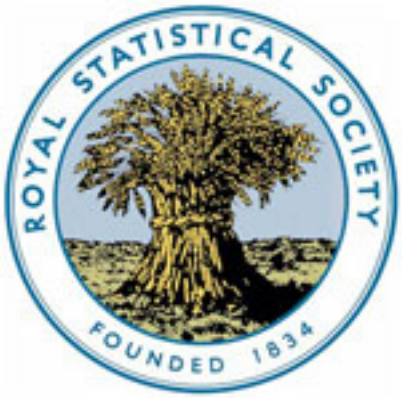

Appl. Statist. (2007)
56, *Part* 5, *pp.* 587–606

An analysis of pulsation periods of long period variable stars

Jeffrey D. Hart,

Texas A&M University, College Station, USA

Chris Koen

University of the Western Cape, Cape Town, South Africa

and Fred Lombard

University of Johannesburg, South Africa

[Received September 2006. Final revision June 2007]

Summary. We report the results of a period change analysis of time series observations for 378 pulsating variable stars. The null hypothesis of no trend in expected periods is tested for each of the stars. The tests are non-parametric in that potential trends are estimated by local linear smoothers. Our testing methodology has some novel features. First, the null distribution of a test statistic is defined to be the distribution that results in repeated sampling from a population of stars. This distribution is estimated by means of a bootstrap algorithm that resamples from the collection of 378 stars. Bootstrapping in this way obviates the problem that the *conditional* sampling distribution of a statistic, given a particular star, may depend on unknown parameters of that star. Another novel feature of our test statistics is that one-sided cross-validation is used to choose the smoothing parameters of the local linear estimators on which they are based. It is shown that doing so results in tests that are tremendously more powerful than analogous tests that are based on the usual version of cross-validation. The positive false discovery rate method of Storey is used to account for the fact that we simultaneously test 378 hypotheses. We ultimately find that 56 of the 378 stars have changes in mean pulsation period that are significant when controlling the positive false discovery rate at the 5% level.

Keywords: False discovery rate; Multiple-hypotheses testing; Profile likelihood; Smoothing methods; Trend detection; Variable stars

1. Introduction

Variable stars, of which tens of thousands are known, are characterized by changes in brightness over time. Various physical mechanisms give rise to the variability. In this paper we are concerned with a group of 378 pulsating stars classified as 'long period variable stars.' These stars are distinguished by their substantial changes in brightness, which are roughly sinusoidal with typical periods between 100 and 300 days. The period of a given star is determined by its internal structure. Changes in period are therefore deemed important by astronomers, as these reflect changing physical conditions in the stars. For general background material on variable stars the interested reader is referred to Hoffmeister *et al.* (1985).

Address for correspondence: Jeffrey D. Hart, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
E-mail: hart@stat.tamu.edu

 0035–9254/07/56587

Astronomers' interest in possible variations in the periods of pulsating stars has been long and sustained. The earliest reference of which we are aware is Birt (1831). Entering the phrase 'period change' in the Nasa 'Astrophysics data system' abstract search currently gives 1500 hits; 'period variability' gives 2571 hits, with more than 100 in each case from the last 3 years. The reason for the interest lies primarily in the fact, which was mentioned above, that changes in period contain information about changing physical conditions in stars. This is a specialized topic which is not free of controversy—see, for example, the short but instructive review by Handler (2004). Attention here is restricted to some of the statistical issues. An account of the physics of rapid period changes in long period variable stars (also known as Mira stars) is given in Whitelock (1999).

The data that were available to us for the 378 stars are their times of maximum and minimum brightness, accumulated over approximately 75 years; see Campbell (1955) and Mattei *et al.* (1990) for details on how the data were collected. We use the series of times between successive brightness maxima as a proxy for the time–local pulsation period. (Times between successive minima could, in principle, also have been used, but these tend to be more poorly observed.) The sizes of the data sets range between 32 and 200 observations. Many researchers have published analyses of such data for individual stars—see Whitelock (1999) for a review. Studies of aggregates of such stars were reported by Percy *et al.* (1990) and Percy and Colivas (1999). The methods that were used by Percy and Colivas (1999) can generally detect only monotonic trends. Koen and Lombard (2004) reported the results of a frequency domain analysis of stars from the American Association of Variable Star Observers database that overcomes this obstacle. Our approach is also designed to detect virtually any sort of trend but is quite different from that of Koen and Lombard (2004). First of all, we operate in the time domain, and for each star we smooth the series of times between successive maxima by using a local linear estimate. A test of no trend is based on this smooth, using a generalized profile likelihood ratio as test statistic. Our analysis also differs from that of Koen and Lombard (2004) in that we explicitly model heteroscedasticity in the measured times of maximum brightness. Furthermore, we take into account the fact that many hypotheses are tested simultaneously by using the method of Storey (2002) for controlling the false discovery rate. An intriguing pattern that we ultimately discover is that the strength of trend in times between maxima is an increasing function of the mean period length.

The paper will proceed as follows. In the next section we describe the basic model that is used for the series of times between maxima, discuss the history of the model and state some assumptions. Section 3 describes how our model is fitted to each star and how we test the no-trend hypothesis. In Section 4, results of applying the methodology of Section 3 to our database of 378 stars are discussed. Here we also apply Storey's method to account for testing multiple hypotheses. Finally, concluding remarks are given in Section 5, and supplementary material is provided in Appendix A.

2. Model used for each star

The term *epoch* will refer to one complete cycle of a star's periodic variation, and $Y_1, \ldots, Y_n$ will denote the observed lengths of time between successive maximum brightnesses at the chronologically ordered epochs $1, \ldots, n$. The expected value of Y_j will be modelled as a function of the standardized epoch $x_j = (j - \frac{1}{2})/n$, $j = 1, \ldots, n$. A model for the data of a single star is

$$Y_j = \mu(x_j) + I_j + \varepsilon_j - \varepsilon_{j-1}, \qquad j = 1, \ldots, n, \tag{1}$$

where μ is a function (defined on $[0, 1]$) that accounts for trend, $I_1, \ldots, I_n$ are random deviations that are intrinsic to the star and $\varepsilon_0, \ldots, \varepsilon_n$ are the errors that are made in determining times of

maximum brightness. By 'trend', we mean not just a systematic increase or decrease in periods, but *any* sort of systematic departure from constancy. If no trend is present, the mean function μ is identical to a constant. The two processes $\{I_j\}$ and $\{\varepsilon_j\}$ are assumed to be independent of each other.

Eddington and Plakidis (1929) appear to have been the first to have proposed model (1) as an explanation for random cycle-to-cycle variations in the periods of Mira stars. It was further explored by Sterne (1934). Both Eddington and Plakidis (1929) and Sterne (1934) assumed homoscedasticity and estimated σ_I and σ_ε for a few example cases with $\mu \equiv$ constant. They showed that the presence of the intrinsic noise terms I_j could give rise to patterns in cumulative sum diagrams that mimic systematic period changes. More recently, in a series of papers Koen and Lombard have explored the consequences of assuming model (1); see Koen and Lombard (2004) and references therein.

Plots of pulsation periods for the long period variable stars *Omicron Ceti*, *R Aquilae*, *R Bootis* and *R Camelopardalis* are shown in Fig. 1. An important aspect of each data set is the negative autocorrelation between observations that are one lag apart, as evidenced by the high frequency behaviour in each plot. This is a result of the last two terms in equation (1).

Each star is characterized by μ, its mean function, and η, parameters of its error process $\{Y_j - \mu(x_j) : j = 1, \ldots, n\}$. Each μ is allowed to be arbitrary and is estimated non-parametrically, as discussed in Section 3. In Section 3.2, two assumptions are made about the distribution of (μ, η), but these still allow for certain dependences between the parameters that are dictated

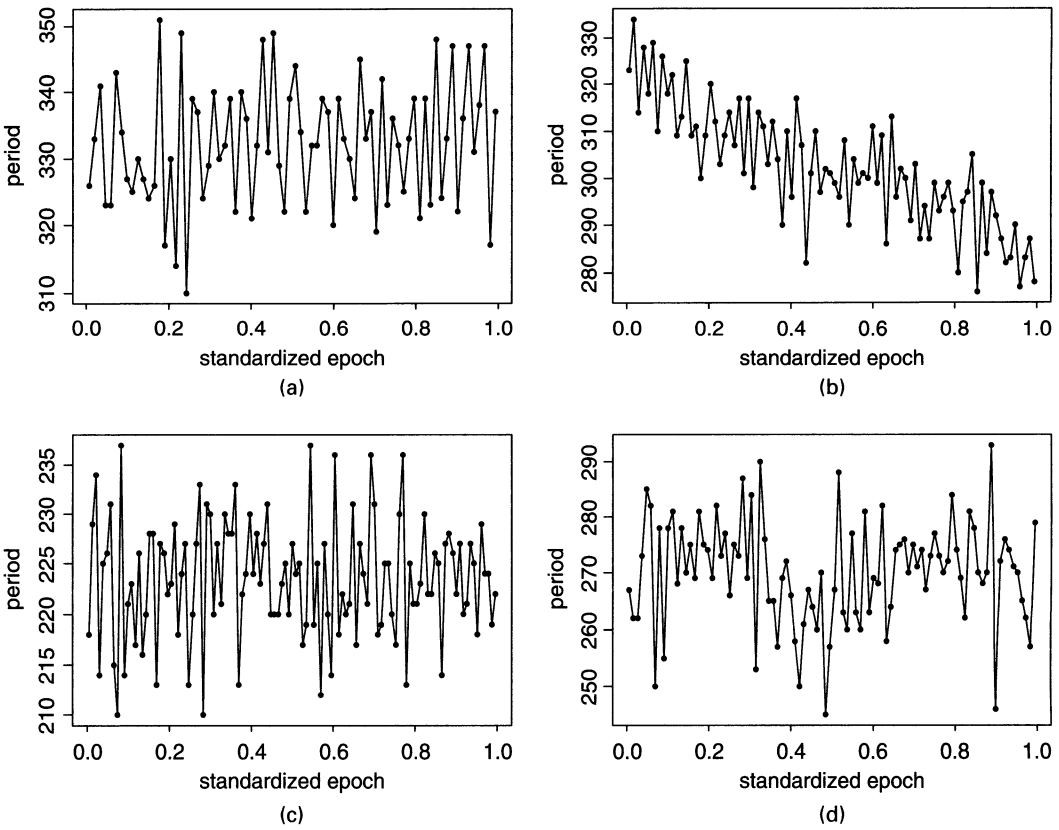

Fig. 1. Typical plots of pulsation periods: (a) *Omicron Ceti*; (b) *R Aquilae*; (c) *R Bootis*; (d) *R Camelopardalis*

by the nature of the data (see Appendix A). Concerning the conditional distribution of errors, given any particular star, we make the following assumptions.

Assumption 1. $I_1, \ldots, I_n$ and $\varepsilon_0, \ldots, \varepsilon_n$ are mutually independent.

Assumption 2. $I_j \sim N(0, \sigma_I^2)$, $j = 1, \ldots, n$, where $\sigma_I^2 < \infty$.

Assumption 3. $\varepsilon_j \sim N\{0, \exp(\beta_0 + \beta_1 x_j)\}$, $j = 0, \ldots, n$, where $|\beta_i| < \infty$, $i = 1, 2$.

Under these assumptions Pokta and Hart (2007) show that model (1) is identifiable.

Some fairly extensive preliminary analysis of the data was done to determine the reasonableness of assumptions 1–3. Independence of $I_1, \ldots, I_n$ was investigated by entertaining a first-order autoregressive (AR1) model for the intrinsic errors of each star. The resulting distribution of AR1 parameter estimates was consistent with what would be expected if each intrinsic series was white noise. A detailed analysis of residuals gave little reason to doubt the normality assumption but did indicate a prevailing pattern of heteroscedasticity in which the variance decreased over time. Since there was little evidence that the variance pattern was complicated, we adhere to Occam's razor and use the variance model in assumption 3, which allows for both decreases and increases over time. Attributing the heteroscedasticity to the experimental errors seems reasonable since advances over the last half-century have meant that later data are more accurate, i.e. that $\mathrm{var}(\varepsilon_j)$ has decreased. If one *did* take a non-parametric approach to estimating $\mathrm{var}(\varepsilon_j)$, a good way to do so in model (1) would be to use the local likelihood approach of Yu and Jones (2004).

3. Method of testing

Our goal is to test the null hypothesis

$$H_0 : \mu(x_1) = \mu(x_2) = \ldots = \mu(x_n)$$

against the negation of H_0 for each of the 378 stars in our database. The semiparametric model that was introduced in Section 2 is assumed henceforth. The non-parametric part of the model is the function μ, and the parametric part is the error series, which has parameters $\boldsymbol{\eta} = (\sigma_I^2, \beta_0, \beta_1)$. An efficient way of estimating the parametric part of a semiparametric model is to use *generalized profile likelihood*, as defined in Severini and Wong (1992). The likelihood for a given star may be expressed as $L_n(\boldsymbol{\eta}, \mu)$, where $\boldsymbol{\eta}$ and μ are candidates for the true parameters $\boldsymbol{\eta}_0$ and the true function μ_0 respectively. In our model, the generalized profile likelihood of Severini and Wong (1992) is $L_n(\boldsymbol{\eta}) = L_n(\boldsymbol{\eta}, \hat{\mu})$, where $\hat{\mu}$ is any consistent non-parametric estimator of μ_0. Severini and Wong (1992) argued that the maximizer of $L_n(\boldsymbol{\eta})$ is an asymptotically efficient estimator of $\boldsymbol{\eta}_0$.

If we use generalized profile likelihood to estimate parameters, it seems only natural to use a *generalized profile likelihood ratio* to test the hypotheses of interest. Under the null hypothesis of no trend, the (constant) function μ_0 will be estimated by $\bar{Y}$, the sample mean of $Y_1, \ldots, Y_n$, and our test statistic will be

$$S = \sup_{\boldsymbol{\eta}}[\log\{L_n(\boldsymbol{\eta}, \hat{\mu})\}] - \sup_{\boldsymbol{\eta}}[\log\{L_n(\boldsymbol{\eta}, \bar{Y})\}]. \tag{2}$$

Our choice for the non-parametric smooth $\hat{\mu}$ is a local linear estimate (Cleveland and Devlin, 1988; Fan, 1992) applied to the regression data (x_j, Y_j), $j = 1, \ldots, n$. The likelihood function has the form

$$L_n(\boldsymbol{\eta}, \hat{\mu}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}_\eta|^{-1/2} \exp\{-\tfrac{1}{2}(\mathbf{Y} - \hat{\boldsymbol{\mu}})^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} (\mathbf{Y} - \hat{\boldsymbol{\mu}})\},$$

where $\mathbf{Y}^{\mathrm{T}} = (Y_1, \ldots, Y_n)$, $\hat{\boldsymbol{\mu}}^{\mathrm{T}} = (\hat{\mu}(x_1), \ldots, \hat{\mu}(x_n))$ and $\boldsymbol{\Sigma}_\eta$ is the covariance matrix of $\mathbf{Y}$ assuming that $\boldsymbol{\eta}$ is the true parameter vector. This likelihood is minimized with respect to $\boldsymbol{\eta}$ by using a PORT routine, as supplied by the R function `nlminb` (R Development Core Team, 2007).

Having specified the nature of the test statistic, we are faced with two important problems:

(a) how will the smoothing parameter of the local linear estimate be chosen and
(b) how will the distribution of the test statistic be determined?

3.1. *Choice of smoothing parameter*

The choice of smoothing parameter is always a crucial issue when applying a non-parametric function estimator. In our analysis a version of *one-sided cross-validation* (OSCV) that takes into account the 1-dependent nature of the errors in model (1) will be used to choose the bandwidth of a local linear smooth. (By 1 dependent, we mean that data that are one epoch apart are possibly dependent and data that are more than one epoch apart are independent.) We shall refer to this version of OSCV as OSCV1. In the setting of regression with independent errors, Hart and Yi (1998) have shown that OSCV yields a more efficient estimator of an optimal bandwidth than does ordinary cross-validation (CV). Similar results have been established by Zhao (2003) for 1-dependent data when comparing the OSCV1 and CV1 methods, a modification of ordinary CV for 1 dependence. Henceforth, S denotes statistic (2) in which $\hat{\mu}$ is a local linear estimate with smoothing parameter chosen by the OSCV1 method.

We now elaborate on the methods mentioned in the preceding paragraph. Let $\hat{\mu}(x; h)$ denote a local linear estimate of $\mu(x)$ having bandwidth h. A popular method of choosing h is CV, the most often used form of which selects h to minimize

$$\mathrm{CV}(h) = \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \hat{\mu}_i(x_i; h)\}^2, \tag{3}$$

where $\hat{\mu}_i(\cdot; h)$ is a local linear estimate that is computed from all the observations except Y_i, $i = 1, \ldots, n$. This version of CV is appropriate for independent observations, since in that case the predictor $\hat{\mu}_i(x_i; h)$ is independent of Y_i. However, data from model (1) are *negatively* serially correlated, implying that the minimizer of equation (3) will tend to be too *large* (Hart, 1994). To adapt CV to our model, we make use of the fact that data from model (1) are 1-dependent data. We thus define the curve $\mathrm{CV}_1(h)$ precisely as in equation (3) except that $\hat{\mu}_i(\cdot; h)$ is a local linear smooth computed from all the observations except Y_{i-1}, Y_i and Y_{i+1}. In so doing the predictor $\hat{\mu}_i(x_i; h)$ is independent of Y_i, and $\mathrm{CV}_1(h)$ is an approximately unbiased estimator of

$$\sum_{i=1}^{n} E\{\hat{\mu}(x_i; h) - \mu(x_i)\}^2 / n,$$

i.e. the optimality criterion of mean average squared error.

Now, as mentioned before, Hart and Yi (1998) have shown that OSCV is a more efficient bandwidth selector than ordinary CV in the setting of independent observations. The OSCV curve is defined by

$$\mathrm{OSCV}(h) = \frac{1}{n} \sum_{i=5}^{n} \{Y_i - \tilde{\mu}_i(x_i; h)\}^2, \tag{4}$$

where $\tilde{\mu}_i(\cdot; h)$ is a local linear estimator that is based on the observations $Y_1, \ldots, Y_{i-1}$. The minimizer of $\mathrm{OSCV}(h)$ multiplied by an appropriate known constant (Hart and Yi, 1998) is the

OSCV bandwidth that is used in an ordinary, i.e. two-sided, local linear smooth. The sum in equation (4) starts at 5 (rather than 1) since a local linear estimate that is based on fewer than four observations can be rather unstable. It may seem paradoxical that OSCV is more efficient than CV, since OSCV is based on predictors that use fewer observations than do the two-sided CV predictors. An explanation for this phenomenon was provided by Hart and Lee (2005).

The OSCV curve for 1-dependent data is defined exactly as in equation (4) but with $\tilde{\mu}_i(\cdot;h)$ computed from $Y_1,\ldots,Y_{i-2}$. The minimizer of this curve is multiplied by the same constant as in the case of independent data (Zhao, 2003). We apply the OSCV1 method in carrying out our tests of the no-trend hypothesis. Simulation is used to justify that tests based on the OSCV1 method are more powerful than those using the CV1 method.

3.2. Bootstrapping the test statistic

Approximating the distribution of S assuming that H_0 is true is a non-standard exercise for at least two reasons. First of all, the limit distribution (as $n \to \infty$) of S is non-standard because of the fact that OSCV1 bandwidths have complicated asymptotic distributions. A second problem is that, at least in small samples, the null distribution of S may depend on unknown parameters of the error process. A common way of dealing with such problems is to use the bootstrap. In our problem one could bootstrap on a per-star basis by drawing samples from an error process with parameters equal to those that maximize $L_n(\boldsymbol{\eta}, \hat{\mu})$ with respect to $\boldsymbol{\eta}$. We prefer to use a bootstrap method that produces just a few reference distributions to which test statistics are compared. The advantage of this approach is that a few reference distributions can be more efficiently estimated than can 378 distinct sampling distributions.

To define our reference distributions we first need to define a few objects. Write

$$\rho = \frac{\sigma_I}{\exp(\beta_0/2)},$$

and let $\boldsymbol{\eta}' = (\rho, \beta_1)$. The cumulative distribution function (CDF) $G_n(\cdot|\boldsymbol{\eta}'_0)$ denotes the distribution of our test statistic S when the data are a sample of size n from model (1) with μ identical to a constant, assumptions 1–3 in force and $\boldsymbol{\eta}' = \boldsymbol{\eta}'_0$. We note that the distribution of S is invariant to the value of $\exp(\beta_0)$ (and hence β_0) whenever μ is constant, and hence G_n represents the null distribution of S given n and all three error parameters. Assuming our 378 stars to be a random sample from a population of similar stars, D will denote the distribution of $\boldsymbol{\eta}'$ in the subpopulation of stars that do not have trends. This subpopulation and its complement are denoted $\mathcal{T}^c$ and $\mathcal{T}$ respectively. Letting H denote the parameter space of $\boldsymbol{\eta}'$, we then define a sample size n reference distribution F_n by

$$F_n(t) = \int_H G_n(t|\boldsymbol{\eta}')\, \mathrm{d}D(\boldsymbol{\eta}').$$

To interpret F_n properly, we introduce the following assumption.

Assumption 4. Within the subpopulation $\mathcal{T}^c$, $\boldsymbol{\eta}'$ and sample size are independent.

Now, if assumption 4 is true, then, for a randomly selected star with no trend, F_n is the conditional distribution of statistic S given that the selected star has sample size n. It is important to note that we condition on n in our definition of F_n, but we average over all $\boldsymbol{\eta}'$. Averaging over $\boldsymbol{\eta}'$ avoids the problem that was mentioned before that the conditional distribution of S might depend on $\boldsymbol{\eta}'$. Our approach in this regard is analogous to an approach which was proposed by Bayarri and Berger (2000) for dealing with composite hypotheses. When testing a single

composite null hypothesis, they proposed a reference distribution of the same form as our F_n. The main distinction between their formulation and ours is that their D is a prior distribution that must be specified by the data analyst, whereas our D has objective reality and can be estimated from the data. So, the distinction between our method and that of Bayarri and Berger (2000) is equivalent to the distinction between empirical Bayes (Robbins, 1956) and Bayes methods. As indicated below, we estimate D by the empirical distribution of estimates of η'.

The formal validity of our approach for estimating F_n involves another assumption, which we now state.

Assumption 5. The distribution of η' is the same within the two subpopulations $\mathcal{T}$ and $\mathcal{T}^c$.

Under assumptions 4 and 5, we propose the following bootstrap algorithm for estimating F_n.

(a) Randomly select one of the 378 stars in our database. Associated with that star is $\hat{\eta} = (\hat{\sigma}_I^2, \hat{\beta}_0, \hat{\beta}_1)$, the maximizer of the star's generalized profile likelihood.

(b) Generate independent and identically distributed standard normal variates $Z_1, \ldots, Z_{2n+1}$, let $\varepsilon_i^* = \exp(\hat{\beta}_1 x_i/2) Z_{n+i+1}$, $i = 0, \ldots, n$, and define

$$Y_i^* = \hat{\rho} Z_i + \varepsilon_i^* - \varepsilon_{i-1}^*, \qquad i = 1, \ldots, n,$$

where

$$\hat{\rho} = \frac{\hat{\sigma}_I}{\exp(\hat{\beta}_0/2)}.$$

(c) Compute S^* from $Y_1^*, \ldots, Y_n^*$ in exactly the same way that the statistic S is computed from observations $Y_1, \ldots, Y_n$.

(d) Repeat steps (a)–(c) B times, resulting in bootstrap statistics $S_1^*, \ldots, S_B^*$.

(e) The p-value for a star with sample size n and S observed to be s is then approximated by $\sum_{i=1}^{B} I_{[s,\infty)}(S_i^*)/B$.

A theoretical justification for this bootstrap algorithm is given in Appendix A, where we also discuss how the algorithm could be modified to avoid assumptions 4 and 5. This modification involves estimating the joint distribution of n, 'size' of trend and η'. Obviously, this distribution cannot be estimated as efficiently as can the marginal distribution of η', and hence, when it is justified, the above algorithm will be preferable. In Appendix A.2 we provide some evidence that assumptions 4 and 5 are at least approximately valid for the data under consideration.

4. Data analysis

The procedure that was described in the previous section was carried out for the 378 stars in our database. The local linear smooths employ a standard Gaussian kernel and have the form

$$\hat{\mu}(x; h) = \sum_{i=1}^{n} w_{i,n}(x; h) Y_i$$

for weights $w_{i,n}(x; h)$ that depend only on x, the design points $x_1, \ldots, x_n$ and bandwidth h. The OSCV1 curve was computed at the same set of 30 bandwidths for each star. These bandwidths were chosen so that the effective numbers of parameters of the 30 smooths were $2, 3, \ldots, 31$. Define $\mathbf{W}_n(h)$ to be the $n \times n$ matrix having element $w_{j,n}(x_i; h)$ in the ith row and jth column, $i = 1, \ldots, n$, $j = 1, \ldots, n$. This is the so-called hat matrix, i.e. the matrix which when multiplied by the data vector $(Y_1, \ldots, Y_n)^T$ produces the predicted values $\hat{\mu}(x_1; h), \ldots, \hat{\mu}(x_n; h)$. Hastie and

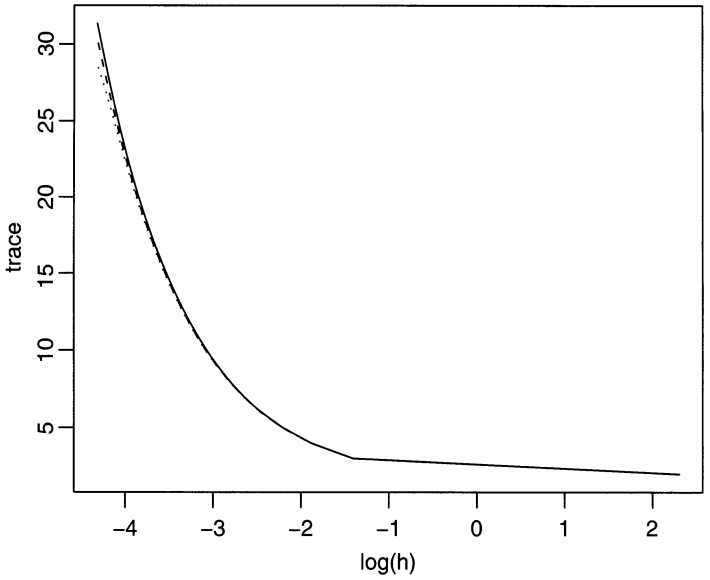

Fig. 2. Traces of the hat matrix: ·····, $n = 32$; -------, $n = 40$; ———, $n = 212$

Tibshirani (1990), pages 52–55, argued that $\mathrm{tr}\{\mathbf{W}_n(h)\}$ provides a good proxy for the degrees of freedom of a smooth.

We chose our 30 bandwidths $h_1, \ldots, h_{30}$ in such a way that, at $n = 74$, the median sample size of all stars, $\mathrm{tr}\{\mathbf{W}_n(h_i)\} = i + 1$, $i = 1, \ldots, 30$. The quantity $\mathrm{tr}\{\mathbf{W}_n(h)\}$ does vary with n but, for the range of sample sizes in our database, all traces are in remarkable agreement, as shown in Fig. 2. For this reason we use the same set of bandwidths for every star. The local linear smooth with bandwidth h_i provides roughly the same fit as a least squares polynomial of degree i. This interpretation is virtually exact in the case of $h_1 = 10$, since it is well known that as $h \to \infty$ the local linear estimate tends to the least squares straight line.

Fig. 3 shows the distribution of $\mathrm{tr}\{\mathbf{W}_n(\hat{h})\}$ over all 378 stars, where $\hat{h}$ is the OSCV1 bandwidth. The distribution of $\mathrm{tr}\{\mathbf{W}_n(\hat{h})\}$ assuming that hypothesis H_0 is true is also shown for comparison. The latter distribution is a weighted average of bootstrap distributions corresponding to different sample sizes between 32 and 212. The bootstrap scheme that was used was precisely the scheme that was described in Section 3, and the weights were chosen proportional to the frequency with which sample sizes appeared in our database. The most obvious difference between the distributions in Fig. 3 is how often the largest bandwidth (i.e. $\mathrm{tr}\{\mathbf{W}_n(h)\} = 2$) is chosen. Under hypothesis H_0, this bandwidth is chosen about 71% of the time, whereas only about 56% of the stars had $\mathrm{tr}\{\mathbf{W}_n(\hat{h})\} = 2$. This in itself is evidence that some non-trivial proportion of stars has trends.

We also computed the CV1 bandwidth for each of the stars and used simulation to approximate the CV1 null distribution at the median sample size of $n = 74$. The two trace distributions are shown in Fig. 4. Under hypothesis H_0, the CV1 distribution is much less concentrated near 2 than is the OSCV1 null distribution. The mean and standard deviation of the latter distribution are 2.39 and 0.75 respectively, whereas the same quantities for the CV1 distribution are 4.38 and 5.04. A simulation study which is described in Appendix A shows that no-trend tests based on CV1 bandwidths have extremely poor power in comparison with those based on OSCV1 bandwidths. The long right-hand tail of the null CV1 distribution explains this.

Our use of a heterodscedastic model for the experimental errors was justified by the fact that 332 of the 378 stars, or 88%, had generalized profile likelihood estimates of β_1 that were less

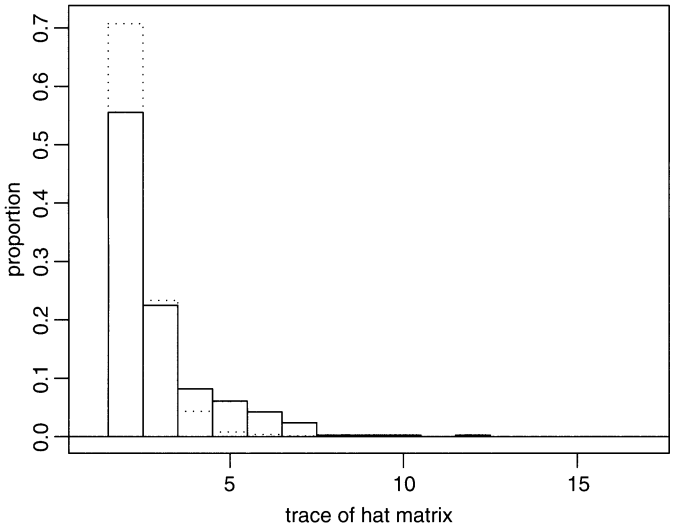

Fig. 3. OSCV1 bandwidth distributions: □, distribution of $\mathrm{tr}\{\mathbf{W}_n(\hat{h})\}$ over all 378 stars; ⋯, distribution of $\mathrm{tr}\{\mathbf{W}_n(\hat{h})\}$ under the null hypothesis of no trend

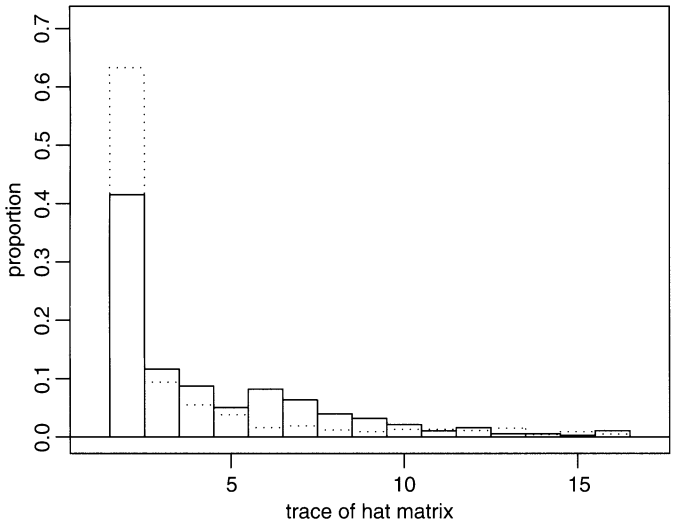

Fig. 4. CV1 bandwidth distributions: □, distribution of $\mathrm{tr}\{\mathbf{W}_n(\tilde{h})\}$ over all 378 stars; ⋯, distribution of $\mathrm{tr}\{\mathbf{W}_n(h)\}$ under the null hypothesis of no trend, where $\tilde{h}$ is the CV1 bandwidth

than 0. This is consistent with our observation that scatter in most data sets appears to *decrease* over time. The median value of β_1 among stars with negative estimates of β_1 was -2.18, whereas the median among stars with positive estimates was 0.70.

4.1. Testing the no-trend hypothesis

We turn now to the question of testing the no-trend hypothesis for each of the stars. The first step is to obtain bootstrap reference distributions, as described in Section 3. This was done for

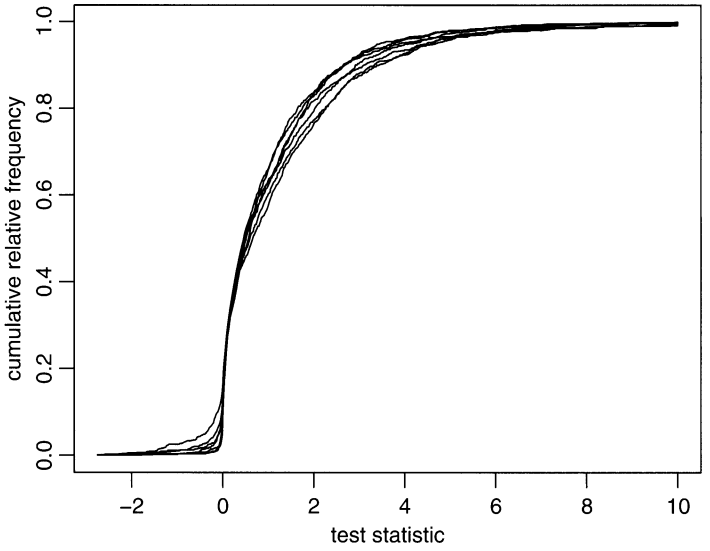

Fig. 5. Bootstrap reference distributions: the various distributions correspond to sample sizes 32, 50, 68, 86, 128, 170 and 212

sample sizes of $n = 32, 50, 68, 86, 128, 170, 212$. The number of bootstrap samples at each n was 1000, and the resulting reference distributions are shown in Fig. 5. The main point of showing Fig. 5 is to indicate how similar the seven distributions are, especially at larger quantiles, which determine a statistic's rejection region. Roughly speaking, quantiles decrease with sample size, as we would expect. For example, the 95th quantiles are 4.59, 4.69, 4.20, 4.49, 3.89, 3.46 and 3.71 for $n = 32, 50, 68, 86, 128, 170, 212$ respectively.

Let $\hat{F}_n$ be the bootstrap distribution of statistic S for sample size n, and let $n_1 < \ldots < n_7$ be the seven sample sizes in the simulation. For a star with sample size n and $S = s$, let n_j be such that $n_j \leqslant n < n_{j+1}$. Then the p-value is taken to be

$$p = 1 - \frac{(n_{j+1} - n)\hat{F}_{n_j}(s) + (n - n_j)\hat{F}_{n_{j+1}}(s)}{n_{j+1} - n_j},$$

which is simply a linear interpolation. A kernel density estimate for the 378 p-values so computed is shown in Fig. 6. The preponderance of values near 0 is evidence that there are significant trends, since otherwise the p-values would be approximately uniformly distributed. Indeed, 101 stars, or 27%, have p-values that are smaller than 0.05 and 56, or 15%, have values that are smaller than 0.01.

4.2. *Accounting for a multiplicity of tests*

When testing multiple hypotheses, one should address the problem of the experimentwise error rate, i.e. the probability of making multiple type I errors. With the advent of microarrays, large scale testing problems, in which hundreds or thousands of tests are conducted simultaneously, have become commonplace. A seminal paper on dealing with such situations is that of Benjamini and Hochberg (1995), who introduced the notion of the *false discovery rate*. Their method provided a great advance over classical methods that are much too conservative in large scale testing problems. Recently, even more powerful methods that still control appropriate error

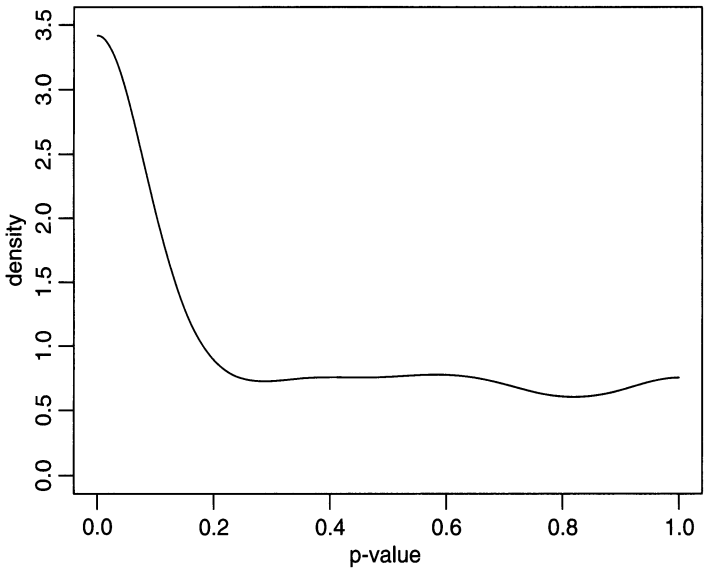

Fig. 6. Estimated density of *p*-values

rates have been devised. One such method is that of Storey (2002), which we shall apply in our analysis.

In a large scale testing problem, let R denote the number of null hypotheses that are rejected and V the number of false positive results. Storey (2002) defined the *positive false discovery rate* to be

$$\text{pFDR} = E(V/R|R>0).$$

Suppose that $P_1,\ldots,P_n$ are independent and identically distributed p-values corresponding to independent tests such that null hypothesis H_{0i} is rejected when $P_i \leqslant \gamma$, $i=1,\ldots,n$. In this situation Storey (2002) argued that

$$\begin{aligned}\text{pFDR}(\gamma) &= P(H_{0i} \text{ is true}|P_i \leqslant \gamma)\\ &= \frac{\pi_0 P(P_i \leqslant \gamma|H_{0i} \text{ is true})}{P(P_i \leqslant \gamma)},\end{aligned}$$

where π_0 is the proportion of all null hypotheses that are true. Since P_i is uniformly distributed under the null hypothesis, we may express this result as

$$\text{pFDR}(\gamma) = \pi_0\gamma/P(P_i \leqslant \gamma). \tag{5}$$

If π_0 and the CDF of P_i were known, equation (5) could be used to choose γ to achieve a desired positive false discovery rate. The next best option would be to estimate the unknown quantities. The CDF of P_i is easily estimated from the observed p-values, and Storey (2002) proposed a scheme for estimating π_0. We use a slightly different method for estimating π_0 that is based on density estimation. In so doing, we are in essence estimating a component of the *local false discovery rate*, as defined by Efron *et al.* (2001)

We have

$$\begin{aligned}F(\gamma) &\equiv P(P_i \leqslant \gamma)\\ &= \pi_0\gamma + (1-\pi_0)\, G(\gamma),\end{aligned}$$

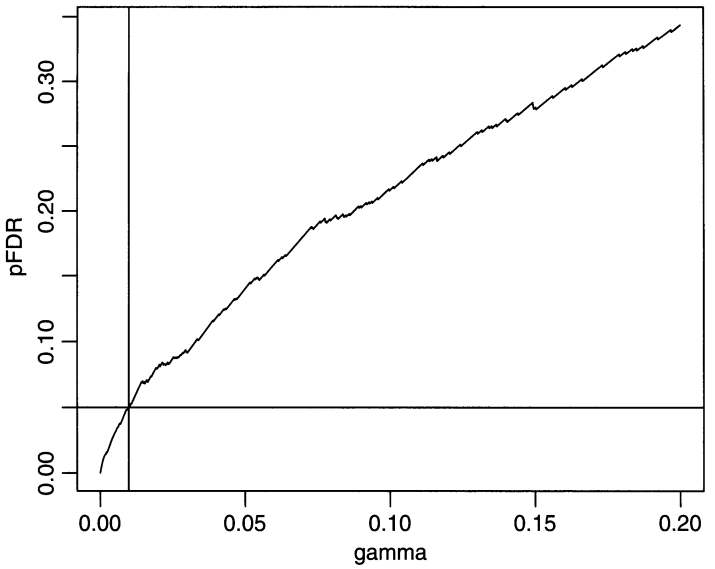

Fig. 7. Estimated positive false discovery rate as a function of critical value γ

where G is the CDF of a p-value given that the alternative hypothesis is true. Assuming that G has derivative g, the p-value density f is thus

$$f(\gamma) = \pi_0 + (1 - \pi_0)\, g(\gamma).$$

We first note that $f(\gamma) \geqslant \pi_0$, and so, for *any* γ, replacing π_0 by $\min\{1, \hat{f}(\gamma)\}$ in equation (5) would be a conservative procedure for estimating $\mathrm{pFDR}(\gamma)$. Obviously, however, it is desirable to try to obtain a realistic estimate of π_0. Typically, the density g will be small near $\gamma = 1$, and hence $f(\gamma) \approx \pi_0$ for γ near 1. Fig. 6 suggests that in our case $f(\gamma)$ is approximately constant for $\gamma \geqslant 0.3$. These observations lead us to estimate π_0 by an average of kernel density estimates at $\gamma \geqslant 0.6$. We used a Gaussian kernel and employed data reflection (as in Cline and Hart (1991)) to deal with boundary effects. The estimates so computed were not overly sensitive to bandwidth. For bandwidths $h = 0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32$ the estimates of π_0 were $0.657, 0.660, 0.667, 0.673, 0.675, 0.682$ and 0.740 respectively.

Taking a conservative route, we estimate π_0 in equation (5) by 0.75. Defining $\hat{F}$ to be the empirical CDF of our 378 p-values, our estimate of pFDR is

$$\widehat{\mathrm{pFDR}}(\gamma) = 0.75\gamma / \hat{F}(\gamma).$$

A plot of this estimate for $0 \leqslant \gamma \leqslant 0.2$ is shown in Fig. 7. Also indicated in Fig. 7 is the value of γ that is associated with a pFDR of 0.05. This value is about 0.00987, meaning that we could reject hypothesis H_0 for any p-value that is smaller than 0.00987 with the assurance that there are only about 5% false positive results. There were 56 stars with p-values smaller than 0.00987. The more conservative method of Benjamini and Hochberg (1995) yields only 52 significant p-values when controlling the false discovery rate at the 5% level. When using an error rate of 1% the methods of Storey and Benjamini and Hochberg yield 27 and 26 significant trends respectively. The p-values for all 378 stars are available from the authors.

4.3. Nature of trends

It is of some interest to know the nature of the trends for the stars that had small p-values. An inspection of the OSCV1 local linear smooths revealed that relatively few had a substantial

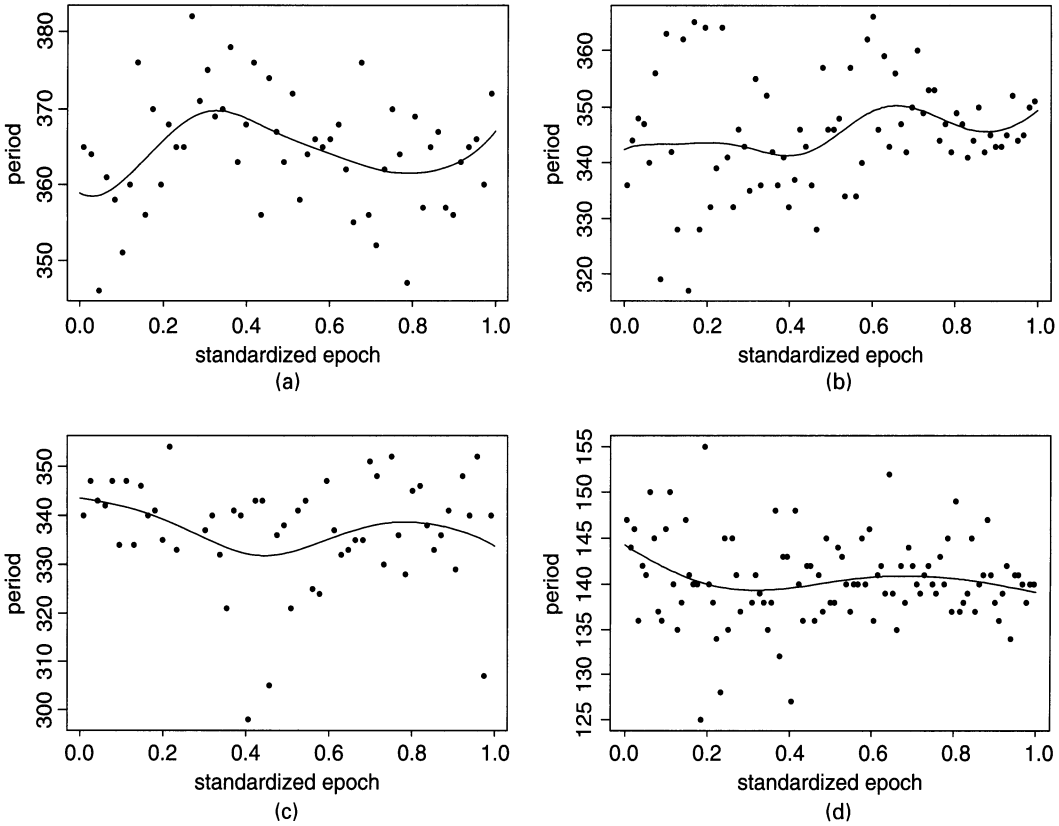

Fig. 8. Scatterplots for four variable stars and OSCV1 local linear smooths: (a) *S Sculptoris*; (b) *W Pegasi*; (c) *RS Hydrae*; (d) *SS Cassiopeiae*

linear component. To quantify this behaviour, we computed, for each of the 101 stars with p-values smaller than 0.05, the statistic

$$R = \sum_{i=1}^{n} (a + bx_i - \bar{Y})^2 \Big/ \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2,$$

where $a + bx$ is the least squares line and $\hat{Y}_1, \ldots, \hat{Y}_n$ are values of the OSCV1 local linear smooth. This statistic is the fraction of total fitted variation that is due to a linear component. The 10th, 25th, 50th and 75th percentiles of R for the 101 stars in question were 0.006, 0.054, 0.140 and 0.464 respectively. Fig. 8 shows scatterplots and local linear smooths for four of the 56 stars that were judged to have trends significant at the 5% pFDR level. The four stars were chosen by the sizes of their p-values, which had ranks 11, 22, 34 and 45 among all 378 stars. The trends are typical. When there is a systematic drift away from a constant period, the process tends to return towards its previous level.

The wavelike trend behaviour in Fig. 8 hints at periodicity. However, in most cases we observe at most one 'period', if indeed we are observing periodic behaviour at all. Similar wavelike behaviour in long-term mean pulsation periods has been remarked on by Berdnikov *et al.* (2000) in a different context: that of three 'Cepheid' pulsating stars, with periods in the range 5.4–10.2 days. Although the data that were analysed by them were more sparse than ours (owing to missing values), more than 5000 pulsation cycles were covered by the long time base-line of

the measurements. The waves were clearly *not* strictly periodic. Interpretation of these results is obviously in the domain of an astrophysicist dealing with stellar pulsation theory.

4.4. Relationship between strength of trend and mean period

Also of some interest is investigating how the likelihood of a trend is related to various characteristics of variable stars. As a first step in this direction, we plotted p-values of trend tests against average observed period length. The latter quantity is simply

$$\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n},$$

where $Y_1, \ldots, Y_n$ are the observations for a given star. A local linear smooth of this scatterplot indicated a tendency for longer period stars to have smaller p-values. This tendency is highly statistically significant. When the order selection test (with cosine basis) of Eubank and Hart (1992) is used to test the null hypothesis of no relationship between p-values and period means, the resulting observed significance level is 2×10^{-6}. To appreciate what this test means, let us consider $E(p|\mu)$, the expected p-value for stars having a mean period of μ. This may be expressed as

$$E(p|\mu) = \tfrac{1}{2} p(\mu) + \{1 - p(\mu)\} E(p|\mu, \text{trend}), \tag{6}$$

where $p(\mu)$ is the proportion of all stars with mean period μ that do not have trends and $E(p|\mu, \text{trend})$ is the expected p-value among stars having trends and mean period μ. Can a negative relationship between p-values and mean period exist when trend behaviour is completely unrelated to mean period? The answer is no, as we shall now argue on the basis of equation (6). If trend behaviour is not dependent on μ, then

(a) $p(\mu) \equiv p_0$ for all μ and
(b) $E(p|\mu, \text{trend})$ is dependent only on sample size.

If statement (b) is true, then $E(p|\mu, \text{trend})$ increases with μ. This is because the longer period stars are those with the smaller sample sizes, since fewer pulsation cycles have elapsed since initiation of the observations; Fig. 10(a) in Appendix A.2. If both statement (a) and statement (b) are true, it thus follows that $E(p|\mu)$ is a non-decreasing function of μ. In summary, the negative association that has been established between p-values and mean periods is the *opposite* of what would be expected in the scientifically uninteresting case where neither relative frequency nor strength of trend depend on mean period.

The negative relationship between p-value and mean period still allows for the possibility that the relative frequency of trends is related to μ, but *strength of trend* is not. To investigate the latter type of relationship, we consider the signal-to-noise ratio statistic

$$\text{SNR} = n^{-1} \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \Big/ \hat{\sigma}_I^2,$$

where $\hat{Y}_1, \ldots, \hat{Y}_n$ are values of the OSCV1 local linear smooth and $\hat{\sigma}_I^2$ is the maximum likelihood estimate of σ_I^2. Fig. 9 is a plot of $\text{SNR}^{1/5}$ *versus* mean period length. Here there is a clear indication of a positive relationship between strength of trend and mean period length. The pictured line is the least squares line, whose slope is significantly different from 0 at level of significance 10^{-8}. It is worth noting that this conclusion does not change if 'outliers' with $\text{SNR}^{1/5} > 1.2$ are eliminated from the data set.

The time period that is covered by our data is, in astronomical terms, but the blink of an eye. It therefore seems unlikely that the changes in the mean pulsation periods that were found here

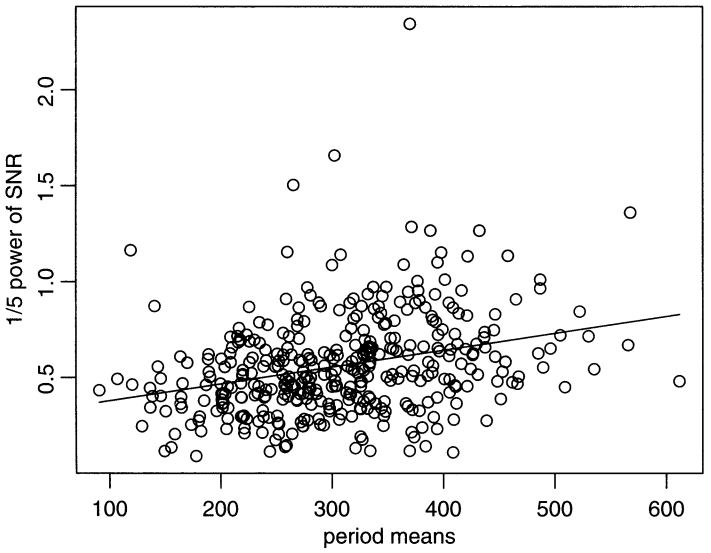

Fig. 9. Power-transformed signal-to-noise ratios *versus* period means

are due to systematic evolutionary changes in the stars. They appear more likely to be due to small secular changes in the structures of the stars. The long period pulsating stars are known to suffer loss of mass, which generally increases with increasing period among members of this stellar type. This leads to fairly rapid changes in the environment defining the outer boundaries of these stars. We speculate that this could explain why the relatively short-term trends in mean pulsation are more pronounced in the longer period stars.

5. Concluding remarks

This paper has undertaken an analysis of statistical properties of times between maximum brightnesses for a database of 378 long period variable stars. The main conclusions of our analysis are as follows.

(a) We provide evidence that the observed pulsation periods of most of our long period variables are heteroscedastic.
(b) Using a method that controls the positive false discovery rate to be 0.05, 56 of the 378 stars in our database have significant trends in times between maximum brightness.
(c) Most of the trends that are statistically significant are non-monotonic, with at most a very weak linear component.
(d) There is a clear tendency for the strength of trend to be positively related to the mean period of a star.

In the course of our analysis we have proposed and made use of a bootstrap test that accounts for both variation between stars and variation in the observations for each individual star. A theoretical justification has been given for this test, which would be useful in many random-effects models and repeated measures settings. Another important statistical finding concerns smoothing-based lack-of-fit tests that employ local linear estimators. We have shown that such tests are *much* more powerful when the smoothing parameter of the local linear estimator is chosen by one-sided CV rather than ordinary CV.

Acknowledgements

We are grateful to the American Association of Variable Star Observers for supplying the machine readable data that were used in this paper. We also thank three reviewers for their insightful comments that led to this improved version of our paper. The work of Professor Hart was supported by National Science Foundation grants DMS-0203884 and DMS-0604801.

Appendix A

Here we provide supplementary material justifying certain claims that are made in the paper, and we also describe an extension of our bootstrap algorithm that avoids assumptions 4 and 5.

A.1. Bootstrap justification

Suppose that assumptions 1–5 hold, a random sample of N stars is obtained and estimates $\hat{\boldsymbol{\eta}}_1, \ldots, \hat{\boldsymbol{\eta}}_N$ of error parameters are computed for these stars. Let $J(y_1, \ldots, y_n|\boldsymbol{\eta}')$ denote the joint distribution of $Y_j = I_j + \varepsilon_j - \varepsilon_{j-1}$, $j = 1, \ldots, n$, when $\sigma_I = 1$ and $(\rho, \beta_1) = \boldsymbol{\eta}'$. The bootstrap that was employed in Section 3.2 provides a Monte Carlo approximation to

$$\hat{F}_n(t) = \int_H \int_{\Re^n} I_{(-\infty,t]}\{s(y_1, \ldots, y_n)\}\, \mathrm{d}J(y_1, \ldots, y_n|\boldsymbol{\eta}')\, \mathrm{d}\hat{D}_N(\boldsymbol{\eta}'),$$

where $s(y_1, \ldots, y_n)$ is the value of the test statistic S when the observations $Y_1, \ldots, Y_n$ take on values $y_1, \ldots, y_n$ and $\hat{D}_N$ is the empirical distribution of $\hat{\boldsymbol{\eta}}'_1, \ldots, \hat{\boldsymbol{\eta}}'_N$. For a star with sample size n, the true null distribution of test statistic S is F_n, and we wish to establish conditions under which $\hat{F}_n$ converges in probability to F_n as $N \to \infty$.

Since $\hat{F}_n$ may be approximated arbitrarily well by generating sufficiently many samples, we assume that it is known. Let D_N denote the empirical distribution of the *true* parameters $\boldsymbol{\eta}'_1, \ldots, \boldsymbol{\eta}'_N$ that are associated with the N stars in the random sample. We may write

$$\hat{F}_n(t) = F_n(t) + E_1(t) + E_2(t),$$

where

$$E_1(t) = \int_H \int_{\Re^n} I_{(-\infty,t]}\{s(y_1, \ldots, y_n)\}\, \mathrm{d}J(y_1, \ldots, y_n|\boldsymbol{\eta}')\{\mathrm{d}D_N(\boldsymbol{\eta}') - \mathrm{d}D(\boldsymbol{\eta}')\}$$

and

$$E_2(t) = \int_H \int_{\Re^n} I_{(-\infty,t]}\{s(y_1, \ldots, y_n)\}\, \mathrm{d}J(y_1, \ldots, y_n|\boldsymbol{\eta}')\{\mathrm{d}\hat{D}_N(\boldsymbol{\eta}') - \mathrm{d}D_N(\boldsymbol{\eta}')\}.$$

Now, $E_1(t)$ is the classical sort of bootstrap error, which will converge almost surely to 0 as $N \to \infty$ since D_N is a strongly consistent estimator of D.

The error $E_2(t)$ is completely a function of the difference between the actual error parameters $\boldsymbol{\eta}'_1, \ldots, \boldsymbol{\eta}'_N$ and their estimates $\hat{\boldsymbol{\eta}}'_1, \ldots, \hat{\boldsymbol{\eta}}'_N$. Let n_i, $i = 1, \ldots, N$, be the numbers of observations for the N stars in our sample. Using a straightforward argument that is available from the authors, it can be shown that $E_2(t)$ converges to 0 in probability if the following conditions hold.

(a) $\min_{1 \leqslant i \leqslant N}(n_i) \to \infty$.
(b) The distribution D has a bounded density.
(c) For every N, $E(\hat{\rho}_i - \rho_i)^2 + E(\hat{\beta}_{1i} - \beta_{1i})^2 \leqslant C/n_i$, $i = 1, \ldots, N$, where $C < \infty$.
(d) Define the function

$$g_{n,t}(\boldsymbol{\eta}') = \frac{\partial}{\partial \rho} \frac{\partial}{\partial \beta_1} \int_{\Re^n} I_{(-\infty,t]}\{s(y_1, \ldots, y_n)\}\, \mathrm{d}J(y_1, \ldots, y_n|\boldsymbol{\eta}').$$

Then $g_{n,t}$ is absolutely integrable (with respect to $\boldsymbol{\eta}'$) for each n and the integrals are uniformly bounded in n.

A.2. Assumptions 4 and 5

Assumption 4 asserts that n and η' are independent. To investigate this assumption we used the estimates $\hat{\eta}_1, \ldots, \hat{\eta}_N$ from our analysis as proxies for the true parameter values. Scatterplots of $z_1 = \log(\hat{\rho})$ *versus* n and $z_2 = \text{sgn}(\hat{\beta}_1)\log(|\hat{\beta}_1|)$ *versus* n are shown in Fig. 10. (The same two outliers are excluded from each of Figs 10(a)–10(d).) At best a very weak association appears to exist between n and either of $\hat{\rho}$ and $\hat{\beta}_1$.

It is fortunate that the test statistic S is invariant to the value of σ_I, since, as seen from Fig. 10(b), there *is* a non-trivial association between n and $\hat{\sigma}_I$. The negative association between n and σ_I can be traced to the fact that n is intimately related to period length. Since all stars are observed over roughly the same time period, the sample size is nearly determined by the average period length, as seen in Fig. 10(a). So, stars with larger values of n have shorter mean periods, and shorter random periods tend to be less variable.

Fig. 11 is provided as evidence that assumption 5 holds to a reasonable approximation. The statistic

$$V = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \Big/ \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

is used as a proxy for whether or not the null hypothesis is true. Indeed, V estimates the parameter

$$\xi = \sum_{i=1}^{n}\{\mu(x_i) - \bar{\mu}^2\} \Big/ \sum_{i=1}^{n}\text{var}(Y_i),$$

which is 0 if and only if the no-trend hypothesis is true. Each of Figs 11(a)–11(f) has the same (x, y)-scale. Figs 11(a), 11(c) and 11(e) respectively provide a comparison of the distribution of (z_1, z_2) under the null hypothesis with those of Figs 11(b), 11(d) and 11(f) under the alternative. Figs 11(a), 11(c) and 11(e) each

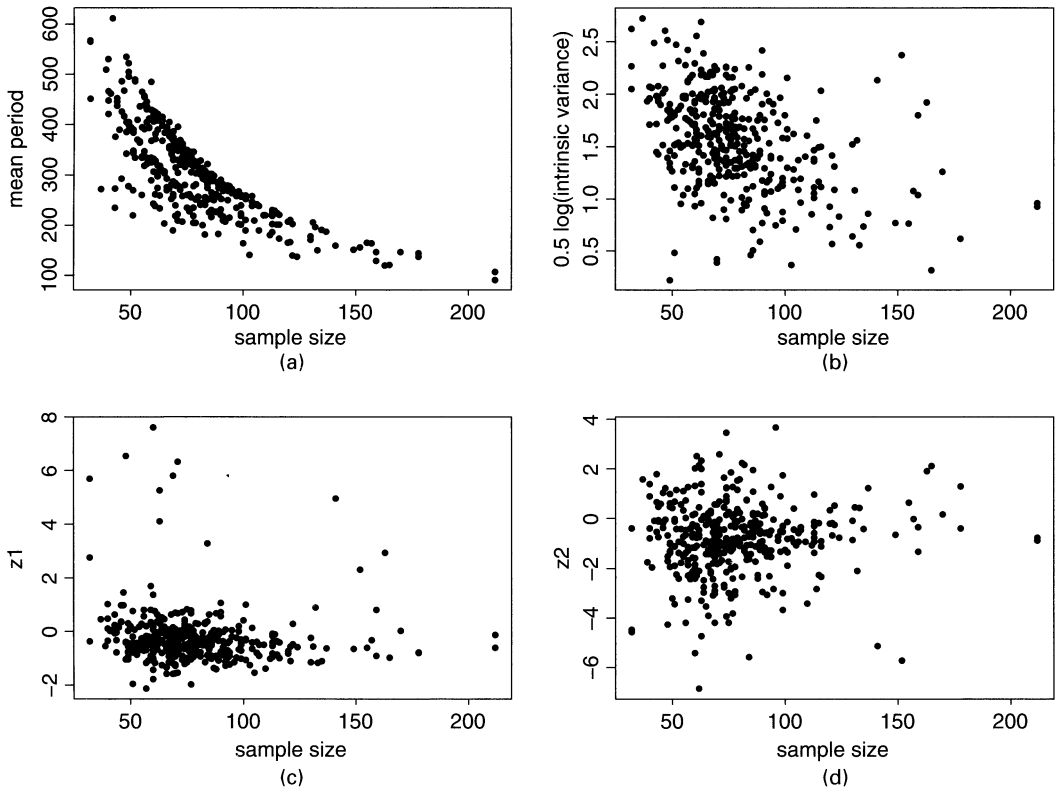

Fig. 10. Scatterplots of various parameter estimates *versus* sample size: (a) mean period; (b) log(intrinsic variance); (c) z_1 ($\log(\hat{\rho})$); (d) z_2 ($\text{sgn}(\hat{\beta}_1)\log(|\hat{\beta}_1|)$)

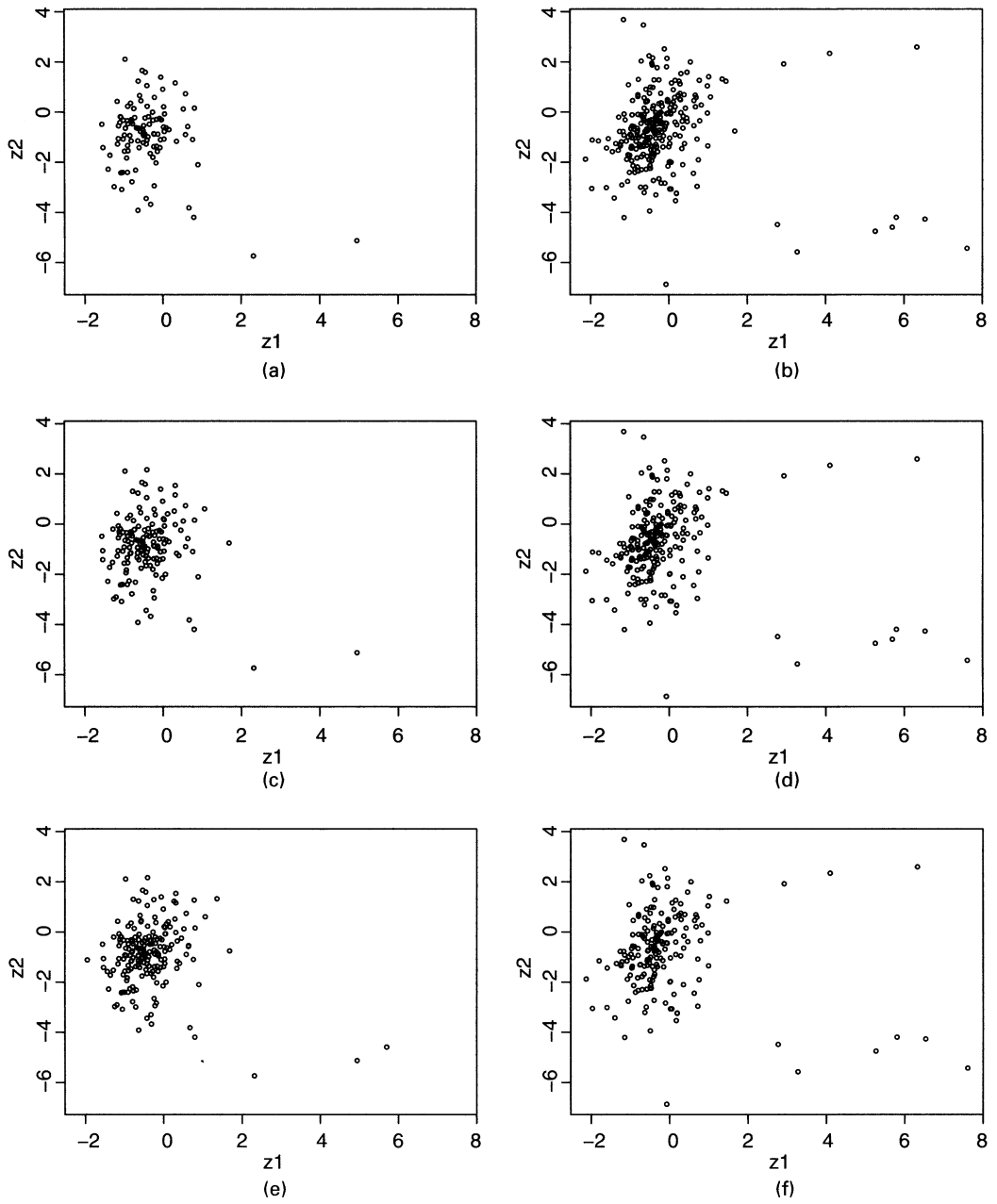

Fig. 11. Plots addressing the validity of assumption 5 ($z_1 = \log(\hat{\rho})$ and $z_2 = \text{sgn}(\hat{\beta}_1)\log(|\hat{\beta}_1|)$): (a) $V < 0.004$; (b) $V > 0.004$; (c) $V < 0.008$; (d) $V > 0.008$; (e) $V < 0.016$; (f) $V > 0.016$

correspond to all stars whose values of V are less than the indicated amount, and Figs 11(b), 11(d) and 11(e) correspond to the complementary stars.

The scatterplots in each row are quite similar, and the plots do not depend substantially on the value of V that is used as a cut-off. On this basis it appears that, even if assumption 5 is violated, the violation is not sufficiently substantial to affect our analysis materially.

A.3. Generalizing the bootstrap

Here we propose a bootstrap algorithm that avoids assumptions 4 and 5. Applying the Bayarri–Berger principle, the null distribution of the test statistic S for a star with sample size n is

$$P(S \leqslant t|n, \xi = 0) = \int_H P(S \leqslant t|n, \xi = 0, \eta')\, \mathrm{d}G(\eta'|n, \xi = 0),$$

where ξ is defined in the previous section and $G(\cdot|n, \xi = 0)$ is the conditional distribution function of η' among stars in subpopulation $\mathcal{T}^c$ that have sample size n.

Under assumptions 4 and 5, the conditional distribution of η' given n and $\xi = 0$ is equal to the unconditional distribution of η', and hence we may simply resample from all N stars to approximate the requisite sampling distribution. More generally it is necessary to estimate $G(\cdot|n, \xi = 0)$. We may use smoothing to do so. First, the parameter ξ may be estimated by V. Then, we may estimate $G(\cdot|n, \xi = 0)$ by using the empirical CDF of a subset of $\{\hat{\eta}'_1, \ldots, \hat{\eta}'_N\}$ corresponding to stars with samples sizes and values of $\hat{\xi}$ 'near' n and 0 respectively. Of course, the tricky problem here is deciding what near is. A solution to this problem in the present context of estimating a conditional CDF was provided by Hall *et al.* (1999). Li and Racine (2006) proposed methodology for estimating conditional CDFs when some covariates are discrete, which is relevant to our case where n is discrete.

A.4. Relative power of one-sided cross-validation and cross-validation tests

Our generalized profile likelihood ratio test is remarkably more powerful when the smoothing parameter of the local linear smooth is chosen by the OSCV1 as opposed to the CV1 method. We demonstrate this fact here by means of a simulation study. To make the study relevant to our variable star analysis, we use trend functions that are actually trend *estimates* from our data analysis. 101 stars had test statistics with p-values that were smaller than 0.05. Let $(\hat{\mu}_i, \hat{\eta}_i)$, $i = 1, \ldots, 101$, be the trend estimates and estimates of η for these 101 stars.

The following process was repeated independently 1000 times.

(a) Randomly select one of the 101 stars that were described immediately above.
(b) If star j was selected, generate $n = 74$ observations from model (1) under assumptions (1)–(3) with $\mu \equiv \hat{\mu}_j$ and $(\sigma_I, \beta_0, \beta_1) = (\hat{\sigma}_{Ij}, \hat{\beta}_{0j}, \hat{\beta}_{1j})$.
(c) Compute the test statistic S from the 74 observations from step (b). In addition, compute S_{CV}, which is identical to S except that the bandwidth of the local linear estimate is chosen by the CV1 rather than OSCV1 method.

The sample size 74 was used since that was the median sample size in our database. The bootstrap method of Section 3.1 was used to estimate the 95th percentiles of the null distributions of S and S_{CV} at $n = 74$. The two estimated percentiles were 3.93 and 26.99 respectively. The large discrepancy between these two suggests that S_{CV} may lead to a less powerful test. In fact, in the 1000 replications of the power study, the empirical powers of the OSCV1- and CV1-based tests were 0.779 and 0.072 respectively.

References

Bayarri, M. J. and Berger, J. O. (2000) *P* values for composite null models. *J. Am. Statist. Ass.*, **95**, 1127–1142.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B, **57**, 289–300.

Berdnikov, L., Ignatova, V. V., Caldwell, J. A. R. and Koen, C. (2000) Long-term waves on Cepheid pulsation period trends. *New Astronom.*, **4**, 625–639.

Birt, W. J. (1831) Observations upon the period of the variable star β Lyrae. *Mnthly Not. R. Astronom. Soc.*, **1**, 192.

Campbell, L. (1955) *Studies of Long Period Variables*. Cambridge: American Association of Variable Star Observers.

Cleveland, W. S. and Devlin, S. J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Statist. Ass.*, **83**, 596–610.

Cline, D. B. H. and Hart, J. D. (1991) Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics*, **22**, 69–84.

Eddington, A. S. and Plakidis, S. (1929) Irregularities of period of long-period variable stars. *Mnthly Not. R. Astronom. Soc.*, **90**, 65–71.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Statist. Ass.*, **96**, 1151–1160.
Eubank, R. L. and Hart, J. D. (1992) Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.*, **20**, 1412–1425.
Fan, J. (1992) Design-adaptive nonparametric regression. *J. Am. Statist Ass.*, **87**, 998–1004.
Hall, P., Wolff, R. C. L. and Yao, Q. (1999) Methods for estimating a conditional distribution function. *J. Am. Statist. Ass.*, **94**, 154–163.
Handler, G. (2004) Amplitude and frequency variability of pulsating stars. *Communs Asteroseism.*, **145**, 71–73.
Hart, J. D. (1994) Automated kernel smoothing of dependent data by using time series cross-validation. *J. R. Statist. Soc.* B, **56**, 529–542.
Hart, J. D. and Lee, C.-L. (2005) Robustness of one-sided cross-validation to autocorrelation. *J. Multiv. Anal.*, **92**, 77–96.
Hart, J. D. and Yi, S. (1998) One-sided cross-validation. *J. Am. Statist. Ass.*, **93**, 620–631.
Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized Additive Models*. New York: Chapman and Hall.
Hoffmeister, C., Richter, G. and Wenzel, W. (1985) *Variable Stars*. Berlin: Springer.
Koen, C. and Lombard, F. (2004) The analysis of indexed astronomical time series—IX: a period change test. *Mnthly Not. R. Astronom. Soc.*, **353**, 98–104.
Li, Q. and Racine, J. S. (2006) Nonparametric estimation of conditional cdfs and quantile functions with mixed categorical and continuous data. *J. Bus. Econ. Statist.*, to be published.
Mattei, J. A., Mayall, M. W. and Waagen, E. O. (1990) *Maxima and Minima of Long Period Variables, 1949-1975*. Cambridge: American Association of Variable Star Observers.
Percy, J. R. and Colivas, T. (1999) Long-term changes in Mira stars: I, period fluctuations in Mira Stars. *Publ. Astronom. Soc. Pacif.*, **111**, 94–97.
Percy, J. R., Colivas, T., Sloan, W. B. and Mattei, J. (1990) Long-term changes in Mira variables. *Astronom. Soc. Pacif. Conf. Ser.*, **11**, 446–449.
Pokta, S. and Hart, J. D. (2007) Approximating posterior probabilities in a linear model with possibly noninvertible moving average errors. *J. Multiv. Anal.*, doi: 10.1016/j.jmva.2007.04.004, to be published.
R Development Core Team (2007) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Robbins, H. (1956) An empirical Bayes approach to statistics. In *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, pp. 157–163. Berkeley: University of California Press.
Severini, T. A. and Wong, W. H. (1992) Profile likelihood and conditionally parametric models. *Ann. Statist.*, **20**, 1768–1802.
Sterne, T. E. (1934) The errors of period of variable stars—part I: the general theory, illustrated by RR Scorpii. *Circular 386*. Harvard College Observatory, Cambridge.
Storey, J. D. (2002) A direct approach to false discovery rates. *J. R. Statist. Soc.* B, **64**, 479–498.
Whitelock, P. A. (1999) "Real-time" evolution in Mira variables. *New Astronom. Rev.*, **43**, 437–440.
Yu, K. and Jones, M. C. (2004) Likelihood-based local linear estimation of the conditional variance function. *J. Am. Statist. Ass.*, **99**, 139–144.
Zhao, C. (2003) One-sided cross-validation for a model motivated by variable star data. *PhD Dissertation*. Department of Statistics, Texas A&M University, College Station.