Nonparametric Inference for Periodic Sequences

Ying Sun, Jeffrey D. Hart, and Marc G. Genton

Department of Statistics
Texas A&M University
College Station, TX 77843-3143
(*sunwards@stat.tamu.edu; hart@stat.tamu.edu; genton@stat.tamu.edu*)

This article proposes a nonparametric method for estimating the period and values of a periodic sequence when the data are evenly spaced in time. The period is estimated by a "leave-out-one-cycle" version of cross-validation (CV) and complements the periodogram, a widely used tool for period estimation. The CV method is computationally simple and implicitly penalizes multiples of the smallest period, leading to a "virtually" consistent estimator of integer periods. This estimator is investigated both theoretically and by simulation. We also propose a nonparametric test of the null hypothesis that the data have constant mean against the alternative that the sequence of means is periodic. Finally, our methodology is demonstrated on three well-known time series: the sunspots and lynx trapping data, and the El Niño series of sea surface temperatures.

KEY WORDS: Consistency; Cross-validation; Model selection; Nonparametric estimation; Period.

1. INTRODUCTION

Sequences that are periodic or nearly so appear in many disciplines, including astronomy, meteorology, environmetrics, and economics. When little is known about the structure of such a sequence, it is important to have nonparametric methods of estimating both its values and its period. This article is concerned with these two estimation problems in the situation where one has observations at equally spaced time points. We consider the following model

$$Y_t = \mu_t + \varepsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

where $Y_1, \ldots, Y_n$ are the observations; $\mu_1, \ldots, \mu_n$ are unknown constants; and the errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent, identically distributed (iid) and mean-zero random variables. It is assumed that $\mu_t = \mu_{t+mp}$ for $t = 1, \ldots, p$, $m = 1, 2, \ldots$, and some integer $p \geq 2$. (It is implicit that p is the smallest such integer.) Of interest is estimating the period p and the sequence values $\mu_1, \ldots, \mu_p$. Our approach to the problem will be nonparametric. Even though p is finite, we assume no upper bound on it, and hence, there is no bound on how many sequence values have to be estimated.

One motivation for our methodology is the study of periodicity of El Niño effects, which are defined as sustained increases of at least 0.5 °C in average sea surface temperatures (SSTs) over the east-central tropical Pacific Ocean. Such occurrences that last less than 5 months are classified as El Niño conditions. An anomaly persisting for 5 months or longer is called an El Niño episode. Typically, this happens at irregular intervals of 2–8 years and lasts 9 months to 2 years; see Torrence and Webster (1999) and references therein. El Niño is associated with floods, droughts, and other weather disturbances in many regions of the world, particularly those bordering the Pacific Ocean. We will apply our proposed methodology to a series of SSTs to estimate the El Niño period.

Our method of estimating the period is based on cross-validation (CV). A candidate period q is evaluated by first "stacking," at the same time point, all data which are separated in time by a multiple of q and then computing a "leave-out-one-cycle" version of the variance for each of the q stacks of data. This variance will tend to be smallest when q equals p. A sequence of period p is also periodic of period mp for $m = 1, 2, \ldots$, and hence, a cross-validatory assessment of variance is required to avoid overfitting, that is, overestimating the period. It is shown that this method of estimating p is asymptotically equivalent to Akaike's information criterion (AIC) (Akaike 1973) when the errors in (1) are assumed to be iid Gaussian.

Our CV method is closely related to a method of Whittaker and Robinson (WR) (1967), which in turn is related to the Buys-Ballot (1847) table. The WR method (described in Section 4) is based on an R^2-like ratio of variances. The key difference between our method and that of WR is that theirs provides no objective way of comparing periods q and kq, where $k = 2, 3, \ldots$. As noted previously, our method provides protection against overfitting, whereas the WR R^2 criterion is guaranteed to be larger for period kq ($k = 2, 3, \ldots$) than it is for q.

We show that when p is sufficiently large, our period estimator $\hat{p}$ is *virtually* consistent, in the sense that $\lim_{n\to\infty} P(\hat{p} = p)$ increases to 1 as p increases. The fact that p is discrete makes the virtual consistency result remarkably strong. To wit, for all n sufficiently large, the probability that our period estimator is exactly equal to the truth is at least 0.99 for each $p \geq 16$.

Recent articles on nonparametric methods for estimating periodic functions include Hall, Reimann, and Rice (2000), Hall and Yin (2003), Hall and Li (2006), Genton and Hall (2007), and Hall (2008), the last of which is an excellent review of the subject. The assumption in these articles is that the underlying function, μ, of interest is defined at all the reals and has period that could be any positive number. As pointed out by Hall (2008), an appropriate *random* spacing of time points is needed in such problems to ensure consistent estimation of the period of

TECHNOMETRICS, FEBRUARY 2012, VOL. 54, NO. 1
DOI: 10.1080/00401706.2012.650499

μ. It is important to appreciate that the problem addressed in the current article is different from that just discussed. Our goal is to estimate the period of the sequence $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots\}$ even in cases where $\boldsymbol{\mu}$ may be a sampling of an underlying function μ. On the other hand, as will be discussed in Section 2.4, a period of p for $\boldsymbol{\mu}$ identifies a finite set of all possible *rational* periods for the sampled function μ. Furthermore, our methodology effectively eliminates some of these rational possibilities, unless the estimated period of $\boldsymbol{\mu}$ is prime. For equally spaced data, Quinn and Thomson (1991) proposed a periodogram-based method for period estimation, which is discussed in Section 2.5. The problem of estimating the various frequencies that make up a signal is given a general treatment in Quinn and Hannan (2001).

The rest of our article is organized as follows. Section 2.1 describes the CV method of estimating the period and sequence values, and discusses existing period estimation methodology based on the periodogram. Section 3 deals with asymptotic properties of the method, while Section 4 proposes a nonparametric test for periodicity. Simulations motivated by real data applications are reported in Section 5, and applications of our methods to well-known time series, including the El Niño series of SSTs, are presented in Section 6. Concluding remarks are given in Section 7, and a proof of virtual consistency is provided in the Appendix.

2. METHODOLOGY

Suppose we observe a time series $\{Y_t : t = 1, \ldots, n\}$ from the model (1), where the sequence $\boldsymbol{\mu}$ is periodic with (smallest) period p and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. random variables with zero means and finite variances σ^2. We propose a methodology for estimating p in Section 2.1 and discuss some related model selection criteria in Section 2.2. In Section 2.3, we propose methods for estimating the sequence values $\mu_1, \ldots, \mu_p$ within one period.

2.1 CV Method for Estimating an Integer Period

Let q be a candidate integer period with $2 \le q \le M_n$, where M_n is of smaller order than $\sqrt{n}$. For each $i = 1, \ldots, q$, we construct an estimator of μ_i by stacking all data that are separated in time by a multiple of q. So, at time point i, we have data $Y_i, Y_{i+q}, \ldots, Y_{i+qk_{q,i}}$, where $k_{q,i}$ is the largest integer such that $i + qk_{q,i} \le n$. For each relevant q, i, and j, define $Y_{qij} = Y_{i+(j-1)q}$ and $\mu_{qij} = \mu_{i+(j-1)q}$. Let

$$\mathrm{CV}(q) = \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} \left(Y_{qij} - \bar{Y}_{qi}^{j}\right)^2, \qquad (2)$$

where $\bar{Y}_{qi}^{j}$ is the average of $Y_{qi\ell}$, $\ell = 1, \ldots, k_{q,i}$, excluding Y_{qij}. We define a period estimator $\hat{p}$ to be the minimizer of $\mathrm{CV}(q)$ for $2 \le q \le M_n$.

The version of CV used in (2) is more closely related to that of Hart and Wehrly (1993) in the setting of functional data than to the version used in smoothing independent data. The method of Hart and Wehrly (1993) leaves out one curve in a sample of curves and then predicts the deleted curve by using all the others. Analogously, in our setting, all the data in one cycle (corresponding to a putative cycle length) are deleted and then data from other cycles are used to predict the omitted cycle. One may regard different cycles as independent copies of a multivariate random variable in the same way that functional data are independent copies of a curve.

The CV method may be motivated by taking the expectation of (2). Let $\varepsilon_{qij} = Y_{qij} - \mu_{qij}$ and $\bar{\varepsilon}_{qi}^{j} = \bar{Y}_{qi}^{j} - \bar{\mu}_{qi}^{j}$, where $\bar{\mu}_{qi}^{j}$ is the average of $\mu_{qi\ell}$, $\ell = 1, \ldots, k_{q,i}$, excluding μ_{qij}. Then, (2) may be written as

$$\mathrm{CV}(q) = \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} \left(\varepsilon_{qij} - \bar{\varepsilon}_{qi}^{j} + \mu_{qij} - \bar{\mu}_{qi}^{j}\right)^2.$$

Now, there are at most two distinct values of $k_{q,i}$ and these values differ by only 1. So,

$$\begin{aligned} E\{\mathrm{CV}(q)\} &\approx E\left\{\left(\varepsilon_{q11} - \bar{\varepsilon}_{q1}^{1}\right)^2\right\} + \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} \left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right)^2 \\ &\approx \sigma^2 \left(1 + \frac{1}{k_{q,1} - 1}\right) + \frac{k_{q,1}^2}{(k_{q,1} - 1)^2} \frac{1}{n} \\ &\quad \times \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} (\mu_{qij} - \bar{\mu}_{qi})^2 \\ &= \sigma^2 \left(1 + \frac{1}{k_{q,1} - 1}\right) + \frac{k_{q,1}^2}{(k_{q,1} - 1)^2} C_q \\ &= E_q, \end{aligned}$$

where $\bar{\mu}_{qi}$ is the average of $\mu_{qi\ell}$, $\ell = 1, \ldots, k_{q,i}$.

A motivation for our CV method stems from the following facts:

F1: The term C_q is 0 if and only if q is a multiple of p, as shown in the Appendix.
F2: By fact F1, $\min_{2 \le q < p} E_q > E_p$ for all n sufficiently large.
F3: Fact F1 and the definition of $k_{q,i}$ imply that $\min_{p < q \le M_n} E_q > E_p$ for all n sufficiently large.

More succinctly, we may say that over any bounded set of q-values containing p, the minimizer of $E\{\mathrm{CV}(q)\}$ is p for all sufficiently large n.

The *necessity* of using CV follows upon considering what happens when $\bar{Y}_{qi}^{j}$ in (2) is replaced by $\bar{Y}_{qi}$, the average of $Y_{qi\ell}$, $\ell = 1, \ldots, k_{q,i}$. Call the resulting criterion $V(q)$. Then, for $q = kp$ and $k = 1, 2, \ldots$,

$$\begin{aligned} E\{V(q)\} &\approx E\{(\varepsilon_{q11} - \bar{\varepsilon}_{q1})^2\} = \sigma^2 \left(1 - \frac{1}{k_{q,i}}\right) \\ &\approx \sigma^2 \left(1 - \frac{kp}{n}\right). \end{aligned}$$

The rightmost quantity immediately above is *decreasing* in k, and so, the minimizer of $V(q)$ will tend to be a large multiple of p.

Of course, the considerations above do not prove that CV yields a good estimator of p. That question will be addressed in Section 3.

2.2 Model Selection Criteria for Period Estimation

Cross-validation (CV) is typically used as a model selection tool. Our use of CV may appear to be different, but in fact, each candidate period q corresponds to a model for the sequence $\boldsymbol{\mu}$ consisting of the q parameters $\mu_1, \ldots, \mu_q$. With this perspective, it is thus natural to consider model selection criteria other than CV in estimating q.

If we assume that the errors in model (1) are i.i.d. Gaussian, then AIC has the form

$$\mathrm{AIC}(q) = n\log\hat{\sigma}_q^2 + 2(q+1), \quad q = 2, \ldots, M_n, \tag{3}$$

where $\hat{\sigma}_q^2 = V(q)$, as defined in Section 2.1. Minimizing AIC(q) with respect to q provides an estimator of p. It is not difficult to argue that under appropriate conditions on M_n, the estimator of p obtained from AIC is asymptotically equivalent to the CV estimator of Section 2.1.

The usual variations of AIC are also possible. These have the general form

$$C(q; c_n) = n\log\hat{\sigma}_q^2 + c_n(q+1), \quad q = 2, \ldots, M_n, \tag{4}$$

for some (specified) penalty c_n. Bayes's information criterion (Schwarz 1978), or BIC, corresponds to $c_n = \log n$. Another possibility is a Hannan–Quinn-type criterion with $c_n = 2c\log\log n$ ($c > 1$) (Hannan and Quinn 1979). Both these methods have $c_n > 2$ for all n large enough and hence have a smaller probability of overestimating the true model dimension than does AIC. In fact, BIC and the Hannan–Quinn criterion produce consistent estimators of the true model dimension, whereas CV and AIC do not. This is true in the setting of the current article as well as in many scenarios where a sequence of *nested* models is under consideration.

It will be shown in Section 3 that although CV (and hence AIC) does not produce consistent estimators of p, its asymptotic probability of selecting p is very close to 1 for p larger than 15. We call this last property *virtual* consistency.

Taking the penalty c_n in (4) to be fixed but larger than 2 leads to a criterion with an even smaller probability of overestimating the true period as compared with CV. Constants motivated by testing considerations will be proposed in Section 4.

2.3 Estimating Sequence Values

Given that the period is q, we desire estimates for the sequence values $\mu_1, \ldots, \mu_q$. If the errors are assumed to be i.i.d. Gaussian, then it is straightforward to verify that the maximum likelihood estimates of these parameters are the means $\bar{Y}_{q1}, \ldots, \bar{Y}_{qq}$, as defined in Section 2.1. For the moment, we set aside the effect that estimation of p has on estimates of the sequence values and assume that $q = p$. Obviously, whenever the errors have two finite moments, $\bar{Y}_{p1}, \ldots, \bar{Y}_{pp}$ are consistent for $\mu_1, \ldots, \mu_p$, since as n tends to ∞, the number of observed full cycles, $k_{p,i}$, also tends to ∞.

The James–Stein theory (Stein 1964) implies that the estimator $\bar{\boldsymbol{Y}} = (\bar{Y}_{p1}, \ldots, \bar{Y}_{pp})$ is inadmissible when $p \geq 3$. A better estimator of a normal mean vector may be obtained by shrinking the sample mean vector toward a specified point. In our setting where the data are time-ordered, a natural form of shrinkage is to smooth means that are close to each other in time. More precisely, we may estimate μ_i by

$$\hat{\mu}_i = \sum_{t=1}^{q} \bar{Y}_{pt} K_h(|t-i|), \tag{5}$$

where $\sum_{t=1}^{q} K_h(|t-i|) = 1$, K_h decreases monotonically to 0 on $[0, \infty)$, and h is a smoothing parameter that dictates how quickly K_h goes to 0. When the sequence $\boldsymbol{\mu}$ is "smooth," in the sense that $\sum_{t=1}^{p-1}(\mu_{t+1} - \mu_t)^2$ is sufficiently small, then the kernel-type smoother (5) can have smaller mean squared error than $\bar{Y}_{pi}$.

Asymptotically, the use of kernel smoothing cannot be expected to yield a large improvement over the simple mean $\bar{Y}_{pi}$. This is because an estimator of the form (5) must eventually collapse to $\bar{Y}_{pi}$ to be consistent. This entails that an optimal estimator of the form (5) will have mean squared error asymptotic to $p\sigma^2/n$, the same as that of $\bar{Y}_{pi}$. The situation where a substantial improvement could be obtained by smoothing is when n is small enough that p/n is fairly large. In this case, the variance of $\bar{Y}_{pi}$ can be large, and so smoothing, while introducing bias, can be beneficial if it reduces variance substantially.

Obviously, maximum likelihood estimates of sequence values can be profoundly affected by misspecification of the value of p. Fortunately, our CV method is such that asymptotically, it only chooses p or a multiple of p, and if p is more than 15, the probability of choosing a value larger than p is extremely small. Let $\hat{p}$ be the estimate of p obtained with the CV method. A useful check on this estimate is to plot $\hat{\mu}_1, \ldots, \hat{\mu}_{\hat{p}}$ against time and check for evidence of more than one cycle. If, for example, a pattern seems to repeat itself twice in the sequence of $\hat{\mu}_i$s, then there is evidence that $\hat{p} = 2p$.

2.4 Noninteger Periods

In some cases, the observed data are a sampling of some periodic function μ that is defined on all of $\mathbb{R}^+$. In this case, the period of μ could be any positive number, and in particular is not necessarily a multiple of the spacing between consecutive time points.

Hall (2008) remarked that in such cases, the time points at which data are observed should not be equally spaced, since then, "for many values of the period (in particular when the period is a rational multiple of the spacing), consistent estimation is not possible." So, when the period of μ can be arbitrary, *design* (i.e., placement) of the time points at which observations are made is important. Hall (2008) pointed out that the appropriate use of randomness in selecting design points ensures that a consistent period estimator can be constructed.

Our point of view in the current article is that we will try to evaluate periodicity as best as we can when the time points *are* evenly spaced, which is obviously a prevalent situation in time series analysis. Suppose that the observations are taken at time points $1, 2, \ldots$. Then, the estimable parameters are the function values $\mu(1), \mu(2), \ldots$, which need not be periodic, even though μ is. Conversely, if the sequence *is* periodic, its period need not be the same as that of μ. However, when the period of μ is sufficiently large relative to the sampling rate, then practically speaking, the sequence $\mu(1), \mu(2), \ldots$ will have the same period as μ.

If integer $\hat{p}$ is the minimizer of our CV criterion, then one could ask: "What periods for μ are consistent with a period of $\hat{p}$ for the sequence $\mu(1), \mu(2), \ldots$?" The answer is "periods $\hat{p}, \hat{p}/2, \hat{p}/3, \ldots$." If we assume that the true period is at least 2 (a common assumption for evenly spaced data), then there are only finitely many periods consistent with $\hat{p}$, with the smallest being $\hat{p}/k_{\hat{p}}$, where k_p is the largest integer such that $p/k_p \geq 2$.

Let $S_{\hat{p}} = \{\hat{p}, \hat{p}/2, \dots, \hat{p}/k_{\hat{p}}\}$. How well is each element of $S_{\hat{p}}$ supported by the data? In Section 3, we will show how to define a confidence set $\mathcal{C}$ of coefficient $1-\alpha$ for p given $\hat{p}$. A certain subset of the integer elements of $S_{\hat{p}}$ will be in $\mathcal{C}$. Of course, $\hat{p}$ is always in $\mathcal{C}$, and the other members are determined by α. Any noninteger element of $S_{\hat{p}}$ that is expressed in simplest form is just as well supported as is $\hat{p}$. Finally, let $\hat{p}/k$ be a noninteger element of $S_{\hat{p}}$ that is not in simplest form. Let r be the greatest common divisor of $\hat{p}$ and k. Then, a period of p/k for μ would produce the smallest integer period of $\hat{p}/r$ in $\boldsymbol{\mu}$, and so, the likelihood of $\hat{p}/k$ can be judged by whether or not $\hat{p}/r$ is in the confidence set for p.

2.5 The Periodogram

By far, the most popular means of identifying periodic behavior in time series data is the periodogram. Authors have often encountered the attitude that the periodogram is the ideal tool for solving the problem of the current article. In our opinion, this is a misguided attitude. We certainly feel that the periodogram is a useful tool, but we also believe that its usefulness can be enhanced by complementing it with the time domain method introduced in this article. Likewise, our method is well complemented by the periodogram.

First of all, our subsequent comments are predicated on having observations that are evenly spaced in time. In other words, we do not have the luxury of randomly distributed design points over the observation interval. In the latter scenario, Hall and Li (2006) showed that the periodogram can be used to form consistent estimators of an arbitrary period. It is worth noting, however, that even in this scenario, Hall and Li (2006) showed that the periodogram method is inefficient in comparison with one based on a least squares, time domain approach.

It seems that part of the confusion about the periodogram stems from not adequately differentiating between two problems. One of these is the problem of estimating the smallest period of a periodic function, which is the subject of the current article. The second problem is that of identifying sinusoidal functions that are highly correlated with the data. We feel that the periodogram is very well suited for solving the latter problem, but not so well suited for the former.

The periodogram $\hat{f}$ is

$$\hat{f}(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} Y_t e^{2\pi i(t-1)\omega}\right|^2, \quad -1/2 \leq \omega \leq 1/2. \quad (6)$$

It is important to note that the function defined by (6) for *all* ω is periodic with period 1. This is an indication of the so-called *aliasing* effect, that is, frequencies larger than $1/2$ cannot be distinguished from lower frequencies. Now, suppose one wishes to determine the extent to which the data are correlated with a single sinusoidal component of period q. The only limiting factor faced by the periodogram in this regard is that q should be at least 2, since only in that case is the corresponding frequency no bigger than $1/2$.

The situation becomes more complicated when one wishes to use the periodogram to determine the smallest period p of a periodic, sampled function μ. A potential advantage of periodogram-based methods over our CV methodology is that they can be used to evaluate noninteger periods. However, there are also multiple factors that tend to limit this advantage. First of all, the function should not be highly correlated with sinusoids of frequency larger than $1/2$. Otherwise, due to aliasing, the periodogram will have substantial power at frequencies other than the fundamental frequency $1/p$ and its harmonics $2/p, 3/p \dots, m_p/p \leq 1/2$. This additional power can defeat methods that might otherwise be able to consistently estimate p.

The condition that μ be orthogonal to sinusoids of frequency larger than $1/2$ is tantamount to assuming that μ have Fourier series of the form

$$\mu(t) = \rho_0 + \sum_{j=1}^{m} \rho_j \cos\left(\frac{2\pi t j}{p} - \phi_j\right), \quad t \geq 0, \quad (7)$$

where $m \leq p/2$. Suppose this is the case. Then, assuming m to be known, Quinn and Thompson (1991) (QT) proposed that p be estimated by the maximizer, $\hat{p}_{QT}$, of $P(q) = \sum_{j=1}^{m} \hat{f}(j/q)$ over $q \geq 2m$. They showed that if (i) μ has the form (7), (ii) p is at least $2m$, and (iii) at least $[m/2]+1$ of $\rho_1, \dots, \rho_m$ are nonzero, then $\hat{p}_{QT}$ is a consistent estimator of p. But (ii) and (iii) are very restrictive conditions. Only if $m = 1$ is the period effectively unrestricted. Any other choice for m, at the very least, prevents one from evaluating periods of less than 4. In short, a key problem with the QT method is that it prevents objective evaluation of both m and p, since any choice of m automatically excludes the possibility of evaluating periods of less than $2m$.

So, taking m to be large is problematic, but taking it to be too small can also be a problem. A simple example suffices to illustrate this point. Suppose that the true function is

$$\mu(t) = 2\cos\left(\frac{2\pi 2t}{8}\right) + \cos\left(\frac{2\pi 3t}{8}\right), \quad t \geq 0,$$

which is periodic with period 8. Here, the Fourier series is truncated at 3, but suppose that one unwittingly uses $m = 2$ in the QT method. Then, the true period of 8 will not be distinguishable from a period of 4, since $\hat{f}(1/8) + \hat{f}(1/4)$ and $\hat{f}(1/4) + \hat{f}(1/2)$ estimate the same quantity. In contrast, *our time domain method has no problem whatsoever in distinguishing the periods 8 and 4 in this case*. With probability tending to 1 as $n \to \infty$, our method will prefer period 8 to period 4.

The discussion in this section is not meant to imply that the periodogram can never be a useful tool for estimating a period, nor is it meant to imply that our CV method is superior to the periodogram. Both methods have their pros and cons, and neither will always be able to identify the true period when the data are evenly spaced. Our opinion is that both methods should be applied to any given dataset. There are cases where the CV method can shore up deficiencies of the periodogram and vice versa. For example, if the QT method is applied with m larger than 1, then the CV method can be used to investigate the possibility of periods smaller than $2m$. Conversely, if CV is minimized at a rather large value of q, the QT method could be used to investigate the possibility that the large integer period is actually due to a smaller, noninteger period in the sampled function; see the discussion in Section 2.4.

3. ASYMPTOTIC PROPERTIES OF THE CV METHOD

We begin with a theorem that describes the asymptotic behavior of the CV period estimator.

Theorem 1. Suppose that model (1) holds with $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. $N(0, \sigma^2)$ random variables. The sequence $\boldsymbol{\mu}$ is assumed to be periodic with the smallest period p, which is a positive integer. Let $\{M_n : n = 1, 2 \dots\}$ be a sequence of positive integers that tends to infinity and is $o(\sqrt{n})$, and define

$$\hat{p} = \text{argmin}_{q \in \{1,\dots,M_n\}} \text{CV}(q).$$

Let $\{Z_{m,i} : m = 1, 2, \dots; i = 1, \dots, pm\}$ be a collection of random variables such that each $Z_{m,i} \sim N(0, 1)$ and any finite subset has a multivariate normal distribution. Furthermore,

C1. For each m, the random variables $Z_{m,1}, \dots, Z_{m,pm}$ are mutually independent.

C2. For each $r > m$, $Z_{m,i}$ and $Z_{r,j}$ are independent unless $i - j = p(rs - m\ell)$ for some pair of nonnegative integers ℓ and s, in which case

$$\text{corr}(Z_{m,i}, Z_{r,j}) = \frac{1}{\sqrt{mr}}.$$

Define $S_m = \sum_{i=1}^{pm} Z_{m,i}^2$ for $m = 1, 2, \dots$. Then, if $\mathcal{S}_n$ is the set $\{1, 2, \dots, M_n\}$ excluding multiples of p,

$$\lim_{n\to\infty} P(\hat{p} \in \mathcal{S}_n) = 0,$$

and

$$\lim_{n\to\infty} P(\hat{p} = jp) = P\left(\bigcap_{m=1}^{\infty} \{S_j - S_m + 2p(m-j) \geq 0\}\right), \quad j = 1, 2, \dots.$$

Defining $\tau(p) = \lim_{n\to\infty} P(\hat{p} = p)$, we have

$$\tau(p) \geq 1 - P(S_1 - S_2 + 2p < 0) - \sum_{m=3}^{\infty} P(S_m > 2p(m-1)). \tag{8}$$

The random variable $(S_1 - S_2)/p$ converges in probability to -1 as $p \to \infty$ and hence $P(S_1 - S_2 + 2p < 0)$ tends to 0 as $p \to \infty$. By applying Bernstein's inequality to each term of the infinite series on the right-hand side of (8), that series may be bounded by a quantity that tends to 0 as $p \to \infty$. These facts show that $\tau(p)$ tends to 1 as $p \to \infty$, which proves the virtual consistency of $\hat{p}$.

For small values of p, simulation of the asymptotic distribution of $\hat{p}$ was conducted to approximate $P(\hat{p} = p)$. The results are shown in Table 1 and were obtained using $n = 5000$, $M_n = 20p$, and 5000 replications for each p. The asymptotic probability of choosing the correct period increases quickly as p increases. The probability is about 0.96 for p as small as 9 and increases to about 0.99 when p is 15.

Table 1. Asymptotic probability of choosing the correct period for small p

p	$\hat{\tau}(p)$	p	$\hat{\tau}(p)$
1	0.489	9	0.957
2	0.694	10	0.968
3	0.791	11	0.971
4	0.854	12	0.977
5	0.892	13	0.981
6	0.908	14	0.983
7	0.933	15	0.988
8	0.954	16	0.990

A few remarks are in order about the conditions of Theorem 1. The assumption that the errors are Gaussian is important mainly for simplifying the proof. With more effort, the theorem can be extended to cases with nonnormal errors so long as the average of an unbounded number of error terms is asymptotically normal. The requirement that the upper bound, M_n, on the period be $o(\sqrt{n})$ is slightly annoying, but actually less restrictive than conditions in most articles on AIC or related methods. For example, the asymptotics of Hannan and Quinn (1979), Schwarz (1978), and Shibata (1976) all assume a *fixed* upper bound on the model dimension as sample size tends to infinity. In practice, our own approach to choosing M_n is to initially err on the side of a very large value such as $n/2$ or $n/3$. If the resulting minimizer of CV(q) is nowhere near the upper bound, then there is no dilemma. If it is near the upper bound, then one may question whether the result is spurious, that is, whether it is merely due to a large variation in $CV(q)$ for a large q.

Using test inversion and the result of Theorem 1, one may obtain asymptotic $(1-\alpha)100\%$ confidence sets for p. A little thought makes it clear that such a set is a subset of $\{\hat{p}, \hat{p}/k_1, \dots, \hat{p}/k_{j(p)-1}, 1\}$, where $2 \leq k_1 < k_2 < \dots < k_{j(p)} = \hat{p}$ are all the divisors of $\hat{p}$. This entails that for prime $\hat{p}$, the confidence set contains at most $\hat{p}$ and 1. Let p_0 be the observed value of $\hat{p}$ and p_0/k be a candidate for the confidence set. Then, p_0/k is included in this set if and only if $P(\hat{p} \geq p_0 | p = p_0/k) > \alpha$, where P denotes probability for the limiting distribution of $\hat{p}$. Knowledge of the asymptotic distribution of $\hat{p}$ allows one to determine conditions under which the large-sample confidence set contains only $\hat{p}$. For $\alpha = 0.05$, simulation indicates that this occurs when $\hat{p}$ is at least 21.

One may justifiably be skeptical of using large-sample theory to obtain a confidence set when n is not "large." Alternatively, a parametric bootstrap could be used to approximate confidence sets in small samples. As discussed before, this can be done using test inversion. To determine if a given value of p, say $\tilde{p}$, is included in the confidence set, one may approximate the probability distribution of $\hat{p}$ on the assumption that $p = \tilde{p}$. This can be done as follows. Let $\hat{\mu}_1, \dots, \hat{\mu}_{\tilde{p}}$ and $\hat{\sigma}^2$ be the Gaussian maximum likelihood estimates of $\mu_1, \dots, \mu_{\tilde{p}}$ and σ^2, respectively, on the assumption that $\tilde{p}$ is the true period. Then, one may generate many samples of size n from a version of model (1) that has period $\tilde{p}$, $\mu_1, \dots, \mu_{\tilde{p}}$ equal to $\hat{\mu}_1, \dots, \hat{\mu}_{\tilde{p}}$ and $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. as $N(0, \hat{\sigma}^2)$. From each sample, the CV estimate, $\hat{p}^*$, of period is obtained, and the empirical distribution of all these estimates provides an approximation to the required distribution.

4. TESTING FOR PERIODICITY

Our proposed methodology for estimating a period may be easily adapted to test the null hypothesis, H_0, that all data have the same mean against the alternative that the sequence of means is periodic with period larger than 1. We use a so-called *order selection test*, which dates to Fellner (1974) and has been studied extensively by Eubank and Hart (1992) and Hart (1997). The idea is to reject H_0 if and only if the period chosen by a model selection criterion is larger than 1. Here, we propose that AIC be used for this purpose since the resulting statistic is more intuitively appealing than one based on our CV criterion.

It is easy to check that the minimizer of (3) over $q = 1, \ldots, M_n$ is larger than 1 if and only if

$$T_n = \max_{2 \le q \le M_n} \left[\frac{n \log\left(\hat{\sigma}_1^2 / \hat{\sigma}_q^2\right)}{q - 1} \right] > 2. \tag{9}$$

Now, consider the test that rejects H_0 if and only if $T_n > 2$. Table 1 and the asymptotic equivalence of AIC and CV imply that this test has asymptotic level of approximately $1 - 0.489 = 0.511$, which is unacceptably large. This problem is easily dealt with by using the appropriate null percentile of T_n in place of 2.

Let $t_{n,\alpha}$ be the $1 - \alpha$ percentile of the null distribution of T_n. Then, rejecting H_0 if and only if $T_n > t_{n,\alpha}$ is equivalent to rejecting it if and only if the criterion (4) with $c_n = t_{n,\alpha}$ is minimized at a value larger than 1. This provides an intuitively appealing way of choosing the penalty constant c_n. For example, if one uses $C(q; t_{n,0.05})$, then the hypothesis of periodicity is tenable at the 0.05 level when $C(q; t_{n,0.05})$ is minimized at $q > 1$. Furthermore, all Table 1 entries for $C(q; t_{n,0.05})$ are at least 0.95, and the virtual consistency property would hold for small p as well as for large.

The large-sample null distribution of T_n is obtained almost immediately from Theorem 1, although at this point, we do not undertake a description of this distribution. In practice, we propose that simulation be used to approximate the null distribution of T_n for the sample size n at hand. This can be done either by generating random samples of size n from a standard normal distribution or by an appropriate bootstrap algorithm. Such an algorithm could accommodate whatever special features the data are felt to have, such as heteroscedasticity or serial correlation.

We next state a theorem concerning consistency of tests based on T_n, a proof of which is given in the Appendix.

Theorem 2. Let μ be a function of the positive reals that is periodic with rational period $j/m \ge 1$, which is expressed in simplest form. Suppose that model (1) holds with $\mu_t = \mu(t)$ for $t = 1, 2, \ldots$ and $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. $N(0, \sigma^2)$ random variables. Let $\{M_n : n = 1, 2 \ldots\}$ be a sequence of positive integers that tends to infinity and is $o(\sqrt{n})$, define T_n as in (9), and let $\mathcal{T}(B)$ be the test that rejects H_0 : "μ is identical to a constant" if and only if $T_n \ge B$. Then, if $B \ge 2$, $\mathcal{T}(B)$ is consistent against any alternative to H_0 such that (i) the sequence $\mu(1), \mu(2), \ldots$ is periodic with period larger than 1, or (ii) j/m is larger than 1 and there exists no a such that $\mu(x) = a$ at more than $j - 1$ distinct values of x in $[0, j/m)$.

Theorem 2 shows that our proposed test is consistent in situations where it should be, namely when the sequence $\boldsymbol{\mu}$ is periodic with period larger than 1. However, our test can also detect many other types of periodic behavior in the underlying *sampled* function μ. In essence, our test is consistent when μ has rational period larger than 1 and does not oscillate too much within one cycle. For example, suppose μ is periodic with period $1.1 = 11/10$. Then, as long as μ does not take on the same value more than 10 times in one cycle, our test will consistently detect that the sampled function μ is periodic.

Whittaker and Robinson (1967) proposed a useful R^2-like statistic that measures the proportion of total variability in the data due to a particular integer period. Their statistic is

$$R^2(q) = \frac{\sum_{t=1}^n (\bar{Y}_{t,q} - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2},$$

where $\bar{Y}$ is the overall mean, and $\bar{Y}_{t,q}$ is the average of Y_t and all data whose indices differ from t by a multiple of q. Interestingly, one can show that $R^2(q) \le R^2(kq)$ for each $q = 1, 2, \ldots$ and each $k = 2, 3, \ldots$. Therefore, while being a useful descriptive device, $R^2(q)$ suffers from the same defect as does R^2 for a sequence of nested regression models. To wit, a given regression model is guaranteed to have a larger R^2 than a smaller nested model, and hence, R^2 is not a reliable means of model selection.

5. SIMULATION STUDIES

5.1 Virtual Consistency

To study the behavior of the CV period estimator, we performed simulations with an intermediate value of the period, that is, $p = 43$, which is our estimate of the El Niño period. First, we consider model (1) with $Y_t = a \sin(2\pi t/43) + \varepsilon_t$, where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. normal with mean 0 and variance σ^2. In this case, $\mu_t = a \sin(2\pi t/43)$ and $p = 43$ is the true period of the sequence $\boldsymbol{\mu}$. Obviously, strong signals and low error levels would lead to a more efficient estimate, and indeed what matters is the ratio σ/a. Thus, in our simulation studies, we fix $a = 1$ and consider $\sigma = 0.5, 1, 1.5$. For all cases, the set over which the objective function $\mathrm{CV}(q)$ was searched is taken to be $\{12, 13, \ldots, 96\}$, and the number of replications of each setting is 1000.

Figure 1 shows how the probabilities of choosing $p = 43$ increase as the sample size n increases for each $\sigma = 0.5, 1, 1.5$. It is clear that the convergence is slower for larger errors. For $\sigma = 1$, $\hat{p}$ is between 41 and 44 approximately 86% of the time at $n = 200$ and about 99% of the time at $n = 300$.

Averages of CV curves are shown in Figure 2. Here, we see what is typical of individual curves, namely they tend to have local minima at or near multiples of p. For example, at $\sigma = 1$, the plots of $E\{\mathrm{CV}(q)\}$ (left panels of Figure 2) are minimized at the true $p = 43$ and have a local minimum at its multiple 86.

Comparing the top and bottom panels of Figure 2 shows how closely the average CV curves agree with the theoretical expectation when the sample size increases from 200 to 1000.

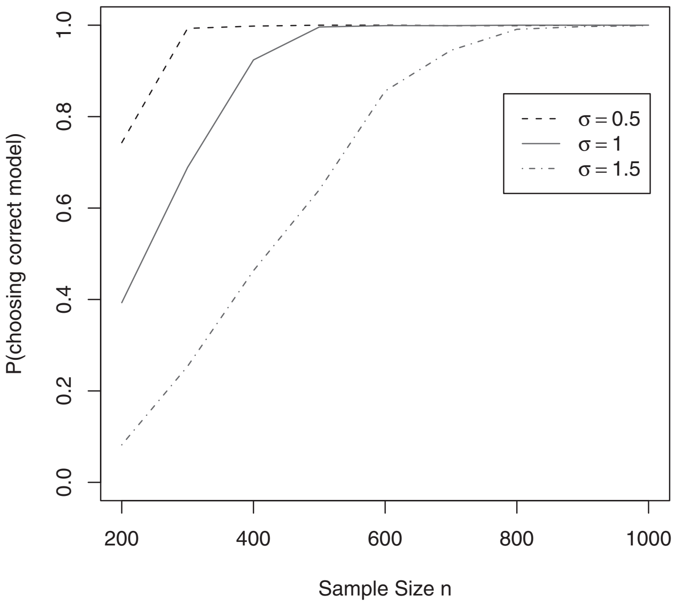

Figure 1. $P(\hat{p} = p)$ for the CV method at different error levels. The online version of this figure is in color.

Also, it is seen that at $n = 1000$, the value of the CV curve at its minimum is very close to σ^2.

In this simulation, we assumed Gaussian error terms, but model (1) does not require this assumption. To confirm the performance of the CV method with non-Gaussian errors, we consider a case where the errors have a Student's t distribution with five degrees of freedom, $t(5)$. Since the previous simulation shows how error levels affect the consistency, we scale the $t(5)$ distribution to have variance 1 and compare the consistency of the CV method to that of the standard normal error case. The probability of choosing the correct period showed no essential difference from the Gaussian error case for different sample sizes n. Therefore, this supports the remark that our CV method is not restricted to Gaussian errors.

5.2 Robustness of the CV Method

Model (1) assumes homogeneity of the error terms. In practice, however, many datasets exhibit periodicity but with cycle amplitudes that appear to change randomly. Such behavior can be modeled with heteroscedastic errors. To show the robustness of the CV method for this situation, we simulate data with a periodic mean function and errors whose standard deviations change randomly from one cycle to the next.

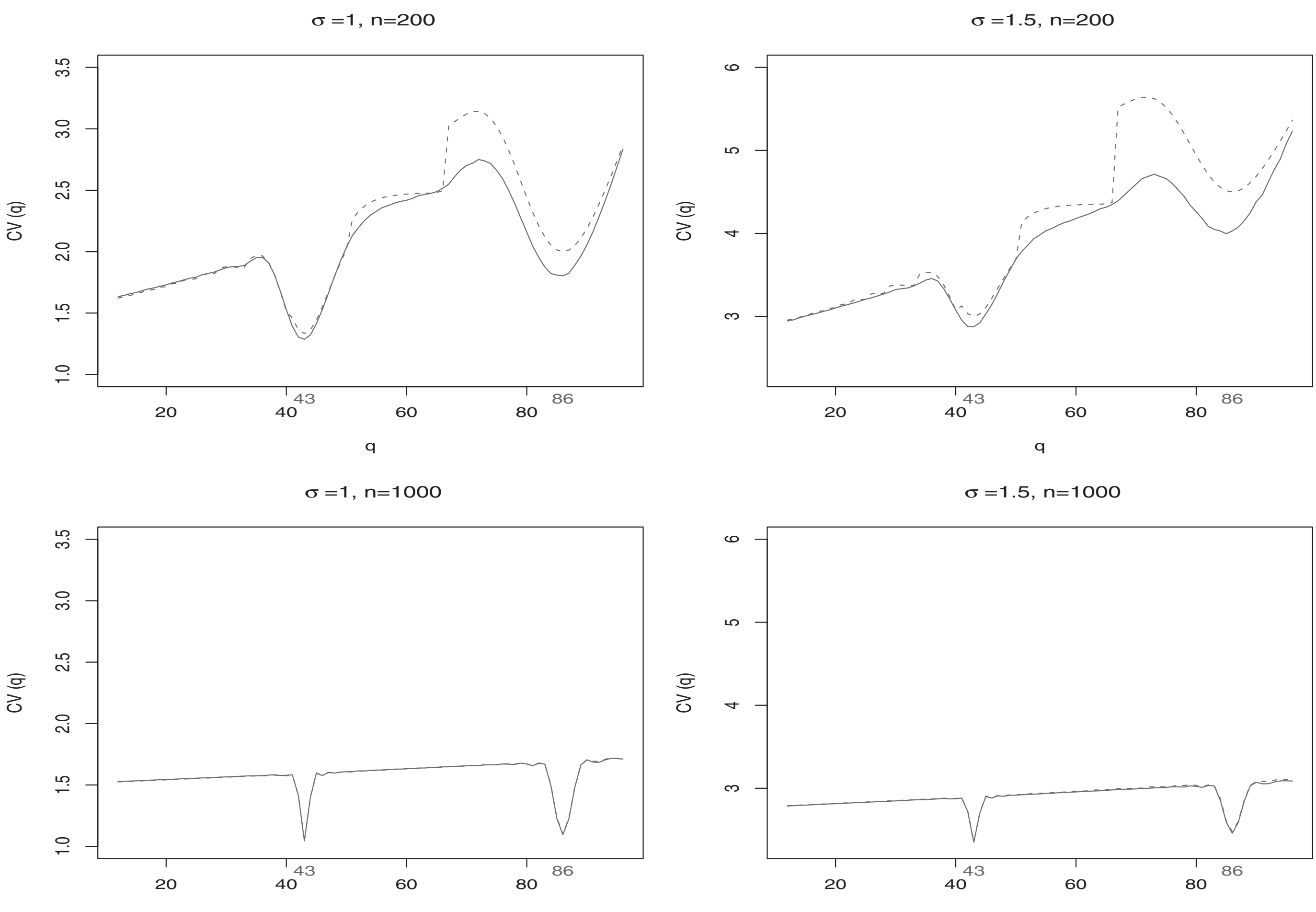

Figure 2. Objective function plot for period candidates q from 12 to 96. The solid curve denotes theoretical expectation of CV(q), and the dashed curve is the average of all 1000 CV curves from a given simulation. The online version of this figure is in color.

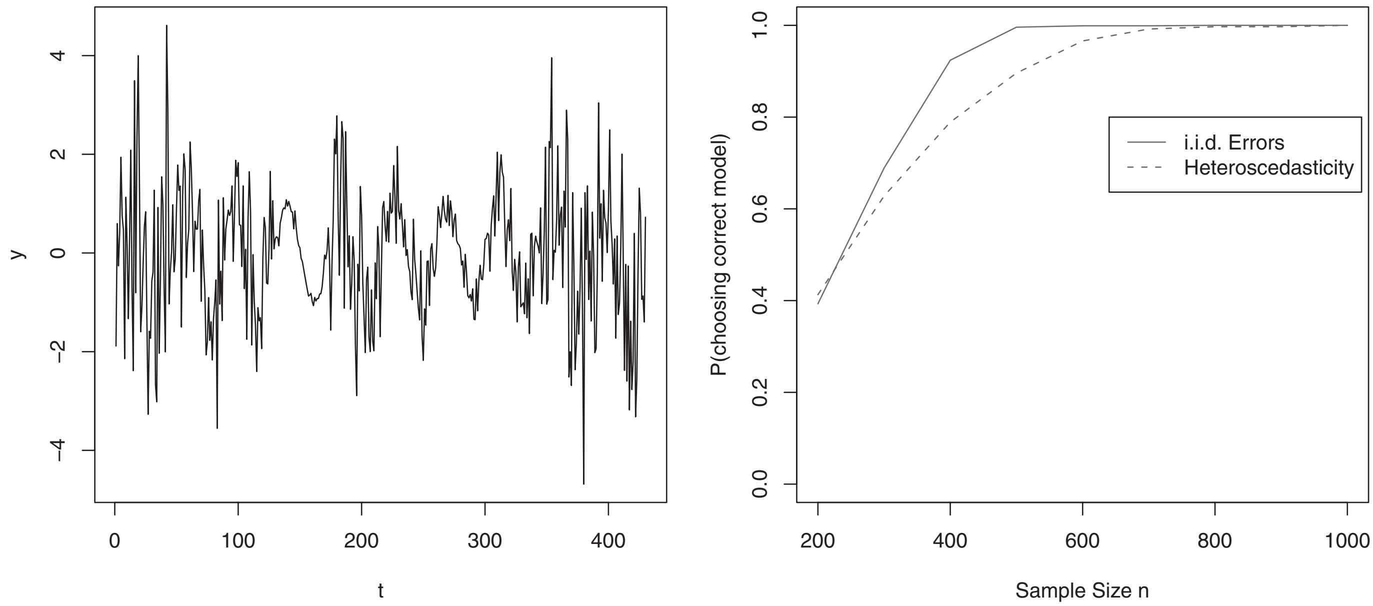

Figure 3. Left panel: an example of heteroscedastic errors with standard deviations $\sigma_1, \ldots, \sigma_k$ taken to be i.i.d. Gamma(1,1) and $p = 43$, $n = 430$. Right panel: virtual consistency of CV method for heteroscedastic errors, $p = 43$. The online version of this figure is in color.

Specifically, suppose that model (1) holds with $\mu_t = \sin(2\pi t/43)$, $n = 43k$, and conditional on $\sigma_1, \ldots, \sigma_k$, $\varepsilon_1, \ldots, \varepsilon_n$ are independent and normally distributed with 0 means and variances as follows:

$$\mathrm{var}(\varepsilon_{i+43(j-1)}) = \sigma_j^2, \quad j = 1, \ldots, k, \ i = 1, \ldots, 43.$$

The standard deviations $\sigma_1, \ldots, \sigma_k$ are taken to be i.i.d. Gamma(1,1) with mean 1. Figure 3 shows one simulated example where the data are generated as just described with $k = 10$, that is, $n = 430$. Simulations at various sample sizes and using our heteroscedastic model were conducted. The results are shown in Figure 3. The proportion of cases in which $\hat{p}$ was equal to p was only slightly less for the heteroscedastic errors than it was for homoscedastic errors with $\sigma = 1$.

Especially for time series data, another interesting question is the robustness of the CV method to serial correlation among the data. We anticipate that the result of our Theorem 1 is not greatly affected by errors that follow an mth order moving average (MA(m)) process as long as $m \leq p$. Presumably, the method will also be fairly robust to other types of covariance stationary errors for which the autocorrelation function damps out quickly at lags larger than p. A simulation example shows this to be the case, at least when the errors are first-order autoregressive. The right

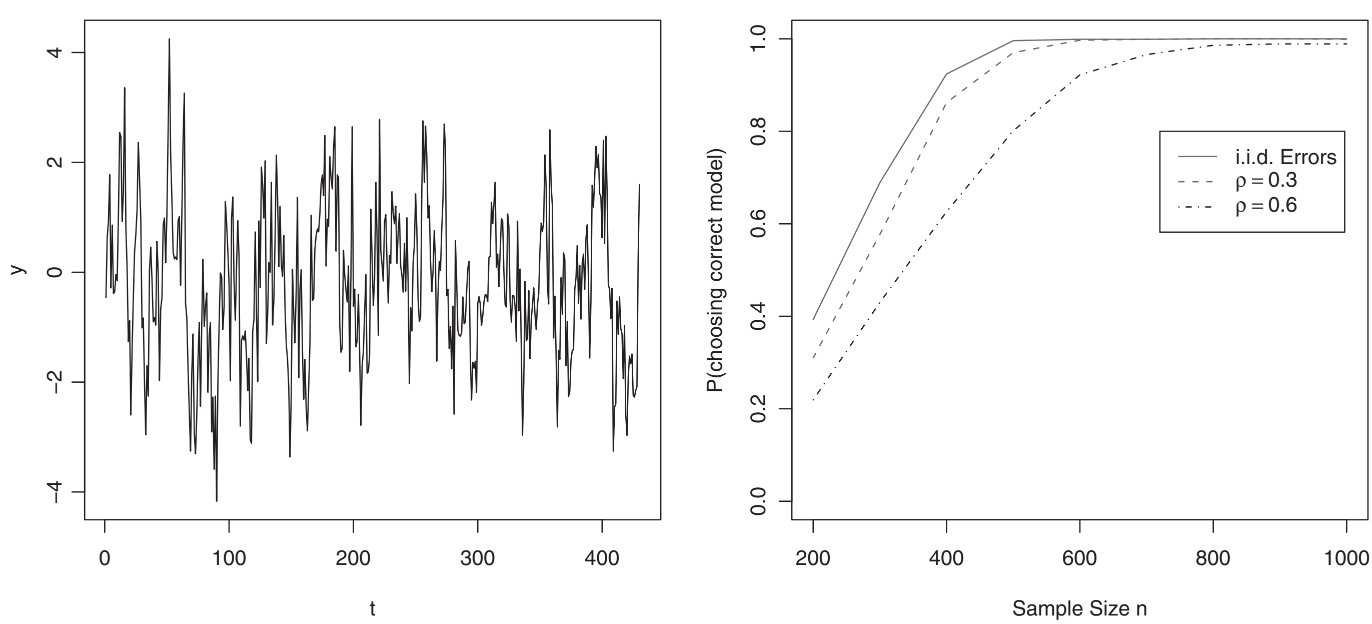

Figure 4. Left panel: an example of autoregressive errors with first-lag correlation equal to 0.6, $p = 43$, $n = 430$, and $\sigma = 1$. Right panel: virtual consistency of the CV method for $p = 43$ and errors that follow a first-order autoregressive process with $\sigma = 1$. The online version of this figure is in color.

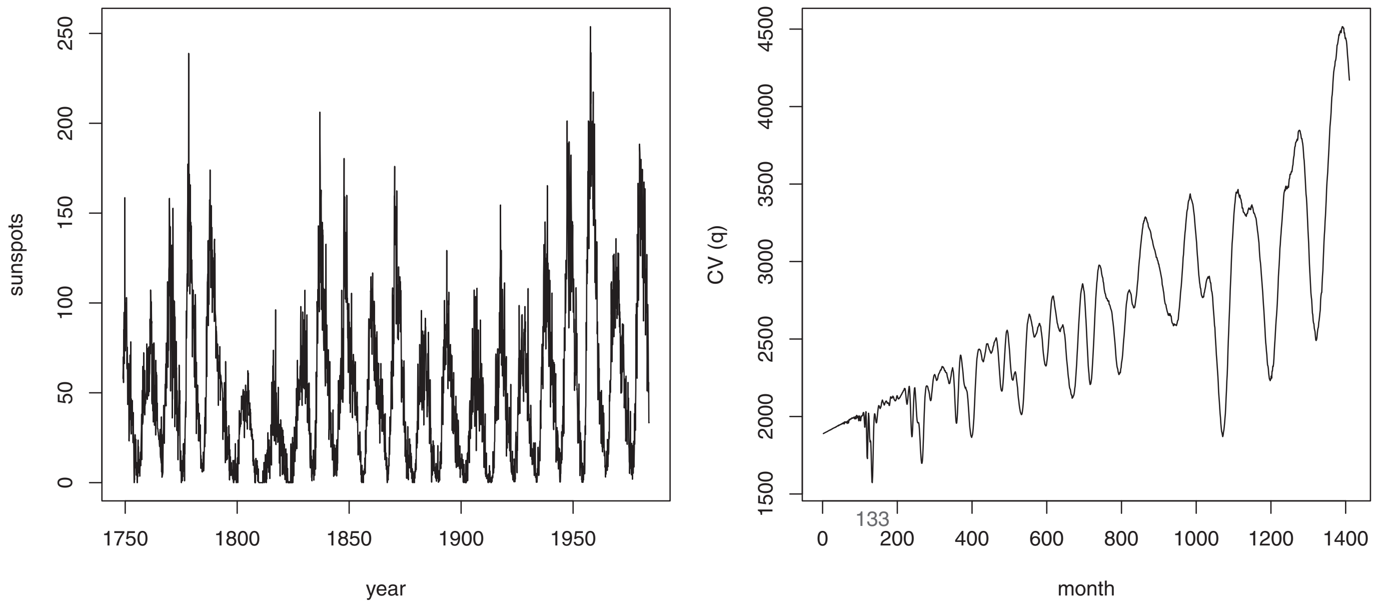

Figure 5. Sunspots data: the left panel is a plot of sunspots numbers from 1749 to 1983, and the right panel is a plot of the CV curve. The online version of this figure is in color.

panel of Figure 4 gives the proportion of cases with $\hat{p} = p$ when the errors follow a first-order autoregressive process with first-lag correlation equal to 0.3 and 0.6. While the proportion of correct identifications is less than in the case of i.i.d. errors, the results are nonetheless encouraging.

6. APPLICATIONS

6.1 Sunspots and Lynx Data

Here, we apply our CV method to two classical time series, the sunspots data and the lynx data, both of which are available in R (R Development Core Team 2012). The sunspots series consists of mean monthly relative sunspot numbers; for example, see Andrews and Herzberg (1985). The data cover the period from 1749 to 1983 for a total of 2820 observations. Sunspots are temporary phenomena on the surface of the sun that appear visibly as dark spots compared with surrounding regions. The number of sunspots peaks periodically, as shown in the left panel of Figure 5. By minimizing the CV objective function (right panel of Figure 5), we get a period estimate of $\hat{p} = 133$ months. This is in close agreement with the "accepted" period estimate of 11 years for the sunspot series. See, for example, the frequency domain analysis of Brockwell and Davis (1991). With such a

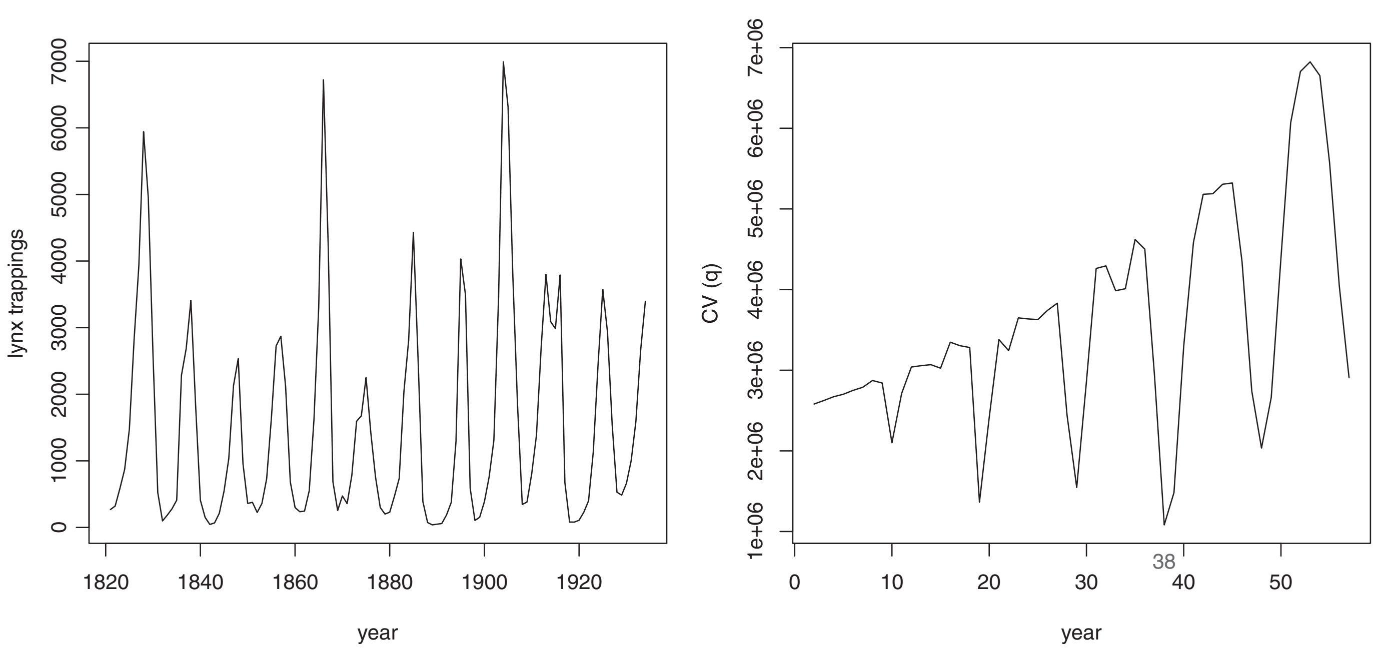

Figure 6. Lynx data: the left panel is a plot of lynx trappings from 1821 to 1934, and the right panel is a plot of the CV curve. The online version of this figure is in color.

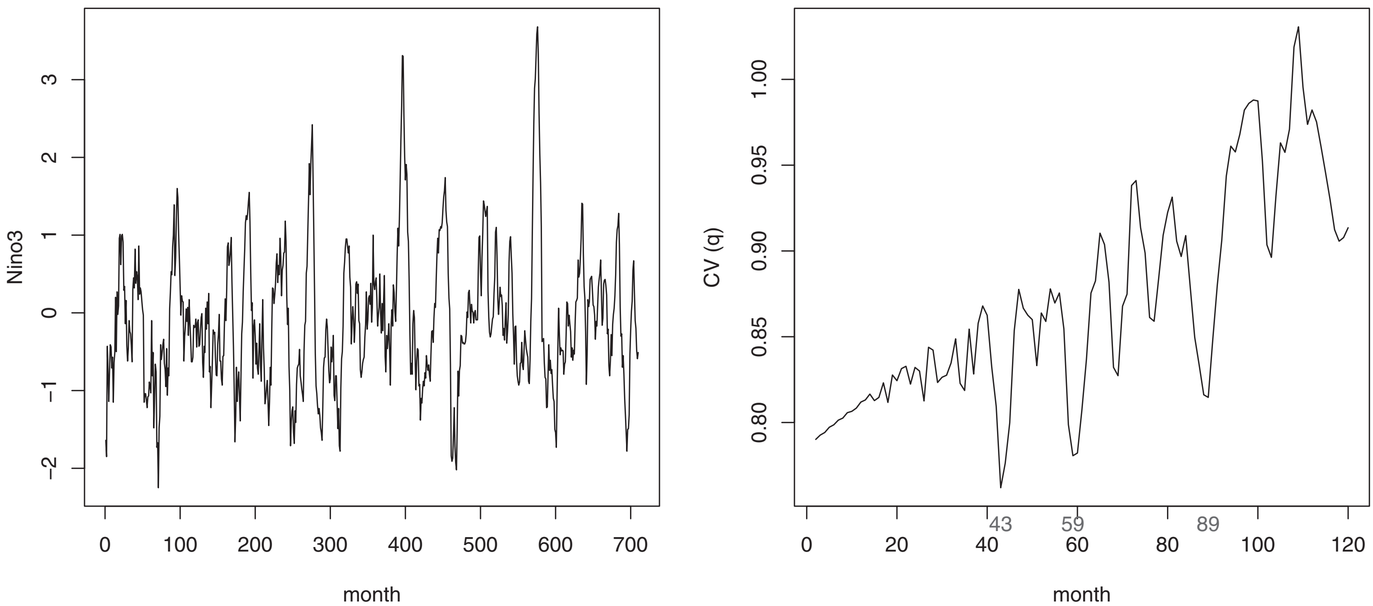

Figure 7. El Niño data: the left panel is a plot of monthly Niño3 sea surface temperatures from 1950–2009, and the right panel is a plot of the CV curve. The online version of this figure is in color.

large $\hat{p}$, the asymptotic 95% confidence set contains only $\hat{p}$, 133 months, as described in Section 3.

Another well-known time series consists of the annual number of lynx trappings in Canada from 1821 to 1934. These data have been analyzed by a number of researchers, including Campbell and Walker (1977). The 114 consecutive observations are plotted in the left panel of Figure 6 and the CV function is shown in the right. The minimizer of the CV function occurs at 38 years, a cycle length that contains four peaks in the time series plot.

The next smallest local minimum is at 19 years, a cycle length that is consistent with a period of 9.5 years. Interestingly, the periodogram for the lynx data is maximized at 9.5 years, which along with the sinusoidal appearance of the data, is highly suggestive of a 9.5-year period. This is a good example of how our CV method and the periodogram complement each other in cases where the period is not necessarily a multiple of the time interval between data points.

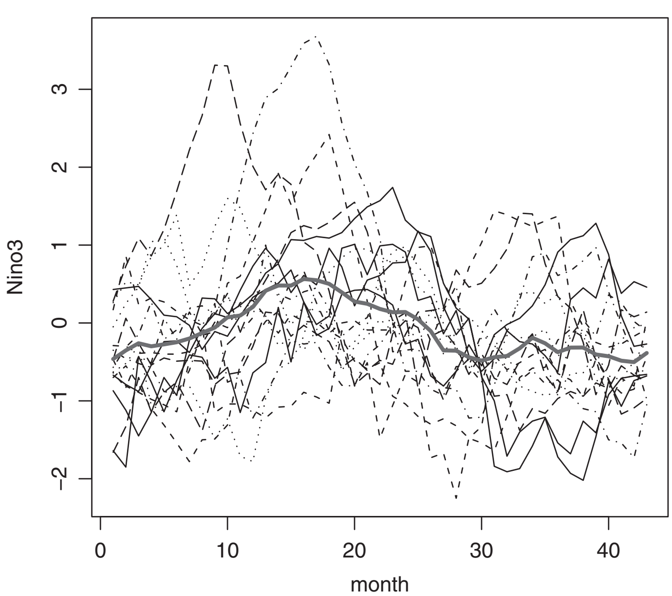

Figure 8. A plot of the 16 1/2 cycles of data (corresponding to a period of 43 months) superimposed on each other. The 43 estimated means are connected by a thicker solid line. The online version of this figure is in color.

How can one reconcile the disparate estimates of 9.5 and 38 years? In a sense, the "correct" estimate depends on how one defines periodicity. If the period is defined as the time between two peaks, then the 9.5-year estimate seems better. However, as can be seen in the time series plot, there is a pattern of one high peak followed by three lower ones that repeats itself for (almost) three cycles. Therefore, a period estimate of 38 years containing four peaks is perhaps more reasonable according to the strict definition of periodicity. Certainly, more observations would be needed to make the 38-year cycle more convincing. In our opinion, it is a strength of the CV method that in conjunction with the periodogram, it identifies both 9.5 and 38 as possible estimates, since those periods are consistent with what a data plot suggests are plausible possibilities.

We also applied our Section 4 test and the QT method and computed R^2 values at the estimated periods. Results are shown in Table 2. In each case, the p-value was obtained by using a parametric bootstrap in which each generated dataset consisted of i.i.d. standard normal variates. The QT method chooses a

Table 2. Tests of periodicity, R^2 values, and Quinn–Thompson period estimates

Data set	p*	R^2	m	Period estimate
Sunspots	0.021	0.243	20	1317.4
Lynx	0.017	0.807	3,4	9.6, 38.5
El Niño	0.260	0.144	10	177.8

NOTE: For each dataset, p* is the approximate p-value of the Section 4 test of periodicity, and m indicates the Fourier series truncation point used in the Quinn–Thompson method.

very large period: for the sunspots data, the estimated period was larger than 1100 for each of $m = 10, 20, 30$, and 40; and interestingly, for the lynx data, the QT period estimates for $m = 3$ and 4 coincided with the two period estimates identified as plausible by our CV method.

6.2 El Niño Effect

El Niño is a phenomenon associated with warmer than normal SST. It occasionally forms across much of the tropical eastern and central Pacific. The times between successive El Niño events are irregular, but the events tend to recur every 2–8 years, as shown by Torrence and Webster (1999). Niño3 SST is an oceanic component index that is one of the measures of variability in the ENSO-Monsoon system (ENSO refers to El Niño-Southern Oscillation). Estimating the period of the El Niño effect by analyzing Niño3 SST data is a very interesting application in climate science. The data we have consist of the area-average SST over the eastern equatorial Pacific, monthly from January 1950 to February 2009, as seen in the left panel of Figure 7. Our CV method yields a period estimate of $\hat{p} = 43$ months; see the right panel of Figure 7 for the CV curve. The 43 estimated means plot is shown in Figure 8. Note that the CV curve is also locally minimized at 59 and 89 months. These results are consistent with the wavelet analysis of Torrence and Webster (1999), who found that power in a wavelet spectrum was "broadly distributed, with peaks in the 2–8-year ENSO band." Their analysis also shows that this power varies over time, with 1960–1990 being a time at which the ENSO band had larger power. In Table 2, note that for the El Niño data, the test shows, at best, weak evidence for periodicity. However, the test is sensitive to periodicity of a *deterministic* nature, whereas periodicity in the El Niño data appears to be more transient, as indicated by the analysis of Torrence and Webster (1999).

6.3 Correlation in Errors

The errors for any one of the sunspots, lynx, or El Niño series are likely to be serially correlated. We thus investigated the effect of correlation by using a bootstrap method. For each series, residuals were computed after obtaining estimates $\hat{p}$ and $\hat{\boldsymbol{\mu}}$ of the period and mean sequence, respectively. Autoregressive moving average (ARMA) models were fitted to these residuals and AIC was used to choose the ARMA order. Then, 1000 independent datasets, each of the same size as the original series, were generated from a model in which the true period and mean sequence were equal to $\hat{p}$ and $\hat{\boldsymbol{\mu}}$, respectively, and the error series followed an ARMA process, with order and parameters estimated from the original series. The period was estimated by the CV method for each of the 1000 datasets. The results are summarized in Table 3.

In the sunspots simulation, in the 12 cases where the estimated period was not 133, it was $2 \times 133 = 266$. The lower probability for the El Niño series is due to the fact that its error standard deviation of 0.327 is relatively large. When this standard deviation was lowered to 0.10 and the simulation repeated, the period estimate was 43 in 933 cases and $2 \times 43 = 86$ in the other 67.

Table 3. Estimated probabilities (from 1000 replications) that $\hat{p} = p$ when the errors are correlated as in the real data examples

Series	Error model	Estimated probability
Sunspots	ARMA(6,4)	0.988
Lynx	ARMA(5,2)	0.999
El Niño	ARMA(5,3)	0.681

These results provide more evidence that the CV method is quite robust to correlation.

7. CONCLUDING REMARKS

We have proposed a CV period estimator for equally spaced data that are the sum of a periodic sequence and noise. The method is computationally simple and implicitly penalizes multiples of the smallest period. Given a particular period, or cycle length, a leave-out-one-cycle version of CV is used to compute an average squared prediction error. The cycle length minimizing this average squared error is the period estimator. Moreover, models corresponding to different periods may be ranked from best to worst by considering values of the objective function, thus extending the possibilities of interpretation. Our theory shows that the CV period estimator $\hat{p}$ is virtually consistent for large p in that its asymptotic probability of equaling p increases monotonically to 1 as p becomes large. When $p = 15$, this probability is approximately 0.99.

We have also proposed a nonparametric test of periodicity that is consistent whenever the observed sequence of means is periodic and, in many cases, where the sampled function is periodic with rational period.

It is also worth noting that the CV method has the advantage that it can easily deal with missing data, as long as the missing data are at random.

APPENDIX: PROOF OF THEOREMS 1 AND 2

We begin with a lemma that addresses the behavior of the deterministic component of the CV criterion.

Lemma 1. Define $\mathcal{S}_n = \{q = 1, 2, \ldots, M_n : q$ is not a multiple of $p\}$, and for $q = 1, 2, \ldots,$

$$C_q = \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} (\mu_{i+(j-1)q} - \bar{\mu}_{qi})^2,$$

where $\bar{\mu}_{qi} = \sum_{j=1}^{k_{q,i}} \mu_{i+(j-1)q} / k_{q,i}$, $i = 1, \ldots, q$, and we assume that $M_n \to \infty$ with $M_n = o(n)$. Then, there exists $\xi > 0$ such that

$$\min_{q \in \mathcal{S}_n} C_q > \xi$$

for all n sufficiently large.

Proof. The proof consists of two main steps:

Step 1: Show that $\min_{q \in \{2,\ldots,p-1\}} C_q$ is bounded away from 0 for all n sufficiently large.

Step 2: Argue that the bound in Step 1 can be applied to C_q for $q > p$ as well.

Step 1: Let q be one of $2, 3, \ldots, p-1$. Then, there is an $\ell \in \{1, 2, \ldots, q\}$ such that not all of $\mu_\ell, \mu_{\ell+q}, \mu_{\ell+2q}, \ldots$ are the same. If there was not such an ℓ, then $\{\mu_j\}$ would be periodic of period q, which contradicts the assumption that p is the smallest period. We have

$$\begin{aligned} C_q &\geq \frac{1}{2nk_{q,q}} \sum_{j=1}^{k_{q,q}} \sum_{k=1}^{k_{q,q}} (\mu_{\ell+(j-1)q} - \mu_{\ell+(k-1)q})^2 \\ &= \frac{1}{2nk_{q,q}} \sum_{r=1}^{p} \sum_{s=1}^{p} n_{qr} n_{qs} (\mu_r - \mu_s)^2, \end{aligned}$$

where n_{qr} is the number of times $\mu_{\ell+(j-1)q}$ equals μ_r for j between 1 and $k_{q,q}$. Since not all of $\mu_{\ell+(j-1)q}$ are the same, there exist r_1 and r_2 such that $\mu_{r_1} \neq \mu_{r_2}$ and $n_{qr_1} > 0$ and $n_{qr_2} > 0$.

Obviously,

$$C_q \geq \frac{1}{2nk_{q,q}} n_{qr_1} n_{qr_2} (\mu_{r_1} - \mu_{r_2})^2.$$

The proof of Step 1 is done if we can show that each of n_{r_1} and n_{r_2} is bounded below by Cn, where $C > 0$. Let r be any integer between 1 and p such that $\mu_{\ell+(j-1)q}$ equals μ_r for some (smallest) j. It follows that there is a nonnegative integer k such that $\ell + (j-1)q - kp = r$. Therefore, for $m = 1, 2, \ldots$, we have

$$\ell + [(j-1) + mp]q - (k + mq)p = r,$$

which implies that $\mu_{\ell+sq} = \mu_r$ at $s = (j-1), (j-1) + p, (j-1) + 2p, \ldots$, and hence, that $n_{qr} \geq (k_{q,q} - j + 1)/p$. The result to be proven in Step 1 follows immediately.

Step 2: Suppose that $\{M_n\}$ is a sequence of integers such that $M_n \to \infty$ with $M_n = o(n)$. Now, let $q = Mp + j$, where $1 \leq M \leq M_n$ and $1 \leq j \leq p-1$. We have

$$\begin{aligned} C_q &= \frac{1}{n} \sum_{i=1}^{Mp+j} \sum_{k=1}^{k_{q,i}} (\mu_{i+(k-1)q} - \bar{\mu}_{qi})^2 \\ &= \frac{1}{2n} \sum_{i=1}^{Mp+j} \frac{1}{k_{q,i}} \sum_{k=1}^{k_{q,i}} \sum_{\ell=1}^{k_{q,i}} (\mu_{i+(k-1)q} - \mu_{i+(\ell-1)q})^2 \\ &\geq \frac{1}{2nk_{q,q}} \sum_{i=1}^{Mp+j} \sum_{k=1}^{k_{q,q}} \sum_{\ell=1}^{k_{q,q}} (\mu_{i+(k-1)j} - \mu_{i+(\ell-1)j})^2 \\ &= \sum_{m=1}^{M} C_{qm} + \frac{1}{2nk_{q,q}} \sum_{i=Mp+1}^{Mp+j} \sum_{k=1}^{k_{q,q}} \sum_{\ell=1}^{k_{q,q}} (\mu_{i+(k-1)j} - \mu_{i+(\ell-1)j})^2, \end{aligned}$$

where

$$\begin{aligned} C_{qm} &= \frac{1}{2nk_{q,q}} \sum_{i=(m-1)p+1}^{mp} \sum_{k=1}^{k_{q,q}} \sum_{\ell=1}^{k_{q,q}} (\mu_{i+(k-1)j} - \mu_{i+(\ell-1)j})^2 \\ &= \frac{1}{2nk_{q,q}} \sum_{r=1}^{p} \sum_{k=1}^{k_{q,q}} \sum_{\ell=1}^{k_{q,q}} (\mu_{r+(k-1)j} - \mu_{r+(\ell-1)j})^2. \end{aligned}$$

It follows that $C_q \geq MC_{q1}$, and hence, that

$$C_q \geq \frac{M}{2nk_{q,q}} \sum_{r=1}^{j} \sum_{k=1}^{k_{q,q}} \sum_{\ell=1}^{k_{q,q}} (\mu_{r+(k-1)j} - \mu_{r+(\ell-1)j})^2.$$

For future reference, we note that

$$k_{q,q} = \left[\frac{n}{q}\right] = \left[\frac{n}{Mp+j}\right] \geq \frac{n}{p(M_n+1)} - 1. \tag{A.1}$$

Using the same type of notation and arguing exactly as in the proof of Step 1,

$$\begin{aligned} C_q &\geq \frac{M}{2nk_{q,q}} n_{q1} n_{q2} (\mu_{r_1(j)} - \mu_{r_2(j)})^2 \\ &\geq \frac{M}{2nk_{q,q}} n_{q1} n_{q2} \delta, \end{aligned}$$

where δ is the smallest nonzero value of $(\mu_i - \mu_j)^2$ for i, j in $\{1, \ldots, p\}$. Now, for $i = 1, 2$, n_{qi} is the number of times that $\mu_{r+(k-1)j}$ equals $\mu_{r_i(j)}$ as k ranges between 1 and $k_{q,q}$. As in the proof of Step 1, we know that

$$n_{qi} \geq \left(\frac{k_{q,q} - k(i,j) + 1}{p}\right),$$

where $k(i,j)$ is the smallest k for which $\mu_{r+(k-1)j} = \mu_{r_i(j)}$. Importantly, $k(i,j)$ depends on j but not M. We thus have

$$\begin{aligned} C_q &\geq \frac{\delta M}{2np^2 k_{q,q}} (k_{q,q} - k^* + 1)^2 \\ &\geq \frac{\delta M}{2(Mp+j)p^2} \left(1 - \frac{Mp+j}{n}\right) \left(1 - \frac{k^*-1}{k_{q,q}}\right)^2, \end{aligned}$$

where k^* is the largest of the integers $k(i,j)$, $i = 1, 2$, $j = 1, \ldots, p-1$.

Now,

$$\frac{M}{Mp+j} \geq \frac{M}{p(M+1)} \geq \frac{1}{2p},$$

and hence,

$$C_q \geq \frac{\delta}{4p^3} \left(1 - \frac{Mp+j}{n}\right) \left(1 - \frac{k^*-1}{k_{q,q}}\right)^2.$$

Recalling (A.1) and the fact that $M_n = o(n)$, it follows that $(Mp+j)/n$ and $(k^*-1)/k_{q,q}$ are smaller than 1/2 for all n sufficiently large, and so,

$$C_q \geq \frac{\delta}{32p^3}$$

for all n sufficiently large. ■

Proof of Theorem 1. To argue that $\lim_{n\to\infty} P(\hat{p} \in \mathcal{S}_n) = 0$, we must show that

$$\lim_{n\to\infty} P\left[\bigcap_{1\leq\ell\leq M_n/p} \bigcap_{q\in\mathcal{S}_n} \{\mathrm{CV}(q) - \mathrm{CV}(\ell p) > 0\}\right] = 1.$$

The periodicity of $\boldsymbol{\mu}$ entails that

$$\begin{aligned} \mathrm{CV}(q) - \mathrm{CV}(\ell p) &= \frac{1}{n}(S_{q,n} - S_{\ell p,n}) \\ &\quad + \frac{2}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} \left(\varepsilon_{qij} - \bar{\varepsilon}_{qi}^{j}\right)\left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right) \\ &\quad + \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{k_{q,i}} \left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right)^2, \end{aligned}$$

where, for each q, $S_{q,n} = \sum_{i=1}^{q}\sum_{j=1}^{k_{q,i}}(\varepsilon_{qij} - \bar{\varepsilon}_{qi}^{j})^2$. Using the result of Lemma 1, it is easily checked that for all n sufficiently large,

$$\min_{q\in\mathcal{S}_n} \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{k_{q,i}}\left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right)^2 \geq \frac{\xi}{2}.$$

It therefore follows that

$$\begin{aligned} P&\left[\bigcap_{1\leq\ell\leq M_n/p}\bigcap_{q\in\mathcal{S}_n}\{\mathrm{CV}(q) - \mathrm{CV}(\ell p) > 0\}\right] \\ &\geq 1 - \sum_{\ell}\sum_{q} P\left(\frac{1}{n}(S_{q,n} - S_{\ell p,n}) \leq -\frac{\xi}{6}\right) \\ &\quad - \sum_{\ell}\sum_{q} P\left(A_{q,n} \leq -\frac{\xi}{6}\right) \\ &\quad - \sum_{\ell}\sum_{q} P\left(B_{q,n} \leq -\frac{\xi}{6}\right), \end{aligned} \tag{A.2}$$

where

$$A_{q,n} = \frac{2}{n}\sum_{i=1}^{q}\sum_{j=1}^{k_{q,i}}\varepsilon_{qij}\left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right) \quad \text{and}$$

$$B_{q,n} = -\frac{2}{n}\sum_{i=1}^{q}\sum_{j=1}^{k_{q,i}}\bar{\varepsilon}_{qi}^{j}\left(\mu_{qij} - \bar{\mu}_{qi}^{j}\right).$$

For t, any positive number, Bernstein's form of the Markov inequality implies that

$$P\left(A_{q,n} \leq -\frac{\xi}{6}\right) \leq \exp\left[-nt\left(\frac{\xi}{6} - 8t\sigma^2 B\right)\right],$$

where $B = \max_{1\leq i\leq p}\mu_i^2$. Taking t to be smaller than $\xi/(48B\sigma^2)$, it is now clear that there exists a positive number C such that $P(A_{q,n} \leq -\xi/6) \leq e^{-Cn}$, and since M_n is smaller than n, $\sum_\ell\sum_q P(A_{q,n} \leq -\xi/6) \to 0$ as $n\to\infty$. The other two terms on the right-hand side of (A.2) can be dealt with in the same way, and therefore, $\lim_{n\to\infty} P(\hat{p}\in\mathcal{S}_n) = 0$.

Now, we consider $P(\hat{p} = p)$, which by the result just proven, is asymptotically equal to

$$P\left(\bigcap_{2\leq m\leq M_n/p}\{\mathrm{CV}(mp) - \mathrm{CV}(p) > 0\}\right). \tag{A.3}$$

(The proof for other $P(\hat{p} = jp)$ is similar and hence omitted.) We have

$$\begin{aligned} \mathrm{CV}(mp) - \mathrm{CV}(p) &= \frac{1}{n}\sum_{i=1}^{mp} C_{mp,i}\left\{\sum_{j=1}^{k_{mp,i}}\varepsilon_{mp,ij}^2 - k_{mp,i}\bar{\varepsilon}_{mp,i}^2\right\} \\ &\quad - \frac{1}{n}\sum_{i=1}^{p} C_{p,i}\left\{\sum_{j=1}^{k_{pi}}\varepsilon_{pij}^2 - k_{pi}\bar{\varepsilon}_{pi}^2\right\}, \end{aligned}$$

where $C_{q,i} = k_{q,i}^2/(k_{q,i}-1)^2$. Writing $\hat{\sigma}^2 = \sum_{i=1}^{n}\varepsilon_i^2/n$, we have

$$\begin{aligned} n(\mathrm{CV}(mp) - \mathrm{CV}(p)) &= \sigma^2\sum_{i=1}^{p} C_{p,i}U_{1,i}^2 - \sigma^2\sum_{i=1}^{mp} C_{mp,i}U_{m,i}^2 \\ &\quad + n\sigma^2(C_{mp,1} - C_{p,1}) + R_{m,n}, \end{aligned}$$

where $U_{m,i} = \sqrt{k_{mp,i}}\bar{\varepsilon}_{mp,i}/\sigma$, $i = 1,\ldots,mp$, $m = 1,\ldots,M_n$, and

$$\begin{aligned} R_{m,n} &= n(\hat{\sigma}^2 - \sigma^2)(C_{mp,1} - C_{p,1}) + \sum_{i=1}^{mp}(C_{mp,i} - C_{mp,1}) \\ &\quad \times \sum_{j=1}^{k_{mp,i}}\varepsilon_{mp,ij}^2 + \sum_{i=1}^{mp}(C_{p,1} - C_{p,i})\sum_{j=1}^{k_{p,i}}\varepsilon_{p,ij}^2. \end{aligned}$$

We remark that for each m, $U_{m,1}^2,\ldots,U_{m,mp}^2$ are i.i.d. χ_1^2 random variables. Subsequently, we use the following facts:

$$\begin{aligned} &n(C_{mp,1} - C_{p,1}) = 2p(m-1) + O\left(\frac{M_n^2}{n}\right), \\ &\qquad C_{mp,1} = 1 + O\left(\frac{M_n}{n}\right), \quad C_{p,1} = 1 + O\left(\frac{1}{n}\right), \\ &|C_{mp,i} - C_{mp,1}| \leq O\left(\frac{M_n}{n}\right)^2, \quad \text{and} \\ &\quad |C_{p,i} - C_{p,1}| \leq O\left(\frac{1}{n}\right)^2, \end{aligned}$$

with the last two inequalities holding uniformly in i, since for each q, $k_{q,1},\ldots,k_{q,q}$ take on at most two distinct values that differ by only 1.

So now, we have

$$\begin{aligned} \frac{n(\mathrm{CV}(mp) - \mathrm{CV}(p))}{\sigma^2} &= \sum_{i=1}^{p} U_{1,i}^2 - \sum_{i=1}^{mp} U_{m,i}^2 \\ &\quad + 2p(m-1) + \tilde{R}_{m,n}, \end{aligned}$$

where, using previously stated facts, we have

$$\tilde{R}_{m,n} = \frac{R_{m,n}}{\sigma^2} + O_p\left(\frac{M_n^2}{n}\right).$$

Note that the term $|R_{m,n}|$ is bounded almost surely by

$$\begin{aligned} &n(C_{mp,1} - C_{p,1})|\hat{\sigma}^2 - \sigma^2| + n\hat{\sigma}^2[\max_i |C_{mp,i} - C_{mp,1}| \\ &\quad + \max_i |C_{p,1} - C_{p,i}|], \end{aligned}$$

which is $O_p(M_n/\sqrt{n}) + O_p(M_n^2/n)$.

Now, let $\{\delta_n\}$ be an arbitrary sequence of positive numbers that tend to 0. The sequence of probabilities (A.3) may be bounded above and below by probability sequences that have the same limits as

$$P\left(\bigcap_{2\leq m\leq M_n/p}\left\{\sum_{i=1}^{p} U_{1,i}^2 - \sum_{i=1}^{mp} U_{m,i}^2 + 2p(m-1) > -\delta_n\right\}\right)$$

and

$$P\left(\bigcap_{2\leq m\leq M_n/p}\left\{\sum_{i=1}^{p} U_{1,i}^2 - \sum_{i=1}^{mp} U_{m,i}^2 + 2p(m-1) > \delta_n\right\}\right),$$

respectively. The last statement is proven by applying Bernstein's form of the Markov inequality to the sequences $P(\bigcap_{2\leq m\leq M_n/p}\tilde{R}_{m,n} > \delta_n)$ and $P(\bigcap_{2\leq m\leq M_n/p}\tilde{R}_{m,n} > -\delta_n)$, and by using the assumption that $M_n = o(\sqrt{n})$ and the fact that

δ_n can be defined to converge arbitrarily slowly to 0. It is now clear that

$$\lim_{n\to\infty} P(\hat{p}=p) = \lim_{n\to\infty} P\left(\bigcap_{2\le m<M_n/p}\left\{\sum_{i=1}^{p} U_{1,i}^2 - \sum_{i=1}^{mp} U_{m,i}^2 + 2p(m-1) > 0\right\}\right).$$

The result will be proven if we can verify that the $U_{m,i}$s have the same limiting correlation structure as that of the $Z_{m,i}$s. By construction, $U_{m,1},\ldots,U_{m,mp}$ are mutually independent. Now, let $r > m$ and consider

$$\begin{aligned}\mathrm{corr}(U_{m,i},U_{r,j}) &= \frac{1}{\sqrt{k_{mp,i}k_{rp,j}}\sigma^2} \\ &\quad \times \sum_{\ell=1}^{k_{mp,i}}\sum_{s=1}^{k_{rp,j}} E\left[\varepsilon_{i+mp(\ell-1)}\varepsilon_{j+rp(s-1)}\right] \\ &= \frac{1}{\sqrt{k_{mp,i}k_{rp,j}}} N_{m,r,i,j},\end{aligned}$$

where $N_{m,r,i,j}$ is the number of times that $j+rp(s-1) = i+mp(\ell-1)$. Now, if there is a pair (ℓ,s) that satisfies this equation, then $(\ell+\eta r, s+\eta m)$, $\eta=1,2,\ldots$, are also solutions and $N_{m,r}\sim k_{rp,j}/m$. Therefore, when $N_{m,r}>0$,

$$\begin{aligned}\mathrm{corr}(U_{m,i},U_{r,j}) &\sim \frac{k_{rp,j}}{m\sqrt{k_{mp,i}k_{rp,j}}} \\ &= \frac{1}{m}\sqrt{\frac{k_{rp,j}}{k_{mp,i}}} \\ &\sim \frac{1}{m}\sqrt{\frac{n/(rp)}{n/(mp)}} \\ &= \frac{1}{\sqrt{mr}}.\end{aligned}$$

■

Proof of Theorem 2. The proof of (i) is straightforward in light of the proof of Theorem 1. The statistic T_n is larger than B if and only if the minimizer, $\hat{p}_B$, of selection criterion $C(q;B)$ is larger than 1. Arguing as in the proof of Theorem 1, $P(\hat{p}_B \ge 2)$ tends to 1 as $n\to\infty$ so long as $\boldsymbol{\mu}$ has period greater than 1.

To prove (ii), we just apply the proof of (i) after showing that $\boldsymbol{\mu}$ is periodic with smallest period at least 2. Since the period of the function μ is j/m, $\boldsymbol{\mu}$ is periodic with period j, which is at least 2 since $j/m>1$. Now, we just need to verify that $\boldsymbol{\mu}$ does not have period 1, which is equivalent to showing that $\mu_1,\ldots,\mu_j$ are all not the same. For each $r=1,\ldots,j$, let ℓ_r be the integer such that $0\le r-\ell_r j/m < j/m$. Since $\mu_r=\mu(r)=\mu(r-\ell_r j/m)$, and μ does not take on the same value more than $j-1$ times in $[0, j/m)$, we are done upon showing that all of $r-\ell_r j/m$, $r=1,\ldots,j$, are distinct.

Let $1\le r<s\le j$ and suppose that $r-\ell_r j/m = s-\ell_s j/m$. Then, $(s-r)/(\ell_s-\ell_r)=j/m$, which is a contradiction since $s-r\le j-1$ and j/m is expressed in simplest form. ■

ACKNOWLEDGMENTS

Professor Hart's research was supported in part by NSF grant DMS-0604801. Professor Genton's research was partially supported by NSF (National Science Foundation) grant DMS-1007504, and award KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

[Received January 2011. Revised November 2011.]

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, eds. B. N. Petrov and F. Csake, Budapest: Akademiai Kiado, pp. 267–281. [83]

Andrews, D. F., and Herzberg, A. M. (1985), *Data*, New York: Springer. [91]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer-Verlag. [91]

Buys-Ballot, C.H. D. (1847), *Les Changements Périodiques de Température*, Utrecht, Kemink et Fils. [83]

Campbell, M. J., and Walker, A. M. (1977), "A Survey of Statistical Work on the Mackenzie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis," *Journal of the Royal Statistical Society:* Series A, 140, 411–431. [92]

Eubank, R. L., and Hart, J. D. (1992), "Testing Goodness-of-Fit in Regression via Order Selection Criteria," *The Annals of Statistics*, 20, 1412–1425. [88]

Fellner, W. H. (1974), "Heuristic Estimation of Probability Densities," *Biometrika*, 61, 485–492. [88]

Genton, M. G., and Hall, P. (2007), "Statistical Inference for Evolving Periodic Functions," *Journal of the Royal Statistical Society:* Series B, 69, 643–657. [83]

Hall, P. (2008), "Nonparametric Methods for Estimating Periodic Functions, With Applications in Astronomy," in *COMPSTAT 2008: Proceedings in Computational Statistics*, ed. P. Brito, Heidelberg: Physika-Verlag, pp. 3–18. [83,85]

Hall, P., and Li, M. (2006), "Using the Periodogram to Estimate Period in Nonparametric Regression," *Biometrika*, 93, 411–424. [83,86]

Hall, P., Reimann, J., and Rice, J. (2000), "Nonparametric Estimation of a Periodic Function," *Biometrika*, 87, 545–557. [83]

Hall, P., and Yin, J. (2003), "Nonparametric Methods for Deconvolving Multiperiodic Functions," *Journal of the Royal Statistical Society:* Series B, 65, 869–886. [83]

Hannan, E. J., and Quinn, B. G. (1979), "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society:* Series B, 41, 190–195. [85,87]

Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer-Verlag. [88]

Hart, J. D., and Wehrly, T. E. (1993), "Consistency of Cross-Validation When the Data Are Curves," *Stochastic Processes and Their Applications*, 45, 351–361. [84]

Quinn, B. G., and Hannan, E. J. (2001), *The Estimation and Tracking of Frequency*, New York: Cambridge University Press. [83]

Quinn, B. G., and Thomson, P. J. (1991), "Estimating the Frequency of a Periodic Function," *Biometrika*, 78, 65–74. [83,86]

R Development Core Team. (2012), *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing. [91]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [85,87]

Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika*, 63, 117–126. [87]

Stein, C. (1964), "Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution With Unknown Mean," *Annals of the Institute of Statistical Mathematics*, 16, 155–160. [85]

Torrence, C., and Webster, P. (1999), "Interdecadal Changes in the ENOS-Monsoon System," *Journal of Climate*, 12, 2679–2690. [83,93]

Whittaker, E., and Robinson, G. (1967), *The Calculus of Observations*, New York: Dover Publications. [83,88]