

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, South Africa

and Fred Lombard
University of Johannesburg, South Africa

Abstract

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 nonparametric in that potential trends are estimated by local linear smoothers. Our testing methodology has a number of 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 upon 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 based on the usual version of cross-validation.

The positive false discovery rate (pFDR) method of Storey (2002) 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 pFDR 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 brightness changes 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 variables.” These stars are distinguished by their substantial brightness changes, which are roughly sinusoidal with typical periods between 100 and 300 days. The period of a given star is determined by its internal structure. Period changes 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 the book by Hoffmeister, Richter and Wenzel (1985).

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 1473 hits; “period variability” gives 2515 hits, with more than 100 in each case from the last 3 years. The reason for the interest lies primarily in the fact, mentioned above, that period changes 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 Variables (also known as Mira stars) is given in Whitelock (1999).

The data 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, Mayall and Waagen (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 authors 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 used by the latter authors can generally detect only monotonic trends. Koen and Lombard (2004) report the results of a frequency domain analysis of stars from the American Association of Variable Star Observers (AAVSO) data base 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 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 false discovery rate. An intriguing pattern 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 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 an appendix, Section 6.

2 Model Used for Each Star

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

$$Y_j = \mu(x_j) + I_j + \epsilon_j - \epsilon_{j-1}, \quad j = 1, \dots, n, \quad (1)$$

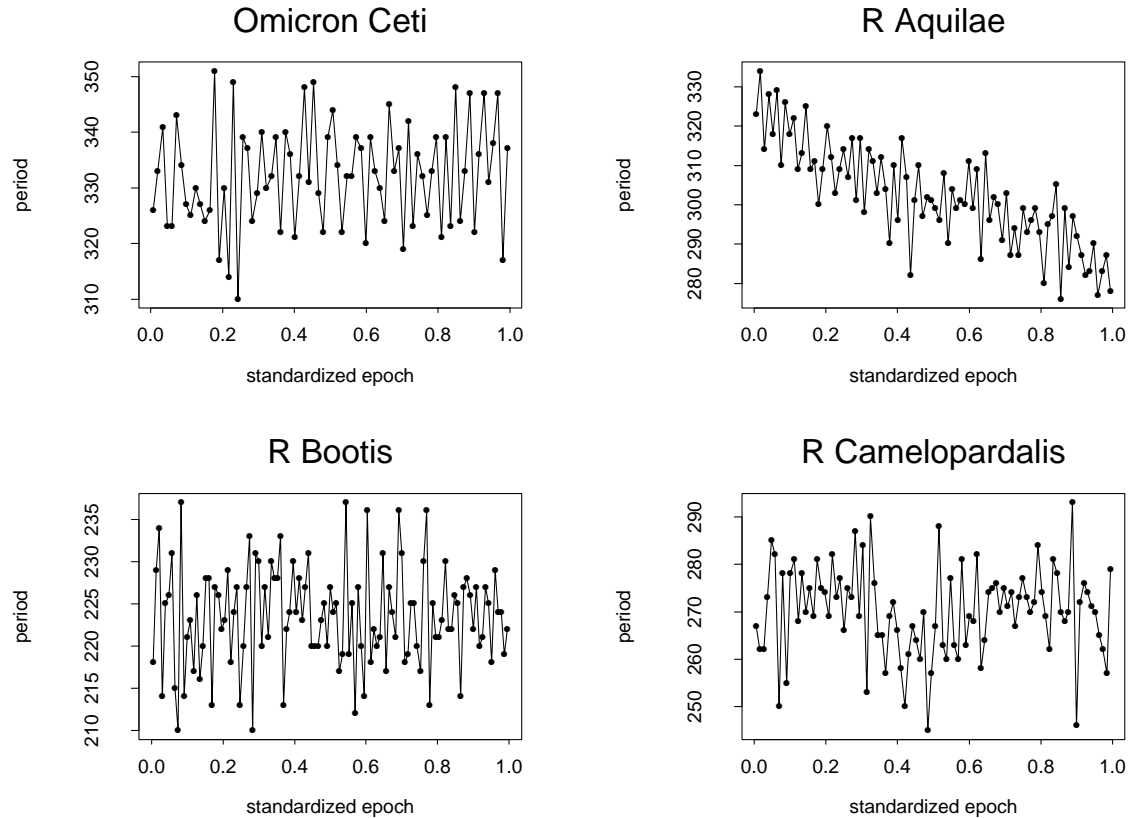


Figure 1: Typical plots of pulsation periods.

where μ is a function (defined on $[0, 1]$) that accounts for trend, I_1, \dots, I_n are random deviations intrinsic to the star, and $\epsilon_0, \dots, \epsilon_n$ are the errors 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.

Blurb here about history of this model in the literature.

Plots of pulsation periods for the long-period variables Omicron Ceti, R Aquilae, R Bootis and R Camelopardalis are shown in Figure 1. An important aspect of each data set is the negative autocorrelation between observations one lag apart, as evidenced by the high frequency behavior 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 $\boldsymbol{\eta}$, parameters of its error process. Each μ is allowed to be arbitrary and is estimated nonparametrically, as discussed in Section 3. In Section 3.2, two assumptions are made about the distribution of $(\mu, \boldsymbol{\eta})$, but these still allow for certain dependencies amongst the parameters that are dictated by the nature of the data (see Appendix). Concerning the conditional distribution of errors, given any particular star, we make the following assumptions:

- A1. $I_1, \dots, I_n, \epsilon_0, \dots, \epsilon_n$ are mutually independent.
- A2. $I_j \sim N(0, \sigma_I^2)$, $j = 1, \dots, n$, where $\sigma_I^2 < \infty$.
- A3. $\epsilon_j \sim N(0, \exp(\beta_0 + \beta_1 x_j))$, $j = 0, \dots, n$, where $|\beta_i| < \infty$, $i = 1, 2$.

Some fairly extensive preliminary analysis of the data was done to determine the reasonableness of A1-A3. Independence of I_1, \dots, 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 A3, 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 $\text{Var}(\epsilon_j)$ has decreased.

3 Method of Testing

Our goal is to test the null hypothesis

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

against the negation of H_0 for each of the 378 stars in our data base. The semiparametric model introduced in Section 2 is assumed henceforth. The nonparametric 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 nonparametric estimator of μ_0 . Severini and Wong (1992) argue 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*, or GPLR, 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, \dots, Y_n , and our test statistic will be

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

Our choice for the nonparametric smooth $\hat{\mu}$ is a local linear estimate (Cleveland and Devlin 1988 and Fan 1992) applied to the regression data (x_j, Y_j) , $j = 1, \dots, n$.

Having specified the nature of the test statistic, we are faced with two important problems: (i) How will the smoothing parameter of the local linear estimate be chosen, and (ii) how will the distribution of the test statistic be determined?

3.1 Choice of smoothing parameter

Choice of smoothing parameter is always a crucial issue when applying a nonparametric function estimator. In our analysis a version of *one-sided cross-validation*, or 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. 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. Similar results have been established by Zhao (2003) for 1-dependent data when comparing OSCV1 and CV1, a modification of ordinary CV for 1-dependence. Henceforth, S denotes (2) in which $\hat{\mu}$ is a local linear estimate with smoothing parameter chosen by OSCV1.

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 cross-validation, the most often used form of which selects h to minimize

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

where $\hat{\mu}_i(\cdot; h)$ is a local linear estimate computed from all the observations except Y_i , $i = 1, \dots, n$. This version of cross-validation 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 (3) will tend to be too *large* (Hart 1994). To adapt cross-validation to our model, we make use of the fact that data from (1) are 1-dependent, meaning that observations more than one lag apart are independent. We thus define the curve $CV_1(h)$ precisely as in (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 $CV_1(h)$ 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, or MASE.

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

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

where $\tilde{\mu}_i(\cdot; h)$ is a local linear estimator based on the observations Y_1, \dots, Y_{i-1} . The minimizer of $OSCV$ 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. 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 is provided by Hart and Lee (2005).

The OSCV curve for 1-dependent data is defined exactly as in (4) but with $\tilde{\mu}_i(\cdot; h)$ computed from Y_1, \dots, 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 OSCV1 in carrying out our tests of the no-trend hypothesis. Simulation is used to justify that tests based on OSCV1 are more powerful than ones using CV1.

3.2 Bootstrapping the test statistic

Approximating the distribution of S assuming H_0 to be true is a non-standard exercise for at least two reasons. First of all, the limit distribution (as $n \rightarrow \infty$) of S is nonstandard 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 upon 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.

In order to define our reference distributions we first need to define a few objects. Write $\rho = \sigma_I / \exp(\beta_0/2)$, and let $\boldsymbol{\eta}' = (\rho, \beta_1)$. The 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 (1) with μ identical to a constant, assumptions A1-A3 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

$$F_n(t) = \int_H G_n(t | \boldsymbol{\eta}') dD(\boldsymbol{\eta}').$$

To properly interpret F_n , we introduce the following assumption:

B1. Within the subpopulation \mathcal{T}^c , $\boldsymbol{\eta}'$ and sample size are independent.

Now, if **B1** 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 mentioned before that the conditional distribution of S might well depend on $\boldsymbol{\eta}'$. Our approach in this regard is analogous to one proposed by Bayarri and Berger (2000) for dealing with composite hypotheses. When testing a single composite null hypothesis, they propose a reference distribution of the same form as our F_n . A key distinction between their formulation and ours is that their D is a prior distribution that must be specified by the data analyst. Having to do so usually involves a certain amount of arbitrariness, and hence could be regarded as a weakness of the Bayarri and Berger (2000) method. In our setting, however, D has objective reality since it is the distribution of $\boldsymbol{\eta}'$ over the subpopulation \mathcal{T}^c and can be estimated from the data by the empirical distribution of estimates of $\boldsymbol{\eta}'$.

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

B2. The distribution of $\boldsymbol{\eta}'$ is the same within the two subpopulations \mathcal{T} and \mathcal{T}^c .

Under assumptions **B1** and **B2**, we propose the following bootstrap algorithm for estimating F_n :

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

- (ii) Generate i.i.d. standard normal variates Z_1, \dots, Z_{2n+1} , let $\epsilon_i^* = \exp[\hat{\beta}_1 x_i/2] Z_{n+i+1}$, $i = 0, \dots, n$, and define

$$Y_i^* = \hat{\rho} Z_i + \epsilon_i^* - \epsilon_{i-1}^*, \quad i = 1, \dots, n,$$

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

- (iii) Compute S^* from Y_1^*, \dots, Y_n^* in exactly the same way that the statistic S is computed from observations Y_1, \dots, Y_n .
- (iv) Repeat steps (i)-(iii) B times, resulting in bootstrap statistics S_1^*, \dots, S_B^* .
- (v) 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 Section 6, where we also discuss how the algorithm could be modified to avoid assumptions **B1** and **B2**. This modification involves estimating the joint distribution of n , “size” of trend and $\boldsymbol{\eta}'$. Obviously, this distribution cannot be estimated as efficiently as can the marginal distribution of $\boldsymbol{\eta}'$, and hence, when it is justified, the above algorithm will be preferable. In Section 6.2 we provide some evidence that assumptions **B1** and **B2** are at least approximately valid for the data under consideration.

4 Data Analysis

The procedure described in the previous section was carried out for the 378 stars in our data base. 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, \dots, 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 $2, 3, \dots, 31$. Define $\mathbf{W}_n(h)$ to be the $n \times n$ matrix having element $w_{j,n}(x_i; h)$ in the i th row and j th column, $i = 1, \dots, n$, $j = 1, \dots, n$. This is the so-called hat matrix, i.e., the matrix which when multiplied by the data vector $(Y_1, \dots, Y_n)^T$ produces the predicted values $\hat{\mu}(x_1; h), \dots, \hat{\mu}(x_n; h)$. Hastie and Tibshirani (1990, pp. 52-55) argue that $\text{trace}(\mathbf{W}_n(h))$ provides a good proxy for the degrees of freedom of a smooth.

We chose our 30 bandwidths h_1, \dots, h_{30} in such a way that at $n = 74$, the median sample size of all stars, $\text{trace}(\mathbf{W}_n(h_i)) = i + 1$, $i = 1, \dots, 30$. The quantity $\text{trace}(\mathbf{W}_n(h))$ does vary with n , but for the range of sample sizes in our data base, all 30 traces are in remarkable agreement, as shown in Figure 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 $h \rightarrow \infty$ the local linear estimate tends to the least squares straight line.

Figure 3 shows the distribution of $\text{trace}(\mathbf{W}_n(\hat{h}))$ over all 378 stars, where \hat{h} is the OSCV1 bandwidth. The distribution of \hat{h} assuming that H_0 is true is also shown for the sake of com-

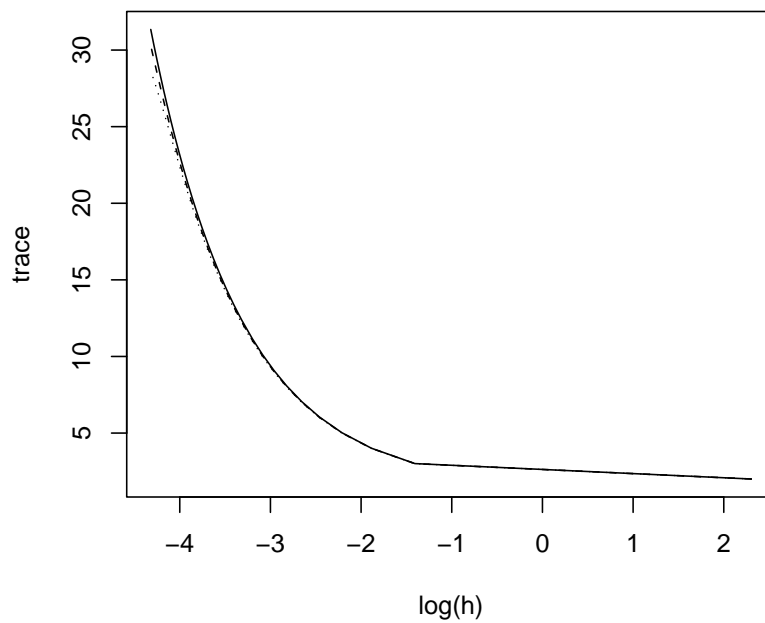


Figure 2: *Traces of hat matrix.* The dotted, dashed and solid lines correspond to $n = 32$, 40 and 212, respectively.

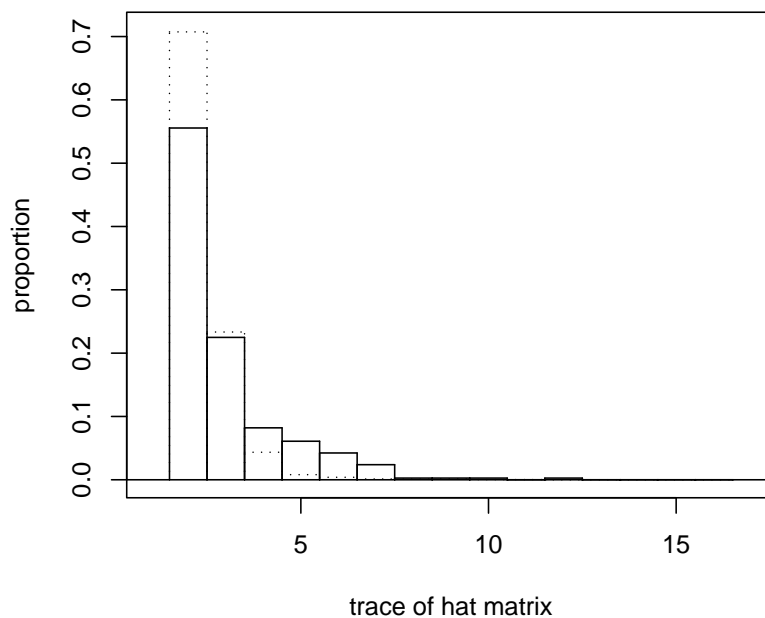


Figure 3: *OSCV1 bandwidth distributions.* The solid and dashed lines are the distribution of $\text{trace}(\mathbf{W}_n(\hat{h}))$ over all 378 stars and under the null hypothesis of no trend, respectively.

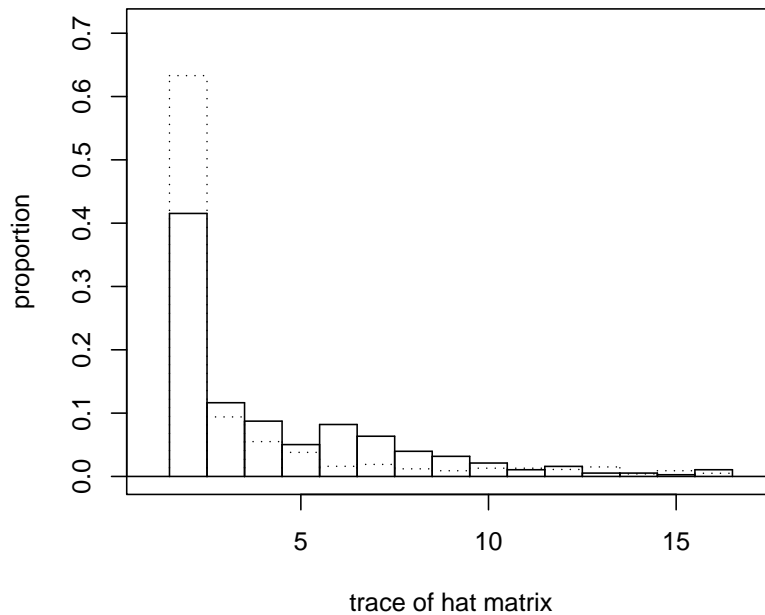


Figure 4: *CV1 bandwidth distributions.* The solid and dashed lines are the distribution of $\text{trace}(\mathbf{W}_n(\hat{h}))$ over all 378 stars and under the null hypothesis of no trend, respectively, where \hat{h} is the CV1 bandwidth.

parison. The latter distribution is a weighted average of bootstrap distributions corresponding to different sample sizes between 32 and 212. The bootstrap scheme used was precisely the one 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 Figure 3 is how often the largest bandwidth (i.e., $\text{trace}(\mathbf{W}_n(h)) = 2$) is chosen. Under H_0 , this bandwidth is chosen about 71% of the time, while only about 56% of the stars had $\text{trace}(\mathbf{W}_n(\hat{h})) = 2$. This in itself is evidence that some nontrivial proportion of stars have 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 Figure 4. Under 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, while the same quantities for the CV1 distribution are 4.38 and 5.04. A simulation study described in the Appendix shows that no-trend tests based on CV1 bandwidths have extremely poor power in comparison to ones based on OSCV1 bandwidths. The long right-hand tail of the null CV1 distribution explains this phenomenon.

Our use of a heteroscedastic 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 less 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, while

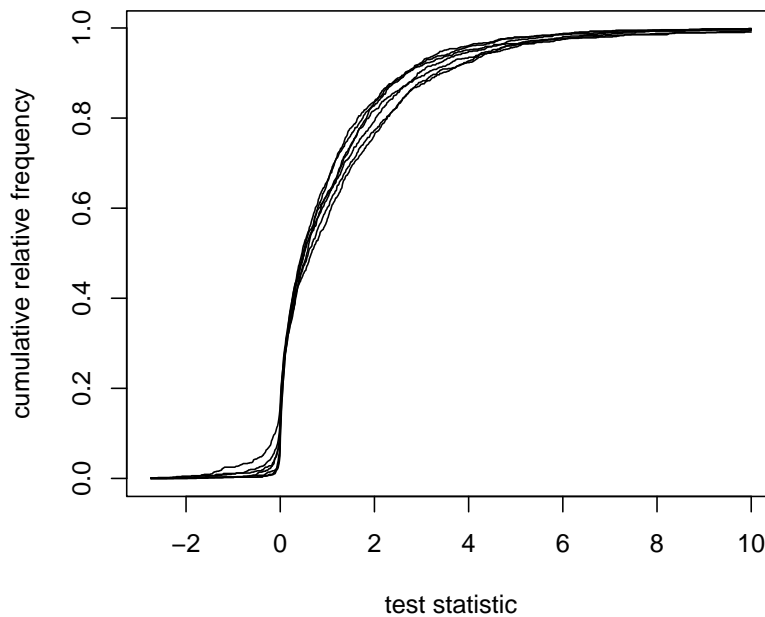


Figure 5: *Bootstrap reference distributions.* The different distributions correspond to sample sizes 32, 50, 68, 86, 128, 170 and 212.

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 sample sizes of $n = 32, 50, 68, 86, 128, 170$ and 212. The number of bootstrap samples at each n was 1000, and the resulting reference distributions are shown in Figure 5. The main point of showing this figure 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 one would expect. For example, the 95th quantiles are 4.59, 4.69, 4.20, 4.49, 3.89, 3.46 and 3.71 for $n = 32, 50, 68, 86, 128, 170$ and 212, respectively.

Let \hat{F}_n be the bootstrap distribution of statistic S for sample size n , and let $n_1 < \dots < n_7$ be the seven sample sizes in the simulation. For a star with sample size n and $S = s$, let n_j be such that $n_j \leq n < n_{j+1}$. 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 Figure 6. The preponderance of values near 0 is evidence that there are significant trends, since otherwise the p -values would be approximately uniformly distributed.

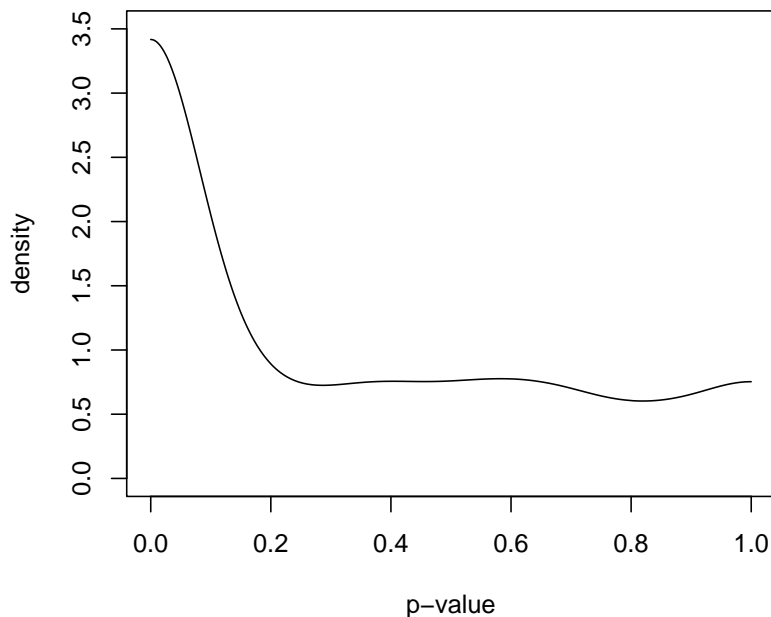


Figure 6: *Estimated density of p-values.*

Indeed, 101 stars, or 27%, have p -values smaller than 0.05 and 56, or 15%, have values smaller than 0.01.

4.2 Accounting for a multiplicity of tests

When testing multiple hypotheses, one should address the problem of 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 *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 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) defines the *positive false discovery rate*, or pFDR, to be $\text{pFDR} = E(V/R | R > 0)$. Suppose that P_1, \dots, P_n are i.i.d. p -values corresponding to independent tests such that null hypothesis H_{0i} is rejected when $P_i \leq \gamma$, $i = 1, \dots, n$. In this situation Storey (2001) argues that

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

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

$$\text{pFDR}(\gamma) = \frac{\pi_0 \gamma}{P(P_i \leq \gamma)}. \quad (5)$$

If π_0 and the cdf of P_i were known, (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) proposes a scheme for estimating π_0 . We use a slightly different method for estimating π_0 based on density estimation.

We have

$$F(\gamma) \equiv P(P_i \leq \gamma) = \pi_0 \gamma + (1 - \pi_0)G(\gamma),$$

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 $f(\gamma) \geq \pi_0$, and so for an *any* γ , replacing π_0 by $\min(1, \hat{f}(\gamma))$ in (5) would be a conservative procedure for estimating $\text{pFDR}(\gamma)$. 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. Figure 6 suggests that in our case $f(\gamma)$ is approximately constant for $\gamma \geq 0.3$. These observations lead us estimate π_0 by an average of kernel density estimates at $\gamma \geq 0.6$. 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, 0.740, respectively.

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

$$\widehat{\text{pFDR}}(\gamma) = \frac{0.75\gamma}{\hat{F}(\gamma)}.$$

A plot of this estimate for $0 \leq \gamma \leq 0.2$ is shown in Figure 7. Also indicated on this plot is the value of γ associated with a pFDR of 0.05. This value is about 0.00987, meaning that we could reject H_0 for any p -value smaller than 0.00987 with the assurance that there are only about 5% false positives. 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 FDR at the 5% level. When using an error rate of 1% the methods of Storey and Benjamini-Hochberg yield 27 and 26 significant trends, respectively. All stars with p -values smaller than 0.05 are given in Table ??, which also indicates the stars that were significant at pFDR levels of 5% and/or 1%.

4.3 Nature of trends

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

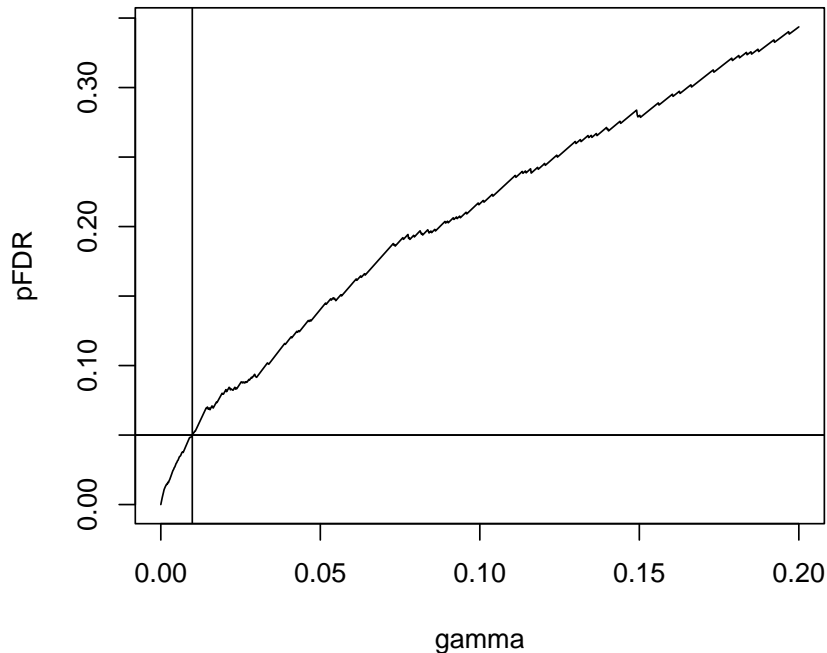


Figure 7: *Estimated positive false discovery rate as a function of critical value γ .*

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

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

where $a + bx$ is the least squares line and $\hat{Y}_1, \dots, \hat{Y}_n$ 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. Figure 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 toward its previous level.

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 versus average observed period length. 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

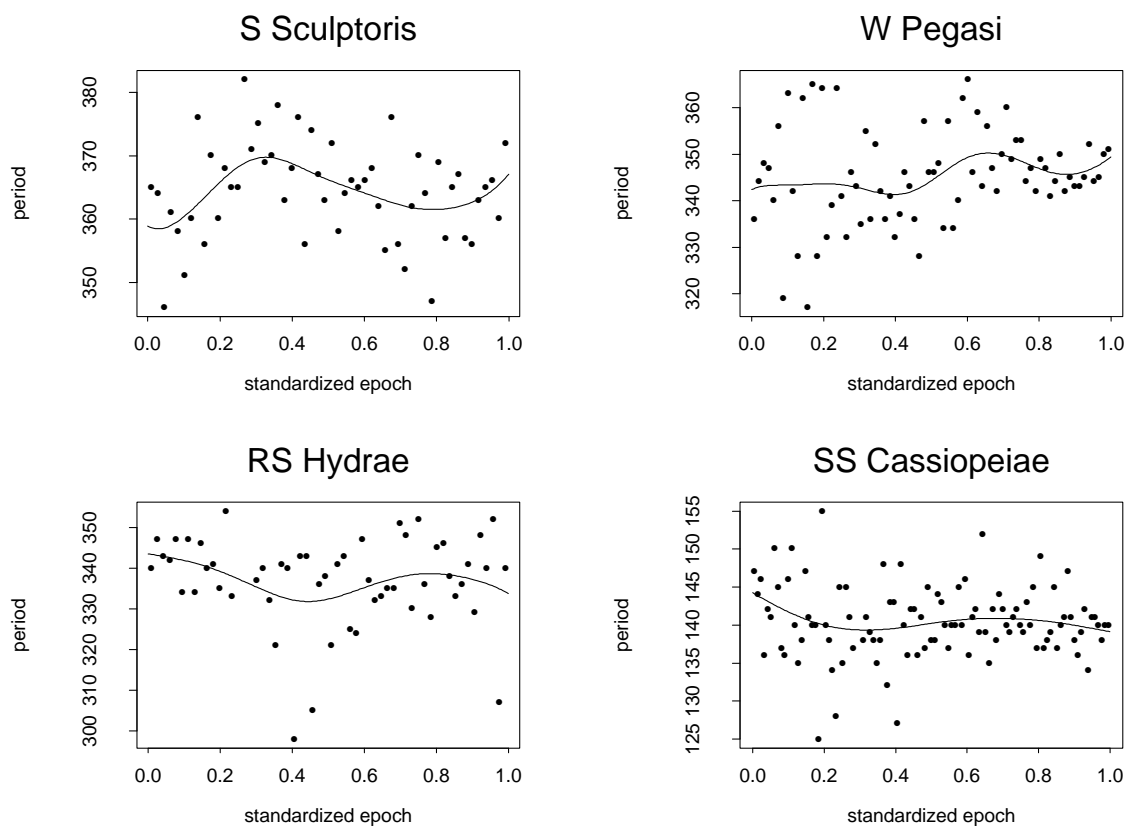


Figure 8: Scatterplots for four variable stars and OSCV1 local linear smooths.

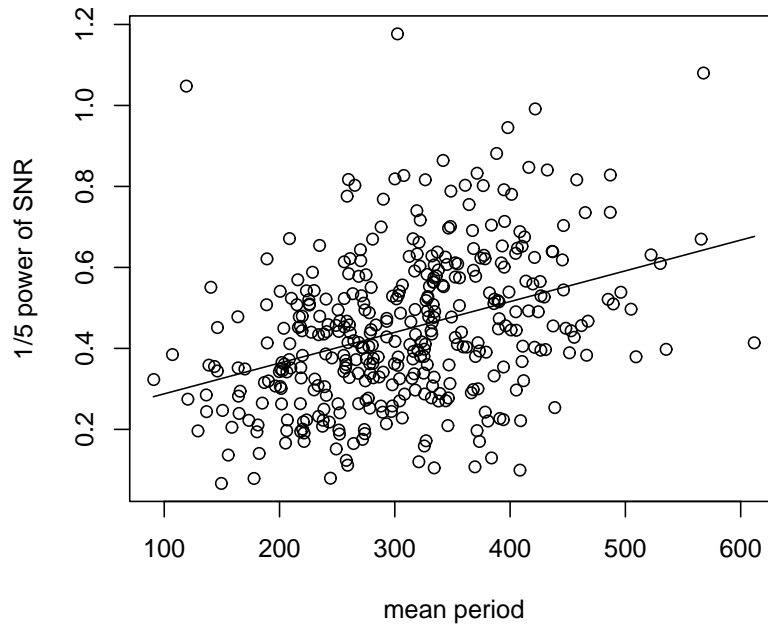


Figure 9: *Power transformed signal-to-noise ratios vs. period means.*

resulting observed significance level is $2 \cdot 10^{-6}$. Of course, the long-period stars are those with the smaller sample sizes, because fewer pulsation cycles have elapsed since initiation of the observations. Hence, the relationship between strength of trend and mean period is somewhat understated in the analysis just described. A better idea of this relationship may be gotten from considering the signal-to-noise ratio statistic

$$SNR = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2},$$

where $\hat{Y}_1, \dots, \hat{Y}_n$ are values of the OSCV1 local linear smooth. Figure 9 is a plot of $SNR^{1/5}$ vs. mean period length. Here there is a very clear indication of the relationship between trend strength and mean period length.

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. Main conclusions of our analysis are as follows:

- We provide evidence that the observed pulsation periods of most of our long period variables are heteroscedastic.
- 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.

- Most of the trends that are statistically significant have little upward or downward tilt, but rather a wavelike behaviour.
- There is a clear tendency for 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 cross-validation rather than ordinary cross-validation.

6 Appendix

Here we provide supplementary material justifying certain claims made in the paper, and also describe an extension of our bootstrap algorithm that avoids assumptions **B1** and **B2**.

6.1 Bootstrap justification

Suppose that assumptions A1-A3 and **B1-B2** hold, a random sample of N stars is obtained, and estimates $\hat{\boldsymbol{\eta}}_1, \dots, \hat{\boldsymbol{\eta}}_N$ of error parameters are computed for these stars. Let $J(y_1, \dots, y_n | \boldsymbol{\eta}')$ denote the joint distribution of $Y_j = I_j + \epsilon_j - \epsilon_{j-1}$, $j = 1, \dots, n$, when $\sigma_I = 1$ and $(\rho, \beta_1) = \boldsymbol{\eta}'$. The bootstrap employed in Section 3.2 provides a Monte Carlo approximation to

$$\hat{F}_n(t) = \int_H \int_{\mathbb{R}^n} I_{(-\infty, t]}(s(y_1, \dots, y_n)) dJ(y_1, \dots, y_n | \boldsymbol{\eta}') d\hat{D}_N(\boldsymbol{\eta}'),$$

where $s(y_1, \dots, y_n)$ is the value of the test statistic S when the observations Y_1, \dots, Y_n take on values y_1, \dots, y_n and \hat{D}_N is the empirical distribution of $\hat{\boldsymbol{\eta}}'_1, \dots, \hat{\boldsymbol{\eta}}'_N$. 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 tends to infinity.

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 $\boldsymbol{\eta}'_1, \dots, \boldsymbol{\eta}'_N$ 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

$$E_1(t) = \int_H \int_{\mathbb{R}^n} I_{(-\infty, t]}(s(y_1, \dots, y_n)) dJ(y_1, \dots, y_n | \boldsymbol{\eta}') [dD_N(\boldsymbol{\eta}') - dD(\boldsymbol{\eta}')]$$

and

$$E_2(t) = \int_H \int_{\mathbb{R}^n} I_{(-\infty, t]}(s(y_1, \dots, y_n)) dJ(y_1, \dots, y_n | \boldsymbol{\eta}') [d\hat{D}_N(\boldsymbol{\eta}') - dD_N(\boldsymbol{\eta}')].$$

Now, $E_1(t)$ is the classical sort of bootstrap error, which will converge almost surely to 0 as $N \rightarrow \infty$ 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 $\boldsymbol{\eta}'_1, \dots, \boldsymbol{\eta}'_N$ and their estimates $\hat{\boldsymbol{\eta}}'_1, \dots, \hat{\boldsymbol{\eta}}'_N$. Let n_i , $i = 1, \dots, N$, be the numbers of observations for the N stars in our sample. Using a straightforward argument available from the authors, it can be shown that $E_2(t)$ converges to 0 in probability if the following conditions hold:

- (i) $\min_{1 \leq i \leq N} n_i \rightarrow \infty$
- (ii) The distribution D has a bounded density.
- (iii) For every N , $E(\hat{\rho}_i - \rho_i)^2 + E(\hat{\beta}_{1i} - \beta_{1i})^2 \leq C/n_i$, $i = 1, \dots, N$, where $C < \infty$.
- (iv) Define the function

$$g_{n,t}(\boldsymbol{\eta}') = \frac{\partial}{\partial \rho} \frac{\partial}{\partial \beta_1} \int_{\mathbb{R}^n} I_{(-\infty, t]}(s(y_1, \dots, y_n)) dJ(y_1, \dots, y_n | \boldsymbol{\eta}').$$

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

6.2 Assumptions B1 and B2

Assumption **B1** asserts that n and $\boldsymbol{\eta}'$ are independent. To investigate this assumption we used the estimates $\hat{\boldsymbol{\eta}}_1, \dots, \hat{\boldsymbol{\eta}}_N$ from our analysis as proxies for the true parameter values. Scatterplots of $z_1 = \log(\hat{\rho})$ vs. n and $z_2 = \text{sign}(\hat{\beta}_1) \log(|\hat{\beta}_1|)$ vs. n are shown in Figure 10. (The same two outliers are excluded from each of the four plots in Figure 10.) 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 Figure 10, there is a nontrivial 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 Figure 10. So, stars with larger values of n have shorter mean periods, and shorter random periods tend to be less variable.

Figure 11 is provided as evidence that assumption **B2** holds to a reasonable approximation. The variable SNR defined in Section 4.4 is used as a proxy for whether or not the null hypothesis is true. Indeed, SNR estimates the parameter

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

which is 0 if and only if the no-trend hypothesis is true. Each of the six graphs in Figure 11 have the same (x, y) -scales. The two graphs in the same row provide a comparison of the distribution of (z_1, z_2) under the null hypothesis (left-hand plot) and under the alternative (right-hand plot). Each left hand plot corresponds to all stars whose values of SNR are less than the indicated amount, and the right hand plot corresponds to the complementary stars.

The scatterplots in each row are quite similar, and the plots do not depend substantially on the value of SNR used as a cutoff. On this basis it appears that even if **B2** is violated, the violation is not substantial enough to materially affect our analysis.

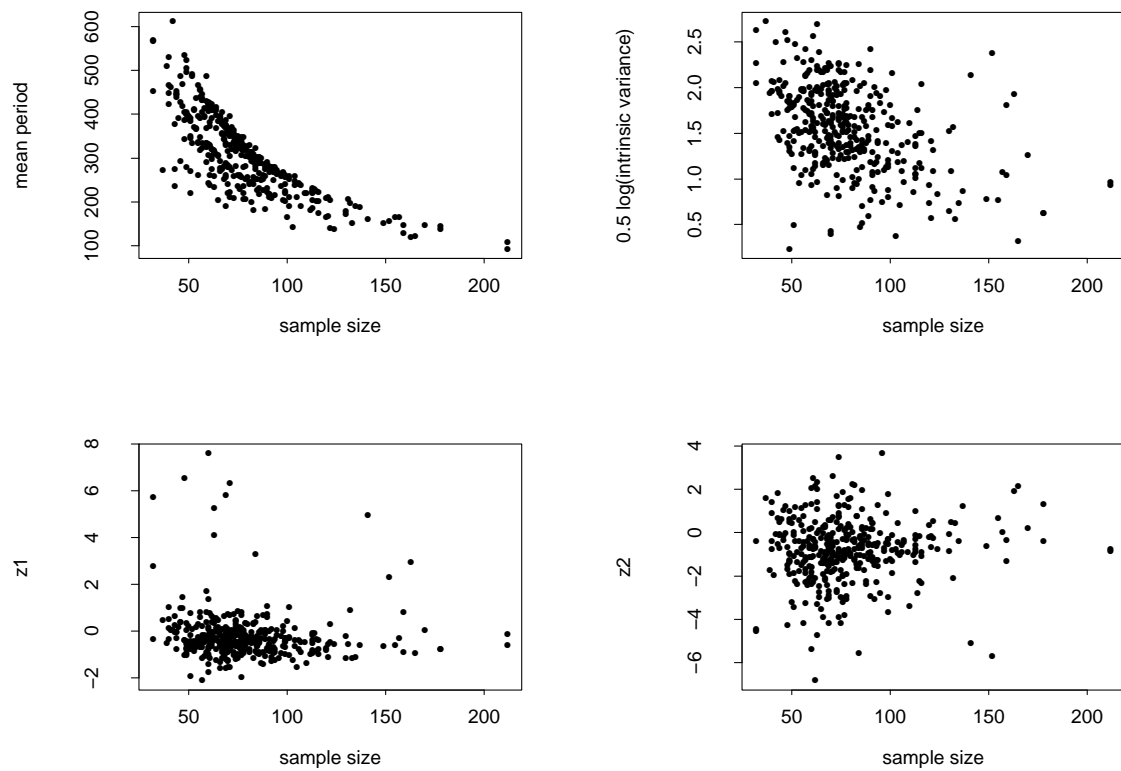


Figure 10: Scatterplots of various parameter estimates vs. sample size. The variables z_1 and z_2 are $\log(\hat{\rho})$ and $\text{sign}(\hat{\beta}_1) \log(|\hat{\beta}_1|)$, respectively.

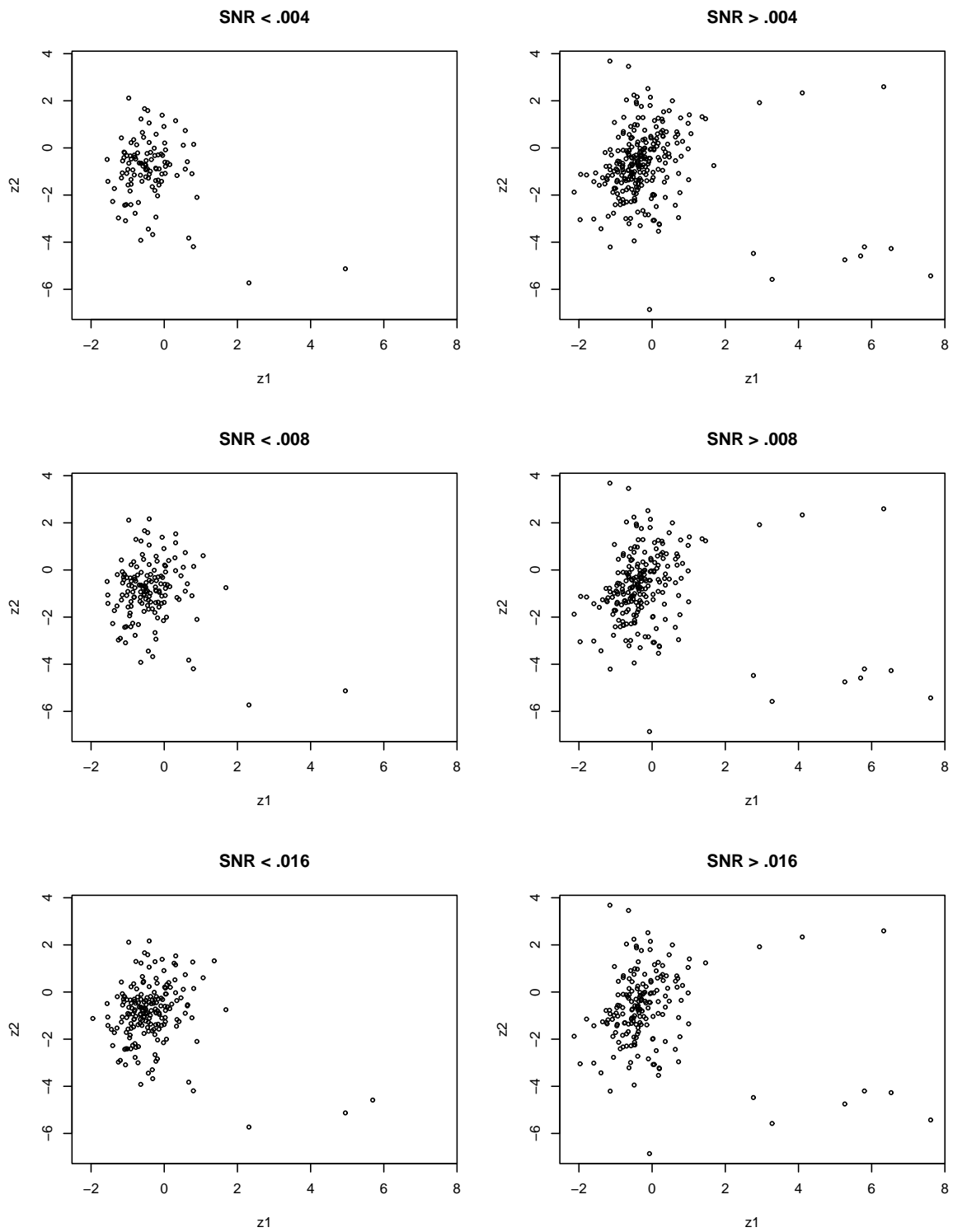


Figure 11: *Plots addressing the validity of assumption B2.* See Figure 10 for a definition of z_1 and z_2 .

6.3 Generalizing the bootstrap

Here we propose a bootstrap algorithm that avoids assumptions **B1** and **B2**. Applying the Bayarri-Berger principle, the null distribution of the test statistic S for a star with sample size n is

$$P(S \leq t | n, \xi = 0) = \int_H P(S \leq t | n, \xi = 0, \boldsymbol{\eta}') dG(\boldsymbol{\eta}' | n, \xi = 0),$$

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

Under assumptions **B1-B2**, the conditional distribution of $\boldsymbol{\eta}'$ given n and $\xi = 0$ is equal to the unconditional distribution of $\boldsymbol{\eta}'$, and hence one may simply resample from all N stars to approximate the requisite sampling distribution. More generally it is necessary to estimate $G(\cdot | n, \xi = 0)$. One may use smoothing to do so. First, the parameter ξ may be estimated by SNR . Then, one may estimate $G(\cdot | n, \xi = 0)$ by using the empirical cdf of a subset of $\{\hat{\boldsymbol{\eta}}'_1, \dots, \hat{\boldsymbol{\eta}}'_N\}$ 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 is provided by Hall, Wolff and Yao (1999). Li and Racine (2006) propose methodology for estimating conditional cdfs when some covariates are discrete, which is relevant to our case where n is discrete.

6.4 Relative power of OSCV and CV tests

Our generalized profile likelihood ratio test is remarkably more powerful when the smoothing parameter of the local linear smooth is chosen by OSCV1 as opposed to CV1. 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. One hundred and one stars had test statistics with P -values smaller than 0.05. Let $(\hat{\mu}_i, \hat{\boldsymbol{\eta}}_i)$, $i = 1, \dots, 101$, be the trend estimates and estimates of $\boldsymbol{\eta}$ for these 101 stars.

The following process was repeated independently 1000 times:

- (1) Randomly select one of the 101 stars described immediately above.
- (2) If star j was selected, generate $n = 74$ observations from model (1) under (A1)-(A3) with $\mu \equiv \hat{\mu}_j$ and $(\sigma_I, \beta_0, \beta_1) = (\hat{\sigma}_{Ij}, \hat{\beta}_{0j}, \hat{\beta}_{1j})$.
- (3) Compute the test statistic S from the 74 observations from step (2). In addition, compute S_{CV} , which is identical to S except that the bandwidth of the local linear estimate is chosen by CV1 rather than OSCV1.

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.

7 Acknowledgements

We are grateful to the American Association of Variable Star Observers for supplying the machine readable data used in this paper.

References

- Bayarri, M.J. and Berger, J.O. (2000). P values for composite null models. *Journal of the American Statistical Association* **95**, 1127-1142.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289-300.
- Birt, W.J. (1831). Observations upon the period of the variable star β Lyrae. *Monthly Notices of the Royal Astronomical Society* **1**, 192.
- Campbell, L. (1955). *Studies of Long Period Variables*. Cambridge (Mass.): AAVSO.
- Cleveland, W.S. and Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* **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.
- Eubank, R.L. and Hart, J.D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics* **20**, 1412-1425.
- Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998-1004.
- Hall, P., Wolff, R.C.L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94**, 154-163.
- Handler, G. (2004). Amplitude and frequency variability of pulsating stars. *Communications in Asteroseismology* **145**, 71-73.
- Hart, J.D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society B* **56**, 529-542.
- Hart, J.D. and Lee, C.-L. (2005). Robustness of one-sided cross-validation to autocorrelation. *Journal of Multivariate Analysis* **92**, 77-96.
- Hart, J.D. and Yi, S. (1998). One-sided cross-validation. *Journal of the American Statistical Association* **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-Verlag.
- Koen, C. and Lombard, F. (2001). The analysis of indexed astronomical time series – VII. Simultaneous use of times of maxima and minima to test for period changes in long-period variables. *Monthly Notices of the Royal Astronomical Society* **325**, 1124-1132.

- Koen, C. and Lombard, F. (2004). The analysis of indexed astronomical time series – IX. A period change test. *Monthly Notices of the Royal Astronomical Society* **353**, 98-104.
- Mattei, J.A., Mayall, M.W. and Waagen, E.O. (1990). *Maxima and Minima of Long Period Variables, 1949-1975*. Cambridge (Mass.): AAVSO.
- Percy, J.R., Colivas, T., Sloan, W.B. and Mattei, J. (1990). Long-term changes in Mira variables. *Astronomical Society of the Pacific Conference Series* **11**, 446-449: Confrontation between Stellar Pulsation and Evolution. Eds. C. Cacciari and C. Clementini.
- Percy, J.R. and Colivas, T. (1999). Long-term changes in Mira stars. I. Period fluctuations in Mira Stars. *Publications of the Astronomical Society of the Pacific* **111**, 94-97.
- Qi, L. and Racine, J.S. (2006). Nonparametric estimation of conditional cdfs and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics*, to appear.
- Severini, T. A. and Wong, W.H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics* **20**, 1768-1802.
- Storey, J.D. (2001). The positive False Discovery Rate: a Bayesian interpretation and the q -value. Unpublished manuscript. *Need to check whether this has been published*.
- Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society B* **64**, 289-300.
- Whitelock, P.A. (1999). “Real-time” evolution in Mira variables. *New Astronomy Reviews* **43**, 437-440.
- Zhao, C. (2003). One-sided cross-validation for a model motivated by variable star data. Unpublished Ph.D. dissertation, Department of Statistics, Texas A&M University.