

ON THE STRUCTURE AND ESTIMATION OF REFLECTION POSITIVE PROCESSES

R. McVINISH,* *Queensland University of Technology*

Abstract

The class of processes formed as the aggregation of Ornstein–Uhlenbeck processes has proved useful in modeling time series from a number of areas and includes several interesting special cases. This paper examines the second-order properties of this class. Bounds on the one-step prediction error variance are proved and consistency of the minimum contrast estimation is demonstrated.

Keywords: Aggregation; long memory; minimum contrast estimation; prediction error

2000 Mathematics Subject Classification: Primary 62M15

Secondary 62G05

1. Introduction

Observational time series from many areas including economics [40] and high-speed networks [36] are formed by the contemporaneous aggregation of a large number of heterogeneous units. For these time series, the behavior at the unit level can often be adequately described by a simple model such as an (autoregressive) AR(1) time series or an on–off process. However, as demonstrated in [22], complex behavior such as long memory may still be observed at the aggregate level depending on the heterogeneity of unit-level parameters.

The phenomenon of complex behavior arising through the aggregation of simple models has been used by researchers in two ways. Firstly, some researchers have exploited this idea in the construction of new models. In particular, Barndorff-Nielsen and coworkers have made considerable use of sums of Lévy driven Ornstein–Uhlenbeck processes to provide a class of processes with flexible marginal distributions and covariance functions (see [5], [6], [8], and [9], among others by this group). These processes have been applied to model data where there is no obvious aggregation in forming the data such as stochastic volatility and turbulence. In some instances the aggregation of as few as two processes were necessary to model the data. Other papers in this line of research include [15], [16], [25], and [37]. Secondly, aggregation has been used to provide a physical interpretation of an existing complex model. This was first done in [22], where the fractional (autoregressive integrated moving average) ARIMA(0, d , 0) model ($0 < d < \frac{1}{2}$) [23] was ‘disaggregated’ into AR(1) unit-level models. Dacunha-Castelle and Fernín [17] have since considered ‘disaggregating’ complex models into other types of unit-level models.

Owing to their popularity in modeling and the richness of the related theory, this paper will focus on the aggregation of independent Ornstein–Uhlenbeck processes, that is, the process Y_t

Received 20 April 2006; revision received 14 November 2007.

* Postal address: School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4001, Australia. Email address: r.mcvinish@qut.edu.au

defined by

$$Y_t := M^{-1/2} \sum_{i=1}^M X_t^{(i)}, \quad dX_t^{(i)} = -\lambda_i X_t^{(i)} dt + \sqrt{2\lambda_i} dB_t^{(i)}, \tag{1.1}$$

where $M \in \mathbb{N}$, $X^{(i)}$ is the Ornstein–Uhlenbeck process with parameter λ_i , and $B_t^{(i)}$ are independent copies of a zero mean, uncorrelated increment process. As described in Section 2, these processes, and their weak limits, will be referred to as reflection positive (RP) processes. Some basic properties of this series will be seen to follow directly from properties of the Laplace transform. In Section 3 the time series formed from instantaneous sampling of the continuous-time process is studied. The one-step prediction error is studied in Section 4 with particular emphasis on the limit of predictability for a sampled RP process. It is demonstrated that a tight lower bound on the one-step prediction error variance can be obtained using only the variance and covariance at lag 2. Finally, in Section 5 weak consistency of the minimum contrast estimate for the aggregating measure of an RP process is established. Estimation of the aggregating measure is important in certain applications where the unit-level models can be interpreted, such as the growth and investment models described in [40]. Moreover, a consistent estimate of the aggregating measure can be used to provide a consistent estimate of the spectral density. It is interesting to note that these prediction error and estimation results are not affected by the presence of long memory.

2. Continuous-time, RP processes

Following [33], a process is said to be RP if its covariance function satisfies

$$\sum_i \sum_j a_i \gamma(t_i + t_j) \bar{a}_j \geq 0, \quad a \in \mathbb{C}^m, t \in \mathbb{R}^m.$$

Let \bar{N} be the class of finite Borel measures on $[0, \infty)$, and let N be the restriction of \bar{N} to $(0, \infty)$. This condition was shown in [24] to be equivalent to $\gamma(t)$ having the representation

$$\gamma(t) = \int_0^\infty e^{-\lambda|t|} \nu(d\lambda) \tag{2.1}$$

for some $\nu \in \bar{N}$. Since the covariance function of the aggregate processes (1.1) is given by

$$\gamma(t) = M^{-1} \sum_{i=1}^M \exp(-\lambda_i |t|) = \int_0^\infty e^{-\lambda|t|} \nu_M(d\lambda),$$

where ν_M is a finite measure supported only on M points, it follows that the aggregate process has the RP property. Furthermore, as the set of finite measures, which are supported on a finite number of points, is dense in the space of finite measures (see [10, Theorem 4]), it follows that all RP processes may be obtained as the weak limit of such an aggregation.

Let Y_t and Z_t be uncorrelated processes denoting the deterministic component and nondeterministic component of X_t , respectively. The nondeterministic component Z_t has moving average representation

$$Z_t = \int_{-\infty}^t G(t-s) dB_s, \tag{2.2}$$

where $G \in L_2[0, \infty)$, B_t is an uncorrelated increment process, and the integral is interpreted in mean square. Let M denote the set of Borel measures on $(0, \infty)$ such that, for any $\mu \in M$,

$$S(\mu) = \left(\int_0^\infty \frac{1}{\lambda + \lambda'} \mu(d\lambda') \right) \mu(d\lambda) \in N.$$

Now assume that the moving average kernel $G(t)$ can be represented as the Laplace transform of some $\mu \in M$. A function defined on $(0, \infty)$ that can be represented in this form is called completely monotone. Interchanging the order of integration, it is seen that the covariance function of Z_t is given by

$$\gamma(t) = \int_0^\infty e^{-\lambda|t|} \left(\int_0^\infty \frac{1}{\lambda + \lambda'} \mu(d\lambda') \right) \mu(d\lambda).$$

Theorem 2.5 of [26] states that $S: M \mapsto N$ is bijective, so the nondeterministic component of any RP process has moving average representation with completely monotone kernel. It now follows that the deterministic component Y_t is a random variable with variance $\nu(\{0\})$ and is constant as a function of time. Unless otherwise stated, it will be assumed that $\nu(\{0\}) = 0$, so $Y = 0$ almost surely.

Of particular importance to modeling is the behavior of the covariance function as t approaches ∞ and as t approaches 0. The former determines the presence of long memory in the time series and can be inferred from the measure ν by application of an Abelian theorem for the Laplace transform. Applying Corollary 1b of [38], it is seen that if

$$\nu(\lambda) \sim \frac{A\lambda^\beta}{\Gamma(\beta + 1)} \quad \text{as } \lambda \rightarrow 0+$$

for some $A > 0$ and $\beta > 0$, then

$$\gamma(t) \sim At^{-\beta} \quad \text{as } t \rightarrow \infty.$$

A Tauberian theorem [38, Theorem 4.3] establishes the converse result, therefore, long memory for the RP process can be characterized by finiteness of the integral $\int_0^\infty \lambda^{-1} \nu(d\lambda)$ (see also [26] for a discussion of this point). It should be noted that even if the process does not possess long memory in the strict sense of the covariance not being in L_1 , it will still display some type of power-law decay unless ν is smaller than any polynomial in a neighborhood of 0. The behavior of the covariance function as t approaches 0 determines, at least in Gaussian processes, the smoothness of the sample paths. Recalling the results discussed in [1, p. 204], if $\gamma(0) - \gamma(t) \sim At^{2\beta}$ then the sample paths are Hölder continuous of order β and have Hausdorff dimension $2 - \beta$. For RP processes, the index β is bounded by $\frac{1}{2}$. This follows since

$$\gamma(0) - \gamma(t) = \int_0^\infty \frac{1 - e^{-\lambda t}}{\lambda t} \lambda t \nu(d\lambda) \geq ct$$

for some $c > 0$ depending only on ν . It is possible to infer β directly from ν by applying an Abelian theorem, so that if

$$\int_0^\lambda x \nu(dx) \sim \frac{A\lambda^{1-2\beta}}{\Gamma(2 - 2\beta)} \quad \text{as } \lambda \rightarrow \infty$$

then

$$\gamma(0) - \gamma(t) \sim At^{2\beta} \quad \text{as } t \rightarrow 0.$$

As noted in the introduction, aggregation can provide a physical interpretation for a number of complex models. We now give a few examples of models in the literature that possess the RP property and, hence, that can be represented as an aggregation. The first example is the class of processes introduced in [4], which are asymptotic stationary solutions to the fractional differential equation

$$(\mathcal{D}^{\beta_n} + A_{n-1}\mathcal{D}^{\beta_{n-1}} + \dots + A_0\mathcal{D}^{\beta_0})X_t = \dot{L}_t, \quad \beta_n > \dots > \beta_0 \geq 0, A_i > 0,$$

where \mathcal{D}^β is the Riemann–Liouville fractional derivative defined by

$$\mathcal{D}^\beta f(t) = \frac{1}{\Gamma(1-\beta)} \frac{d^n}{dt^n} \int_0^t (t-\tau)^{n-\beta-1} f(\tau) d\tau, \quad \beta \in [n-1, n),$$

and \dot{L}_t is the Lévy noise, that is the derivative of the Lévy process L_t in the distribution sense. In the simplest case we assume that $E L_t = 0$ and $E L_t^2 < \infty$, so that if $\beta_0 < \frac{1}{2}$ and $\beta_n > \frac{1}{2}$ then the solution has representation (2.2), where G is the Green function of a deterministic fractional differential equation and the integral is interpreted in mean square. The Green function has Laplace transform

$$g(p) = (p^{\beta_n} + A_{n-1}p^{\beta_{n-1}} + \dots + A_0p^{\beta_0})^{-1}.$$

It has been shown in Theorem 1 of [3] that the Green function is a completely monotone function if and only if $\beta_n \leq 1$ and, hence, it has the RP property.

Remark 2.1. At this point it is necessary to correct some statements concerning this class of processes. Anh *et al.* [4] proposed using asymptotic solutions to the fractional differential equations with L_t an increasing Lévy process to model the stochastic volatility. However, in this case, for the integral defining the solution to exist, it is necessary that $G \in L_1(0, \infty)$. When $G \in L_1(0, \infty)$, it follows that the covariance function is in $L_1[0, \infty)$ and so the process cannot possess long memory, contrary to the statements in Section 5 of [4]. Note that, for G to be in $L_1[0, \infty)$, it is necessary that $\beta_0 = 0$.

A further difficulty with their proposal is that, for the process to have nonnegative sample paths, it is necessary that the Green function is nonnegative. As the example given in Equation (5.6) of [4] was demonstrated to be ill-defined, consider the case given by their Equation (4.9). In this case $n = 1, \beta_0 = 0$, and the Green function is

$$G(t) = t^{\beta_1-1} E_{\beta_1, \beta_1}(-A_0 t^{\beta_1}), \quad t > 0,$$

where $E_{\alpha, \beta}(x)$ is the two parameter Mittag–Leffler function

$$E_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{\Gamma(\alpha k + \beta)}.$$

If $\beta_1 \leq 1$ then $G(t)$ is completely monotone and, hence, nonnegative. However, if $2 > \beta_1 > 1$ then applying Theorem 1.3-4 of [18] it is seen that, for sufficiently large t , $G(t) < 0$. In summary, unless we are able to establish the nonnegativity of $G(t)$ directly, we should restrict the parameters to $\beta_n \leq 1$ and $\beta_0 = 0$ if they are to be used in the modeling of stochastic volatility.

A related example is the model

$$\mathcal{D}^\beta(\alpha I + \mathcal{D})^\gamma X_t = \dot{B}_t, \quad \alpha > 0, \beta + \gamma > \frac{1}{2}, 0 < \beta < \frac{1}{2},$$

where I is the identity operator, which was proposed in [20]. The solution is interpreted as (2.2) with $G(t)$ having the Laplace transform $g(p) = p^{-\beta}(\alpha + p)^{-\gamma}$. From Equation 2.1.2-1 of [34],

$$G(t) = \frac{t^{\beta+\gamma-1}}{\Gamma(\beta + \gamma)} {}_1F_1(\gamma; \beta + \gamma; -\alpha t),$$

where ${}_1F_1$ is Kummer’s confluent hypergeometric function, and from Equation 3.33.1-3 of [34], it is seen that G is completely monotone for $\beta + \gamma \leq 1$. Thus, the process is RP under this additional restriction on the parameters.

Other examples include Lamperti’s transformation of fractional Brownian motion with $H \leq \frac{1}{2}$ (see [7] and [14]), certain time-reversible Markov chains [28, Subsection 7.4], and observations of some random fields from a fixed spatial location [2]. Finally, the KMO-Langevin equation studied in [26] and [32] provides a model which characterizes the class of RP processes.

3. Discrete time and sampled processes

From a practical perspective, even though X_t is a continuous-time process, it is only possible to observe it at discrete points in time. The simplest sampling scheme involves observing X_t without error at times $t = h, 2h, 3h, \dots$, where h is the sampling period. Let x_t denote the sampled time series and let $\{\gamma_k\}_0^\infty$ denote its covariance sequence. While for many processes the choice of sampling period can dramatically affect subsequent analysis of the time series, this is not the case for RP processes. It is a direct consequence of a result of [19] that the covariance function $\gamma(t)$ of the continuous-time process can be uniquely determined by $\{\gamma_k\}_0^\infty$. Without loss of generality, the subsequent analysis will assume that the sampling period is set to $h = 1$. Applying the change of variable $u = -\log \rho$ to (2.1), it is seen that $\{\gamma_k\}_0^\infty$ has the representation

$$\gamma_k = \int_0^1 \rho^{|k|} \sigma(d\rho) \quad \text{for all } k \in \mathbb{Z}, \tag{3.1}$$

where σ and ν are related by

$$\nu(\lambda) = -\sigma(e^{-\lambda}). \tag{3.2}$$

Since the support of ν is $[0, \infty)$, σ must be continuous at 0. Any time series with covariance function (3.1) can be represented as the aggregation of an AR(1) time series and, hence, as a sampled RP process. One example of this is the Cauchy model proposed in [21] ($\alpha \leq 1$), which is just a sampled version of the continuous-time RP process where ν is the distribution of the Mittag–Leffler process [30]. Another example is the ARIMA(0, d , 0), $0 < d < \frac{1}{2}$, model [22], although the measure ν does not have a simple form.

From the decomposition of the continuous-time process, an RP time series may have a deterministic component which is a random variable of variance $\sigma(\{1\})$ and constant over time. The spectral density of the nondeterministic component may be written as

$$f(\omega; \sigma) = \frac{1}{2\pi} \int_0^1 \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\omega)} \sigma(d\rho) = \int_0^1 g(\omega; \rho) \sigma(d\rho),$$

where $g(\omega; \rho)$ denotes the spectral density function of a standard AR(1) time series with parameter ρ . The spectral density can be bounded below by

$$f(\omega; \sigma) \geq \frac{1}{8\pi} (\gamma_0 - \gamma_2), \tag{3.3}$$

and, for ω in a neighborhood of 0, it is bounded above by

$$f(\omega; \sigma) \leq \frac{\gamma_0}{2\pi(1 - \cos^2(\omega))}. \tag{3.4}$$

While this paper focuses on instantaneous sampled processes, there are two other common sampling schemes. In random sampling the sampling times t_i are defined by $t_i = t_{i-1} + z_i$, where $t_0 = 0$ and the z_i are independent and identically distributed random variables. The covariance sequence for the observed time series is then

$$\gamma_k = \int_0^\infty \varphi_z(\lambda)^{|k|} \nu(d\lambda),$$

where φ_z is the Laplace transform of the distribution of z . Since φ_z is a completely monotone function, the covariance sequence has representation (3.1); however, σ and ν are now related by

$$\nu(\lambda) = -\sigma(\varphi_z(\lambda)).$$

The other common approach to sampling is local average sampling, that is, observing

$$x_t = c_h \int_{(t-1/2)h}^{(t+1/2)h} X_s ds$$

for some constant $c_h > 0$. This sampling scheme is particularly relevant to stochastic volatility modeling. It is known that when an Ornstein–Uhlenbeck process is sampled by local average sampling, the observations form an ARMA(1, 1) time series, where the AR parameter is the same as under instantaneous sampling and the MA parameter is constrained by the AR parameter. Therefore, under local average sampling, γ_k has representation (3.1) for some measure σ but only for $|k| \geq 1$. It is interesting to note that both the power-law model of [31] and the fractional Gaussian noise can be viewed as the result of local average sampling of an RP process.

4. Prediction error

Proposition 4.1, below, gives two basic properties of the one-step linear prediction error for a sampled RP process. These properties will be used in this section to determine bounds on the linear prediction error and in the following section to prove consistency of the minimum contrast estimate of the aggregating measure.

Proposition 4.1. *Let $\psi(\sigma)$ denote the one-step linear prediction error variance of an RP time series with measure σ .*

- (i) $\psi(\sigma)$ is a convex function of σ .
- (ii) If σ_m is a sequence of measures converging weakly to σ then $\psi(\sigma_m) \rightarrow \psi(\sigma)$.

Proof. Let $y_t^{(1)}$ and $y_t^{(2)}$ be two independent RP time series with measures σ_1 and σ_2 , respectively. For any $\xi \in [0, 1]$, define $x_t = \xi^{1/2} y_t^{(1)} + (1 - \xi)^{1/2} y_t^{(2)}$. Clearly, the measure

for x_t is $\xi\sigma_1 + (1 - \xi)\sigma_2$. It follows that $\psi(\xi\sigma_1 + (1 - \xi)\sigma_2)$ is bounded below by

$$\begin{aligned} \psi(\xi\sigma_1 + (1 - \xi)\sigma_2) &= \text{var}(x_{t+1} - \mathbb{E}(x_{t+1} \mid x_t, \dots)) \\ &\geq \text{var}(x_{t+1} - \mathbb{E}(x_{t+1} \mid y_t^{(1)}, y_t^{(2)}, \dots)) \\ &\geq \xi \text{var}(y_{t+1}^{(1)} - \mathbb{E}(y_{t+1}^{(1)} \mid y_t^{(1)}, \dots)) \\ &\quad + (1 - \xi) \text{var}(y_{t+1}^{(2)} - \mathbb{E}(y_{t+1}^{(2)} \mid y_t^{(2)}, \dots)) \\ &\geq \xi\psi(\sigma_1) + (1 - \xi)\psi(\sigma_2), \end{aligned}$$

where the first inequality follows as the σ -field generated by (x_t, x_{t-1}, \dots) is a subset of the σ -field generated by $(y_t^{(1)}, y_t^{(2)}, \dots, y_{t-1}^{(1)}, \dots)$, and the second inequality follows from the independence of $y_t^{(1)}$ and $y_t^{(2)}$. The proof of (i) is complete.

Now consider a sequence of measures σ_m converging weakly to σ . Using the Kolmogorov formula to determine $\psi(\sigma_m)$, we have

$$\psi(\sigma_m) = 2\pi \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega; \sigma) \, d\omega\right) \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left[\frac{f(\omega; \sigma_m)}{f(\omega; \sigma)}\right] \, d\omega\right).$$

We need to show that the second factor converges to 1 as σ_m converges weakly to σ . Take any $\varepsilon > 0$. As a family of functions of ρ on $[0, 1]$, parameterized by ω , $g(\omega; \rho)$ is bounded and equicontinuous for all $|\omega| > \varepsilon$. Applying Theorem 3.2 of [35] (see Appendix A), we have $f(\omega; \sigma_m) \rightarrow f(\omega; \sigma)$ uniformly on $|\omega| > \varepsilon$. From (3.3) and (3.4), $f(\omega; \sigma)$ is bounded from above and below, hence, we may apply the dominated convergence theorem to obtain

$$\lim_{m \rightarrow \infty} \int_{\varepsilon}^{\pi} \log \left[\frac{f(\omega; \sigma)}{f(\omega; \sigma_m)}\right] \, d\omega = 0.$$

From (3.3) and (3.4), there exist finite constants $c_1, c_2 > 0$ such that

$$\left| \log \left[\frac{f(\omega; \sigma_m)}{f(\omega; \sigma')}\right] \right| \leq c_1 |\log \omega| + c_2.$$

Applying the dominated convergence theorem, the integral

$$\int_0^{\varepsilon} \log \left[\frac{f(\omega; \sigma)}{f(\omega; \sigma_m)}\right] \, d\omega$$

can be made arbitrarily small by taking sufficiently small ε . The proof of (ii) is complete.

In [13] it was shown that given γ_0 and γ_1 the time series with the largest prediction error variance is a Gaussian AR(1) time series. This result implies that, for any finite Borel measure σ , $\psi(\sigma) \leq \gamma_0(1 - (\gamma_1/\gamma_0)^2)$. Furthermore, this upper bound is tight since an AR(1) time series has representation (3.1). Proposition 4.2, below, provides a tight lower bound on $\psi(\sigma)$.

Proposition 4.2. *Let X_t be a sampled RP process with measure σ . Then $\psi(\sigma) \geq \gamma_0 - \gamma_2$.*

Proof. Suppose that σ is an atomic measure with a finite number of atoms $\{\rho_i\}$ with masses $\{\xi_i/(1 - \rho_i^2)\}$. Let y_t^i be an independent AR(1) time series such that

$$y_{t+1}^i = \rho_i y_t^i + e_{t,i}, \quad e_{t,i} \sim N(0, 1).$$

The times series defined by $x_t = \sum_i \xi_i^{1/2} y_t^i$ has an RP covariance sequence determined by σ . From Proposition 4.1(i) we have the inequality $\psi(\sigma) \geq \sum_i \xi_i$. Noting that

$$\gamma_0 = \sum_i \xi_i (1 - \rho_i^2)^{-1} \quad \text{and} \quad \gamma_2 = \sum_i \xi_i \rho_i^2 (1 - \rho_i^2)^{-1},$$

the inequality is proved for finite atomic measures. This can be extended to all Borel measures on $[0, 1]$ by considering a sequence of finite atomic measures converging weakly to σ with the same zeroth and second moment and applying Proposition 4.1(ii). This completes the proof.

In an attempt to determine tighter lower bounds we could consider the set of measures for which $\gamma_k = c_k, k = 0, \dots, m$, and then minimize $\psi(\sigma)$ over this set. Proposition 4.3, below, gives the general structure of an RP time series with minimum prediction error.

Proposition 4.3. *The RP time series satisfying $\gamma_k = c_k, k = 0, 1, \dots, m$, and having the smallest prediction error has a measure which is concentrated on no more than $m + 1$ points.*

Proof. Let $\mathcal{D} | c$ denote the set of measures satisfying the constraints, and let $\mathcal{D}_A | c$ denote the subset of measures whose mass is concentrated on a finite number of points. The extreme points of $\mathcal{D}_A | c$ are those measures whose mass is concentrated on no more than $m + 1$ points; see [27, Theorem 21.1]. As ψ is a convex function and $\mathcal{D}_A | c$ is a convex set, the minimum of ψ on $\mathcal{D}_A | c$ occurs at one of the extreme points. Now let σ^* be a measure which minimizes ψ on $\mathcal{D} | c$, and let σ_n^* be a sequence of measures in $\mathcal{D}_A | c$ converging weakly to σ^* . From Proposition 4.1(ii), for $\varepsilon > 0$, there exists an N such that, for all $n \geq N, \psi(\sigma^*) + \varepsilon \geq \psi(\sigma_n^*)$. Noting that there exists a measure $\tilde{\sigma}$ concentrated on no more than $m + 1$ points such that $\psi(\sigma_n^*) \geq \psi(\tilde{\sigma})$, and letting ε tend to 0, completes the proof.

This result can be applied to determine a tight lower bound on $\psi(\sigma)$ using γ_0 and γ_1 . From Proposition 4.3, the measure which minimizes $\psi(\sigma)$ subject to the constraint on γ_0 and γ_1 has its mass concentrated on no more than two points. Minimizing $\psi(\sigma)$ yields the measure $\sigma^*(d\rho) = (\gamma_0 - \gamma_1)\delta(\rho) d\rho + \gamma_1\delta(\rho - 1) d\rho$. The corresponding time series is $X_t = Y_t + Z$, where Y_t is white noise with variance $\gamma_0 - \gamma_1$ and Z is a random variable with variance γ_1 . The prediction error of this time series is obviously $\gamma_0 - \gamma_1$, and so we have the bound $\psi(\sigma) \geq \gamma_0 - \gamma_1$. Note that this bound does not improve on the bound in Proposition 4.2 since $\gamma_0 - \gamma_2 \geq \gamma_0 - \gamma_1$ for any RP time series.

5. Minimum contrast estimation

Now consider the problem of estimating the measure ν from a sampled RP process. From (3.2), this can be achieved by estimating σ in (3.1) and then applying the transformation. As noted in the introduction, estimation of ν can be important in certain applications where the unit-level models can be interpreted. As a simple example, consider the employment status of an individual as being modeled by a continuous-time, homogeneous, two-state $\{0, 1\}$ Markov chain. The correlation function for an individual’s employment history is of the form $e^{-\lambda|t|}$ for some $\lambda > 0$ and, hence, the aggregate employment rate in the economy would have a correlation function of the form (2.1). Being able to determine the aggregating measure would allow certain questions to be addressed such as ‘what percentage of the population are unemployed for more than x weeks in a year?’. While this model ignores certain important features of unemployment, it does demonstrate the importance of estimating the aggregating measure.

It is noted that Leipus *et al.* [29] have recently considered estimation of the aggregating measure for sums of AR(1) time series where the correlation parameter can take any value in $(-1, 1)$. Assuming that the measure has a density, they studied the convergence of the density in the L_2 norm and uniform norm. While this is interesting, the assumption of a density excludes a number of important cases, in particular the simple AR(1), and weak convergence of the estimated measure may be more relevant.

Let x_t be the sampled RP process where the covariance sequence has representation (3.1) with measure σ^* . In this section we consider the estimate of σ^* obtained from the minimization of the contrast function

$$U_n(\sigma) = \frac{1}{n} \sum_j \frac{I_n(\omega_j)}{f(\omega_j; \sigma)} + \int_{-\pi}^{\pi} \log f(\omega; \sigma) d\omega, \tag{5.1}$$

over the set of finite Borel measures on $[0, 1]$, where

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n e^{-i\omega t} x_t \right|^2$$

and ω_j are the Fourier frequencies. In order to prove weak convergence of the sequence of measures minimizing (5.1), we need to introduce a mild assumption on x_t . Let $\hat{F}_n(\omega)$ and $\hat{\gamma}_k^{(n)}$ denote the empirical spectral distribution function and empirical covariance, respectively, where

$$\hat{F}_n(\omega) = \frac{1}{n} \sum_{\omega_j < \omega} I_n(\omega_j) \quad \text{and} \quad \hat{\gamma}_k^{(n)} = \frac{1}{n} \sum_{t=1}^{n-|k|} x_t x_{t+|k|}.$$

We introduce the following assumption.

Assumption 5.1. *For any $\omega \in (-\pi, \pi)$, $\hat{F}_n(\omega)$ is a consistent estimate of $F(\omega)$ and, for any $k \in \mathbb{Z}$, $\hat{\gamma}_k^{(n)}$ is a consistent estimate of γ_k .*

This assumption is satisfied by any sampled completely nondeterministic Gaussian RP process. Sufficient conditions for Assumption 5.1 to hold in non-Gaussian settings can be found, for example, in [11, Subsection 5.10]. The proofs of the following results parallel the corresponding results for parametric spectral densities (see [12, Proposition 10.8.2 and Theorem 10.8.1]).

Proposition 5.1. *Let Σ_ξ be the subset of finite Borel measures on $[0, 1]$ such that*

$$\int_0^1 (1 - \rho^2) \sigma(d\rho) \geq \xi^{-1} \quad \text{and} \quad \int_0^1 \sigma(d\rho) \leq \xi.$$

Then, under Assumption 5.1,

$$\frac{1}{n} \sum_j \frac{I_n(\omega_j)}{f(\omega_j; \sigma)} \rightarrow \int_{-\pi}^{\pi} \frac{f(\omega; \sigma^*)}{f(\omega; \sigma)} d\omega \quad \text{a.s.}$$

uniformly on Σ_ξ .

Proof. Define $q_m(\omega; \sigma)$ as the Cesaro mean of the first m Fourier approximations to $f^{-1}(\omega; \sigma)$. This may be written as

$$q_m(\omega; \sigma) = \sum_{|k| < m} \left(1 - \frac{|k|}{m}\right) b_k e^{-ik\omega} = \int_{-\pi}^{\pi} f^{-1}(\omega - y; \sigma) K_m(y) dy,$$

where $b_k = (2\pi)^{-1} \int e^{ik\omega} f^{-1}(\omega; \sigma) d\omega$ and K_m is the Fejer kernel. The main part of this proof involves determining how well $f^{-1}(\omega; \sigma)$ is approximated by $q_m(\omega; \sigma)$. For any $\varepsilon > 0$,

$$\begin{aligned} &|q_m(\omega; \sigma) - f^{-1}(\omega; \sigma)| \\ &\leq \int_{(-\varepsilon, \varepsilon)} |f^{-1}(\omega - y; \sigma) - f^{-1}(\omega; \sigma)| K_m(y) dy \\ &\quad + \int_{(-\varepsilon, \varepsilon)^c} |f^{-1}(\omega - y; \sigma) - f^{-1}(\omega; \sigma)| K_m(y) dy \\ &\leq \int_{(-\varepsilon, \varepsilon)} |f^{-1}(\omega - y; \sigma) - f^{-1}(\omega; \sigma)| K_m(y) dy + 16\pi\xi \left(1 - \int_{-\varepsilon}^{\varepsilon} K_m(y) dy\right), \end{aligned}$$

where the second inequality follows from (3.3). A modulus of continuity must be determined for $f^{-1}(\omega; \sigma)$ when $|\omega| \geq \phi > 0$. Again, using (3.3),

$$\begin{aligned} &|f^{-1}(\omega - y; \sigma) - f^{-1}(\omega; \sigma)| \\ &\leq 64\pi^2 \xi^2 |f(\omega - y; \sigma) - f(\omega; \sigma)| \\ &\leq 64\pi^2 \xi^2 \int_0^1 \left| \frac{2\rho(\cos(\omega) - \cos(\omega - y))(1 - \rho^2)}{(1 + \rho^2 - 2\rho \cos(\omega))(1 + \rho^2 - 2\rho \cos(\omega - y))} \right| \sigma(d\rho) \\ &\leq C(\phi) \xi^2 |y|, \end{aligned}$$

where $C(\phi)$ is a constant depending only on ϕ . Therefore, for all $\sigma \in \Sigma_\xi$, $|\omega| \geq \phi > 0$, and $\varepsilon > 0$,

$$|q_m(\omega; \sigma) - f^{-1}(\omega; \sigma)| \leq C(\phi) \xi^2 \varepsilon + 16\pi\xi \left(1 - \int_{-\varepsilon}^{\varepsilon} K_m(y) dy\right) \leq C(\phi, \varepsilon, m, \xi),$$

where $C(\phi, \varepsilon, m, \xi) \rightarrow 0$ for all $\phi > 0$ and $\xi < \infty$ as $m \rightarrow \infty$ and $\varepsilon \rightarrow 0$ at an appropriate rate. The error incurred by substituting $q_m(\omega; \sigma)$ for $f^{-1}(\omega; \sigma)$ in the contrast function is

$$\left| \frac{1}{N} \sum_j I(\omega_j) (f^{-1}(\omega_j; \sigma) - q_m(\omega_j; \sigma)) \right| \leq C(\phi, \varepsilon, m, \xi) \hat{F}_n(\pi) + 16\pi\xi (\hat{F}_n(\phi) - \hat{F}_n(-\phi)). \tag{5.2}$$

Under Assumption 5.1, \hat{F}_n is consistent and so this error can be made arbitrarily small by an appropriate sequence in $(m, \varepsilon, \phi) \rightarrow (\infty, 0, 0)$. From the definition and (3.3), it is seen that $|b_k| \leq 8\pi\xi$ for all $\sigma \in \Sigma_\xi$. The empirical covariance $\hat{\gamma}_k^{(n)}$ is a consistent estimate and so following the parametric case (see [12, p. 379])

$$\frac{1}{N} \sum_j I(\omega_j) q_m(\omega_j; \sigma) \rightarrow \sum_{|k| < m} \gamma_k \left(1 - \frac{|k|}{m}\right) b_k \quad \text{a.s.} \tag{5.3}$$

uniformly on Σ_ξ . Finally,

$$\begin{aligned} & \left| \sum_{|k|<m} \gamma_k \left(1 - \frac{|k|}{m}\right) b_k - \int_{-\pi}^{\pi} \frac{f(\omega; \sigma^*)}{f(\omega; \sigma)} d\omega \right| \\ & \leq \int_{-\pi}^{\pi} |q_m(\omega; \sigma) - f^{-1}(\omega; \sigma)| f(\omega; \sigma^*) d\omega \\ & \leq 4 \int_{-\phi}^{\phi} f(\omega; \sigma^*) d\omega + \varepsilon C(\phi) \xi \int_{-\pi}^{\pi} f(\omega; \sigma^*) d\omega. \end{aligned} \tag{5.4}$$

Combining (5.2)–(5.4), the proposition is established.

Theorem 5.1. *Let σ_n be the sequence of measures on Σ_ξ which minimize the contrast function. If $\sigma^* \in \Sigma_\xi$ then σ_n converges weakly to σ^* .*

Proof. Suppose that σ_n does not converge weakly to σ^* . Then, by the Helly–Bray lemma (see, for example, [39, Section 17.4]), there is a weakly convergent subsequence σ_{n_k} such that $\sigma_{n_k} \rightarrow \sigma'$:

$$\begin{aligned} & |U_{n_k}(\sigma_{n_k}) - U_{n_k}(\sigma')| \\ & \leq \left| \frac{1}{n_k} \sum_j I(\omega_j) \frac{f(\omega_j; \sigma_{n_k}) - f(\omega_j; \sigma')}{f(\omega_j; \sigma_{n_k}) f(\omega_j; \sigma')} \right| + \left| \int_{-\pi}^{\pi} \log \left[\frac{f(\omega; \sigma_{n_k})}{f(\omega; \sigma')} \right] d\omega \right|. \end{aligned} \tag{5.5}$$

From Proposition 4.1(ii), if σ_{n_k} converges weakly to σ' then the second term in (5.5) converges to 0. For any $\phi > 0$, the first term in (5.5) is bounded above by

$$64\pi^2 \xi^2 \sup_{\omega \notin (-\phi, \phi)} |f(\omega_j; \sigma_{n_k}) - f(\omega_j; \sigma')| \hat{F}_{n_k}(\pi) + 16\pi \xi (\hat{F}_{n_k}(\phi) - \hat{F}_{n_k}(-\phi)). \tag{5.6}$$

From Theorem A.1 in Appendix A, $f(\omega_j; \sigma_{n_k})$ converges uniformly to $f(\omega_j; \sigma')$ on $|\omega| \geq \phi > 0$. Thus, the first term in (5.6) converges to 0 as n_k tends to ∞ . As ϕ is arbitrary and \hat{F} is consistent, it then follows from Proposition 5.1 that

$$U_{n_k}(\sigma_{n_k}) \rightarrow \int_{-\pi}^{\pi} \frac{f(\omega; \sigma^*)}{f(\omega; \sigma')} d\omega + \int_{-\pi}^{\pi} \log f(\omega; \sigma') d\omega \quad \text{a.s.}$$

and, hence, $U_n(\sigma^*) \rightarrow 2\pi + \int_{-\pi}^{\pi} \log f(\omega; \sigma^*) d\omega$ a.s. As

$$x - 1 - \log(x) \geq 0 \quad \text{for all } x > 0$$

with equality if and only if $x = 1$, then, for sufficiently large n_k , $U_{n_k}(\sigma^*) < U_{n_k}(\sigma_{n_k})$. A contradiction is obtained and this completes the proof.

Corollary 5.1. *The estimate of the spectral density $f(\omega; \sigma_n)$, where σ_n is the measure which minimizes (5.1), is consistent in L_1 for $f(\omega; \sigma^*)$.*

Proof. This result follows immediately from the weak consistency of the estimate of σ^* and the fact that $g(\omega; \rho)$ is a symmetric unimodal and equicontinuous function for all $|\omega| > \phi$, any $\phi > 0$, and for all $\rho \in [0, 1]$.

Remark 5.1. To apply Theorem 5.1 to the estimation of σ^* , it is necessary to determine a ξ such that $\sigma^* \in \Sigma_\xi$. It is possible to replace ξ by $\xi_n = 2((\hat{\gamma}_0^{(n)} - \hat{\gamma}_2^{(n)})^{-1} \vee \hat{\gamma}_0^{(n)})$. Weak consistency will hold for this choice of ξ by application of the Borel–Cantelli lemma.

Remark 5.2. The minimization of (5.1) is a difficult problem as we are minimizing over a set of measures. From the proof of Theorem 5.1, it is seen that (5.1) is a continuous function of σ with respect to the metric of weak convergence. It is also known that any measure can be approximated in the weak sense by one comprised of a finite number of point masses. Therefore, (5.1) can be minimized over the set of measures formed from a finite number of point masses. Standard algorithms can now be applied to minimize (5.1) over this set. The number of point masses can be increased to check for convergence, at least to a local minima.

Appendix A.

Theorem A.1. (Theorem 3.2 of [35].) *Let \mathcal{A} be a class of functions on the separable metric space X possessing the following properties:*

- (i) \mathcal{A} is uniformly bounded, i.e. there exists a constant M such that $|f(x)| \leq M$ for all $f \in \mathcal{A}$ and $x \in X$;
- (ii) \mathcal{A} is equicontinuous.

If $\mu_n, \mu \in \mathcal{M}$ and $n = 1, 2, \dots$ (\mathcal{M} is the class of measures on the Borel subsets of X), then μ_n implies μ if and only if, for each family \mathcal{A} satisfying (i) and (ii) above, we have

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{A}} \left| \int f d\mu_n - \int f d\mu \right| = 0.$$

Acknowledgements

This research was supported by the ARC Centre for Complex Dynamic Systems and Control CEO348165. The author gratefully acknowledges helpful discussions with N. Leonenko and K. Mengersen.

References

- [1] ADLER, R. J. (1981). *The Geometry of Random Fields*. John Wiley, New York.
- [2] ANGULO, J. M., ANH, V. V., MCVINISH, R. AND RUIZ-MEDINA, M. D. (2005). Fractional kinetic equations driven by Gaussian and infinitely divisible noise. *Adv. Appl. Prob.* **37**, 366–392.
- [3] ANH, V. V. AND MCVINISH, R. (2003). Completely monotone property of fractional Green functions. *Frac. Calc. Appl. Anal.* **6**, 157–173.
- [4] ANH, V. V., HEYDE, C. C. AND LEONENKO, N. N. (2002). Dynamic models of long memory processes driven by Lévy processes. *J. Appl. Prob.* **39**, 730–747.
- [5] BARNDORFF-NIELSEN, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2**, 41–68.
- [6] BARNDORFF-NIELSEN, O. E. AND LEONENKO, N. N. (2005). Spectral properties of superpositions of Ornstein–Uhlenbeck type processes. *Methodology Comput. Appl. Prob.* **7**, 335–352.
- [7] BARNDORFF-NIELSEN, O. E. AND PÉREZ-ABREU, V. (1999). Stationary and self-similar processes driven by Lévy processes. *Stoch. Process. Appl.* **84**, 357–369.
- [8] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2001). Non-Gaussian Ornstein–Uhlenbeck based models and some of their uses in financial econometrics (with discussion). *J. R. Statist. Soc. Ser. B* **63**, 167–241.
- [9] BARNDORFF-NIELSEN, O. E., JENSEN, J. L. AND SØRENSEN, M. (1998). Some stationary processes in discrete and continuous time. *Adv. Appl. Prob.* **30**, 989–1007.
- [10] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- [11] BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, CA.
- [12] BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.
- [13] BURG, J. P. (1975). *Maximum Entropy and Spectral Analysis*. Doctoral Thesis, Stanford University.
- [14] CHERIDITO, P., KAWAGUCHI, H. AND MAEJIMA, M. (2003). Fractional Ornstein–Uhlenbeck processes. *Electron. J. Prob.* **8**, 1–14.

- [15] COMTE, F. AND RENAULT, E. (1996). Long memory continuous-time models. *J. Econometrics* **73**, 101–149.
- [16] COX, D. R. (1991). Long-range dependence, non-linearity and time irreversibility. *J. Time Ser. Anal.* **12**, 329–335.
- [17] DACUNHA-CASTELLE, D. AND FERMÍN, L. (2006). Disaggregation of long memory processes on C^∞ class. *Electron. Commun. Probab.* **11**, 35–44.
- [18] DIRBASHIAN, M. M. (1993). *Harmonic Analysis and Boundary Value Problems in the Complex Domain*. Birkhäuser, Basel.
- [19] FELLER, W. (1939). Interpolation of completely monotone functions. *Duke Math. J.* **5**, 661–674.
- [20] GAO, J. (2004). Modelling long-range dependent Gaussian processes with application in continuous-time financial models. *J. Appl. Prob.* **41**, 467–482.
- [21] GNEITING, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *J. Appl. Prob.* **37**, 1104–1109.
- [22] GRANGER, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14**, 227–238.
- [23] GRANGER, C. W. J. AND JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1**, 15–29.
- [24] HIDA, T. AND STEIT, L. (1977). On quantum theory in terms of white noise. *Nagoya Math. J.* **68**, 21–34.
- [25] IGLOI, E. AND TERDIK, G. (1999). Long-range dependence through Gamma-mixed Ornstein–Uhlenbeck processes. *Electron. J. Probab.* **4**, 1–33.
- [26] INOUE, A. (1993). On the equations of stationary processes with divergent diffusion coefficients. *J. Fac. Sci. Univ. Tokyo Sect. IA* **40**, 307–336.
- [27] KARLIN, S. AND SHAPLEY, L. S. (1953). *Geometry of Moment Spaces* (Amer. Math. Soc. Memoirs **12**). American Mathematical Society, Providence, RI.
- [28] KEILSON, J. (1979). *Markov Chain Models—Rarity and Exponentiality* (Appl. Math. Sci. **28**). Springer, New York.
- [29] LEIPUS, R., OPPENHEIM, G., PHILIPPE, A. AND VIANO, M.-C. (2006). Orthogonal series density estimation in a disaggregation scheme. *J. Statist. Planning Infer.* **136**, 2547–2571.
- [30] LIN, G. D. (1998). On the Mittag–Leffler distributions. *J. Statist. Planning Infer.* **74**, 1–9.
- [31] MARTIN, R. J. AND WALKER, A. M. (1997). A power-law model and other models for long-range dependence. *J. Appl. Prob.* **34**, 657–670.
- [32] OKABE, Y. (1986). On KMO-Langevin equations for stationary Gaussian process with T-positivity. *J. Fac. Sci. Univ. Tokyo Sect. IA* **33**, 1–56.
- [33] OSTERVALDER, K. AND SCHRADER, R. (1973). Axioms for Euclidean Green functions. *Commun. Math. Phys.* **31**, 83–112.
- [34] PRUDIKOV, A., BRYCHKOV, Y. AND MARICHEV, O. (1990). *Integrals and Series*, Vol. 5. Gordon and Breach, New York.
- [35] RANGA RAO, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **32**, 659–680.
- [36] TAQQU, M. S., WILLINGER, W. AND SHERMAN, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Comput. Commun. Rev.* **27**, 5–23.
- [37] VIANO, M.-C., DENIAU, C. AND OPPENHEIM, G. (1995). Long-range dependence and mixing for discrete time fractional processes. *J. Time Series Anal.* **16**, 323–338.
- [38] WIDDER, D. V. (1941). *The Laplace Transform*. Princeton University Press.
- [39] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge University Press.
- [40] ZAFFARONI, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *J. Econometrics* **120**, 75–82.