Functional Hierarchical Linear Models with Application to Analysis of Ecology Momentary Assessment Data *

Runze Li
Department of Statistics
Pennsylvania State University
University Park, PA16802-2111

June 3, 2003

Abstract

Intensively measured data were frequently collected in prevention studies using a modern data collection device. This kind of data pose many challenges for statisticians and data analysts due to its unbalanced and irregular data structure. In this report, we propose various methods to analyze an intensively collected data set in the study of ecology momentary assessment (EMA), collected by Smoking Research Group at the University of Pittsburgh. Of interest is to examine the relationship between *urge to smoke* and *affect*. We first conducted cluster analysis for only the data subset collected by random prompts, based on various scores summarized from the data subset using varying coefficient models for each subject. It was found from our analysis that there may exist 4 possible clusters among the collected samples. We further examined the data using functional linear model with all linear, quadratic and interaction terms. From our analysis, we found that overall trends for different clusters are very different, and the quadratic terms and interaction terms are not statistically significant at level 0.05. Finally, functional mixed effects models are proposed to fit the EMA data. We discovered that effects of negative affect change over time; the patterns of effects of arousal dramatically change after participants quit smoking, and effects for different individuals can be very different.

This research was supported by a National Institute on Drug Abuse (NIDA) Grant 1-P50-DA10075. The author is grateful to Professor Saul Shiffman for providing the EMA data and Professor Linda Collins for her encouragement. The author is also indebted to Professor Joseph L. Schafer and Dr. Ted Walls for their motivating discussions, and to Brian Flaherty for his comments.

1 Introduction

Nonparametric regression techniques have been used in the analysis of longitudinal data (Diggle, Liang and Zeger, 1994) and functional data (Ramsey and Silverman, 1997). Various nonparametric models have been proposed to fit longitudinal data when the data were collected over dense time points. These models mainly include varying-coefficient models (Hastie and Tibshirani, 1993), functional linear models (Hoover, et al 1998) and nonparametric mixed effects models (Wu and Zhang, 2002). Modern electronic devices now allow researchers to collect data intensively over time to assess momentary impacts in prevention studies. However, the collected data pose certain difficulties for data analysts because many traditional statistical methods cannot be applied directly for such kinds of data. In this paper, we explore the possibility of applying the models aforementioned for the analysis of intensively collected data in prevention studies to assess momentary impacts.

This work was motivated by the analysis of an Ecological Momentary Assessment (EMA) data set, which was collected by Smoking Research Group led by Professor Saul Shiffman at the University of Pittsburgh. During the study, participants were provided with a hand-held computer, such as a Palm Pilot. The computer beeped the participants at intervals and presented them with questions. It also presented questions to individuals upon demand. For instance, subjects may have been instructed to start a particular data entry program when they felt the urge to smoke a cigarette. Typically, an individual may have been asked as many as 50 questions at 250 time points. Because of the intensive nature of the data collection, data were collected on no more than a few hundred participants.

One of our goals in this paper is to introduce kernel regression and local polynomial regression techniques to readers in the fields of prevention studies. To this end, we present the fundamental ideas of local polynomial regression in Section 2. As an extension of classical nonparametric regression models, varying coefficient models are also introduced in this section. Issues in practical implementation are discussed, and SAS codes to run local polynomial regression for the varying-coefficient models are given in the Appendix. Functional mixed effects models are presented in Section 3. We further proposed an easily implemented estimation procedure for the functional mixed effects models using *Proc mixed* in SAS. This will help readers in the fields of prevention studies to apply the functional mixed effects models for their own research without fully understanding the statistical theoretic backups of the estimation procedures.

Perhaps the most prominent and theoretically important situational association with smoking is that between smoking and mood. Both negative and positive affect have been proposed as important triggers for smoking, and most theories of smoking emphasize the role of affect in driving smoking (Shiffman, *et al* , 2002). In general, it is believed that negative affect is the most commonly attributed cause of smoking relapse (Cummings, Gordon and Marlatt, 1980 and Shiffman, 1982). Using ecological momentary assessment, Shiffman *et al* (2002) found that after controlling for smoking restrictions, smoking was strongly related to smoking urges and modestly related to consumption of coffee and food, the presence of other smokers, and several activities. The striking finding in Shiffman *et al* (2002) is that smoking was unrelated to negative or positive affect or to arousal, although it was associated with restlessness. A natural question arising from these results is whether affect or arousal has impacts on variables which are related to smoking, such as impacts on the urge to smoke and impacts on restlessness. A positive answer to this question implies that affect may have indirectly impacted on smoking through the urge to smoke or restlessness, although it may not have a direct impact on smoking. To address this question, we investigate the relationship between the urge to smoke and mood: affect, arousal and attention disturbance. To get insights into how mood associates with the urge to smoke in the ecology momentary assessment, we concentrate on the data subset collected at a random prompt. Cluster analyses were first conducted for the data subset, and the collected samples were classified into clusters. Using functional linear models, graphical tools are proposed to identify potential impact covariates. Finally, we demonstrate the functional mixed effects model by the analysis of EMA data. From our analysis, we found that affect, arousal and attention disturbance all have significant impact on urge to smoke. Furthermore their effects indeed are time-dependent. We also found that the effects may vary from individual to individual as the variance functions of the random effects in the functional mixed effects model are significantly away from zero.

The rest of this paper is organized as follows. Section 2 presents the basic concepts and fundamental ideas of kernel regression, including running mean and k-nearest neighborhood estimator, and local polynomial regression, including the running line in the literature. Functional mixed effects models and their estimation procedure are proposed in Section 3. In Section 4, we give a thorough analysis of EMA data. Conclusions and discussions are given in Section 5. SAS codes and some supporting figures are given in the Appendix.

2 Nonparametric Regression and Local Polynomial Fitting

In this section, we introduce fundamental ideas of local polynomial regression and illustrate the local polynomial fitting by applications to nonparametric regression models. Suppose that a random sample consisting of $(x_1, y_1), \cdots, (x_n, y_n)$ is taken from the nonparametric regression model:

$$Y = m(X) + \varepsilon, \tag{2.1}$$

where $E(\varepsilon|X) = 0$ and $\text{var}(\varepsilon|X = x) = \sigma^2(x)$. The function $m(x) = E(Y|X = x)$ is called a regression function. In the context of nonparametric regression, it is assumed only that $m(x)$ is a smooth regression of x. Our goal here is to estimate $m(x)$ without imposing a parametric form on the regression function. In this section, we first discuss the case of X being a scalar variable. Extension to multivariate x will be also studied in Section 2.3.

2.1 Kernel Regression

For a given data set, one may try to fit the data by using a linear regression model. If a nonlinear pattern appears in the scatter plot of Y against X, one may employ polynomial regression to reduce the modeling bias of linear regression. To see modeling biases for polynomial regression, a random sample consists 100 observations from the following model

$$y_i = \frac{x_i}{1 - x_i} + \varepsilon_i,$$

where $x_i = i/(n+1)$ is fixed design point, and ε_i's are iid $N(0, 1)$. The scatter plot of the data is depicted in Figure 1, from which it can be seen that the relationship between x and y is highly nonlinear. Polynomial regression is used to fit the data. Figure 1 presents the resulting fits by using four different degrees of polynomials. One can easily see that all resulting fits have substantial biases. Because the degree of polynomial regression cannot be controlled continuously, polynomial functions are not very flexible in modeling features encountered in practice. Furthermore, individual observations can have a large influence on remote parts of the curve in polynomial regression models. Nonparametric regression techniques can be used to repair the drawbacks of polynomial fitting. Fan and Gijbels (1996) give a detailed background and an excellent overview on various nonparametric regression techniques.

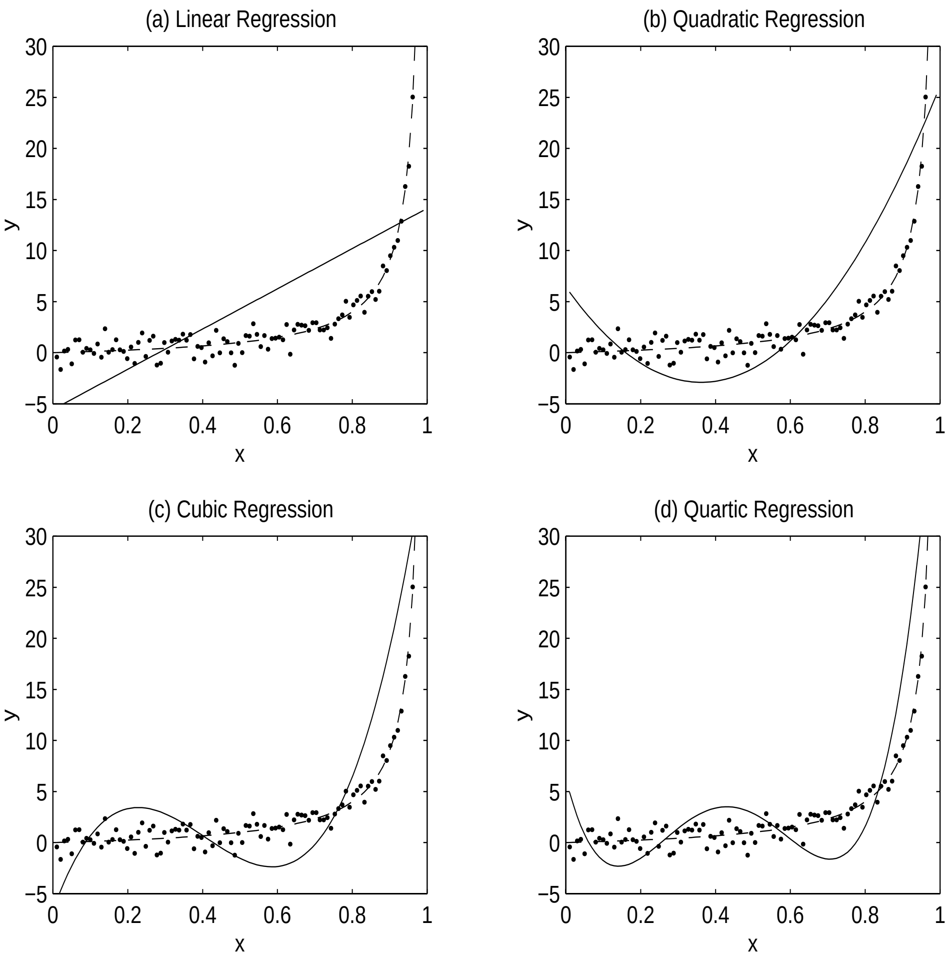

Figure 1: Polynomial fits to the a simulated data set. In (a) — (d), the dots are scatter plot of data, dashed line stands for the true mean function, and solid line is polynomial fit.

Usually, a datum point closer to x carries more information about the value of $m(x)$.

Therefore, an intuitive estimator for the regression function $m(x)$ is the running local average. An improved version of this is the locally weighted average. That is,

$$\widehat{m}(x) = \sum_{i=1}^{n} w_i(x) y_i / \sum_{i=1}^{n} w_i(x).$$

An alternative interpretation of the locally weighted average estimator is that the resulting estimator is the solution to a weighted least-squares problem:

$$\min_{\theta} \sum_{i=1}^{n} (y_i - \theta)^2 w_i(x).$$

In other words, the kernel regression estimator is a weighted least squares estimate at the point x using a local constant approximation.

Let $K(x)$ be a function satisfying $\int K(x)\,dx = 1$, called a *kernel* function, and h is a positive number called a *bandwidth* or a *smoothing parameter*. Take the weights $w_i(x)$ to be $h^{-1}K\{(x_i - x)/h\}$, denoted by $K_h(x_i - x)$. Then it results in a kernel regression estimator, which is given by

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x_i - x) y_i}{\sum_{i=1}^{n} K_h(x_i - x)}. \tag{2.2}$$

See Nadaraya (1964) and Watson (1964).

It is well known that the choice of K is not very sensitive, scaled in a canonical form as discussed by Marron and Nolan (1988), to the estimate $m(x)$. Thus, it is assumed in this paper that the kernel function is a symmetric probability density function. The most commonly used kernel function is the Gaussian density function given by

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2). \tag{2.3}$$

Other popular kernel functions include the symmetric beta family

$$K(x) = \frac{1}{Beta(1/2, \gamma+1)} (1 - x^2)_+^{\gamma}, \qquad \gamma = 0, 1, \cdots, \tag{2.4}$$

where + denotes the positive part, which is assumed to be taken before exponentiation. The support of K is $[-1, 1]$, and $Beta(\cdot, \cdot)$ is a beta function. The corresponding kernel functions when $\gamma = 0, 1, 2$ and 3 are the uniform, the Epanechnikov, the biweight and the triweight kernel functions. Figure 2 shows these kernel functions.

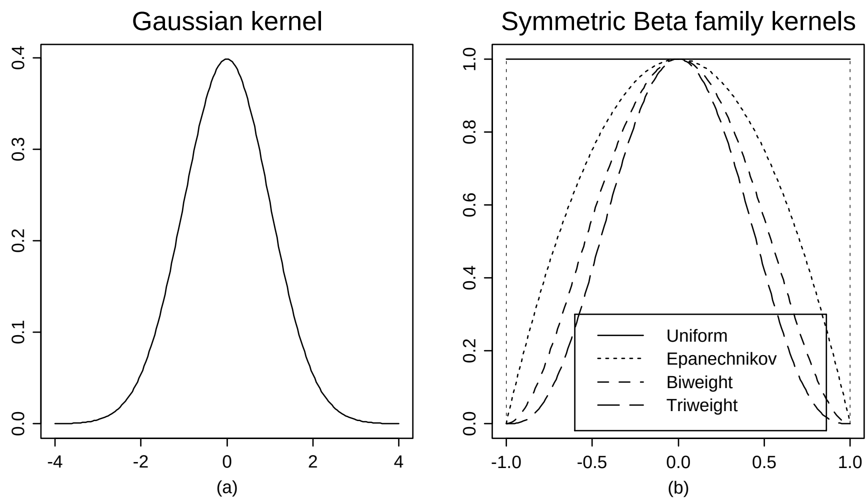

Figure 2: Commonly-used kernels. (a) Gaussian kernel; (b) Symmetric Beta family of kernels that are renormalized to have maximum height 1.

The smoothing parameter h controls the smoothness of regression function. The choice of the bandwidth is of crucial importance. If h is chosen too large, then the resulting estimate misses fine features of the data, while if h is selected too small, then spurious sharp structure become visible. There are works on bandwidth selection in the literature. Jones, Marron and Sheater (1996a, b) give a systematic review on this topic. In practice, one may use a data driven method to choose the bandwidth or select a bandwidth by visualizing the resulting estimated regression function.

Two special cases of the kernel estimator are running means and k-nearest neighborhood estimator. If we take the kernel function to be the density of uniform distribution over $[-1, 1]$, the resulting estimate in (2.2) is the average of y_i's whose associated x_i falls in the interval $[x - h, x + h]$. This estimate is referred to as the *running mean* estimate in the literature. The k *nearest neighborhood* method corresponds to a kernel regression using the uniform kernel with a bandwidth h_x such that there are exactly k x_i's between $x - h_x$ and $x + h_x$. In such a case, the bandwidth h does depend on x. Since we always use the k observations nearest to x to estimate the regression function $m(x)$, the resulting estimate is referred to as the k *nearest neighborhood* estimate.

So far, we have discussed only the estimation of $m(x)$ for a given x. In practice, we may want to estimate $m(x)$ over an interval of x. It is impossible to evaluate $\widehat{m}(x)$ for all

x over the interval. As usual, we only evaluate $\widehat{m}(x)$ over a certain of grid points of x. For instance, we want to obtain $\widehat{m}(x)$ over $[a, b]$. We calculate $\widehat{m}(x)$ over $x = a + (b - a) * i/N$, for $i = 1, \cdots, N$. Here, N is the number of grid points and usually is set to be 200 or 400 in practice. Finally, we plot x against $\widehat{m}(x)$ over the grid points.

We now apply the kernel regression for the data set plotted in Figure 1. The resulting estimate is depicted in Figure 3(a). Compared with the polynomial fits in Figure 1, the kernel estimate catches the overall trend much better than the polynomial fits. To assess the fitting, we compute the average of absolute error (AAE), defined as

$$\mathrm{AAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i|,$$

where $\widehat{y}_i$ is the underlying fit. The AAEs for the linear, quadratic, cubic and quartic polynomial fits are 4.9146, 4.4094, 3.8850, and 3.3227, respectively; the AAE for the kernel regression with bandwidth 0.02 is 0.9119. The AAE decreases as the degree of polynomial fit increases, but the kernel regression provides us a much better fit.

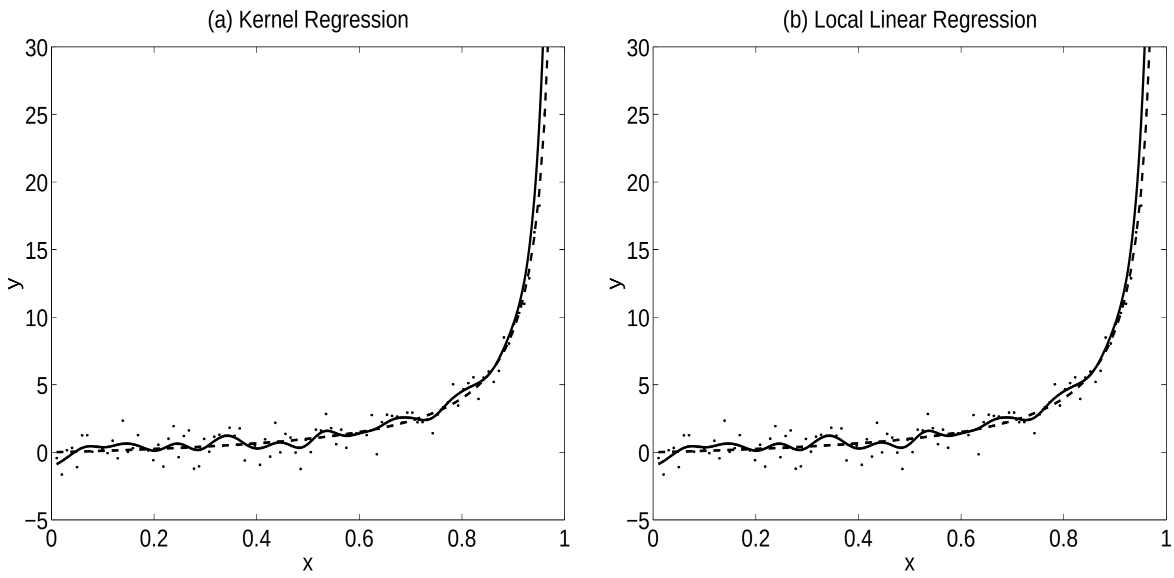

Figure 3: *Plot of Kernel Regression Estimate and Local Linear Estimate. In (a) and (b), the dots are scatter plot of the simulated data, dashed line is the true mean function and solid line is the kernel estimate in (a) and the local linear fit in (b).*

2.2 Local Polynomial Regression

The kernel regression has been extended to local polynomial regression. Suppose that the regression function m is smooth. Applying the Taylor expansion for x in a neighborhood of x, we have

$$m(z) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(z-x)^j \equiv \sum_{j=0}^{p} \beta_j (z-x)^j. \tag{2.5}$$

Thus, for x_i close enough to x,

$$m(x_i) \approx \sum_{j=0}^{p} \beta_j (x_i - x)^j \equiv \mathbf{x}_i^T \boldsymbol{\beta},$$

where $\mathbf{x}_i = (1, (x_i - x), \cdots, (x_i - x)^p)^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$. Intuitively datum points close to x carry more information about $m(x)$. This suggests using a locally weighted polynomial regression

$$\sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 K_h(x_i - x). \tag{2.6}$$

Denote by $\widehat{\beta}_j$ $(j = 0, \cdots, p)$ the minimizer of (2.6). The above exposition suggests that an estimator for the regression function $m(x)$ is

$$\widehat{m}(x) = \widehat{\beta}_0(x). \tag{2.7}$$

Furthermore, an estimator for the ν-th order derivative of $m(x)$ at x is

$$\widehat{m}_\nu(x) = \nu! \widehat{\beta}_\nu(x).$$

When $p = 1$, the local polynomial fitting is referred to as local linear fitting. Particularly, with the uniform kernel, the local linear fitting indeed is the *running line* fitting. Using the uniform kernel, the regression curve is estimated by the intercept in the linear regression based on the data whose x_i's lie between $x - h$ and $x + h$.

When $p = 0$, the local polynomial fitting becomes kernel regression fitting. Hence, the kernel regression may be viewed as a local constant fitting. The local polynomial fitting not only estimates regression curves, but also its derivatives. This is one advantage of the local polynomial fitting over the kernel regression fitting.

The idea of local polynomial regression has been around for a long time. Since both a local constant and local polynomial fit use effectively datum points in a local neighborhood, this idea is referred to as "local modeling". It appeared in the statistical literature in Stone (1977) and Cleveland (1979). Local polynomial regression has many advantages, such as adapting automatically to the boundary (Fan and Gijbels, 1996).

Now we apply the local linear fitting for the data plotted in Figure 1. The estimated regression curve is depicted in Figure 3(b), and looks like the kernel regression fit. We further compute the MAD for the local linear fit, which equals 0.8368. So, the local linear fitting is slightly better than the kernel regression fitting.

2.3 Varying Coefficient Models

We have discussed how to estimate the regression function when x is univariate. The local polynomial regression can be extended straightforwardly to the case of multivariate x. However, it cannot work well in practice due to the so-called *curse of dimensionality*. Therefore, statisticians have studied nonparametric regression models with a certain mean function structure. Two popular classes of nonparametric regression models with multivariate x are additive models (Hastie and Tibshirani, 1990) and varying coefficient models (Hastie and Tibshirani, 1993).

Suppose that $(u_i, \mathbf{x}_i, y_i)$, $i = 1, \cdots, n$, is a random sample from a *varying coefficient model*:

$$Y = \beta_0(U) + \beta_1(U)X_1 + \cdots + \beta_d(U)X_d + \varepsilon, \tag{2.8}$$

where $E(\varepsilon|U, \mathbf{x}) = 0$ and $\text{var}(\varepsilon|U, \mathbf{x}) = \sigma^2$. We in this section illustrate how to apply the local polynomial regression to the varying coefficient model (2.8). Here we focus on the case of univariate U. Due to the curse of dimensionality, the varying coefficient model with multivariate U may not be very useful.

For u in a neighborhood of u_0, it follows by the Taylor expansion that for $l = 0, 1, \ldots, d$,

$$\beta_l(u) \approx \sum_{j=0}^{p} \frac{\beta_l^{(j)}(u_0)}{j!}(u - u_0)^j \equiv \sum_{j=0}^{p} \beta_{lj}(u - u_0)^j.$$

Let $\{\widehat{\beta}_{lj}, l = 0, \cdots, d, j = 0, \cdots, p\}$ be the solution of the following least squares:

$$\sum_{i=1}^{n} \left[y_i - \sum_{l=0}^{d} \{\sum_{j=0}^{p} \beta_{lj}(u_i - u_0)^j\} x_{il} \right]^2 K_h(u_i - u_0), \tag{2.9}$$

where $x_{i0} \equiv 1$. Then $\widehat{\beta}_l(u_0) = \widehat{\beta}_{l0}$, and $\widehat{\beta}_l^{\nu}(u_0) = \nu!\widehat{\beta}_{l\nu}$ for $\nu = 1, \cdots, p$.

2.4 Implementation and Computation

In this section, we address some issues to implement local polynomial regression. Although the choice of kernel function is not very sensitive, the Gaussian kernel is commonly used in practice. However, it has been shown that the Epanechnikov kernel

$$K(x) = 0.75(1 - x^2)_+ \tag{2.10}$$

is optimal under certain criteria (Fan and Gijbels, 1996). Thus, we suggest the use of Epanechnikov kernel.

The optimal bandwidth can be derived, but it involves the unknown regression function and the unknown density function of X. Hence it cannot be applied directly. There are many references on the topic of bandwidth selection. See review paper by Jones, Marron and Sheater (1996a, b) and references therein. In practice, one may subjectively choose a bandwidth by visualization. Here we introduce the cross-validation method, which is a data driven method and conceptually simple, but it needs intensive computation. We illustrate the idea only for the local polynomial regression (2.7). The cross-validation method is directly applicable for the varying coefficient model. Let $\widehat{m}_{h,(-i)}(x)$ be the local linear regression estimator (2.7) using a bandwidth h without using the i^{th}-observation (x_i, y_i). We now analogously validate the "goodness-of-fit" by measuring the "prediction error" $y_i - \widehat{m}_{h,(-i)}(x_i)$. The cross-validation criterion measures the overall "prediction errors", which is defined by

$$CV(h) = n^{-1} \sum_{i=1}^{n} \{y_i - \widehat{m}_{h,(-i)}(x_i)\}^2. \tag{2.11}$$

The cross-validation bandwidth selector $\widehat{h}_{CV}$ chooses the one that minimizes $CV(h)$. It is clear that one may extend the above leave-one-out cross-validation to leave-several-out cross-validation. The latter may save much computational time.

Some computer codes are available through internet. For example, S-plus codes can be downloaded from Matt Wand's homepage at

http://www.biostat.harvard.edu/˜mwand/software.html

while Matlab codes can be downloaded from James S. Marron's homepage at

http://www.stat.unc.edu/faculty/marron/marron_software.html or through the authors. These codes can be easily implemented by directly plugging-in data.

In the rest of this section, we demonstrate how to write SAS codes for the varying coefficient model. The kernel regression and local polynomial regression are its special cases.

Let us focus on local linear regression. To find the local linear estimate for the varying coefficient model (2.8), we solve the following least squares problem:

$$\sum_{i=1}^{n} \{y_i - \beta_{00} - \beta_{10}x_{i1} - \cdots - \beta_{d0}x_{id} - (\beta_{01} + \beta_{11}x_{i1} + \cdots + \beta_{d1}x_{id})(u_i - u_0)\}^2 K_h(u_i - u_0). \tag{2.12}$$

Let $k_i = \sqrt{K_h(u_i - u_0)}$, and define $y_i^* = y_i * k_i$ and $x_{il}^* = x_{il} * k_i$, $l = 1, \cdots, d$. Thus, solving the least squares (2.12) is equivalent to regressing y_i^* on $k_i, x_{i1}^*, \cdots, x_{id}^*$, $k_i * (u_i - u_0), x_{i1}^* * (u_i - u_0), \cdots, x_{id}^* * (u_i - u_0)$ without an intercept. That is, to find the least squares estimate for the following linear regression

$$y_i^* = \beta_{00}k_i + \beta_{10}x_{i1}^* + \cdots + \beta_{d0}x_{id}^* + \beta_{01}k_i(u_i - u_0) + \beta_{11}x_{i1}^*(u_i - u_0) + \cdots + \beta_{d1}x_{id}^*(u_i - u_0) + \varepsilon_i^*, \tag{2.13}$$

where $E(\varepsilon_i^* | \mathbf{x}_i, u_i) = 0$. Hence it may be easily implemented in SAS by using *proc reg*. Section A.1.1 presents a typical SAS code for fitting the varying coefficient over $u \in [a, b]$ with $a = 0$, $b = 10$, $h = 1.2$ and the number of grid point equals 200.

3 Functional Hierarchical Models

Mixed effects models, also called hierarchical linear models, have been frequently used to fit longitudinal data. Due to advance of modern technology of data collection, longitudinal data can be collected at irregular and possibly subject-specific time points. Typical examples are data collected in the study of ecology momentary assessment. The unbalanced structure creates a challenging in the analysis of longitudinal data collected at irregular time points. Hierarchical linear models may not be appropriate in the presence of time-dependent effects. In this section, we introduce functional hierarchical linear models, which are useful extension of hierarchical models.

Before we pursue this further, let us introduce some notation for longitudinal data. Suppose that a random sample consists of n subjects. For i-th subject, $i = 1, \cdots, n$, the response variable $y_i(t)$ and the covariates $\{\mathbf{x}_i(t), \mathbf{z}_i(t)\}$ are collected at time points $t = t_{ij}$, $j = 1, \cdots, J_i$, where J_i is the total number of observations on the i-th subject. We further assume that the random sample is taken from a functional hierarchical model

$$y_i(t_{ij}) = \boldsymbol{\beta}(t_{ij})^T \mathbf{x}_i(t_{ij}) + \boldsymbol{\gamma}_i(t_{ij})^T \mathbf{z}_i(t_{ij}) + \varepsilon_i(t_{ij}), \tag{3.1}$$

where $E\{\varepsilon_i(t)|\mathbf{x}_i(t), \mathbf{z}_i(t)\} = 0$ and $\text{var}\{\varepsilon_i(t)|\mathbf{x}_i(t), \mathbf{z}_i(t)\} = \sigma^2(t)$.

Similar to hierarchical models, it is assumed that $\boldsymbol{\beta}(t)$ consists of p_1 unknown smoothing functions; $\boldsymbol{\gamma}_i(t)$ consists of p_2 unknown smoothing random functions, and for $l = 1, \cdots, p_2$, the l-component of $\boldsymbol{\gamma}_i(t)$, $i = 1, \cdots, n$ is independent and identically distributed zero mean, variance $\sigma_l^2(t)$, and the correlation function $\rho_l(s, t)$ at time s and t. In order to include an intercept and a random intercept in model (3.1), we may make the first components of $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ equal 1. We further allow $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ to have some common covariates.

Functional hierarchical linear model are extensions of hierarchical models by allowing regression coefficients depending on time. It keeps the explanatory power of hierarchical linear models. In the absence of random effects, the functional hierarchical models coincide with the functional linear models, studied by Fan and Zhang (2000) for longitudinal data. The varying coefficient model (2.8) can be viewed as a special case of the functional linear model. In the absence of covariates $\mathbf{x}$ and $\mathbf{z}$, the functional linear models are identical to the nonparametric mixed effects model, studied by Wu and Zhang (2002). Therefore, the functional hierarchical linear models provide a unified framework for many existing models in the literature.

The ideas of local polynomial regression are applicable for the functional mixed effects models. Here we focus on the local linear regression. For any given t_0 in which we want to estimate $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\gamma}_i(\cdot)$, mimic (2.13), and let $k_{ij} = \sqrt{K_h(t_{ij} - t_0)}$. It follows that

$$k_{ij} y_i(t_{ij}) = \sum_{l=1}^{p_1} \{\beta_{l0} + \beta_{l1}(t_{ij} - t_0)\} k_{ij} x_{il}(t_{ij}) + \sum_{l=1}^{p_2} \{\gamma_{l0} + \gamma_{il1}(t_{ij} - t_0)\} k_{ij} z_{il}(t_{ij}) + k_{ij} \varepsilon_i(t_{ij}). \tag{3.2}$$

This can be viewed as a linear mixed effects model with new response variable $k_{ij} y_i(t_{ij})$, new fixed effect covariates $k_{ij} x_{il}(t_{ij})$ and $k_{ij} x_{il}(t_{ij})(t_{ij} - t_0)$, $l = 1, \cdots, p_1$ and new random effect covariates $k_{ij} z_{il}(t_{ij})$ and $k_{ij} z_{il}(t_{ij})(t_{ij} - t_0)$, $l = 1, \cdots, p_2$. Thus, this can be easily implemented using existing software packages, such as *proc mixed* in SAS and *lme* in S-plus.

The theoretical properties of the resulting estimate in (3.2) can be justified since the resulting estimate essentially is the same as that of the local likelihood estimate for nonparametric mixed effects model (Wu and Zhang, 2002). Section A.1.2 presents a typical SAS code for fitting the varying coefficient over $u \in [a, b]$ with $a = 1$, $b = 5$, $h = 2.5$ and the number of grid point equals 200.

4 Analysis of EMA data

Shiffman *et al.* (2002) demonstrated that the urge to smoke variable is substantially associated with smoking, while smoking is unrelated to negative affect, arousal, or attention disturbance. It is of interest to study how negative affect, arousal or attention disturbance associate with the urge to smoke. To this end, we take the *urge to smoke* variable as response variable, $y(t)$, and *negative affect*, denoted by $x_1(t)$, *arousal*, $x_2(t)$, and *attention disturbance*, $x_3(t)$. Further, the three x variables are centralized. It is known that some participants succeeded in quitting smoking, while some did not. This implies that the data may consist of several clusters. It is of important to separate them by cluster analysis first, and then analyze them one cluster by one cluster. We investigate this issue in Section 4.1.

It is always a tricky task to identify important variables or covariates and then include them in the data analysis. A useful strategy is to introduce many potential candidate variables at the initial stage of modeling, and then apply some variable screening techniques to exclude insignificant variables. Section 4.2 will provide graphical tools for this purpose.

In Section 4.3, we analyze an EMA data subset collected at random prompts in details using functional mixed effects models. We will present only results of one cluster to illustrate how to use SAS to estimate the fixed smooth functions and the random smooth functions.

4.1 Cluster analysis

It is challenging to directly conduct cluster analysis on the EMA data due to the nature of data structure. Before conducting cluster analysis, summarizing data is necessary.

We conducted cluster analysis on *difference of score for urge to smoke between pre-quit*

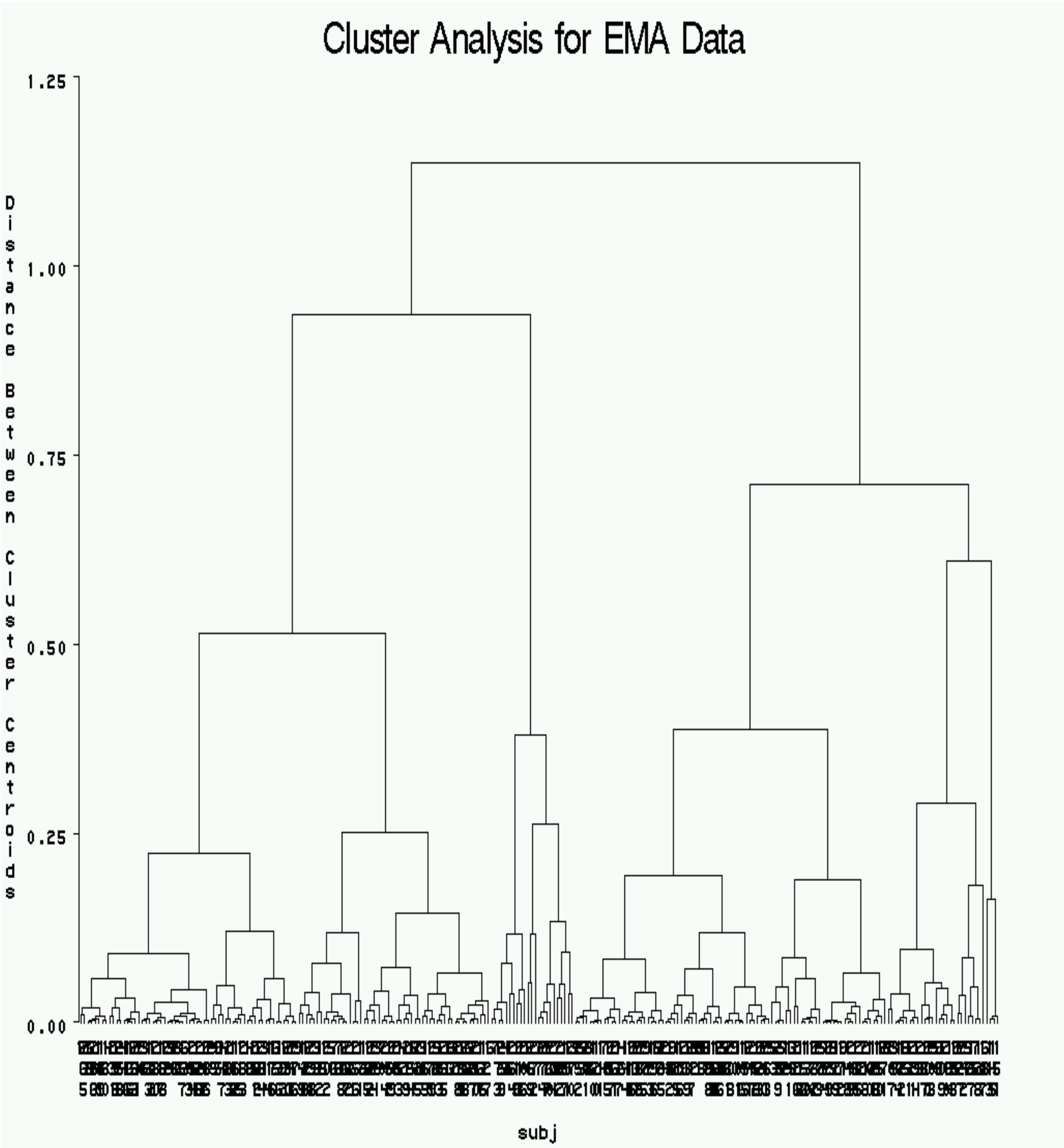

Figure 4: *Tree Plot of Clusters for α_i's.*

and post-quit. Specifically, for subject i, let

$$\alpha_i = \bar{y}_{ib} - \bar{y}_{ia},$$

where $\bar{y}_{ib}$ and $\bar{y}_{ia}$ are the sample means of urge to smoke during pre-quit period and during post-quit period, respectively. We conducted cluster analysis on α_i. This is a simple variable, but it provides us with meaningful results. See next section for results of each cluster. The cluster tree resulting from the cluster analysis based on α_i is depicted in Figure 4, from which it can be seen that the population is classified into 4 groups. The summary for each cluster is presented in Table 1, in which N stands for the number of subjects in each cluster. Figure 5 depicts the density curve of α_i's estimated by kernel density estimation method (Wand and Jones, 1995).

Table 1: Summary of the Four Clusters

Cluster	N	Mean	Std Dev	Minimum	Maximum
1	113	2.8317575	0.8393703	1.6347	4.4145
2	86	0.5484047	0.6318319	-0.7257	1.4708
3	31	-1.5105129	0.7251084	-3.4009	-0.7794
4	23	5.5376391	0.6416067	4.6014	6.7842

We also have tried cluster analysis on other scores. For instance, we fit time-varying coefficient models (2.8) to data of each subject. Thus, we have the estimated regression coefficients for each subject. We then performed a cluster analysis using the regression coefficients. A typical regression tree is depicted in Figure 9 in Appendix. Figure 9 does not provide much information about clusters. Results of cluster analysis for other regression coefficient functions are similar, and we omit them here.

4.2 Graphical tools for model specification

Model selection and model specification are fundamental in statistical modeling. This section provides some graphical tools for model specification. To explore possible time-dependent effects of mood on the urge to smoke, we consider a time-dependent coefficient linear model

$$\begin{aligned} y(t) &= \beta_0(t) + \beta_1(t)x_1(t) + \beta_2(t)x_2(t) + \beta_3(t)x_3(t) + \beta_4(t)x_1^2(t) + \beta_5(t)x_2^2(t) + \beta_6(t)x_3^2(t) \\ &\quad + \beta_7(t)x_1(t)x_2(t) + \beta_8(t)x_1(t)x_3(t) + \beta_3(t)x_2(t)x_3(t) + \varepsilon(t). \end{aligned} \tag{4.1}$$

This model is also referred to as functional linear model (Fan and Zhang, 2000). Here we also include all quadratic terms and all interaction terms of the three factor scores in order to reduce possible model bias. Model (4.1) can be viewed as a varying coefficient model (2.7).

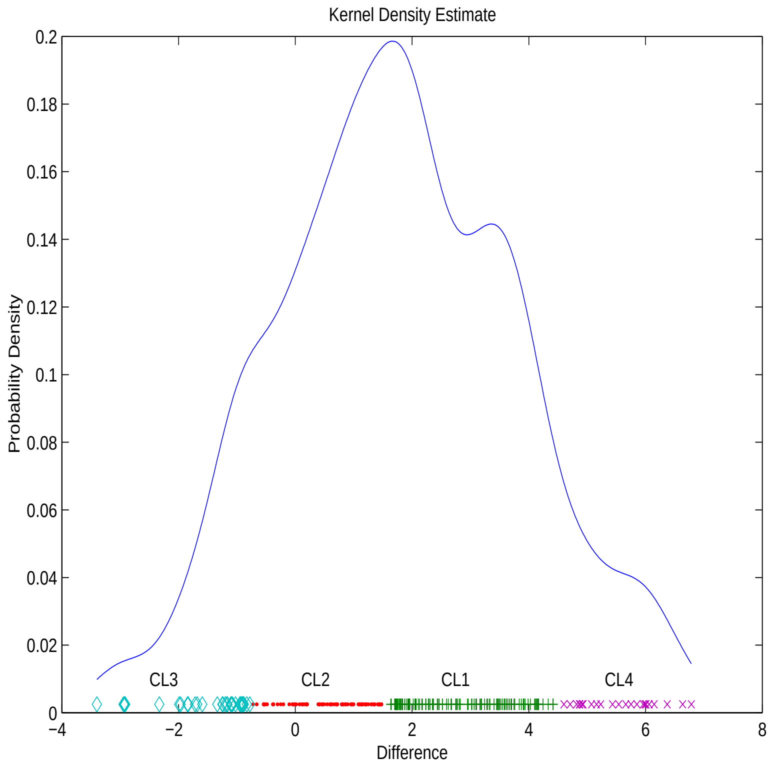

Figure 5: *Estimated Probability Density Curve. The solid curve is the estimated density curve. The symbols "+", ".", "◇", "x" are scatter plots of* α_i *in Cluster 1, 2, 3 and 4, respectively.*

Hence the estimation procedure for varying coefficient models can be extended directly for model (4.1). However, for the purpose of model specification, a simple but quick approach may be developed. Motivating the idea of a running mean, we take the kernel to the uniform kernel and set the bandwidth equal 0.5 (day). With this special kernel and bandwidth, one can easily implement the kernel regression using SAS *proc glm*. The SAS codes to estimate the regression function are given in Section A.1.3. The estimated intercept function for each cluster is depicted in Figure 6. The estimated intercept functions are not smooth because they are evaluated in a time point per day, and data for each day are not overlapped. Note that the x variables have been centralized before conducting analysis, Figure 6 provides us overall trends when the mood of smokers is on average level. It is interesting to observe that the overall trends for different clusters are quite different. On average, subjects in Clusters 1 and 4 made great progress after they quit smoking, subjects in Cluster 2 made small progress, and subjects in Cluster 3 made no progress. The overall trends almost remain constant during the pre-quit period; however, they change over time during post-quit period. For Clusters 1 and 4, the overall trends significantly decrease over the post-quit period. For Cluster 2, the overall trend slightly decreases over the post-quit period, while for Cluster 3, the overall trend slightly increases post-quit period. Although the cluster analysis is based on the simple score α_i, it provides us reasonable results.

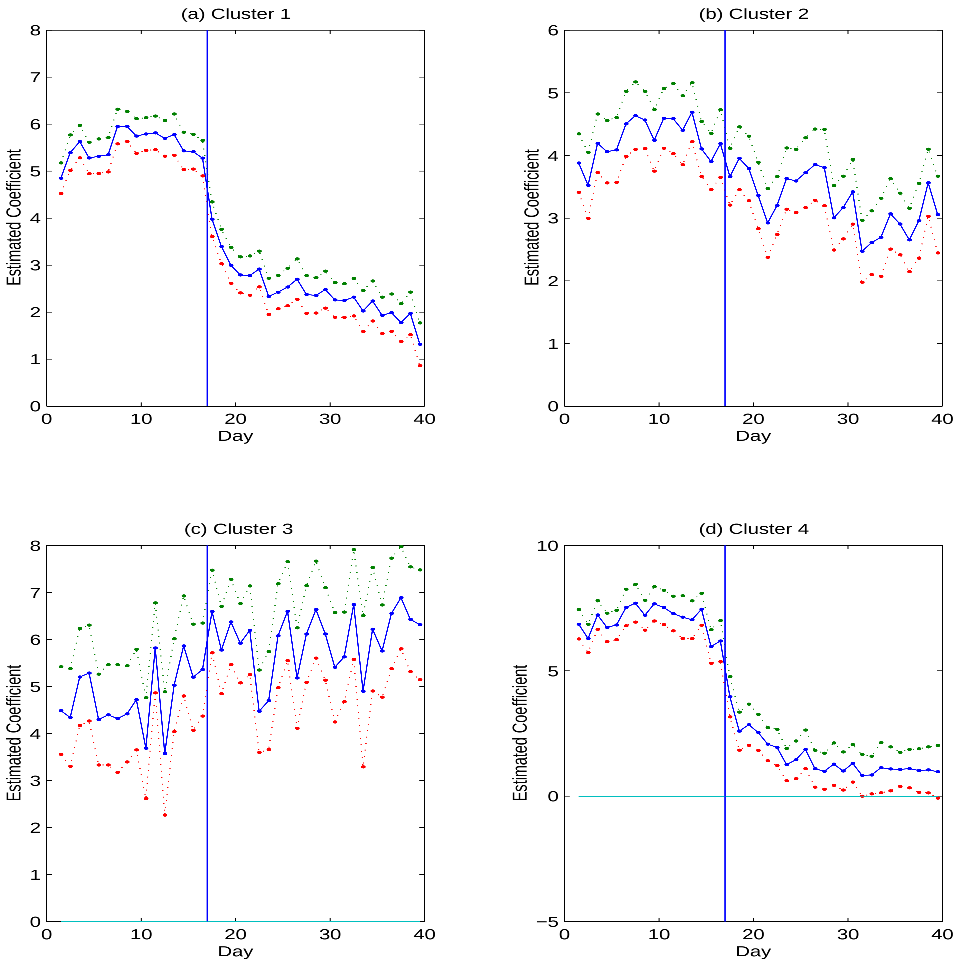

Figure 6: *Plot of Estimated Intercept Functions. Solid curves stand for estimate and dotted curves are 95% pointwise confidence intervals.*

Figure 7 depicts the estimated coefficients of *negative affect*. From Figure 7, it can be seen that the patterns for effects of negative affect for different clusters look quite different. For Cluster 1, the effect of negative affect seems to be constant over the pre-quit period; however, it decreases over post-quit period. On the other hand, for Cluster 4, the effect of negative affect seems to be statistically insignificant over the pre-quit period, has significant jump up just after quit date, decreases over post-quit period, and comes eventually back to an insignificant level.

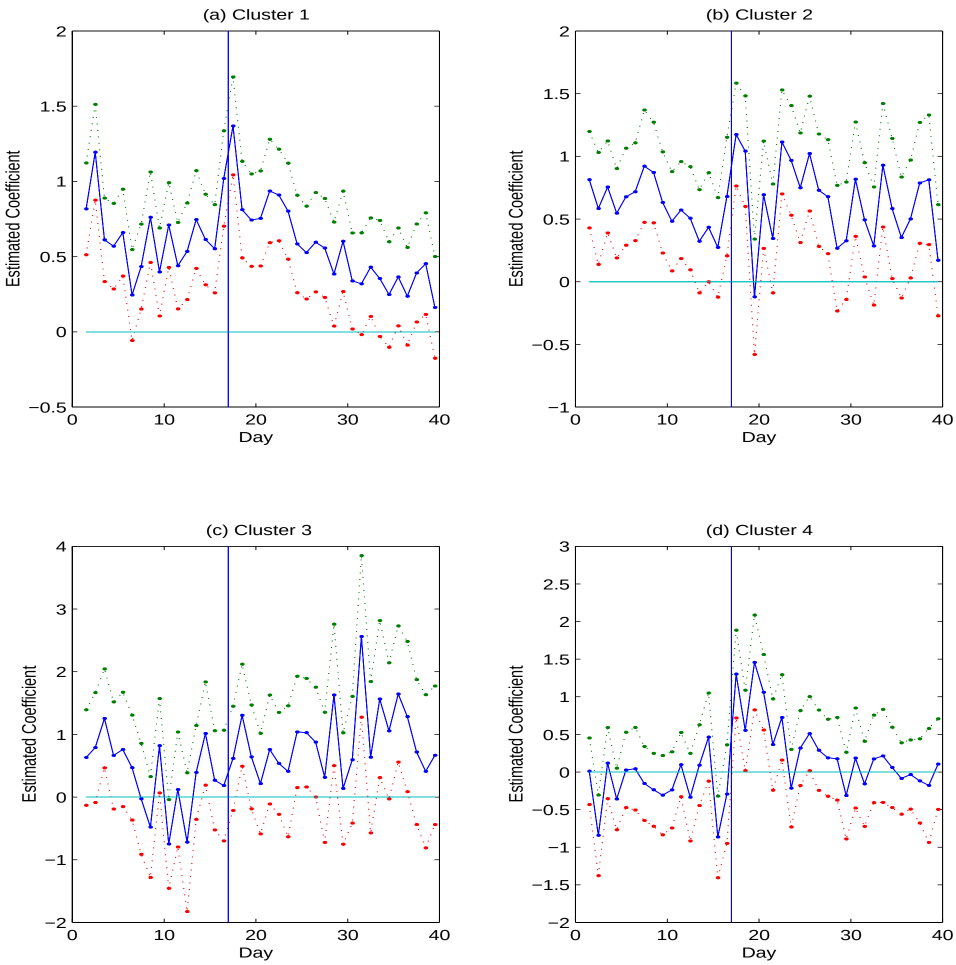

Figure 7: *Plot of Estimated Effects of Negative Affect. Solid curves stand for estimate and dotted curves are 95% pointwise confidence intervals.*

The estimated coefficients of other effects, quadratic terms and interaction terms are presented in Figures 10 to 17 in Section A.2 of the Appendix. From those figures, it can be seen that the effects of quadratic terms and interaction terms are not statistically significant. Therefore, they may be excluded from the model in the analysis in next section.

4.3 Analysis via functional hierarchical linear models

This section provides a detailed analysis of EMA data using functional mixed effects models. We will focus on only the data in Cluster 1. As mentioned in the last section, all quadratic terms and interaction terms are not statistical significant. Therefore, we consider the following functional hierarchical model:

$$\begin{aligned} y_i(t_{ij}) &= \beta_0(t_{ij}) + \beta_1(t_{ij})x_{i1}(t_{ij}) + \beta_2(t_{ij})x_{i2}(t_{ij}) + \beta_3(t_{ij})x_{i3}(t_{ij}) + \beta_4(t_{ij})x_{i1}(t_{ij})x_{i2}(t_{ij}) \\ &\quad + \gamma_{i0}(t_{ij}) + \gamma_{i1}(t_{ij})x_{i1}(t_{ij}) + \gamma_{i2}(t_{ij})x_{i2}(t_{ij}) + \varepsilon_i(t_{ij}), \end{aligned} \tag{4.2}$$

where $\beta_i(\cdot)$, $i = 0, ..., 4$ are fixed effects, $\gamma_{il}(\cdot)$, $l = 0, 1, 2$ are random effects, $\varepsilon_i(t_{ij})$ is a Guassian process with mean zero, variance function $\sigma^2(t_{ij})$, and the following correlation structure: equally correlated within a day and an arbitrary correlated between days. We only include the interaction term between negative affect and arousal because the possible interaction may play an important role in how to interpret the impact of negative affect on urge to smoke associated with arousal. Based on our limited modeling experience using *proc mixed* in SAS, we found that including many random effects in the model may result in non-convergent solutions. Therefore, the interaction term and attention were not included as random effects.

Proc mixed in SAS is employed to estimate the regression coefficient functions. The code is displayed in Section A.1.4. Figure 8 depicts the plot of the estimated intercept function $\widehat{\beta}_0(t)$. Comparing with Figure 6(a) and Figure 8, we found that overall trends are the same, but Figure 8 provides us a smooth function. Furthermore, the local linear regression also provides us a more efficient estimate as it also used the data in other days.

The plot of $\widehat{\beta}_1(t)$, the regression function of negative affect, is displayed in Figure 18, from which we can see that there exists a dramatic change around quit date, and the effect does vary over time. Figure 19 depicts the estimated $\widehat{\beta}_2(t)$. We can see the pattern changes over time from this figure. The effect of arousal is positive over the pre-quit period, while the effect of arousal become negative or non-significant over the post-quit period.

Figure 20 shows the estimated $\widehat{\beta}_3(t)$. The effect of attention varies around 0.25. From Figure 20, attention might have only constant effect. Figure 21 presents the estimated $\widehat{\beta}_4(t)$.

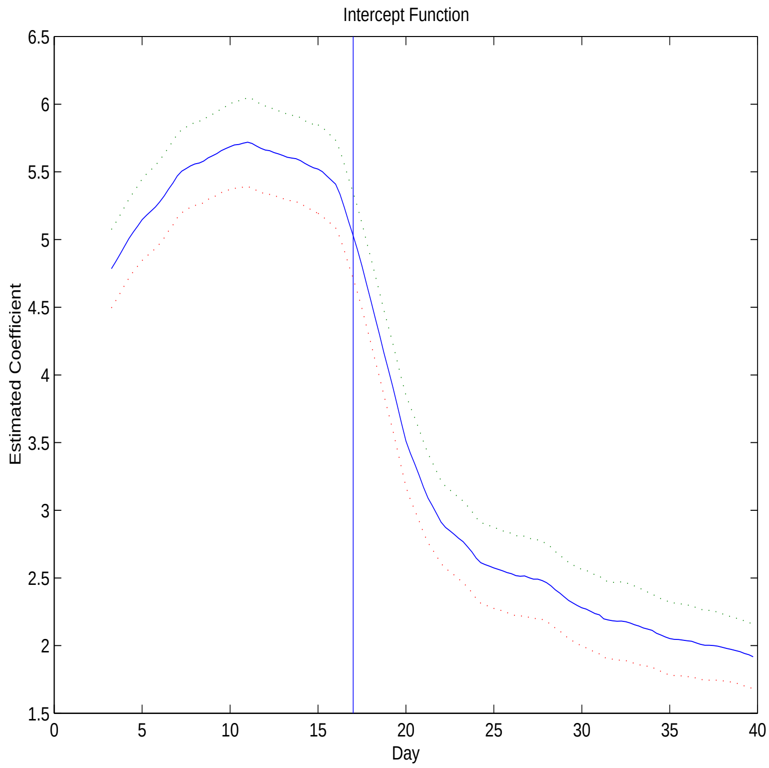

Figure 8: *Plot of Estimated Intercept Function. The solid curve is the estimated regression coefficient function, and the dotted curves are 95% pointwise confidence interval.*

It is clear that the interaction effect between negative affect and arousal is not statistically significant at level 0.05.

Conditional effects of negative affect, given low or high arousal level, which corresponds to the 1st and 3rd quartiles of arousal scores, are depicted in Figure 22. The overall pattern is similar to that of effect of negative affect. This is because the interaction term is not statistical significant.

We next present results for the random effects. The estimated variance functions are depicted in Figures 23—25. All random effects are significant. The variance function for the random intercept may be a constant. The variance function for the random effect of negative affect becomes a constant after quitting smoking. Figure 26—28 depicts the plot of $\beta_l(t)+\beta_{il}(t)$, $l=0,1,2$, for the first 3 subjects in Cluster 1. From these figures, it can be seen that individual effects may be dramatically different from the fixed effects. The estimated variance function $\widehat{\sigma}^2(t)$ for the random error is depicted in Figure 29, which also shows the variance function significantly varies over time.

5 Conclusions and Discussions

In this paper, we have introduced the local polynomial regression techniques, and we have applied them for varying coefficient models and functional mixed effects models. A thorough analysis of EMA has been conducted using functional linear models and functional mixed effects models. Patterns of change on important smoking related dimensions were identified in this work. To our best knowledge, these patterns were not found in previous work.

References

Cummings, C., Gordon, J. R. and Marlatt, G. A. (1980). Relapse: prevention and prediction, In W. R. Willer (ed.) *The addictive behaviors* (pp. 291-322). Oxford, England, Pergamon.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Oxford, U. K. Oxford University Press.

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

Fan, J. and Zhang, J. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Jour. Royal Statist. Soc., B*, **62**, 303-322.

Hastie, T. J. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

Hastie, T. J. and Tibshirani, R. (1993). Varying-coefficient models (with discussions), *J. Royal Statist. Soc, B.*, **55**, 757-796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient model with longitudinal data, *Biometrika*, **85**, 809-822.

Jones, M. C., Marron, J. S. and Sheater, S. J. (1996a). A brief survey of bandwidth selection for density estimation, *J. Amer. Statist. Assoc.*, **91**, 401-407.

Jones, M. C., Marron, J. S. and Sheater, S. J. (1996b). Progress in data-based bandwidth selection for kernel density estimation, *Computational Statistics*, **11**, 337-381.

Marron, J.S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statist. Prob. Lett.*, **7**, 195–199.

Nadaraya, E.A. (1964). On estimating regression. *Theory Prob. Appl.*, **9**, 141–142.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*, Springer, New York.

Shiffman, S. (1982). Relapse following smoking cessation: A situational analysis, *Journal of Consulting and Clinical Psychology*, **50**, 71-86.

Shiffman, S., Gwaltney, C. J., Balabanis, M. H., Liu, K. S., Paty, J. A., Kassel, J. D., Hickcox, M. and Gnys, M. (2002). Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. *J. Abnormal Psychology*, **111**, 531-545.

Wand, M. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā* Ser. A, **26**, 359–372.

Wu, F. and Zhang, J. (2001). Local polynomial mixed-effects model for longitudinal data. *Jour. Amer. Statist. Asso.*, **97**, 883-897.

Appendix

A.1 SAS codes

A.1.1 SAS code for varying coefficient models

```
options linesize= 95  ps=70 pageno=1 mprint mlogic symbolgen formdlim='-';

data data1;
   infile 'C:\data\dataa.dat';
   input  y u x1-xd;
run;

%macro varycoef(startv,endv);
  %do i=&startv %to &endv %by 1;
    data data0;
        set data1;
        if (u>(0+&i*(10-0)/200+1.2))|(u<(0+&i*(10-0)/200-1.2)) then delete;
            /* u0=a+(b-a)*i/200, i=1,...,200; h=1.2; a=0, b=10 */
        ui0 = u - 0-&i*((10-0)/200); /* ui0= ui-u0 */
        k = sqrt(0.75*(1-(ui0/1.2)*(ui0/1.2))); /* Epanechnikov Kernel is
            used here */
        xk1 = x1*k;
        xk2 = x2*k;
            .
            .
            .
        xkd = xd*k;
        uxk1= ui0*xk1;
        uxk2= ui0*xk2;
            .
            .
            .
        uxkd= ui0*xkd;
     run;

    title2 "i= &i";

    proc reg data=data0;
        model y= k xk1-xkd k*ui0 uxk1-uxkd/noint;
    run;
    %end;
%mend varycoef;
%varycoef(1,200);  /* run macro varycoef */
```

A.1.2 SAS code for functional mixed effect models

```
options linesize= 95  ps=70 pageno=1 mprint mlogic symbolgen formdlim='';

data data1;
   infile 'C:\data\datab.dat';
   input  subj y t x1-xp1 z1-zp2;
run;

%macro fmixed(startv,endv);
  %do i=&startv %to &endv %by 1;
    data data0;
        set data1;
        if (t>(1+&i*((5-1)/200)+2.5))|(t<(1+&i*((5-1)/200)-2.5)) then delete;
            /* t0=a+(b-a)*i/200, i=1,...,200; h=2.5; a=1, b=5 */
        ti0 = t - 1-&i*((5-1)/200); /* ti0= ti-u0 */
        k = sqrt(0.75*(1-(ti0/2.5)*(ti0/2.5))); /* Epanechnikov Kernel is
            used here   */

        xk1 = x1*k;;
        xk2 = x2*k;;
            .
            .
            .
        xkp1 = xd*k;

        txk1= ti0*xk1;
        txk2= ti0*xk2;;
            .
            .
            .
        txkp1= ti0*xkp1;

        zk1 = z1*k;;
        zk2 = z2*k;;
            .
            .
            .
        zkp2 = zd*k;

        tzk1= ti0*zk1;
        tzk2= ti0*zk2;;
            .
            .
            .
```

```
        tzkp2= ti0*zkp2;
       run;

    title2 "i= &i";

    proc mixed covtest data=data0 method=REML;
        class subj;
        model y = xk1-xkp1 uxk1-uxkp1/noint solution;
        random zk1-zkp2 tzkp1-tzkp2/subject=subj solution;
    run;
    %end;
%mend fmixed;

%fmixed(1,200);  /* run macro fmixed */
```

A.1.3 SAS code for graphical tools

```
options linesize= 95  ps=70 pageno=1;
libname ema "c:\data\outdata\";
data qttype1;
    infile 'C:data\QTDATAmtype1.m';
    input  subj time0 day dow B_urge NegAff Arousal Atten;
    NegAff2 = NegAff*NegAff;
    Arousal2 = Arousal*Arousal;
    Atten2 = Atten*Atten;
    Neg_Arousal =NegAff*Arousal;
    Neg_Atten =NegAff*Atten;
    Arousal_Atten =Arousal*Atten;
    time=time0+day;
    if (dow<6) then wk=0; else wk=1;
run;
proc sort data=ema.out;
    by subj;
run;

title1 "Analysis of EMA Data using clustering";
 data qttype1out;
    merge  qttype1 ema.out;
    by subj;
    drop clusname;
    if day>40 then delete;
    /* Only analyze data before day 40 as data are sparse after day 40*/
run;

data cluster1;
    set qttype1out;
    where cluster=1;
run;

title2 "Output of Cluster 1";
    proc glm data=cluster1;
    class day;
    model B_urge = day day*NegAff day*Arousal day*Atten day*NegAff2 day*Neg_Arousal
        day*Neg_Atten day*Arousal2 day*Arousal_Atten day*Atten2/noint solution;
run;
quit;

data cluster2;
    set qttype1out;
    where cluster=2;
```

```
run;

title2 "Output of Cluster 2";
proc glm data=cluster2;
    class day;
    model B_urge = day day*NegAff day*Arousal day*Atten day*NegAff2 day*Neg_Arousal
        day*Neg_Atten day*Arousal2 day*Arousal_Atten day*Atten2/noint solution;
run;
quit;

data cluster3;
    set qttype1out;
    where cluster=3;
run;

title2 "Output of Cluster 3";
proc glm data=cluster3;
    class day;
    model B_urge = day day*NegAff day*Arousal day*Atten day*NegAff2 day*Neg_Arousal
        day*Neg_Atten day*Arousal2 day*Arousal_Atten day*Atten2/noint solution;
run;
quit;

data cluster4;
    set qttype1out;
    where cluster=4;
run;

title2 "Output of Cluster 4";
proc glm data=cluster4;
     class day;
     model B_urge = day day*NegAff day*Arousal day*Atten day*NegAff2 day*Neg_Arousal
        day*Neg_Atten day*Arousal2 day*Arousal_Atten day*Atten2/noint solution;
run;
quit;
```

A.1.4 SAS code for graphical tools

```
options linesize= 95  ps=70 pageno=1 mprint mlogic symbolgen formdlim=' ';
libname ema "c:\data\outdata\";

title1 "Analysis of EMA Data Using Linear Regression";
data qttype1;
    infile 'C:\data\QTDATAmtype1.m';
    input  subj time0 day dow B_urge NegAff Arousal Atten;
    time=time0+day;
run;

proc sort data=ema.out;
    by subj;
run;

title1 "Analysis of EMA Data using clustering";
data qttype1out;
    merge  qttype1 ema.out;
    by subj;
    drop clusname;
    if day>43 then delete;
run;

data cluster1;
    set qttype1out;
    where cluster=1;
run;

data qttype10;
    set cluster1;
    NegAff = NegAff  +0.0661303 ;  /* centralized */
    Arousal = Arousal-0.0485865;
    Atten = Atten-0.0174763;
    NegAff2 = NegAff*NegAff;
    Arousal2 = Arousal*Arousal;
    Atten2 = Atten*Atten;
    Neg_Arousal =NegAff*Arousal;
    Neg_Atten =NegAff*Atten;
    Arousal_Atten =Arousal*Atten;
run;

%macro kernals(startv,endv);
  %do i=&startv %to &endv %by 1;
    data qttype2;
```

```
        set qttype10;
        if (time>(&i*0.25+2.5))|(time<(&i*0.25-2.5)) then delete;
                /* grid=0.25 bandwidth=2.5 */
        t0 = &i*0.25-time;
        w = sqrt(0.75*(1-(t0/2.5)*(t0/2.5)));
        B_urge = w*B_urge;
        NegAff = w*NegAff;
        Arousal = w*Arousal;
        Atten =  w*Atten;
        Neg_Arousal = w*Neg_Arousal;
    run;

    title2 "i= &i";

    ods listing close;

    proc mixed covtest data=qttype2 method=REML
         noclprint noitprint noinfo ;
        class subj day;
        model B_urge = w NegAff Arousal Neg_Arousal Atten w*t0 t0*NegAff t0*Arousal
                       t0*Neg_Arousal t0*Atten/noint solution ;
        random w  w*day NegAff Arousal/ subject=subj solution;
        ods output CovParms=cov&i SolutionF=est&i SolutionR=pre&i;
    run;

    ods listing;
   %end;
%mend kernals;

%kernals(3*4+1,40*4-1);
    /* Not to compute the estimate at boundary due to boundary
    effects */

%macro p1(startv,endv);
  %do i=&startv %to &endv %by 1;
    title2 "i= &i";
    proc print data=est&i;
    run;
  %end;
%mend p1;

%p1(3*4+1,40*4-1);

%macro p2(startv,endv);
  %do i=&startv %to &endv %by 1;
```

```
    title2 "i= &i";
    proc print data=cov&i;
    run;
  %end;
%mend p2;

%p2(3*4+1,40*4-1);

%macro p3(startv,endv);
  %do i=&startv %to &endv %by 1;
    title2 "i= &i";
    proc print data=pre&i;
    where subj<9;
    run;
  %end;
%mend p3;

%p3(3*4+1,40*4-1);
```

A.2 More figures for analysis of EMA data

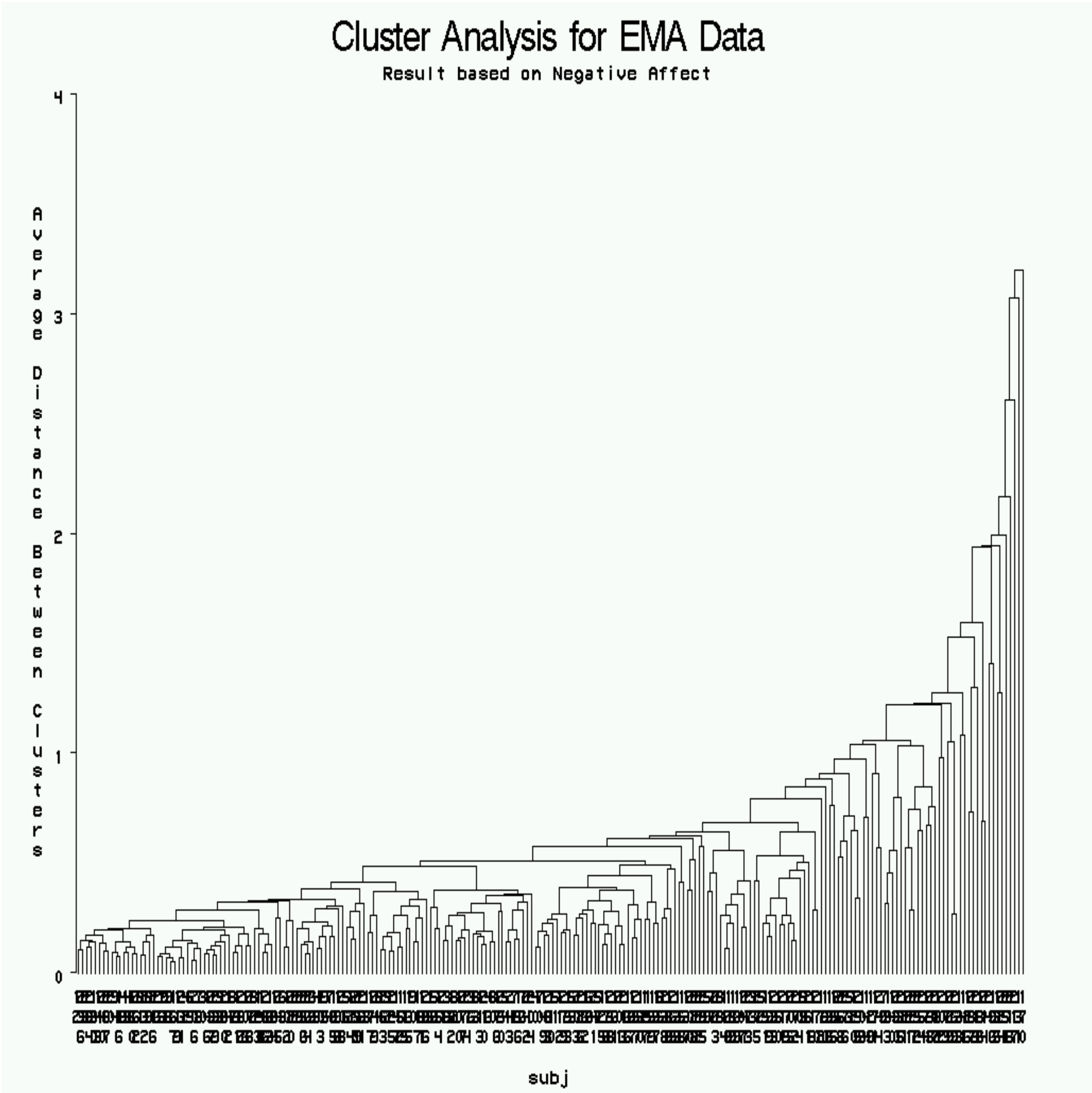

Figure 9: *Tree Plot of Clusters based on Coefficient of Negative Affects.*

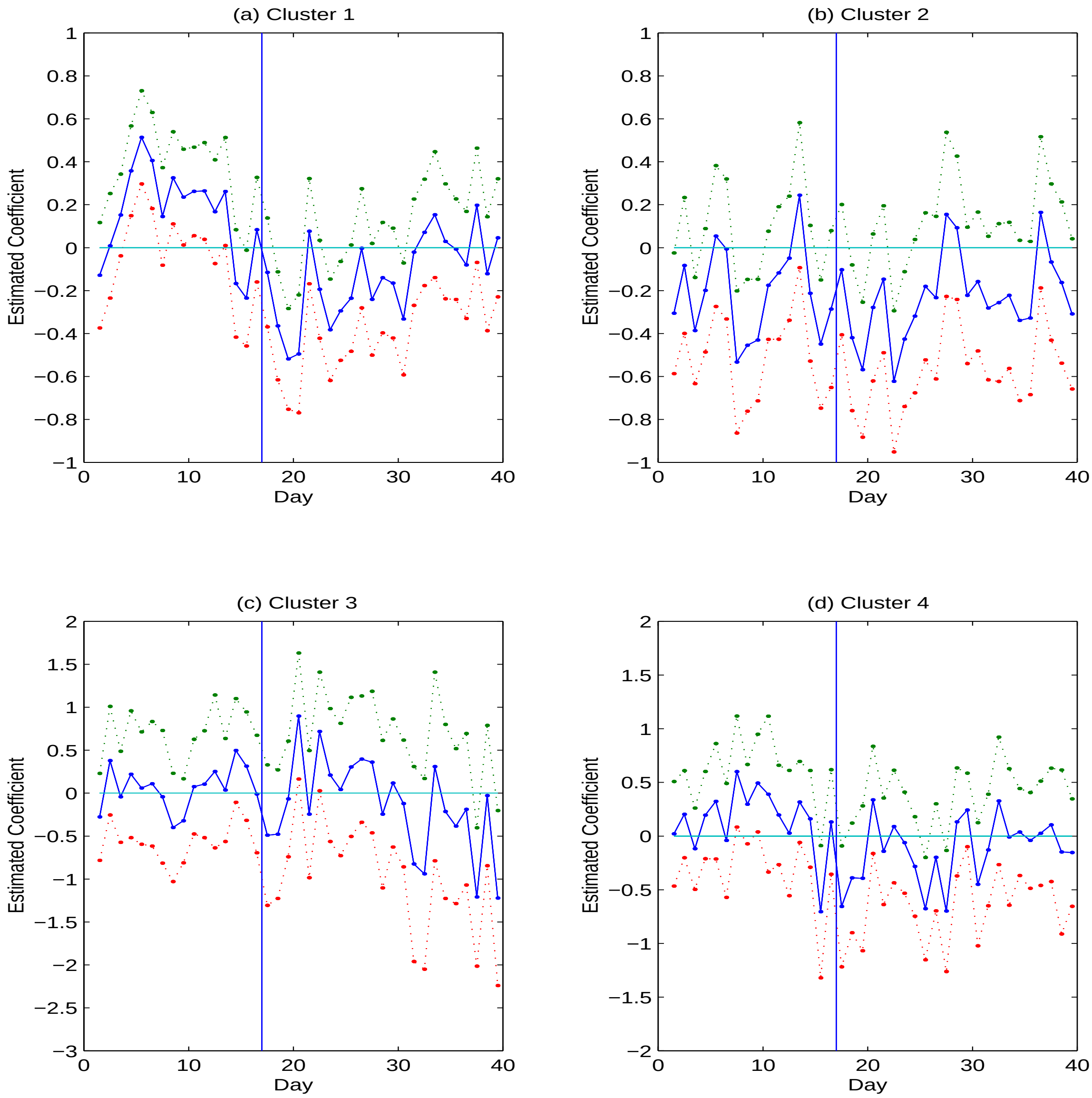

Figure 10: *Plot of Estimated Effects of Arousal. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

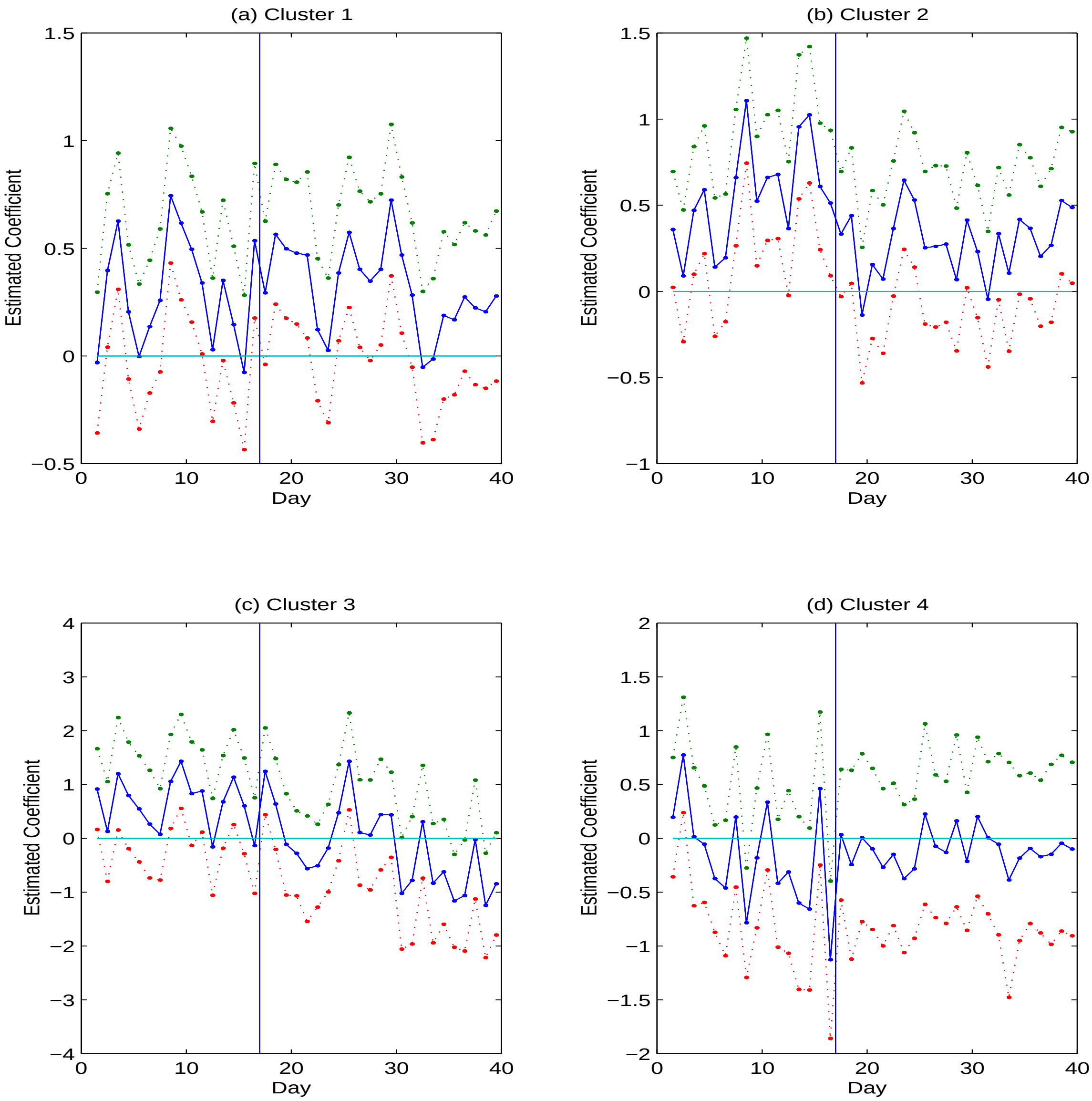

Figure 11: *Plot of Estimated Effects of Attention. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

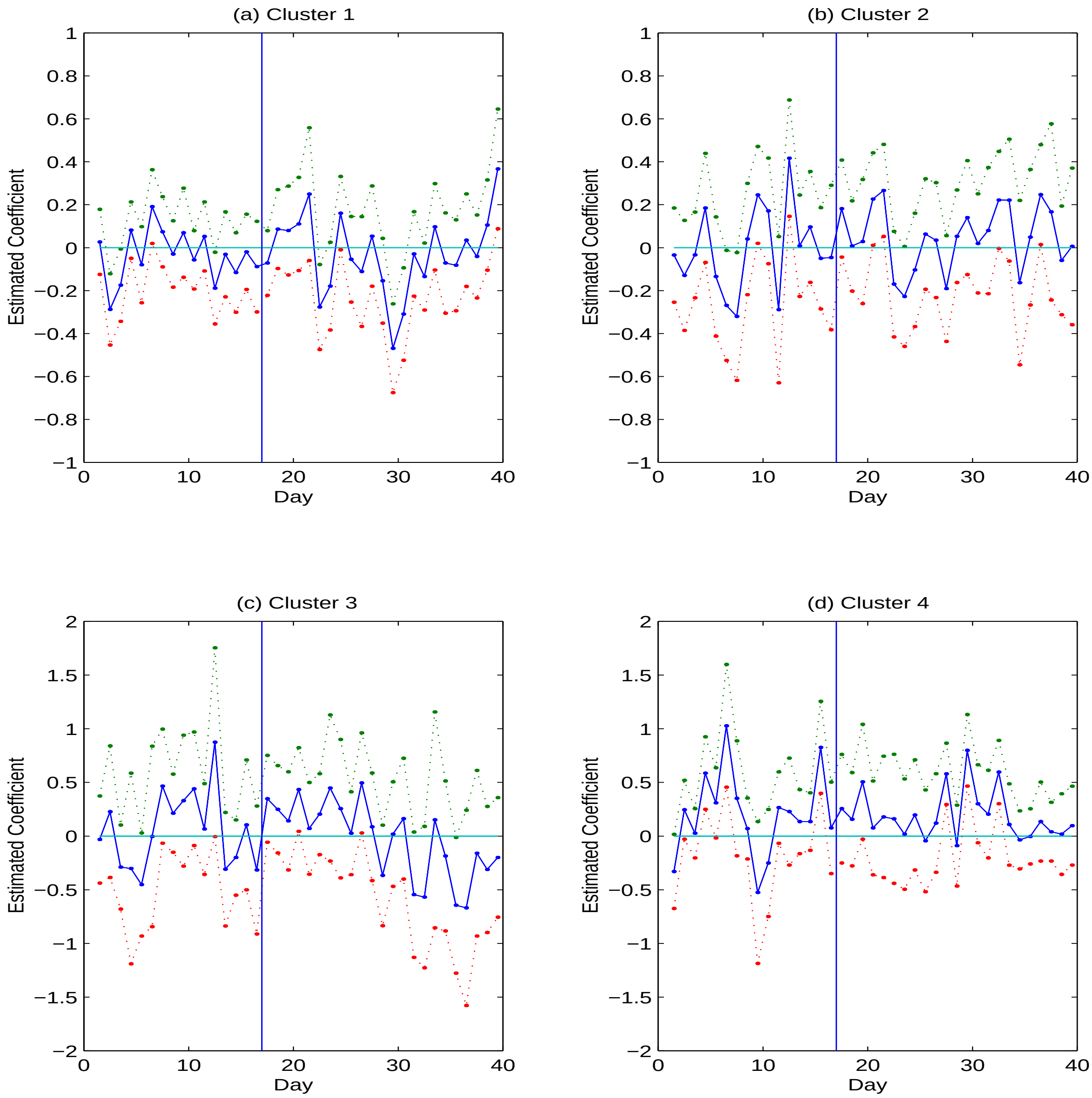

Figure 12: *Plot of Estimated Effects of Square of Negative Affect. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

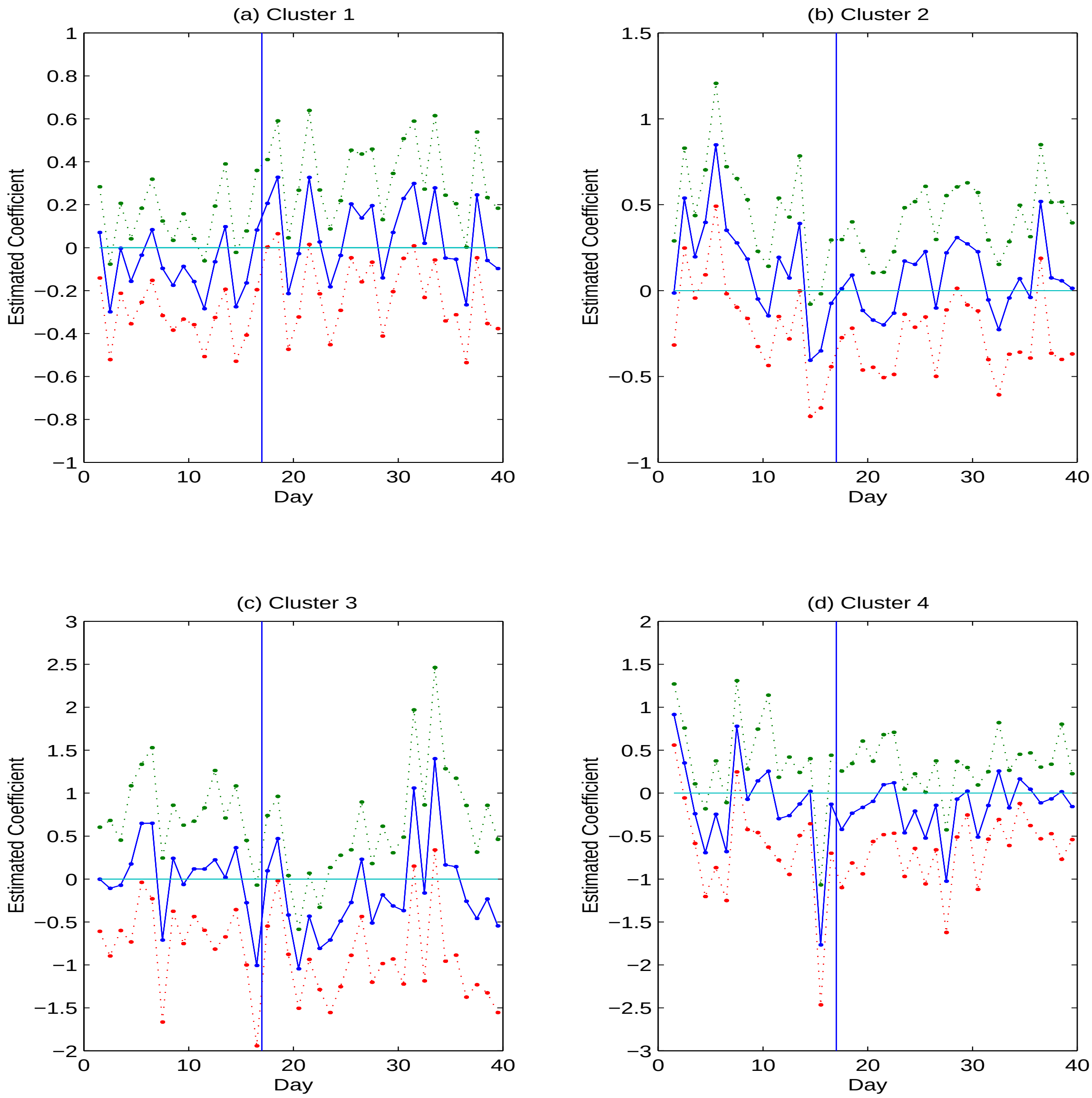

Figure 13: *Plot of Estimated Effects of Interaction between Negative Affect and Arousall. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

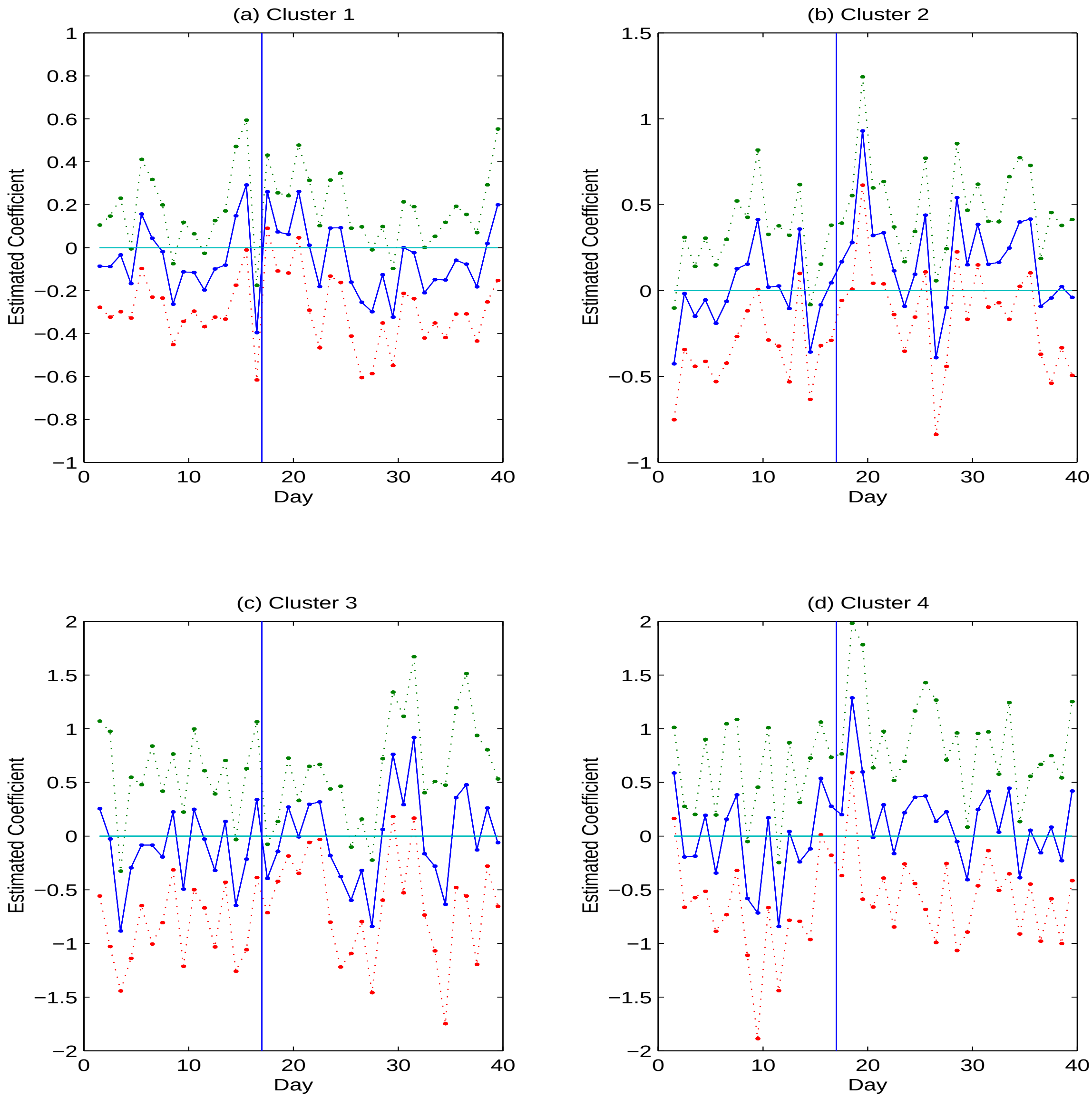

Figure 14: *Plot of Estimated Effects of Interaction between NA and Attention. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

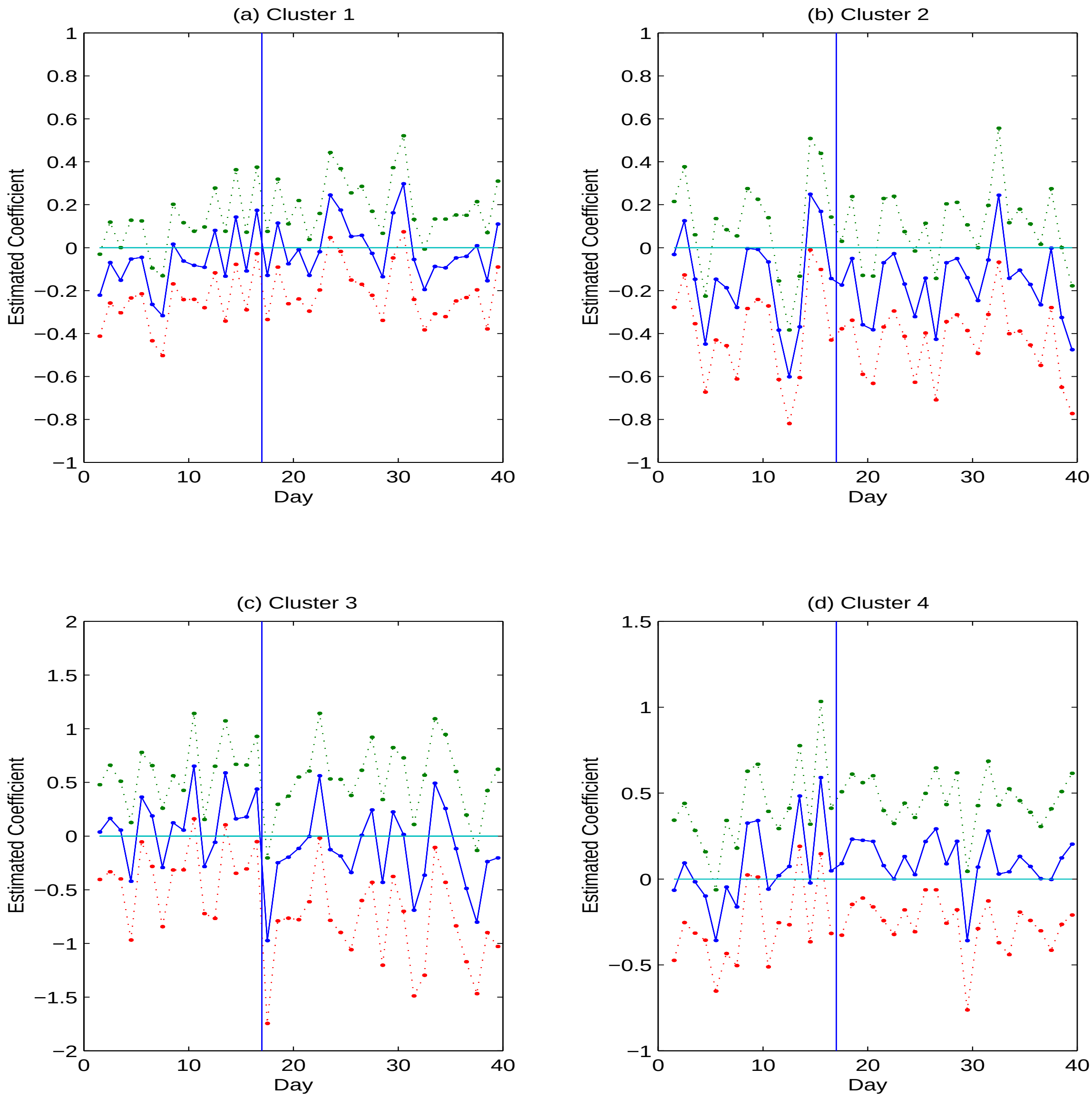

Figure 15: *Plot of Estimated Effects of Square of Arousal. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

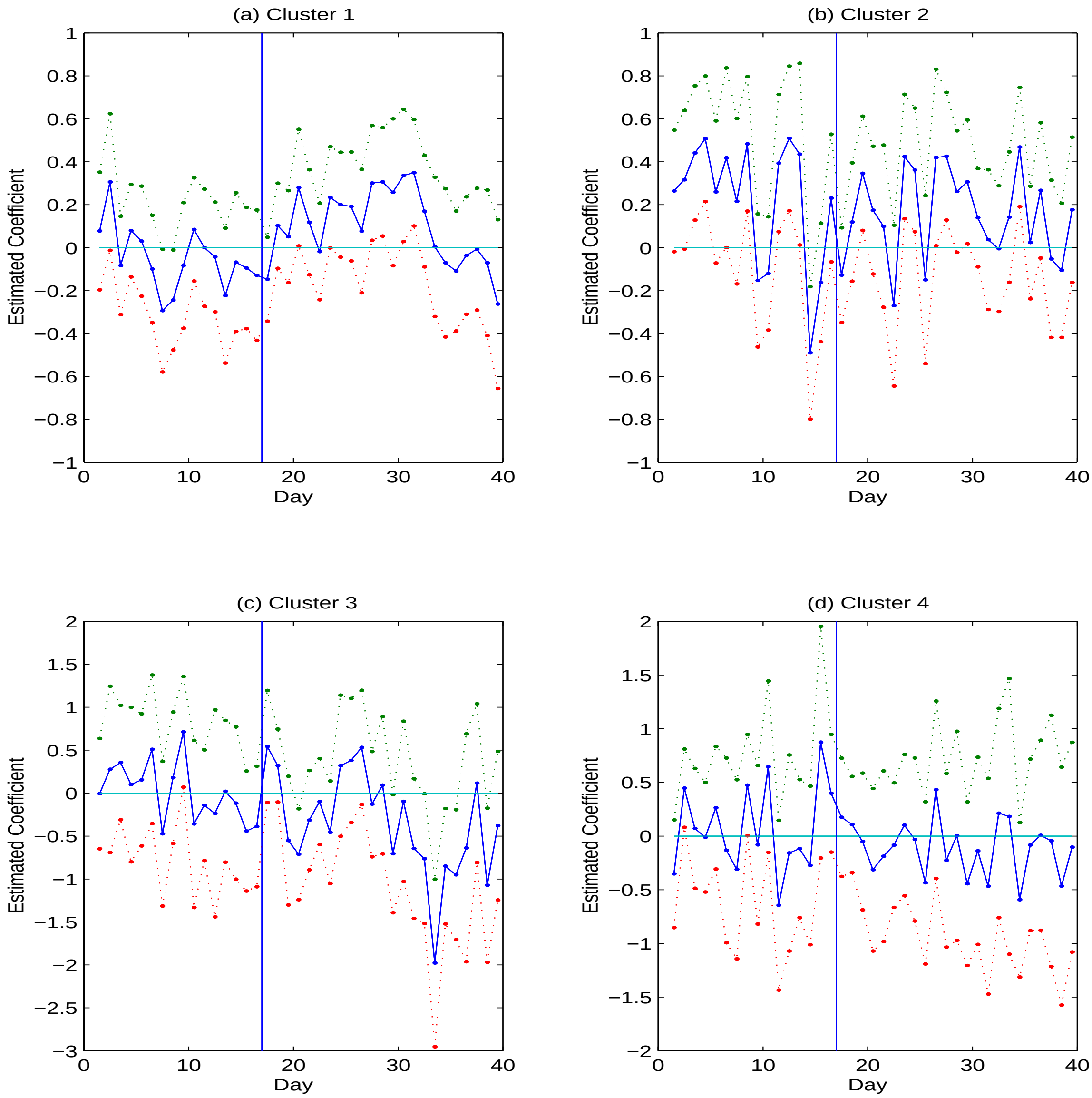

Figure 16: *Plot of Estimated Effects of Interaction between Arousal and Attention. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

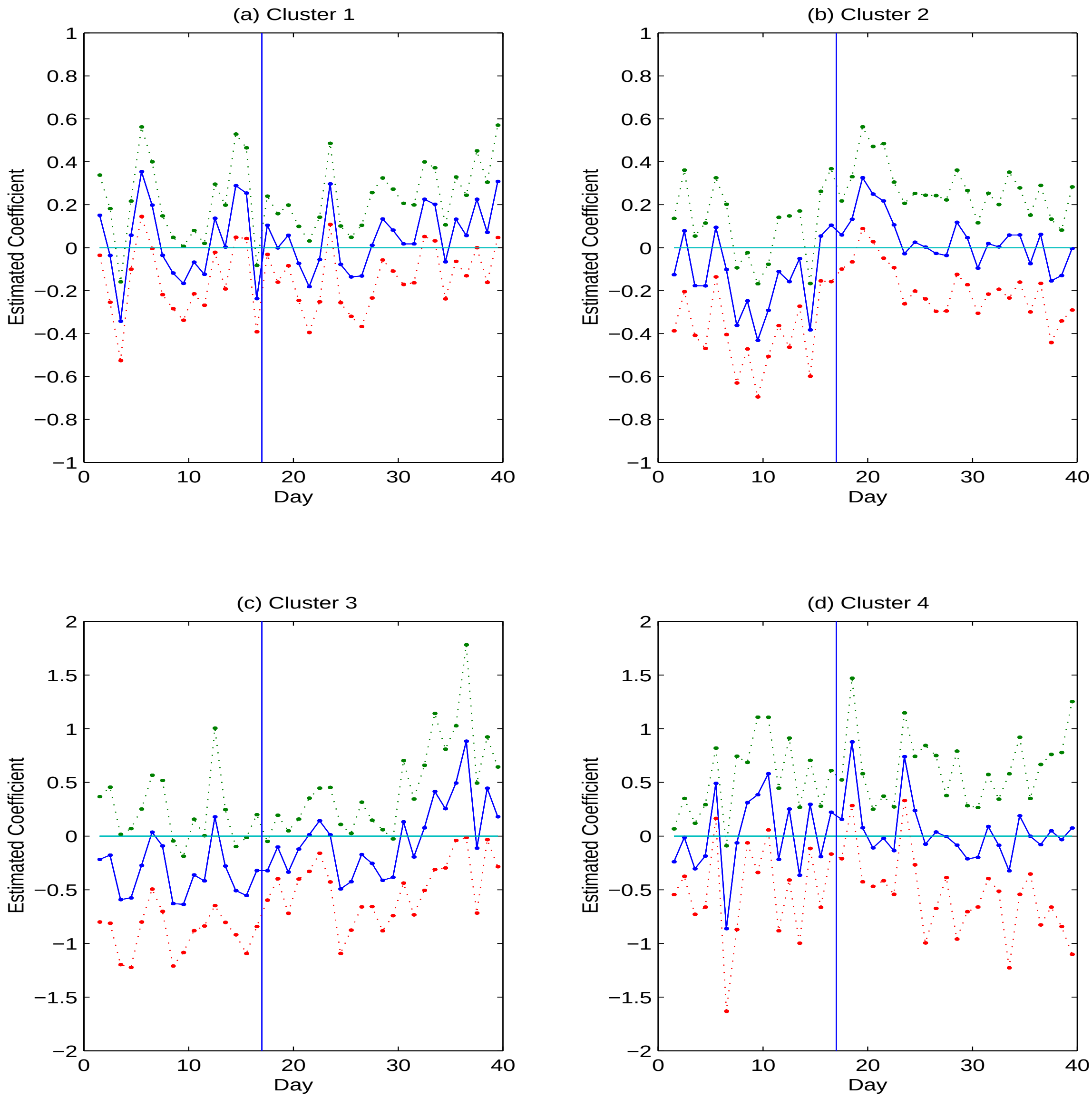

Figure 17: *Plot of Estimated Effects of Square of Attention. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

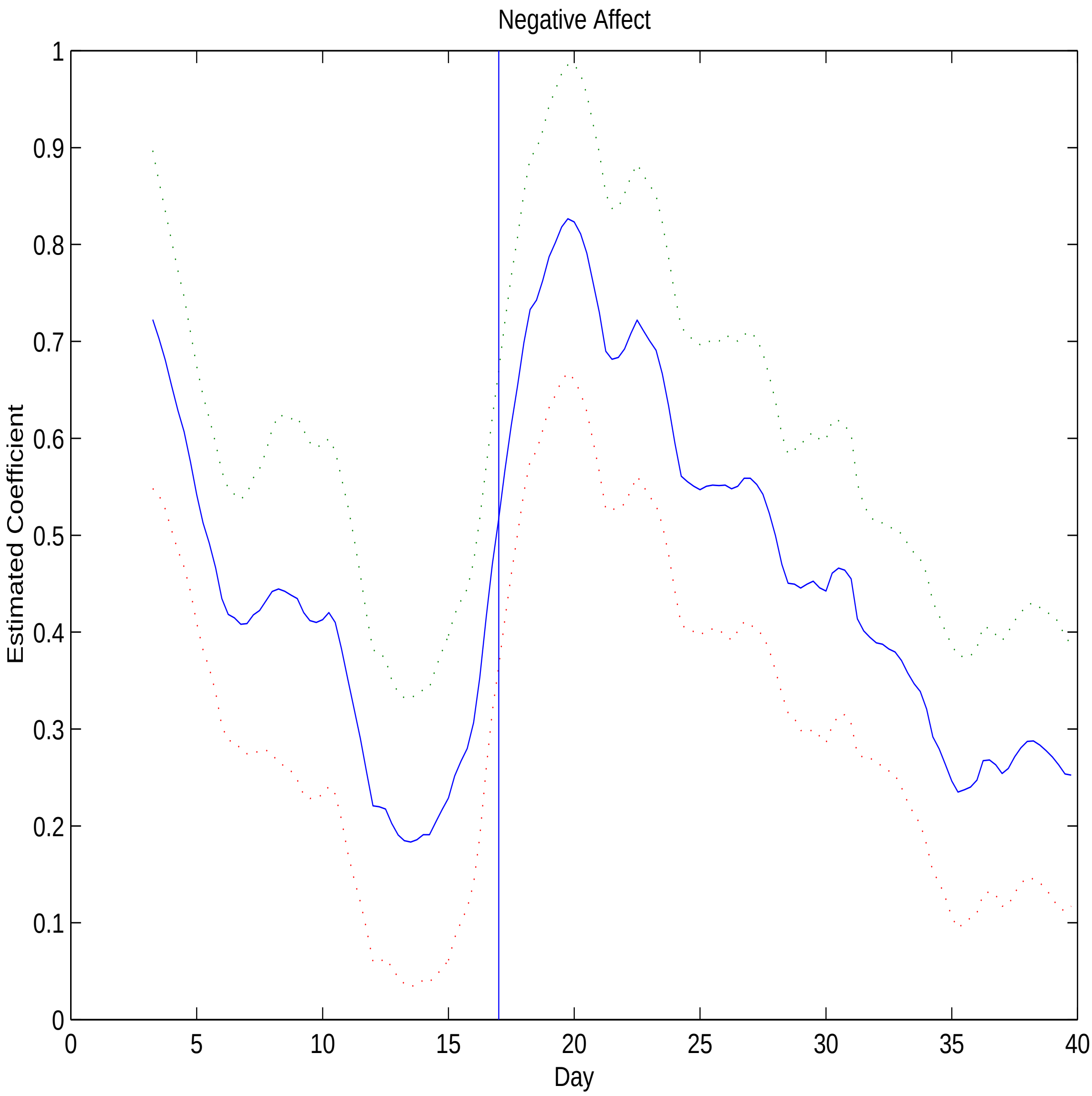

Figure 18: *Plot of Estimated Regression Function of Negative Affect. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

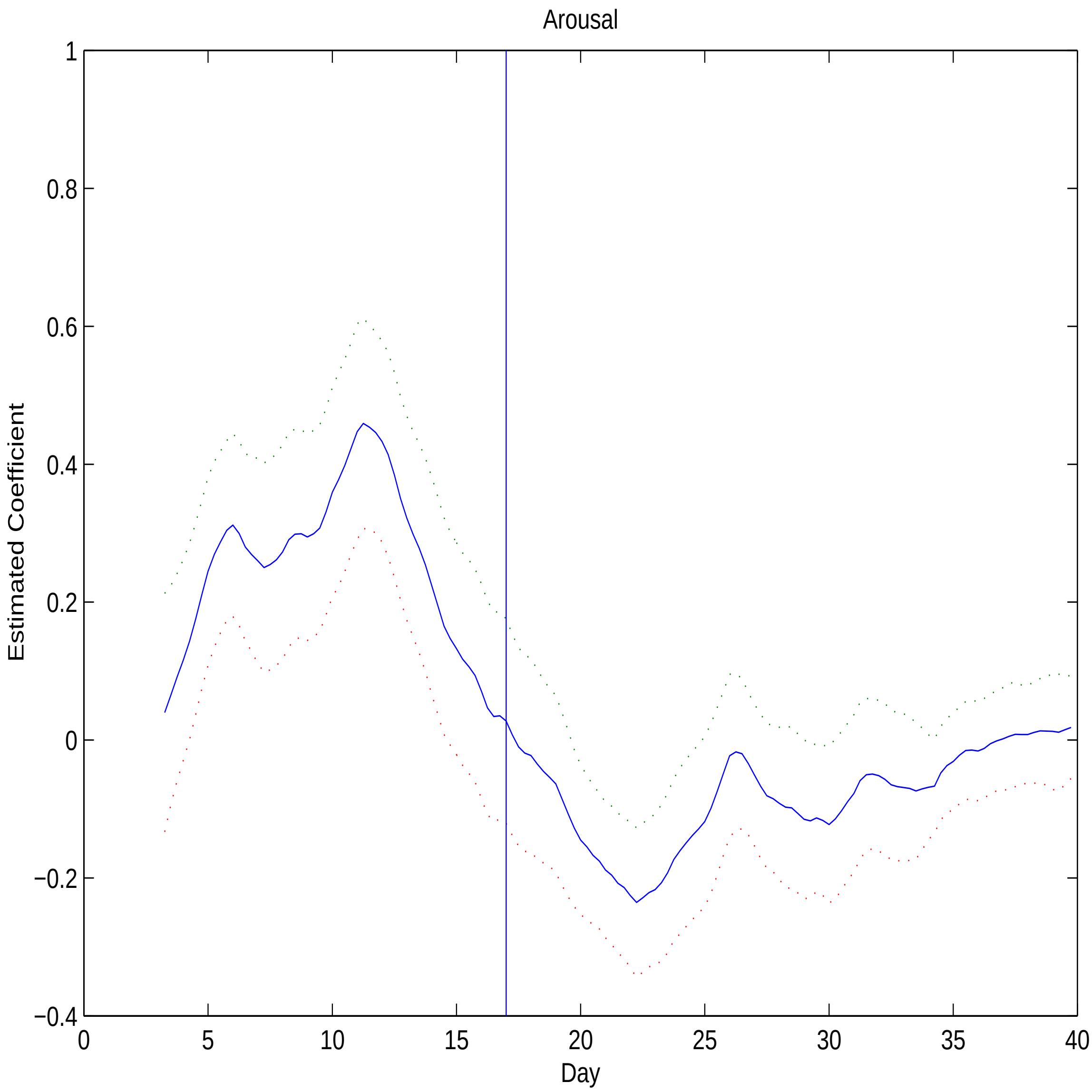

Figure 19: *Plot of Estimated Regression Function of Arousal. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

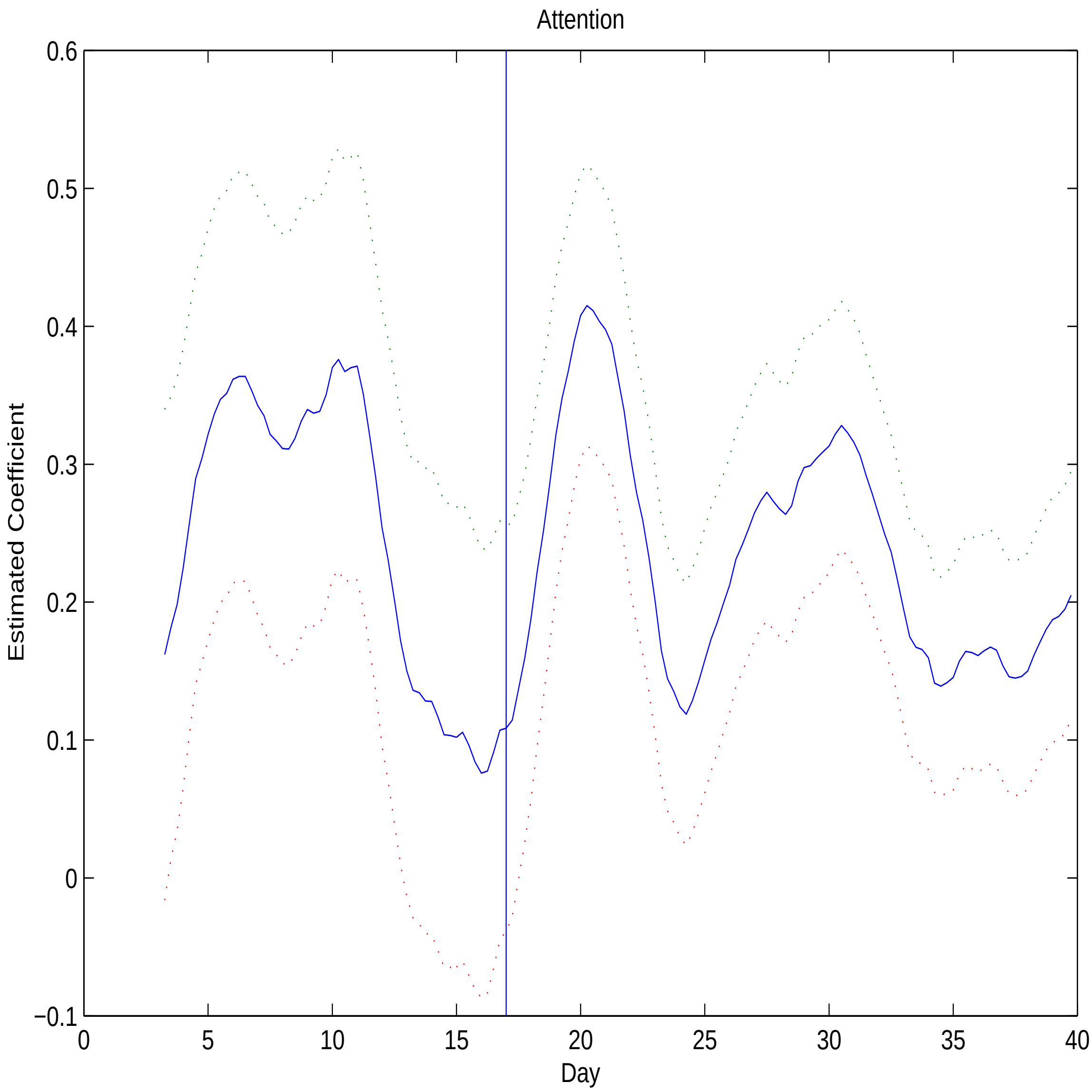

Figure 20: *Plot of Estimated Regression Function of Attention. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

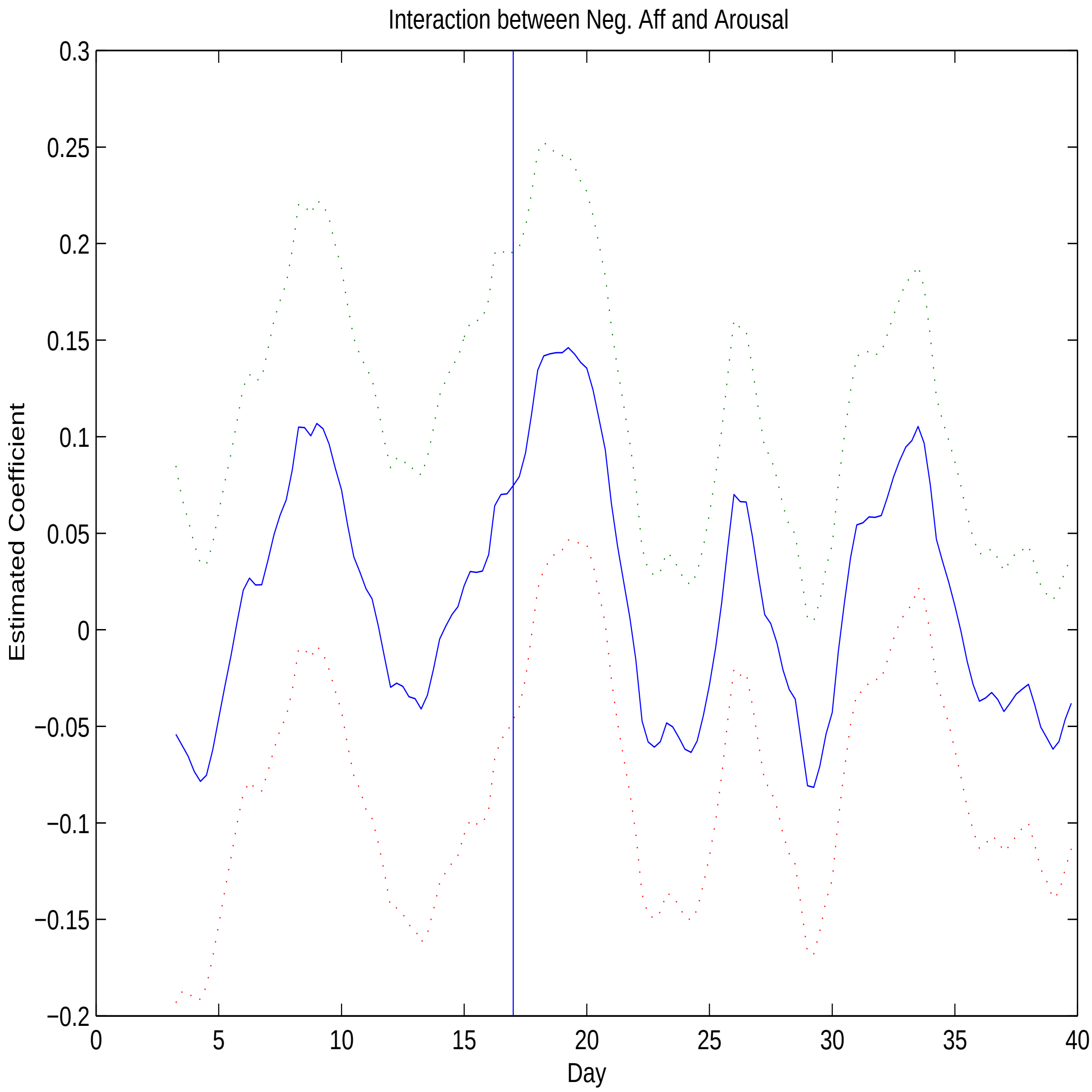

Figure 21: *Plot of Estimated Regression Function of Interaction between Negative Affect and Arousal. The solid line stands for estimated regression coefficient, and the dotted lines are 95% pointwise confidence interval.*

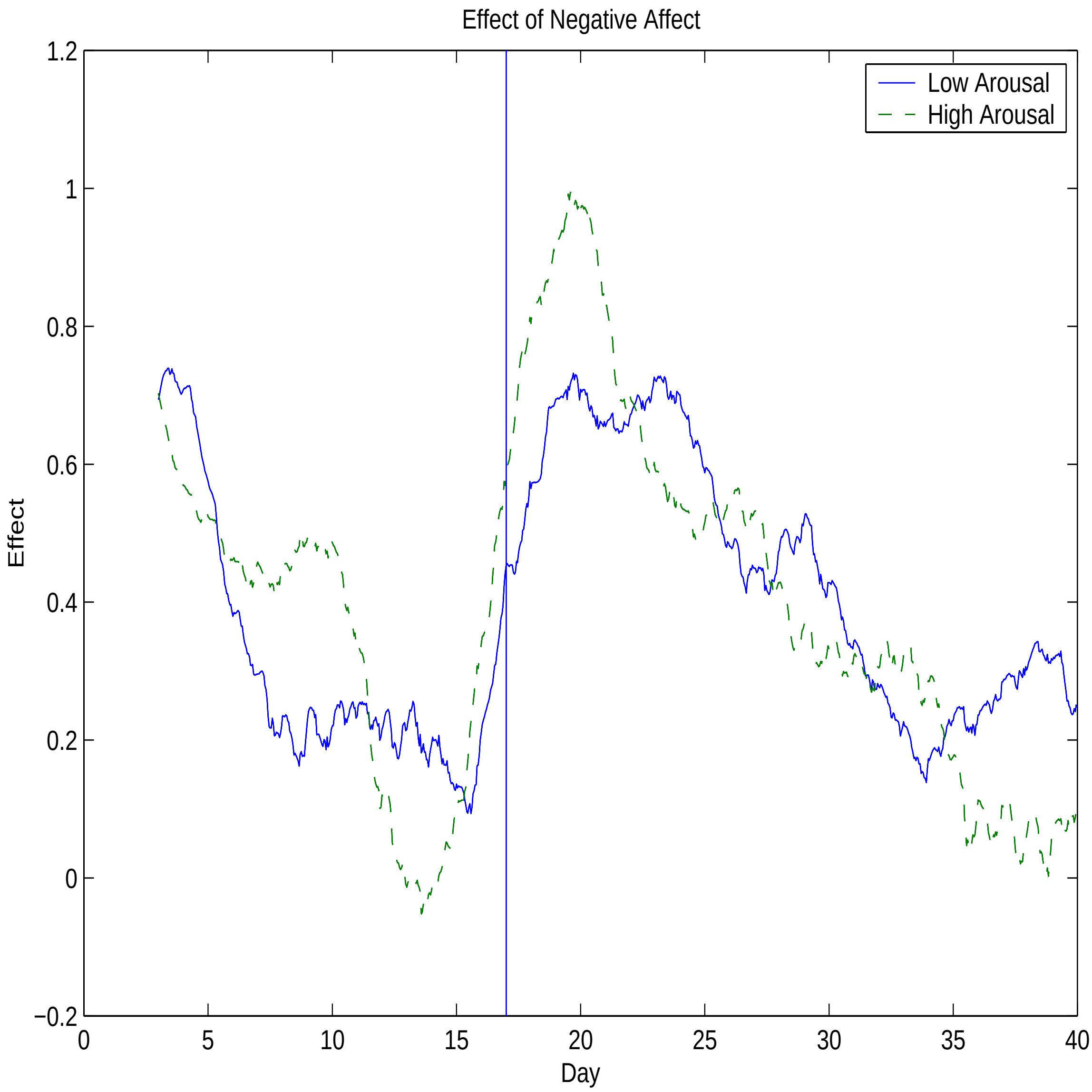

Figure 22: *Plot of Effects of Negative Affect.*

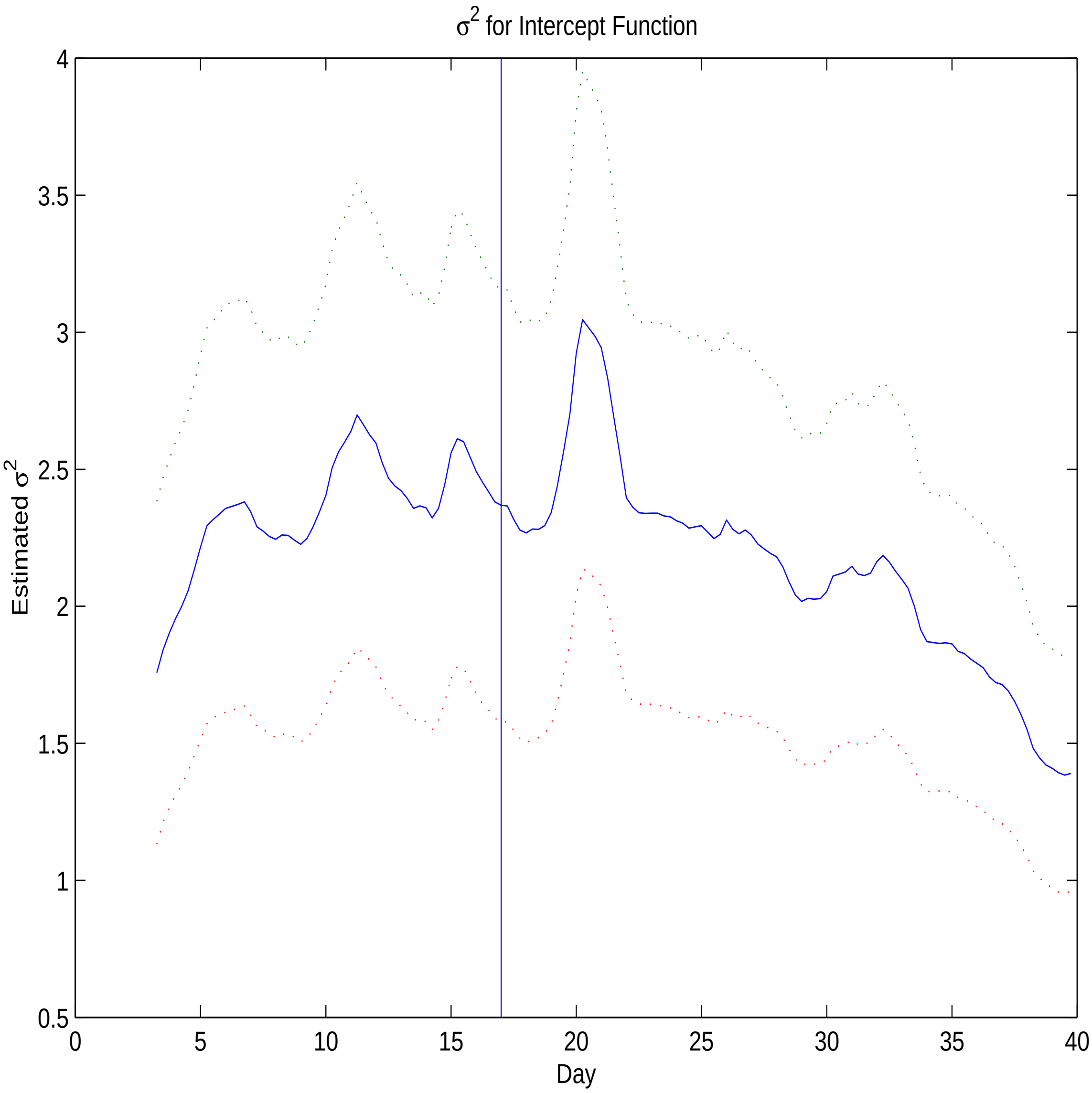

Figure 23: *Plot of $\widehat{\sigma}^2$ for Random Intercept. The solid line stands for estimated variance function, and the dotted lines are 95% pointwise confidence interval.*

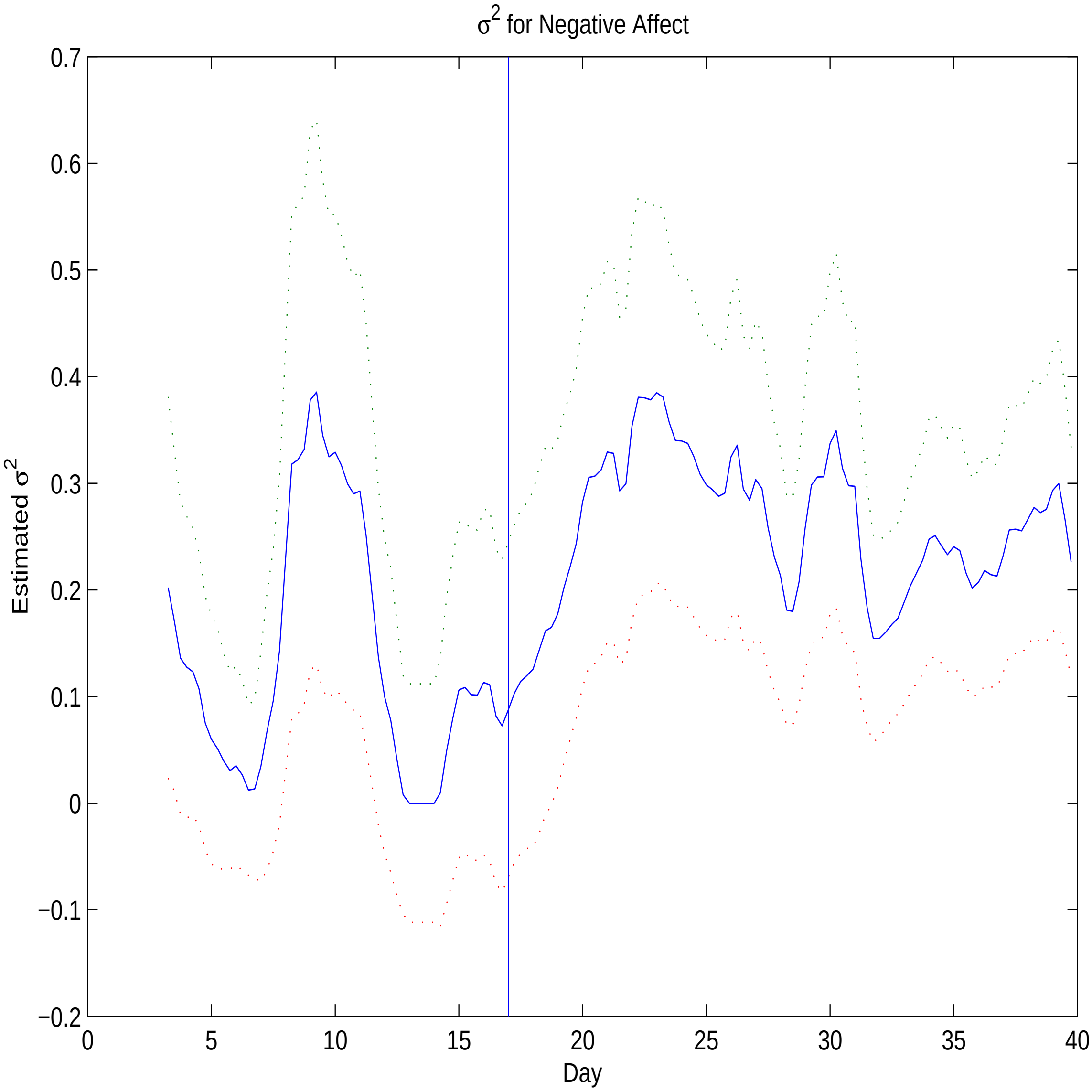

Figure 24: *Plot of $\widehat{\sigma}^2$ for Negative Affect. The solid line stands for estimated variance function, and the dotted lines are 95% pointwise confidence interval.*

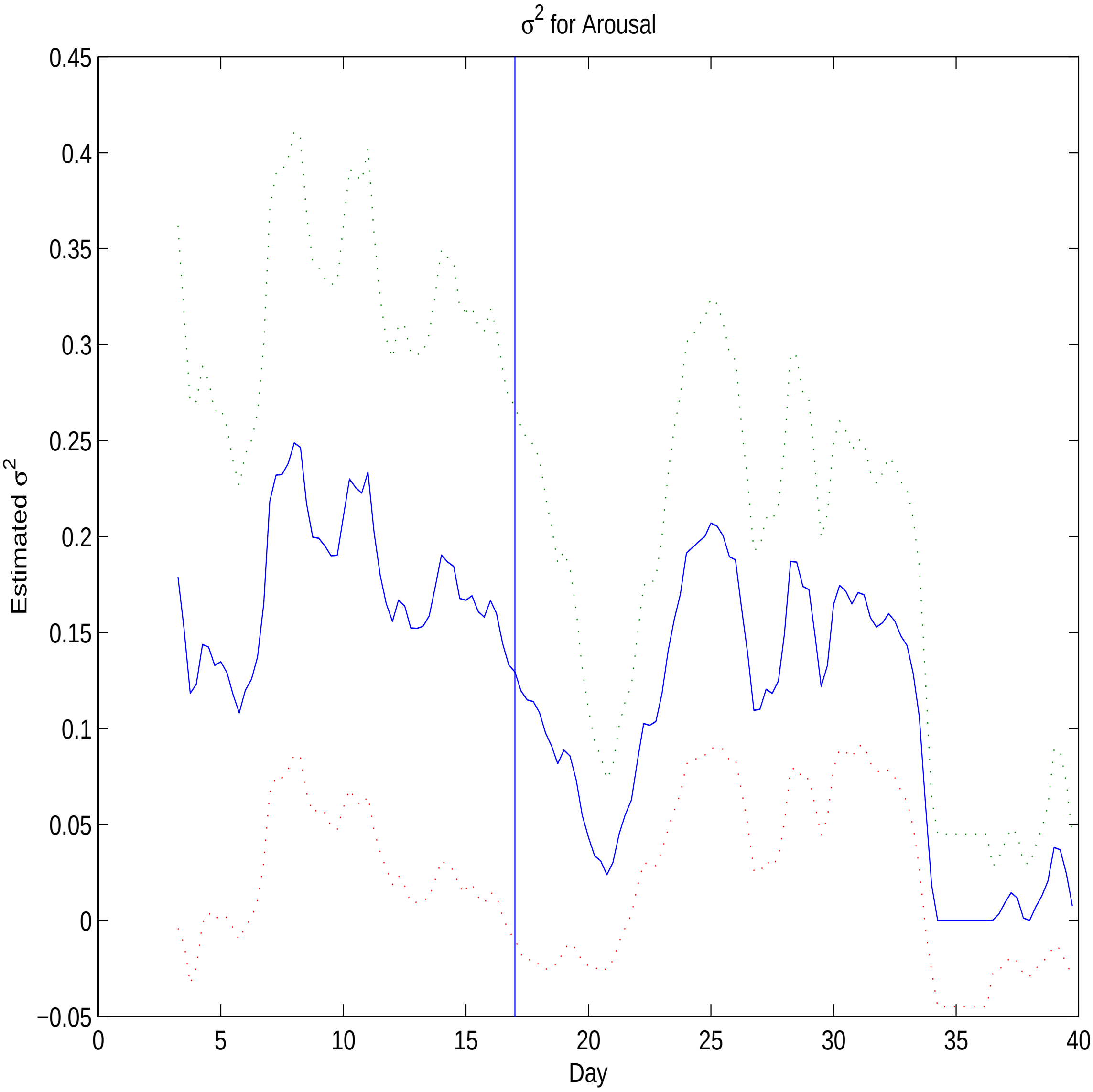

Figure 25: *Plot of $\widehat{\sigma}^2$ for Arousal. The solid line stands for estimated variance function, and the dotted lines are 95% pointwise confidence interval.*

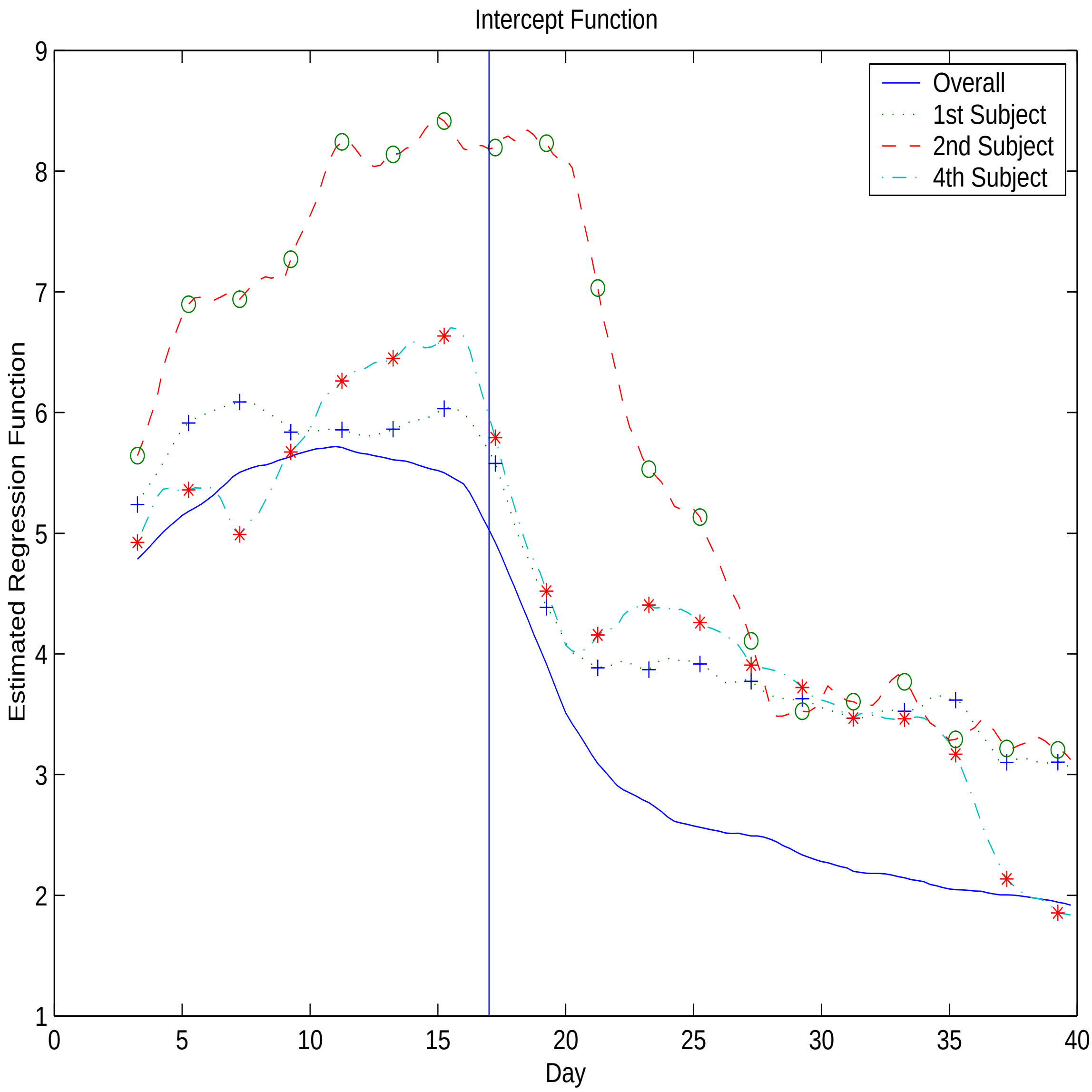

Figure 26: *Plot of* $\widehat{\beta}_0(t) + \widehat{\gamma}_{i0}(t)$.

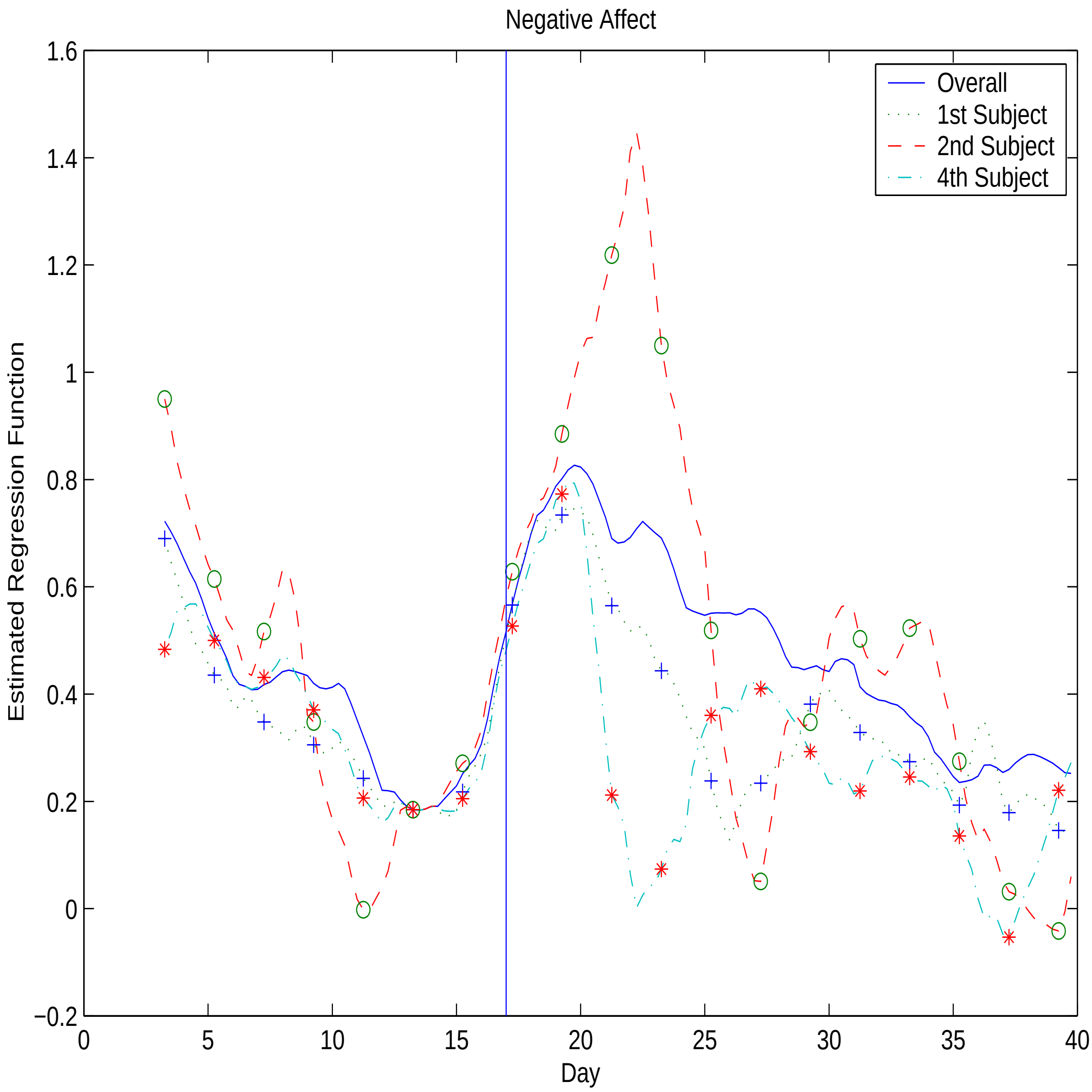

Figure 27: *Plot of* $\widehat{\beta}_1(t) + \widehat{\gamma}_{i1}(t)$.

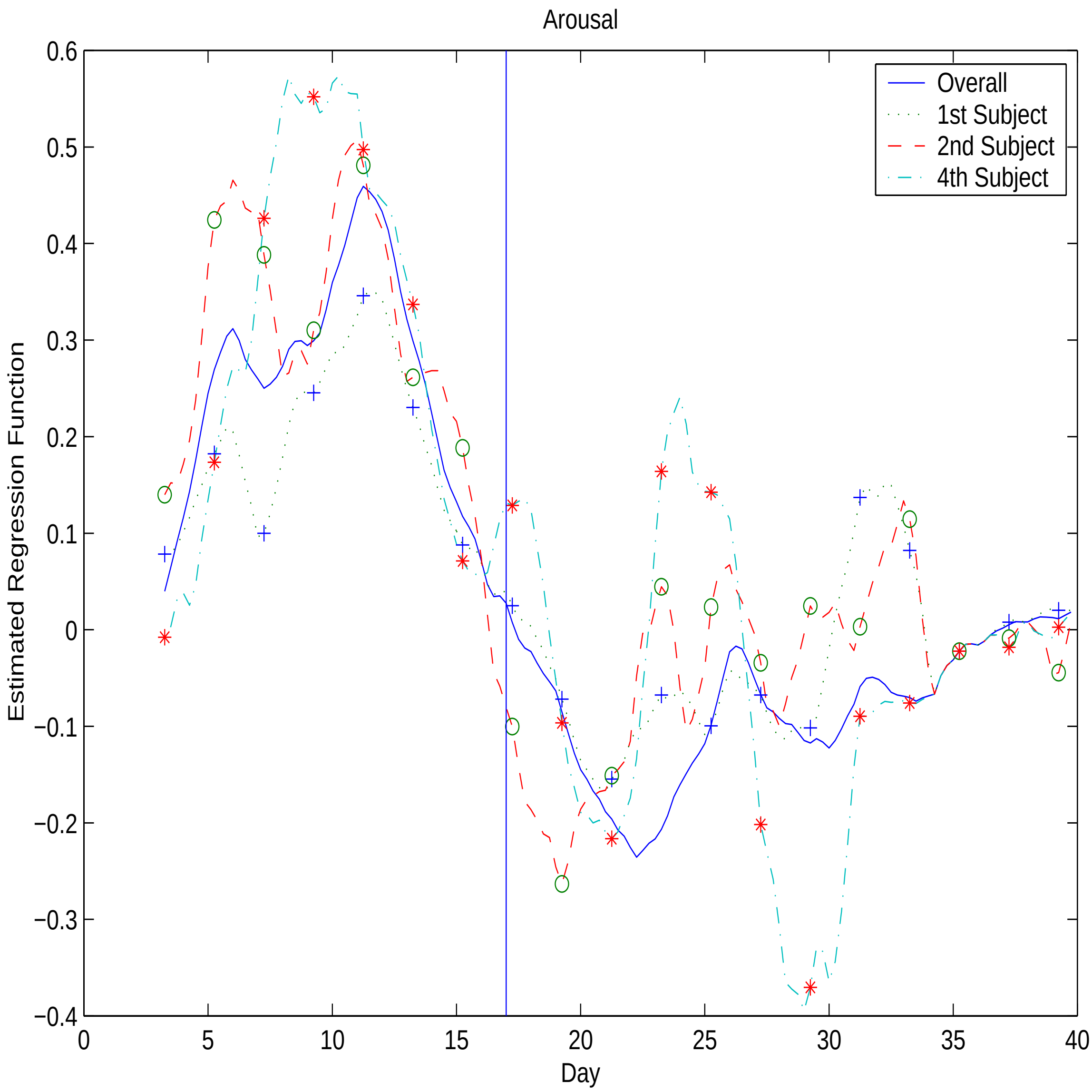

Figure 28: *Plot of* $\widehat{\beta}_2(t) + \widehat{\gamma}_{i2}(t)$.

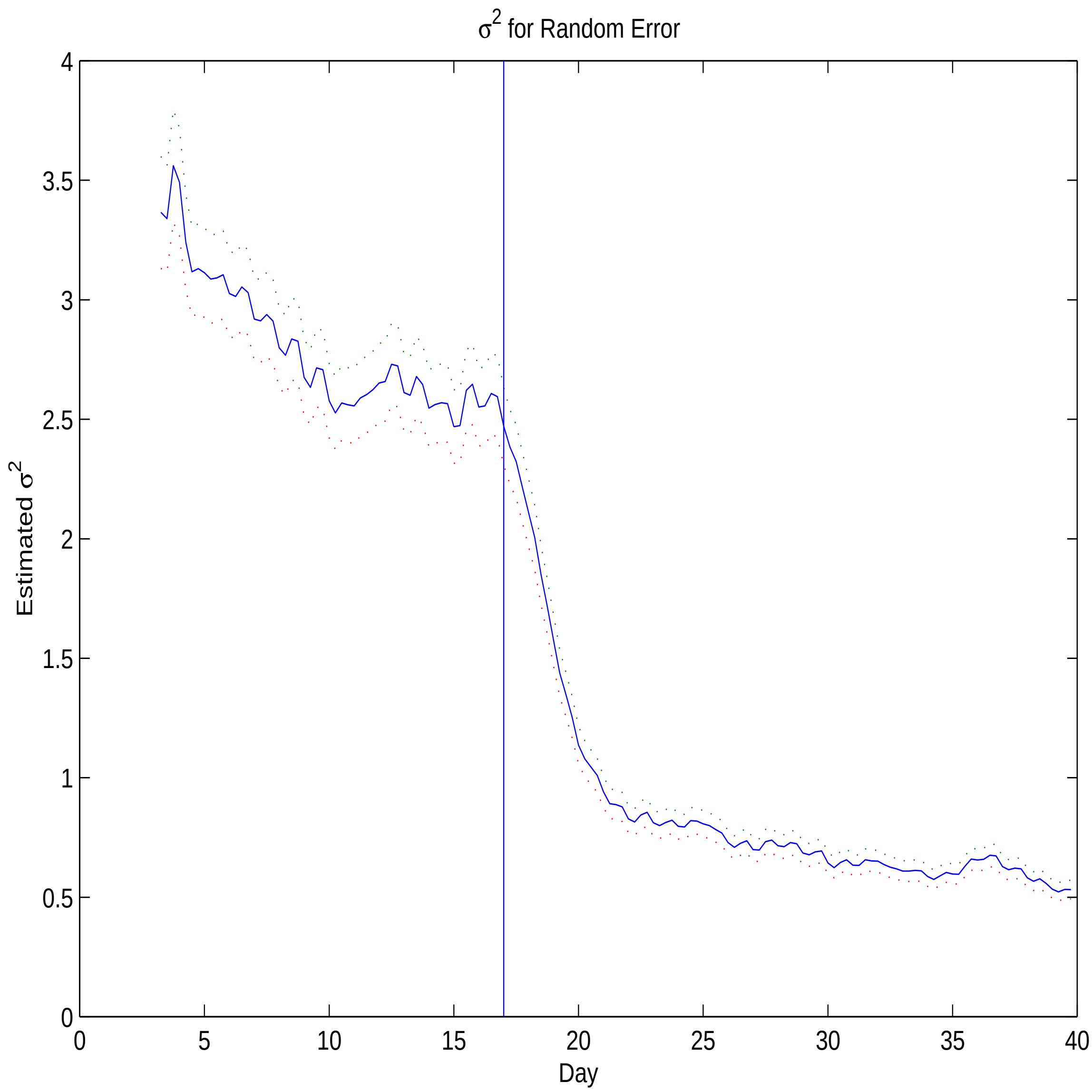

Figure 29: *Plot of $\widehat{\sigma}^2$ for Random Error. The solid line stands for estimated variance function, and the dotted lines are 95% pointwise confidence interval.*