

Inference with the Whittle likelihood: a tractable approach using estimating functions

Joao Jesus and Richard E. Chandler

Department of Statistical Science

University College London, Gower Street, London WC1E 6BT, UK

October 14, 2016

Abstract

The theoretical properties of the Whittle likelihood have been studied extensively for many different types of process. In applications however, the utility of the approach is limited by the fact that the asymptotic sampling distribution of the estimator typically depends on third- and fourth-order properties of the process that may be difficult to obtain. In this paper, we show how the methodology can be embedded in the standard framework of estimating functions, which allows the asymptotic distribution to be estimated empirically without calculating higher-order spectra. We also demonstrate that some aspects of the inference, such as the calculation of confidence regions for the entire parameter vector, can be inaccurate but that a small adjustment, designed for application in situations where a misspecified likelihood is used for inference, can lead to marked improvements.

1 Introduction

Since its introduction by Whittle (1953) as a computationally efficient way to approximate the likelihood of a stationary Gaussian process, the ‘Whittle’ or ‘spectral’ likelihood function has at times attracted considerable interest. In the mainstream of modern applied time series analysis, it is no longer widely used: modern computing power, along with efficient algorithms for computing the exact likelihood of a Gaussian process (summarised, for example, in Chandler and Scott 2011, Section 5.1.5), mean that the use of such approximations is no longer necessary for ‘standard’ models. However, when working with processes for which an exact likelihood is hard to compute or is not analytically tractable but where the second-order spectral density is easy to compute, Whittle estimation remains a useful addition to the time series analyst’s toolkit. Indeed, over the last decade or so there has been a resurgence of interest in the methodology, when using nonstandard models in application areas such as hydrology (Montanari and Toth, 2007), biomedicine (Krafty and Hall, 2013) and oceanography (Sykulski et al., 2015); although

it is becoming accepted that the resulting estimators can be inefficient in non-Gaussian settings (Contreras-Cristán et al., 2006).

The properties of Whittle estimation have been studied extensively. Thus, Hannan (1973) proved the consistency and asymptotic normality of the Whittle estimator for the class of linear processes, the results subsequently being extended by Rice (1979). Fox and Taqqu (1986) and Giraitis and Taqqu (1999) studied its properties in the presence of long-range dependence, and showed that both for long-memory Gaussian processes and non-linear processes the estimator is consistent but does not necessarily have a limiting multivariate normal distribution after standardisation. More recently, there has been substantial work in applying Whittle likelihood principles to particular classes of nonlinear processes such as ARCH (Giraitis and Robinson, 2001); and Dahlhaus (2000, 2009) examines extensions to locally stationary processes, both for the univariate and multi-variate cases. The approach has also been applied to spatial processes (e.g. Fuentes 2002). These examples are not intended to be exhaustive, but rather to give a flavour of the range of processes for which Whittle estimation can be useful.

Although the Whittle estimator has been studied extensively from a theoretical perspective, a difficulty is that its asymptotic covariance matrix typically depends on the fourth-order spectral density of the process being studied (Robinson, 1978). This is a major drawback in practical applications, because expressions for fourth-order spectra are unavailable for many processes of interest (Chandler, 1997; Giraitis and Robinson, 2001). Perhaps as a result of this, there are few published applications of Whittle estimation in which standard errors are given for the parameter estimates, or in which formal model comparisons (e.g. based on changes in the log Whittle likelihood) are carried out. This is clearly undesirable.

Our aim in the present paper is to highlight a straightforward solution to this problem, with empirical estimation of the required asymptotic covariance matrix so that there is no need to evaluate fourth-order spectra explicitly: this is done by casting Whittle estimation in the general framework of estimating functions. A similar approach was taken by Heyde (1997), who focused on stationary processes with zero mean. Indeed, much of the literature on the topic is concerned with zero-mean processes: the original justification for this in Whittle (1953) was that ‘corrections for the mean do not affect the different formulae’, but in fact this is not necessarily the case when working with models for which both the first- and second-order properties are functions of the same parameter(s). Perhaps the simplest example of such a process is a sequence of independent Poisson variates in which the variance and mean are equal (to λ , say): in this situation the sample mean is sufficient for λ , but a ‘standard’ application of Whittle estimation would subtract the sample mean and then estimate λ using what is effectively the variance of the observations. Obviously, nobody would seriously contemplate Whittle estimation in this situation; but our own interest in the topic stems from work with stochastic models based on Poisson processes, for which exact maximum likelihood is not feasible and where the sample mean may carry considerable information about the parameter(s) of interest: an example is given in Section 2 below. One of our subsidiary contributions, therefore, is to extend the existing theory of Whittle estimation to processes with non-zero mean. Of more general interest perhaps, we also show via simulations that standard estimating function asymptotics can perform poorly in this context for some aspects of inference in finite

samples; and we demonstrate how recent developments, aimed at adjusting likelihood-based inference in mis-specified models, can rectify this problem.

After introducing our motivating application in the next section, Section 3 briefly outlines the main ideas of Whittle estimation. Section 4 then reviews the relevant theory of estimating functions. Section 5 demonstrates that the Whittle estimator does indeed satisfy the requirements of this theory. Finite-sample performance is assessed in Section 6, using simulations of the model from Section 2. Section 7 concludes.

2 Motivation: inference for a stochastic rainfall model

To provide some context for the subsequent contributions of this paper, in this section we give an example of the type of model that motivated our own interest in the topic. It is a stochastic model for rainfall time series, and is very different from the kinds of model for which Whittle estimation is typically used. Although the specific model considered here is rather simplified, more complex models of this type are widely used by hydrologists to simulate artificial rainfall time series (see Wheater et al. 2005 for a review), for purposes such as the assessment of flood risk and the impacts of climate change.

We consider the simplest of a class of continuous-time rainfall models introduced by Rodriguez-Iturbe et al. (1987). In this model, the rainfall intensity at any time is considered to be a sum of contributions from rain ‘cells’ that are ‘active’ at that time. Each cell is initiated by an event of a stationary Poisson process $N(t)$ of rate λ , and has a random duration during which it deposits rain with a constant intensity (also random) so that its temporal profile is rectangular. The rainfall intensity at time t is thus

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u) ,$$

where $X_{\tau}(r)$ is the random intensity of a cell that arrived at time τ , observed r time units later. In the simplest case, the intensities of all cells are taken to be independent exponential variables with mean μ_X ; the durations are taken to be independent exponential variables with mean μ_L ; and the intensities and durations are taken to be mutually independent. The model thus has three parameters λ , μ_X and μ_L .

In practice of course, rainfall intensities cannot be recorded in continuous time. Observations typically consist of time series at a specific temporal resolution, often daily or hourly, representing rainfall accumulations within the corresponding intervals. Denoting by Y_t the rainfall accumulation within the t th interval, we therefore have

$$Y_t = \int_{(t-1)\Delta}^{t\Delta} Y(u) du , \tag{1}$$

where Δ is the duration of the aggregation interval.

Parameter estimation and inference for such a model is difficult. It is not feasible to evaluate the exact log-likelihood on the basis of aggregated data, although Northrop (2006) has obtained a marginal likelihood for λ and μ_L based on the binary sequence of

‘dry’ and ‘wet’ intervals (i.e. intervals in which Y_t is either zero or non-zero). Usually, such models are fitted using a generalised method of moments that aims to match, as closely as possible, the observed values of selected summary statistics. A difficulty with this approach is that the choice of summary statistics is potentially arbitrary. We show below that Whittle estimation offers an appealing alternative, and that it has some justification despite the highly non-Gaussian nature of the process. We will also see that the parameters λ and μ_X are not separately identifiable using standard Whittle estimation, and that the problem can be resolved only by accounting explicitly for the non-zero mean of the process. A related point is that the parameters all combine to influence both the first- and second-order properties of the process: this contrasts with most models where Whittle estimation has been applied, where different parameters control the first- and second-order structure. Identifiability problems aside therefore, for this class of model it is not appropriate simply to subtract the mean and then carry out standard Whittle estimation for the parameters, because the mean itself carries relevant information. In particular, the calculation of standard errors for the parameter estimates requires that the first- and second-order properties are considered simultaneously.

We return to this class of models for our simulation study in Section 6.

3 The Whittle likelihood

Let (Y_t) be a stationary stochastic process corresponding to some model with parameter vector $\boldsymbol{\theta} \in \Theta$; and consider conducting inference about $\boldsymbol{\theta}$ on the basis of observations Y_1, \dots, Y_n . If the process is Gaussian, then the log-likelihood for $\boldsymbol{\theta}$ involves both the determinant and inverse of the covariance matrix of $\mathbf{Y} = (Y_1 \dots Y_n)'$ under the model. The main contribution of Whittle (1953) was to provide an approximation to the Gaussian log-likelihood that avoids the costly evaluation of this determinant and inverse.

The same approximate log-likelihood can be obtained using a different argument, without assuming that the process itself is Gaussian but rather appealing to a central limit theorem on the sample Fourier coefficients (Subba Rao and Chandler, 1996; Chandler, 1997). This derivation provides some useful insights into the nature of the approximations involved, and we therefore sketch it informally here. Denote by $\mu = \mu(\boldsymbol{\theta})$ the mean of the process (Y_t) , by $c(r; \boldsymbol{\theta})$ its autocovariance function at lag r , by

$$h(\omega; \boldsymbol{\theta}) = \sum_{r=-\infty}^{\infty} c(r; \boldsymbol{\theta}) e^{-i\omega r} ,$$

its second-order spectral density at frequency $\omega \in [-\pi, \pi]$ and by $c_\kappa(r_1, \dots, r_{\kappa-1})$ the joint κ th cumulant of $Y_t, Y_{t+r_1}, \dots, Y_{t+r_{\kappa-1}}$. Further, for $p = 0, \dots, \lfloor n/2 \rfloor$ and $\omega_p = 2\pi p/n$, define the sample Fourier coefficients as

$$A_p = \sum_{t=1}^n Y_t \cos(\omega_p t) \quad \text{and} \quad B_p = \sum_{t=1}^n Y_t \sin(\omega_p t) , \quad (2)$$

and define the periodogram at frequency ω_p as

$$I(\omega_p) = \frac{1}{n} (A_p^2 + B_p^2) .$$

Note that the observations can be recovered from the Fourier coefficients by applying an inverse Fourier transform; hence the coefficients may be regarded as data in their own right, with no loss of information. Moreover, if the process has short-range dependence so that the joint cumulants of all orders $\kappa \geq 2$ are absolutely summable:

$$\sum_{r_1=-\infty}^{\infty} \cdots \sum_{r_{\kappa-1}=-\infty}^{\infty} |c_{\kappa}(r_1, \dots, r_{\kappa-1})| < \infty , \quad (3)$$

then in large samples the Fourier coefficients are approximately normally distributed and pairwise independent (Brillinger, 1975, Section 4.4):

$$\begin{aligned} A_0 &\sim N(n\mu, nh(0; \boldsymbol{\theta})) ; & A_p &\sim N\left(0, \frac{n}{2}h(\omega_p; \boldsymbol{\theta})\right) \quad (p \neq 0) ; \\ B_p &\equiv 0 \quad (\omega_p = 0, \pi) ; & \text{and } B_p &\sim N\left(0, \frac{n}{2}h(\omega_p; \boldsymbol{\theta})\right) \quad (\omega_p \neq 0, \pi) . \end{aligned} \quad (4)$$

If these results were to hold exactly for all of the Fourier coefficients, then a likelihood for $\boldsymbol{\theta}$ could be written as a product of the corresponding densities. Some algebra confirms that the corresponding log-likelihood is

$$\begin{aligned} \log L(\boldsymbol{\theta}; \mathbf{Y}) = & - \sum_{p=0}^{\lfloor n/2 \rfloor} \left[1 - \frac{1}{2} \delta_{p, n/2} \right] \left[\frac{I(\omega_p)}{h(\omega_p; \boldsymbol{\theta})} + \log h(\omega_p; \boldsymbol{\theta}) \right] \\ & - \left[\frac{1}{2} \log h(0; \boldsymbol{\theta}) + \frac{(A_0 - n\mu(\boldsymbol{\theta}))^2}{2h(0; \boldsymbol{\theta})} \right] , \end{aligned} \quad (5)$$

where $\delta_{i,j}$ is the Kronecker delta. The final term, and the $p = 0$ contribution to the first term, are contributions from the zero-frequency coefficient A_0 : if these terms are omitted then (5) reduces to $\sum_{p=1}^{\lfloor n/2 \rfloor} [I(\omega_p)/h(\omega_p; \boldsymbol{\theta}) + \log h(\omega_p; \boldsymbol{\theta})]$, which is the usual form of the Whittle log-likelihood (Hauser, 1998). The inclusion of the zero-frequency term, however, can be helpful in some situations as outlined in Section 1.

According to this derivation, the main sources of approximation in (5) relate to the marginal normality and independence of the Fourier coefficients. Although these approximations may be asymptotically negligible when considering the coefficients at a finite, fixed set of frequencies, in practice the number of frequencies grows with n : the increase in the number of approximation errors thus offsets the reduction in their size. For further discussion of this point, see Chandler (1997), who provides some suggestions for modifying the log-likelihood so that the approximation errors become negligible. These suggestions all involve discarding information, however, which is unsatisfactory. We take an alternative viewpoint in the present paper, therefore: we consider (5) as a log-likelihood based on a mis-specified joint distribution for the sample Fourier coefficients, and we study its properties using an estimating functions approach.

4 Estimating functions

If the derivatives $\partial h(\omega; \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $\partial \mu(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ exist for all $\omega \in [0, \pi]$ and $\boldsymbol{\theta} \in \Theta$, then (5) can be maximised by solving $\partial \log L/\partial \boldsymbol{\theta} = \mathbf{0}$. This is an example of an estimating equation, and the function $\partial \log L/\partial \boldsymbol{\theta}$ is an estimating function: a review and discussion of the relevant theory is provided in Jesus and Chandler (2011). We here summarise, without proof, the main definitions and results presented in that paper.

Consider carrying out inference for a parameter vector $\boldsymbol{\theta} \in \Theta$ on the basis of a vector \mathbf{Y}_n of n random variables; and suppose that $\boldsymbol{\theta}$ is estimated by solving an equation

$$\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}_n) = \mathbf{0} , \quad (6)$$

which is assumed to have a unique root. Many common estimation methods can be cast in this framework, notably those where the estimating function $\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ is the gradient of some objective function. For example, the normal equations in least-squares estimation have the form (6), as do the score equations in maximum likelihood estimation. However, the framework also covers situations, such as method of moments estimation, where the estimating equations are obtained directly without reference to an optimisation problem.

The sequence $\{\mathbf{g}_n(\cdot; \cdot) : n \in \mathbb{N}\}$ of estimating functions is assumed to satisfy:

Assumption 1: There exists a sequence $(\boldsymbol{\eta}_n)$ of $p \times p$ matrices, independent of $\boldsymbol{\theta}$ and such that as $n \rightarrow \infty$, $\boldsymbol{\eta}_n \mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ converges uniformly in probability to a non-random function, $\mathbf{g}_\infty(\boldsymbol{\theta})$ say, such that (i) the equation $\mathbf{g}_\infty(\boldsymbol{\theta}) = \mathbf{0}$ has a unique root at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is an interior point of Θ (ii) $\mathbf{g}_\infty(\cdot)$ is bounded away from zero except in the neighbourhood of $\boldsymbol{\theta}_0$.

Assumption 2: The $p \times p$ matrices $\{\mathbf{G}_n(\boldsymbol{\theta}) = \partial \mathbf{g}_n/\partial \boldsymbol{\theta}\}$ exist, and each of them is invertible. Furthermore, there exist sequences $(\boldsymbol{\gamma}_n)$ and $(\boldsymbol{\delta}_n)$ of invertible $p \times p$ matrices that do not depend on $\boldsymbol{\theta}$ and are such that:

1. At $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the covariance matrix of $\tilde{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{Y}_n) = \boldsymbol{\gamma}_n \mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ exists and converges to a limiting matrix, $\tilde{\boldsymbol{\Sigma}}$ say, as $n \rightarrow \infty$. We note that the limit of a sequence of positive definite matrices is itself positive semi-definite, and hence there is no need for the additional requirement in Jesus and Chandler (2011) that $\tilde{\boldsymbol{\Sigma}}$ is itself a valid covariance matrix.
2. Defining $\tilde{\mathbf{G}}_n(\boldsymbol{\theta}) = \partial \tilde{\mathbf{g}}_n/\partial \boldsymbol{\theta}$, there exists $c > 0$ such that, for all $\boldsymbol{\theta}$ with $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < c$, the matrix $\tilde{\mathbf{G}}_n(\boldsymbol{\theta})\boldsymbol{\delta}_n$ converges uniformly in probability to an invertible matrix $\mathbf{M}(\boldsymbol{\theta})$ with elements that are continuous functions of $\boldsymbol{\theta}$.

Under Assumption 1, as $n \rightarrow \infty$ the estimating equation (6) defines a unique estimator $\hat{\boldsymbol{\theta}}_n$ that converges in probability to $\boldsymbol{\theta}_0$. Moreover, under Assumptions 1 and 2 the covariance matrix of $\boldsymbol{\delta}_n^{-1} \hat{\boldsymbol{\theta}}_n$ converges to $\mathbf{C} = \mathbf{M}_0^{-1} \tilde{\boldsymbol{\Sigma}} (\mathbf{M}_0^{-1})'$ where $\mathbf{M}_0 = \mathbf{M}(\boldsymbol{\theta}_0)$. If, in addition, $\tilde{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ has a limiting multivariate normal (MVN) distribution then the distribution of $\boldsymbol{\delta}_n^{-1} \hat{\boldsymbol{\theta}}_n$ is itself multivariate normal in the limit. Operationally, this result says that the distribution of $\hat{\boldsymbol{\theta}}_n$ is approximately MVN $\left(\boldsymbol{\theta}_0, \mathbf{G}_0^{-1} \boldsymbol{\Sigma}_n [\mathbf{G}_0^{-1}]'\right)$ in large samples,

where Σ_n is the covariance matrix of the unnormalised estimating function $\mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y}_n)$ and $\mathbf{G}_0 = \text{E}[\mathbf{G}_n(\boldsymbol{\theta}_0)]$, which is assumed to exist. This provides a means of testing hypotheses and calculating confidence intervals about elements of $\boldsymbol{\theta}$.

4.1 Estimating functions as gradient vectors

In the present paper, the estimating function of interest is the gradient vector of the approximate log-likelihood (5). In generic situations where an estimating function is obtained by differentiating some objective function $Q_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ say, inference about $\boldsymbol{\theta}$ can also be based on the objective function itself: the relevant theory comes from the properties of the estimating functions, however (Jesus and Chandler, 2011). We outline the main results here, for situations in which estimators are obtained by maximising (rather than minimising) $Q_n(\boldsymbol{\theta}; \mathbf{Y}_n)$.

Consider testing hypotheses about, or calculating confidence regions for, some linear combination $\Xi\boldsymbol{\theta}$ of the parameters, where Ξ is a matrix of appropriate dimension. Often one may be interested in testing a hypothesis of the form $H_0 : \Xi\boldsymbol{\theta} = \boldsymbol{\xi}_0$. In classical time series analysis for example, such a test could be used to compare autoregressive models of different orders by fitting a high-order model and testing the hypothesis that one or more coefficients are zero: in this case, Ξ is a matrix of ones and zeroes picking out the coefficients to be tested, and $\boldsymbol{\xi}_0$ is a vector of zeroes. Similarly, for the kind of stochastic rainfall model described in Section 2 we might ask whether the exponential distribution provides an adequate representation of rain cell intensity and duration: this could be done by fitting a more complex model in which both quantities have gamma distributions, and then testing the hypothesis that the shape parameters of these distributions are both equal to unity. Alternatively, it may be of interest to calculate confidence intervals for the shape parameters, to gain insights into the kinds of variation in rain cell properties that are consistent with the data.

For the hypothesis testing problem, let $\tilde{\boldsymbol{\theta}}_n$ be the value of $\boldsymbol{\theta}$ maximising $Q_n(\boldsymbol{\theta}; \mathbf{Y}_n)$ subject to the restriction $\Xi\boldsymbol{\theta} = \boldsymbol{\xi}_0$. Under Assumptions 1 and 2, and if $\text{E}[\mathbf{G}_n(\boldsymbol{\theta}_0)] = \mathbf{G}_0$ exists, then under H_0 the quantity $2 \left[Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}_n) - Q_n(\tilde{\boldsymbol{\theta}}_n; \mathbf{Y}_n) \right]$ has approximately the same distribution as $\mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z}$, where $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, -\Xi\mathbf{G}_0^{-1}\Sigma_n[\Xi\mathbf{G}_0^{-1}]')$ and $\mathbf{A} = -\Xi\mathbf{G}_0^{-1}\Xi'$: the negative signs here arise because the expected Hessian matrix \mathbf{G}_0 is negative definite at a maximum of $Q_n(\cdot; \cdot)$. This generalises the result in Kent (1982) on the distribution of likelihood ratios in mis-specified models. In principle, it yields a straightforward procedure for testing the hypothesis of interest: accept if

$$2 \left[Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}_n) - Q_n(\tilde{\boldsymbol{\theta}}_n; \mathbf{Y}_n) \right] \tag{7}$$

is less than the appropriate percentile of the distribution of $\mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z}$ and reject otherwise. Similarly, a confidence region or interval for $\Xi\boldsymbol{\theta}$ is the set of values for which (7) is less than the corresponding percentile.

Unfortunately, distributions of quadratic forms such as $\mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z}$ are intractable in general: they are often approximated using scaled and shifted chi-squared distributions. Jesus and Chandler (2011) provide further details, and also highlight situations where the

required distribution reduces to the usual chi-squared result for likelihood ratio testing in a correctly specified model. One such situation arises when $\mathbf{G}_0 = -\boldsymbol{\Sigma}_n$ (i.e. the expected negative Hessian of the objective function is equal to the covariance matrix of its gradient vector), in which case the large-sample covariance matrix of the estimator $\hat{\boldsymbol{\theta}}_n$ is itself equal to $\boldsymbol{\Sigma}_n^{-1} = -\mathbf{G}_0^{-1}$. In classical likelihood inference, this equality is the second Bartlett identity (Severini, 2000, Section 3.5). If it holds, then the distribution of \mathbf{Z} above reduces to $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, -\boldsymbol{\Xi}\mathbf{G}_0^{-1}\boldsymbol{\Xi}') = \text{MVN}(\mathbf{0}, \mathbf{A})$; and the distribution of $\mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z}$ is exactly chi-squared with q degrees of freedom, where q is the rank of $\boldsymbol{\Xi}$.

When the second Bartlett identity does not hold, inference based on (7) has a further drawback: the objective function threshold required to defined a confidence region for a rank- q transform of $\boldsymbol{\theta}$ varies depending on the matrix $\boldsymbol{\Xi}$. In particular, different thresholds are required to define confidence intervals for different elements of $\boldsymbol{\theta}$: thus, a 95% confidence interval for θ_1 (for which $\boldsymbol{\Xi} = (1 \ 0 \ \dots \ 0)$) might consist of all values for which (7) is less than some threshold τ_1 say, whereas a 95% confidence interval for θ_2 (for which $\boldsymbol{\Xi} = (0 \ 1 \ \dots \ 0)$) might consist of all values for which (7) is less than a different threshold τ_2 . This is unintuitive. An alternative is to work with an adjusted objective function that satisfies the second Bartlett identity, such as

$$Q_n^{(ADJ)}(\boldsymbol{\theta}; \mathbf{Y}) = Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}) - \left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \mathbf{G}_0 \boldsymbol{\Sigma}_n^{-1} \mathbf{G}_0 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) \right\} \frac{Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}) - Q_n(\boldsymbol{\theta}; \mathbf{Y})}{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \mathbf{G}_0 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)} \quad (8)$$

which is the ‘vertically adjusted’ objective function defined at equation (25) of Chandler and Bate (2007). When $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$, (8) yields $Q_n^{(ADJ)}(\boldsymbol{\theta}; \mathbf{Y}) = Q_n(\boldsymbol{\theta}; \mathbf{Y})$: the original and adjusted objective functions share the same maximum. The denominator in the final term is the second-order Taylor expansion of $Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}) - Q_n(\boldsymbol{\theta}; \mathbf{Y})$: thus, to second order, the final ratio in (8) is unity, the final term in its entirety is $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \mathbf{G}_0 \boldsymbol{\Sigma}_n^{-1} \mathbf{G}_0 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)$, and $Q_n^{(ADJ)}(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}) - Q_n^{(ADJ)}(\boldsymbol{\theta}; \mathbf{Y})$ is $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \mathbf{G}_0 \boldsymbol{\Sigma}_n^{-1} \mathbf{G}_0 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)$. The Hessian of this quadratic form (i.e. the matrix $\mathbf{G}_0 \boldsymbol{\Sigma}_n^{-1} \mathbf{G}_0$) is the inverse of the large-sample covariance matrix of $\hat{\boldsymbol{\theta}}_n$ as given earlier: thus $Q_n^{(ADJ)}(\boldsymbol{\theta}; \mathbf{Y})$ does indeed satisfy the second Bartlett identity as claimed.

Combining (8) with equations (18) and (20) of Chandler and Bate (2007), it is easy to show that a test of $H_0 : \boldsymbol{\Xi}\boldsymbol{\theta} = \boldsymbol{\xi}_0$ can be carried out by comparing

$$\frac{2 \left(\boldsymbol{\Xi} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\xi}_0 \right)' \left[\boldsymbol{\Xi} \mathbf{G}_0^{-1} \boldsymbol{\Sigma}_n \left(\boldsymbol{\Xi} \mathbf{G}_0^{-1} \right)' \right]^{-1} \left(\boldsymbol{\Xi} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\xi}_0 \right)}{\left(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n \right)' \mathbf{G}_0 \left(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n \right)'} \left[Q_n(\hat{\boldsymbol{\theta}}_n; \mathbf{Y}) - Q_n(\tilde{\boldsymbol{\theta}}_n; \mathbf{Y}) \right] \quad (9)$$

with the appropriate percentage point of a χ_q^2 distribution. Confidence regions can be constructed similarly.

To apply the results outlined above in practice, it is necessary to estimate the matrices \mathbf{G}_0 and $\boldsymbol{\Sigma}_n$. Suggestions for achieving this in the context of Whittle estimation are provided at the end of the next section.

5 Asymptotic distribution of the Whittle estimator

This section is the core of the paper, showing that for the Whittle log-likelihood defined by (5), the estimating function $\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}) = \partial \log L(\boldsymbol{\theta}; \mathbf{Y}) / \partial \boldsymbol{\theta}$ satisfies Assumptions 1 and 2 in Section 4. Having done this, the generic theory above follows immediately. We reiterate that the properties of the usual Whittle estimator have been established previously by Heyde (1997) using an approach that we regard as similar in spirit, although this work does not appear to have been widely noticed in the time series community and does not consider the contribution from frequency $\omega_0 = 0$. In fact, one of the key steps in our approach is to work with the centred process

$$Y_t^* = Y_t - \mu(\boldsymbol{\theta}_0) = Y_t - \mu_0 \text{ say,} \quad (10)$$

where $\boldsymbol{\theta}_0$ now is the true value of $\boldsymbol{\theta}$ (i.e. the value corresponding to the process that generated the data). This enables us to treat the zero-frequency contributions separately from the remainder.

We make no pretence that the asymptotic results in this section will be ‘news’: indeed, we borrow heavily from the established literature. Our aim is rather to frame the results in a form that is suitable for routine practical application. Before continuing, it is helpful to collate some results that will be used in the subsequent derivations.

5.1 Some useful convergence results

The first result in this section provides a central limit theorem for a weighted sum of periodogram ordinates.

Result 1: Let (Y_t) be a zero-mean stationary process with second-order spectral density $h(\omega)$ satisfying the following conditions:

1. $0 < h(\omega) < \infty$ for all $\omega \in [-\pi, \pi]$.
2. $h(\omega)$ satisfies a Hölder(ξ) condition for some $\xi > 1/2$ i.e. there exists a constant K such that for all $\omega_1, \omega_2 \in [-\pi, \pi]$, $|h(\omega_2) - h(\omega_1)| \leq K |\omega_2 - \omega_1|^\xi$.

Moreover, suppose that the fourth-order spectral density of (Y_t) is square-integrable, and that the cumulant condition (3) is satisfied. Then, if $\boldsymbol{\phi}(\omega)$ is a vector-valued function of $\omega \in [-\pi, \pi]$, each element of which is an even function that is Hölder(ξ) for some $\xi > 1/2$, the quantity

$$n^{-1/2} \sum_{p=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} \boldsymbol{\phi}(\omega_p) [I(\omega_p) - h(\omega_p)] \quad (11)$$

has a limiting multivariate normal distribution as $n \rightarrow \infty$, with mean vector zero and finite covariance matrix. ■

This result is Theorem 4 of Robinson (1978), rephrased slightly for current purposes. In fact, Robinson’s statement omits the zero-frequency term in the sum; but an inspection

of the proof confirms that the result is unaffected by this. Robinson also summarises some alternatives to the cumulant condition (3), and the same or similar results regarding limiting multivariate normality have been provided by several authors in other situations, including long-memory processes for which (3) does not hold: see, for example, Heyde and Gay (1993), Giraitis et al. (2001) and Giraitis and Koul (2013). Bardet et al. (2008) provide a good review of the area, highlighting a variety of different conditions that can be used to establish this kind of result.

In Result 1, the required Hölder condition is trivially satisfied if $h(\omega)$ has a finite derivative for all $\omega \in [-\pi, \pi]$: for in this case, by the mean value theorem, $h(\omega_2) - h(\omega_1) = (\omega_2 - \omega_1)\partial h/\partial\omega|_{\omega=\omega^\dagger}$ for some $\omega^\dagger \in [\omega_1, \omega_2]$, whence $|h(\omega_2) - h(\omega_1)| \leq |\omega_2 - \omega_1| \sup_{\omega \in [-\pi, \pi]} |\partial h/\partial\omega|$ and $h(\omega)$ is Hölder(1).

The second result is Lemma 1 of Brillinger and Rosenblatt (1967) and provides an approximation for the cumulants of the discrete Fourier transform of a stationary process:

Result 2: Let (Y_t) be a stationary process, with κ th-order spectral density defined on $[-\pi, \pi]^{\kappa-1}$ as

$$h^{(\kappa)}(\omega_1, \dots, \omega_{\kappa-1}) = \sum_{r_1=-\infty}^{\infty} \dots \sum_{r_{\kappa-1}=-\infty}^{\infty} c_\kappa(r_1, \dots, r_{\kappa-1}) e^{-i \sum_{j=1}^{\kappa-1} \omega_j r_j} \quad (12)$$

where $c_\kappa(r_1, \dots, r_{\kappa-1})$ is the joint κ th-order cumulant of $Y_t, Y_{t+r_1}, \dots, Y_{t+r_{\kappa-1}}$ as before. Suppose that for each $j \in \{1, \dots, \kappa-1\}$,

$$\sum_{r_1=-\infty}^{\infty} \dots \sum_{r_{\kappa-1}=-\infty}^{\infty} |r_j c_\kappa(r_1, \dots, r_{\kappa-1})| < \infty, \quad (13)$$

which is a slightly stronger version of the short-range dependence condition (3); and define $J_p = \sum_{t=1}^n Y_t e^{-i\omega_p t} = A_p + iB_p$, with A_p and B_p as in (2). Then the joint κ th-order cumulants of the $\{J_p\}$ satisfy

$$c_\kappa(J_{p_1}, \dots, J_{p_\kappa}) = (2\pi)^{\kappa-1} \Delta^{(n)} \left(\sum_{j=1}^{\kappa} p_j \right) h^{(\kappa)}(w_{p_1}, \dots, w_{p_{\kappa-1}}) + O(1), \quad (14)$$

where $\Delta^{(n)}(p) = n$ if $p = 0 \pmod{n}$ and zero otherwise. ■

We are now in a position to derive the estimating functions defined by the Whittle likelihood, and to show that under some constraints on the process (Y_t) the Whittle estimating function satisfies the conditions of Section 4.

5.2 The spectral score function

The gradient vector of the Whittle log-likelihood (5) will be referred to as the vector of spectral scores, with i th component

$$g_i(\boldsymbol{\theta}; \mathbf{Y}) = \frac{\partial \log L}{\partial \theta_i}$$

$$\begin{aligned}
&= - \sum_{p=1}^{\lfloor n/2 \rfloor} \left[1 - \frac{1}{2} \delta_{p,n/2} \right] \left[\frac{\partial h(\omega_p; \boldsymbol{\theta})}{\partial \theta_i} \frac{1}{h(\omega_p; \boldsymbol{\theta})^2} (h(\omega_p; \boldsymbol{\theta}) - I(\omega_p)) \right] \\
&\quad - \frac{1}{2} \left[\frac{\partial h(0; \boldsymbol{\theta})}{\partial \theta_i} \frac{1}{h(0; \boldsymbol{\theta})} - \frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_i} \frac{2(A_0 - n\mu(\boldsymbol{\theta}))}{h(0; \boldsymbol{\theta})} - \frac{\partial h(0; \boldsymbol{\theta})}{\partial \theta_i} \frac{(A_0 - n\mu(\boldsymbol{\theta}))^2}{nh(0; \boldsymbol{\theta})^2} \right]. \quad (15)
\end{aligned}$$

The first line here is the contributions from all frequencies except zero; the last line is the contribution from the zero-frequency term in (5). By dealing with the terms separately, we can exploit existing results for the situation in which the zero-frequency term is excluded.

For the subsequent development, it will be helpful to express (15) in terms of the periodogram of the centred process (10). Let A_0^* denote the zero-frequency Fourier coefficient for the sequence Y_1^*, \dots, Y_n^* , and let $I^*(\omega_p)$ denote the periodogram at frequency ω_p . Then, from the definitions (2) and (3) applied to the centred process, we have $A_0^* = \sum Y_t^* = A_0 - n\mu_0$, $I^*(0) = A_0^{*2}/n = I(0) - 2A_0\mu_0 + n\mu_0^2$ and $I^*(\omega_p) = I(\omega_p)$ for $p \neq 0$. Note that only the Fourier coefficient at zero frequency is affected by the centring. Of course, the quantities A_0^* and $I^*(0)$ cannot be calculated in practice since they require knowledge of μ_0 , which itself depends on the unknown parameter $\boldsymbol{\theta}_0$; nonetheless, for theoretical purposes it is convenient to work with these quantities.

Rearranging the expressions in the previous paragraph, we find that $A_0 = A_0^* + n\mu_0$ and that $I(0) = I^*(0) + 2A_0^*\mu_0 - n\mu_0^2$. Plugging these in (15), and defining

$$a_i(\omega; \boldsymbol{\theta}) = \frac{\partial h(\omega; \boldsymbol{\theta})}{\partial \theta_i} \frac{1}{h(\omega; \boldsymbol{\theta})^2} = -\frac{\partial}{\partial \theta_i} h^{-1}(\omega; \boldsymbol{\theta}), \quad (16)$$

we obtain

$$\begin{aligned}
g_i(\boldsymbol{\theta}; \mathbf{Y}) &= - \sum_{p=1}^{\lfloor n/2 \rfloor} \left[1 - \frac{1}{2} \delta_{p,n/2} \right] [a_i(\omega_p; \boldsymbol{\theta}) (h(\omega_p; \boldsymbol{\theta}) - I^*(\omega_p))] \\
&\quad - \frac{1}{2} \left[a_i(0; \boldsymbol{\theta}) h(0; \boldsymbol{\theta}) - \frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_i} \frac{2(A_0^* + n\mu_0 - n\mu(\boldsymbol{\theta}))}{h(0; \boldsymbol{\theta})} - a_i(0; \boldsymbol{\theta}) \frac{(A_0^* + n\mu_0 - n\mu(\boldsymbol{\theta}))^2}{n} \right].
\end{aligned}$$

The summands in the first line of this expression are all even functions of ω , and therefore the sum can be written as

$$-\frac{1}{2} \sum_{p=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} a_i(\omega_p; \boldsymbol{\theta}) (h(\omega_p; \boldsymbol{\theta}) - I^*(\omega_p)) + \frac{1}{2} a_i(0; \boldsymbol{\theta}) (h(0; \boldsymbol{\theta}) - I^*(0)) ,$$

the final term arising because there was no $p = 0$ term in the original sum. Some further algebraic manipulation, using the fact that $I^*(0) = A_0^{*2}/n$, leads eventually to the result

$$\begin{aligned}
g_i(\boldsymbol{\theta}; \mathbf{Y}) &= -\frac{1}{2} \sum_{p=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} a_i(\omega_p; \boldsymbol{\theta}) [h(\omega_p; \boldsymbol{\theta}) - I^*(\omega_p)] + \frac{A_0^* + n[\mu_0 - \mu(\boldsymbol{\theta})]}{h(0; \boldsymbol{\theta})} \frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_i} \\
&\quad + \frac{a_i(0; \boldsymbol{\theta})}{2} [n(\mu_0 - \mu(\boldsymbol{\theta}))^2 + 2A_0^*(\mu_0 - \mu(\boldsymbol{\theta}))]. \quad (17)
\end{aligned}$$

This expresses the spectral scores as a function of the periodogram and sample mean of a zero-mean process, whence existing results from the literature can be applied.

5.3 Assumptions on (Y_t)

For the remainder of the paper, we assume that (Y_t) is a stationary process such that the centred process (Y_t^*) , defined in (10), satisfies the requirements of Result 1. Note that the zero-mean process (Y_t^*) has the same second-order spectral density as (Y_t) .

As well as the requirements of Result 1, we impose several additional conditions on the model for (Y_t) , expressed in terms of its properties at individual values of $\boldsymbol{\theta} \in \Theta$. These conditions are reasonably innocuous: they mostly relate to existence, smoothness, differentiability and finiteness of relevant quantities. They are as follows:

C1 If $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ then at least one of the following conditions holds:

- $h(\omega; \boldsymbol{\theta}_1)$ and $h(\omega; \boldsymbol{\theta}_2)$ differ on a set of non-zero measure in $[-\pi, \pi]$.
- $\mu(\boldsymbol{\theta}_1) \neq \mu(\boldsymbol{\theta}_2)$.

C2 The $\boldsymbol{\theta}$ -derivatives of all quantities appearing in (17) exist and are finite.

C3 The quantities $h(\omega; \boldsymbol{\theta})$ and $\{\partial h / \partial \theta_i\}$ have bounded ω -derivatives of order up to 2, on $[-\pi, \pi]$.

C4 The third-order spectral density, $h^{(3)}(\cdot, \cdot; \boldsymbol{\theta})$ say, is such that $h^{(3)}(0, \omega; \boldsymbol{\theta}_0)$ is twice differentiable in ω and $\partial^2 h^{(3)}(0, \omega; \boldsymbol{\theta}_0) / \partial \omega^2$ is finite for all $\omega \in [-\pi, \pi]$.

Condition C1 ensures that the parameters are identifiable, in the sense that different values of $\boldsymbol{\theta}$ lead to asymptotically distinct estimating function values. C2 is trivial: note that if the spectral density is bounded away from zero and infinity on $[-\pi, \pi]$ as required by Result 1, and if the mean $\mu(\boldsymbol{\theta})$ is finite, then finite derivatives of the various quantities in (17) are guaranteed if the elements of $\partial \mu / \partial \boldsymbol{\theta}$ and $\partial h(\omega_p; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ themselves exist and are finite. C3 ensures that the quantity $a_i(\omega; \boldsymbol{\theta}) h(\omega; \boldsymbol{\theta})$ has finite second derivatives on $[-\pi, \pi]$, with $a_i(\omega; \boldsymbol{\theta})$ as defined in (16). Finally, C4 is needed to handle the zero-frequency term when demonstrating convergence of the covariance matrix of the normalised estimating function as shown in Appendix A. From definition (12), it follows that $\partial^2 h^{(3)}(0, \omega; \boldsymbol{\theta}_0) / \partial \omega^2 = -\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} r_2^2 c_3(r_1, r_2) \exp(i\omega r_2)$ so that C4 will be satisfied if there is no periodic variation in $c_3(r_1, r_2)$ and if $\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} r_2^2 |c_3(r_1, r_2)| < \infty$. This holds trivially if the third-order cumulants are all zero, which is the case for Gaussian processes.

An immediate consequence of the above conditions is that each of the even, vector-valued functions $\{\partial a_i(\omega; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}\}$ exists and satisfy the requirements of the function $\phi(\cdot)$ in Result 1. This is because C3, together with the requirements of Result 1, guarantees that the ω -derivatives of $\partial a_i(\omega; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ are finite so that $\partial a_i(\omega; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ is itself Hölder(1). It follows that under these conditions, Result 1 can be applied to the first term of (17) after appropriate normalisation.

Conditions C1 to C4 exclude some important classes of processes: for example, long-memory processes in which the spectral density is infinite at $\omega = 0$. However, they do include a wide variety of non-Gaussian processes for which Whittle estimation has not been widely used: the rainfall model of Section 2 is an example.

Finally, we assume that the parameter space Θ is compact. Although this is slightly restrictive, it is a ubiquitous device in the literature for establishing uniform convergence as required to obtain consistency results. For a more detailed discussion including alternative possibilities, see Jesus and Chandler (2011) and citations therein. Moreover, for many models this assumption is a convenient way to guarantee the finiteness or boundedness of quantities appearing in conditions C2 and C3 above.

5.4 Conditions for consistency

The Whittle estimator is consistent if the root of the estimating equation $\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y}) = \mathbf{0}$ converges in probability to $\boldsymbol{\theta}_0$, the true parameter value. From Section 4, this occurs if a normalised version of $\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y})$ converges to a deterministic function with a unique root at $\boldsymbol{\theta}_0$ and which is bounded away from zero away from $\boldsymbol{\theta}_0$. We now show that under the conditions set out in Section 5.3, this is achieved by choosing the normalizing matrix (in the notation of Assumption 1) $\boldsymbol{\eta}_n = n^{-1}\mathbf{I}$, where \mathbf{I} is the identity matrix. With this choice of $\boldsymbol{\eta}_n$, the i th element of $\boldsymbol{\eta}_n\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y})$ is just $n^{-1}g_i(\boldsymbol{\theta}; \mathbf{Y})$, with $g_i(\boldsymbol{\theta}; \mathbf{Y})$ given by (17). Notice first that A_0^*/n is the sample mean of the centred observations and, under the stated assumptions, this tends to zero as $n \rightarrow \infty$. Moreover, from Result 1 applied to the centred process (Y_t^*) , the quantity $n^{-1}\sum_p a_i(\omega_p; \boldsymbol{\theta}) [I^*(\omega_p) - h(\omega_p; \boldsymbol{\theta}_0)]$ converges in probability to zero as $n \rightarrow \infty$ under the conditions of Section 5.3. Thus we can write (here and throughout the remainder of the paper, all p -sums are from $-\lfloor \frac{n-1}{2} \rfloor$ to $\lfloor \frac{n}{2} \rfloor$)

$$\begin{aligned} n^{-1} \sum_p a_i(\omega_p; \boldsymbol{\theta}) [I^*(\omega_p) - h(\omega_p; \boldsymbol{\theta})] &= n^{-1} \sum_p a_i(\omega_p; \boldsymbol{\theta}) [h(\omega_p; \boldsymbol{\theta}_0) - h(\omega_p; \boldsymbol{\theta})] + o_p(1) \\ &= \int_{-\pi}^{\pi} a_i(\omega; \boldsymbol{\theta}) [h(\omega; \boldsymbol{\theta}_0) - h(\omega; \boldsymbol{\theta})] d\omega + o_p(1) , \end{aligned}$$

the second step following because the difference between the integral and its Riemann sum is $O(n^{-2})$ providing the integrand has bounded second derivative (condition C3 of Section 5.3 guarantees this). It follows that, for each $\boldsymbol{\theta}$,

$$\begin{aligned} \lim_{n \rightarrow \infty} [\boldsymbol{\eta}_n\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{y})]_i &= -\frac{1}{2} \int_{-\pi}^{\pi} a_i(\omega; \boldsymbol{\theta}) [h(\omega; \boldsymbol{\theta}) - h(\omega; \boldsymbol{\theta}_0)] d\omega \\ &\quad + \frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_i} \frac{(\mu_0 - \mu(\boldsymbol{\theta}))}{h(0; \boldsymbol{\theta})} + \frac{a_i(0; \boldsymbol{\theta})}{2} (\mu_0 - \mu(\boldsymbol{\theta}))^2 . \end{aligned}$$

Pointwise convergence to a limiting function with a zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is now obvious; the uniqueness of this zero is guaranteed by condition C1. Uniform convergence follows from the compactness of Θ .

5.5 Limiting distribution

To obtain a limiting covariance matrix for the estimator, in addition to the conditions for consistency we need to show that the requirements of Assumption 2 are satisfied. In fact, for the class of processes considered here these requirements can be met by choosing $\gamma_n = n^{-1/2}\mathbf{I}$ and $\delta_n = n^{-1/2}\mathbf{I}$. This is demonstrated in Appendix A, using Result 2.

If we can also show that $\gamma_n \mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y})$ has a limiting multivariate normal distribution, then a central limit theorem for the estimator is assured. Again, Result 1 is key: it provides limiting joint normality of the p -sums in each component of $\gamma_n \mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y})$ in (17). The remainder of the normalised estimating function is a linear function of $n^{-1/2}A_0^*$, which itself is normal in the limit according to (4). However, the limiting marginal normality of two non-degenerate quantities is not sufficient to ensure the limiting normality of their sum; and it is not obvious how to provide an easily verifiable condition that ensures this in the present context. We therefore leave this as an additional assumption, noting that it is unlikely to be contentious except in pathological examples. If the assumption holds, it follows that $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} MVN(\mathbf{0}, \mathbf{M}_0 \tilde{\boldsymbol{\Sigma}} \mathbf{M}_0')$, where $\mathbf{M}_0 = \lim_{n \rightarrow \infty} n^{-1} \partial^2 \log L / \partial \boldsymbol{\theta} \boldsymbol{\theta}' |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\tilde{\boldsymbol{\Sigma}} = n^{-1} \text{Var}(\mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y}))$.

To use this result in practice, for finite n it is convenient to proceed as though $\hat{\boldsymbol{\theta}} \sim MVN(\boldsymbol{\theta}_0, \mathbf{G}_0^{-1} \boldsymbol{\Sigma}_n \mathbf{G}_0^{-1})$ where $\mathbf{G}_0 = \partial^2 \log L / \partial \boldsymbol{\theta} \boldsymbol{\theta}' |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\boldsymbol{\Sigma} = \text{Var}(\partial \log L / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0})$. It is then necessary to estimate the matrices \mathbf{G}_0 and $\boldsymbol{\Sigma}_n$. As noted in Jesus and Chandler (2011), in principle any estimator can be used for this purpose providing the estimation error is asymptotically negligible compared with the quantity being estimated. Thus, \mathbf{G}_0 can be estimated as $\hat{\mathbf{G}}_0 = \partial^2 \log L / \partial \boldsymbol{\theta} \boldsymbol{\theta}' |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, which can be calculated either analytically or numerically.

The estimation of $\boldsymbol{\Sigma}_n$ is more challenging. A direct approach would use analytical approximations to the covariance matrix of the spectral scores, similar to the results provided in Appendix A. However, these approximations involve higher-order properties of the process (Y_t) which are rarely available except for rather simple processes. Alternatives are to use empirical or simulation-based estimators. The latter are constructed by simulating a large number of replicates, each of size n , of the process (Y_t) , calculating $\partial \log L / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ for each replicate and then calculating the sample covariance matrix of these simulated spectral score vectors. In many situations, this is the most straightforward way to proceed. However, there are other settings in which it might be appropriate to consider that the observations consist of R (say) independent replicates of an underlying process: an example is given in Section 6. In such settings, a log-likelihood and associated spectral score vector can be computed separately for each replicate; the overall log-likelihood and spectral score are then obtained by summing over the replicates, and the covariance matrix of the spectral scores for each replicate can be estimated as the sample covariance matrix of the individual score vectors. The independence of the replicates then ensures that the covariance matrix of the total score vector is just R times the covariance matrix for an individual replicate.

Whatever approach is used to estimate $\boldsymbol{\Sigma}_n$, the estimating functions framework provides an opportunity to estimate the covariance matrix of $\hat{\boldsymbol{\theta}}$ without recourse to higher-order spectra; the results from Section 4 can then be used for inference about $\boldsymbol{\theta}_0$.

6 Simulation study

We now investigate the finite-sample performance of the theory outlined above, using a simulation study based on the stochastic rainfall model from Section 2. Recall that this model has parameters λ , μ_X and μ_L which are all non-negative. For ease of numerical optimisation, and also to avoid inferential problems if any of the parameter estimates is close to zero, for our inference we take logarithms throughout, so that our parameter vector is $\boldsymbol{\theta} = (\log \lambda, \log \mu_X, \log \mu_L)' \in \Theta$, where Θ is an appropriate compact subspace of \mathbb{R}^3 . By ‘appropriate’ here, we mean ‘large enough to be unrestrictive in practical terms’.

We consider the process (Y_t) of rainfall accumulations aggregated over intervals of length Δ time units, as in equation (1). Without loss of generality, we take $h = 1$. In this case (Y_t) is stationary, with mean

$$\mu(\boldsymbol{\theta}) = \lambda\mu_X\mu_L \quad (18)$$

and second-order spectral density

$$h(\omega; \boldsymbol{\theta}) = \sum_{k=-\infty}^{\infty} s(\omega + 2k\pi) \left(\frac{\sin([\omega + 2k\pi]/2)}{[\omega + 2k\pi]/2} \right)^2 \quad (|\omega| \leq \pi), \quad (19)$$

where

$$s(\omega) = 2\lambda\mu_L^2\mu_X^2/\pi(1 + \mu_L^2\omega^2) \quad (20)$$

is the spectral density of the underlying continuous-time process (the spectral densities for this, and several other similar models, are given by Chandler 1997). Result (19), on the spectral density of the aggregated process, follows from Priestley (1981, equations (4.12.24) and (7.1.12)): in practice, the infinite sum can be truncated at values $\pm K$ beyond which $s(\omega + 2|k|\pi) / (\omega + 2|k|\pi)^2$ is negligible. For the process studied here (and other similar processes), with realistic parameter values we find that $K = 10$ usually suffices.

These closed-form expressions for the mean and spectral density ensure that Whittle estimation is feasible for this particular model and, moreover, that most of the conditions in Section 5.3 can be verified directly. Thus, for any finite $\boldsymbol{\theta}$ the spectral density (19) is finite, bounded away from zero and has finite $\boldsymbol{\theta}$ - and ω -derivatives: hence the requirements of Result 1 are satisfied. Moreover, the spectral density has finite ω -derivatives and the mean (18) has finite $\boldsymbol{\theta}$ -derivatives: thus conditions C2 and C3 are also satisfied. Moving to C1, note that λ and μ_X contribute the same factor $\lambda\mu_X^2$ to the spectral density throughout its range: these parameters are not identifiable from the usual Whittle likelihood therefore, because distinct values of $\boldsymbol{\theta}$ yield identical spectra. However, if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are distinct with $h(\omega; \boldsymbol{\theta}_1) = h(\omega; \boldsymbol{\theta}_2)$ then (18) shows that $\mu(\boldsymbol{\theta}_1) \neq \mu(\boldsymbol{\theta}_2)$ as required by C1: by incorporating the zero frequency term as in (5) therefore, the model becomes identifiable.

The remaining condition to check is C4. This is demonstrated in Appendix B.

6.1 Design of the simulation study

The simulation setup used here is similar to that described in Jesus and Chandler (2011), and is designed to be typical of a situation in which this kind of stochastic rainfall model

is often used. Thus, because the models are stationary they are usually fitted separately to hourly data for each calendar month (Rodriguez-Iturbe et al., 1988; Wheater et al., 2005): data from different years are considered to provide independent replicates of the rainfall process. Typically, such models are fitted using records of least 20 years’ duration. Therefore, for each of our simulations we generate $m = 20$ independent sets of 30 days’ worth of hourly values, each from a model with parameter vector $\boldsymbol{\theta}_0 = (-3.5 \ 0 \ 1.1)'$. These values (with time measured in hours and intensity in mm h^{-1}) can be considered typical of January rainfall in the UK.

We generate 1000 simulated datasets with the structure described above. For each dataset, we compute a log-likelihood (5) separately for each of the 20 replicates; and an overall log-likelihood is then defined as a sum of these individual contributions as described in Section 5.5.

To study the properties of the estimators, in the first instance we examine the distributions of estimates for each parameter. However, the focus of the present paper is not so much upon the estimators *per se* as upon the associated inference. To examine this therefore, we compare the ‘empirical’ and ‘theoretical’ standard errors of the estimators. Specifically, for each parameter we calculate an ‘empirical’ standard error as the standard deviation of the estimates obtained from the 1000 simulations. For comparison, we calculate ‘theoretical’ standard errors as follows: for each simulation we calculate estimates of \mathbf{G}_0 and $\boldsymbol{\Sigma}_n$ as described in Section 5.5, use these to estimate the covariance matrix $\mathbf{G}_0^{-1}\boldsymbol{\Sigma}_n\mathbf{G}_0^{-1}$ and then calculate the sample mean of these estimated covariance matrices as a summary measure. Our ‘theoretical’ standard errors are the square roots of the diagonal elements of this matrix.

As well as comparing the empirical and theoretical standard errors, we examine the coverages of 95% and 99% confidence intervals and regions. Specifically, for each simulated dataset, confidence intervals are calculated according to the large-sample multivariate normal distribution from Section 5.5. For each parameter, the proportion of intervals containing the true parameter value is calculated. If the theory holds in practice, then a $100(1 - \alpha)\%$ confidence interval should contain the true parameter value roughly $100(1 - \alpha)\%$ of the time. Similarly, confidence regions for the entire parameter vector can be computed, either on the basis of the statistic $2 \left[\log L(\hat{\boldsymbol{\theta}}; \mathbf{y}) - \log L(\boldsymbol{\theta}_0; \mathbf{y}) \right]$ (this is the test statistic (7) with $\boldsymbol{\Xi} = \mathbf{I}_{3 \times 3}$ and $\boldsymbol{\xi} = \boldsymbol{\theta}_0$), or the adjusted version of this statistic corresponding to (9).

A further question of interest relates to the estimation of $\boldsymbol{\Sigma}_n$ in Section 5.5. As noted there, the availability of multiple independent replicates allows the possibility of estimating $\boldsymbol{\Sigma}_n$ empirically by computing the empirical covariance matrix of the spectral scores from each replicate. In a similar setting involving generalised method of moments estimation however, Jesus and Chandler (2011) found that this empirical estimator led to very poorly calibrated inference. For comparison therefore, we also investigate the performance of a simulation-based estimator as described in Section 5.5. For each simulation, 1000 further replicates of the process are generated using the parameter vector $\hat{\boldsymbol{\theta}}$; the spectral scores at $\hat{\boldsymbol{\theta}}$ are calculated for each new replicate (note that no optimisation is required here: merely the evaluation of (17) at $\hat{\boldsymbol{\theta}}$ for each simulated dataset); and the results are used to estimate the covariance matrix of the spectral scores for a single replicate. In the

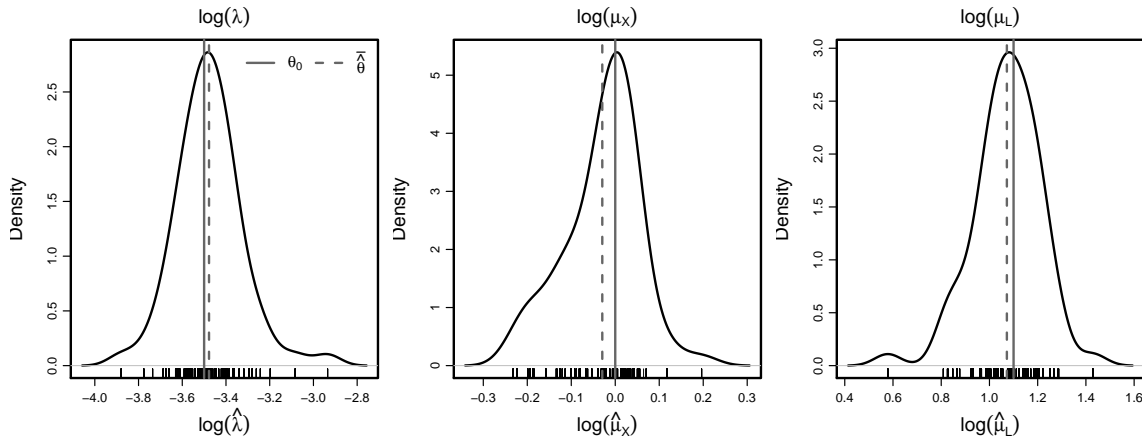


Figure 1: Sampling distributions of estimators for the parameters of a Poisson pulse model for rainfall. Distributions are obtained from 1000 simulated data sets, each containing 20 independent 30-day sequences of hourly values and generated using parameter values $\log(\lambda) = -3.5, \log(\mu_X) = 0, \log(\mu_L) = 1.1$. Vertical solid and dashed lines represent respectively the true value of each parameter and the mean of the simulated estimates. The individual estimates are indicated at the bottom of each plot; and their distributions are shown as kernel density estimates.

results reported below, these two approaches are referred to respectively as “Empirical” and “Simulation-based”.

6.2 Results

Figure 1 shows the distributions of the 1000 sets of estimates. The distributions for $\log \hat{\lambda}$ and $\log \hat{\mu}_L$ seem reasonably symmetrical so that a normal approximation may be adequate here; the distribution of $\log \hat{\mu}_X$ is negatively skewed. Moreover, the estimates appear slightly biased: the empirical biases for the three parameters are 0.022, -0.029 and -0.028 respectively, with associated standard errors 0.005, 0.003 and 0.004. Taken together, these results suggest that larger sample sizes may be required before the asymptotics for this particular process become highly accurate. Nonetheless, the biases are on the order of just 2% to 3% on the original (non-logarithmic) scale for each parameter: this is probably acceptable for many applications. Moreover, the skewness of the sampling distribution for $\log \mu_X$ is arguably problematic only if it affects the substantive conclusions derived from the asymptotic theory, for example if the associated standard errors and confidence interval coverages are inaccurate. This can be investigated using the results in Table 1.

Focusing first on the empirical and theoretical standard errors in Table 1, there seems to be reasonable agreement: the empirical estimation of Σ_n seems to yield lower standard errors than the simulation-based estimation, although there is no particular reason on the basis of these results to prefer one method over the other. A different story emerges when we consider the coverages of confidence intervals for individual parameters, however: here, the empirical estimates of Σ_n lead to undercoverage in all cases, and the performance of the confidence intervals using simulation-based estimates is much better. However, the

	Parameter							
	λ		μ_X		μ_L		Θ	
	$\hat{\Sigma}_{\text{Emp}}$	$\hat{\Sigma}_{\text{Sim}}$	$\hat{\Sigma}_{\text{Emp}}$	$\hat{\Sigma}_{\text{Sim}}$	$\hat{\Sigma}_{\text{Emp}}$	$\hat{\Sigma}_{\text{Sim}}$	$\hat{\Sigma}_{\text{Emp}}$	$\hat{\Sigma}_{\text{Sim}}$
Bias	0.022		−0.029		−0.028		—	
Empirical std err	0.151		0.083		0.136		—	
Theoretical std err	0.144	0.165	0.087	0.096	0.122	0.133	—	—
95% coverage	0.93	0.96	0.84	0.91	0.87	0.88	0.75	0.78
(adjusted)	—	—	—	—	—	—	0.84	0.93
99% coverage	0.96	0.97	0.93	0.97	0.94	0.99	0.90	0.91
(adjusted)	—	—	—	—	—	—	0.90	0.97

Table 1: Performance of asymptotic inference in 1000 simulations of a Poisson pulse model for rainfall. “Empirical” standard errors are derived directly from the 1000 sets of estimates; “theoretical” ones from the mean of the estimated large-sample covariance matrices. Coverages indicate the proportions of simulations for which asymptotic confidence intervals or regions at the nominal level contained the true parameter value: for individual parameters, confidence intervals are directly from the limiting normal distribution of the estimator, while for the entire parameter vector they are derived either from (7) or (for the ‘adjusted’ rows) from (9). In columns ‘ $\hat{\Sigma}_{\text{Emp}}$ ’, Σ_n is estimated from the empirical covariance matrix of the spectral scores in each simulation; in columns ‘ $\hat{\Sigma}_{\text{Sim}}$ ’, it is estimated by resimulating from the fitted model.

coverage of confidence regions for the entire parameter vector, based on (7), is rather poor whichever estimate is used (e.g. 75% or 78% for a nominal 95% confidence region). The coverage of confidence regions based on (9) is not much improved when Σ_n is estimated empirically; but for simulation-based estimation of Σ_n these coverages improve dramatically, to 93% and 97% for nominal 95% and 99% confidence regions. For nonlinear models where inference is challenging, and a sample size at the lower end of what is typical when applying this kind of model, this kind of performance is quite impressive.

Although these results are from a single set of simulations of one particular model, the messages are consistent with our experience and understanding of similar problems elsewhere. Thus, when studying the performance of the generalised method of moments, we find that empirical covariance estimation can yield very poor confidence interval coverage in finite samples and that simulation-based estimation is much more reliable (Jesus and Chandler, 2011). Similarly, when calculating confidence regions based on a mis-specified likelihood, arguments in Chandler and Bate (2007) suggest that in general the contours of the mis-specified likelihood are the wrong shape so that it is hard to achieve accurate inference in finite samples; adjustments such as (8) are designed to overcome this.

7 Discussion

We have extended the work of Heyde (1997) to show how Whittle estimation may be viewed from an estimating functions perspective, which enables formal asymptotic in-

ference to be carried out without evaluating higher-order properties of the model under consideration; to show how some aspects of the inference can be improved by working with an easily-computed adjustment to the Whittle log-likelihood (although this adjustment does not change the estimator itself); and to provide asymptotics for use in situations where the mean of the process depends on the parameter(s) of interest so that a zero frequency term should be included in the Whittle likelihood. In a special issue devoted to the memory of Maurice Priestley, it is perhaps worth pointing out that our contribution here can be regarded as a (very minor, compared with his own work) practical contribution to spectral estimation.

From an applied perspective, perhaps the most interesting contribution of the present article is to provide a feasible way of calculating standard errors and confidence regions for Whittle estimation. This is particularly useful given the concerns expressed by some authors (e.g. Contreras-Cristán et al. 2006) about its inefficiency: if standard errors can be estimated reliably, then the analyst can at least be made aware that parameter uncertainty may be large. The use of an adjusted log-likelihood surface to define confidence regions may also help to resolve the issue, reported in Contreras-Cristán et al. (2006), that the Whittle log-likelihood surface may have a substantially different shape from the real log-likelihood. A reasonable estimator of the covariance matrix of the spectral scores is needed, however, for the inference to be accurate. In many applications of estimating functions, an ‘empirical’ estimator of this covariance matrix is used; however, we find that this can perform poorly for Whittle estimation, in which case an alternative should be sought. Our experience is that simulation-based estimators perform well, and can be calculated cheaply on modern computers for any process that is easy to simulate.

We have restricted our attention to processes satisfying the conditions of Section 5.3. These conditions allow the application of Whittle estimation to a much wider class of situations than is often appreciated (for example, the highly non-Gaussian rainfall model of our example), but nonetheless they exclude some important situations — for example, long-memory processes where the spectral density $h(\omega; \boldsymbol{\theta})$ is unbounded at the origin. They are imposed in order to satisfy Assumptions 1 and 2 which are taken from Jesus and Chandler (2011) as a fairly generic set of conditions under which ‘classical’ asymptotic results hold. Some of the conditions can be weakened in specific situations (see, for example, the earlier discussion of alternative conditions under which Result 1 holds), and hence the estimating functions approach can be applied to other classes of process as well; however, such classes typically need to be studied individually to establish the required results. We nonetheless hope that our contribution will stimulate an interest in this kind of approach to frequency-domain inference, and that it will highlight the availability of easily-computed standard errors and confidence regions in this setting. For the specific case of long-memory processes, it is perhaps worth noting that the large-sample distribution (4) of the Fourier coefficients, that was used to motivate the log-likelihood expression (5), is no longer valid for the lowest frequencies: the justification for using (5) therefore seems questionable, with or without the zero-frequency term. An alternative is to form a log-likelihood from the correct large-sample distributions of the Fourier coefficients, as implied by results in Hurvich and Beltrao (1993) for example. This is essentially the idea of Sykulski et al. (2015), who recommend replacing the spectral density in the Whittle likelihood by the expected value of the periodogram. It would be of interest to use an

estimating functions approach to explore the properties of such methods.

Appendix A: Requirements of Assumption 2

In Section 5, we claimed that with the choice $\gamma_n = n^{-1/2}\mathbf{I}$ and $\delta_n = n^{-1/2}\mathbf{I}$ and under the conditions of Section 5.3, the estimating functions $\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y})$ with i th element given by (17) meet the requirements of Assumption 2. We here justify this claim.

First, we note that condition C2 (Section 5.3) ensures that the matrices $\mathbf{G}_n(\boldsymbol{\theta}) = \partial \mathbf{g}_n / \partial \boldsymbol{\theta}$ exist. Our first task, therefore, is to demonstrate that the covariance matrix of $\gamma_n \mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y})$ converges to a limiting positive definite matrix $\tilde{\Sigma}$. Consider the (i, j) element of $\gamma_n \mathbf{g}_n(\boldsymbol{\theta}; \mathbf{Y})$: $\text{Cov}(n^{-1/2}g_i(\boldsymbol{\theta}; \mathbf{y}), n^{-1/2}g_j(\boldsymbol{\theta}; \mathbf{y}))$ which, from (17), can be written as

$$\frac{1}{4} \text{Cov} \left(n^{-1/2} \sum_p a_i(\omega_p; \boldsymbol{\theta}) [I^*(\omega_p) - h(\omega_p; \boldsymbol{\theta})], \right. \\ \left. n^{-1/2} \sum_p a_j(\omega_p; \boldsymbol{\theta}) [I^*(\omega_p) - h(\omega_p; \boldsymbol{\theta})] \right) \quad (21)$$

$$+ \frac{b_{ij}(\boldsymbol{\theta})}{n} \text{Var}(A_0^*) \quad (22)$$

$$+ \frac{1}{n} \text{Cov} \left(\sum_p [d_{ij}(\boldsymbol{\theta})a_i(\omega_p; \boldsymbol{\theta}) + d_{ji}(\boldsymbol{\theta})a_j(\omega_p; \boldsymbol{\theta})] I^*(\omega_p), A_0^* \right), \quad (23)$$

where $b_{ij}(\boldsymbol{\theta})$ and $d_{ij}(\boldsymbol{\theta})$ are non-random functions of $\boldsymbol{\theta}$.

The convergence of (21) follows immediately from Result 1 given earlier, and the limit of (22) follows directly from (4):

$$\lim_{n \rightarrow \infty} \frac{b_{ij}(\boldsymbol{\theta}_0)}{n} \text{Var}(A_0^*) = b_{ij}(\boldsymbol{\theta}_0) h(0; \boldsymbol{\theta}_0).$$

The treatment of (23) is less straightforward. Note first that $E(A_0^*) = 0$ so that the covariance is just an expectation. Note also that $A_0^* = J_0^*$ and that $I^*(\omega_p) = n^{-1} J_p^* J_{-p}^*$ where J_p^* is the p th complex Fourier coefficient of the centred process, defined analogously to J_p in Result 2 (Section 5.1). Hence, focusing on the first term in (23), we have

$$\begin{aligned} \frac{1}{n} \text{Cov} \left[\sum_p d_{ij}(\boldsymbol{\theta}) a_i(\omega_p; \boldsymbol{\theta}) I^*(\omega_p), A_0^* \right] &= \frac{d_{ij}(\boldsymbol{\theta})}{n} E \left[A_0^* \sum_p a_i(\omega_p; \boldsymbol{\theta}) I^*(\omega_p) \right] \\ &= \frac{d_{ij}(\boldsymbol{\theta})}{n^2} \sum_p a_i(\omega_p; \boldsymbol{\theta}) E [J_0^* J_p^* J_{-p}^*] \\ &= \frac{d_{ij}(\boldsymbol{\theta})}{n^2} \sum_p a_i(\omega_p; \boldsymbol{\theta}) c_3(J_0^*, J_p^*, J_{-p}^*). \end{aligned}$$

The last step here again makes use of the fact that $E[J_p^*] = 0$; as in Result 2, $c_3(\cdot, \cdot, \cdot)$ is

a generic notation for the joint third cumulant. Applying (14) now yields

$$\frac{d_{ij}(\boldsymbol{\theta})}{n^2} \sum_p a_i(\omega_p; \boldsymbol{\theta}) [(2\pi)^2 n h^{(3)}(0, \omega_p) + O(1)] = \frac{(2\pi)^2 d_{ij}(\boldsymbol{\theta})}{n} \sum_p a_i(\omega_p; \boldsymbol{\theta}) h^{(3)}(0, \omega_p) + O(n^{-1}),$$

where $h^{(3)}(\cdot, \cdot)$ is the third-order spectral density of (Y_t) . Boundedness of the second derivatives (Section 5.3, condition C4) ensures that the sum approximates an integral with negligible error, so that the entire expression converges to $d_{ij}(\boldsymbol{\theta}) (2\pi)^2 \int_{-\pi}^{\pi} a_i(\omega; \boldsymbol{\theta}) h^{(3)}(0, \omega) d\omega$ as $n \rightarrow \infty$. Similarly, the second term in (23) tends to $d_{ji}(\boldsymbol{\theta}) (2\pi)^2 \int_{-\pi}^{\pi} a_j(\omega; \boldsymbol{\theta}) h^{(3)}(0, \omega) d\omega$.

Combining all of these results, as $n \rightarrow \infty$ the elements of $\text{Cov}(\boldsymbol{\gamma}_n \mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y}), \boldsymbol{\gamma}_n \mathbf{g}_n(\boldsymbol{\theta}_0; \mathbf{Y}))$ are seen to converge to quantities which, under the restrictions of Section 5.3, are all finite. The limiting matrix is the required $\tilde{\boldsymbol{\Sigma}}$ of Assumption 2.

We now turn to the second requirement of Assumption 2: that for $\tilde{\mathbf{g}}(\boldsymbol{\theta}; \mathbf{y}) = \boldsymbol{\gamma}_n \mathbf{g}_n(\boldsymbol{\theta}; \mathbf{y})$ there is a sequence $(\boldsymbol{\delta}_n)$ of matrices such that $[\partial \tilde{\mathbf{g}}(\boldsymbol{\theta}; \mathbf{Y}) / \partial \boldsymbol{\theta}] \boldsymbol{\delta}_n$ converges in probability to a non-random matrix $\mathbf{M}(\boldsymbol{\theta})$, continuous in $\boldsymbol{\theta}$, in a neighbourhood of $\boldsymbol{\theta}_0$. Under the conditions of Section 5.3, the choice $\boldsymbol{\delta}_n = n^{-1/2} \mathbf{I}$ satisfies this requirement. For, in this case, recalling that $\boldsymbol{\gamma}_n = n^{-1/2} \mathbf{I}$, the $(i, j)^{\text{th}}$ element of $[\partial \tilde{\mathbf{g}}(\boldsymbol{\theta}; \mathbf{Y}) / \partial \boldsymbol{\theta}] \boldsymbol{\delta}_n$ is $n^{-1} \partial g_i / \partial \theta_j$, with $g_i(\boldsymbol{\theta}; \mathbf{Y})$ defined in (17). The required convergence to a continuous function can be shown using arguments that are very similar to those presented above (notably invoking Result 1, the convergence of $n^{-1} A_0^*$ to zero and the approximation of a p -sum with an ω -integral). The details are omitted because the algebra, although straightforward, is lengthy and not particularly instructive.

Appendix B: Condition C4 for the rainfall model

To apply the results of Section 5 to the rainfall model of Section 2, it is necessary to demonstrate that condition C4 holds for the discretised process in which Y_t represents the aggregated rainfall over some time interval Δ as in equation (1). Without loss of generality, in this appendix we take $\Delta = 1$ time unit. As discussed in Section 5.3, for a process with no periodic variation it is sufficient to demonstrate that $\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} r_2^2 |c_3(r_1, r_2)| < \infty$, where $c_3(\cdot, \cdot)$ is the joint third-order cumulant of Y_t, Y_{t+r_1} and Y_{t+r_2} . The derivation below makes use of two preliminary results.

Lemma 1: Let $Y = \sum_{i=1}^N X_i$ where $N \sim \text{Poi}(\mu)$, the $\{X_i\}$ are independent and identically distributed random variables with finite positive moments of all orders and $Y := 0$ if $N = 0$. Let $\alpha_k(\mu) = \text{E}(Y^k | N > 0)$ for $k \in \mathbb{N}$. Then $\alpha_k(\mu)$ is an increasing function of μ .

Proof: Unconditionally, Y has a compound Poisson distribution and $\text{E}(Y^k)$ is a k th-order polynomial in μ , with positive coefficients that depend on the moments of X . We have $\alpha_k(\mu) = \text{E}(Y | N > 0) = \text{E}(Y^k) / (1 - e^{-\mu})$, which is an increasing function of μ because $\mu^n / (1 - e^{-\mu})$ is increasing in $\mu > 0$ for $n \in \mathbb{N}$. ■

Lemma 2: for any finite $m > 0$, $A > 0$ and $B > 0$, $0 < \sum_{r=0}^{\infty} r^m [1 - \exp(-Ae^{-Br})] < \infty$.

Proof: the lower bound is trivial since $\exp(-Ae^{-Br}) < 1$ when $A > 0$ and $B > 0$. It follows that $\sum_{r=0}^{\infty} r^m [1 - \exp(-Ae^{-Br})] = |\sum_{r=0}^{\infty} r^m [1 - \exp(-Ae^{-Br})]|$

$$\begin{aligned} &= \left| -\sum_{r=0}^{\infty} r^m \sum_{k=1}^{\infty} (-Ae^{-Br})^k / k! \right| = \left| -\sum_{k=1}^{\infty} \frac{(-A)^k}{k!} \sum_{r=0}^{\infty} r^m e^{-Brk} \right| \\ &< \sum_{k=1}^{\infty} \frac{A^k}{k!} \sum_{r=0}^{\infty} r^m e^{-Br} = (e^A - 1) \sum_{r=0}^{\infty} r^m e^{-Br} , \end{aligned} \quad (24)$$

the strict inequality following from the fact that $e^{-Brk} < e^{-Br}$ for $k > 1$ and $B > 0$. The r -sum in (24) is convergent by the ratio test, and the result is shown. ■

Turning now to the cumulants for the rainfall model: note first that $c_3(r_1, r_2)$ can alternatively be written as $E(Y_t^* Y_{t+r_1}^* Y_{t+r_2}^*)$, where $Y_t^* = Y_t - E(Y_t)$. Also, recall that rain cells arrive in a Poisson process and that their intensities are independent: dependence in the process can only arise, therefore, from cells that affect multiple intervals simultaneously. Accordingly, let $U_{t,t+r_1,t+r_2}$ be a random variable taking the value 1 if at least one cell is active during all three intervals, and zero otherwise. Then $E(Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 0) = 0$ because, conditional upon $U_{t,t+r_1,t+r_2} = 0$, at least one of the intervals must be independent of the other two and the process (Y_t^*) has zero mean. Unconditionally therefore, we have

$$\begin{aligned} c_3(r_1, r_2) &= E(Y_t^* Y_{t+r_1}^* Y_{t+r_2}^*) = E_{U_{t,t+r_1,t+r_2}} [E(Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2})] \\ &= E(Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 1) \times P(U_{t,t+r_1,t+r_2} = 1) , \end{aligned}$$

so that

$$|c_3(r_1, r_2)| = E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 1) \times P(U_{t,t+r_1,t+r_2} = 1) . \quad (25)$$

The next step is to show that $E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 1)$ is bounded uniformly in r_1 and r_2 . A full derivation is neither brief nor instructive, so we merely sketch the argument. First, Y_t is non-negative and $E(Y_t)$ is finite so that unboundedness of $E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 1)$ can arise only due to contributions from the upper tail of the joint distribution. A sufficient condition for $E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^* | U_{t,t+r_1,t+r_2} = 1)$ to be uniformly bounded is therefore that $E(Y_t Y_{t+r_1} Y_{t+r_2} | U_{t,t+r_1,t+r_2} = 1)$ itself is uniformly bounded. Now write $Y_t = \tilde{Y}_t + \check{Y}_t$, where \tilde{Y}_t is the contribution from cells that are active during all three intervals and \check{Y}_t is the contribution from all other cells. \tilde{Y}_t and \check{Y}_t are independent due to the independence of cell characteristics; and $E(Y_t Y_{t+r_1} Y_{t+r_2} | U_{t,t+r_1,t+r_2} = 1) = E(Y_t Y_{t+r_1} Y_{t+r_2} | \tilde{Y}_t > 0)$. The value of $\tilde{Y}_t \tilde{Y}_{t+r_1} \tilde{Y}_{t+r_2}$ cannot exceed that obtained when the contributing cells are all active throughout each interval: in this case \tilde{Y}_t , \tilde{Y}_{t+r_1} and \tilde{Y}_{t+r_2} are all equal and their marginal distribution is compound Poisson with mean at most $\lambda(\mu_L + 1)\mu_X$ (because the exponential cell intensity distribution has finite moments, and the rate of cells that simultaneously affect three intervals is no greater than the

rate of cells affecting a single interval which is $\lambda(\mu_L + 1)$). Similar uniform bounds are obtained for higher moments. Lemma 1 now guarantees uniform boundedness of the conditional moments given $\tilde{Y}_t > 0$. Next consider the joint moments of \check{Y}_t , \check{Y}_{t+r_1} and \check{Y}_{t+r_2} . Independence of \check{Y}_t and \check{Y}_t implies that conditioning on $\check{Y}_t > 0$ does not affect these joint moments, which cannot exceed the corresponding (unconditional) joint moments of the original process (Y_t) because the non-negative contributions from some cells have been subtracted. The original process has finite joint moments that are uniformly bounded (for example, a Chebyshev-style argument coupled with stationarity can be used to show that $E(Y_t Y_{t+r_1} Y_{t+r_2}) \leq 2E(Y_t^3)$). Conditional on $U_{t,t+r_1,t+r_2} = 1$ therefore, the processes (\check{Y}_t) and (\check{Y}_t) have uniformly bounded joint third moments and are independent: uniform boundedness of $E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^*| | U_{t,t+r_1,t+r_2} = 1)$ follows, yielding $E(|Y_t^* Y_{t+r_1}^* Y_{t+r_2}^*| | U_{t,t+r_1,t+r_2} = 1) < M$ for some finite M .

Having established this, from (25) we have

$$|c_3(r_1, r_2)| \leq MP(U_{t,t+r_1,t+r_2} = 1) . \quad (26)$$

If r_1 and r_2 are not both equal to zero, $P(U_{t,t+r_1,t+r_2} = 1)$ is the probability that at least one cell active at the end of the earliest time interval $t, t + r_1, t + r_2$ is still active at the start of the latest interval. By stationarity, this is equal to the probability that at least one cell active at time zero is still active at time r , where $r = \max(|r_1|, |r_2|, |r_2 - r_1|) - 1$. Denote this event by E_r , and let N_0 be the number of cells active at time zero. The model specification implies that $N_0 \sim \text{Poi}(\lambda\mu_L)$ (Rodriguez-Iturbe et al., 1987).

In this case, for $r \geq 0$ we have $P(E_r) = \sum_{k=0}^{\infty} P(E_r | N_0 = k) P(N_0 = k)$

$$\begin{aligned} &= \sum_{k=0}^{\infty} [1 - P(\text{all } k \text{ active cells terminated by time } r) | N_0 = k] (\lambda\mu_L)^k e^{-\lambda\mu_L} / k! \\ &= \sum_{k=0}^{\infty} \left[1 - (1 - e^{-r/\mu_L})^k \right] (\lambda\mu_L)^k e^{-\lambda\mu_L} / k! , \end{aligned}$$

the last step following from the independence of the exponentially distributed cell durations. This expression simplifies to $1 - \exp[-\lambda\mu_L e^{-r/\mu_L}]$ so that, if r_1 and r_2 are not both equal to zero, $P(U_{t,t+r_1,t+r_2} = 1) = 1 - \exp[-\lambda\mu_L e^{-[\max(|r_1|, |r_2|, |r_2 - r_1|) - 1]/\mu_L}]$. From (26) therefore, $|c_3(r_1, r_2)| \leq M \{1 - \exp[-\lambda\mu_L e^{-[\max(|r_1|, |r_2|, |r_2 - r_1|) - 1]/\mu_L}]\}$ so that $\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} r_2^2 |c_3(r_1, r_2)|$ will be finite if

$$\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} r_2^2 \{1 - \exp[-\lambda\mu_L e^{-[\max(|r_1|, |r_2|, |r_2 - r_1|) - 1]/\mu_L}]\} < \infty . \quad (27)$$

Strictly speaking, the term for $r_1 = r_1 = 0$ here does not correspond to $P(U_{t,t,t} = 1)$; however, the difference between the two quantities is finite so it suffices to demonstrate (27). Now,

$$\max(|r_1|, |r_2|, |r_2 - r_1|) = \begin{cases} r_1 & 0 \leq r_2 \leq r_1 \\ -r_1 & r_1 \leq r_2 \leq 0 \\ r_2 & 0 \leq r_1 < r_2 \\ -r_2 & r_2 < r_1 \leq 0 \\ r_1 - r_2 & r_1 > 0, r_2 < 0 \\ r_2 - r_1 & r_1 < 0, r_2 > 0 \end{cases}$$

We therefore split the sum on the left-hand side of (27) into components corresponding to each of these six regions. It is straightforward to show that the contributions from the first two regions ($0 \leq r_2 \leq r_1$ and $r_1 \leq r_2 \leq 0$) are equal; similarly that the contributions from the other two pairs of regions are equal. Thus the left-hand side of (27) can be written as

$$\begin{aligned}
& 2 \left\{ \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{r_1} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(r_1-1)/\mu_L})] + \sum_{r_2=1}^{\infty} \sum_{r_1=0}^{r_2-1} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(r_2-1)/\mu_L})] \right. \\
& \qquad \qquad \qquad \left. + \sum_{r_2=-\infty}^{-1} \sum_{r_1=1}^{\infty} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(r_1-r_2-1)/\mu_L})] \right\} \\
= & 2 \left\{ \sum_{r_1=0}^{\infty} \frac{r_1(r_1+1)(2r_1+1)}{6} [1 - \exp(-\lambda\mu_L e^{-(r_1-1)/\mu_L})] \right. \\
& \left. + \sum_{r_2=1}^{\infty} r_2^3 [1 - \exp(-\lambda\mu_L e^{-(r_2-1)/\mu_L})] + \sum_{r_2=1}^{\infty} \sum_{r_1=1}^{\infty} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(r_1+r_2-1)/\mu_L})] \right\}
\end{aligned}$$

where we have reversed the sign of r_2 in the final term. The first two terms are finite by Lemma 2. Substituting $v = r_1 + r_2$ in place of r_1 in the final term, we get

$$\begin{aligned}
& \sum_{r_2=1}^{\infty} \sum_{v=r_2+1}^{\infty} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(v-1)/\mu_L})] \\
= & \sum_{v=2}^{\infty} \sum_{r_2=1}^{v-1} r_2^2 [1 - \exp(-\lambda\mu_L e^{-(v-1)/\mu_L})] \\
= & \sum_{v=2}^{\infty} \frac{v(v-1)(2v-1)}{6} [1 - \exp(-\lambda\mu_L e^{-(v-1)/\mu_L})] ,
\end{aligned}$$

which is also finite by Lemma 2. Condition C4 is thus established.

References

- Bardet, J.-M., Doukhan, P., and León, J. R. (2008). Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. *Journal of Time Series Analysis*, 29:906–945.
- Brillinger, D. R. (1975). *Time Series - Data Analysis and Theory*. Holt, Rinehart and Winston, Inc.
- Brillinger, D. R. and Rosenblatt, M. (1967). Asymptotic theory of estimates of k-th order spectra. *Proc Nat Acad Sciences USA*, 57(2):206–210.
- Chandler, R. E. (1997). A spectral method for estimating parameters in rainfall models. *Bernoulli*, 3(3):301–322.

- Chandler, R. E. and Bate, S. M. (2007). Inference for clustered data using the independence log-likelihood. *Biometrika*, 94:167–183.
- Chandler, R. E. and Scott, E. M. (2011). *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. Wiley, Chichester.
- Contreras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. (2006). A note on Whittle’s likelihood. *Communications in Statistics — Simulation and Computation*, 35:857–875.
- Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Stat*, 28(6):1762–1794.
- Dahlhaus, R. (2009). Local inference for locally stationary time series based on empirical spectral measure. *Journal of Econometrics*, 151:101–112.
- Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent Gaussian time series. *The Annals of Statistics*, 14:517–532.
- Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, 89:197–210.
- Giraitis, L., Hidalgo, J., and Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *The Annals of Statistics*, 29:987–1023.
- Giraitis, L. and Koul, H. L. (2013). On asymptotic distributions of weighted sums of periodograms. *Bernoulli*, 19:2389–2413.
- Giraitis, L. and Robinson, P. (2001). Whittle estimation of ARCH models. *Econometric Theory*, 17:608–631.
- Giraitis, L. and Taqqu, M. S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *The Annals of Statistics*, 27(1):178–203.
- Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, 10:130–145.
- Hauser, M. (1998). Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. *J. Statistical Planning and Inference*, XX.
- Heyde, C. C. (1997). *Quasi-Likelihood and its applications*. New York: Springer.
- Heyde, C. C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications*, 45:169–182.
- Hurvich, C. M. and Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory process. *J. Time Series Analysis*, 14:455–472.
- Jesus, J. and Chandler, R. E. (2011). Estimating functions and the generalized method of moments. *Interface Focus*, 1:871–885.
- Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69(1):19–27.

- Krafty, R. T. and Hall, M. (2013). Canonical correlation analysis between time series and static outcomes, with application to the spectral analysis of heart rate variability. *The Annals of Applied Statistics*, 7:570–587.
- Montanari, A. and Toth, E. (2007). Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? *Water Resources Research*, 43:W05434. DOI:10.1029/2006WR005184.
- Northrop, P. J. (2006). Estimating the parameters of rainfall models using maximum marginal likelihood. *Student*, 5(3/4):173–183.
- Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.
- Rice, J. (1979). On the estimation of the parameters of a power spectrum. *Journal of Multivariate Analysis*, 9:378–392.
- Robinson, P. M. (1978). Alternative models for stationary stochastic processes. *Stochastic Processes and their Applications*, 8:141–152.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proc R. Soc. Lond.*, A410:269–288.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1988). A point process model for rainfall: further developments. *Proc R. Soc. Lond.*, A417:283–298.
- Severini, T. A. (2000). *Likelihood methods in Statistics*. Oxford University Press, New York.
- Subba Rao, T. and Chandler, R. E. (1996). A frequency domain approach for estimating parameters in point process models. In Robinson, P. M. and Rosenblatt, M., editors, *Athens Conference on Applied Probability and Time Series, Vol. II: Time Series Analysis (in memory of E.J. Hannan)*, pages 392–405. Springer-Verlag, New York.
- Sykulski, A. M., Olhede, S. C., Lilly, J. M., and Danioux, E. (2015). Lagrangian time series models for ocean surface drifter trajectories. *J. Roy. Statist. Soc. Series C (Applied Statistics)*, to appear. doi:10.1111/rssc.12112.
- Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M. L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Env. Res. & Risk Ass*, 19:403–416.
- Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, 2:423–434.