Should we sample a time series more frequently? Decision support via multirate spectrum estimation

Guy P. Nason

University of Bristol, Bristol, UK.

Ben Powell

University of Bristol, Bristol, UK.

Duncan Elliott

Office for National Statistics, Newport, UK.

Paul Smith

University of Southampton, Southampton, UK.

Summary. Suppose we have a historical time series with samples taken at a slow rate, e.g. quarterly. This article proposes a new method to answer the question: is it worth sampling the series at a faster rate, e.g. monthly? Our contention is that classical time series methods are designed to analyse a series at a single and given sampling rate with the consequence that analysts are not often encouraged to think carefully about what an appropriate sampling rate might be. To answer the sampling rate question we propose a novel Bayesian method that incorporates the historical series, cost information and small amounts of pilot data sampled at the faster rate. The heart of our method is a new Bayesian spectral estimation technique that is capable of coherently using data sampled at multiple rates and is demonstrated to have superior practical performance compared to alternatives. Additionally, we introduce a method for hindcasting historical data at the faster rate. A freeware R package, regspec, is available that implements our methods. We illustrate our work using official statistics time series including the United Kingdom consumer price index and counts of United Kingdom residents travelling abroad, but our methods are general and apply to any situation where time series data are collected.

Keywords: Aliasing, Bayesian statistics, Multirate, Spectrum estimation, Time series.

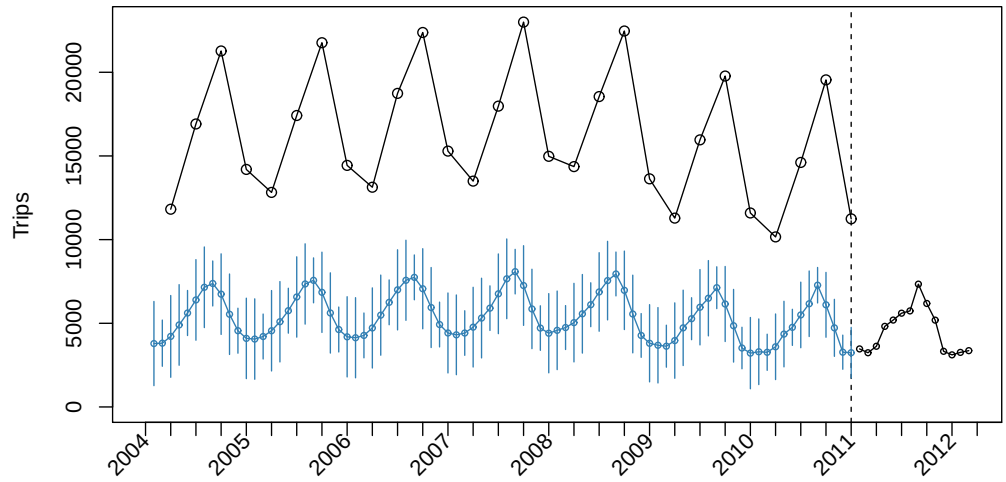

Fig. 1. Black points represent quarterly and monthly (estimated) counts of trips abroad made by UK residents. Blue points are predictions (hindcasts, explained later) for the monthly series given all the data and the estimated spectrum. The vertical blue lines centred on these points have lengths four times the prediction standard deviation here. The solid lines interpolating the points are intended solely to guide the eye, while the vertical dashed line indicates when the sampling rate was increased. Source: Office for National Statistics: Monthly Overseas Travel and Tourism, May 2014; Overseas Travel and Tourism Q1 2014.

1. Introduction

1.1. Practical context

In time series analysis it is well known that the sampling rate is an important consideration. If one samples infrequently, then information that changes rapidly (high frequencies) can be missed. On the other hand, if one samples too frequently then one runs the risk of paying for sampling and storage of unnecessarily detailed information.

Often, one has access to a time series sampled at a slow rate and a proposal is made to obtain more information by sampling at a faster rate. An equivalent situation arises when we consider rates and costs for publishing or communicating observations. Records of the number of United Kingdom (UK) residents travelling abroad, see Figure 1, provide one such example. One cannot decrease the amount of information obtained by collecting samples more frequently. However, as increasing the sampling rate typically has real extra cost, the question "Is it worth sampling more frequently?" is a valid one. Further analysis of the series of trips abroad by UK residents appears in Example 3 below.

The concept of 'worth' is, in general, complex and situation dependent, depending not only on the measured variable in question but also the utility function of the observer and other stakeholders. Section 3 introduces a model that accounts for the various costs involved in sample rate switching, and uses it to assess the resulting net benefit or loss of doing so, in a rational decision-theoretic manner. This is achieved by concentrating on the second-order statistics of the process, with which we describe the joint population of observed and unobserved data. For second-order stationary time series these translate into the mean and the autocovariance, or its Fourier dual, the spectral density function, or spectrum. We focus, in particular, on the latter as its behaviour under different sampling rates is well understood.

For this paper we assume that the underlying quantity of interest is a (second-order) stationary stochastic process X_t sampled on the integers t with finite mean and variance. Now, purely for the sake of exposition, since we will eventually consider more general sampling schemes, suppose also that, maybe due to financial or technical constraints, we have only been able to collect a long series of observations at the even integers, $y_t = X_{2t}$, and there is a proposal to move to sampling $x_t = X_t$ on the integers. We would like to know the true spectrum $f_X(\omega)$ where ω is a frequency variable with domain $[0, \frac{1}{2})$.

Since y_t is collected on the even integers, with sampling rate half that of the proposed x_t, an estimator of the spectrum based on y_t, $\hat{f}_y(\omega)$ say, is typically considered to 'see' only frequencies in the range $[0, \frac{1}{4})$ because, due to the well-known phenomenon of aliasing (discussed at length in Section 2), the top half of the frequencies in X_t are folded onto and added to the bottom half. Clearly, the spectrum we want, $f_X(\omega)$, cannot be uniquely identified from $f_Y(\omega)$.

On the other hand, if we *only* use data sampled at the faster rate, which we will refer to as pilot or trial data, it could take some time before we can collect enough to obtain a reliable estimate of $f_X(\omega)$. A key statistical innovation of this article enables the incorporation of both the new (short) pilot data *and* the less frequently sampled established series *together* to obtain an estimate of $f_X(\omega)$ better than those obtained with competing methods. Section 2 reviews aliasing and describes our new method for spectral estimation for multirate time series data.

There appears to be a relative paucity of classical statistical theory to support the many analysts who increasingly find themselves in receipt of time series at multiple rates. However, recently, the analysis of mixed frequency data has received a great deal of attention, mostly in econometrics. For example, mixed frequency data is used for forecasting in Ghysels et al. (2006), Clements and Galvão (2008), Armesto et al. (2010), Andreou et al. (2011), Schorfheide and Song (2015); for VAR modelling in Foroni et al. (2013), Cimadomo and D'Agostino (2015), Eraker et al. (2015), Foroni et al. (2015), Götz and Hauzenberger (2015); and for assessing Granger causality in Ghysels et al. (2015b), Ghysels et al. (2015a). A recent overview of models for mixed frequency data is given by Foroni and Marcellino (2013). The goal of our paper is completely different in that we use a frequency domain approach to estimate spectra based on data taken at multiple rates, often with small amounts of faster data, and try to ascertain whether it is worth continuing to sample at the faster rate.

1.2. Modelling Strategy

Our strategy for modelling time series in the type of situation sketched out in Section 1.1, centres on what we will call *linear Bayes spectrum estimation*. This section briefly lays out our motivations for adopting this strategy, whilst the rest of the paper reinforces them.

Firstly, we approach the parameterization of second-order stationary time series in terms of a spectral density rather than, e.g., a set of seasonal autoregressive integrated moving average (SARIMA) parameters because the aliasing phenomenon, which will turn out to be so dominant in our problem, is well understood and easily described in the frequency domain. In the time domain, where the SARIMA parameters effectively live, it is not.

Secondly, we adopt a Bayesian mindset for the project as a whole because of its capacity to coherently structure inferences combining heterogeneous sources of information, and to make sense of inferences as components of decision problems.

Thirdly, we employ simple and fast linear Bayes methods, described in Goldstein and Wooff (2007), to structure our inferences because, in the absence of expedient conjugate relationships, implementation of an exact full Bayes inference for a high-dimensional object, such as a spectrum within the span of a large number of basis functions, threatens to become computationally burdensome. Certainly, a Monte Carlo method could also be used for approximating full Bayes estimates. The scale and subtlety of the work required for implementation in such problems is discussed in Ceperley et al. (2012), who highlight, in particular, the unresolved challenges faced when dealing with multimodal posterior density functions in high dimension. Example 2, which we present in Section 2.5, exhibits both features. As with the classical periodogram, we will also want to re-fit our models with more degrees of freedom as more data become available. Incorporating this would necessitate a reversible jump-like step, which further complicates issues of algorithm mixing and convergence, as discussed in Green and Hastie (2009).

One can identify a number of previous contributions to Bayesian spectrum estimation, ranging from the classical work of Whittle (1957) to contemporary Bayesian analyses including Mallick et al. (2002), Choudhuri et al. (2004), Pensky et al. (2007) and Rousseau et al. (2012). We would, however, place our work closest to the, chronologically intermediate and not expressly Bayesian, papers from Jones (1976) and Wahba (1980). Their application of simple, yet powerful smoothing formulae to the log-spectrum is built upon here in the way we apply similar formulae to the assimilation of data at different sampling rates, and in our interpretation of mean squared error-type statistics as posterior or adjusted variances, which allow us to quantify expected costs associated with uncertainties for the spectrum.

1.3. Review of Spectrum Estimation for Different Types of Missing Data

Multirate data can be viewed as an example of single rate data with systematically missing observations. Good reviews of established methods for spectrum estimation with missing data can be found in Broersen and Bos (2006) or Section 10.4 of Broersen (2006) who divide the field into three categories.

The first is Least-Squares Spectral Analysis (LSSA), developed by Vaníček (1971),

and also known as the Lomb-Scargle periodogram of Lomb (1976) and Scargle (1982). Broersen and Bos (2006) state that this method works well for processes with strong sinusoidal behaviour contaminated with additive noise, reproducing spectral peaks accurately but not spectral slopes. Conventional implementations of LSSA ascertain an operational frequency-range according to an average spacing between all (fast- and slow-sampled) available data. They then proceed to compute periodogram-type statistics by projecting the data onto a set of sinusoids spread over these frequencies. As a result, and by default, LSSA methods do begin to look for higher frequencies as the fast-sampled data are introduced by gradually moving the upper bound on the spectrum's effective support upwards. This is far too slow for us, however: as soon as we know that the sampling rate may be doubled, regardless of the number of fast-sampled data we currently have, we know that we should be thinking about the spectrum at all frequencies identifiable at the higher rate. To be clear, this is not a criticism of the Lomb-Scargle method itself, but of its applicability to the type of problem we are most interested in.

The second category centres around state space models, which encompass all SARIMA models. These are naturally untroubled by missing data, which are effectively integrated out as and when the filter-type algorithm, responsible for computing parameter likelihoods, reaches them. The problem here, which we discuss at length in Nason et al. (2014), is that without explicit formulation of a process's spectrum we have no way to encode and prepare for the known consequences of the aliasing phenomenon. Postponing a mathematical description, one can probably recall, from watching car wheels turn on a television programme perhaps, how the subsampling of a highly periodic signal gives rise to illusory oscillations far below the frequencies of the original signal. It will turn out that the true and illusory frequencies correspond to well-separated modes in the likelihood that state-space calibration methods conventionally rely on. Without explicit knowledge and consideration of this multimodality, all likelihood-based SARIMA methods, from likelihood maximization to sophisticated Markov chain Monte Carlo (MCMC), run the risk of favouring one mode over another purely on the basis of arbitrary initial conditions and/or random MCMC proposals. Data sampled at a higher rate ought to gradually smooth out one of these modes, but will only slowly alleviate this fundamental problem.

The third group of methods relies on expectation-maximisation algorithms to repeatedly impute missing data based on a current estimate for the spectrum, before re-estimating that spectrum with conventional methods, see Wang et al. (2005) for an example. Such alternating estimation procedures may also be adapted into Gibbs samplers, leading to the possibility of spectrum variance estimation via a MCMC sample of simulated spectra. By employing a spectral parameterization, the distinct modes of the SARIMA likelihood become continuous ridges that reflect the idea that spectral mass can be transferred directly from one aliased frequency to another. The ambiguity for the second-order structure is now better behaved, but is still present and provides a large volume for stochastic inference algorithms to explore. For algorithms that are very fast or ingeniously adaptive such exploration need not be demanding, but we will argue that the aliasing-induced ambiguity does not need thorough exploration because we already understand it perfectly.

The next section introduces our new method which is tailored for the slow/fast sampling scheme that we are considering. Its superiority over the first two established methods is

explored below in Example 2. We will not, however, compare it with the third group of methods since we do not see them as competing models, but rather as a large class of alternate numerical methods for approximating the same Bayesian inference.

2. Infrequent Sampling, Aliasing and Spectral Estimation

2.1. Aliasing

Firstly, we establish notation for a long history of observations, with a subscript denoting low sampling rate, taken at every K^{th} time point,

$$D_{low} = \{x_K, x_{2K}, \ldots, x_{N_{low}K}\}, \tag{1}$$

and the subsequent series of observations sampled at integer time points,

$$D_{high} = \{x_S, x_{S+1}, x_{S+2}, \ldots, x_{S-1+N_{high}}\}.$$

where S is an integer used to locate D_{high} in time, and N_{low} and N_{high} give the lengths of the series.

The key conceptual object for describing the variability within and between these datasets is the spectrum (scaled to exist on $(0, 1/2)$ here, though $(0, \pi)$ or $(-\pi, \pi)$ are often-used alternatives). As is well-known, the spectrum, f, is related to the autocovariance function, γ, for a stationary process according to

$$f(\omega) = 2\gamma(0) + 4\sum_{\tau=1}^{\infty} \gamma(\tau)\cos(2\pi\tau\omega), \qquad \omega \in (0, 1/2) \tag{2}$$

$$\gamma(\tau) = \mathrm{cov}(x_t, x_{t+\tau}) = 2\pi \int_0^{1/2} f(\omega)\cos(2\pi\tau\omega)dw, \qquad \tau \in \mathbb{N}, \tag{3}$$

and variants of these formulae are found in the literature.

The effect of under-sampling a stationary process is often interpreted in terms of the 'aliasing phenomenon', which causes components of a signal at certain frequencies to be confounded, or aliased, with others. The aliases of a particular frequency are pairs of reflections in a set of points given by integer multiples of the Nyquist frequency $\omega_N = (2K)^{-1}$, where K is the time interval between observations. It follows that from D_{low} we can only learn about a folded or aliased version of f,

$$s(K\omega) = K^{-1} \sum_{x \in \mathcal{A}(\omega)} f(x) \tag{4}$$

where $\mathcal{A}(\omega) = \{x = j/K \pm \omega \mid j \in \mathbb{N}, 0 \leq x \leq 1/2\}$ is the set of frequencies aliased with ω. It may be shown that (4) is the spectral density of the subsampled process for which time is rescaled so that observations appear to occur at integer time points. The extra information in data set D_{high}, concerning the full, unfolded spectrum, can be appreciated as filling in the relative sizes of the summands in (4).

It may also be shown that observations of known linear combinations of the process at

every K^{th} time point are similarly only informative for the K-folded spectrum, and that only minor adjustments are needed below to apply it to the situation where a low sampling rate regime involves observing sums or averages of the process at intervals of K time units. This is the case since the act of observing a particular known linear combination of values is equivalent to passing the signal through a filter, whose effect is to multiply the original spectrum by a second, known, spectrum, see, e.g. Chatfield (2003, Chapter 9).

Section 2.2, next, describes how we propose to estimate the spectrum using single rate data taken at every time point and then Section 2.3 adapts our methodology for every K^{th} time point. Section 2.4 explains how we cope with subsampled and aggregated time series. Finally, Section 2.5 then explains how we can combine these techniques in a Bayesian learning framework to deal with multirate data and gives some examples.

2.2. Spectral estimation for a series observed at every time point

Our chosen method for estimating the spectrum of a stationary process treats the logged periodogram ordinates of a time series as noisy observations of the log spectrum, with an additive bias and (almost) homogeneous noise variance following Wahba (1980). This leads to linear regression-type calculations for learning about the coefficients in a basis representation of the log-spectrum.

The regression parameters can be obtained from the D_{high} time series of length $T = N_{high}$ by the raw periodogram, e.g. Brockwell and Davis (2009),

$$I_T(\omega_j) = T^{-1} \left| \sum_{t=1}^{T} x_t \exp(2\pi\imath\omega_j t) \right|^2, \tag{5}$$

where

$$\omega_j = \frac{j}{2n_\omega}, \qquad j = 0, \ldots, n_\omega, \qquad n_\omega = \left\lceil \frac{T+1}{2} \right\rceil.$$

The $I_T(\omega_j)$ are asymptotically independently distributed as exponential or scaled χ^2 random variables, depending on their index, see Chapter 10 of Brockwell and Davis (2009) or Chapter 4 of Shumway and Stoffer (2006), for example. Both distributions are special cases of Gamma distributions whose log moments are available in analytic form: for a general Y distributed according to a Gamma distribution with shape parameter α and rate parameter β we have

$$\mathbb{E}\{\log(Y)\} = \psi(\alpha) - \log(\beta), \qquad \mathrm{var}\{\log(Y)\} = \psi_1(\alpha), \tag{6}$$

where ψ is the digamma function and ψ_1 is the trigamma function, see identities 4.352(1) and 4.358(2) from Gradsteyn and Ryzhik (2007). This result suggests to us that we may usefully model the logged periodogram values as observations of their mean value, obscured by additive noise with variance independent of that mean value. The asymptotics thus lead us to the following model of the log-periodogram as proposed by Wahba (1980),

page 123

$$\log\{I_T(\omega_j)\} = c_j + \log\{f(\omega_j)\} + e_j, \tag{7}$$

with $c_j = \psi(1) \approx -0.58$, $\mathbb{E}(e_j) = 0$ and $\mathrm{var}(e_j) = \psi_1(1) \approx 1.64$ for the interior frequencies that are asymptotically exponential distributed frequencies, and $c_j = \psi(1/2) + \log(2) \approx -1.27$, $\mathbb{E}(e_j) = 0$ and $\mathrm{var}(e_j) = \psi_1(1/2) \approx 4.93$ for the 'boundary frequencies' that are asymptotically scaled χ^2-distributed.

We then choose to model the log-spectrum as a superposition of M basis functions,

$$\log[f(\omega)] = \sum_{k=0}^{M-1} \beta_k b_k(\omega), \tag{8}$$

where the β_k are scalar constants and the $b_k(\omega)$ are scalar functions that we will later gather into (unsubscripted) M-dimensional vector quantities β and $b(\omega)$. In the examples of Section 4 we employ a Fourier series representation for the log-spectrum, with the effect that the zeroth basis coefficient, which multiplies a unit constant basis function, attains special significance since it is equal to the logged forecast variance

$$\mathrm{var}_H(X_t) = \exp\left\{2\int_0^{1/2} \log f(\omega) d\omega\right\} = \exp(\beta_0), \tag{9}$$

which is Kolmogorov's formula. More precisely, (9) quantifies the mean squared error of a best linear prediction for a process value given an infinite history of observations from it, which is alluded to by the subscript H, for history. We will use Kolmogorov's formula extensively in Section 3 where it is used to inform a cost analysis, enabling us to deduce whether switching to a faster sampling rate is expected to be worthwhile.

Choosing such a basis of sinusoids means that our parameterization coincides with a quantity known as the cepstrum, see, e.g. Childers et al. (1977), and renders the model particularly adept at describing spectra with regularly spaced peaks, which occur often in physical processes in the form of harmonics. Nevertheless, it would still be interesting, and unproblematic, to consider alternative bases that are more efficient at spanning other function spaces considered relevant for different problems.

Prior moments for the logged spectrum are induced by specifying prior moments for the coefficients β_k. In the examples to come, for example, we specify that the coefficients are all uncorrelated and have variances that decay as the frequency of the basis functions that they multiply increases. A background and further justification for this sort of specification, which favours smoother log-spectra, can be found in Wahba (1981), who reserves particular attention for variance specifications of the form

$$\mathrm{cov}(\beta_j, \beta_k) \propto \delta_{jk}(m + \lambda\nu_j)^{-m} \qquad \text{for} \qquad j, k = 0, \ldots, M-1, \tag{10}$$

where ν_j is the frequency of the j-th basis function; $\lambda, m > 0$ act as roughness and shape (hyper)parameters; and δ_{jk} is the Kronecker delta. By also specifying that only the expectation of β_0 is non-zero, we encode the belief that the log-spectrum will likely deviate

smoothly from that of a white noise process. We are free to modify this specification however, with the addition of log-spectra of other processes, to the right-hand side of (8) for example.

After calculation of the periodogram ordinates I_T, remembering $T = N_{high}$, we adjust the basis coefficients, β by plugging them in, along with prior moments, $\mathbb{E}(\beta)$ and V, for β and those for the noise terms, namely the e_j in (7), derived from (6), into linear Bayes adjustment equations to produce the adjusted expectation and variance quantities

$$\mathbb{E}_I(\beta) = \mathbb{E}(\beta) + VB^T \left(BVB^T + \Sigma\right)^{-1} [\log(I_T) - \mathbb{E}\{\log(I_T)\}], \tag{11}$$

$$\mathrm{var}_I(\beta) = V - VB^T \left(BVB^T + \Sigma\right)^{-1} BV, \tag{12}$$

where superscript T signifies transpose and, using squared brackets and subscripts to identify the scalar elements of an array,

$$[\beta]_k = \beta_k, \qquad [I_T]_k = I_T(\omega_k), \tag{13}$$

$$[V]_{j,k} = \mathrm{cov}(\beta_j, \beta_k), \qquad [B]_{j,k} = b_k(\omega_j). \tag{14}$$

When $T = N_{high}$ is odd

$$\mathbb{E}[\log\{I_T(\omega_j)\}] = \begin{cases} b(\omega_j)^T \mathbb{E}(\beta) + \psi(1/2) + \log(2) & j = 1 \\ b(\omega_j)^T \mathbb{E}(\beta) + \psi(1) & j = 2, \ldots, T, \end{cases}$$

with $[\Sigma]_{j,k} = \pi^2/2$ if $j = k = 1$; $[\Sigma]_{j,k} = \pi^2/6$ if $j = k = 2, \ldots, T$ and zero otherwise. When T is even

$$\mathbb{E}[\log\{I_T(\omega_j)\}] = \begin{cases} b(\omega_j)^T \mathbb{E}(\beta) + \psi(1/2) + \log(2) & j = 1, T, \\ b(\omega_j)^T \mathbb{E}(\beta) + \psi(1) & j = 2, \ldots, T-1, \end{cases}$$

with $[\Sigma]_{j,k} = \pi^2/2$ if $j = k = 1, T$; $[\Sigma]_{j,k} = \pi^2/6$ if $j = k = 2, \ldots, T-1$ and zero otherwise.

Equations (11)-(12) may be found in their more general form on pages 56–57 of Goldstein and Wooff (2007). Even without an appreciation of linear Bayes methodology, however, the general form of (11) and (12) will be familiar from their frequent recurrence as constrained estimators in the regression literature, in the form of Tikhonov regularization or ridge regression formulae for example.

To compute approximate credible intervals for the values of the log-spectrum, we will later associate the adjusted expectations and variances for them, which are induced by the adjusted moments for the basis coefficients, with modes and expected squared-deviations-from-the-modes of probability distributions. Then, assuming unimodality of these hypothesised distributions, we use Gauss's inequality (Pukelsheim, 1994) to compute conservative credible intervals for the variables they describe. Using this inequality provides us with intervals that accommodate a wide range of posterior distributions while being considerably narrower than those derived for more general settings, such as Chebyshev's inequality.

2.3. Spectral estimation with series of observations at every K^{th} time point

The trick to extending our methodology so that we can make inferences from time series of observations at every K^{th} time point is to employ a first-order Taylor expansion of the log of the folded spectrum, s from (4), in terms of the log of the unfolded spectrum, f.

To use the trick in practice, we begin by taking the discrete Fourier transform of the subsampled data D_{low} of length $R = N_{low}$, just as we did with D_{high} in the previous section,

$$J_R(w_j) = R^{-1} \left| \sum_{r=1}^{R} x_{rK} \exp(2\pi \imath w_j r) \right|^2 , \tag{15}$$

with

$$w_j = \frac{j}{2n_w}, \qquad j = 0, \ldots, n_w, \qquad n_w = \left\lceil \frac{R+1}{2} \right\rceil ,$$

and we understand these periodogram values as being informative for the folded spectrum.

The role of the Taylor expansion is to provide an expression for the log of the folded spectrum that is linear in the log of the unfolded spectrum at the aliased frequencies, and so also linear in the basis coefficients β. This means keeping the expansion's constant term and K linear terms corresponding to each of the aliased frequencies, while the higher-order terms are discarded on the assumption that they are negligibly small. So, using a subscript $*$ to denote the central values (of the spectrum, $f_*(\omega)$, and folded-spectrum, $s_*(\omega)$) around which we expand, we write the Taylor expansion as,

$$\begin{aligned} \log\{s(K\omega)\} = &- \log(K) + \log \left\{ \sum_{x \in \mathcal{A}(\omega)} f(x) \right\} \Bigg|_{f_*} \\ &+ \sum_{x \in \mathcal{A}(\omega)} \frac{\partial \log\{s(K\omega)\}}{\partial \log\{f(x)\}} \Bigg|_{f_*} [\log\{f(x)\} - \log\{f_*(x)\}] \\ &+ \mathcal{O}\left(\sum_{x,x' \in A(\omega)} [\log\{f(x)\} - \log\{f_*(x)\}][\log\{f(x')\} - \log\{f_*(x')\}] \right) , \end{aligned}$$

and, filling in the derivatives using the chain rule and truncating the series, we derive the expression

$$\begin{aligned} \log\{s(K\omega)\} \approx &- \log(K) + \log\{ \sum_{x \in \mathcal{A}(\omega)} f_*(x)\} \\ &+ \sum_{x \in \mathcal{A}(\omega)} \frac{f_*(x)}{K s_*(K\omega)} [\log\{f(x)\} - \log\{f_*(x)\}], \end{aligned} \tag{16}$$

where

$$s_*(K\omega_j) = K^{-1} \sum_{x \in \mathcal{A}(\omega_j)} f_*(x), \tag{17}$$

$$f_*(\omega) = \exp\left\{ \sum_{k=0}^{M-1} \mathbb{E}(\beta_k) b_k(\omega) \right\}. \tag{18}$$

The vector of log-periodogram values $J_R(w_j)$ with $j = 0, \ldots, n_w$, calculated from D_{low} in (15), is understood to be a noisy observation of $\log(s)$. More precisely, the log-periodogram values $J_R(w_j)$ are modelled by exactly the same kind of model as in (7) except that f is replaced by the folded spectrum s. Further, $\log(s)$ is approximately a biased weighted average of certain values of $\log(f)$, with the weights of the average determined by the values of the expansion's centre, f_*. Specifically, the ratio of the central spectrum at a particular frequency to that of the sum from all its aliases determines the degree to which D_{low} modifies the estimate of the spectrum there. For example, when the central spectrum is constant over $(0, 1/2)$, all the K coefficients in (16) are equal so that observation of $J_R(w_j)$ affects our estimate for the spectrum equally at each of the aliased frequencies, resulting in the type of symmetry seen in the top-left panel of Figure 3. When the central spectrum puts the majority of its spectral mass on one aliased frequency, the effect of the periodogram ordinates $J_R(w_j)$ is to adjust this value while leaving the others relatively unchanged.

The expansion point f_* needs to be specified before we can construct the equations for adjusting our beliefs about β given the subsampled data, and the natural candidate for this is the spectrum described by the prior expectation for β. Consequently, our prior expectation for β, informed by intuition, expert knowledge, or other data, not only serves to inform estimates for the spectrum, it directs the information in D_{low} to certain parts of the spectral domain and so determines how we will use it to adjust our beliefs.

In later versions of our code for spectral estimation we have also investigated recomputing the linearization at the posterior expectation derived from the previous expansion point. With this strategy, each adjustment of the basis coefficients begins to look like a Newton optimization of an approximate posterior density. While this development shows great promise, it is not integral to the principles underlying the method described here, and further exposition is reserved for future work.

The adjustment equations (11)–(12) for odd $R = N_{low}$, for the subsampled data, become

$$\mathbb{E}_J(\beta) = \mathbb{E}(\beta) + VC^T \left(CVC^T + \Sigma\right)^{-1} \{\log(J_R) - \mathbb{E}[\log(J_R)]\}, \tag{19}$$

$$\mathrm{var}_J(\beta) = V - VC^T \left(CVC^T + \Sigma\right)^{-1} CV, \tag{20}$$

where

$$[C]_{j,k} = \sum_{x \in \mathcal{A}(w_j)} \frac{f_*(x)}{K s_*(K w_j)} b_k(x), \tag{21}$$

and

$$\mathbb{E}[\log\{J_R(w_j)\}] = \begin{cases} \log\{s_*(w_j)\} + \psi(1/2) + \log(2) & j = 1 \\ \log\{s_*(w_j)\} + \psi(1) & j = 2, \ldots, R. \end{cases} \tag{22}$$

The constants in the expectations (22) and Σ in (19)–(20) are derived from the same results for logged Gamma variables as those in (11)–(12) but are necessarily of different dimension, depending on the lengths of the respective series. Analogous modifications are made for even R.

2.4. Spectral estimation with other derived series

Until now we have assumed that the historical data have arisen as the result of subsampling a finer resolution time series. In many situations, however, coarsely sampled data arise as the result of aggregation of finer series. For example, for the UK traveller data the quarterly number of trips by residents will be the sum taken over three months, and not the value in a given month.

We incorporate both subsampling and aggregation in the last stage of our method by considering the observations, z_t, to be the filtered time series

$$z_t = \theta_0 x_t + \theta_1 x_{t-1} + \ldots + \theta_U x_{t-U},$$

where x_t are the values of the process whose spectrum, f, we are trying to estimate and the $\{\theta_u \mid u = 0, \ldots U\}$ are known constants. To tackle this problem, we introduce additional notation for the spectrum of the filtered process and its folded version

$$g(\omega) = \Theta(\omega) f(\omega), \tag{23}$$

$$t(K\omega) = K^{-1} \sum_{x \in \mathcal{A}(\omega)} g(x), \tag{24}$$

respectively, where $\Theta(\omega) = \left|\sum_{u=0}^{U} \theta_u \exp(-2\pi\imath\omega u)\right|^2$ is the squared gain of the filter $\{\theta_u\}_{u=0}^{U}$, see Chatfield (2003, Section 9.3). We then denote the vector of periodogram values calculated from the series $\{z_K, z_{2K}, \ldots, z_{RK}\}$ by

$$[L]_j = L_R(w_j) = R^{-1} \left| \sum_{r=1}^{R} z_{rK} \exp(-2\pi\imath w_j r) \right|^2, \tag{25}$$

where $\{w_j\}$ is as in (15). The Bayes linear adjustment equations in this case are

$$\mathbb{E}_L(\beta) = \mathbb{E}(\beta) + VF^T \left(FVF^T + \Sigma\right)^{-1} [\log(L_R) - \mathbb{E}\{\log(L_R)\}], \tag{26}$$

$$\mathrm{var}_L(\beta) = V - VF^T \left(FVF^T + \Sigma\right)^{-1} FV, \tag{27}$$

and include arrays which are populated analogously to those in Sections 2.2 and 2.3: Σ is

a diagonal matrix with values $\pi^2/2$ or $\pi^2/6$, V and β have exactly the same interpretation. The objects

$$[F]_{j,k} = \sum_{x \in \mathcal{A}(w_j)} \frac{g_*(x)}{K t_*(K w_j)} b_k(x), \tag{28}$$

where

$$g_*(\omega) = \Theta(\omega) f_*(\omega) = \Theta(\omega) \exp\left\{ \sum_{k=0}^{M-1} \mathbb{E}(\beta_k) b_k(\omega) \right\}, \tag{29}$$

$$t_*(K\omega) = K^{-1} \sum_{x \in \mathcal{A}(\omega)} g_*(x) \tag{30}$$

denote the values of the aggregated and folded aggregated spectra corresponding to the Taylor expansion's centre, and

$$\mathbb{E}[\log\{L_R(w_j)\}] = \begin{cases} \log\{t_*(w_j)\} + \psi(1/2) + \log(2) & j = 1, \\ \log\{t_*(w_j)\} + \psi(1) & j = 2, \ldots, R, \end{cases} \tag{31}$$

(for odd R) are defined analogously to (21) and (22).

2.5. Estimation with a mixture of types of series: sampling slow then fast.

To begin with, before we have looked at any data, we specify a prior mean and variance for β. Then, adjustment of these quantities given a time series requires that we pass them through equations (26)–(27), or the more specific alternatives (19)–(20) or (11)–(12), depending on whether one is dealing with the situations described in Sections 2.4, 2.3 or 2.2. The adjusted expectation and variance quantities that result from these equations are then carried forward to be plugged into those for the next adjustment as prior moments. In the present context, the 'next adjustment' means incorporating information obtained from the fast-sampled data

This simple picture is complicated slightly by the linearisation of the folded spectrum however. In a full Bayes analysis, the type in which there are no numerical constraints or approximations, the order in which adjustments are made ought not to affect their cumulative effect. This is not the case here because the value of $\mathbb{E}(\beta)$ influences adjustments via the value of f_*.

Although it is difficult to provide an authoritative answer to which order is best, we recommend, if possible, adjusting by the high frequency time series first so that the linearisation for the folded spectrum, which tends to provide most of the spectrum shape information, provides a more faithful approximation to the true folded spectrum. In this way, we use D_{high} to tell us roughly where the spectral power is located before using D_{low} to tell us more precisely what shape the spectrum takes there.

The methodology described in Section 2 has been encapsulated in a freely available R (R Development Core Team, 2009) package called `regspec`. We now describe two

examples showing `regspec` in action.

EXAMPLE 1 (SPECTRUM ESTIMATION FOR SLOW/FAST SAMPLED DATA). *This example illustrates the type of inference made possible by our new procedure's ability to assimilate time series observations at two different sampling rates. Specifically, the task is to estimate an ARMA(3,1) process's spectrum from a thinned (slow-sampled) series of* 256 *observations at even integer time points, followed by an un-thinned (fast-sampled) series of only* 16 *observations at integer time points. The process's parameters are* $\phi = (-0.5, 0.4, 0.8)$ *for the AR part and* $\theta = 0.2$ *for the MA part. Our prior for this example asserts that*

$$\mathbb{E}(\beta_j) = 0 \text{ and } \operatorname{cov}(\beta_j, \beta_k) = 9\delta_{jk}\left\{1 - \exp(-\lambda)\right\}\exp(-\lambda j), \tag{32}$$

where $j, k = 0, \ldots, 99$ *and* $\exp(-\lambda) = 0.9$. *This covariance specification comes from taking the limit of* (10) *as* $m \to \infty$. *Figure 2 shows estimates for the spectrum based on just the* 16 *fast-sampled observations for our method along with a standard out-of-the box AR and kernel-smoothed periodogram spectral estimates, giving an impression of their comparative poor performance. The wide credible intervals are crucial in diagnosing the imprecision of all the estimates given so few data.*

Estimates based on few data are clearly going to be poor. So, as we have only few fast-sampled observations, can we improve our estimate by incorporating information from the 'previously collected' slow-sampled data (sampled every second integer)? The top-left plot in Figure 3 shows a spectral estimate using only the slow-sampled data but with the knowledge that we will progressively incorporate the fast-sampled data. Hence, the horizontal axis, rather than being on $(0, 0.25)$ *which would be the normal frequency range observable (for samples on the even integers), is plotted on* $(0, 0.5)$*, the range associated with the incoming fast-sampled data. Figure 3 has unfolded the spectral estimate from* $(0, 0.25)$ *in a symmetric way about* 0.25*: as we have not yet incorporated any higher frequency data, we have no reason to upweight contributions to the* $(0, 0.25)$ *spectrum from either the lower or higher halves of the new spectrum.*

The top-right plot in Figure 3 shows the result of incorporating the additional 16 fast-sampled observations. The spectral peak in Figure 3 (top-right) is estimated to a much higher degree of accuracy compared to that in Figure 2 and the credible intervals are dramatically narrower.

The bottom plots in Figure 3 show a similar situation to the top plots except that the thinning procedure takes only every third observation. The result is similar, incorporating additional less frequently sampled data can boost estimation performance for a small amount of more frequently sampled data.

Despite the usefulness of the plots in Figure 3, there is a great deal of information that they do not communicate. Specifically, while the intervals here reflect the pointwise variance of the spectrum, they say nothing about the covariances between the values of the spectrum at different frequencies. These covariances quantify ideas along the lines that: if the spectrum is smaller than expected at one place it ought to be correspondingly larger than expected in another. Utilizing this covariance information is important when aliasing is present, Section 2 of Nason et al. (2014) provides more details of how this might be done.

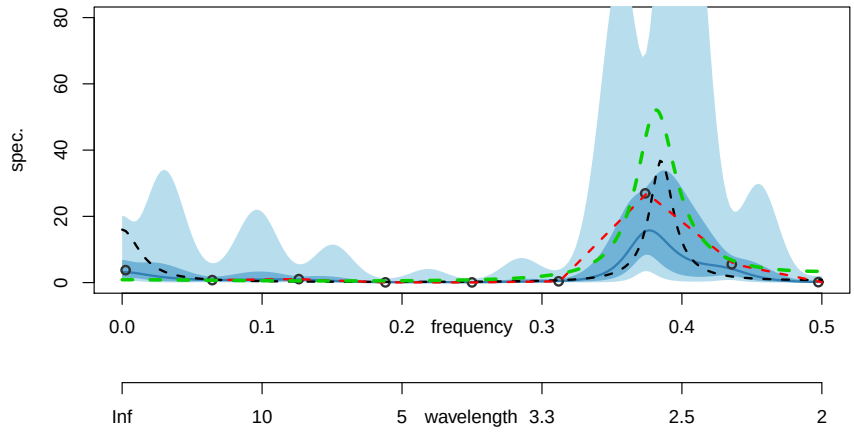

Fig. 2. The true spectrum for the ARMA process is plotted here with a black dashed line. The spectral estimate produced by our new method is plotted with a blue solid line, the AR estimate with a green dashed line and the kernel-smoothed periodogram estimate with a red dashed line. The light blue regions define our approximate 50% and 90% credible intervals around, and the black circles show the periodogram values calculated from the short fast-sampled series.

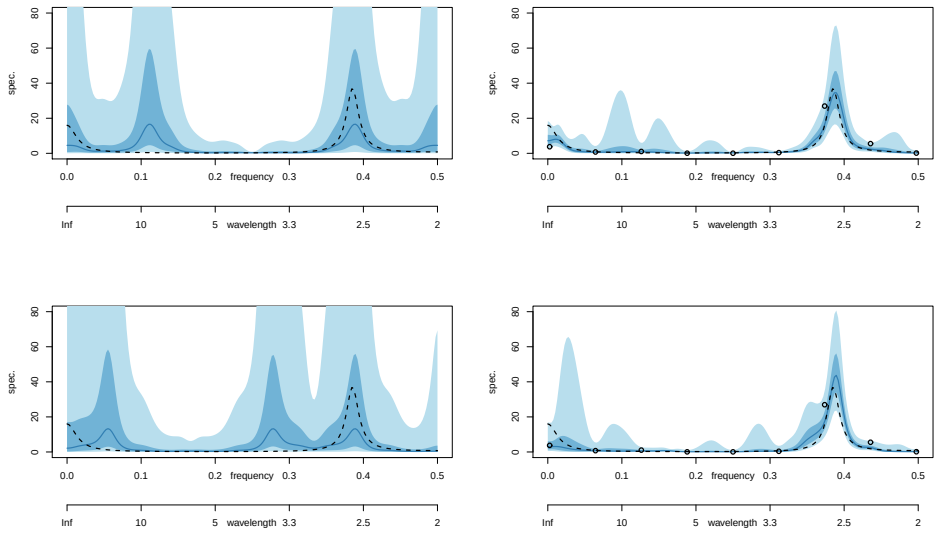

Fig. 3. These figures use the same formatting as Figure 2 and show equivalent estimates based on slow-sampled series (left column), and the slow-sampled and fast-sampled series together (right column) in the cases when the thinning leaves only every second observation (top row) and every third observation (bottom row).

EXAMPLE 2 (COMPARISON OF SPECTRUM ESTIMATORS). *We now compare three methods in the Example 1 scenario with* $N_{low} = 256$ *observations sampled on the even integers followed by integer-sampled observations of varying lengths. The methods used are the kernel-smoothed Lomb-Scargle periodogram (LSSA), the spectrum corresponding to the ARMA model fitted by maximum likelihood (ML ARMA) and our linear Bayes adjusted expectation (LB) that is tailored for this scenario.*

Our performance metric for each simulation $m = 1, \ldots, 1000$ *is the discrepancy statistic*

$$d_m = N_\omega^{-1} \sum_{i=1}^{N_\omega} \{\log[f(\omega_i)] - \log[\hat{f}_m(\omega_i)]\}^2,$$

where $\hat{f}_m(\omega_i)$ *is a point estimate of* $f(\omega_i)$ *for simulation* m*, and the* $\{\omega_i\}$ *is an equally spaced set of* $N_\omega = 256$ *points across* $(0, 1/2)$*.*

We set $N_{low} = 256$ *and repeat the simulation* 1000 *times for each of a range of values for* N_{high}*. Table 1 shows the superiority of our new linear Bayes estimate over the LSSA and ML ARMA methods.*

As alluded to in Section 1.3, and described extensively in Nason et al. (2014), we expect the ML ARMA method to do badly since it can easily be tricked into latching onto illusory frequencies when it follows its likelihood surface into a local mode. This is reflected nicely

Table 1. Monte Carlo estimates of the expected discrepancy statistic, $\mathbb{E}(d_m)$.

N_{high}	8	16	32	64	128	256	512	1024
LSSA	1.53	1.50	1.45	1.35	1.24	1.12	0.99	0.92
ML ARMA	1.90	1.75	1.80	1.55	1.44	1.50	1.28	1.07
LB	1.48	1.03	0.65	0.35	0.21	0.13	0.08	0.05

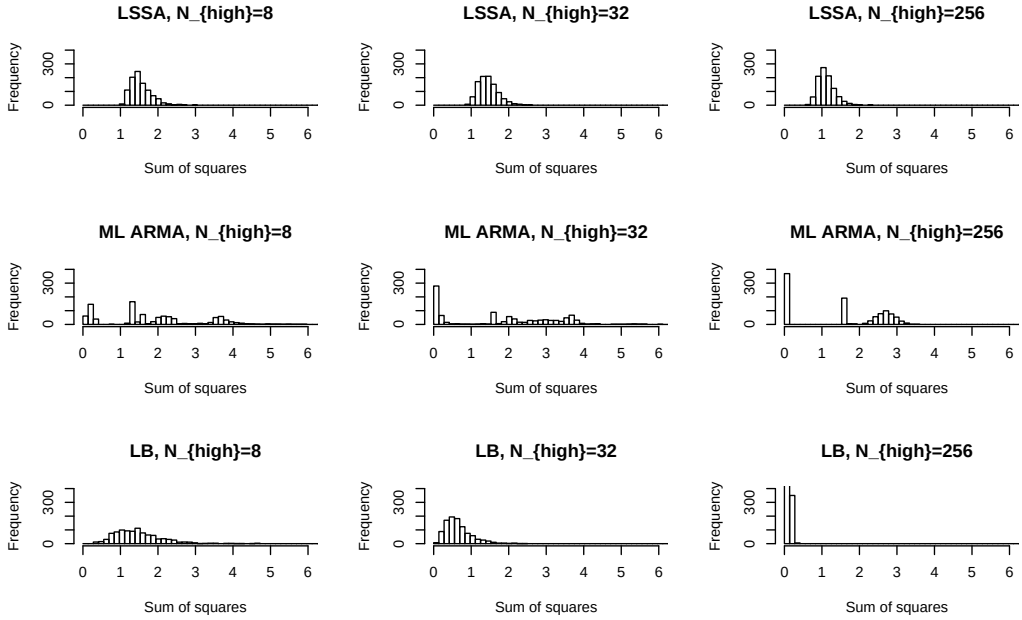

Fig. 4. A selection of histograms of the discrepancy statistic for the estimated spectra over 1000 realisations of the time series D_{high} and D_{low}.

in Figure 4 in the form of distinct modes in the discrepancy statistic that correspond to the differences between the modes of the likelihood.

The relatively poor performance of the LSSA method here is exacerbated by the true spectrum having a peak near the upper limit of the spectral domain. Only in the limit when the fast-sampled data vastly outnumber the slow-sampled data can the method even look at the highest frequencies.

EXAMPLE 3 (UK RESIDENTS' TRIPS ABROAD). *We revisit the number of UK residents travelling abroad series shown in Figure 1 where* $N_{low} = 28$ *and* $N_{high} = 14$*. The plot is informal insofar as plotting data points with different meanings: the upper series consists of sums over quarters, while the lower values are sums over months.*

For this example, we use our model to encode the difference of the log-spectrum of the observed process to that of a baseline SARIMA$(1, 0, 0) \times (1, 0, 0)$ *model with both AR coefficients equal to* 0.6 *with seasonal period twelve months, constant intercept of 5200 and innovation variance of* 3600*. This prior baseline expectation recognizes the dominant*

annual periodicity anticipated and observed in this data. The priors for the difference are of the same form as (32) *with* $\exp(-\lambda) = 0.8$ *and* $\delta_{j,k}$ *instead of* $9\delta_{j,k}$.

Figure 5 shows estimates of the log-spectrum for the monthly series based on the D_{low} *and* D_{high} *individually and combined* $D_{low} \cup D_{high}$. *One should compare the improved estimate (middle) which combines the slow- and fast-sampled data with the estimate constructed just from the fast-sampled data (top). Incorporating the slow-sampled data has tightened the annual peak at wavelength twelve months and appreciably improved the estimation of the lower frequencies in, approximately, the left third of the plot. More precisely, the log-spectral estimate in the middle plot follows more closely the accurate low frequency information obtained from the quarterly data plot and, additionally, the credible intervals are noticeably tighter in the lower frequency region. At first glance it can be difficult to differentiate between the estimates in Figure 5. Nason et al. (2014) Figure 12 shows the differences between the adjusted means of the estimates, more clearly demonstrating the effects of combining the slow and fast-sampled data.*

Figure 6 plots approximate means and 90% *credible intervals for the autocorrelations of the trips abroad series. These are derived via a three-step approximation procedure where a multivariate normal distribution on the log-spectrum induces a log-normal distribution on the spectrum; a normal approximation of this log-normal spectrum induces a multivariate normal distribution on the autocovariances; finally, from the pragmatic moment approximations for ratios of correlated normal variables, Marsaglia (2006), we derive means and variances for the autocorrelations, which inform the credible regions.*

The most obvious change observed when comparing the autocorrelations estimated from just the monthly data (top) with those using the combined series (middle) is the tightening of the credible intervals, reflecting the increased information supplied by the quarterly data. We also note that the top plot features estimated autocorrelations that are revised upwards in the middle plot; this initial underestimation can be attributed to the inability of the short monthly series to contribute to estimates of the longer lag autocorrelations. Also, the shape of the peak in the autocorrelations around twelve months is noticeably sharper in the middle plot than that shown in the bottom plot.

The bottom subfigure of 6 is interesting, as it shows how slow-sampled data are leveraged against our prior for the log-spectrum (in this case, one that implies that the log-spectrum deviates smoothly from that of a seasonal autoregressive process) to produce estimates for autocorrelations at completely unobserved lags (highlighted in Figure 6 using red crosses). While the precise nature of this leveraging, whose legitimacy rests on Bayes theorem and our approximations of its implications, is difficult to describe concisely in this instance, it is perhaps useful now to discuss it by providing a commentary on the flow of information through our model:

Our prior belief for a smooth log spectrum translates to a prior for quickly decaying autocorrelations, which underlies the bounds of the autocorrelations tightening around zero in Figure 6. The expected location of spectral power determines how the autocorrelations are effectively interpolated between values for which slow-sampled data are immediately informative. Here, the filter gain takes the form of a spectral density with the greatest part of its mass over the first third of the spectral domain. With an implicit prior assumption that the spectral mass of the monthly process is distributed evenly. Hence, periodogram

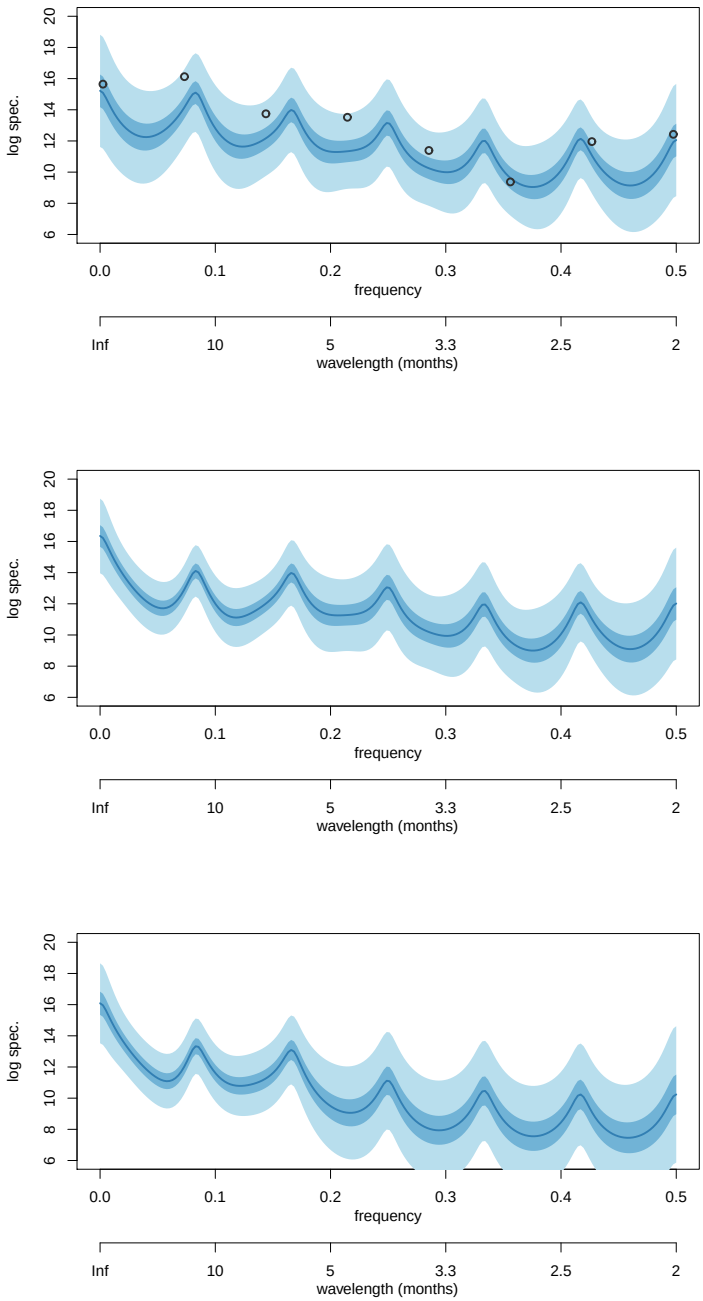

Fig. 5. Estimates of the log-spectrum for the number of trips abroad made by UK residents each month given different training data. The top estimate is informed only by 14 monthly data points, the bottom estimate is informed only by the quarterly data and the middle estimate is informed by both data sets. The black circles in the top plot show the log-periodogram values calculated from the monthly data.

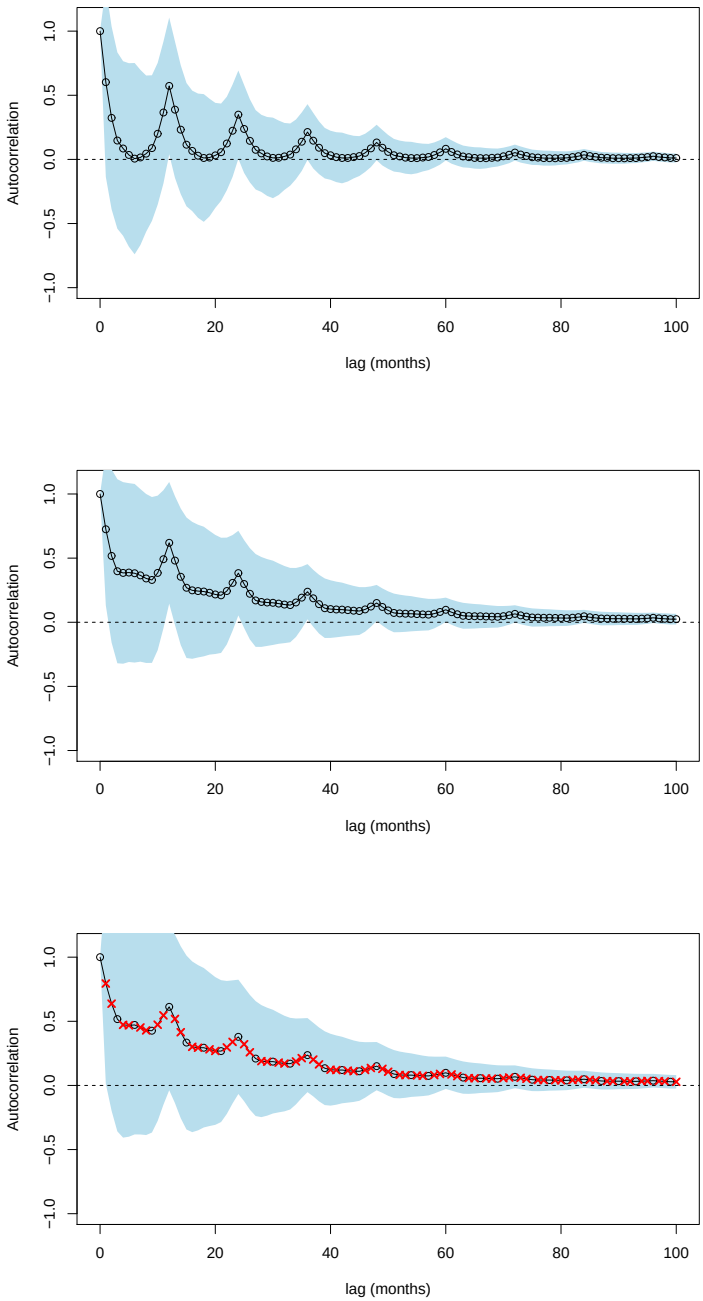

Fig. 6. Estimates of the autocorrelation values for the number of trips abroad made by UK residents each month (corresponding to the spectral estimates shown in Figure 5). The top estimate is informed only by the monthly data, the bottom estimate is informed only by the quarterly data and the middle estimate is informed by both data sets as in Figure 5. The height of the shaded region at each lag is approximately four standard deviations of the autocorrelation value there. For the bottom subfigure, autocorrelations (obtained by interpolation of the log-spectrum and assuming smoothness) not at lags of multiples of three months are denoted with red crosses.

ordinates from the aggregated series are predominantly informative for the spectrum over this third. When these ordinates are observed to be greater than expected, our power estimates over the lower frequencies are revised upwards. The accumulation of power at low frequencies then translates into the smooth interpolation of the autocorrelations (rather than an oscillatory one), observed in the bottom subplot of Figure 6.

Our work with this example underlines the importance of combining information from lower and higher sampling rates after a change in sampling strategy, in particular for publication purposes. Doing so will facilitate the estimation of trends in the data, represented at lower frequencies and also of regular seasonal fluctuations that are often estimated and removed in official statistics publications. Discarding the quarterly information here can be seen to lead to wider credible intervals at seasonal and lower frequencies for example.

One particularly important use for estimates of the second order quantities that we study here is the prediction of unobserved process values. To demonstrate our ability to do this, we derive monthly hindcasts for the 'trips abroad series' as best linear predictors using all the available data (the quarterly pre-2011, and monthly post-2011) and autocovariance information derived from the middle spectral estimate in Figure 5, which itself uses all the data. The hindcasts and error bars (of length four times the prediction standard deviation, for visibility) are plotted in blue on Figure 1. Note that the direction of time here is unimportant, and the same method could be used to infer a process's values given a densely sampled past and sparsely sampled future; a calculation relevant for scenarios in which the sampling rate is decreased.

Bayesian hindcasting methods have been developed in the fields of climatology and meteorology, see, for example, Solomon (2007, p. 689), Werner and Holbrook (2011), Ortego et al. (2014) and references therein. However, hindcasting using multiple sample rate assimilation via a multirate spectral approach as developed here appears to be new and fully coherent in the Bayesian sense.

3. Pricing sampling strategies

Above, we established a strategy for learning about a spectrum in a multirate context. Now we seek to formulate and solve the decision problem of determining whether a switch to a higher rate is cost effective. To structure our understanding of possible realities, and so to provide a rationale for making decisions, we propose studying the cost function

$$C = \underbrace{N_{trial}C_{trial}}_{\text{term 1}} + \underbrace{\mathbf{1}_{change}C_{change}}_{\text{term 2}} + \underbrace{N_{future}[C_{samp} + F_{low} + \mathbf{1}_{change}\{(K-1)C_{samp} + F_{high} - F_{low}\}]}_{\text{term 3}}, \quad (33)$$

for which the notation is summarized in the following table.

Notation	Meaning
C_{samp}	Standard observation cost
C_{trial}	Trial observation cost
C_{change}	One-off change cost
F_{low}	Forecast cost in the low-frequency regime
F_{high}	Forecast cost in the high-frequency regime
K	Slow sampling interval from Section 2.3
N_{trial}	No. of trial observations
N_{future}	No. of future time intervals which contain either one or K observations depending on whether we switch or not.
$\mathbf{1}_{change}$	The sampling strategy change indicator function.

While specification of the cost function serves to endow our uncertainties with meaning, in doing so, it necessarily forces us to entertain certain assumptions. Implicit to (33), for example, is the notion that our current default strategy is to collect data at the low sampling rate and that resources would need to be expended to switch to the higher sampling rate. Further implications are discussed in Nason et al. (2014) Section 1.2.

Returning to our motivating problem, we would like to be able to advise a client considering a change to the measurement of a time series on whether they ought to sample more often, and on how easy it would be for them to collect evidence informing this decision. In terms of the cost function, we interpret these goals as providing recommendations for an optimal value for $\mathbf{1}_{change} \in \{0, 1\}$ and for $N_{trial} \in \mathbb{N}$. For now, we assume the costs C_{samp}, C_{trial} and C_{change}, and the number N_{future}, to be fixed constants that are given to us. We will return to the plausibility of this assumption in paragraph five of Section 5.2. We also anticipate that specifications for the forecast costs F_{low} and F_{high} are likely to be highly context specific. Certainly, in real applications, one can easily imagine discussing, and attempting to elicit, complex relationships between forecasting skill and actual costs. While it will be interesting to investigate more realistic relationships in the future, we concentrate now on a widely applicable and, crucially, tractable relationship whereby forecast costs are considered to be proportional to the log of the forecast variance, which is derived from the log-spectrum via Kolmogorov's formula,

$$F_{low} = C_u \log[\mathrm{var}_{H_{low}}(x_t)] = 2C_u \int_0^{1/2} \log[s(\omega)] d\omega,$$
$$F_{high} = C_u \log[\mathrm{var}_{H_{high}}(x_t)] = 2C_u \int_0^{1/2} \log[f(\omega)] d\omega,$$

where $f(\omega)$ and $s(\omega)$ are the spectrum and folded spectrum as defined in (2) and (4) respectively, and C_u is a calibration constant that converts prediction success into monetary terms based on one's utility. The intuition behind this selection is that the ability to forecast the future provides a tangible benefit. These cost specifications coincide up to proportionality with the expected log-likelihoods for process values given infinite histories of observations from the process, which we have denoted here with H_{low} and H_{high}.

To make a specification for $\mathbf{1}_{change}$, and so for the future sampling strategy that we

currently favour, we must ask whether the cost of switching relative to not switching,

$$Q := C_{change} + N_{future}\{(K-1)C_{samp} + F_{high} - F_{low}\}, \tag{34}$$

is greater than zero, in which case we stick with the low frequency sampling regime, or less than zero, in which case we switch to the high frequency sampling regime. In practice however, since $F_{high} - F_{low}$ is unknown, we can only estimate the cost and base our decision for which sampling strategy to choose upon this estimate. This decision process gives rise to the additional random variable that defines a future cost reduction based on our current beliefs for Q,

$$R_D := \begin{cases} Q & \mathbb{E}_D(Q) \leq 0, \\ 0 & \mathbb{E}_D(Q) > 0, \end{cases} \tag{35}$$

where $\mathbb{E}_D(Q)$ denotes our expectation for Q given a generic dataset D. We also note that Q is bounded above since $F_{high} < F_{low}$, reflecting the fact that better informed forecasts should always lead to smaller costs,

$$Q < Q_{max} = C_{change} + (K-1)N_{future}C_{samp}. \tag{36}$$

The same Taylor expansion that enabled us to incorporate under-sampled time series into our inferences in Section 2.3 now allows us produce an approximate mean and variance for $F_{high} - F_{low}$, and thus for Q, since we approximate it with the following linear combination of the elements of β,

$$\begin{aligned} F_{high} - F_{low} \approx & C_u \left(\beta_0 - 2\int_0^{1/2} \log\{s_*(\omega)\} d\omega \right. \\ & \left. -2K \int_0^{1/(2K)} \sum_{x \in \mathcal{A}(\omega)} \frac{f_*(x)}{K s_*(K\omega)} \left[\log\{f(x)\} - \log\{f_*(x)\}\right] d\omega \right) \\ = & C_u \left(\beta_0 - 2\int_0^{1/2} \log\{s_*(\omega)\} d\omega \right. \\ & \left. -2K \left\{ \int_0^{1/(2K)} \sum_{x \in \mathcal{A}(\omega)} \frac{f_*(x)}{K s_*(K\omega)} b(x) d\omega \right\}^T (\beta - \beta_*) \right), \end{aligned} \tag{37}$$

where $b(x)$ is the same vector of basis function values introduced in (8), and superscript T signifies the matrix transpose. To formulate the expected or anticipated cost reduction afforded by a trial, we need only consider the events which will cause us to adjust our expectation for Q from one side of zero to the other. These are the instances in which the information gathered in the trial causes us to change our mind for the long-term sampling strategy to employ. To proceed, we must recognise that our adjusted expectation for Q given the trial data is a random variable, since we have not yet performed the trial. We then

need to calculate the expected cost reduction as the probability that our adjusted expectation for Q crosses the decision threshold, multiplied by the average expected cost reduction given that this happens.

To gather information for the behaviour of our future adjusted expectation we return to the variance estimate (12). This formula tells us how our prior variance for the coefficients β ought to be reduced after observing some data, or, seen in another way, how our prior variance for β can be partitioned into a portion that disappears once data have been observed and a portion that remains. The part that disappears is the variance for what our adjusted expectation will be after observing the data. Additionally, we also know our current expectation for this adjusted expectation: it must be the same as our current expectation for β in order for our beliefs to be coherent.

Since we have specified covariance properties for $\log(f)$ *a priori* through the coefficient vector β, and since (12) does not depend on the periodogram values, we can anticipate how the variance for our beliefs for $\log(f)$ and the linearised version of $F_{high} - F_{low}$ (37) will evolve. Consequently, given the expectation and variance of Q informed by some (possibly empty) historical dataset D_0, both the expectation and variance of our adjusted expectation given the additional series D_{trial} (with length N_{trial}), are known in advance of performing the trial, taking us a step closer to calculating the expected benefit of trials of various length.

To explicitly quantify the expected cost reduction afforded by a trial, we load the moments of the adjusted expectation into a particular distribution and calculate

$$\mathbb{E}(R_{D_0 \cup D_{trial}} - R_{D_0}) = \begin{cases} \int_0^{Q_{max}} \mathbb{E}_{D_0 \cup D_{trial}}(Q)\ \phi\{\mathbb{E}_{D_0 \cup D_{trial}}(Q)\} & \mathbb{E}_{D_0}(Q) \leq 0, \\ -\int_{-\infty}^{0} \mathbb{E}_{D_0 \cup D_{trial}}(Q)\ \phi\{\mathbb{E}_{D_0 \cup D_{trial}}(Q)\} & \mathbb{E}_{D_0}(Q) > 0. \end{cases} \tag{38}$$

where ϕ is the probability density function assigned to $\mathbb{E}_{D_0 \cup D_{trial}}(Q)$.

In our work so far with this methodology, we have found it effective to plug the moments for the future expectation for Q into a reflected and translated Gamma distribution with support $(-\infty, Q_{max}]$. We have also experimented with using a normal distribution here, ignoring the upper bound Q_{max}. Our findings appear reasonably robust to this choice, in so far as producing trial length recommendations that still lead to average reduced costs in simulations, but since calculations are straightforward in both cases, we favour the Gamma distribution in order to account for the upper bound (36).

Identifying an appropriate value for N_{trial} requires that we compute approximate values (38) for all feasible values of N_{trial}, and adding to them the costs required to pay for the trials. The trial length corresponding to the minimal net expected cost then constitutes our recommendation for N_{trial}. To clarify, according to our reasoning framework we always have an expectation for Q and, consequently, always have a preference between high and low sampling rates. By identifying a non-zero value for N_{trial} at which the expectation for (33) is minimized, however, we are identifying that the risk associated with acting on our current preference equates to a cost greater than that of a trial. If the minimizing trial length is zero, then we are ready to present our recommendation for the switching decision variable, $\mathbf{1}_{change} = \mathbf{1}\{\mathbb{E}(Q) \leq 0\}$, immediately without collecting any trial observations.

4. Examples

EXAMPLE 4 (ESTIMATING FORECAST VARIANCES). *In this example we test the accuracy of our log-spectrum derived estimate for the statistic* (37)*, which quantifies the potential for increasing the precision of forecasts. The investigation takes the form of a Monte Carlo simulation experiment in which we draw a large ensemble of log-spectra from a multivariate normal prior with moments of the form* (10)*, and simulating two time series of normally distributed values from each one. We then use these time series to recover the log-spectra. The first series here is of length* 90 *and consists of observations at every other time point, while the second is of length* 48 *and consists of observations at every time point. Our estimates for* (37) *are their adjusted expectations, which follows from that of* β*. They are accompanied by approximate conservative* 90% *credible intervals calculated using Gauss's inequality.*

Figure 7 compares our estimates against the truth but also against alternative estimates, computed using more conventional time series methods, that do not take the structural relationship between the forecast variances in the two sampling regimes into account. We will refer to these as AR estimates. These are derived by, firstly, fitting AR models by maximum-likelihood to the pairs of time series, and calculating corresponding pairs of AR residuals. By conditioning on the maximum-likelihood estimates for the AR coefficients and entertaining improper uniform priors for the innovation variances for the two processes, we induce inverse-gamma posterior distributions for them and a scaled F-distribution posterior for their ratio. We proceed to truncate the F-distribution in recognition of the idea that the forecast variance in the high sampling-frequency regime cannot exceed that in the low one. Finally, the median, which we take as a point estimator, and a central credible interval for this approximate posterior are computed from particular quantiles of the F-distribution, which are easily evaluated with standard statistical software.

Straight away, we can see that the AR estimates perform relatively badly, achieving a mean absolute error over the simulations of 0.24 *against an equivalent figure of* 0.11 *for the linear Bayes estimates. We also note that over the ensemble of* 1000 90%*-approximate credible intervals calculated with the linear Bayes and AR approaches, the former are on average* 0.58 *as wide as the latter, while they capture the true value of the log variance ratio in* 86% *and* 81% *of cases respectively.*

EXAMPLE 5 (THE UK CONSUMER PRICE INDEX). *Here, we analyse a monthly Consumer Price Index (CPI) series produced by the Office for National Statistics which measures the change in prices using a chain-linked fixed-basket approach. Informally, we understand the CPI as being informative for typical consumer day-to-day spending as a percentage of a 2005 baseline. We preprocess the data firstly by passing them through the log transform and then by removing a linear trend. Following this, we model the difference from the spectrum of a white noise process with variance* 0.03^2 *and prior specification for the* β*s as in* (32) *but with* $\exp(-\lambda) = 0.95$ *and* $4\delta_{j,k}$ *instead of* $9\delta_{j,k}$*.*

Our next step is to modify the data set to recreate the type of scenario we anticipated in Section 2. This involves separating the first 270 data points and thinning them to every third value so that our historical low frequency data set, D_{low}*, consists of quarterly observations. For* D_{high} *we take the subsequent 4 years of data at the monthly rate. From*

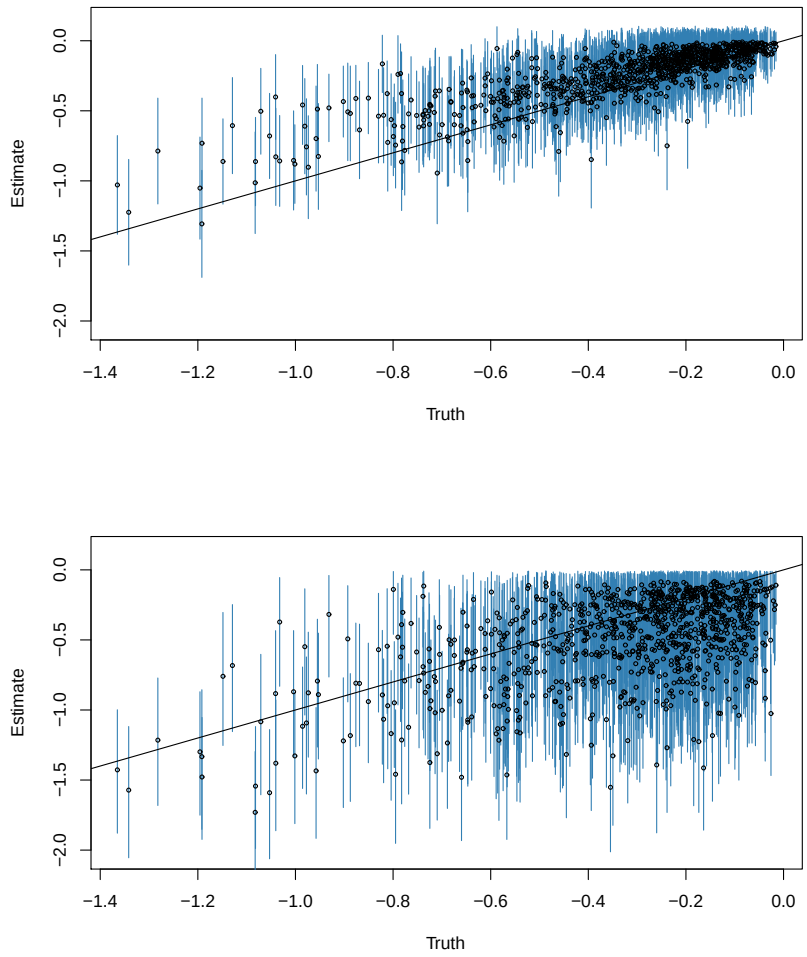

Fig. 7. Plots of the true and estimated log forecast variance ratios. The black diagonal line shows the estimates' target. The upper subfigure here shows the estimates derived via our linear Bayes model for the log-spectrum, while the lower subfigure shows estimates computed from AR residuals as described in Example 4. The blue vertical lines represent approximate 90% credible intervals for the true values.

Fig. 8. The thinned (slow-sampled) de-trended log-CPI time series.

these two time series (plotted in Figure 8) we produce a spectral estimate that encodes an expectation and variance for the linearised version of $F_{high} - F_{low}$*, which is interpretable as a log ratio of the forecast variance in the high sampling-frequency regime to that in the low sampling-frequency regime. Using the same strategy used to produce plots for the log-spectrum, we associate these two moments with a mode and expected squared-deviation-from-the-mode before employing Gauss's inequality to compute bounds for a conservative credible interval. By exponentiating the expectation and the bounds we produce forecast variance ratio values* $(0.34, 0.47, 0.65)$*, meaning that our analysis implies increasing the sampling rate will lead to a forecast variance of between* 34% *and* 65% *of that achieved otherwise, with probability at least* 0.9*. By looking at residuals from AR models fitted to the data sets* D_{low} *and* D_{high}*, as we did in the example 4, we compute a point estimate for* $F_{high} - F_{low}$ *of* -1.48*, which, upon exponentiation, suggests a forecast variance reduction to* 21% *of the low-sampling rate variance. The accompanying approximate credible interval, derived using the F-distribution, suggests that the forecast variance can be reduced to between* 15% *and* 35% *with probability* 0.9*.*

Both these estimates may be compared to the exponentiated bounds of the 90% *conservative credible interval calculated using the linear Bayes model for the log-spectrum given the whole, fast-sampled time series, which we refer to as the Oracle estimate. Informally, we understand the Oracle estimate as being as close as we can get to the sort of hidden true values available in our simulation experiments. The three variance ratio estimates and their associated prediction intervals are compared graphically in Figure 9. We can see here that the AR estimate appears to overestimate the extra precision made possible by high-frequency sampling. The linear Bayes point estimate for the forecast variance ratio is closer to that of the Oracle, and is accompanied by an interval which overlaps with the*

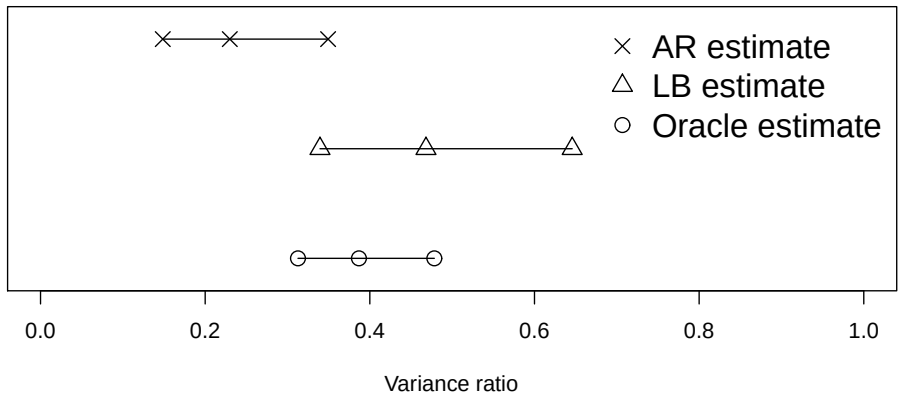

Fig. 9. Point estimates and 90% credible intervals for the ratio of the prediction variance at high and low CPI sampling rates as calculated using: the thinned data and an analysis of residuals to fitted AR models; the thinned data and the linear Bayes model for the log-spectrum; and the un-thinned data and the linear Bayes model for the log-spectrum.

greater part of the Oracle's interval.

So we suggest here, without proposing specific but essentially arbitrary cost parameters that would serve to quantify the argument, that in the scenario described in this example, we would have been able to specify a more appropriate strategy in a variety of situations using the linear Bayes model rather than the calculations based on the AR fits. Firstly, we would argue that the evidence in favour of sampling at the monthly rate is not as strong as the AR estimate suggests. Thus the linear Bayes inference may have prevented us from incurring sampling costs which were not recouped by more precise forecasts in the future. Secondly, the wider credible interval for the linear Bayes inference encodes the idea that we would have been able to make the switching decision with less confidence than indicated by the AR calculations. The implication this time is that additional data and a postponement of a sampling strategy decision would have been of greater value than suggested by the AR analysis.

5. Discussion and conclusions

5.1. Methodology Overview

We have shown the power and simplicity of log-spectrum smoothing techniques and how they can be extended to conduct inference using time series data sampled at multiple rates in a Bayesian framework. Our methods permit us to anticipate the set of plausible spectra confounded by the aliasing phenomenon in a way that existing procedures fail to do. Further, we have developed a method for quantifying our uncertainty for the spectrum in

terms of its implications for forecasting and hindcasting, and proceeded to translate this uncertainty into a cost that may be weighed against the costs of obtaining further information, thus informing a decision problem that is genuinely relevant to applied statisticians and official statistics in particular.

An R Development Core Team (2009) package, provisionally titled `regspec`, for implementing the calculations is available from the authors and will be released on the Comprehensive R Archive Network (CRAN) in due course.

5.2. Further Avenues and Applications

One area where our central problem is very real is in official statistics, where surveys are often relatively large and costly, but there is constant pressure for more timely information to support evidence-based decision making in government, balanced against the need for efficiency in the use of public funds. In 1994-5, for example, there was a review of the options for making the Labour Force Survey into a true monthly survey by Steel (1997), but the subsequent public consultation on the options determined that there was insufficient benefit for the cost of £7-8m (in 1996 prices) as discussed by Werner (2006). However, a reduced-cost solution of producing rolling three-monthly estimates each month was introduced from April 1998, costing around £0.2m (much of the cost was to produce an integrated Labour Market Statistics release, not all for calculation of the rolling estimates). An experimental monthly series has been calculated from the existing (quarterly) design since 2004 as an aid to interpretation, but is not considered to be of sufficient quality to stand in its own right.

All of these steps in the evolution of the Labour Force Survey could be assessed retrospectively using the methods developed here; many of the data are available, and some of the cost estimates. The issue of whether a true monthly design would be beneficial remains topical, but the high cost of data collection for such a design has tended to make it an unlikely development. However, the topic is often revisited whenever labour market statistics receive heightened attention, such as when the Bank of England considered linking interest rate change decisions to the rate of unemployment during 2013. So there is a possibility of prospective analysis for future change.

Section 3 discussed the importance of the smoothness of the log-spectrum for the value of pilot observations to the decision. This is also related to the cost and design of a faster rate survey (whether the pilot or a full survey). It is unusual for a survey design to be carried forward wholesale for implementation at a faster rate; usually there is a design change involved, often a change to the sample size, which affects the quality of the new series. For example, the construction output statistics produced by the Office for National Statistics changed from a sample of 12,000 business per quarter to 8,000 businesses per month in 2010 as described by Crook and Sharp (2010). This change would be expected to make the sampling variance of the high frequency (monthly) series higher than that of the low-frequency series.

A change such as that in the construction output statistics suggests that other forms of the cost model (33) might be worthy of study. We could begin, for example, by separating out the costs for high frequency and low frequency observations to allow for differences in

design. This would mean replacing the third term in (33) to give

$$C = N_{trial}C_{trial} + \mathbf{1}_{change}C_{change} + N_{future}\{\mathbf{1}_{change}K(C_{high} + F_{high}) + (1 - \mathbf{1}_{change})(C_{low} + F_{low})\}. \quad (39)$$

The cost function (39) then has some interesting special cases: (a) $C_{trial} = C_{high}$; when the trials use the proposed new method, rather than being a smaller pilot. This allows straightforward post-hoc evaluation of the net benefit of a change already made to a statistical collection; (b) when there is no change in data collection cost, but there is a change in publication frequency, that is when $KC_{high} = C_{low}$. This is like the Labour Force Survey example, where the outputs change periodicity without a change in the ongoing collection costs (in practice there is usually a smaller cost to changing the estimation and compilation processes, and the publication schedule). There are doubtless many alternative ways in which the costs of changes could be measured.

In both (33) and (39), N_{future} codifies the period over which the cost/benefit trade-off is to be evaluated for the decision process. It is only possible to recommend change when the third term in the cost function is negative, and given that it is indeed negative, the larger N_{future} , the more often we will conclude that change is worthwhile. Some general guidance on appraisal of decisions like this is available from HM Treasury (2003), and is based on the service life of the asset under consideration. There is no general guidance on the service life of a statistical survey, but using information on service lives of research and development assets in Ker (2013) for Information and Communication Services industries we can use 8-20 years as a rule of thumb. In fact 20 years corresponds with the current life of some redesigns (such as the Labour Force Survey) and therefore feels appropriate, though this is not based on a comprehensive analysis of survey changes. The trade-off between cost and benefit is also affected by the relative size of the C and F parameters in the cost function; in general there may be a challenge in translating the variances (the F parameters) into suitable monetary values.

Further motivation for the synthesis of data at multiple sampling rates can be identified in recent reports from the ESSnet (European Statistical System network). Among the network members' documented concerns are issues relating to company Value-Added Tax (VAT) registrations, which provide administrative data for monthly and quarterly turnover estimates. The report by Vlag et al. (2010), in particular, discusses the difficulties faced by Statistics Netherlands in 2009 after abolition of legislation that obliged companies to produce monthly VAT declarations. Although the precise details are too complex to describe fully here, it is clear that: the informativeness of their monthly VAT series was severely degraded; the annual series was not, since annual reports remained compulsory; and monthly turnover estimates were still required in order to fulfil commitments to national and European authorities. We contend that our methodology would have been of value here for its ability to propagate knowledge for a process's variability from one sampling regime to another. Specifically, we would have been able to estimate both the spectral properties and actual values of the process underlying the VAT figures from the historic monthly series and contemporary annual series. The deliverable of this procedure would have been a 'nowcast' accompanied by coherent credible intervals, resulting in a series of estimates

qualitatively equivalent to the hindcast of Figure 1 but with the time axis reversed. Squeezing information from the degraded contemporary monthly series would be challenging and, together with analysis of the UK construction output survey, represents an important direction for further work.

With an eye to the future, the consideration of changes to the way the population Census is conducted in England and Wales also fits within this framework. The ONS (2013) report sets out options for a census using existing government data and compulsory annual surveys, which has considerable transition and ongoing costs but has an advantage in allowing higher-frequency (possibly annual) updates to some statistics on the population relative to the low frequency (decennial) population census. The cost estimates for the different approaches are also set out in the consultation document, so an appropriate cost function could be used to bring this information together to aid decision-making.

It appears that the applications of this approach are many in topical areas of official statistics, and there are other areas of statistics where similar applications would also be valuable in evidence-based decision-making.

Acknowledgements

This work was supported by EPSRC Grant ER/K020951/1.

References

Andreou, E., Ghysels, E., and Kourtellos, A. (2011) Forecasting with mixed-frequency data, in M. P. Clements and D. F. Hendry, eds., *The Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.

Armesto, M. T., Engemann, K. M., and Owyang, M. T. (2010) Forecasting with mixed frequencies, *Federal Reserve Bank of St. Louis Review*, **92**, 521–36.

Brockwell, P. J. and Davis, R. A. (2009) *Time Series: Theory and Methods*, Springer, New York, second edition.

Broersen, P. M. T. (2006) *Automatic Autocorrelation and Spectral Analysis*, Springer, London.

Broersen, P. M. T. and Bos, R. (2006) Time-series analysis of data are randomly missing, *IEEE Trans. Instrum. Meas.*, **55**, 79–84.

Ceperley, D., Chen, Y., Craiu, R. V., Meng, X.-L., Mira, A., and Rosenthal, J. (2012) Challenges and advances in high dimensional and high complexity Monte Carlo computation and theory, in *Proceedings of the Workshop at the Banff International Research Station for Mathematical Innovation Discovery*, Final Report.

Chatfield, C. (2003) *The Analysis of Time Series: An Introduction*, Chapman and Hall/CRC, London, sixth edition.

Childers, D. G., Skinner, D. P., and Kemerait, R. C. (1977) The cepstrum: A guide to processing, *Proc. IEEE*, **65**, 1428–1443.

Choudhuri, N., Ghosal, S., and Roy, A. (2004) Bayesian estimation of the spectral density of a time series, *J. Am. Statist. Ass.*, **99**, 1050–1059.

Cimadomo, J. and D'Agostino, A. (2015) Combining time variation and mixed frequencies: an analysis of government spending multipliers in Italy, *J. Appl. Econometrics*, **30**, (to appear).

Clements, M. and Galvão, A. B. (2008) Macroeconomic forecasting with mixed-frequency data, *J. Bus. Econ. Stat.*, **26**, 546–554.

Crook, T. and Sharp, G. (2010) Development of construction statistics, *Economic and Labour Market Review*, **4**, 52–58.

Eraker, B., Chiu, C. W., Foerster, A. T., Kim, T., and Seoane, H. D. (2015) Bayesian mixed frequency VARs, *J. Fin. Econom.*, **13**, 698–721.

Foroni, C. and Marcellino, M. (2013) A survey of econometric methods for mixed-frequency data, Technical Report 2013-06, Norges Bank Research, PO Box 1179 Sentrum, 0107 Oslo, Norway.

Foroni, C., Ghysels, E., and Marcellino, M. (2013) Mixed-frequency vector autoregressive models, in T. B. Fomby, L. Kilian, and A. Murphy, eds., *VAR models in Macroeconomics. Advances in Econometrics*, volume 32, pp. 247–272, Emerald Group Publishing, Bingley, UK.

Foroni, C., Guérin, P., and Marcellino, M. (2015) Markov-switching mixed-frequency VAR models, *Int. J. Forecasting*, **31**, 692–711.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006) Predicting volatility: Getting the most out of return data sampled at different frequencies, *J. Econometrics*, **131**, 59–95.

Ghysels, E., Hill, J. B., and Motegi, K. (2015a) Simple Granger causality tests for mixed frequency data, (submitted for publication).

Ghysels, E., Hill, J. B., and Motegi, K. (2015b) Testing for Granger causality with mixed frequency data, *J. Econometrics*, **190**, (to appear).

Goldstein, M. and Wooff, D. (2007) *Bayes Linear Statistics, Theory and Methods*, Wiley, Chichester.

Götz, T. B. and Hauzenberger, K. (2015) Time-varying mixed-frequency vector autoregressive models, `https://sites.google.com/site/khauzenberger/`.

Gradsteyn, I. and Ryzhik, I. (2007) *Table of Integrals, Series and Products*, Elsevier, Amsterdam.

Green, P. J. and Hastie, D. I. (2009) Reversible jump MCMC.

HM Treasury (2003) *The Green Book: Appraisal and Evaluation in Central Government*, Stationary Office Books, London.

Jones, R. H. (1976) Estimation of the innovation generalized variance of a multivariate stationary time series, *J. Am. Statist. Ass.*, **71**, 386–388.

Ker, D. (2013) Service lives of R&D assets: background and comparison of approaches, Technical report, Office for National Statistics.

Lomb, N. R. (1976) Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, **39**, 447–462.

Mallick, B. K., Denison, D. G. T., and Gangopadhyay, A. K. (2002) A Bayesian curve fitting approach to power spectrum estimation, *J. Nonparam. Statist.*, **14**, 141–153.

Marsaglia, G. (2006) Ratios of normal variables, *Statist. Sci.*, **16**, 1–10.

Nason, G. P., Powell, B., Elliott, D., and Smith, P. (2014) Supplementary material for "Should We Sample More Frequently? decision support via mumultirate spectrum estimation", Technical Report 14:02, Statistics Group, University of Bristol.

ONS (2013) The census and future provision of population statistics in England and Wales, Public consultation report, Office for National Statistics.

Ortego, M., Egozcue, J., and Tolosana-Delgado, R. (2014) Bayesian trend analysis of extreme wind using observed and hindcast series off the Catalan coast, NW Mediterranean sea., *Nat. Hazards Earth Syst. Sci.*, **14**, 2387–2397.

Pensky, M., Vidakovic, B., and De Canditiis, D. (2007) Bayesian decision theoretic scale-adaptive estimation of a log-spectral density, *Statistica Sinica*, **17**, 635–666.

Pukelsheim, F. (1994) The three sigma rule, *Am. Stat.*, **48**, 88–91.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rousseau, J., Chopin, N., and Liseo, B. (2012) Bayesian nonparametric estimation of the spectral density of a long or intermediate memory Gaussian time series, *Ann. Statist.*, **40**, 964–995.

Scargle, J. D. (1982) Studies in astronomical time series ii. statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J.*, **263**, 835–853.

Schorfheide, F. and Song, D. (2015) Real-time forecasting with a mixed-frequency VAR, *J. Bus. Econ. Stat.*, **33**, 366–380.

Shumway, R. and Stoffer, D. (2006) *Time Series Analysis and Its Applications with R Examples*, Springer, New York.

Solomon, S. (2007) *Climate Change 2007 — The Physical Science Basis*, Cambridge University Press, Cambridge.

Steel, D. (1997) Producing monthly estimates of unemployment and employment according to the international labour office definition, *J. R. Statist. Soc.* A, **160**, 5–46.

Vaníček, P. (1971) Further development and properties of the spectral analysis by least-squares, *Astrophys. J.*, **12**, 10–33.

Vlag, P., Ortega Azurdy, S., van Loon, A., and Scholtus, S. (2010) Production of monthly turnover estimates with VAT in the Netherlands: Challenges and first results, Technical Report Deliverable 4.1-SGA 2010, Statistics Netherlands, The Hague.

Wahba, G. (1980) Automatic smoothing of the log periodogram, *J. Am. Statist. Ass.*, **75**, 122–132.

Wahba, G. (1981) Data-based optimal smoothing of orthogonal series density estimates, *Ann. Statist.*, **9**, 146–156.

Wang, Y., Stoica, P., Li, J., and Marzetta, T. L. (2005) Nonparametric spectral analysis with missing data via the EM algorithm, *Digit. Signal Process.*, **15**, 191–206.

Werner, A. and Holbrook, N. (2011) A Bayesian forecast model of Australian region tropical cyclone formation, *J. Climate*, **24**, 6114–6131.

Werner, B. (2006) Reflections on fifteen years of change in using the Labour Force Survey, *Labour Market Trends*, **114**, 257–277.

Whittle, P. (1957) Curve and periodogram smoothing, *J. R. Statist. Soc.* B, **19**, 38–63.