Fitting A Semiparametric Model Based On Two Sources Of Information

Peihua Qiu
School of Statistics
University of Minnesota
313 Ford Hall
224 Church Street S.E.
Minneapolis, MN 55455

Abstract

We discuss fitting a semiparametric model based on two sources of information, which is motivated by a rat sleep dataset. In a recent study, rats were exposed to two different lighting conditions. The first, baseline condition, was a standard 24-hour schedule of 12 hours lights on followed by 12 hours lights off; the second, test condition, exposed rats to a continuous 3 hours lights on / 3 hours lights off schedule. Rat sleep was believed to be affected mainly by the circadian rhythm under the baseline lighting condition and by both the circadian rhythm and light under the test lighting condition. In this article, we suggest fitting a nonparametric model for the dataset under the baseline lighting condition. For the dataset under the test lighting condition, a two-part model is suggested. The first part equals an unknown coefficient multiplied by the nonparametric function used for modeling the dataset under the baseline lighting condition, explaining the remnant of the circadian rhythm under the test lighting condition. The second part is a periodic nonparametric function which would explain the effect of the test lighting condition. This modeling procedure can be used to model other physiological parameters affected by both intrinsic and extrinsic factors.

Key Words and Phrases: Circadian rhythm; Kernel estimation; Lighting condition; Periodic function space; Semiparametric model; Rat sleep.

1 Introduction

Animal sleep is believed to be affected by both intrinsic factors like circadian rhythm and extrinsic factors like lighting condition and other environmental conditions (Benca *et al.* 1993). Psychiatrists are interested in identifying these two kinds of effects in order to understand more about the animal sleep mechanism. In a recent study, experiments were performed on rats which were subject to two different lighting conditions. The first, "baseline" condition, was a standard 24-hour schedule of 12 hours lights on followed by 12 hours lights off; the second, "test" condition, exposed rats to a continuous 3 hours lights on / 3 hours lights off schedule. The baseline condition was used to approximate the natural lighting condition under which rat sleep was believed to be affected mainly by the circadian rhythm. The test condition was used to add an external stimulus to rat sleep. Scientists were interested in knowing how this lighting condition and the circadian rhythm affected rat sleep simultaneously. A major response variable of the study was the percentage of time that a rat was in sleep in each 5-minute interval of the 24-hour day.

The problem described above could be formulated as follows. Under the baseline condition, the model is

$$Y_i^* = f_1(x_i^*) + \varepsilon_i^*, \;\; i = 1, \cdots, n^*, \tag{1.1}$$

where f_1 is a regression function and $\{\varepsilon_i^*\}$ are random errors. Under the test condition, the model is

$$Y_i = cf_1(x_i) + f_2(x_i) + \varepsilon_i, \;\; i = 1, \cdots, n, \tag{1.2}$$

where c is an unknown parameter, f_2 is a function that differs from f_1 and $\{\varepsilon_i\}$ are random errors independent of $\{\varepsilon_i^*\}$. In (1.2), f_2 and cf_1 represent the effects of the external stimulus and the circadian rhythm, respectively. The parameter c reflects the latter effect: the bigger the value of c, the more important the effect of the circadian rhythm. Design points $\{x_i^*\}$ and $\{x_i\}$ could be different but they come from a same design space. For simplicity, we assume that the design space is $[0, 1]$ and the related design points are equally spaced in $[0, 1]$.

The sample size n^* in (1.1) is often moderately large because it is not difficult to obtain observations under the baseline condition. Therefore f_1 could be estimated well from the model (1.1) alone. After f_1 is replaced by its estimator $\widehat{f}_1$ (which is estimated from $\{Y_i^*\}$ alone) in (1.2), the model (1.2) is similar to a semiparametric model. We use the word "similar" here because a conventional semiparametric model in the literature has the form

$$Y_i = \alpha\, x_i + f(t_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.3}$$

where α is an unknown parameter, f is an unknown smooth function, x_i and t_i are related by another model

$$x_i = h(t_i) + \eta_i, \tag{1.4}$$

where h is a function of t, $\{\eta_i\}$ are uncorrelated random variables with zero mean and constant variance, and $\{\eta_i\}$ are independent of $\{\varepsilon_i\}$ (cf. e.g., Chen 1988; Cuzick 1992; Heckman 1986; Speckman 1988).

We notice that the models (1.1) and (1.2) differ from the models (1.3) and (1.4) in several aspects. First, there are two different samples in the models (1.1) and (1.2) while estimation of (1.3) is based on a single sample $\{(Y_i, x_i, t_i)\}$. Second, if f_1 is replaced by $\widehat{f}_1$ in (1.2), then $\widehat{f}_1$ can be written in the form of (1.4) as follows:

$$\widehat{f}_1(x_i) = f_1(x_i) + \delta(x_i), \tag{1.5}$$

where $\delta(x_i) = \widehat{f}_1(x_i) - f_1(x_i)$. But $\delta(x_i)$ differs from η_i in that the variance of η_i is often assumed to be a constant while the variance of $\delta(x_i)$ converges to zero when n^* tends to infinity if $\widehat{f}_1$ is a strong consistent estimator such as the kernel estimator (Härdle 1990) and the local polynomial kernel estimator (Cleveland 1979; Fan 1992). As Eubank and Whitney (1989) pointed out, many semiparametric methods in the literature (e.g., Rice 1986; Heckman 1986) depend heavily on the variance structure of (1.4). In general these methods can not be applied directly to the case in which the variance of η_i is zero. It is therefore not appropriate to apply these methods to the current problem either.

The functions f_1 and f_2 in model (1.2) may not be estimable if they belong to a same function space. For example, $\{c = c^{(1)}, f_1(x) = f_1^{(1)}(x), f_2(x) = f_2^{(1)}(x)\}$ and $\{c = c^{(1)} + a, f_1(x) =$

$f_1^{(1)}(x), f_2(x) = f_2^{(1)}(x) - af_1^{(1)}(x)\}$ determine a same distribution for the response variable for any real number a. For the rat sleep data (cf. Figure 5.1 in Section 5), we have the following observation. Rats are nocturnal animals who tend to go to sleep when exposed to light. The rat sleep measurement would have a jump at the time when the light is switched. Therefore it is reasonable to assume that f_1 is a piecewisely continuous function with a possible jump at $x = 0.5$ and f_2 is a periodic function with 4 periods in $[0, 1]$. In each period, it is continuous except at the middle position of the period where it might have a jump. Under these assumptions, it can be checked that f_1 and f_2 are estimable.

The remaining part of the article is organized as follows. In next section, we discuss the estimation of c, f_1 and f_2 in models (1.1) and (1.2). Some of their statistical properties are presented in Section 3. In Section 4, we discuss several simulation examples concerning the performance of the fitted models. Finally we apply our modeling procedure to the rat sleep dataset in Section 5.

2 Model Fitting

In model (1.1), we assume that the regression function f_1 belongs to the function space $\mathcal{F}_1 := \{g :$ g is continuous in $[0, 0.5)$ and $(0.5, 1]$, and $g_+(0.5) - g_-(0.5) \neq 0\}$. We further assume that K^* is a $n^* \times n^*$ matrix which is a strong consistent smoother in $\mathcal{F}_1$ in the sense that $\underline{\widehat{f_1}}^* := K^*\underline{Y^*}$ is a strong consistent estimator of $\underline{f_1}^*$, where $\underline{f_1}^* = (f_1(x_1^*), \ldots, f_1(x_{n^*}^*))'$, $\underline{\widehat{f_1}}^* = (\widehat{f}_1(x_1^*), \ldots, \widehat{f}_1(x_{n^*}^*))'$ and $\underline{Y^*} = (Y_1^*, \ldots, Y_{n^*}^*)'$. In model (1.2), f_1 is replaced by its estimator $\widehat{f}_1$ and f_2 is assumed in the function space $\mathcal{F}_2 := \{g : g$ is a periodic function with four periods in $[0, 1]$; g is continuous in each period except at the middle point of the period where g might have a jump$\}$. Then the Speckman-type estimators of c and $\underline{f_2} := (f_2(x_1), \ldots, f_2(x_n))'$ (cf. Speckman 1988) can be written as:

$$\widehat{c} = \frac{\underline{\widehat{f_1}}'(I - K)^2\underline{Y}}{\underline{\widehat{f_1}}'(I - K)^2\underline{\widehat{f_1}}}, \tag{2.1}$$

$$\underline{\widehat{f_2}} = K(\underline{Y} - \widehat{c}\underline{\widehat{f_1}}), \tag{2.2}$$

where K is a $n \times n$ matrix which is assumed to be a strong consistent smoother in $\mathcal{F}_2$, $\underline{\widehat{f_1}} := (\widehat{f}_1(x_1), \ldots, \widehat{f}_1(x_n))'$, $\underline{\widehat{f_2}} := (\widehat{f}_2(x_1), \ldots, \widehat{f}_2(x_n))'$, and $\underline{Y} = (Y_1, \ldots, Y_n)'$.

Different smoothing methods could be used here to estimate the related nonparametric functions. For simplicity of presentation, we use the kernel smoothing method in this paper. Let $ker^*(x)$ be a density kernel function (a non-negative function with unit integration) with support $[-1/2, 1/2]$. Then K^* for $\mathcal{F}_1$ could be defined by:

$$K^* = \begin{pmatrix} K^{*(11)} & 0 \\ 0 & K^{*(22)} \end{pmatrix}, \tag{2.3}$$

where $K^{*(11)}$ and $K^{*(22)}$ are kernel smoothers of the observations in $[0, 0.5)$ and $[0.5, 1]$, respectively. By using the Nadaraya-Watson estimator (cf. Härdle 1990), the (i,j)th element of $K^{*(11)}$ is defined by

$$K^{*(11)}(i,j) := \frac{ker^*((x_j - x_i)/h^*)}{\sum_{\ell=1}^{n^*/2} ker^*((x_\ell - x_i)/h^*)}, \quad \text{for } 0 \leq x_i, x_j < 0.5, \tag{2.4}$$

where h^* is a bandwidth parameter. The matrix $K^{*(22)}$ could be similarly defined. We should point out that some quantities used in this paper such as $K^*, h^*, \underline{\widehat{f_1}}$ and $\underline{\widehat{f_2}}$ depend on the sample sizes. For simplicity, we did not make this explicit in notation.

The matrix K could be similarly defined. When the design points are equally spaced in the design space, which is true in the rat sleep example, a simple way to estimate a regression function in the function space $\mathcal{F}_2$ is to merge the observations in different periods $[0, 0.25)$, $[0.25, 0.5)$, $[0.5, 0.75)$ and $[0.75, 1.0]$ first as if they were located in a single period $[0, 0.25)$. Then a kernel smoother can be applied to the merged data. This is equivalent to choosing the matrix K to be

$$K = \frac{1}{4} \begin{pmatrix} K^{(p)} & K^{(p)} & K^{(p)} & K^{(p)} \\ K^{(p)} & K^{(p)} & K^{(p)} & K^{(p)} \\ K^{(p)} & K^{(p)} & K^{(p)} & K^{(p)} \\ K^{(p)} & K^{(p)} & K^{(p)} & K^{(p)} \end{pmatrix}, \tag{2.5}$$

where $K^{(p)}$ is a kernel smoother for observations in $[0, 0.25)$ and it can be defined similarly to (2.3)-(2.4) by a bandwidth h and a density kernel function $ker(x)$ with support $[-1/2, 1/2]$. That is, $K^{(p)}$ could be defined by

$$K^{(p)} = \begin{pmatrix} K^{(p11)} & 0 \\ 0 & K^{(p22)} \end{pmatrix}, \tag{2.6}$$

where the (i,j)th element of $K^{(p11)}$ is defined by

$$K^{(p11)}(i,j) := \frac{ker((x_j - x_i)/h)}{\sum_{\ell=1}^{n/2} ker((x_\ell - x_i)/h)}, \quad \text{for } 0 \leq x_i, x_j < 0.125, \tag{2.7}$$

and the matrix $K^{(p22)}$ can be similarly defined for design points in $[0.125, 0.25)$.

In reality, the bandwidths h^* and h need to be specified. There are several bandwidth selectors in the literature (cf. Chu and Marron 1991). In our numerical examples, the cross-validation procedure is used to determine the bandwidths. That is, h^* is determined by minimizing the cross-validation score $CV(h^*)$ which is defined by:

$$CV(h^*) = \frac{1}{n^*} \sum_{i=1}^{n^*} \left(\widehat{f}_{1i}(x_i^*) - Y_i^* \right)^2, \tag{2.8}$$

where $\widehat{f}_{1i}(x)$ is the "leave-1-out" estimator of $f_1(x)$, namely, the observation (x_i^*, Y_i^*) is left out in constructing $\widehat{f}_{1i}(x)$, for $i = 1, \cdots, n^*$ (cf. Wu and Chu 1993). The bandwidth h could be determined similarly.

At the end of this section, we would like to point out that the method presented here is not restricted to $\mathcal{F}_1$ and $\mathcal{F}_2$. As long as the two function spaces differ in that the strong consistent local smoother chosen for the second function space yields an inconsistent estimator in the first one, this method can be applied, which is further explained by the auguments in Section 3.

3 Properties of the Fitted Models

We now discuss the consistency of $\widehat{c}$ and $\underline{\widehat{f}_2}$. Speckman (1988) discussed the conditional L^2 consistency of the fitted model of (1.3) conditional on $\{x_i\}$ in (1.4). In our case, there are two different samples. It might be more appropriate to discuss the unconditional consistency with respect to both samples. In (2.1), $\underline{\widehat{f}_1}$ appears in both the numerator and the denominator. It may be difficult to discuss the L^2 consistency directly. Our strategy for overcoming this difficulty is to discuss the almost sure consistency first and then to relate the L^2 consistency to the almost sure consistency by using the Dominated Convergence Theorem.

Theorem 3.1 In models (1.1) and (1.2), suppose that f_1 and f_2 belong to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively; they are Lipschitz (α) continuous in the intervals of continuity for some $0 < \alpha \leq 1$; $E|\varepsilon_i^*|^p < \infty$, $E|\varepsilon_i|^p < \infty$ for some $p \geq 2$. We further assume that the kernel functions ker^* and ker are both Lipschitz (β) continuous for some $0 < \beta \leq 1$; there exists a positive constant ν and a sequence $\{\beta_{n^*}\}$ satisfying $\lim_{n^* \to \infty} \beta_{n^*} = \infty$ such that the window width h^* satisfies the following conditions: (i) $(n^*)^\nu (\beta_{n^*} \log n^*)^{-1} \{(h^*)^\alpha + [(n^*)^\lambda h^*]^{-1} + [(n^*)^\beta (h^*)^{\beta+1}]^{-1}\} = o(1)$, (ii) $(n^*)^{2\nu} (n^* \beta_{n^*} h^*)^{-1} = O(1)$, and (iii) $(n^*)^{\nu+1/\rho-1} (h^* \beta_{n^*} \log n^*)^{-1} = o(1)$; and the window width h satisfies $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh = \infty$. Then for the estimators defined by (2.1)-(2.7), we have

$$\widehat{c} - c = O\{(n^*)^{-\nu} \log n^*\} + O(h^\alpha) + o(n^{-1/2} \log n \log\log n), \quad a.s.$$

and

$$\max_{1\leq i\leq n} |\widehat{f}_2(x_i) - f_2(x_i)| = O(c - \widehat{c}) + o\{(nh)^{-1/2} \log(nh) \log\log(nh)\}, \quad a.s.$$

The proof of Theorem 3.1 is given in Appendix A. In the theorem, if $\alpha = \beta = 1$, $\nu = 1/3$, $h = n^{-1/2}$ and $n^* = O(n^{3/2})$, then $\widehat{c} - c = o(n^{-1/2} \log n \log\log n)$, $a.s.$ This agrees with the strong convergence rate of the least squares regression (see e.g., Lai and Robbins 1977). If $h = O(n^{-1/3})$ and $n^* = O(n)$, then $\max_{1\leq i\leq n} |\widehat{f}_2(x_i) - f_2(x_i)| = o(n^{-1/3} \log n \log\log n)$, $a.s.$ This is the strong convergence rate of the Nadaraya-Watson kernel estimator (cf. e.g., Cheng and Lin 1981).

From (A.3) in Appendix A, we know that

$$E(\widehat{c}|\underline{Y^*}) = c + \frac{\underline{\widehat{f}_1}'(I-K)^2(\underline{f_1} - \underline{\widehat{f}_1})}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}} c + \frac{\underline{\widehat{f}_1}'(I-K)^2\underline{f_2}}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}} = c + O\{(n^*)^{-\nu} \log n^*\} + O(h^\alpha)$$

and

$$Var(\widehat{c}|\underline{Y^*}) = \frac{\underline{\widehat{f}_1}'(I-K)^4\underline{\widehat{f}_1}}{\{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}\}^2} \sigma^2 = O(n^{-1}),$$

where σ^2 is the variance of ε_1 in (1.2). It is not difficult to check that both $|E(\widehat{c}|\underline{Y^*})|$ and $|Var(\widehat{c}|\underline{Y^*})|$

are almost surely bounded. Hence by the Dominated Convergence Theorem,

$$E(\widehat{c}) = E\{E(\widehat{c}|\underline{Y^*})\} = c + O\{(n^*)^{-\nu}\log n^*\} + O(h^\alpha)$$

and

$$Var(\widehat{c}) = Var\{E(\widehat{c}|\underline{Y^*})\} + E\{Var(\widehat{c}|\underline{Y^*})\} = O\{(n^*)^{-2\nu}(\log n^*)^2\} + O(h^{2\alpha}) + O(n^{-1}).$$

Similar arguments can be applied to $\widehat{f}_2$. Therefore we have

Theorem 3.2 Under the conditions stated in Theorem 3.1, $\widehat{c}$ is L^2 consistent with rate $O\{(n^*)^{-2\nu}(\log n^*)^2\}+O(h^{2\alpha})+O(n^{-1})$; $\widehat{f}_2$ is uniformly L^2 consistent with rate $O\{(n^*)^{-2\nu}(\log n^*)^2\}+O(h^{2\alpha})+o\{n^{-1}(\log n \log\log n)^2\} + O\{(nh)^{-1}\}$.

In Theorem 3.2, if $\alpha = \beta = 1$, $\nu = 1/3$, $h = n^{-1/2}$ and $n^* = O(n^{3/2})$, then the L^2 convergence rate of $\widehat{c}$ is $O\{n^{-1}(\log n)^2\}$ which is close to the least squares convergence rate $O(n^{-1})$. If $h = O(n^{-1/3})$ and $n^* = O(n)$, then the uniform L^2 convergence rate of $\widehat{f}_2$ is $O\{n^{-2/3}(\log n)^2\}$ which is nearly the optimal convergence rate $O(n^{-2/3})$ of the nonparametric regression estimators of regression functions with continuous first order derivatives. There are two possible reasons why both rates are $(\log n)^2$ slower than the optimal rates mentioned above. One is that the regression functions are assumed to be Lipschitz (1) continuous in our case instead of being assumed to have continuous first order derivatives. The second possible reason is that the L^2 convergence rates in Theorem 3.2 are derived from the almost sure convergence rates in Theorem 3.1 and the factor $\log(n)$ is commonly seen in the almost sure convergence rates. The following theorem establishes the asymptotic normality for $\widehat{c}$. It can be proved based on (A.3)-(A.6) in Appendix A.

Theorem 3.3 Besides the conditions in Theorem 3.1, if $n^{1/2-\nu}\beta_{n^*}\log n^* = O(1)$ and $n^{1/2}h^\alpha = o(1)$, then the distribution of $n^{1/2}(\widehat{c} - c)$ converges to $N(0, \sigma^2/\tau^2)$ where $\tau^2 = \int_0^1[f_1(x) - g(f_1)(x)]^2dx$, $g(f_1)(\cdot)$ is a periodic function with four periods in $[0,1]$ and $g(f_1)(x) = [f(x) + f(x+0.25) + f(x+0.5) + f(x+0.75)]/4$ when $0 \le x < 0.25$.

4 A Simulation Study

In this section, we present some simulation results. In models (1.1) and (1.2), suppose that $f_1(x) = 3\sin(\pi x) + \sin(8\pi x)$ when x is in $[0, 0.5]$, $f_1(x) = 1.5\sin(\pi x) + \sin(8\pi x)$ when x is in $(0.5, 1]$; $f_2(x)$ is a periodic function with four periods in $[0, 1]$ and $f_2(x) = \sin(4\pi x)$ when x is in $(0, 0.25]$; $c = 0.5$; $\{\varepsilon_i^*\}$ and $\{\varepsilon_i\}$ are i.i.d. random errors with $\varepsilon_1^* \sim N(0, 0.5^2)$ and $\varepsilon_1 \sim N(0, 0.5^2)$; and $n^* = n = 404$. The true regression functions and their noisy versions are shown in Figure 4.1 by the solid curves and the small dot points.

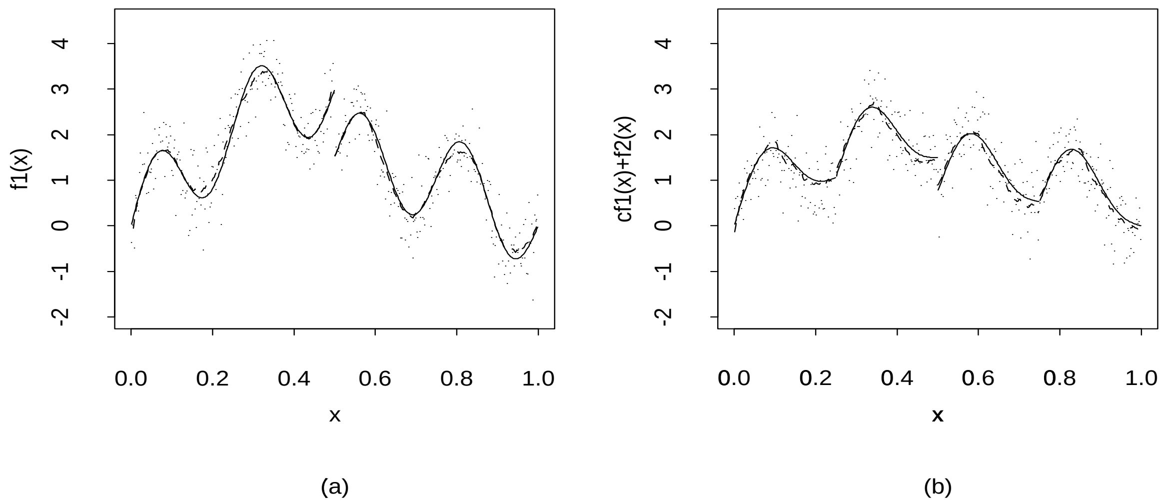

Figure 4.1: The solid curves, dashed curves and tiny dots represent the regression functions, their estimators and the noisy data, respectively. Plots (a) and (b) correspond to models (1.1) and (1.2), respectively.

We then fit the models (1.1) and (1.2) by the procedure described in Section 2 (cf. (2.1)-(2.7)). Both $ker^*(x)$ and $ker(x)$ are chosen to be the Epanechnikov kernel function $\frac{12}{11}(1 - x^2)I_{[-0.5,0.5]}(x)$ (see e.g., Härdle 1990). The bandwidths are determined by the cross-validation procedure (2.8). The estimated functions are presented in Figure 4.1 by the dashed curves. The estimated value of c is 0.5059.

The above simulation is then repeated 1000 times. The histogram of the 1000 $\widehat{c}$ values is displayed in Figure 4.2(a). It can be seen that the distribution of $\widehat{c}$ looks normal. The sample mean and the sample standard deviation of the 1000 $\widehat{c}$ values are 0.5010 and 0.0326, respectively. We then test $H_0 : c = 0.5$ *vs* $H_a : c \neq 0.5$ at significance level 0.05 based on the asymptotic normality of $\widehat{c}$ (cf. Theorem 3.3). For a given true value of c, the simulation is repeated 1000 times. In a specific simulation, H_0 is rejected if $|(\widehat{c}-0.5)/std(\widehat{c})| > z_{0.975}$ where $std(\widehat{c})$ denotes the standard deviation of $\widehat{c}$ and $z_{0.975}$ denotes the 0.975 quantile of the standard normal distribution. The relative frequency of rejection in the 1000 simulations is used to estimate the power of the test at the given c value. The solid curve in Figure 4.2(b) displays the power of the test when the true value of c varies among 0.4, 0.42, 0.44, 0.46, 0.48, 0.5, 0.52, 0.54, 0.56, 0.58 and 0.6. The corresponding power of the test when $n = n^* = 1096$ is displayed by the dashed curve. The dotted line indicates the significance level. We can see that overall the test performs quite reasonably.

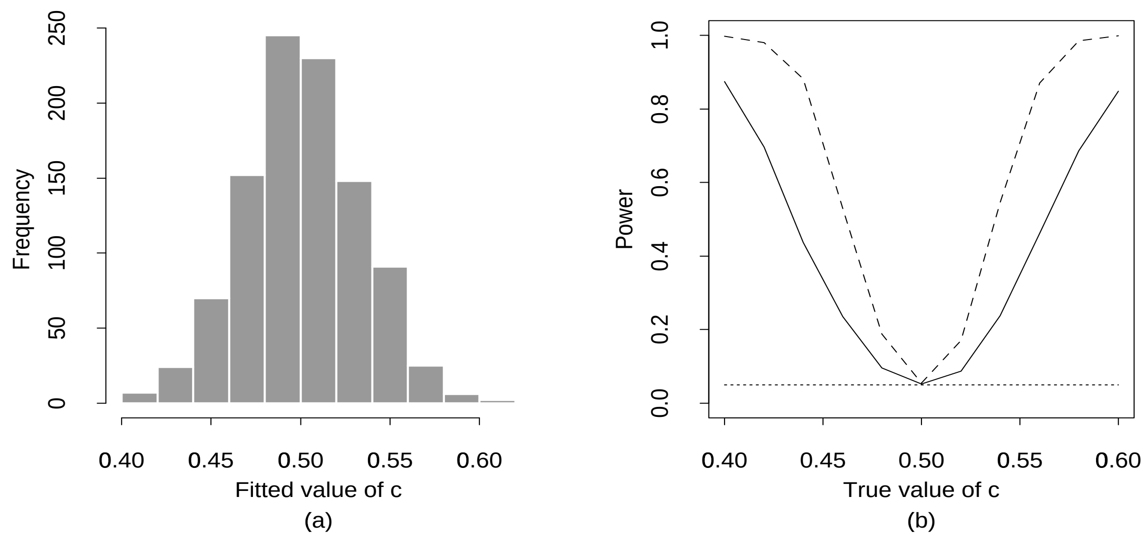

Figure 4.2: (a) The histogram of 1000 $\widehat{c}$ values when $n = n^* = 404$. (b) The solid curve denotes the power of the test when $n = n^* = 404$, the dashed curve denotes the power of the test when $n = n^* = 1096$, and the dotted line indicates the significance level 0.05.

Next we investigate the large-sample properties of the fitted models by choosing $n^* = n = 148, 404, 1096, 2980$ and 8104, respectively. Please note that $\log n$ roughly equals 5, 6, 7, 8 and 9 in these cases. For each value of n^* (=n), the averaged values of $|\widehat{c}-c|$ and $MSE(\widehat{f}_2)$ are computed from 100 replications, where $MSE(\widehat{f}_2)$ denotes the mean squared error of $\widehat{f}_2$. These values are plotted in Figure 4.3 in log-scale. When $n = 1096$, for example, $|\widehat{c}-c| = 0.0166$ and $MSE(\widehat{f}_2) = 0.0040$, which indicates that the estimators are quite accurate. We further notice that $(\log|\widehat{c}-c|, \log n)$ in plot (a) and $(\log\{MSE(\widehat{f}_2)\}, \log n)$ in plot (b) roughly follow two straight lines with slopes -0.5 and -0.8, respectively. This indicates that (1) the strong convergence rate of $\widehat{c}$

could reach the least squares rate $o(n^{-1/2}\log n \log\log n)$; and (2) the L^2 convergence rate of $\widehat{f}_2$ could be $O(n^{-4/5})$ which is the optimal rate when the regression function has continuous second-order derivative. The second conclusion suggests that the convergence rate given in Theorem 3.2 about $\widehat{f}_2$ could be improved if the regression functions are assumed to be smoother in their intervals of continuity.

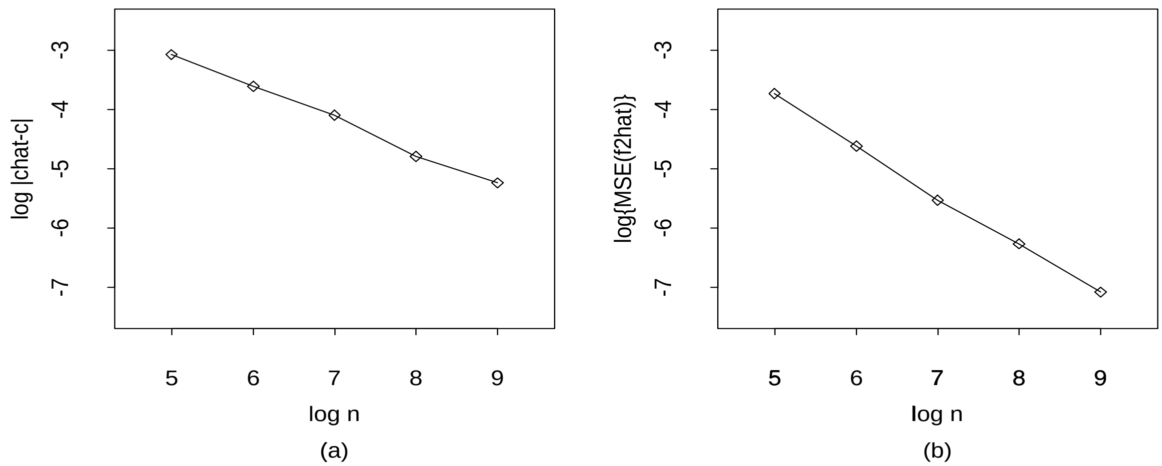

Figure 4.3: (a) $\log|\widehat{c}-c|$ vs $\log n$; (b) $\log\{MSE(\widehat{f}_2)\}$ vs $\log n$.

5 Modeling the Rat Sleep Data

We now return to the rat sleep example described in Section 1. In the psychiatric experiment mentioned there, there were 8 pigmented Brown Norway rats exposed to the baseline lighting condition and another 8 rats of the same strain exposed to the test lighting condition. The illuminance level was set at 1000 lux. The percentage of time that each rat was in sleep in each 5-minute interval of the day was recorded using standard techniques. The investigators then averaged the data over a group of rats exposed to the same lighting condition. Figure 5.1(a) shows the baseline data by the dot points. It can be seen that there is a discontinuity at time=12 o'clock when the light is switched. The data under the test lighting condition are displayed in Figure 5.1(b). We then use the modeling procedure (2.1)-(2.7) to fit these two data sets. The kernel functions are chosen to be the same as those in Section 4. The bandwidths are determined by the cross-validation procedure. The fitted function $\widehat{f}_1$ is plotted in Figure 5.1(a) by the solid curves. Figures 5.1(c) and 5.1(d) show

$\widehat{c}\widehat{f}_1$ and $\widehat{f}_2$, respectively. After combining these two parts, the fitted model of (1.2) is presented in Figure 5.1(b) by the solid curves.

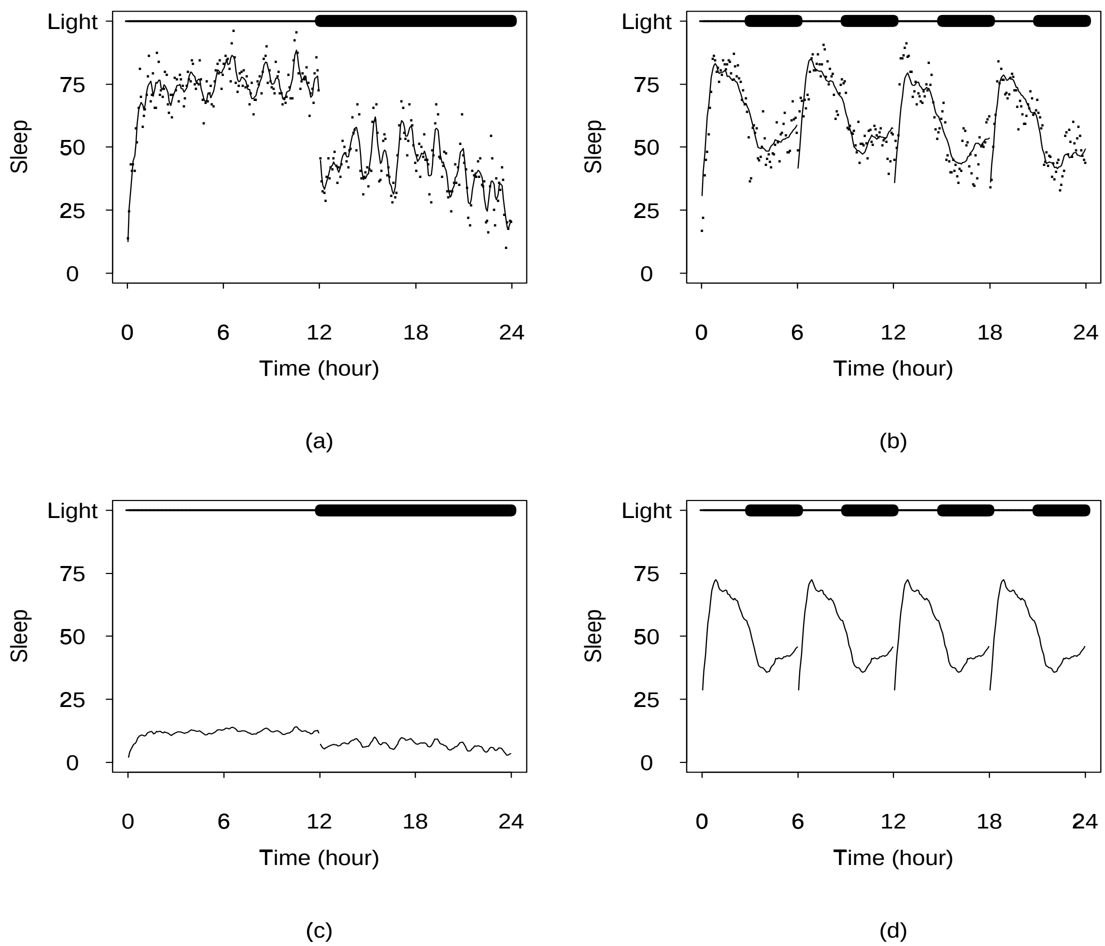

Figure 5.1: (a) Small dots represent the rat sleep data under the baseline lighting condition. "Sleep" in the vertical axis denotes the percentage of time that a rat is in sleep state in each 5-minute interval of the day. Solid curves represent $\widehat{f}_1$. (b) Small dots represent the rat sleep data under the test lighting condition. Solid curves represent the fitted model of (1.2). (c) Solid curves denote $\widehat{c}\widehat{f}_1$. (d) Solid curves denote $\widehat{f}_2$. In each plot, thin lines at top indicate periods of light. Thick lines represent darkness.

The estimated value of c is 0.1602. The values of σ^2 and τ^2 (defined in Theorem 3.3) can be estimated by the MSE of the model (1.2) and $\frac{1}{n}\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}$, respectively, where $\underline{\widehat{f}_1}$ and K are defined in Section 2. The values of these two estimators are $\widehat{\sigma}^2 = 48.6566$ and $\widehat{\tau}^2 = 333.5467$. By using a simple T-test based on Theorem 3.3, the p-value for testing $H_0 : c = 0$ vs $H_a : c \neq 0$ is less than 0.0001. Therefore we can conclude that the 24-hour cycle effect is statistically significant under the 6-hour cycle lighting condition. That is, the circadian rhythm does affect rat sleep under the test lighting condition.

It is noteworthy that the following model might be more appropriate for the test data if the data were not averaged over a group of rats:

$$Y_{ij} = cf_1(x_i) + f_2(x_i) + A_j + \varepsilon_{ij}, \quad i = 1, \cdots, n, \; j = 1, 2, \cdots, s, \tag{5.1}$$

where A_j is a random variable denoting the random effect of the jth rat, ε_{ij} is the random error, and s is the number of rats under the test lighting condition. In the case that the non-random (or the "fixed-effect") part of the model is a continuous nonparametric function, Wang (1998) suggested a procedure to fit the so-called "nonparametric mixed effect model" by using penalized maximum likelihood estimation and smoothing splines. It might be a challenging problem to fit model (5.1) because its "fixed-effect" part includes a nonparametric function and an unknown parameter c. We leave this problem for our future research.

Acknowledgements The author is grateful to a referee for many insightful suggestions. This research was supported in part by a Grant-in-Aid of Research, Artistry and Scholarship of University of Minnesota.

APPENDIX

A Proof of Theorem 3.1

Under the conditions stated in the theorem and by using the similar arguments to those in the proof of Theorem 3 of Cheng and Lin (1981), we know that

$$\frac{\tilde{n}^{\nu}}{\beta_{\tilde{n}} \log \tilde{n}} \| \tilde{K}\tilde{Y} - \underline{\tilde{f}} \| = o(1), \quad a.s. \tag{A.1}$$

and

$$\| \tilde{K}\underline{\tilde{f}} - \underline{\tilde{f}} \| = O(\tilde{h}^{\alpha}) \tag{A.2}$$

where $\tilde{n}, \tilde{K}, \tilde{Y}, \tilde{f}$ and $\tilde{h}$ denote either $(n^*, K^*, Y^*, f_1, h^*)$ or (n, K, Y, f_2, h), and $\|\underline{x}\| \triangleq \max_{1 \le i \le \tilde{n}} |x_i|$.

We can express $\widehat{c}$ as follows:

$$\begin{aligned}
\widehat{c} &= \frac{\underline{\widehat{f}_1}'(I-K)^2(c\underline{f_1}+\underline{f_2}+\underline{\varepsilon})}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}} \\
&= c+\frac{\underline{\widehat{f}_1}'(I-K)^2(\underline{f_1}-\underline{\widehat{f}_1})}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}}c+\frac{\underline{\widehat{f}_1}'(I-K)^2\underline{f_2}}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}}+\frac{\underline{\widehat{f}_1}'(I-K)^2\underline{\varepsilon}}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}} \\
&\triangleq c+T_1+T_2+T_3. \qquad (A.3)
\end{aligned}$$

Let f_1^* be a periodic function in function space $\mathcal{F}_2$ which is defined by

$$f_1^*(x)=\{f_1(x)+f_1(x+0.25)+f_1(x+0.5)+f_1(x+0.75)\}/4, \quad \text{for } x \text{ in } [0,0.25).$$

Then

$$(I-K)\underline{f_1}=O(\underline{f_1}-\underline{f_1^*}),$$

$$|\underline{\widehat{f}_1}'(I-K)^2(\underline{\widehat{f}_1}-\underline{f_1})|=O(n\|\underline{\widehat{f}_1}-\underline{f_1}\|\|\underline{f_1}-\underline{f_1^*}\|),$$

$$|(\underline{\widehat{f}_1}-\underline{f_1})(I-K)^2(\underline{\widehat{f}_1}-\underline{f_1})|=O(\|\underline{\widehat{f}_1}-\underline{f_1}\|^2)$$

and

$$\underline{f_1}'(I-K)^2\underline{f_1}=O[n\int_0^1\{f_1(x)-f_1^*(x)\}^2dx]$$

Hence

$$\begin{aligned}
&\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1} \\
=\;& \underline{f_1}'(I-K)^2\underline{f_1}+2\underline{\widehat{f}_1}'(I-K)^2(\underline{\widehat{f}_1}-\underline{f_1})+(\underline{\widehat{f}_1}-\underline{f_1})(I-K)^2(\underline{\widehat{f}_1}-\underline{f_1}) \\
=\;& O(n)
\end{aligned}$$

Consequently,

$$\begin{aligned}
|T_1| &\le |c|\left\{\frac{(\underline{\widehat{f}_1}-\underline{f_1})(I-K)^2(\underline{\widehat{f}_1}-\underline{f_1})}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}}\right\}^{1/2} \\
&= O(\|\underline{\widehat{f}_1}-\underline{f_1}\|). \qquad (A.4)
\end{aligned}$$

It is quite obvious that

$$|T_2|\le\left\{\frac{\underline{f_2}'(I-K)^2\underline{f_2}}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}}\right\}^{1/2}=\left\{\frac{(K\underline{f_2}-\underline{f_2})'(K\underline{f_2}-\underline{f_2})}{\underline{\widehat{f}_1}'(I-K)^2\underline{\widehat{f}_1}}\right\}^{1/2}=O(h^\alpha), \qquad (A.5)$$

where the result (A.2) has been used.

For T_3, we have

$$\begin{aligned} T_3 &= \frac{\underline{\widehat{f_1}}'(I-K)\underline{\varepsilon} + \underline{\widehat{f_1}}'(I-K)K\underline{\varepsilon}}{\underline{\widehat{f_1}}'(I-K)^2\underline{\widehat{f_1}}} \\ &= \frac{\underline{\widehat{f_1}}'(I-K)\underline{\varepsilon} + o(1)}{\underline{\widehat{f_1}}'(I-K)^2\underline{\widehat{f_1}}} \\ &= o[\{\underline{\widehat{f_1}}'(I-K)^2\underline{\widehat{f_1}}\}^{-1/2}\log\{\underline{\widehat{f_1}}'(I-K)^2\underline{\widehat{f_1}}\}] \\ &= o(n^{-1/2}\log n) \qquad \text{(A.6)} \end{aligned}$$

In (A.6), we have used the result that $\underline{d}'\underline{\varepsilon} = o\{(\underline{d}'\underline{d})^{1/2}\log(\underline{d}'\underline{d})\}$ which can be found in Theorem 1 of Lai and Robbins (1977). By combining (A.3)-(A.6), the conclusion about $\widehat{c}$ in the theorem is obtained.

Now

$$\begin{aligned} \underline{\widehat{f_2}} &= K(c\underline{f_1} + \underline{f_2} + \underline{\varepsilon} - \widehat{c}\underline{\widehat{f_1}}) \\ &= K\underline{f_2} + K\underline{\varepsilon} + K(c\underline{f_1} - \widehat{c}\underline{\widehat{f_1}}) \\ &= K\underline{f_2} + K\underline{\varepsilon} + Kc(\underline{f_1} - \underline{\widehat{f_1}}) + K(c-\widehat{c})\underline{\widehat{f_1}} \\ &= \underline{f_2} + (K\underline{f_2} - \underline{f_2}) + K\underline{\varepsilon} + O(\|f_1 - \widehat{f_1}\|) + O(c-\widehat{c}), \quad a.s. \qquad \text{(A.7)} \end{aligned}$$

By noticing the fact that $K\underline{\varepsilon} = o\{(nh)^{-1/2}\log(nh)\}$ and by using (A.1)-(A.2), we have the conclusion for $\underline{\widehat{f_2}}$.

REFERENCES

Benca, R., Obermeyer, W., Bergmann, B., Lendvai, N., and Gilliland, M. (1993). Failure to induce rapid eye movement sleep by dark pulses in pigmented inbred rat strains. *Physiology and Behavior* **54**, 1211-1214.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics* **16**, 136-146.

Cheng, K.F., and Lin, P.E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. verw. Gebiete* **57**, 223-233.

Chu, C.K., and Marron, J.S. (1991). Choosing a kernel regression estimator. *Statistical Science* **6**, 404-419.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing of scatterplots. *Journal of the American Statistical Association* **74**, 829-836.

Cuzick, J. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *The Annals of Statistics* **20**, 1129-1136.

Eubank, R.L., and Whitney, P. (1989). Convergence rates for estimation in certain partially linear models. *Journal of Statistical Planning and Inference* **23**, 33-43.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998-1004.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications,* Chapman & Hall: London.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Heckman, N. (1986). Spline smoothing in partially linear models. *Journal of the Royal Statistical Society - B* **48**, 244-248.

Lai, T.L., and Robbins, H. (1977). Strong consistency of least squares estimators in regression models. *Proceedings of the National Academic Sciences of USA* **74**, 2667-2669.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters* **4**, 203-208.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society - B* **50**, 413-436.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society - B* **60**, 159-174.

Wu, J.S., and Chu, C.K. (1993). Nonparametric function estimation and bandwidth selection for discontinuous regression functions. *Statistica Sinica* **3**, 557-576.