}-\left\{\frac{\beta}{T}+O\left(T^{-2}\right)\right\} \\
= & \left\{\frac{\beta}{T}+O\left(T^{-3/2}\right)\right\}-\left\{\frac{\beta}{T}+O\left(T^{-2}\right)\right\} \\
= & O\left(T^{-3/2}\right).
\end{aligned}
\tag{3.5.23}
$$

As $\sum_{s=0}^{t} \cos(2(\omega s+\phi)) \leq \frac{1}{|\sin(\omega)|}$, we have that

$$\begin{aligned}
&\sum_{t=0}^{T-1} \Delta\left[\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1\right)\exp\left(\alpha+\beta t/T+\gamma u_t\right)\right]\sum_{s=0}^{t}\cos(2(\omega s+\phi)) \\
\leq\ & \frac{1}{|\sin(\omega)|}\sum_{t=0}^{T-1} O\left(\frac{1}{T^{3/2}}\right)\exp\left(\gamma u_{t+1}\right)\exp\left(\alpha+\beta t/T\right) \\
\leq\ & \frac{\exp\left(|\gamma|+|\alpha|+|\beta|\right)}{|\sin(\omega)|}\sum_{t=0}^{T-1} O\left(\frac{1}{T^{3/2}}\right) \\
=\ & O\left(\frac{1}{\sqrt{T}}\right) \\
\longrightarrow\ & 0 \text{ as } T \longrightarrow \infty. \qquad (3.5.24)
\end{aligned}$$

By Abel's lemma, for any A, B

$$\begin{aligned}
&\sum_{t=0}^{T-1} \exp\left(A+Bt/T\right)\left(\exp\left(\gamma u_{t+1}\right)-\exp\left(\gamma u_t\right)\right) \\
=\ & \big(\exp\left(A+B\right)\exp\left(\gamma u_T\right)-\exp(A)\exp\left(\gamma u_0\right)\big) \\
-\ & \sum_{t=0}^{T-1}\left(\exp\left(A+B(t+1)/T\right)-\exp\left(A+Bt/T\right)\right)\exp\left(\gamma u_{t+1}\right). \qquad (3.5.25)
\end{aligned}$$

Thus

$$\begin{aligned}
&\sum_{t=0}^{T-1}\Bigg[\exp\left(i\lambda(a+bt/T)/\sqrt{T}\right)\big\{\exp\left((\gamma+i\lambda c/T)u_{t+1}\right)-\exp\left((\gamma+i\lambda c/T)u_t\right)\big\} \\
&-\big\{\exp\left(\gamma u_{t+1}\right)-\exp\left(\gamma u_t\right)\big\}\Bigg]\exp\left(\alpha+\beta t/T\right) \\
=\ & \Bigg[\big(\exp\left(\alpha+\beta\right)\exp\left(i\lambda(a+b)/\sqrt{T}\right)\exp\left((\gamma+i\lambda c/T)u_T\right) \\
&-\exp\left(\alpha\right)\exp\left(i\lambda a/\sqrt{T}\right)\exp\left((\gamma+i\lambda c/T)u_0\right)\big) \\
&-\big(\exp\left(\alpha+\beta\right)\exp\left(\gamma u_T\right)-\exp\left(\alpha\right)\exp\left(\gamma u_0\right)\big)\Bigg] \\
+\ & \Bigg\{\sum_{t=0}^{T-1}\exp\left(\alpha+\gamma u_{t+1}+\beta t/T\right)\left(\exp\left(\beta/T\right)-1\right) \\
-\ & \exp\left(\alpha+\gamma u_{t+1}+\beta t/T\right)\exp\left(\frac{\lambda i\left(a+cu_{t+1}+bt/T\right)}{\sqrt{T}}\right)\left(\exp\left(\frac{\beta+\lambda ib/\sqrt{T}}{T}\right)-1\right)\Bigg\}.
\end{aligned}$$
(3.5.26)

Hence

$$
\begin{aligned}
& \Big[\big(\exp(\alpha+\beta)\exp\left(i\lambda(a+b)/\sqrt{T}\right)\exp\left((\gamma+i\lambda c/T)u_T\right) \\
& -\exp(\alpha)\exp\left(i\lambda a/\sqrt{T}\right)\exp\left((\gamma+i\lambda c/T)u_0\right)\big) \\
& -\big(\exp(\alpha+\beta)\exp(\gamma u_T)-\exp(\alpha)\exp(\gamma u_0)\big)\Big] \\
= \; & \exp(\alpha+\beta+\gamma u_T)\left(\exp\left(\frac{\lambda i\,(a+b+cu_T)}{\sqrt{T}}\right)-1\right) \\
- \; & \exp(\alpha+\beta+\gamma u_0)\left(\exp\left(\frac{\lambda i\,(a+b+cu_0)}{\sqrt{T}}\right)-1\right) \\
\longrightarrow \; & 0 \text{ as } T\longrightarrow\infty. \qquad (3.5.27)
\end{aligned}
$$

$$
\begin{aligned}
& \sum_{t=0}^{T-1}\exp(\alpha+\gamma u_{t+1}+\beta t/T)\exp\left(\frac{\lambda i\,(a+cu_{t+1}+bt/T)}{\sqrt{T}}\right)\left(\exp\left(\frac{\beta+\lambda ib/\sqrt{T}}{T}\right)-1\right) \\
- \; & \sum_{t=0}^{T-1}\exp(\alpha+\gamma u_{t+1}+\beta t/T)\left(\exp(\beta/T)-1\right) \\
= \; & \sum_{t=0}^{T-1}\left[\exp\left(\frac{\lambda i\,(a+cu_{t+1}+bt/T)}{\sqrt{T}}\right)\left(\exp\left(\frac{\beta+\lambda ib/\sqrt{T}}{T}\right)-1\right)-\left(\exp(\beta/T)-1\right)\right] \\
\times \; & \exp(\alpha+\gamma u_{t+1}+\beta t/T). \qquad (3.5.28)
\end{aligned}
$$

Replacing each exponential term by a Taylor series expansion,

$$
\begin{aligned}
& \left[\exp\left(\frac{\lambda i\,(a+cu_{t+1}+bt/T)}{\sqrt{T}}\right)\left(\exp\left(\frac{\beta+\lambda ib/\sqrt{T}}{T}\right)-1\right)-\left(\exp(\beta/T)-1\right)\right] \\
= \; & \left\{1+\frac{\lambda i\,(a+cu_{t+1}+bt/T)}{\sqrt{T}}-\frac{\lambda^2\,(a+cu_{t+1}+bt/T)^2}{T}+O\left(T^{-3/2}\right)\right\} \\
\times \; & \left[\left\{\frac{\beta}{T}+\frac{\lambda ib}{T\sqrt{T}}+O\left(T^{-2}\right)\right\}-\left\{\frac{\beta}{T}+O\left(T^{-2}\right)\right\}\right] \\
= \; & O\left(T^{-3/2}\right). \qquad (3.5.29)
\end{aligned}
$$

Therefore

$$\begin{aligned}
& \sum_{t=0}^{T-1}\left[\exp\left(\frac{\lambda i\,(a+cu_{t+1}+bt/T)}{\sqrt{T}}\right)\left(\exp\left(\frac{\beta+\lambda ib/\sqrt{T}}{T}\right)-1\right)-(\exp(\beta/T)-1)\right] \\
\times\quad & \exp(\alpha+\gamma u_{t+1}+\beta t/T) \\
=\quad & \sum_{t=0}^{T-1} O\left(T^{-3/2}\right)\exp(\alpha+\gamma u_{t+1}+\beta t/T) \\
\leq\quad & \exp(|\alpha|+|\gamma|+|\beta|)\sum_{t=0}^{T-1} O\left(T^{-3/2}\right) \\
=\quad & O\left(\frac{1}{\sqrt{T}}\right) \\
\longrightarrow\quad & 0 \quad \text{as } T\longrightarrow\infty. \qquad (3.5.30)
\end{aligned}$$

Thus

$$\begin{aligned}
& \exp\left[\sum_{t=1}^{T}\left(\exp\left(\frac{i\lambda}{\sqrt{T}}(a+bt/T+cu_t)\right)-1\right)\exp(\alpha+\beta t/T+\gamma u_t)\cos(2(\omega t+\phi))\right] \\
\longrightarrow\quad & 1 \ \text{ as } T\longrightarrow\infty, \qquad (3.5.31)
\end{aligned}$$

so

$$E\left[\exp(i\lambda P_T(\boldsymbol{\theta}))\right]\longrightarrow \lim_{T\to\infty}\left\{\exp\left(\frac{-\lambda^2}{2}\sum_{t=1}^{T}\left[\frac{(a+bt/T+cu_t)^2}{T}\right]\exp(\alpha+\beta t/T+\gamma u_t)\right)\right\}$$

which is the unique characteristic function of a mean zero Gaussian random variable.

Chapter 4

Models with two latent processes and their properties

4.1 Introduction

The basic model discussed so far, with its log-linear mean and trigonometric latent process, accounts well for a single unknown periodicity in a poisson count time series while also providing a correlation structure, but it is limited in both conditional expectation and unconditional covariance. Models with more than one unknown period, additional local dependence or a correlation function which decreases in the long term would therefore be useful extensions. In this chapter, such models will be constructed by addition of a second latent process to the basic model and their unconditional means and covariance functions computed. The asymptotic properties of the mean parameter estimator and the RFT-based frequency estimator will then be studied.

4.2 Adding a second periodic latent process to the model

There are many situations which could result in data with more than one unknown period. Astronomical observations such as sunspots or Jupiter's storms could be dependent on the interacting cycles of an unknown subset of satellites. Another example is the case counts of a highly infectious and rapidly evolving class of diseases like influenza. These could have both a medium-term "predator-prey" cycle of one, two or three years and a long-term cycle indicative of the time taken for a particularly malignant variety of pathogen to evolve. The length of both cycles would be of interest, the first when deciding how to optimally utilise drugs or vaccines in the short term, the second to predict when the next epidemic is likely to occur. Clearly a model which can capture the behaviour of data with more than one period is needed Therefore we shall examine models with two unknown periods incorporated via a two-period trigonometric latent process. In later sections we shall try to extend the results on asymptotic behaviour of $\hat{\boldsymbol{\theta}}_{GLM}$ and $\hat{\omega}$ to cover the two-period models.
An extra period can be added to the basic model either multiplicatively, with

$$\{y_t \mid \phi, \psi\} \sim Po\left(\exp\left(\mathbf{x}_t^T \boldsymbol{\beta}\right) \cos^2\left(\omega t + \phi\right) \cos^2\left(\delta t + \psi\right)\right)$$

or additively, with

$$\{y_t \mid \phi, \psi\} \sim Po\left(\exp\left(\mathbf{x}_t^T \boldsymbol{\beta}\right) \left(\cos^2\left(\omega t + \phi\right) + A^2 \cos^2\left(\delta t + \psi\right)\right)\right).$$

In both cases ϕ and ψ are independent Uniform$(0, 2\pi)$ random variables. Examining the multiplicative model first,

$$E\left(y_t\right) = \exp\left(\mathbf{x}_t^T \boldsymbol{\beta}\right) E\left(\cos^2\left(\omega t + \phi\right)\right) E\left(\cos^2\left(\delta t + \psi\right)\right) = \frac{\exp\left(\mathbf{x}_t^T \boldsymbol{\beta}\right)}{4} = \mu_t,$$

$$\begin{aligned} V\left(y_t\right) &= \exp\left(2\mathbf{x}_t^T \boldsymbol{\beta}\right) V\left(\cos^2\left(\omega t + \phi\right) \cos^2\left(\delta t + \psi\right)\right) + E\left(y_t\right) \\ &= \mu_t^2 \left[E\left(4\cos^4\left(\omega t + \phi\right)\right) E\left(4\cos^4\left(\delta t + \psi\right)\right) - 1\right] + \mu_t \\ &= \mu_t^2 \left[E\left(\frac{1}{2}\left(\cos 4\left(\omega t + \phi\right) + 4\cos 2\left(\omega t + \phi\right) + 3\right)\right)^2 - 1\right] + \mu_t \\ &= \mu_t + \mu_t^2 \left[\left(\frac{3}{2}\right)^2 - 1\right] = \mu_t + \frac{5\mu_t^2}{4}, \end{aligned} \quad (4.2.1)$$

$$\begin{aligned}
Cov\left(y_t, y_{t+s}\right) &= \mu_t\mu_{t+s}Cov\big(4\cos^2\left(\omega t+\phi\right)\cos^2\left(\delta t+\psi\right), 4\cos^2\left(\omega t+\omega s+\phi\right) \\
&\quad \times\cos^2\left(\delta t+\delta s+\psi\right)\big) \\
&= \Big[E\big(4\cos^2\left(\omega t+\phi\right)\cos^2\left(\omega t+\omega s+\phi\right)\big)\times \\
&\quad E\big(4\cos^2\left(\delta t+\psi\right)\cos^2\left(\delta t+\delta s+\psi\right)\big)-1\Big]\mu_t\mu_{t+s} \\
&= \mu_t\mu_{t+s}\left[\left(\frac{1}{2}\cos\left(2\omega s\right)+1\right)\left(\frac{1}{2}\cos\left(2\delta s\right)+1\right)-1\right] \\
&= \frac{\mu_t\mu_{t+s}}{4}\left(2\cos\left(2\omega s\right)+2\cos\left(2\delta s\right)+\cos\left(2\omega s\right)\cos\left(2\delta s\right)\right) \\
&= \frac{\mu_t\mu_{t+s}}{8}\left(4\cos\left(2\omega s\right)+4\cos\left(2\delta s\right)+\cos\left(2\left(\omega-\delta\right)s\right)+\cos\left(2\left(\omega+\delta\right)s\right)\right).
\end{aligned} \tag{4.2.2}$$

Thus the covariance function of $\varepsilon_t = 4\cos^2\left(\omega t+\phi\right)\cos^2\left(\delta t+\psi\right)$ is a harmonic function with frequencies equal to twice the difference and sum of the initial frequencies, as well as 2ω and 2δ themselves. Due to these extra frequencies, the multiplicative model might have potential as a reduced-parameter method of fitting data with an apparent four-period covariance function. In the case where two of the frequencies are close to the sum and difference of the others and the amplitudes of the four terms have a ratio close to 4:4:1:1, at least two fewer parameters would be needed to fit the multiplicative model opposed to an additive one with four trigonometric terms. However, the multiplicative model is limited in terms of amplitude - this is a quantity which is impossible to uniquely separate from the intercept term. The additive model does not have this limitation, although only the ratio of amplitudes is separable. For

$$\begin{aligned}
\{y_t \mid \phi, \psi\} &= \exp\left(\mathbf{x}_t^T\boldsymbol{\beta}\right)\left(\cos^2\left(\omega t+\phi\right)+A^2\cos^2\left(\delta t+\psi\right)\right) \\
E\left(y_t\right) &= \exp\left(\mathbf{x}_t^T\boldsymbol{\beta}\right)\left[E\left(\cos^2\left(\omega t+\phi\right)\right)+E\left(A^2\cos^2\left(\delta t+\psi\right)\right)\right] \\
&= \frac{1+A^2}{2}\exp\left(\mathbf{x}_t^T\boldsymbol{\beta}\right)=\mu_t,
\end{aligned} \tag{4.2.3}$$

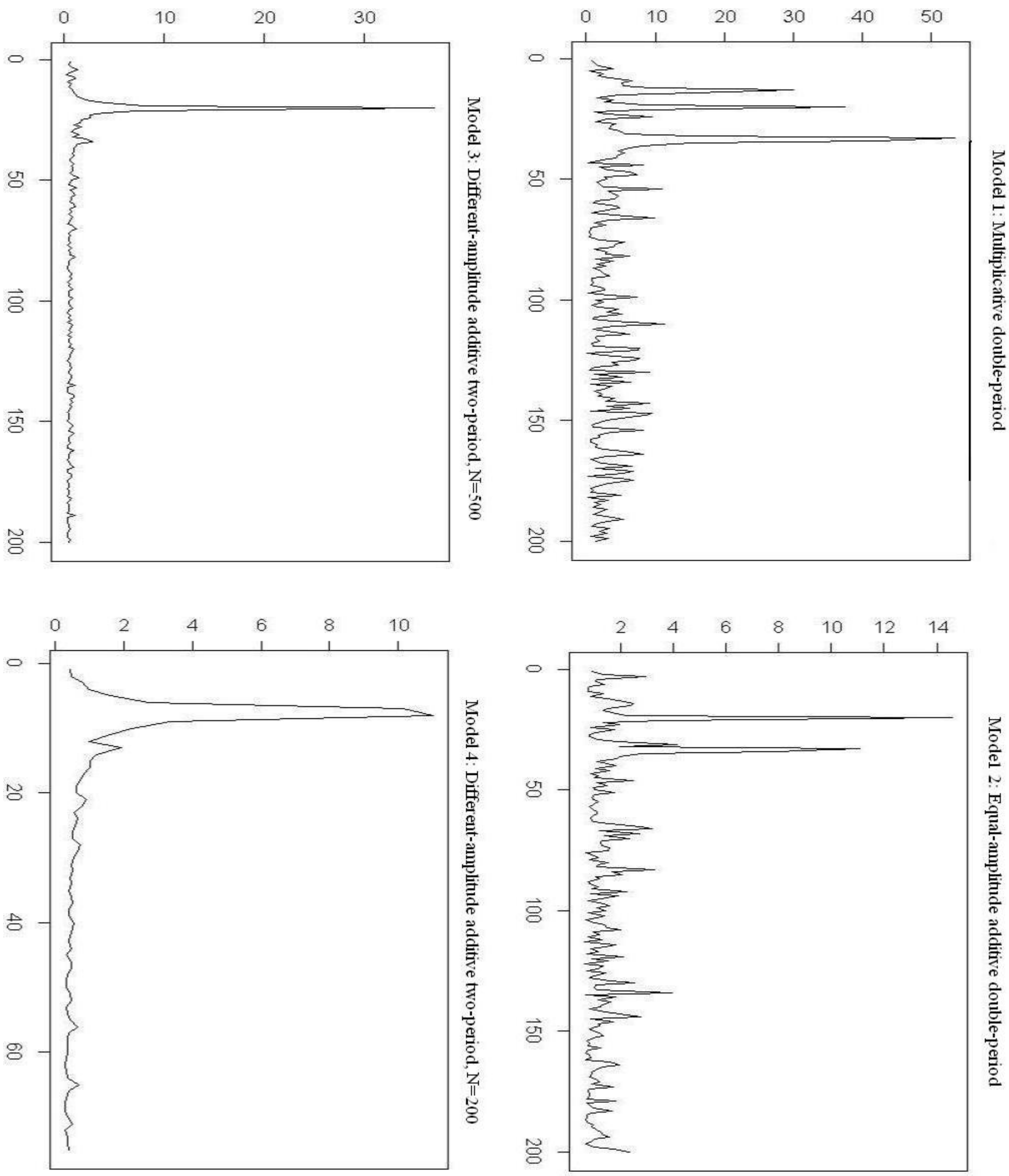
$$\begin{aligned}
V(y_t) &= \mu_t + \exp\left(2\mathbf{x}_t^T\boldsymbol{\beta}\right)\left[E\left(\cos^2(\omega t+\phi)+A^2\cos^2(\delta t+\psi)\right)^2 - \left(\frac{1+A^2}{2}\right)^2\right] \\
&= \mu_t + \Big[E\left(\cos^4(\omega t+\phi)\right) + A^4 E\left(\cos^4(\delta t+\psi)\right) - \frac{1}{4}\left(1+2A^2+A^4\right) \\
&\quad + 2A^2 E\left(\cos^4(\omega t+\phi)\right)E\left(\cos^4(\delta t+\psi)\right)\Big]\exp\left(2\mathbf{x}_t^T\boldsymbol{\beta}\right) \\
&= \mu_t + \exp\left(2\mathbf{x}_t^T\boldsymbol{\beta}\right)\left[\frac{3}{8}+\frac{3A^4}{8}+\frac{2A^2}{4}-\frac{1}{4}\left(1+2A^2+A^4\right)\right] \\
&= \mu_t + \exp\left(2\mathbf{x}_t^T\boldsymbol{\beta}\right)\left(\frac{1+A^4}{8}\right) \\
&= \mu_t + \frac{\mu_t^2}{2}\left[\frac{1}{(1+A^2)^2}+\frac{A^4}{(1+A^2)^2}\right], \qquad (4.2.4)
\end{aligned}$$

$$\begin{aligned}
Cov(y_t, y_{t+s}) &= Cov\Big(\cos^2(\omega t+\phi)+A^2\cos^2(\delta t+\psi), \cos^2(\omega(t+s)+\phi) \\
&\quad + A^2\cos^2(\delta(t+s)+\psi)\Big)e^{\mathbf{x}_t^T\boldsymbol{\beta}}e^{\mathbf{x}_{t+s}^T\boldsymbol{\beta}} \\
&= \Big[A^2E\big(\cos^2(\omega t+\phi)\big)E\big(\cos^2(\delta t+\delta s+\psi)\big) - \left(\frac{1+A^2}{2}\right)^2 + \\
&\quad E\Big(\cos^2(\omega t+\phi)\cos^2(\omega t+\omega s+\phi)+A^4\cos^2(\delta t+\psi)\cos^2(\delta t+\delta s+\psi)\Big) \\
&\quad +A^2E\big(\cos^2(\omega t+\omega s+\phi)\big)E\big(\cos^2(\delta t+\psi)\big)\Big]e^{\mathbf{x}_t^T\boldsymbol{\beta}}e^{\mathbf{x}_{t+s}^T\boldsymbol{\beta}} \\
&= \left[\frac{1}{8}\left(2+\cos(2\omega s)\right)+\frac{A^4}{8}\left(2+\cos(2\delta s)\right)+\frac{A^2}{2}-\frac{1}{4}-\frac{A^2}{2}-\frac{A^4}{4}\right] \\
&\times\ e^{\mathbf{x}_t^T\boldsymbol{\beta}}e^{\mathbf{x}_{t+s}^T\boldsymbol{\beta}} \\
&= \frac{\cos(2\omega s)+A^4\cos(2\delta s)}{8}e^{\mathbf{x}_t^T\boldsymbol{\beta}}e^{\mathbf{x}_{t+s}^T\boldsymbol{\beta}} \\
&= \frac{\mu_t\mu_{t+s}}{2\left(1+A^2\right)^2}\left(\cos(2\omega s)+A^4\cos(2\delta s)\right). \qquad (4.2.5)
\end{aligned}$$

Due to the possible difference in amplitudes, the variance of the latent process in the additive model is slightly more complex than that in the multiplicative model. The correlation function however is simpler, using only the initial frequencies.

Several sets of simulated two-period data have been generated and analysed to find their periodicities using the DFT of the latent covariance estimator. The peaks in all the graphs of the DFT moduli are close to the expected values from the theoretical covariance of their corresponding latent process, whether ε_t is two trigonometric terms multiplied together as in Model 1, two equal-amplitude terms added together as in Model 2 or two terms with different amplitudes added together as in Models 3 and 4.

Figure 4.1: DFT plots for covariance estimates from four simulations of size 500, 500, 500, and 200 respectively. All have unconditional mean $\exp\left(1 - log\left(2\right) - 0.004t\right)$. The latent processes of Model 1, Model 2, Model 3 and Model 4 are $4\cos^2\left(2\pi t/24\right)\cos^2\left(2\pi t/40\right)$, $2\left(\cos^2\left(2\pi t/24\right) + \cos^2\left(2\pi t/40\right)\right)$,$2\left(\cos^2\left(2\pi t/24\right) + A^2\cos^2\left(2\pi t/40\right)\right)$ and $2\left(\cos^2\left(2\pi t/24\right) + A^2\cos^2\left(2\pi t/40\right)\right)$ respectively.



The only drawback, at least empirically, is that in the last case, as could be predicted from the theoretical covariance function, the term with the higher amplitude has a frequency displayed on the graph as a much higher peak than that corresponding to the frequency of the low-amplitude term. The first peak in the DFT moduli of Models 3 and 4 is so tall that it would be difficult to ascertain whether the second peak is more than just the product of random noise if it was not known to correspond to an important frequency. One possible way of resolving this is to difference the covariance estimates by a lag equal to the inverse of the estimate of the high-amplitude frequency. It is simple to show that this new sample has expectation proportional to $\cos(2\ell t)$ where ℓ is the low-amplitude frequency. ℓ can then be estimated using the DFT as for the frequency of a single-period data set. Hence differencing the covariance estimates is, in theory, a plausible strategy for removing the high-amplitude frequency. In practice, it depends on the sample size, the length of the low-amplitude period relative to the number of covariances estimated and the relative amplitude sizes. The larger the sample size is and the closer the amplitude ratio is to unity, the more differentiable from random peaks the peak at the low-amplitude frequency appears. The former effect is due to the reduced influence of white noise, which reduces the relative size of truly random peaks, as can be seen by comparing the DFT plots for models 3 and 4. The latter effect is because the low-amplitude peak is less overshadowed by the high-amplitude peak. This can be inferred from the differences between the DFT plots of Models 2 and 3, which differ only in that one has terms with equal amplitude and the other an amplitude ratio of 4:1. The other factor mentioned, the length of the low-amplitude period relative to the number M of covariances estimated, is potentially important when period q is neither close or equal to a divisor of M. In this case, M is between kq and $(k+1)\,q$ for some k, leading to q being estimated as either $\frac{M}{k}$ or $\frac{M}{k+1}$ when in fact it is somewhere between them. To illustrate the degree of inaccuracy this could lead to, consider the case where $q = 20$ and $M = 90$. The true frequency ω is thus $\frac{2\pi}{20} = \frac{\pi}{10} = 2\pi\frac{4.5}{90}$ so would be likely to cause a peak on the DFT at either 4 or 5. As $\hat{q} = 2\pi M/\hat{\omega}$, the period would either be estimated as 18 or 22.5, either of which are poor estimates of 20. For large values of M, the

$M : q$ ratio problem could be avoided by recalculating the DFT for subsets of the covariance estimators - for example over the first 60 and 80 covariance estimators for the previous example. It is possible that whenever a period is long enough relative to the sample size and at the same time small enough in amplitude to cause such an inaccuracy, it has little enough effect on the behaviour of the observations for an inaccurate estimate to make much difference to the model.

4.3 An ARMA latent process for short-term, non-periodic dependence

Under the basic single-period model, the covariance structure is $Cov\left(y_t, y_{t+s}\right) = \frac{1}{2}\mu_t\mu_{t+s}\cos\left(2\omega s\right)$. Clearly correlation is highest when the lag is close to zero or a multiple of the period and lowest when the lag is near to a half-period. This structure accounts well both for high short-term correlation and for correlation between two maximum or minimum values being highly positive while that between a maximum and a minimum is highly negative. However, when the trend is positive, the correlation between two observations lagged by a multiple of the period actually increases as the multiple does. It is highly unlikely that the sunspot numbers or disease case counts observed at two consecutive time points would be less correlated than those observed several years apart, so a source of short-term, monotonically decreasing variation is needed to create a model with a more realistic covariance function.

As when extending the single-period model to a two-period one, we shall add short-term correlation to the basic model by adding a second latent process. The double latent process $\{\varepsilon_t\} = 2\exp(z_t)\cos^2\left(\omega t + \phi\right)$ is suggested, where z_t is a weakly stationary zero-mean ARMA(p,q) process. This gives the model

$$\{y_t|\phi, z_t\} \quad \sim \quad Po\left(2\exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right)\cos^2(\omega t + \phi)\right) \tag{4.3.1}$$

with

$$E\left(y_t|z_t\right) \quad = \quad \exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right). \tag{4.3.2}$$

Thus

$$\begin{aligned} E\left(y_t\right) &= \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right) E\left(\exp\left(z_t\right)\right) \\ &= \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right) \exp\left(\sigma_z^2/2\right) \end{aligned} \quad (4.3.3)$$

where $\sigma_z^2 = V\left(z_t\right)$.

$$\begin{aligned} V\left(y_t\right) &= V\left(E\left(y_t|z_t\right)\right) + E\left(V\left(y_t|z_t\right)\right) \\ &= V\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right)\right) + E\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right) + \frac{1}{2}\exp\left(2\mathbf{x}_t^T\boldsymbol{\theta} + 2z_t\right)\right), \end{aligned} \quad (4.3.4)$$

$$\begin{aligned} & E\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right) + \frac{1}{2}\exp\left(2\mathbf{x}_t^T\boldsymbol{\theta} + 2z_t\right)\right) \\ &= \exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + \sigma_z^2/2\right) + \frac{1}{2}\exp\left(2\mathbf{x}_t^T\boldsymbol{\theta} + 2z_t\right), \end{aligned} \quad (4.3.5)$$

$$\begin{aligned} V\left(\exp\left(\mathbf{x}_t^T\boldsymbol{\theta} + z_t\right)\right) &= \exp\left(2\mathbf{x}_t^T\boldsymbol{\theta}\right)\left[E\left(\exp\left(2z_t\right) - \exp\left(\sigma_z^2\right)\right)\right] \\ &= \exp\left(2\mathbf{x}_t^T\boldsymbol{\theta}\right)\left[E\left(\exp\left(2\sigma_z^2\right) - \exp\left(\sigma_z^2\right)\right)\right].. \end{aligned} \quad (4.3.6)$$

Thus $V\left(y_t\right) = \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2/2\right) + \exp\left(2\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2\right)\left[\frac{3\exp\left(\sigma_z^2/2\right)}{2} - 1\right]$.
Also

$$\begin{aligned} Cov\left(y_t, y_{t+s}\right) &= E\left(y_t y_{t+s}\right) - E\left(y_t\right)E\left(y_{t+s}\right) \\ &= E\big(E\left(y_t y_{t+s}|z_t, z_{t+s}\right)\big) - \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_{t+s}^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2\right) \\ &= E\big[Cov\left(y_t, y_{t+s}|z_t, z_{t+s}\right) + E\left(y_t|z_t\right)E\left(y_{t+s}|z_{t+s}\right)\big] \\ &- \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_{t+s}^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2\right) \\ &= E\left[\frac{\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_{t+s}^T\boldsymbol{\theta}\right)\exp\left(z_t\right)\exp\left(z_{t+s}\right)\left(\cos(2\omega s) + 2\right)}{2}\right] \\ &- \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_{t+s}^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2\right) \\ &= \exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)\exp\left(\mathbf{x}_{t+s}^T\boldsymbol{\theta}\right)\exp\left(\sigma_z^2\right)\left[\frac{\cos(2\omega s) + 2}{2}\exp\left(\gamma_z(s)\right) - 1\right]. \end{aligned} \quad (4.3.7)$$

4.4 The strong mixing of a multivariate AR process

In this section we shall prove that stationary autoregressive (AR) processes are strongly mixing in the sense of Peligrad et al. (1997), Definition 2.3, commonly known as α-mixing. Although strong mixing of certain ARMA processes has been proven in Berkes et al. (1981), we shall establish α-mixing of AR processes with an exponentially decreasing rate. Let $\{X_t\}$ denote a stationary AR process of order p. Then α-mixing implies that a central limit theorem (clt) exists for $S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - E[X_t])$, using Peligrad et al. (1997), Theorem 2.2. Although this result is not of direct use, as a clt for $\{S_n\}$ can be proved straightforwardly without recourse to α-mixing, we will also show that the α-mixing of $\{X_t\}$ implies the α-mixing of $\{Y_t\}$. $\{Y_t|X_t, \phi\} \sim D\left(F_t(X_t, \phi)\right)$. D denotes an arbitrary but specified probability distribution and $F_t(\cdot)$ denotes a deterministic function (possibly) dependent upon t. In our periodic model for Poisson counts, $F_t(X_t)$ is taken to be $\exp\left(\mathbf{x}_t^T \boldsymbol{\theta}\right)(1 + cos\left(2\omega t + 2\phi\right))$, where $\mathbf{x}$ and $\boldsymbol{\theta}$ denote vectors, of covariates associated with time point t and of the corresponding parameters respectively. Given $\{X_t\}$ and ϕ, the $\{Y_t\}$'s are conditionally independent. The α-mixing of $\{Y_t\}$ then can be used to establish a central limit theorem for

$$\tilde{S}_n = \sum_{t=1}^{n} a_{nt}(Y_t - E[Y_t]) \qquad (4.4.1)$$

conditional on ϕ, under weak conditions upon $\{a_{nt}\}$ and $\{Y_t\}$ provided that there exists $0 < \sigma^2 < \infty$ such that $var(\sum_{t=1}^{n} a_{nt} Y_t) \to \sigma^2$ as $n \to \infty$. Such results will be very useful when proving asymptotic results for estimators of parameters in the double-latent-process model.

4.4.1 Rewriting $MAR(m, p)$ processes as $MAR(mp, 1)$ processes

An m-dimensional multivariate autoregressive process of order p ($MAR(m, p)$), $\{\mathbf{X}_t\}$, satisfies

$$\mathbf{X}_t = \sum_{i=1}^{p} \Phi_i \mathbf{X}_{t-i} + \boldsymbol{\epsilon}_t, \qquad (t \in \mathbb{Z}) \qquad (4.4.2)$$

where Φ_i is an $m \times m$ matrix and $\{\boldsymbol{\epsilon}_t\}$ are independent and identically distributed according to $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \Gamma)$, an m-dimensional multivariate normal with mean $\mathbf{0}$ and covariance matrix Γ. The $MAR(m, p)$ process can be embedded in an $MAR(mp, 1)$

process as follows. Let $\mathbf{Z}_t = (\mathbf{X}_t^T, \mathbf{X}_{t-1}^T, \ldots, \mathbf{X}_{t-p+1}^T)^T$, $\tilde{\boldsymbol{\epsilon}}_t$ be a vector of length mp with $\tilde{\boldsymbol{\epsilon}}_t = (\boldsymbol{\epsilon}_t^T, 0, 0, \ldots, 0)^T$ and

$$A = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots I & 0 & \end{pmatrix} \qquad (4.4.3)$$

with I denoting an m-dimensional identity matrix. Then $\{\mathbf{X}_t\}$ is embedded in $\{\mathbf{Z}_t\}$, where

$$\mathbf{Y}_t = A\mathbf{Y}_{t-1} + \tilde{\boldsymbol{\epsilon}}_t \qquad (t \in \mathbb{Z}).. \qquad (4.4.4)$$

Let λ denote the Perron-Frobenius (largest-magnitude) eigenvalue of A, then $\{\mathbf{Z}_t\}$ and hence $\{\mathbf{X}_t\}$ is stationary if $\lambda < 1$. Note that the eigenvalues of A can be complex. Moreover, we assume that $|A| \neq 0$, where $|A|$ denotes the determinant of A. Throughout we restrict attention to stationary processes. Furthermore, since any $MAR(m,p)$ process can be embedded in an $MAR(mp,1)$ process, all results presented are for $MAR(m,1)$ processes.

4.4.2 Mixing for $MAR(m,1)$ processes

Let $\{\mathbf{Z}_t\}$ satisfy

$$\mathbf{Z}_t = A\mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t \qquad (t \in \mathbb{Z}), \qquad (4.4.5)$$

where $\boldsymbol{\epsilon}_t \sim MVN(\mathbf{0}, \Gamma)$ and $\lambda < 1$. Let $\mathcal{A}$ and $\mathcal{B}$ be two σ-algebras of events and define

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(AB) - P(A)P(B)|.. \qquad (4.4.6)$$

Let $\mathcal{F}_a^b = \sigma(\mathbf{Z}_t, a \leq t \leq b)$. Then $\{\mathbf{Z}_t\}$ is said to be a strongly mixing sequence if $\alpha_k \to 0$ as $k \longrightarrow \infty$, where

$$\alpha_k = \sup_n \alpha(\mathcal{F}_{-\infty}^n, \mathcal{F}_{n+k}^\infty).. \qquad (4.4.7)$$

Equation (4.4.7) can be simplified by making two simple observations concerning $\{\mathbf{Z}_t\}$. Firstly, $\{\mathbf{Z}_t\}$ is stationary, so

$$\alpha_k = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty).. \tag{4.4.8}$$

Secondly, for any $l \in \mathbb{Z}$, $\mathbf{Z}_{l+1}, \mathbf{Z}_{l+2}, \ldots$, are independent of $\mathbf{Z}_{l-1}, \mathbf{Z}_{l-2}, \ldots$ given $\mathbf{Z}_l$. Thus

$$\begin{aligned} \alpha_k &= \alpha(\mathcal{F}_0^0, \mathcal{F}_k^k) \\ &= \sup_{A,B \subset \mathbb{R}^m} |P(\mathbf{Z}_0 \in A, \mathbf{Z}_k \in B) - P(\mathbf{Z}_0 \in A)P(\mathbf{Z}_k \in B)| \\ &= \sup_{A,B \subset \mathbb{R}^m} P(\mathbf{Z}_0 \in A)|P(\mathbf{Z}_k \in B|\mathbf{Z}_0 \in A) - P(\mathbf{Z}_k \in B)|.. \end{aligned} \tag{4.4.9}$$

We shall exploit (4.4.9) to prove the α-mixing of $\{\mathbf{Z}_t\}$.

Let $\Sigma = \sum_{j=0}^{\infty} A^j \Gamma (A^T)^j$ and let $\Lambda_k = \sum_{j=0}^{k-1} A^j \Gamma (A^T)^j = \Sigma - A^k \Sigma (A^T)^k$. Inverting the equation $(1 - A\mathcal{B})\mathbf{Z}_t = \boldsymbol{\epsilon}_t$, where $\mathcal{B}$ is the difference operator, gives $\mathbf{Z}_t = \sum_{j=0}^{\infty} A^j \boldsymbol{\epsilon}_t$. Then observing that $Var(\mathbf{Z}_t) = \sum_{j=0}^{\infty} Var(A^j \boldsymbol{\epsilon}_t)$ makes it clear that

$$\mathbf{Z}_k \sim MVN(\mathbf{0}, \Sigma) \tag{4.4.10}$$

and

$$\mathbf{Z}_k | \mathbf{Z}_0 = \mathbf{z} \sim MVN(A^k \mathbf{z}, \Lambda_k).. \tag{4.4.11}$$

Let $\mathbf{X}$ and $\mathbf{Y}$ be m-dimensional random variables with probability density functions (pdfs) $g(\mathbf{x})$ and $h(\mathbf{x})$, respectively. The total variation distance between $\mathbf{X}$ and $\mathbf{Y}$, denoted $d_{TV}(\mathbf{X}, \mathbf{Y})$ is given by

$$\begin{aligned} d_{TV}(\mathbf{X}, \mathbf{Y}) &= \sup_{C \subset \mathbb{R}^m} |P(\mathbf{X} \in C) - P(\mathbf{Y} \in C)| \\ &= \frac{1}{2} \int |g(\mathbf{x}) - h(\mathbf{x})| \, d\mathbf{x}. \end{aligned} \tag{4.4.12}$$

Combining (4.4.9) with the above definition, we have that

$$\alpha_k \leq \int f(\mathbf{z}) d_{TV}(\{\mathbf{Z}_k | \mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k) \, d\mathbf{z}, \tag{4.4.13}$$

where $f(\mathbf{z})$ is the pdf of $\mathbf{Z}_0 \sim MVN(\mathbf{0}, \Sigma)$.

We continue by studying $d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k)$ before returning to (4.4.13). Let $\mathbf{Y}(k, \mathbf{z}) \sim MVN(A^k\mathbf{z}, \Sigma)$. By the triangle inequality,

$$d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k) \leq d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Y}(k, \mathbf{z})) + d_{TV}(\mathbf{Y}(k, \mathbf{z}), \mathbf{Z}_k). \tag{4.4.14}$$

We proceed by obtaining bounds for the two terms on the right hand side of (4.4.14).

Let $\mathbf{U} \sim MVN(\mathbf{0}, \Sigma)$ and $\mathbf{U}_k \sim MVN(\mathbf{0}, \Lambda_k)$. A simple change of variable argument shows that

$$d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Y}(k, \mathbf{z})) = d_{TV}(\mathbf{U}, \mathbf{U}_k).. \tag{4.4.15}$$

Note that

$$\begin{aligned} 2d_{TV}(\mathbf{U}, \mathbf{U}_k) &= (2\pi)^{-m/2} \int \left| |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) - |\Lambda_k|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right)\right| d\mathbf{x} \\ &\leq (2\pi)^{-m/2} \int \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) \left||\Sigma|^{-1/2} - |\Lambda_k|^{-1/2}\right| d\mathbf{x} \\ &\quad + \frac{21}{(2\pi)^{m/2}|\Lambda_k|^{1/2}} \int \left|\exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) - \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right)\right| \mathrm{d}\mathbf{x}. \end{aligned} \tag{4.4.16}$$

The first term on the right hand side of (4.4.16) is equal to

$$\begin{aligned} \frac{|\Sigma|^{1/2}||\Lambda_k|^{1/2} - |\Sigma|^{1/2}|}{|\Sigma|^{1/2}|\Lambda_k|^{1/2}} &= \frac{|\Sigma|^{-1/2}||\Lambda_k|^{1/2} - |\Sigma|^{1/2}|}{|\Sigma|^{-1/2}|\Lambda_k|^{1/2}} \\ &= \frac{||\Lambda_k\Sigma^{-1}|^{1/2} - 1|}{|\Lambda_k\Sigma^{-1}|^{1/2}} \\ &= |\Lambda_k\Sigma^{-1}|^{-1/2} \left|1 - |\Lambda_k\Sigma^{-1}|^{1/2}\right|, \end{aligned} \tag{4.4.17}$$

where $\Lambda_k\Sigma^{-1} = I - A^k\Sigma(A^T)^k\Sigma^{-1} = I - R_k$, say.

To obtain bounds for (4.4.17), we first need to establish the following two lemmas.

Lemma 1: Let F be a $m \times m$ matrix such that $\hat{f} = \max_{1\leq i,j\leq m} |f_{ij}| \leq \epsilon$ for some $\epsilon \geq 0$, where f_{ij} is the $(i, j)^{th}$ of F. Then

$$||I + F| - 1| \leq m\epsilon + O(\epsilon^2). \tag{4.4.18}$$

Proof: For any two $n \times n$ matrices A and B and scalar k, it is well-established that

$$|A + kB| - |A| = |A|tr\left(A^{-1}B\right)k + O(k^2).$$

A special case of this is where $A = I_n$, giving

$$|I_n + kB| = 1 + tr\,(B)\,k + O(k^2).$$

Thus after rewriting $F = \hat{f}X$, where all elements of X are less than or equal to 1 in absolute value, we have that

$$|I_n + \hat{f}X| - 1 \quad = \quad tr\,(X)\,\hat{f} + O(\hat{f}^2). \tag{4.4.19}$$

Thus $\Big||I_n + \hat{f}X| - 1\Big| \leq m\epsilon + O(\epsilon^2)$. □

Hence there exists $\epsilon_0 > 0$, such that for all $\epsilon \leq \epsilon_0$,

$$||I + F| - 1| \leq (m+1)\epsilon.. \tag{4.4.20}$$

Since $|A|$ is non-singular, there exists a diagonal matrix D consisting of the eigenvalues of A and invertible matrix L such that

$$A = L^{-1}DL, \tag{4.4.21}$$

where D and L are possibly complex matrices, depending upon whether or not the eigenvalues of A are real or not. Then the maximal absolute value of an element of A^k is less than or equal to $m^2\hat{q}\hat{l}\lambda^k$, where $\hat{q}$ is the maximum absolute value of $Q = L^{-1}$.
Lemma 2: For $n \geq 1$, let $B_1, B_2, \ldots, B_n$ be $m \times m$ matrices and $F = \prod_{j=1}^{n} B_j$. Then

$$\hat{f} \leq m^{n-1}\prod_{j=1}^{n}\hat{b}_j.. \tag{4.4.22}$$

Proof. The ij^{th} value of a product of an $l \times m$ matrix A and an $m \times p$ matrix B is always smaller than ma_ib_j, where a_i and b_j are the largest absolute values in the i^{th}row of A and j^{th} column of B respectively. Using the case where $l = m = n$, together with induction upon n the proof is very straightforward and is hence omitted. □

Hence, there exists $\beta < \infty$, such that for all $k \geq 0$, the maximal element of $R_k = A^k\Sigma(A^T)^k\Sigma^{-1}$ is less than $\beta\lambda^{2k}$.

Thus there exists $k_1 \in \mathbb{N}$, such that for all $k \geq k_1$,

$$\big||\Lambda_k \Sigma^{-1}| - 1\big| \leq (m+1)\beta\lambda^{2k}.. \tag{4.4.23}$$

Since for all $x \geq 0$, $|\sqrt{x} - 1| \leq |x - 1|$,

$$\big||\Lambda_k \Sigma^{-1}|^{1/2} - 1\big| \leq (m+1)\beta\lambda^{2k}.. \tag{4.4.24}$$

Therefore there exists $k_2 \in \mathbb{N}$, such that for all $k \geq k_2$,

$$|\Lambda_k \Sigma^{-1}|^{1/2} \geq \frac{1}{2}. \tag{4.4.25}$$

and hence, for all sufficiently large k, the first term on the right hand side of (4.4.16) is less than $2(m+1)\beta\lambda^{2k}$.

Turning to the second term on the right hand side of (4.4.16),

$$\begin{aligned} & \int \left| \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) - \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right) \right| d\mathbf{x} \\ \leq & \int \max\left\{ \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right), \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right) \right\} \\ \times & \left| \exp\left(-\frac{1}{2}\left|\mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right|\right) - 1 \right| d\mathbf{x} \\ \leq & \frac{1}{2}\int \left\{ \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) + \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right) \right\} \left|\mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right| d\mathbf{x}, \end{aligned} \tag{4.4.26}$$

since for any $y > 0$, $|1 - e^{-y}| \leq y$.

Note that

$$\mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{x}^T\Lambda_k^{-1}\mathbf{x} = \mathbf{x}^T\left(\Sigma^{-1} - \Lambda_k^{-1}\right)\mathbf{x} \tag{4.4.27}$$

and $\Lambda_k^{-1} = \Sigma^{-1}(I - R_k)^{-1}$. Thus

$$\begin{aligned} \mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{x}^T\Lambda_k^{-1}\mathbf{x} &= \mathbf{x}^T\Sigma^{-1}\left\{I - (I - R_k)^{-1}\right\}\mathbf{x} \\ &= \mathbf{x}^T\Sigma^{-1}\left(\left\{(I - R_k) - I\right\}(I - R_k)^{-1}\right)\mathbf{x} \\ &= -\mathbf{x}^T\Sigma^{-1}R_k(I - R_k)^{-1}\mathbf{x}. \end{aligned} \tag{4.4.28}$$

For any matrix B, it is clear that

$$|\mathbf{x}^T B\mathbf{x}| \leq \sum_{i=1}^{m} m\hat{b}x_i^2.. \tag{4.4.29}$$

Thus letting $B_k = \Sigma^{-1}R_k(I - R_k)^{-1}$, we have that

$$\left|\mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right| \leq m\hat{b}_k \sum_{i=1}^{m} x_i^2.. \tag{4.4.30}$$

As noted above, the maximum value $\hat{r}_k$ of R_k is bounded above by $\beta\lambda^{2k}$. Therefore the exists $k_3 \in \mathbb{N}$ such that, for all $k \geq k_3$, $\hat{r}_k \leq 1/(2m)$.

Now

$$\begin{aligned} B_k &= \Sigma^{-1}R_k(I - R_k)^{-1} \\ &= \Sigma^{-1}R_k \sum_{j=0}^{\infty} R_k^j. \end{aligned} \tag{4.4.31}$$

By Lemma 2, the maximal absolute value of an element of R_k^j is less than $2^{-(j-1)}$ for all $k \geq k_3$. Hence, the maximal absolute value of an element of $(I - R_k)^{-1}$ is less than $\sum_{j=0}^{\infty} 2^{-(j-1)} = 4$, for all $k \geq k_3$. Therefore there exists $\beta_1 < \infty$ such that for all $k \geq k_3$, $\hat{b}_k \leq \beta_1\lambda^{2k}$.

Hence for all $k \geq k_3$,

$$\begin{aligned} &\int \left|\exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) - \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right)\right| d\mathbf{x} \\ \leq\ & m\beta_1\lambda^{2k}\sum_{i=1}^{m}\int x_i^2\left\{\exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) + \exp\left(-\frac{1}{2}\mathbf{x}^T\Lambda_k^{-1}\mathbf{x}\right)\right\} d\mathbf{x}. \end{aligned} \tag{4.4.32}$$

For $i = 1, 2, \ldots, m$, let σ_i^2 be the $(i,i)^{th}$ element of Σ. By construction it is straightforward to see that the i^{th} component of $\mathbf{U}_k$ has smaller variance than the i^{th} component of $\mathbf{U}$. Hence, the right hand side of (4.4.17) is less than

$$m\beta_1\lambda^{2k}\left\{(2\pi)^{m/2}|\Sigma|^{1/2} + (2\pi)^{m/2}|\Lambda_k|^{1/2}\right\}\sum_{i=1}^{m}\sigma_i^2. \tag{4.4.33}$$

Thus the second term on the right hand side of (4.4.16) is less than

$$\left(m\beta_1\left\{1 + |\Lambda_k\Sigma^{-1}|^{-1/2}\right\}\sum_{i=1}^{m}\sigma_i^2\right)\lambda^{2k}. \tag{4.4.34}$$

Since $|\Lambda_k\Sigma^{-1}|^{-1/2} \to 1$ as $k \to \infty$, we can conclude that there exists $k^* \in \mathbb{N}$ and $\beta^* < \infty$, such that, for all $k \geq k^*$,

$$d_{TV}(\mathbf{U}, \mathbf{U}_k) \leq \beta^*\lambda^{2k}. \tag{4.4.35}$$

Next let $\mathbf{V} \sim MVN(\boldsymbol{\mu}, \Sigma)$, for some m-dimensional mean vector $\boldsymbol{\mu}$. Let $\mathcal{B}_d = \{\mathbf{x}; |\mathbf{x}| \leq d\}$.

Let $S = \Sigma^{-1}$ and for $d \geq 0$, let $\mathbf{B}_d = \{\mathbf{x}; \sum_{i=1}^m x_i^2 \leq d^2\}$, a ball of radius d centered at the origin. Then

$$\begin{aligned} d_{TV}(\mathbf{V}, \mathbf{Z}_k) &= (2\pi)^{-m/2}|\Sigma|^{-1/2} \int \left| \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T S(\mathbf{x}-\boldsymbol{\mu})\right) - \exp\left(-\frac{1}{2}\mathbf{x}^T S\mathbf{x}\right)\right| d\mathbf{x} \\ &\leq \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \int_{\mathbf{x}\in\mathcal{B}_d} \left| \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T S(\mathbf{x}-\boldsymbol{\mu})\right) - \exp\left(-\frac{1}{2}\mathbf{x}^T S\mathbf{x}\right)\right| d\mathbf{x} \\ &\quad + P(\mathbf{V} \notin \mathcal{B}_d) + P(\mathbf{Z}_k \notin \mathcal{B}_d). \end{aligned} \tag{4.4.36}$$

Looking at the second part of the right-hand-side of (4.4.36), observe that

$$\begin{aligned} P(\mathbf{V} \notin \mathcal{B}_d) &= P(|\mathbf{V}| > d) \\ &\leq \sum_{i=1}^m P(|V_i| > d/m) \\ &= \sum_{i=1}^m \left\{\Phi\left(-\frac{d - m\mu_i}{m\sigma_i}\right) + \Phi\left(-\frac{d + m\mu_i}{m\sigma_i}\right)\right\}, \end{aligned} \tag{4.4.37}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of a standard univariate normal random variable. Similarly, $P(\mathbf{Z}_k \notin \mathcal{B}_d) \leq 2\sum_{i=1}^m \Phi(-d/(m\sigma_i))$.

Turning to the first term on the right hand side of (4.4.36), note that

$$(\mathbf{x}-\boldsymbol{\mu})^T S(\mathbf{x}-\boldsymbol{\mu}) = \mathbf{x}^T S\mathbf{x} - 2\boldsymbol{\mu}^T S\mathbf{x} + \boldsymbol{\mu}^T S\boldsymbol{\mu}. \tag{4.4.38}$$

Therefore since for all $y \geq 0$, $|1 - e^{-y}| \leq y$, we have that the first term on the right hand side of (4.4.36) is less than

$$\begin{aligned} & (2\pi)^{-m/2}|\Sigma|^{-1/2} \int_{\mathbf{x}\in\mathcal{B}_d} \max\left\{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T S(\mathbf{x}-\boldsymbol{\mu})\right), \exp\left(-\frac{1}{2}\mathbf{x}^T S\mathbf{x}\right)\right\} \\ & \times \left\{|\boldsymbol{\mu}^T S\mathbf{x}| + \frac{1}{2}|\boldsymbol{\mu}^T S\boldsymbol{\mu}|\right\} d\mathbf{x} \\ \leq\; & 2\sup_{\mathbf{x}\in\mathbf{B}_d} |\boldsymbol{\mu}^T S\mathbf{x}| + |\boldsymbol{\mu}^T S\boldsymbol{\mu}|. \end{aligned} \tag{4.4.39}$$

Putting the two upper bounds together, it is clear that

$$\begin{aligned} d_{TV}(\mathbf{V}, \mathbf{Z}_k) &\leq \sum_{i=1}^m \left\{\Phi\left(-\frac{d - m\mu_i}{m\sigma_i}\right) + \Phi\left(-\frac{d + m\mu_i}{m\sigma_i}\right) + 2\Phi\left(-\frac{d}{m\sigma_i}\right)\right\} \\ &+ 2\sup_{\mathbf{x}\in\mathcal{B}_d} |\boldsymbol{\mu}^T S\mathbf{x}| + |\boldsymbol{\mu}^T S\boldsymbol{\mu}|. \end{aligned} \tag{4.4.40}$$

Now that bounds for $d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Y}(k,\mathbf{z}))$ and $d_{TV}(\mathbf{Y}(k,\mathbf{z}), \mathbf{Z}_k)$ have been obtained, we can establish the α-mixing of $\{\mathbf{Z}_t\}$. Let $\tilde{\mathcal{B}}_k = \mathcal{B}_{\lambda^{-k/4}}$. Since the total variation distance between any two random variables is at most 1, it follows from (4.4.13) that

$$\begin{aligned}\alpha_k &\leq P(\mathbf{Z}_0^d \notin \tilde{\mathcal{B}}_k) + \int_{\mathbf{z}\in\tilde{\mathcal{B}}_k} f(\mathbf{z}) d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k)\, d\mathbf{z} \\ &= \sum_{i=1}^{m} \Phi\left(-\frac{\lambda^{-k/4}}{m\sigma_i}\right) + \int_{\mathbf{z}\in\tilde{\mathcal{B}}_k} f(\mathbf{z}) d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k)\, d\mathbf{z}. \quad (4.4.41)\end{aligned}$$

Note that $\sup_{\mathbf{z}\in\tilde{\mathcal{B}}_k} |A^k\mathbf{z}| \to 0$. Therefore there exists $k_4 \in \mathbb{N}$ such that for all $k \geq k_4$ and $\mathbf{z} \in \tilde{\mathcal{B}}_k$, $A^k\mathbf{z} \in \mathcal{B}_1$.

From the bounds proven for $d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k)$, we have that for all $k \geq \max\{k^*, k_4\}$ and $\mathbf{z} \in \tilde{\mathcal{B}}_k$,

$$\begin{aligned}d_{TV}(\{\mathbf{Z}_k|\mathbf{Z}_0 = \mathbf{z}\}, \mathbf{Z}_k) &\leq \beta^*\lambda^{2k} + 2\sup_{\mathbf{z},\mathbf{x}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S\mathbf{x}| + \sup_{\mathbf{z}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S A^k\mathbf{z}| \\ &+ 4\sum_{i=1}^{m} \Phi\left(-\frac{\lambda^{-k/4} - m}{m\sigma_i}\right). \quad (4.4.42)\end{aligned}$$

As stated earlier, there exists a diagonal matrix D consisting of the eigenvalues of A and invertible matrix L such that $A = L^{-1}DL$. Hence $A^k = L^{-1}D^kL$, where D^k is a diagonal matrix consisting of the k^{th} powers of the eigenvalues of A.

Therefore, as L and S are matrices of fixed terms, using (4.4.29) , (4.4.22) and the definition of $\tilde{\mathcal{B}}_k$, there exists $c_1, c_2, c_3 < \infty$, such that

$$\sup_{\mathbf{z},\mathbf{x}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S\mathbf{x}| \leq c_1\lambda^k c_2\lambda^{-k/4} c_3\lambda^{-k/4}. \quad (4.4.43)$$

Similarly, there exists $d_1, d_2, d_3 < \infty$

$$\sup_{\mathbf{z}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S A^k\mathbf{z}| \leq \left(d_1\lambda^k d_2\lambda^{-k/4}\right)^2 d_3. \quad (4.4.44)$$

Thus there exists $H_1, H_2 < \infty$, such that for all sufficiently large k,

$$\sup_{\mathbf{z},\mathbf{x}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S\mathbf{x}| \leq H_1\lambda^{k/2}, \quad (4.4.45)$$

and

$$\sup_{\mathbf{z}\in\tilde{\mathcal{B}}_k} |(A^k\mathbf{z})^T S A^k\mathbf{z}| \leq H_2\lambda^{3k/2}. \quad (4.4.46)$$

Examining the terms $\sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}\mp m\mu_i}{m\sigma_i}\right)$ and $\sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}}{m\sigma_i}\right)$ next,

$$\sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}+m\mu_i}{m\sigma_i}\right) < \sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}}{m\sigma_i}\right) < \sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}-m\mu_i}{m\sigma_i}\right) \to 0. \tag{4.4.47}$$

There exists k_0 such that for all $k \geq k_0$,

$$\sum_{i=1}^{m}\Phi\left(-\frac{\lambda^{-k/4}-m\mu_i}{m\sigma_i}\right) \leq \sum_{i=1}^{m}\Phi\left(-\lambda^{-k/8}\right). \tag{4.4.48}$$

$$\begin{aligned}\Phi\left(-\lambda^{-k/8}\right) &= \int_{-\infty}^{-\lambda^{-k/8}}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\mathrm{d}y \\ &\leq \int_{-\infty}^{-\lambda^{-k/8}}\frac{1}{\sqrt{2\pi}}e^{-|y|}\mathrm{d}y \\ &= \int_{\lambda^{-k/8}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y}\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}}e^{-\lambda^{-k/8}} \\ &\leq \lambda^{k/2},\end{aligned} \tag{4.4.49}$$

for all sufficiently large k.

Combining (4.4.42), (4.4.45) and (4.4.46) with (4.4.41), we have that there exists $H < \infty$, such that for all sufficiently large k,

$$\alpha_k \leq H\lambda^{k/2}. \tag{4.4.50}$$

Thus $\{Z_t\}$ is α-mixing with an geometrically decreasing α_k. This exponential rate of convergence to zero is what gives the strong mixing of $\{Z_t\}$ potential for extension to unbounded functions of $\{Z_t\}$, such as exponentials or infinite summations.

4.4.3 Strong mixing of periodic Poisson regression models with a secondary AR latent process

Now the α-mixing of $\{\mathbf{Z}_t\}$ has been established, we shall try to extend it to periodic Poisson regression models with a secondary latent AR process. For each time point

$t \in \mathbb{Z}$, we assume that there exists a vector of covariates, $\mathbf{x}_t$ and an $MAR(m,1)$ process $\mathbf{Z}_t$ such that

$$\{Y_t|\phi, \mathbf{Z}_t = \mathbf{z}\} \sim Po(\exp(\mathbf{x}_t^T\boldsymbol{\theta})\exp(\mathbf{Z}_t)), \tag{4.4.51}$$

where the dimensionality of $F_t(\cdot)$ depends upon the distribution D. The following Theorem shows that $\{Y_t\}$ are α-mixing. Suppose that $\{Y_t\}$ satisfies (4.4.51) and $\{\mathbf{Z}_t\}$ is an $MAR(m,1)$ process. Then $\{Y_t\}$ is α-mixing with mixing coefficients $\{\tilde{\alpha}_k\}$ such that for all sufficiently large k, $\tilde{\alpha}_k \leq H\lambda^{k/2}$.

Proof. We show that the mixing coefficients for $\{Y_t\}$ satisfies the bounds obtained for $\{\mathbf{Z}_t\}$. Fix $n \in \mathbb{Z}$ and $k \in \mathbb{N}$. For $l, m \in \mathbb{Z}$, let $\tilde{\mathcal{F}}_l^m = \sigma(Y_i; l \leq i \leq m)$. Let

$$\begin{aligned} A = \{(\ldots, \mathbf{Y}_{n-1}, \mathbf{Y}_n)\} \quad &, \quad B = \{(\mathbf{Y}_{n+k}, \mathbf{Y}_{n+k+1}, \ldots)\} \\ C = \{(\ldots, \mathbf{Z}_{n-1}, \mathbf{Z}_n)\} \quad &, \quad D = \{(\mathbf{Z}_{n+k}, \mathbf{Z}_{n+k+1}, \ldots)\} \\ E = \{\mathbf{Z}_n\} \quad &, \quad F = \{\mathbf{Z}_{n+k}\}\,. \end{aligned} \tag{4.4.52}$$

Let $f_{n,k}(\cdot,\cdot)$ denote the joint pdf of $\mathbf{Z}_n$ and $\mathbf{Z}_{n+k}$ and let $f(\cdot)$ denote the pdf of $\mathbf{Z}_n$ ($\mathbf{Z}_{n+k}$). Then

$$\begin{aligned} & \sup_{A \in \tilde{\mathcal{F}}_{-\infty}^n, B \in \tilde{\mathcal{F}}_{n+k}^\infty} |P(AB) - P(A)P(B)| \\ = & \sup_{A \in \tilde{\mathcal{F}}_{-\infty}^n, B \in \tilde{\mathcal{F}}_{n+k}^\infty} \left| \int \{f_{n,k}(\mathbf{x}, \mathbf{y}) P(AB|\mathbf{Z}_n = \mathbf{x}, \mathbf{Z}_{n+k} = \mathbf{y}) \right. \\ & \left. \quad - f(\mathbf{x}) f(\mathbf{y}) P(A|\mathbf{Z}_n = \mathbf{x}) P(B|\mathbf{Z}_{n+k} = \mathbf{y})\}\, d\mathbf{x} d\mathbf{y} \right|. \end{aligned} \tag{4.4.53}$$

Note that since A, B are independent given C and D.

$$\begin{aligned} P(A, B|C, D) &= P(A|C, D)\, P(B|C, D) \\ &= P(A|C)\, P(B|D)\,. \end{aligned} \tag{4.4.54}$$

Hence,

$$\begin{aligned} P(A,B|E,F) &= \int\int P(A,B,C=x,D=y|E,F)\,\mathrm{d}x\mathrm{d}y \\ &= \int\int P(A,B|C=x,D=y,E,F)\,f(x,y|E,F)\,\mathrm{d}x\mathrm{d}y \\ &= \int\int P(A|C=x,D=y,E,F)\,P(B|C=x,D=y,E,F) \\ &\quad \times f(x,y|E,F)\,\mathrm{d}x\mathrm{d}y \\ &= \int\int P(A|C=x,E)\,P(B|D=y,F)\,f(x,y|E,F)\,\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{4.4.55}$$

Now

$$\begin{aligned} f(y|C=x,E,F) &= f(y|F) \\ f(x,y|E,F) &= f(x|E,F)\,f(y|C=x,E,F) \\ &= f(x|E)\,f(y|C=x,E,F) \end{aligned} \tag{4.4.56}$$

implying that

$$\Longrightarrow f(x,y|E,F) = f(x|E)\,f(y|F).$$

Therefore

$$\begin{aligned} &\int\int P(A|C=x,E)\,P(B|D=y,F)\,f(x,y|E,F)\,\mathrm{d}x\mathrm{d}y \\ =&\int\int P(A|C=x,E)\,P(B|D=y,F)\,f(x|E)\,f(y|F)\,\mathrm{d}x\mathrm{d}y \\ =&\left\{\int P(A|C=x,E)\,f(x|E)\,\mathrm{d}x\right\}\left\{\int P(B|D=y,F)\,f(y|F)\,\mathrm{d}y\right\} \\ =&\ P(A|E)\,P(B|F). \end{aligned} \tag{4.4.57}$$

Thus A and B are independent given $\mathbf{Z}_n$ and $\mathbf{Z}_{n+k}$. Therefore

$$P(AB|\mathbf{Z}_n=\mathbf{x},\mathbf{Z}_{n+k}=\mathbf{y}) = P(A|\mathbf{Z}_n=\mathbf{x})P(B|\mathbf{Z}_{n+k}=\mathbf{y}). \tag{4.4.58}$$

It follows from (4.4.58) that

$$\begin{aligned} &-\int |f_{n,k}(\mathbf{x},\mathbf{y})-f(\mathbf{x})f(\mathbf{y})|\,d\mathbf{x}d\mathbf{y} \\ \leq &\int \{f_{n,k}(\mathbf{x},\mathbf{y})P(AB|\mathbf{Z}_n=\mathbf{x},\mathbf{Z}_{n+k}=\mathbf{y}) - f(\mathbf{x})f(\mathbf{y})P(A|\mathbf{Z}_n=\mathbf{x})P(B|\mathbf{Z}_{n+k}=\mathbf{y})\}\,d\mathbf{x}d\mathbf{y} \\ \leq &\int |f_{n,k}(\mathbf{x},\mathbf{y})-f(\mathbf{x})f(\mathbf{y})|\,d\mathbf{x}d\mathbf{y}. \end{aligned} \tag{4.4.59}$$

Note that

$$\frac{1}{2}\int\left|\frac{f_{n,k}(\mathbf{x},\mathbf{y})}{f(\mathbf{x})}-f(\mathbf{y})\right|d\mathbf{y}=d_{TV}(\{\mathbf{Z}_{n+k}|\mathbf{Z}_n=\mathbf{x}\},\mathbf{Z}_{n+k}). \tag{4.4.60}$$

Therefore the right hand side of (4.4.53) is bounded above by

$$\int f(\mathbf{x})d_{TV}(\{\mathbf{Z}_{n+k}|\mathbf{Z}_n=\mathbf{x}\},\mathbf{Z}_{n+k})\,d\mathbf{x}, \tag{4.4.61}$$

which is the bound obtained for α_k in (4.4.13) as required. □

For example, suppose that

$$Y_t|\mathbf{Z}_t\sim\mathrm{Po}(\exp(\mathbf{x}_t^T\beta+\mathbf{Z}_t^T\theta)), \tag{4.4.62}$$

where Po denotes a Poisson random variable, $\mathbf{x}_t$ are covariates associated with time point t and β and θ are vectors of parameters. Since for any $\vartheta\in\mathbb{R}^m$, $E[\exp(\mathbf{Z}_t^T\vartheta)]<\infty$, it is straightforward to show that $\sup_t E[Y_t^4]<\infty$ if and only if $\sup_t\exp(\mathbf{x}_t^T\beta)<\infty$. The model given by (4.4.62) is the Poisson regression model considered in Davis et al. (2000).

4.5 Extension of results on $\hat{\boldsymbol{\theta}}_{GLM}$ for double latent processes

We shall now extend the uniform convergence in probability and asymptotic normality of the linear parameter estimator and the pointwise convergence of the frequency estimator, each established for the basic model in previous chapters, to three models with two latent processes. These are the two-period models with additive and with multiplicative trigonometric latent processes, respectively, and the single-period model with an ARMA process accounting for non periodic, monotonically decreasing dependence.

4.5.1 Consistency of $\hat{\boldsymbol{\theta}}_{GLM}$

The convergence of the variance of the log-likelihood $V(l_T(\boldsymbol{\theta}))\longrightarrow 0$ as $T\longrightarrow 0$ is a sufficient condition for $\{\hat{\boldsymbol{\theta}}_{GLM}-\boldsymbol{\theta}_0\}$ to be pointwise convergent to 0 and equicontinuous in probability, which together establish uniform convergence of $\hat{\boldsymbol{\theta}}_{GLM}$ to $\boldsymbol{\theta}_0$. Thus for each of the double-latent-process models we shall examine the limiting behaviour of its likelihood variance. In each case $y_t \sim Po\left(\exp(\mathbf{x}_t^T\boldsymbol{\theta})\varepsilon_t\right)$ where $\exp(\mathbf{x}_t^T\boldsymbol{\theta})$ is the unconditional mean of y_t and ε_t denotes the latent process. Let $\tilde{l}_T(\boldsymbol{\theta},\vartheta)$ denote the log-likelihood for a single-process model with frequency ϑ.

For the additive model, with latent process $\varepsilon_t = \dfrac{2}{1+A^2}\big(\cos^2(\omega t+\phi)+A^2\cos^2(\delta t+\psi)\big)$

$$\begin{aligned}
V\left(l_T(\boldsymbol{\theta},\omega,\delta)\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(y_t,y_s)\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta} \\
&= \frac{1}{2T^2(1+A^2)^2}\Bigg[\sum_{t=1}^{T}(\mathbf{x}_t^T\boldsymbol{\theta})^2 e^{\mathbf{x}_t^T\boldsymbol{\theta}} + \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}e^{\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}} \\
&\times \big(\cos(2(t-s)\omega)+A^4\cos(2(t-s)\delta)\big)\Bigg] \\
&\propto V\left(\tilde{l}_T(\boldsymbol{\theta},\omega)\right) + A^2 V\left(\tilde{l}_T(\boldsymbol{\theta},\delta)\right) \\
&\longrightarrow 0 \text{ as } T \longrightarrow \infty. \qquad (4.5.1)
\end{aligned}$$

For the multiplicative model, with latent process $\varepsilon_t = 4\cos^2(\omega t+\phi)\cos^2(\delta t+\psi)$

$$\begin{aligned}
V\left(l_T\left(\boldsymbol{\theta}\right)\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(y_t,y_s)\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta} \\
&= \frac{1}{8T^2}\Bigg[\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2 e^{\mathbf{x}_t^T\boldsymbol{\theta}} + \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}e^{\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}} \\
&\times \Big\{4\cos(2(t-s)\omega)+\cos\left(2(t-s)(\omega+\delta)\right)+4\cos(2(t-s)\delta) \\
&\quad +\cos\left(2(t-s)(\omega-\delta)\right)\Big\}\Bigg] \\
&\propto V\left(l_{T1}\left(\boldsymbol{\theta},\omega\right)\right)+V\left(l_{T2}\left(\boldsymbol{\theta},\delta\right)\right)+V\left(l_{T3}\left(\boldsymbol{\theta},(\omega+\delta)\right)\right)+V\left(l_{T4}\left(\boldsymbol{\theta},(\omega-\delta)\right)\right) \\
&\longrightarrow 0 \text{ as } T \longrightarrow \infty. \qquad (4.5.2)
\end{aligned}$$

For the model with trigonometric-log-ARMA latent process

$\varepsilon_t = 2\exp(z_t - \sigma^2/2)\cos^2(\omega t + \phi)$

$$\begin{aligned}
V\left(l_T\left(\boldsymbol{\theta}\right)\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}\right)\\
&\times \left[\frac{\cos(2(t-s)\omega)+2}{2}\left(\exp(\gamma(t-s))-1\right)+\frac{\cos(2(t-s)\omega)}{2}\right]\\
&= \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)^2\left\{\exp\left(\mathbf{x}_t^T\boldsymbol{\theta}\right)-\exp\left(2\mathbf{x}_t^T\boldsymbol{\theta}\right)\left[\frac{3\exp\left(\sigma_z^2/2\right)}{2}-1\right]\right\}\\
&+ \frac{2}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}\right)\\
&\times \left[\frac{\cos(2(t-s)\omega)}{2}\exp(\gamma(t-s))+\left(\exp(\gamma(t-s))-1\right)\right]. \qquad (4.5.3)
\end{aligned}$$

Let $\gamma(h)$ denote the autocorrelation function of an arbitrary ARMA process. For a stationary process, there exists $K < \infty$ and $0 < r < 1$, such that for all $h \in \mathbb{N}, |\gamma(h)| \leq Kr^h$

$$\begin{aligned}
&\frac{2}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T}\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}\right)\left(\exp(\gamma(t-s))-1\right)\\
\leq\ & \frac{K\sup\left(\mathbf{x}_s^T\boldsymbol{\theta}\mathbf{x}_t^T\boldsymbol{\theta}\exp\left(\mathbf{x}_s^T\boldsymbol{\theta}+\mathbf{x}_t^T\boldsymbol{\theta}\right)\right)}{T^2}\sum_{s=1}^{T}\sum_{t=s}^{T}r^{t-s}\\
\propto\ & \frac{1}{T(1-r)}+\frac{1}{T^2(1-r)^2}\\
\longrightarrow\ & 0 \text{ as } T\longrightarrow\infty. \qquad (4.5.4)
\end{aligned}$$

4.5.2 Asymptotic normality of $\hat{\boldsymbol{\theta}}_{GLM}$

Let $P_T\left(\boldsymbol{\theta}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)\left(y_t-\exp(\mathbf{x}_t^T\boldsymbol{\theta})\right)$ be any linear projection of the rescaled score function $\sqrt{T}S_T\left(\boldsymbol{\theta}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\binom{1}{t/T}\left(y_t-\exp(\mathbf{x}_t^T\boldsymbol{\theta})\right)$. By the Cramer-Wold theorem (Billingsley (1968), pages 48-49), the convergence of the characteristic function $\varphi_{P_T}(\lambda)$ of $P_T\left(\boldsymbol{\theta}\right)$ to $\exp\left(i\lambda\mu-\lambda^2\sigma^2/2\right)$ (the characteristic of a Gaussian random variable $X\sim N(\mu,\sigma^2)$) as $T\longrightarrow\infty$, is a sufficient condition for

$\sqrt{T}S_T\left(\boldsymbol{\theta}\right)\underset{D}{\longrightarrow}N\left(\mathbf{0},\boldsymbol{\Sigma}\right)$.

This in turn, by the Mean-Value theorem, is a sufficient condition for $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{GLM}-\boldsymbol{\theta}\right)$ to converge in distribution to a Gaussian random variable.

The limiting behaviour of the characteristic function of each of the double-process models shall thus be studied - as the expected limit of a conditional Poisson characteristic function for the two double-period models and as the unconditional expectation of a Gaussian characteristic function for the model with a secondary ARMA latent process. The strong mixing of $\{y_t\}$ established in the previous section will be utilised in the last case. In each case $y_t \sim Po\left(\exp(\mathbf{x}_t^T\boldsymbol{\theta})\varepsilon_t\right)$ where $\exp(\mathbf{x}_t^T\boldsymbol{\theta})$ is the unconditional mean of y_t and ε_t denotes the latent process.

$$\begin{aligned}
&E\left[\exp\left(P_T\left(\boldsymbol{\theta}\right)\right)|\phi,\psi\right]\\
=\ &\prod_{t=1}^{T} E\left(\exp\left(\left(\gamma+\eta t/T\right)i\lambda\left(y_t-\mu_t\right)\right) \mid \phi\right)\\
=\ &\prod_{t=1}^{T}\exp\left[\mu_t\varepsilon_t\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1\right)-i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\mu_t\right].
\end{aligned} \tag{4.5.5}$$

For

$$\{y_t|\phi,\psi\} \sim Po\Big(\mu_t\big(1+\cos(2(\omega t+\phi))\big)\big(1+\cos(2(\delta t+\psi))\big)\Big)$$

this is equal to

$$\begin{aligned}
&\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1-i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)\mu_t\right]\\
\times\ &\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1\right)\mu_t\right.\\
&\times\Bigg(\cos\left(2\left(\omega t+\phi\right)\right)+\cos\left(2\left(\delta t+\psi\right)\right)+\frac{1}{2}\cos\left(2\left((\omega+\delta)t+(\phi+\psi)\right)\right)\\
&\left.+\frac{1}{2}\cos\left(2\left((\omega+\delta)t+(\phi+\psi)\right)\right)\Bigg)\right].
\end{aligned} \tag{4.5.6}$$

Similarly for $\{y_t|\phi,\psi\} \sim Po\left(\mu_t\frac{1}{1+A^2}\left(1+\cos(2(\omega t+\phi))+A^2+A^2\cos(2(\delta t+\psi))\right)\right)$,

$$\begin{aligned}
E\left[\exp\left(P_T\left(\boldsymbol{\theta}\right)\right)|\phi,\psi\right] &= \exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1-i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)\mu_t\right]\\
&\times\ \exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1\right)\mu_t\right.\\
&\left.\times\frac{1}{1+A^2}\Bigg(\cos\left(2\left(\omega t+\phi\right)\right)+A^2\cos\left(2\left(\delta t+\psi\right)\right)\Bigg)\right].
\end{aligned} \tag{4.5.7}$$

Both of these conditional characteristic functions are the product of a term independent of $\{\phi, , \psi\}$, identical to that seen for the single-process case, multiplied by several $\{\phi, , \psi\}$-dependent terms, each of the form

$$\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1\right)\mu_t\cos\left(2\left(\vartheta t+\varphi\right)\right)\right] \tag{4.5.8}$$

for some frequency ϑ and random phase-shift φ. We have proven that any such term converges to one, so

$$E\left[\exp\left(P_T\left(\boldsymbol{\theta}\right)\right)\right] \longrightarrow \lim_{T\to\infty}\left\{\exp\left[\sum_{t=1}^{T}\left(\exp\left(i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)-1-i\lambda\left(\frac{\gamma+\eta t/T}{\sqrt{T}}\right)\right)\mu_t\right]\right\}$$

for both double-period models.

The vital factor in evaluating the limit of $\varphi(P_T)$ for the single-process model or either of the double-period models is the factorisation of the conditional characteristic function into the product of a frequency-independent term and a frequency-dependent term. The latter can then be shown to converge to 1, so the limit of the unconditional characteristic function is equal to that of the former, fixed-value term. This is not possible when examining $\varphi(P_T)$ for the model with a secondary ARMA latent process, due to $\varepsilon_t = 2e^{z_t}\cos^2(\omega t+\phi)$ being non-cyclic. Instead we shall use mixing properties of some ARMA processes to help calculate the asymptotic characteristic function.

$$\begin{aligned}
P_T\left(\boldsymbol{\theta}\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)\left(y_t-\mu_t\right)\\
E\left(P_T\left(\boldsymbol{\theta}\right)\middle|\,\phi\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)\left(E\left(y_t|\phi\right)-\mu_t\right)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)\mu_t\cos(2\omega t+\phi)
\end{aligned} \tag{4.5.9}$$

$$\begin{aligned}
V\left(P_T\left(\boldsymbol{\theta}\right)\middle|\,\phi\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\gamma+\delta s/T\right)\left(\gamma+\delta t/T\right)Cov\Big(\left(y_s-\mu_s\right),\left(y_t-\mu_t\right)\Big|\phi\Big)\\
&= \frac{1}{T}\Bigg[\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\gamma+\delta\frac{s}{T}\right)\left(\gamma+\delta\frac{t}{T}\right)\mu_t\mu_s\mathrm{C}_t\mathrm{C}_s\left(e^{\sigma^2\rho(t-s)}-1\right)\\
&+ \sum_{s=1}^{T}\left(\gamma+\delta s/T\right)^2\mu_s\mathrm{C}_s\Bigg].
\end{aligned} \tag{4.5.10}$$

From Section 4.4, $\{Y_t|\phi\}$ is α-mixing with $\tilde{\alpha}_k \leq H\lambda^{k/2}$ for all sufficiently large k. Hence for all $\delta > 0$, $\sum_k k^{2/\delta}\tilde{\alpha}_k < \infty$. From Peligrad and Utev(1997), Theorem 2.2(c) we have the following theorem;

Let $\{a_{ni}; 1 \leq i \leq n\}$ be a triangular array of real numbers such that

$$\sup_n \sum_{i=1}^{n} a_{ni}^2 < \infty \quad \text{and} \quad \max_{1\leq i\leq n} |a_{ni}| \to 0 \quad \text{as} \quad n \longrightarrow \infty. \tag{4.5.11}$$

Then if there exists $\delta > 0$ such that $\{|Y_t|^{2+\delta}\}$ is uniformly integrable and $\inf_t var(Y_t) > 0$,

$$\tilde{S}_n = \sum_{t=1}^{n} a_{nt}\left(Y_t - \nu_t\right) \xrightarrow{D} N(\mu, \sigma^2) \qquad \text{as } n \longrightarrow 0 \tag{4.5.12}$$

provided that there exists $\mu < \infty$ and $\sigma^2 < \infty$ such that $E(\tilde{S}_n|\phi) \to \mu$ and $var(\tilde{S}_n|\phi) \to \sigma^2$ as $n \longrightarrow 0$. $P_T = \dfrac{1}{\sqrt{T}}\sum\limits_{t=1}^{T}\left(\gamma + \delta t/T\right)\left(y_t - \mu_t\right)$ is of the form $\tilde{S}_T$ with $a_{Tt} = \gamma + \eta t/T$. It is clear that P_T satisfies all the conditions required above, so $\{P_T|\phi\} \xrightarrow{D} N\left(\lim\limits_{T\to\infty}\{E\left(P_T|\,\phi\right)\}, \lim\limits_{T\to\infty}\{V\left(P_T|\,\phi\right)\}\right)$. Thus

$$\begin{aligned}\lim_{T\to\infty}\left\{\varphi\big(P_T|\phi\big)\right\} &= \lim_{T\to\infty}\left\{E\left[\exp\left(\lambda i\, P_T|\,\phi\right)\right]\right\} \\ &= \lim_{T\to\infty}\left\{\exp\left\{\lambda i E\left(P_T|\,\phi\right) - \frac{\lambda^2}{2}V\left(P_T|\,\phi\right)\right\}\right\}. \end{aligned} \tag{4.5.13}$$

Thus

$$\begin{aligned}
\varphi(P_T) &= E(\varphi(P_t|\phi)) \\
&= E\Bigg[\exp\Bigg(\frac{\lambda i}{\sqrt{T}}\sum_{t=1}^{T}(\gamma+\delta t/T)\,\mu_t(\mathrm{C}_t-1)-\frac{\lambda^2}{T}\sum_{t=1}^{T}(\gamma+\delta t/T)^2\,\mu_t\mathrm{C}_t \\
&- \frac{\lambda^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\gamma+\delta s/T)(\gamma+\delta t/T)\,\mu_t\mu_s\mathrm{C}_t\mathrm{C}_s\left(e^{\sigma^2\rho(t-s)}-1\right) \\
&= \exp\Bigg[\frac{-\lambda^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\gamma+\delta s/T)(\gamma+\delta t/T)\,\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right) \\
&\quad -\frac{\lambda^2}{T}\sum_{t=1}^{T}(\gamma+\delta t/T)^2\,\mu_t\Bigg] \\
&\times\ E\Bigg[\exp\Bigg\{\frac{\lambda i}{\sqrt{T}}\sum_{t=1}^{T}(\gamma+\delta t/T)\,\mu_t\cos(2(t\omega+\phi)) \\
&\quad -\frac{\lambda^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big[(\gamma+\delta s/T)(\gamma+\delta t/T)\,\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right) \\
&\times\ \Big\{\cos(2(t\omega+\phi))+\cos(2(s\omega+\phi))+\cos(2(t\omega+\phi))\cos(2(s\omega+\phi))\Big\}\Big] \\
&-\ \frac{\lambda^2}{T}\sum_{t=1}^{T}(\gamma+\delta t/T)^2\,\mu_t\cos(2(t\omega+\phi))\Bigg\}\Bigg]. \qquad (4.5.14)
\end{aligned}$$

By Abel's Lemma, $\sum_{t=1}^{T} a_t b_t$ is finite if $B_\tau = \sum_{t=1}^{\tau} b_t$ is bounded and $\sum_{t=1}^{T-1}(a_{t+1}-a_t)$ is absolutely convergent. Thus

$$\sum_{t=1}^{T}(\gamma+\delta t/T)\,e^{\alpha+\beta t/T}\cos(2(\omega t+\phi))$$

and

$$\sum_{t=1}^{T}(\gamma+\delta t/T)^2\,e^{\alpha+\beta t/T}\cos(2(\omega t+\phi))$$

are finite, as both

$$\begin{aligned}
\sum_{t=1}^{T}\Delta\left\{(\gamma+\delta t/T)\,e^{\alpha+\beta t/T}\right\} &= \sum_{t=1}^{T}\Big\{(\gamma+\delta t/T)\,e^{\alpha+\beta t/T}\left(e^{\beta/T}-1\right) \\
&\quad +e^{\alpha+\beta(t+1)/T}\delta/T\Big\}, \qquad (4.5.15)
\end{aligned}$$

and

$$\sum_{t=1}^{T}\Delta\left\{(\gamma+\delta t/T)^2 e^{\alpha+\beta t/T}\right\} = \sum_{t=1}^{T}\Big\{ (\gamma+\delta t/T)^2 e^{\alpha+\beta t/T}\left(e^{\beta/T}-1\right) + 2e^{\alpha+\beta(t+1)/T}\left[(\gamma+\delta t/T)\,\delta/T+\delta^2/T^2\right]\Big\}, \tag{4.5.16}$$

where $\Delta\{f(t)\} = f(t+1)-f(t)$, are absolutely convergent and $\sum_{t=1}^{\tau}\cos\left(2(\omega t+\phi)\right) \leq \frac{1}{\sin(\omega)}$.

The case where ω is a multiple of π is excluded, as then $\cos(2(\omega t+\phi))$ is a fixed term and hence the latent process is not periodic.

Hence $\frac{\lambda i}{\sqrt{T}}\sum_{t=1}^{T}(\gamma+\delta t/T)\,\mu_t\cos\left(2(\omega t+\phi)\right) - \frac{\lambda^2}{T}\sum_{t=1}^{T}(\gamma+\delta t/T)^2\mu_t\cos\left(2(\omega t+\phi)\right) \longrightarrow 0$ as $T\longrightarrow\infty$.

Next we will consider the final frequency-dependent term of $\varphi\left(P_t|\phi\right)$

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{ (\gamma+\delta s/T)(\gamma+\delta t/T)\,\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right)\\
&\quad\times\Big[\cos\left(2(\omega t+\phi)\right)+\cos\left(2(\omega s+\phi)\right)+\cos\left(2\omega(t+s)+4\phi\right)\Big]\Big\}\\
=\;& 2\sum_{s=1}^{T}\sum_{t=s}^{T}\Big\{ (\gamma+\delta s/T)(\gamma+\delta t/T)\,\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right)\\
&\quad\times\left(\cos\left(2(\omega t+\phi)\right)+\cos\left(2(\omega s+\phi)\right)+\frac{1}{2}\cos\left(2\omega(t+s)+4\phi\right)\right)\Big\}\\
-\;& \sum_{t=1}^{T}(\gamma+\delta t/T)^2\mu_t^2\left(2\cos\left(2(\omega t+\phi)\right)+\frac{1}{2}\cos\left(4(\omega t+\phi)\right)\right)\left(e^{\sigma^2}-1\right).
\end{aligned} \tag{4.5.17}$$

After rescaling and adjustment of frequency and exponential parameters,

$$\sum_{t=1}^{T}(\gamma+\delta t/T)^2\mu_t^2\left(2\cos\left(2(\omega t+\phi)\right)+\frac{1}{2}\cos\left(4(\omega t+\phi)\right)\right)\left(e^{\sigma^2}-1\right) \tag{4.5.18}$$

has the same functional form as

$$\sum_{t=1}^{T}(\gamma+\delta t/T)^2\mu_t\cos\left(2(t\omega+\phi)\right) \tag{4.5.19}$$

so it too is finite.

$$\begin{aligned}
&\sum_{s=1}^{T}\sum_{t=s}^{T}\Bigg\{\left(\gamma+\delta\frac{s}{T}\right)\left(\gamma+\delta\frac{t}{T}\right)e^{2\alpha+\beta(t+s)/T}\left(e^{\sigma^2\rho(t-s)}-1\right)\\
&\times\left(\cos\left(2(\omega t+\phi)\right)+\cos\left(2(\omega s+\phi)\right)+\frac{1}{2}\cos\left(2(\omega(t+s)+2\phi)\right)\right)\Bigg\}\\
=\;&\sum_{s=1}^{T}\sum_{h=0}^{T-s}\Bigg\{\left(\gamma+\delta\frac{s}{T}\right)\left(\gamma+\delta\frac{s+h}{T}\right)e^{2\alpha+\beta(h+2s)/T}\left(e^{\sigma^2\rho(h)}-1\right)\\
&\times\left(\cos\left(2(\omega(s+h)+\phi)\right)+\cos\left(2(\omega s+\phi)\right)+\frac{1}{2}\cos\left(2(\omega(h+2s)+2\phi)\right)\right)\Bigg\}.
\end{aligned} \tag{4.5.20}$$

Each part of the above can be rewritten in the form $\sum_{s=1}^{T} a_s B_s$ where
$a_s = e^{2\alpha+2\beta s/T}\cos\left(\omega_1 s+\phi_1\right)\left(\gamma+\delta s/T\right)^{\kappa_1}$ and
$B_s = \sum_{h=0}^{T-s}\left(\delta h/T\right)^{\kappa_2}e^{\beta h/T}\cos\left(\omega_2 s+\phi_2\right)\left(e^{\sigma^2\rho(h)}-1\right)$
where $\{\kappa_1,\kappa_2\}=\{1,1\}$ or $\{2,0\}$.

By Abel's lemma,

$$\begin{aligned}
\sum_{s=1}^{T} a_s B_s &= B_T\sum_{s=0}^{T}a_s - B_0 a_0 - \sum_{s=0}^{T-1}\left(\left[B_{s+1}-B_s\right]\sum_{r=0}^{s}a_r\right)\\
&= b_0\sum_{s=0}^{T}a_s - a_0\sum_{s=0}^{T}b_s + \sum_{s=0}^{T-1}\left(b_{T-s}\sum_{r=0}^{s}a_r\right).
\end{aligned} \tag{4.5.21}$$

Thus

$$\begin{aligned}
&\sum_{s=1}^{T}e^{2\alpha+2\beta s/T}\cos\left(\omega_1 s+\phi_1\right)\left(\gamma+\frac{\delta s}{T}\right)^{\kappa_1}\\
\times\;&\sum_{h=0}^{T-s}\left(\frac{\delta h}{T}\right)^{\kappa_2}e^{\beta h/T}\cos\left(\omega_2 h+\phi_2\right)\left(e^{\sigma^2\rho(h)}-1\right)\\
=\;&\sum_{s=1}^{T}a_s B_s
\end{aligned} \tag{4.5.22}$$

is bounded by a function of parameters for all T if $\sum_{s=1}^{t}a_s$, $\sum_{h=0}^{T}b_h$ and $\sum_{s=0}^{T-1}|b_{T-s}|$ are bounded for all parameter values and all values of T.

$\sum_{s=1}^{t}a_s$ is identical in structure to $\sum_{s=1}^{t}\left(\gamma+\delta t/T\right)e^{\alpha+\beta t/T}\cos\left(2(\omega t+\phi)\right)$ or

$\sum_{s=1}^{t} (\gamma + \delta t/T)^2 \, e^{\alpha+\beta t/T} \cos(2(\omega t + \phi))$, both of which have already been shown to be bounded for all values of T.

$$\begin{aligned}
\sum_{h=0}^{T} b_h &= \sum_{h=0}^{T} (\delta h/T)^{\kappa_2} \, e^{\beta h/T} \cos(\omega_2 s + \phi_2)) \left(e^{\sigma^2 \rho(h)} - 1 \right) \\
&\leq \sum_{h=0}^{T} (|\delta| h/T)^{\kappa_2} \, e^{\beta h/T} \left(e^{\sigma^2 \rho(h)} - 1 \right) \\
&\leq \sum_{h=0}^{T} (|\delta| h/T)^{\kappa_2} \, e^{\beta h/T} K r^h \\
&= \begin{cases} K \left(\dfrac{1 - (e^{\beta/T} r)^{T+1}}{1 - (e^{\beta/T} r)} \right) & \text{for } \kappa_2 = 0, \\ |\delta| K \left(\dfrac{(T+1)(e^{\beta/T} r)^{T+1}}{T\,(e^{\beta/T} r - 1)} - \dfrac{e^{\beta/T} r \left((e^{\beta/T} r)^{T+1} - 1 \right)}{T\,(e^{\beta/T} r - 1)^2} \right) & \text{for } \kappa_2 = 1. \end{cases}
\end{aligned} \tag{4.5.23}$$

both of which are bounded for all T by a function depending only on the parameters. Similarly,

$$\begin{aligned}
\sum_{s=0}^{T-1} |b_{T-s}| &= \sum_{s=0}^{T-1} (|\delta|(T-s)/T)^{\kappa_2} \, e^{\beta(T-s)/T} |\cos(\omega_2 (T-s) + \phi_2)| \left| e^{\sigma^2 \rho(T-s)} - 1 \right| \\
&= \sum_{s=1}^{T} (|\delta| s/T)^{\kappa_2} \, e^{\beta s/T} |\cos(\omega_2 s + \phi_2)| \left| e^{\sigma^2 \rho(s)} - 1 \right| \\
&\leq \sum_{s=1}^{T} (|\delta| s/T)^{\kappa_2} \, e^{\beta s/T} K r^s \\
&= K \left(\frac{1 - (e^{\beta/T} r)^T}{1 - (e^{\beta/T} r} \right) \text{ or } |\delta| K \left(\frac{(e^{\beta/T} r)^T}{e^{\beta/T} r - 1} - \frac{e^{\beta/T} r \left((e^{\beta/T} r)^T - 1 \right)}{T\,(e^{\beta/T} r - 1)^2} \right).
\end{aligned} \tag{4.5.24}$$

Thus $\sum_{s=1}^{T} e^{2\alpha+2\beta s/T} \cos(\omega_1 s + \phi_1)) \left(\gamma + \frac{\delta s}{T} \right)^{\kappa_1} \sum_{h=0}^{T-s} \left(\frac{\delta h}{T} \right)^{\kappa_2} e^{\beta h/T} \cos(\omega_2 h + \phi_2)) \left(e^{\sigma^2 \rho(h)} - 1 \right)$ is bounded for all T, so

$$\begin{aligned}
&\frac{\lambda^2}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \Big[(\gamma + \delta s/T)(\gamma + \delta t/T)\, \mu_t \mu_s \left(e^{\sigma^2 \rho(t-s)} - 1 \right) \\
&\times \big\{ \cos(2(t\omega + \phi)) + \cos(2(s\omega + \phi)) + \cos(2(t\omega + \phi)) \cos(2(s\omega + \phi)) \big\} \Big] \\
\longrightarrow \;& 0 \text{ as } T \longrightarrow \infty.
\end{aligned} \tag{4.5.25}$$

We have shown that all parts of $i\lambda E\left(P_T|\,\phi\right)-\lambda^2 V\left(P_T|\,\phi\right)$ which are dependent upon ϕ converge to zero, so

$$\begin{aligned}
\lim_{T\to\infty}\left\{\varphi\left(P_T\right)\right\} &= \lim_{T\to\infty}\left\{E\left(\varphi\left(P_t|\phi\right)\right)\right\} \\
&= \lim_{T\to\infty}\left\{E\left(i\lambda E\left(P_T|\,\phi\right)-\lambda^2 V\left(P_T|\,\phi\right)\right)\right\} \\
&= \lim_{T\to\infty}\left\{i\lambda E\left(P_T|\,\phi\right)-\lambda^2 V\left(P_T|\,\phi\right)\right\} \\
&= \lim_{T\to\infty}\left\{\exp\left[\frac{-\lambda^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\gamma+\delta s/T\right)\left(\gamma+\delta t/T\right)\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right)\right.\right. \\
&\qquad \left.\left.-\frac{\lambda^2}{T}\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)^2\mu_t\right]\right\} \qquad (4.5.26)
\end{aligned}$$

which is the characteristic function of a zero-meaned Gaussian random variable with variance

$$\lim_{T\to\infty}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\gamma+\delta s/T\right)\left(\gamma+\delta t/T\right)\mu_t\mu_s\left(e^{\sigma^2\rho(t-s)}-1\right)+\sum_{t=1}^{T}\left(\gamma+\delta t/T\right)^2\mu_t\right\}.$$

4.5.3 Consistency of the ACF estimator for a double latent process and its DFT

As for the single-latent-process model, there is strong empirical evidence that the real part of the Discrete Fourier Transform (DFT) of the latent process autocovariance estimator, denoted as $\bar{R}_T(\theta)$, is maximised at or very close to the true frequency values. All moments and statistics calculated previously for either of the two double-period models resemble sums of several corresponding single-period model moments or statistics, allowing direct extension of results established for the single-period model to the double period model. Similarly the consistency of $\bar{R}_T(\theta)$ established for the single-period model can straightforwardly be adapted to prove consistency of $\bar{R}_T$ for the double-period model.

To establish consistency of $\bar{R}_T(\theta)$ for the model with a secondary ARMA latent process, we shall take the same approach as used for the single model. That is, showing pointwise convergence in probability of the DFT of the latent autocovariance calculated using the true means, denoted by $\tilde{R}_T(\theta)$, to its expectation $R_T(\theta)$ after establishing the convergence of $R_T(\theta)$ to a delta function $R(\theta)$ maximised at the true frequency. This is sufficient to prove of the convergence in probability of $\tilde{R}_T(\theta)$ to $R(\theta)$. Then showing that $\bar{R}_T(\theta)$ converges in probability to $\tilde{R}_T(\theta)$ is sufficient to establish the convergence in probability of $\bar{R}_T(\theta)$ to $R(\theta)$. Inserting the secondary ARMA latent process makes calculating third and fourth-order moments of $y_t - \mu_t$ potentially very difficult, so the strong mixing of Y_t shall be used to provide useful upper bounds. Let

$$\begin{aligned}
\tilde{R}_T(\theta) &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+s}\right)\cos(\theta s), \\
\bar{R}_T(\theta) &= \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s}\bar{\varepsilon}_t\bar{\varepsilon}_{t+s}\right)\cos(\theta s) \\
R_T(\theta) &= \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\left[\cos(2\omega s)e^{\gamma(s)} + 2\left(e^{\gamma(s)}-1\right)\right]\cos(\theta s), \qquad (4.5.27)
\end{aligned}$$

where $\tilde{\varepsilon}_t = y_t \exp\left(-\mathbf{x}_t^T\boldsymbol{\theta}_0\right) - 1$ and $\bar{\varepsilon}_t = y_t \exp\left(-\mathbf{x}_t^T\hat{\boldsymbol{\theta}}\right) - 1$.

$$\begin{aligned}
& E\left(\tilde{R}_T(\theta)\right) \\
=\; & \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s} e^{-\mathbf{x}_t^T\boldsymbol{\theta}_0}e^{-\mathbf{x}_{t+s}^T\boldsymbol{\theta}_0}E\left[\left(y_t - e^{\mathbf{x}_t^T\boldsymbol{\theta}_0}\right)\left(y_{t+s} - e^{\mathbf{x}_{t+s}^T\boldsymbol{\theta}_0}\right)\right]\right)\cos(\theta s) \\
=\; & \frac{1}{\tau}\sum_{s=0}^{\tau-1}\left(\frac{1}{T-s}\sum_{t=1}^{T-s} e^{-\mathbf{x}_t^T\boldsymbol{\theta}_0}e^{-\mathbf{x}_{t+s}^T\boldsymbol{\theta}_0}Cov\left(y_t y_{t+s}\right)\right)\cos(\theta s) \\
=\; & \frac{1}{\tau}\sum_{s=0}^{\tau-1}Cov\left(\varepsilon_t\varepsilon_{t+s}\right)\cos(\theta s) \\
=\; & \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\left[\cos(2\omega s)e^{\gamma(s)} + 2\left(e^{\gamma(s)}-1\right)\right]\cos(\theta s). \qquad (4.5.28)
\end{aligned}$$

Thus $\tilde{R}_T(\theta)$ is an unbiased estimate of $R_T(\theta)$. Now

$$\begin{aligned}
& \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\left[\cos(2\omega s)e^{\gamma(s)} + 2\left(e^{\gamma(s)}-1\right)\right]\cos(\theta s) \\
=\; & \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\left(\cos(2\omega s)+2\right)\left(e^{\gamma(s)}-1\right)\cos(\theta s) + \frac{1}{2\tau}\sum_{s=0}^{\tau-1}\cos(2\omega s)\cos(\theta s),
\end{aligned}$$

$$(4.5.29)$$

where

$$\begin{aligned}
& \frac{1}{2\tau}\sum_{s=0}^{\tau-1}(\cos(2\omega s)+2)\left(e^{\gamma(s)}-1\right)\cos(\theta s) \\
\leq\ & \frac{1}{2\tau}\sum_{s=0}^{\tau-1}3\left(e^{\gamma(s)}-1\right) \\
\leq\ & \frac{3}{2\tau}\sum_{s=0}^{\tau-1}\left(e^{\mathcal{K}r^s}-1\right) \\
\leq\ & \frac{3e^{\mathcal{K}}}{2\tau}\sum_{s=0}^{\tau-1}r^s \propto \frac{1-r^\tau}{\tau(1-r)} \\
\longrightarrow\ & 0 \text{ as } \tau \longrightarrow \infty.
\end{aligned} \tag{4.5.30}$$

Thus $R_T(\theta) \longrightarrow R(\theta) = \begin{cases} 0 & \text{if } \theta \neq 2\omega \\ \frac{1}{2} & \text{if } \theta = 2\omega. \end{cases}$

Next assume that $E\left[X_t^4\right] \leq K^2$ for some $K > 0$. Then $E\left[X_t^2\right] \leq K$ by Jensen's inequality. By the Cauchy-Scharz inequality, $E(XY) \leq \sqrt{E\left(X^2\right)E\left(Y^2\right)}$, so $V\left(X_tX_s\right) \leq E\left[X_t^2X_s^2\right] \leq \sqrt{K^2}\sqrt{K^2} = K^2$. Hence

$$\begin{aligned}
& V\left(\frac{1}{T-s}\sum_{t=1}^{T-s}X_tX_{t+s}\right) \\
=\ & \frac{1}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=1}^{T-s}Cov\left(X_tX_{t+s}, X_uX_{u+s}\right) \\
=\ & \frac{1}{(T-s)^2}\sum_{t=1}^{T-s}\left\{V\left(X_tX_{t+s}\right)+2\sum_{u=t+1}^{T-s}Cov\left(X_tX_{t+s}, X_uX_{u+s}\right)\right\} \\
\leq\ & \frac{K^2}{T-s}+\frac{2}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=t+1}^{T-s}Cov\left(X_tX_{t+s}, X_uX_{u+s}\right).
\end{aligned} \tag{4.5.31}$$

Let U_t and U_{t+k} be two elements of a strong mixing process with mixing coefficient $\alpha_u(k)$ such that $E\left(U_t^4\right) < \infty$ and $E\left(U_{t+s}^4\right) < \infty$. Then from Berkes and Morrow (1981),Lemma 2, $Cov\left(U_t, U_{t+k}\right) \leq 10\alpha_u(k)^{1/2}\sqrt{E\left(U_t^4\right)E\left(U_{t+k}^4\right)}$.

Applying this to the previous equation,

$$\begin{aligned}
& \frac{K^2}{T-s}+\frac{2}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=t+1}^{T-s} Cov\left(X_tX_{t+s},X_uX_{u+s}\right) \\
\leq\ & \frac{K^2}{T-s}+\frac{2}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=t+1}^{T-s} 10\alpha_{r-t-s}^{1/2}K^2 \\
=\ & \frac{K^2}{T-s}+\frac{2}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=1}^{T-s-t} 10\alpha_{u-s}^{1/2}K^2 \\
\leq\ & \frac{K^2}{T-s}+\frac{2}{(T-s)^2}\sum_{t=1}^{T-s}\sum_{u=1}^{T-s} 10\alpha_{u-s}^{1/2}K^2 \\
=\ & \frac{K^2}{T-s}+\frac{20K^2(T-s)}{(T-s)^2}\sum_{u=1}^{T-s}\alpha_{u-s}^{1/2} \\
=\ & \frac{K^2}{T-s}\left(1+20\sum_{t=1}^{T-s}\alpha_{t-s}^{1/2}\right) \\
\longrightarrow\ & 0 \text{ as } T\longrightarrow\infty. \qquad (4.5.32)
\end{aligned}$$

Therefore

$$\begin{aligned}
V\left(\tilde{R}_T(\theta)\right) &= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1} Cov\left(\tilde{\gamma}(s),\tilde{\gamma}(r)\right)\cos(\theta s)\cos(\theta r) \\
&\leq \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sum_{r=0}^{\tau-1}\sqrt{V\left(\tilde{\gamma}(s)\right)V\left(\tilde{\gamma}(r)\right)} \\
&= \frac{1}{\tau^2}\sum_{s=0}^{\tau-1}\sqrt{V\left(\tilde{\gamma}(s)\right)}\sum_{r=0}^{\tau-1}\sqrt{V\left(\tilde{\gamma}(r)\right)} \\
&= \left(\frac{1}{\tau}\sum_{s=0}^{\tau-1}\sqrt{V\left(\tilde{\gamma}(s)\right)}\right)^2 \\
&\leq \max_{0\leq s\leq\tau}\left\{V\left(\tilde{\gamma}(s)\right)\right\} \\
&\longrightarrow 0 \text{ as } T\longrightarrow\infty. \qquad (4.5.33)
\end{aligned}$$

Thus by Chebychev's inequality, $\left|\tilde{R}_T(\theta)-R_T(\theta)\right| \underset{P}{\longrightarrow} 0$ as $T\longrightarrow\infty$.

The next step is to prove that $\left|\bar{R}_T(\theta)-\tilde{R}_T(\theta)\right| \underset{P}{\longrightarrow} 0$ as $T\longrightarrow\infty$. We shall do this by splitting the probability into two parts, one where the maximum observed value exceeds T^α for some suitable value of α and the case where all terms are less than or equal to T^α.

$$\begin{aligned} P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon\right) &= P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T \le T^\alpha\right) \\ &+ P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T > T^\alpha\right) \\ &\le P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \cap M_T \le T^\alpha\right) + P\left(M_T > T^\alpha\right) \\ &= P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \mid M_T \le T^\alpha\right) P\left(M_T \le T^\alpha\right) \\ &+ P\left(M_T > T^\alpha\right), \end{aligned} \tag{4.5.34}$$

where $M_T = \max_{1 \le t \le T} \{y_t\}$.

The series of steps taken in Section 3.4 to establish that

$$P\left(\left|\bar{R}_T(\theta) - \tilde{R}_T(\theta)\right| > \epsilon \mid M_T \le T^\alpha\right) P\left(M_T \le T^\alpha\right) \longrightarrow 0 \text{ as } T \longrightarrow \infty \tag{4.5.35}$$

depend only on the maximum y-value and the distribution of $\sqrt{T}\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|$. Thus they also hold for a double-latent-process model, so it remains to prove that $P\left(M_T > T^\alpha\right) \longrightarrow 0$ as $T \longrightarrow \infty$.

$$\begin{aligned} P\left(M_T > T^\alpha\right) &= P\left(M_T > T^\alpha \cap L_T \le \gamma \log(T)\right) + P\left(M_T > T^\alpha \cap L_T > \gamma \log(T)\right) \\ &\le P\left(M_T > T^\alpha \cap L_T \le \gamma \log(T)\right) + P\left(L_T > \gamma \log(T)\right) \\ &= P\left(M_T > T^\alpha \mid L_T \le \gamma \log(T)\right) P\left(L_T \le \gamma \log(T)\right) \\ &+ P\left(L_T > \gamma \log(T)\right), \end{aligned} \tag{4.5.36}$$

where $L_T = \max_{1 \le t \le T} \{z_t\}$.

Let $\tilde{X}_T \sim Po(\lambda_T)$ where $\lambda_T = 2T^\gamma e^{\alpha + |\beta|}$.

Conditional upon $L_T \le \gamma \log(T)$, for all $t \le T$,

$$\lambda_T \ge 2\exp\left(\alpha + \beta t/T + z_t\right) \ \forall\, t \le T, \text{ so } \forall\, k > 0, .$$

Therefore for all $k \ge 0$

$$P\left(y_t > k | L_T \le \gamma \log(T)\phi\right) \le P\left(\tilde{X}_T > k\right)$$

giving by Markov's inequality

$$\begin{aligned} P\left(y_t > T^\alpha | L_T \leq \gamma \log(T), \phi\right) &\leq P\left(\tilde{X}_T > T^\alpha\right) \\ &\leq \frac{E\left(s^{\tilde{X}_T}\right)}{s^{T^\alpha}} \\ &= \frac{\exp((s-1)\lambda_T)}{s^{T^\alpha}}. \end{aligned} \tag{4.5.37}$$

Taking $s = 2$,

$$\begin{aligned} P\left(y_t > T^\alpha | L_T \leq \gamma \log(T), \phi\right) &\leq \frac{\exp(\lambda_T)}{2^{T^\alpha}} \\ &\propto \frac{\exp(T^\gamma)}{2^{T^\alpha}} \\ &= \exp\left(T^\gamma - T^\alpha \log(2)\right). \end{aligned} \tag{4.5.38}$$

Let $\tilde{X}_{1T}, \tilde{X}_{2T}, \ldots, \tilde{X}_{TT}$ be iid according to $\tilde{X}_T$ and let $\tilde{M}_T = \sup_{1\leq t\leq T}\left\{\tilde{X}_{tT}\right\}$.

$$\begin{aligned} P\left(\tilde{M}_T > x\right) &= P\left(\bigcup_{t=1}^{T}\left\{\tilde{X}_T \leq x\right\}\right) \\ &= \sum_{t=1}^{T} P\left(\tilde{X}_{tT} > x\right) \\ &\leq TP\left(\tilde{X}_T > x\right) \leq \frac{T}{2^x}e^\lambda. \end{aligned} \tag{4.5.39}$$

Since $P\left(y_t > T^\alpha | L_T \leq \gamma \log(T), \phi\right) \leq P\left(\tilde{X}_T > T^\alpha\right)$, we have that
$P\left(M_T > T^\alpha | L_T \leq \gamma \log(T), \phi\right) \leq P\left(\tilde{M}_{tT} > T^\alpha\right)$. Hence

$$\begin{aligned} P\left(M_T > T^\alpha | L_T \leq \gamma \log(T), \phi\right) &\leq \frac{Te^{T^\gamma}}{2^{T^\alpha}} \\ &= \frac{T}{\exp\left(T^\alpha \log(2) - T^\gamma\right)}. \end{aligned} \tag{4.5.40}$$

Now removing the dependence on ϕ

$$\begin{aligned} P\left(M_T > T^\alpha | L_T \leq \gamma \log(T)\right) &= \int_0^{2\pi} P\left(M_T > T^\alpha | L_T \leq \gamma \log(T), \phi\right) f(\phi)\, \mathrm{d}\phi \\ &\leq \int_0^{2\pi} P\left(\tilde{M}_{Tt} > T^\alpha\right) f(\phi)\, \mathrm{d}\phi \\ &= P\left(\tilde{M}_T > T^\alpha\right) \int_0^{2\pi} f(\phi)\, \mathrm{d}\phi \\ &= P\left(\tilde{M}_T > T^\alpha\right) \leq \frac{Te^{T^\gamma}}{2^{T^\alpha}}. \end{aligned} \tag{4.5.41}$$

Hence

$$\begin{aligned}\frac{Te^{T^\gamma}}{2^{T^\alpha}} &= \frac{Te^{T^\gamma}}{4^{T^\alpha/2}} \\ &= \frac{T}{(4/e)^{T^\alpha/2}e^{T^\alpha/2-T^\gamma}} \\ &\leq \frac{T}{\left(\frac{4}{e}\right)^{T^\alpha/2}} \end{aligned} \tag{4.5.42}$$

for any $\gamma < \alpha$.

For any $p > 0$ and $|c| > 1$, $\frac{n^p}{c^n} \longrightarrow 0$ as $n \longrightarrow \infty$. Taking $n = T^{\alpha/2}$, $\frac{Te^{T^\gamma}}{2^{T^\alpha}} \leq \frac{n^{2/\alpha}}{\left(\frac{4}{e}\right)^n} \longrightarrow 0$ as $n \longrightarrow \infty$, as $4/e > 1$ and $2/\alpha > 0$

$$P\left(L_T > \gamma \log(T)\right) \leq \sum_{t=1}^{T} P\left(Z_t > \gamma \log T\right). \tag{4.5.43}$$

Thus for all $\gamma < \alpha$, $P\left(M_T > T^\alpha | L_T \leq \gamma \log(T)\right) \longrightarrow 0$ and $P\left(L_T \geq \gamma \log(T)\right) \longrightarrow 0$ as $T \longrightarrow \infty$. Thus $P\left(M_T > T^\alpha \cap L_T \leq \gamma \log(T)\right) \longrightarrow 0$ and $P\left(M_T > T^\alpha \cap L_T \geq \gamma \log(T)\right) \longrightarrow 0$ as $T \longrightarrow \infty$, so $P\left(M_T > T^\alpha\right) \longrightarrow 0$ as $T \longrightarrow \infty$. This is sufficient to conclude that $\left|\hat{R}_T(\theta) - \tilde{R}_T(\theta)\right| \underset{P}{\longrightarrow} 0$ as $T \longrightarrow \infty$.

Chapter 5

Analysis of measles case counts in the UK

5.1 Introduction and preliminary analysis

We will analyse the following data set of measles occurrences in the UK from 1944-1966. From the beginning of 1944 to the end of 1966, the number of measles cases observed in each of sixty UK cities were recorded every two weeks. Various authors have studied this data set or subsets of it including the following; B.T. Grenfell, O.N. Bjørnstad, B.F. Finkenstädt and A. Morton. In Finkenstadt and Grenfell (2000), a discrete-time epidemic model incorporating a time-varying transmission parameter and birth rates, called the Time Series Susceptible-Infected-Recovered (TSIR) model, is developed to explain the predominant biennial pattern and the high variation in peak amplitudes apparent in the 1944-1966 60-city measles data set. The TSIR model is also considered in Grenfell et al (2002), where random immigration and population sizes are added to help simultaneously capture the behaviour of large-city endemic cycles and small-town episodic outbreaks. In Morton and Finkenstädt (2005), Markov-Chain-Monte-Carlo methods are applied to the TSIR model to make inference about the unknown parameters of interest and missing data in the form of unobserved populations. A epidemic metapopulation model, to better capture the spatiotemporal properties in measles epidemics, is constructed in Xia et. al.(2004)

by combining the TSIR model with a gravity model for regional spread.

A very different modelling approach will be taken in this chapter, based on the periodic count time series model for the case counts as numbers, opposed to the SIR model based on specifying the disease progression in each individual.

Studying the multiplot of all sixty time series, there is strong empirical evidence of the same pattern of periodicity across most of the locations. There also appears to be high correlation between population and measles counts. The latter pattern is displayed most clearly by the maxima in London being over twice as large as those in any other city, with Manchester, Birmingham, Leeds and Liverpool also having high numbers. The plots of the total summation of cases (the "total UK counts") and that without London (the "capital-less UK counts") are more similar than one would expect from the high numbers of cases in London. This suggests that either most locations have a very similar oscillatory pattern or the capital-less UK counts are large enough to prevent London dominating the total UK counts, even though the number of cases in each individual location is relatively small compared with London. The observation times of the local maxima and minima ("peaks" and "troughs") of the total UK and capital-less UK counts are displayed in Table 5.1 below.

The average lag between both maxima and minima suggests a two-yearly periodicity. However the pattern is asymmetric, with the lowest counts observed between 12 and 16 fortnights after the highest counts, rather than the 26 one would

Table 5.1: Locations of extreme values

UK maxima	UK exc London maxima	UK minima	UK exc London minima
33	32	47	47
83	83	99	99
135	135	150	150
189	190	203	202
238	238	254	254
294	294	306	306
345	345	360	360
398	398	410	410
448	449	462	463
499	499	515	515
554	551	566	566

Figure 5.1: Multi-plot of the measles counts in 60 UK cities

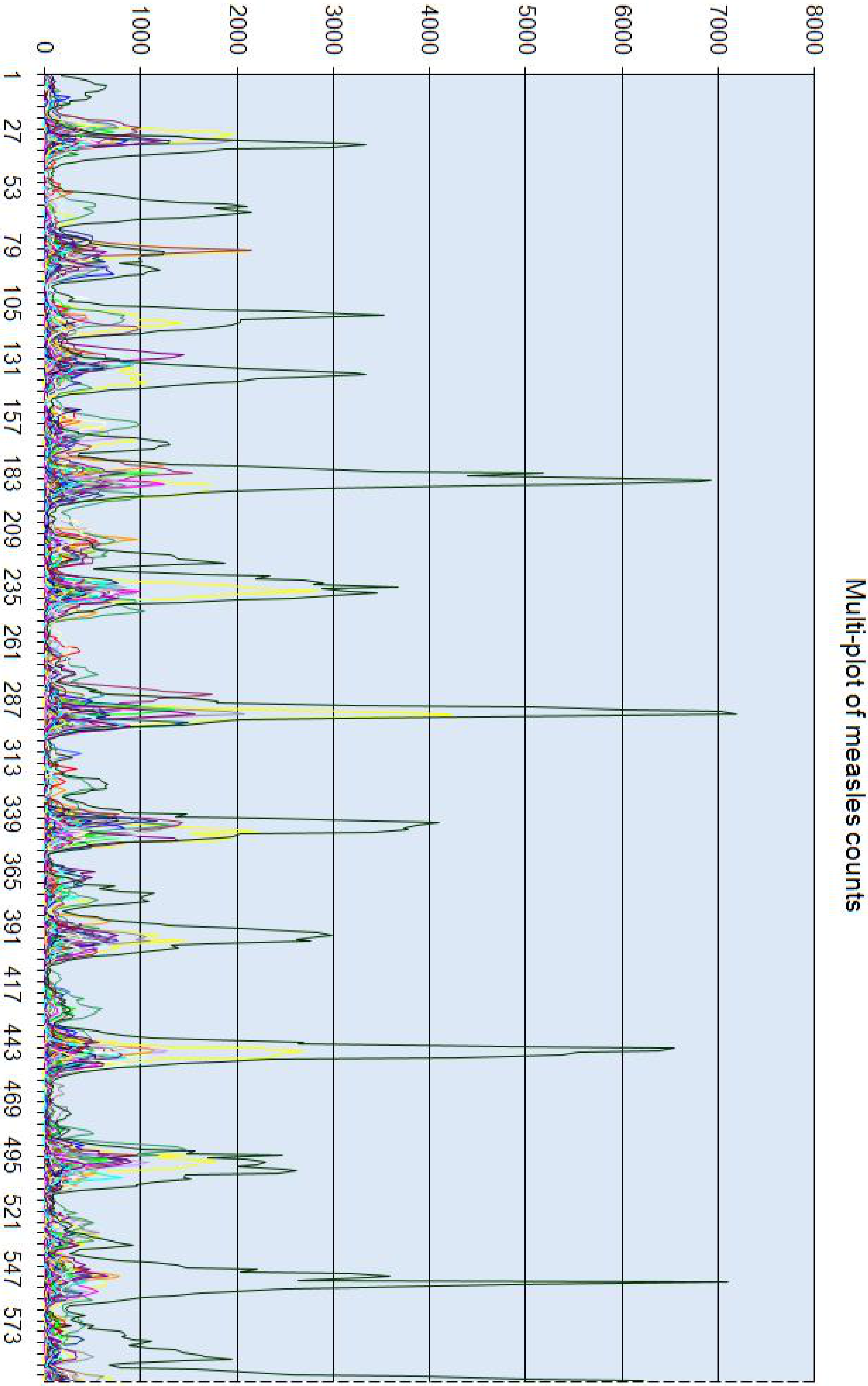
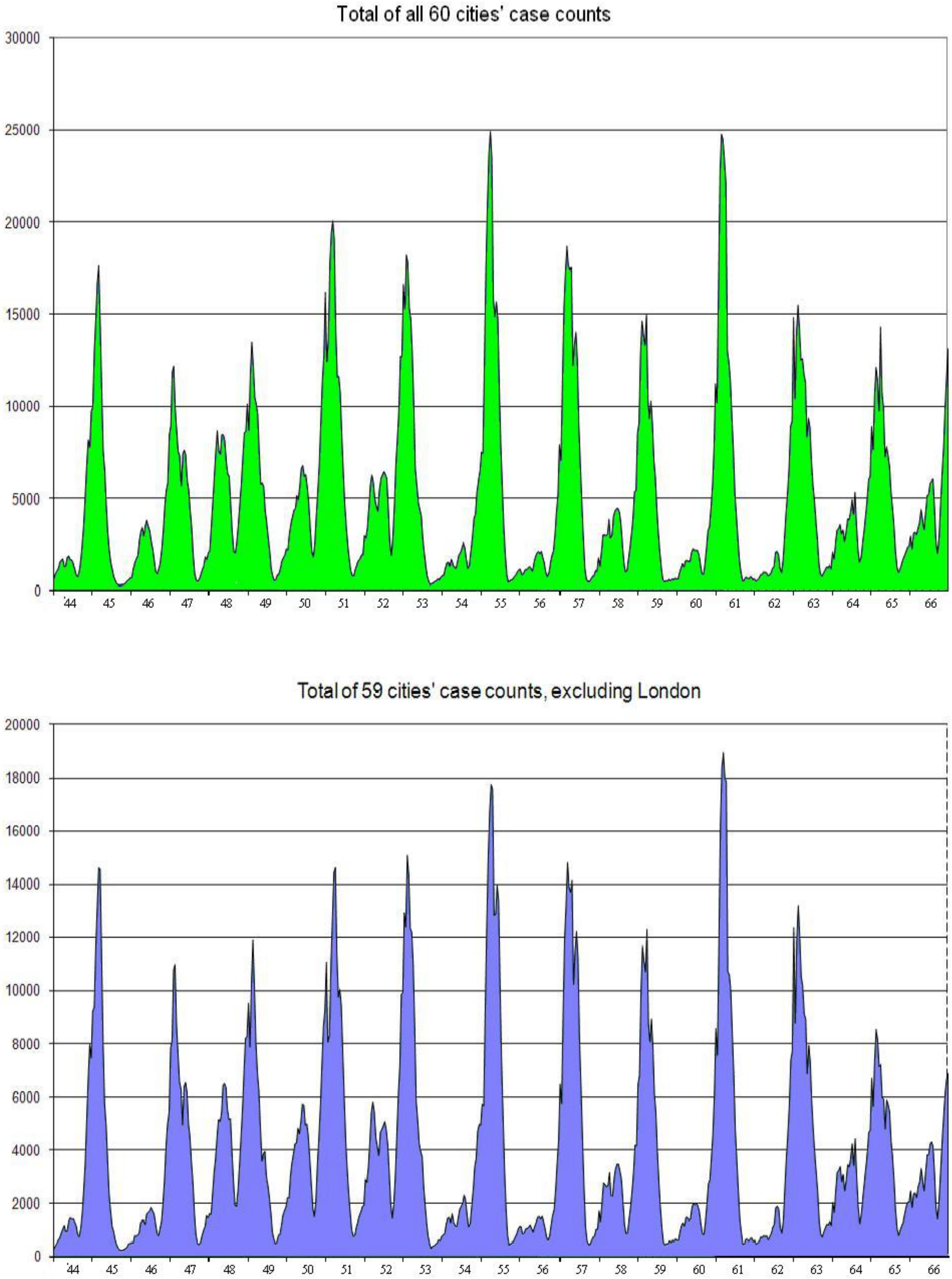


Figure 5.2: Plots of all UK measles counts, including and excluding London



see with a simple sin or cosine curve with period 52. The highest counts are observed in early spring (late February to mid-April) while the lowest counts are 6.5 to 7 months later, in early to mid-October. There is also a varying degree of minor peaks and troughs in the data, giving the plot a similarity to a combination of $\cos\left(\frac{\pi t}{p}\right)$ and $\cos\left(\frac{2\pi t}{p}\right)$. This double periodicity might best be modelled as a latent process of the form $\left(A^2\cos^2\left(\omega t+\phi\right)+\cos^2\left(\delta t+\psi\right)\right)$ or the simpler form $\left(1+\cos\left(2\omega t+2\phi\right)\right) = \frac{1}{2}\left(3+4\cos\left(2\omega t+2\phi\right)+\cos\left(4\omega t+4\phi\right)\right)$ or a combination of latent process and trigonometric regressors. The "major-minor" pattern is probably due to the measles virus "taking" too many "victims" in the first year to maintain a large enough population to infect so many the next year, but the number of susceptible individuals recovering by the third year and so on. In many respects the measles counts can be thought of as predator-prey population statistics, with the infectious individuals acting as the predator and susceptible people as the prey.

A model for measles in the UK would ideally incorporate all potentially important covariates such as weather statistics, such as temperatures or precipitation, and population statistics such as size and birth rates. As measles is a respiratory disease transferred via nasal and oral fluids, infection is dependent on the amount of close contact between susceptible people. This is a factor that is likely to increase as population does or when people spend more time indoors due to cold or wet weather. Considering the role of contact and the fact that the majority of measles infections are in infant-school-age children, an indicator variable to distinguish between observations sampled during term-times and during holidays might also be important. Over twenty-three years, a large number of children will be born and reach the vulnerable age for measles. The birth rate or sub-population size is thus likely to be the main factor apart from the periodic components in explaining large differences in counts during different years. This was the conclusion reached in Finkenstadt and Grenfell (1998) where the effects of population size and both current and delayed birthrates, on both epidemic size and disappearance-reemergence cycles, were studied in depth.

Location-specific covariates such as regional population density and location isolation, which do not apply to the total case counts considered here, were also studied. Overall, potentially important regressors are:

- Time, not only as a linear function. If cases are decreasing in the long term, the rate could be logarithmic or negative exponential.
- Yearly seasonal function of the form $A\cos\left(\frac{2\pi t}{26}\right)+B\sin\left(\frac{2\pi t}{26}\right)$. This could account for much of the variation caused by seasonal climate changes.
- Total population, birth rate and sub-population consisting of the susceptible age group. The third is probably less thoroughly recorded than the other two but would be the most informative. Although both will have some significance, current birth rates would be expected to have a bigger effect on counts several years in the future than on present counts, while changes in the total population do not always match those in the vulnerable sub-population.
- Weather patterns, both seasonal and otherwise. Temperatures are seasonal in the long term but can vary greatly in the short term and precipitation is unpredictable though one might expect more in the winter. Rare occurrences such as heatwaves, droughts or floods could also significantly affect the counts in certain years via shared accommodation or high stress levels.

5.2 Building models for measles case counts in the UK

After consideration, we decided to fit a model to the total measles case counts in the UK rather than at any individual location. There are several types of potentially significant covariate, both seasonal and unseasonal. The classes of seasonal covariates are:

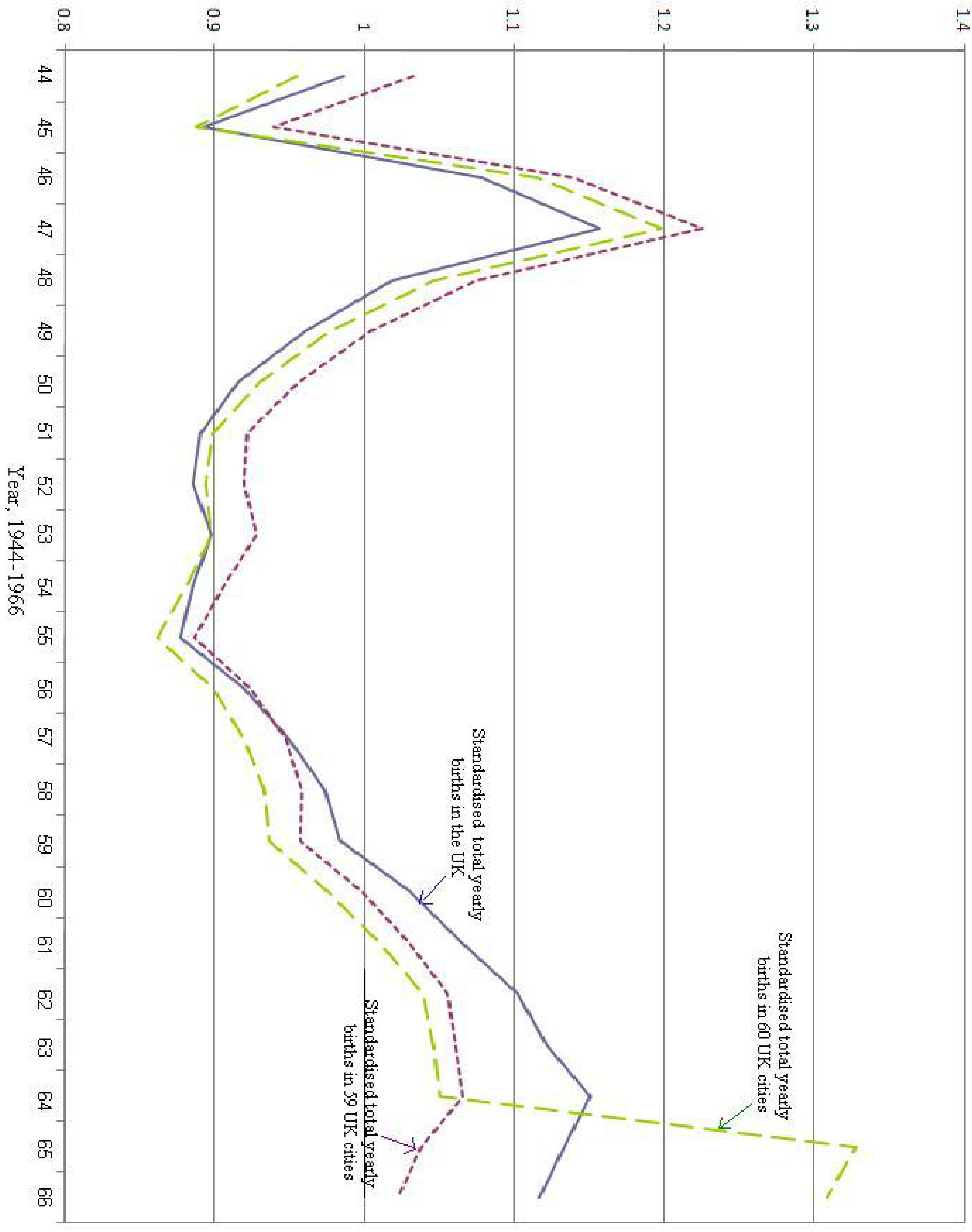
- Trigonometric functions of annual, biennial and maybe other periodicities. The most significant frequencies will be estimated using the RFT approach.

- Previous observation y_{t-1} or its logarithm. The latter gives a model which can be rewritten as $y_t \sim Po(y_{t-1}^{\gamma}\mathbf{x}_t^T\boldsymbol{\theta})$.
- Weather statistics. After preliminary investigation, the Meterological Office Hadley Centre data sets for Central England appear to be the best representations of UK weather patterns. These data sets consist of the average rainfall and average, maximum and minimum temperatures recorded daily in a triangle bordered by Lancaster, Bristol and London. These daily figures will be averaged, maximised or minimised over fortnightly intervals to provide weather regressors for our model.

The non-seasonal covariates also fall into three groups, namely:

- Trend functions. We shall consider the first four powers, the logarithm and the positive and negative exponentials of $\frac{t}{100}$.
- Indicator variable. A binary indicator taking value 1 during time intervals estimated to overlap with school holidays and 0 otherwise.
- Birth statistics, both past and present. The birth counts for each of the sixty locations, recorded annually for the same twenty-three years as the measles survey, and those for the whole of the UK over a much longer time period. Figure 5.3 displays the total birth counts for the sixty locations, the total for the fifty-nine locations excluding London and the UK birth counts, after dividing by their respective means. The sixty-city birth counts show a misleading jump corresponding to the redefinition of London Borough in 1965, while the 59-city and UK counts are similar in pattern. This suggests that analysing just the total of the measles case counts in the fifty-nine cities excluding London as a representation of the whole of the UK would be better if we wish to include birth statistics in the model. The 59-city births are likely to be more closely correlated to the data, but the UK births are not restricted to 1944-1966, allowing for the inclusion of longer-lagged birth statistics without reducing the data set.

Figure 5.3: Plots of the standardised yearly birth counts for the UK, the 60 locations and the 59 locations excluding London



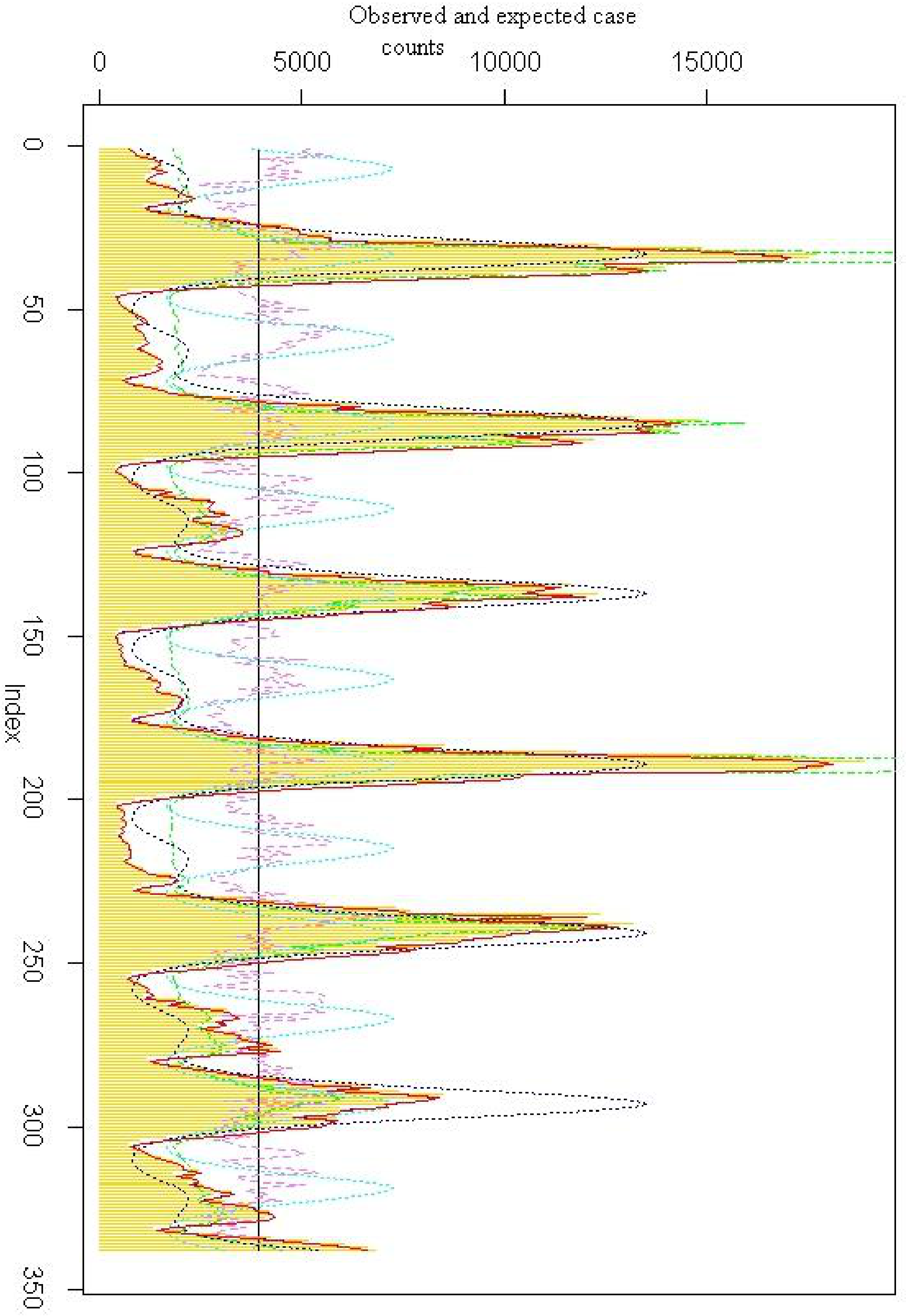
We will start by comparing the null model with the models incorporating one set of each of the covariates with oscillatory behaviour. Table 5.2 below displays the AIC, BIC and log-likelihood of the models with intercept only, with the four weather parameters, with a yearly trigonometric function, with the previous observation and with the logarithm of the previous observation. Figure 5.4 displays the actual measles counts with line plots of the counts predicted by each model. Both

Table 5.2: Goodness-of-fit statistics for periodic models

Model	AIC	BIC	Log-likelihood
Null	1180213	1180549	-590105.6
Weather	1120063	1121407	-560027.5
Annual function	849084.8	850092.8	-424539.4
Annual-biennial	188218.1	189898.1	-94104.07
Previous count	250410.3	251082.3	-125203.1
Previous log	74156.94	74828.94	-37076.47

the goodness-of-fit statistics and the plot of fitted values indicate that the model with single covariate $\log(Y_{t-1})$ is by far the best fitting model. As this model can be rewritten as $E(Y_t|Y_{t-1}) = \mathcal{A}Y_{t-1}^{\gamma}$, we shall refer to this as the linear observation-driven (LOD) model. The model with single covariate Y_{t-1}, denoted the exponential observation-driven (EOD) model, and that with annual and biennial trigonometric covariates, also both approve the model fit to the data, both graphically and statistically compared with the other parameter-driven models. The annual-biennial model is better at capturing the major-minor peak pattern of the data, while the EOD model is more suited to accounting for the variations in biennial maxima. The clear under and over-estimation by both models occurring during the times corresponding to the peaks and troughs of the data however cause large deviance compared with the LOD model. Graphically, the models with weather statistics and with a yearly trigonometric function appear to be equally poor-fitting, but the lower log-likelihood and information criteria of the latter indicate that temperatures and rainfall are less of a source of variation than an annual cycle. Bearing this in mind, we shall look for hidden periodicities in the measles counts under the LOD model, using the Discrete

Figure 5.4: Plot of the UK measles counts in histogram mode and the fitted values for the null model and four single regressor-type models. The values for the previous logarithm model, previous observation model, weather covariate model, annual and annual-biennial trigonometric function models and null model are shown by the solid red line, the dot-dashed green line, the dashed pink line, the dotted cyan line, the dotted navy line and the solid black line respectively.



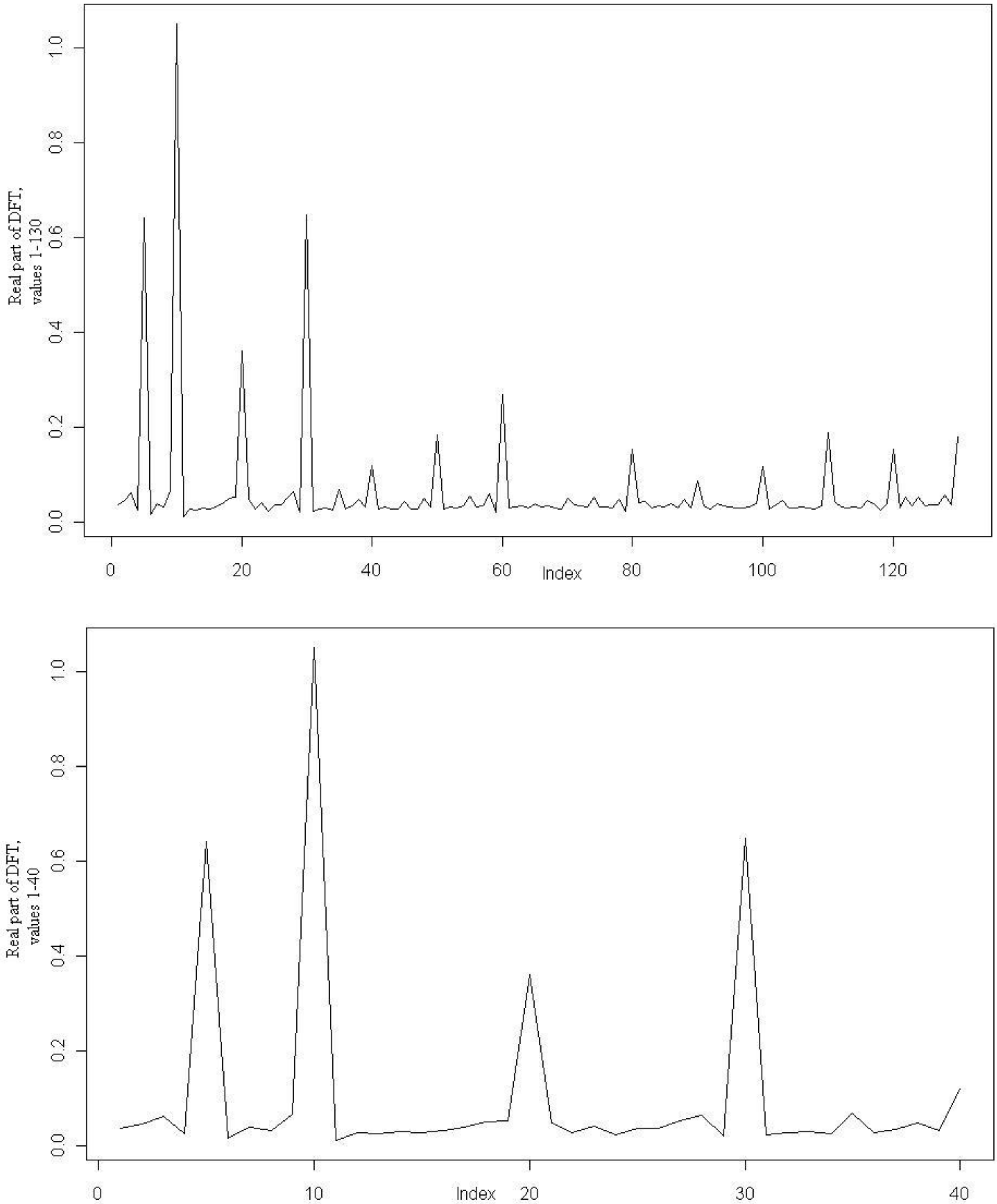
Fourier Transform estimator. We wish to account for the locations and relative amplitudes of maximum and minimum values over time using trigonometric functions rather than the precise shape of the data curve. To simplify phase-shift and goodness-of-fit estimation, any significant frequencies will be added to the model as log-linear trigonometic covariates rather than as trigonometric components of a linear latent process. Figure 5.5 shows the real part of the DFT, along with a subsection to better locate the maximising values. The tallest peak at 10, corresponding to a frequency of $2\pi/26$, indicates that the measles counts have yearly periodicity as well as short-term dependence. The two shorter peaks at 5 and 30 suggest that frequencies of $\pi/26$ and $3\pi/13$ might also be significant. To test this, the GLM estimators of four extensions of the LOD model are calculated and their goodness-of-fit statistics compared. These models incorporate functions with frequency $\pi/13$, with frequencies $\pi/13$ and $\pi/26$, with frequencies $\pi/13$ and $3\pi/13$ and with frequencies $\pi/13$, $\pi/26$ and $3\pi/13$. The results are displayed in Table 5.3 below. The large difference between the basic LOD

Table 5.3: Goodness-of-fit statistics for the LOD model with added trigonometric terms

Model periods	AIC	BIC	Log-likelihood
None	74156.94	74828.94	-37076.47
26	53025.86	54369.86	-26508.93
26 and 52	46644.07	48660.07	-23316.04
26 and 26/3	47359.81	49375.81	-23673.91
26, 52 and 26/3	40865.95	43553.95	-20424.98

model and that with annual periodicity, in the values of all three statistics, supports the inclusion of a trigonometric function with period 26. The peaks on the DFT corresponding to periods 52(two years) and 26/3(four months) were of equal height, suggesting that either both or neither were significant. Each pair of goodness-of-fit statistics for the two corresponding double-period LOD model models support this, being close in value to each other and significantly smaller than those for the annual LOD model. This suggests that three trigonometric functions, with four-monthly, annual and biennial periods respectively, would be optimal. This is supported by the AIC, BIC and log-likelihood for the three-period model being almost equally small relative to those for the annual-biennial model and those for the four-month-annual

Figure 5.5: Plots of values 1-130 and 1-40 for the DFT of the covariance function estimators for the LOD model, calculated up to lag 260



model.

As mentioned previously, the data plot of UK measles counts somewhat resembles a trigonometric function proportional to $A\cos\left(\frac{\pi t}{13}\right)+\cos\left(\frac{\pi t}{26}\right)$ or its exponential and we have strong hypotheses why a childhood respiratory disease would follow such a pattern. Thus the annual and biennial periodicities indicated by the DFT estimator are not unexpected. The four-month period, in contrast, is not apparent in the data plot and there is no obvious explanation why it should be almost as significant a factor as the two-yearly period. It is likely that the absence of an obvious four-month period from the plot is due to much of the difference between variables distributed around $\exp\left(A\cos\left(\frac{\pi t}{13}\right)+B\cos\left(\frac{\pi t}{26}\right)\right)$ and around $\exp\left(A\cos\left(\frac{3\pi t}{13}\right)+B\cos\left(\frac{\pi t}{26}\right)+C\cos\left(\frac{\pi t}{13}\right)\right)$ being small enough to attribute to random variation. Given that most of the minor peaks and a few of the major peaks are rather jagged, part of the significance of the four-month period could be due to much of the deviation from the two-period model being accounted for by a third trigonometric function with a short period, even when this deviation is purely random. Another potential explanation why measles case counts appear to be dependent on a four-month cyclic process is the effect of school holidays on measles infection. Primary-school children are likely to be exposed more to their close kin and less to people their own age during school holidays. Even if a school is very local and a family does not travel during holidays, the structure of the day will also be different, which could alter both the display and the reporting of symptoms. Although the three school terms are not usually of equal length, due to the length of the summer holiday and the high variation in the temporal location of the Easter holiday, the four-month trigonometric function might still be accounting well for variations in measles case counts caused several times annually by changes in disease transmission and expression during school holidays. Taking this into account, it seems reasonable to infer that the three-period LOD model is optimal at this stage. Later we shall investigate whether an alternative measure of holiday effects, such as an indicator variable, could replace the four-month trigonometric functions.

Next we shall see if adding any of the weather statistics to the three-period LOD model improves the fit further. From the statistics in Table 5.2, the weather-only model is better-fitting than the null model, indicating that temperatures and rainfall do have some effect on the measles case counts. However, the relatively poor fit of the weather-statistics model compared with the other single-set models suggests that any improvements in fit will be small compared with those achieved previously by adding trigonometric terms.

The GLM estimates of five triple-period LOD models, with each one of the weather measurements and with all four, are calculated and their goodness-of-fit statistics are displayed in Table 5.4 below, together with those for the three-period LOD model found to be optimal before. As one would expect whenever more covariates are

Table 5.4: Goodness-of-fit statistics for the three-period LOD model with added weather statistics

Model	AIC	BIC	Log-likelihood
Triple period	40865.95	43553.95	-20424.98
Mean temp	40848.61	43872.61	-20415.31
Min temp	40865.1	43889.1	-20423.55
Max temp	40867.6	43891.6	-20424.8
Rain	40867.59	43891.59	-20424.8
All weather	40761.96	44793.96	-20368.98

added to a model, there is some decrease in the log-likelihoods of the models with a single weather statistic and a larger decrease in that of the model with all four. These decreases, however, are very small and all the BICs and the AICs of three of the five models are larger than those of the prior three-period LOD model. This suggests that, despite earlier discussion of the likely effects of rainfall and temperatures on diseases via alterations in contact, none of the weather statistics have any significant effect on the measles case counts. It is possible that this is due to a combination of two factors; much of the variation that could be caused by weather is already being accounted for by the annual trigonometric function, and the two susceptible age groups; pre-school and primary school children, being less exposed to and less affected by the weather respectively than groups of individuals in the UK.

After studying the significance of previous observations, weather statistics and trigonometric functions, the optimal periodic-covariate model appears to be the linear observation-driven model with four-monthly, annual and biennial two-parameter trigonometric functions. We shall now see if the fit can be improved using four different groups of non-periodic covariates - a binary indicator for school holidays, a seven-variable set of the first four powers, the postive and negative exponentials and the logarithm of time, the annual births for the 59 cities from ten years past to the current year and the annual births for the UK from ten years past to the current year. Parameter estimates for the three-period LOD model with the covariates from each group are calculated and the goodness-of-fit statistics for each model are displayed in Table 5.5 below.

It is easy to see that the holiday-indicator is the only type of covariate which, when

Table 5.5: Goodness-of-fit statistics for the three-period LOD model with annual birth rates, time polynomial and holiday-time indicator

Model	AIC	BIC	Log-likelihood
Triple period	40865.95	43553.95	-20424.98
59 city births	38529.73	44913.73	-19245.87
UK births	38475.24	44859.24	-19218.62
Times	40018.34	45058.34	-19994.17
Holiday indicator	40232.76	43256.76	-20107.38

added to the three-period LOD model, reduces both the AIC and BIC. The AICs and log-likelihoods of the model with seven time covariates and the two models with eleven lagged birth counts are lower than those of the prior triple-period LOD model while the BICs are larger. This suggests that, as expected, time or birth rates do have some effect on measles case counts, but models containing all time or birth covariates are somewhat over-parameterized. Bearing this in mind, we shall take the three-period LOD model with holiday indicator as the new optimal model, then investigate the significance of individual lagged births or time covariates one-by-one rather than as a group. The BICs of the eleven models with one k-year lagged 59-city birth statistic (k= $0, 1, \ldots, 10$) are first compared with each other, as are those of the eleven models with one k-year lagged UK birth statistic and those of the seven models with one time measurement. The second, third and fourth rows of Table 5.6 display the goodness-of-fit statistics for the models from each set with lowest BIC. The models which include 8-year-lag UK and 59-city births respectively generate a smaller BIC

Table 5.6: Goodness-of-fit statistics for the holiday-indicator LOD model with the optimal single birth-rate, optimal pair of birth rates and optimal single time regressor.

Model	AIC	BIC	Log-likelihood
Holiday indicator	40232.76	43256.76	-20107.38
UK 8-year lag	39750.66	43110.66	-19865.33
59-city 8-year lag	39730.02	43090.02	-19855.01
Negative exp	39925.88	43285.88	-19952.94
UK 7+8-year lag	39391.37	43087.37	-19684.69
59-city 7+8-year lag	39404.71	43100.71	-19691.35

than the model without births. However the BIC is larger for a model including the negative-exponential time covariate. This suggests that the measles case counts have no significant trend in time but might be dependant on at least one of previous year's births. Thus the next stage is the comparison of the BICs of the 10 models with 8-year-lag UK birth and one other of the k-year lag births and the same for the 59-city births. The BICs in rows 5 and 6 of Table 5.6 suggest 7 and 8-year-lagged UK births are both significant while only the 8-year-lagged UK city births are significant. It is easy to see that all the goodness-of-fit statistics are smallest for the three-period LOD model with holiday indicator and 7 and 8-year-lag UK births, which we shall denote the 3P 7-8-lag LOD model with indicator. It is somewhat surprising that the best-fitting model appears to be that depending on birth rates in the whole of the UK, as one would expect the total measles case counts summed over the fifty nine locations to be more closely correlated with the total number of births in those locations. A possible explanation of this is that, due to daily commuting both of susceptible individuals and carriers, the measles case counts in any one location also depend on the number of births in a certain area around that location. This factor could have the cumulative effect of making UK birth rates apparently a better reflection of measles case counts summed over the fifty nine locations than the total birth rates in those locations. In other words, the birth rates in separate locations are being recorded over too narrow an area. It is also unexpected that the most significant birth rates appear to be those counted seven and eight years prior, representing the current population of seven and eight-year-olds. It could be that, due to the relative lack of information given in annual birth rates when the case counts are fortnightly, there is too little difference between the births in different years to explain much variation in measles case counts, so just one or two lagged birth rates are useful in the model as a representation of the average susceptible population during a particular year. It could be a coincidence that it is the populations of seven and eight-year olds that appear to best fulfil this role, or it could be that between the ages of seven and eight is when sub-population size is most important due to an optimal combination of socialising and susceptibility - seven and eight-year-olds probably have a larger social circle than younger children, making them more exposed, but are less likely to have already had measles than older children.

We mentioned earlier that the four-month trigonometric function might be less significant in a model with a holiday indicator. Other periodic functions might also have phase shifts close to 0 or π, making either the sin or cos term surplus to requirements. Thus we shall compare the goodness-of-fit statistics of the 3P 7-8-lag LOD model without indicator, without the least significant trigonometric term and without one of the two parts of the four-month period function to see if the parsimony of our model can be improved while maintaining a good fit.

The R-output summary of parameter estimates and their properties, for the covari-

Table 5.7: Parameter estimates and their properties

Covariate	Par.estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.04	0.0241	43.01	<2e-16
$\log(y_{t-1})$	0.866	0.00249	348.52	<2e-16
$\cos(\pi t/13)$	0.0791	0.0017	46.52	<2e-16
$\sin(\pi t/13)$	0.129	0.00198	65.08	<2e-16
$\cos(3\pi t/13)$	-0.110	0.00149	-73.87	<2e-16
$\sin(3\pi t/13)$	0.0370	0.00137	27.09	<2e-16
$\cos(\pi t/26)$	-0.181	0.00263	-68.67	<2e-16
$\sin(\pi t/26)$	0.0372	0.00210	17.71	<2e-16
holiday indicator	0.0609	0.00246	24.78	<2e-16
8-year-lag UK birth	0.000608	0.0000215	28.33	<2e-16
7-year-lag UK birth	-0.00059	0.0000312	-18.93	<2e-16

ates in the 3P 7-8-lag LOD model with indicator variable, is displayed in Table 5.7. Although the significance tests shown in the fourth and fifth columns indicate that all six trigonometric covariates are highly significant, $\sin(\pi x/26)$ has the smallest z-value, so we shall see what effects removing it has on goodness-of-fit compared with the four-month period terms.

The statistics in Table 5.8 indicate that all three four-month-period terms - the

Table 5.8: Goodness-of-fit statistics for the 3P 7-8-lag LOD model with and without binary indicator, without $\sin(\pi t/26)$, without $\cos(3\pi t/13)$ and without $\sin(3\pi t/13)$.

Model	AIC	BIC	Log-likelihood
All	39391.37	43087.37	-19684.69
Without indicator	40004.17	43364.17	-19992.08
Without $\sin(\pi t/26)$	39702.63	43062.63	-19841.32
Without $\cos(3\pi t/13)$	44915.36	48275.36	-22447.68
Without $\sin(3\pi t/13)$	40124.23	43484.23	-20052.11

binary indicator and the two trigonometric functions with frequency $3\pi/13$ - are highly significant, as the AICs, BICs and log-likelihoods of the models with one of these covariates removed are all larger than the corresponding statistics of the "full" model. This difference is particularly noticeable when the $\cos(3\pi t/13)$ term is removed from the model. In contrast, the model without $\sin(\pi t/26)$ has a slightly smaller BIC, suggesting that this term may not be significant. Although the AIC and log-likelihood of the model with $\sin(\pi t/26)$ are still slightly smaller, this is only to be expected when comparing two models, one of which is a more parsimonious version of the other. Thus our final optimal model for fitting the fortnightly measles case counts between 1954 and 1966 is the LOD model with trigonometric covariates $\sin(\pi t/13), \cos(\pi t/13), \cos(\pi t/26), \cos(3\pi t/13)$ and $\sin(3\pi t/13)$, binary indicator variable for school holidays and 7 and 8-year-lagged birth statistics for the whole of the UK. The y_{t-1}^{α} term is the only part of this model which also appears in SIR-type models such as those developed in Grenfell et.al. (2002) and in Finkenstadt and Morton (2005). As in the Poisson time series model, y_{t-1}^{α} plays a key role in the total number of infections at time point t in both of these papers and the estimate $\hat{\alpha} = 0.866$ is similar to those estimated in Grenfell et.al. (2002) and in Finkenstadt and Morton (2005), table 2, although there is no formal link between the α values in the different models. The two types of model also both make use of covariates such as lagged birth rates and periodic functions, although the model structures are very different and the SIR models incorporate many localised covariates but not deterministic trigonometric functions. Unlike the annual births for the 59 cities, which are available only for the same time period as the case counts themselves, annual births in the UK are recorded from the eighteenth century to the present. Thus our optimal model has the advantage over any model with lagged births from the 59 locations that it is extendable to all twenty-three years of measles case counts. Figure 5.6 displays the plot of all twenty-three years of UK measles case counts and a line plot of the values predicted by the LOD model with trigonometric covariates $\sin(\pi t/13), \cos(\pi t/13), \cos(\pi t/26), \cos(3\pi t/13)$ and $\sin(3\pi t/13)$, binary indicator variable for school holidays and 7 and 8-year-lagged birth statistics for the whole of the UK.

5.3 Potential extensions of the measles analysis to multivariate data sets and to other diseases

There are several alternatives to the univariate time series for analysing measles counts, including

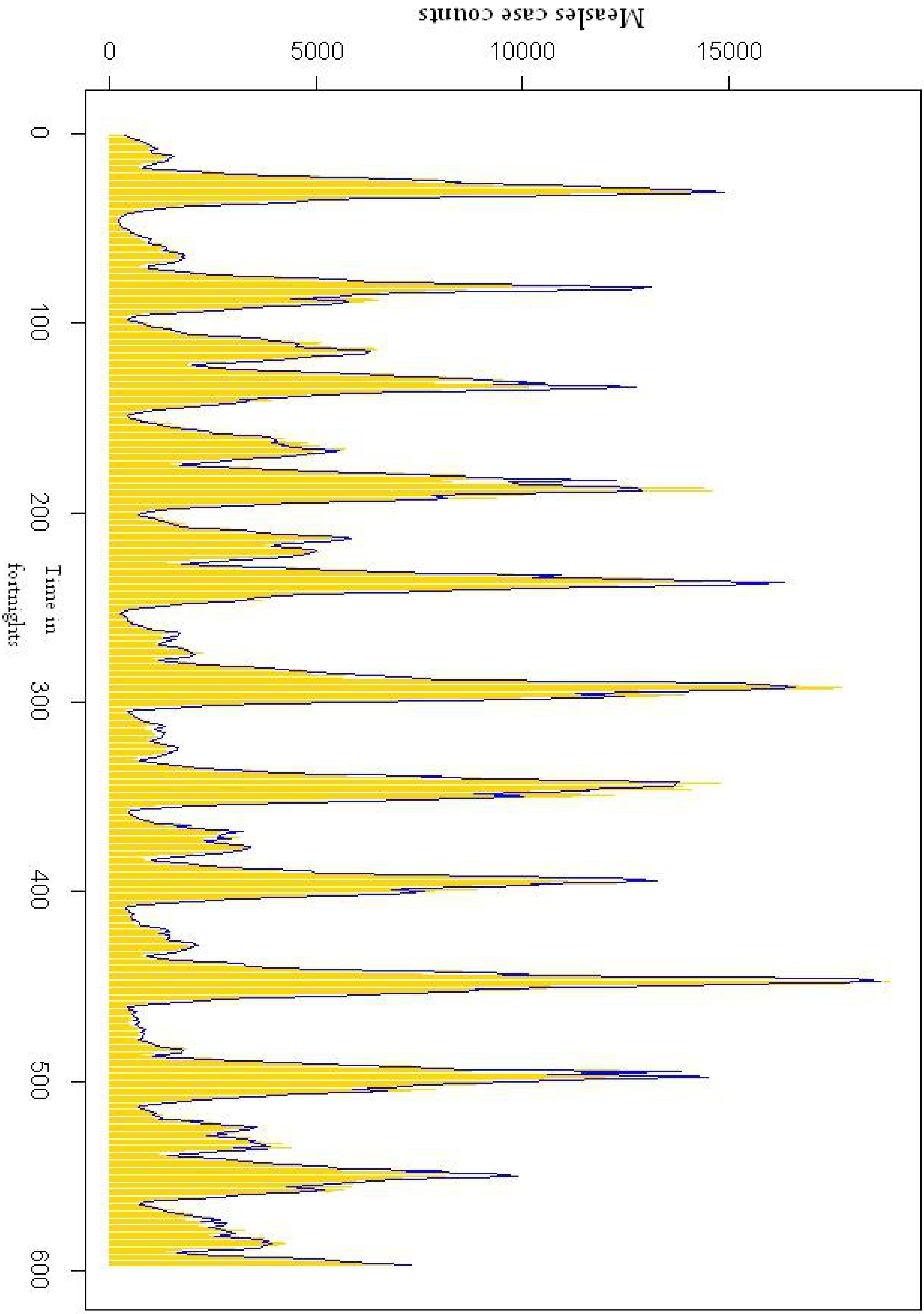
1. Taking a single location and building a model for the counts observed there alone.

2. Combining several locations judged to be in close proximity to create a single time series of larger values and then building a model for this.

3. Modelling all 60×598 observations or a subset as a longitudinal data sample.

4. Taking all 60×598 observations (or a subset) as a data matrix of up to 598 observations of a multidimensional variable or a 598×60 spatial-temporal data set

Figure 5.6: Plot of all 23 years of UK measles counts in histogram mode and the fitted values for the corresponding optimal model



All four models have pros and cons. (4), the multivariate or spatial-temporal model, would be the most thorough as all the individual observations can be used while correlation between counts in different locations can also be analysed. The longitudinal model (3) would also use all the data but would necessitate an assumption that observations are only correlated within locations and through the covariates. This is potentially a big assumption in the late 20th century due to the prevalence of daily commuting throughout modern Britain. Both models would be more complex and so require more computing power than the time-series models (1) and (2) which differ from that developed previously for the whole of the UK only in the subset of the data used. This is more pronounced for the multivariate model than the longitudinal model due to the former's complex, non-block-diagonal covariance matrix. The only extra complications in using model (1) would be the occurence of zeros and lack of smoothing compared with the Total-UK model. (2) would have some smoothing and fewer zeros than (1), but would need preliminary investigation into the proximity of locations before one can infer which groups of cities/towns can be thought of as a larger sub-population.

As mentioned previously, the apparent dependence on UK-wide births rather than those particular to the fifty-nine locations might indicate that daily commuting, children going to school in larger towns and adults commuting to work, has significant influence on measles case counts. Children who travel longer distances to school or have commuting parents could have a higher likelihood of infection due to extra exposure during travel on public transport, especially considering that people do not have to be infected with a respiratory virus to act as short-term carriers. Children's daily commuting also raises a important question about the counts themselves - where are they taken? Are some infected children in Greater Manchester counted as cases in Manchester, their school's location, or as cases in their home-town on the outskirts of Manchester? (Oldham, Bolton, etc) Another type of "movement" in the UK which could cause variation when looking at several locations is seasonal holiday travel. People moving from their home towns to coastal places such as Brighton and Yarmouth for holidays could cause a temporary shift in case counts during July or August, probably best defined either as an indicator variable or as some sort of measure of population shift during summer. Any model for measles in the UK as a whole gives one little idea of the effects of local population or local birth rates on the case counts observed in individual locations. These factors together with the effects of commuting suggest that there is much more potential for extending the analysis performed so far to a longitudinal, multivariate or spatial-temporal data set than to a time series model for a single location or sub-population. Although a longitudinal model would necessitate the unlikely assumption of location independence, one might still use covariates such as the number of neighbours a city has, the population statistics of those neighbours or even the previous case counts of close neighbours to examine the interactions between locations. Covariates from other locations could also be weighted to take into account the different distances between neighbours. The analysis of a multivariate or spatial-temporal data set would be more complicated, but could model the effects of other locations on mean and autocorrelation more efficiently as well as the cross-correlation of locations. The main sources of interest which could be studied in such analyses, opposed to the analysis of measles in just a single location, include the differences in the spread of measles between areas of high and of low population and the factors, such as population, relative isolation or location within the UK, which could affect how quickly measles re-emerges after dying off in a small town and vice-versa. One important modification which would have to be made before we could analyse measles case counts location-by-location is the adjustment of previous case counts. While the use of $\log(y_{t-1})$ as a covariate in the LOD model for the whole of the UK is not problematic due to the absence of zeros in the data, $y_t = 0$ multiple times in many of the smaller UK locations, causing y_s to be mistakingly estimated as 0 for all $s > t$. One way to avoid y_{t-s} repeatedly acting as an absorbing state would be to replace y_{t-1} by a strictly positive transformation y^*_{t-1}. Cameron and Trivedi (1998) suggest one of the two simple transforms

$$y^*_{t-1} = \max(c, y_{t-1})$$

and

$$y^*_{t-1} = c + y_{t-1}$$

, where c could, in increasing complexity, be a known constant, a extra fixed parameter to be estimated, or a process designed to predict the re-emergence of measles after a period of zero case counts.

When one considers how little factors such as weather statistics or birth rates vary from place to place within the UK compared with many other countries, the importance of models which can incorporate location-varying covariates become even more apparent. A developing country such as India, where much of the population is still rural, will have huge differences in wealth-related factors such as family size and health statistics, which in turn will cause significant differences in disease transmission via different levels of disease exposure and resilience. Although the USA is similar to the UK in relative birth rates and health, it has enormous locational differences in weather. Patterns of exposure through socialising are likely to differ greatly between the sub-tropical, humid Florida swamp, the hot, arid deserts of Arizona and sub-arctic Alaska. High differences in precipitation in particular also cause large differences in movement and population spread - scarcity of water means that most people in the arid south-west are clustered in one of a relatively small number of isolated urban areas, while people in wetter regions, particularly those with a mediterranean or sub-tropical climate, are much more evenly spread and much less isolated. These differences in population spread, movement and social patterns are likely to cause significant differences in disease transmission.

The complexity of the models needed to analyse multi-location disease counts in countries such as the USA or India would be very high. Not only would there be a variance matrix to estimate as well as a large number of mean parameters, the correlation between locations could be highly dependent on weather or birth statistics. This complexity suggests an additional direction for expansion; the utilisation of traditionally empirically robust methods such as expectation-maximisation (EM) or Markov-Chain-Monte-Carlo (MCMC) algorithms to make inferences about multivariate or spatial-temporal periodic count data models.

Bibliography

[1] Bebbington, M., Zitikis, R. (2004). A robust heuristic estimator for the period of a Poisson intensity function. *Methodology and Computing in Applied Probability* 6, *pp* 441-462

[2] Berkes, I. , Morrow, G. J. (1981). Strong invariance properties for mixing random fields. *Wahrscheinlichkeitstheorie* 57, *pp*15-37

[3] Berkes, I. , Horvath, L., Kokoszka, P. (2003). GARCH Processes: Structure and Estimation. *International Statistical Institute and Bernoulli Society for Mathematical Statisticsand Probability* Vol. 9, No. 2, *pp.* 201-227

[4] Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

[5] Brockwell, P. J., Davis, R. A. - Time series : theory and methods - 2nd ed. New York; London : Springer, 1991.

[6] Cameron, A. C., Trivedi, P. K. Regression analysis of count data. Cambridge : Cambridge University Press, 1998.

[7] Davis, R. A., Dunsmuir, W. T. M., Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* 87, *pp* 491-505.

[8] Finkenstädt, B., Grenfell, B.T. (1998). Empirical determinants of measles metapopulation dynamics in England and Wales *Proc. Roy. Soc B* 265, *pp* 211-220

[9] Finkenstädt, B., Grenfell, B.T. (2000). Time series modelling of childhood diseases: a dynamical systems approach *JRSSC* 49, *pp* 187-205

[10] Finkenstädt, B., Morton, A.(2005). Discrete Time Modelling of Disease Incidence Time Series by Using Markov Chain Monte Carlo Methods. *JRSSC* 54, *pp* 575-594

[11] Gallant, A. R. Nonlinear statistical models. New York; Chichester : Wiley, 1987.

[12] Genton, M. G., Hall, P.(2007). Statistical inference for evolving periodic functions. *J. R. Statist. Soc* B 69, *pp* 643-657.

[13] Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B. (2002). Endemic and epidemic dynamics of measles:Estimating epidemiological scaling with a time series SIR model *Ecological Monographs* 72 *pp*169-184

[14] Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B. (2002). Endemic and epidemic dynamics of measles:Scaling predictability, noise and determinism with the time-series SIR model. *Ecological Monographs* 72 *pp*185-202

[15] Gummer, A. W.(1991). Probability density functions of successive intervals of a non-homogeneous Poisson process under low-frequency conditions. *Biological Cybernetics* 65, *pp* 23-30.

[16] Han, J., Dong, G., Yin, Y. Efficient mining of partial periodic patterns in time series database. *Data Engineering, 1999. Proceedings.*, 15th International Conference, *pp* 106-115.

[17] Hannan, E. J.(1973). The estimation of frequency. *Journal of Applied Probability* 10, *pp* 510-519.

[18] Helmers, R., Mangku, I. W., Zitikis, R.(2003). Consistent estimation of the intensity function of a cyclic Poisson process. *Journal of Multivariate Analysis* 84, *pp*19-39.

[19] Helmers, R., Mangku, I. W., Zitikis, R.(2005). Statistical properties of a kernel-type estimator of the intensity function of a cyclic Poisson process. *Journal of Multivariate Analysis* 92, *pp*1-23.

[20] Ihler, A., Hutchins, J., Smyth, P. (2006). Adaptive event detection with time-varying Poisson processes. *KDD'06* August, *pp*20-23

[21] Kuhl, M. E., Wilson, J. R., Johnson, M. A. (1997). Estimating and stimulating Poisson Processes having trends or multiple periodicities. *IIE Transactions* 29:3, *pp* 201-211.

[22] Kuhl, M. E., Wilson, J. R., Damerdji, H. (1997). Estimating and stimulating Poisson Processes with trends or asymmetric cyclic effects. *Proceedings of the 1997 Winter Simulation Conference.*

[23] Kuhl, M. E., Wilson, J. R., Damerdji, H. (1998). Least-squares estimation of non-homogeneous Poisson processes. *Proceedings of the 1998 Winter Simulation Conference.*

[24] McCullagh, P., Nelder, J. A. - Generalized linear models - 2nd ed. London : Chapman and Hall, 1989.

[25] Mitchell, B. R. - British Historical Statistics. Cambridge University Press, 1988

[26] Peligrad, M., Utev, S. (1997). Central limit theorem for linear processes. *The Annals of Probability* 25, *pp* 443-456

[27] Priestley, M. B. - Spectral analysis and time series. London : Academic Press, 1981.

[28] Vere-Jones, D. (1982). On the estimation of frequency in point-process data. *J. of Appl. Prob.* 19A, *pp* 383-394.

[29] Willis, D. M. (1964). The statistics of a particular non-homogeneous Poisson process *Biometrika* 51, *pp* 399-404

[30] Wong, W. H. (1986). Theory of partial Likelihood. *Annals of Statistics* 14, *pp* 88-123.

[31] Xia, Y.Grenfell, B.T., Bjørnstad, O.N., (2004). Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics *The American Naturalist* 164, No.2 *pp*267-281

[32] Zeger, S. L.(1988). A regression model for time series of counts. *Biometrika* 75, *pp* 621-9.

[33] Zeger, S. L., Qaqish, B.(1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* 44, *pp* 1019-1031.