Semiparametric Modeling for Longitudinal Data Analysis

Jianqing Fan

Chinese University of Hong Kong

joint work with Runze Li

http://www.stat.unc.edu/faculty/fan.html

October 4, 2002
Longitudinal Data

**Data:** From individual treatment effects over time.

**Nature:** Highly unbalanced, observing at irregular time points.

**Time:** $i^{th}$ individual observed at time $t_{ij}$, $j = 1, \cdots, J_i$.

**Collected data:** covariate vector $\mathbf{x}_i(t)$ along with its associated response $y_i$ at time $t_{ij}$:

\[
\{(t_{ij}, \mathbf{x}_i(t_{ij}), y(t_{ij})), j = 1, \cdots, J_i\}
\]

**Notation:** $\mathbf{y}_i = (y_i(t_{i1}), \cdots, y_i(t_{iJ_i}))^T$, $\mathbf{X}_i = (\mathbf{x}_i(t_{i1}), \cdots, \mathbf{x}_i(t_{iJ_i}))^T$. 
Example 1 — CD4 data

Measurements (from multi-center AIDS cohort study): CD4-cell percentages and other important covariates for 283 homosexual men infected during 84-91 were measured over a period of time in order to monitor AIDS progression (Kaslow et al. 1987).

Covariates: $X_1(t) =$ Smoking, $X_2(t) =$ age, $X_3(t) =$ preCD4.

Missing data: Very heavy; fewer data points at the end

Question: — How to model $X(t)$ and $Y(t)$?

— How to predict individual’s progression?
Example 2 — Hormone Data

**Measurements:** urinary metabolite progesterone curves measured over 21 conception and 70 nonconception women menstrual cycles.

**Nesting:**

\[
\begin{array}{c|c}
\text{conception} & \text{nonconception} \\
\text{subject} & \text{subject} \\
\text{subject} & \text{subject} \\
\text{cycle} & \text{cycle} \\
\text{cycle} & \text{cycle} \\
\end{array}
\]

**Alignment:** aligned and truncated around the day of ovulation.

**Data:** \( Y_{ijk}(t) \) with missing values in some cycles.

**Question:** How to analyze nested functional data?
Statistical models (I)

Classical Linear Models: [Diggle, Liang and Zeger (1994) and Hand and Crowder (1996)]

\[ y(t) = \beta_0 + \beta_1 x_1(t) + \cdots + \beta_p x_p(t) + \varepsilon(t) \]

Coefficients are independent of time \( t \)

**Method:** Weighted LS using a working covariance matrix \( W_i \):

\[
\sum_{i=1}^{n} (y_i - \alpha_i - X_i \beta)^T W_i (y_i - \alpha_i - X_i \beta)
\]

A semiparametric model: allowing baseline to be time-dependent

\[ y(t) = \beta_0(t) + \beta_1 x_1(t) + \cdots + \beta_p x_p(t) + \varepsilon(t) \]

Statistical models (II)

**Functional linear model:** (Varying-coefficient model)

\[
Y(t) = \beta_0(t) + \beta_1(t)X_1(t) + \cdots + \beta_p(t)X_p(t) + \varepsilon(t)
\]


**Comparisons:**

— Linear model is too restrictive, can not capture time effect.

— Functional linear model is very flexible and understandable, but its coefficients can not be well estimated (collinearity), and are not as interpretable.

— The semiparametric model falls between these two. The \( \beta \) admits similar interpretation. It captures some time effect but not as flexible as the FL model.
Lin and Ying’s approach

Model: \( y(t) = \alpha(t) + \beta^T x(t) + \varepsilon(t) \).

Counting process: \( N_i(t) \equiv \sum_{j=1}^{J_i} I(t_{ij} \leq t) \)

Assumption’s:

(a) **Observed time:** \( N_i(t) = N_i^*(t \wedge c_i) \) (somewhat artificial)

(b) **Noninform. censoring:** \( E\{y_i(t)|x_i(t), c_i \geq t\} = E\{y_i(t)|x_i(t)\} \).

(c) **Observations:** The trajectories \([x(t), y(t)]\) are fully observable until the censoring time \(c_i\). This is unrealistic, and can be implemented by **linear interpolation** or **constant interpolation**

(d) **Working independence:** \( W_i \) is diagonal.
LY method — independence

**Assumption**: Observation times are independent of $\mathbf{x}(t)$.

**Direct approach**: Regarding $\alpha(t)$ as a parameter, minimize

$$
\sum_{i=1}^{n} \int_{0}^{+\infty} w(t) \{y_i(t) - \alpha(t) - \beta^T \mathbf{x}_i(t)\}^2 dN_i(t) = \sum_{i} \left\{ \sum_{j=1}^{J_i} w(t_{ij}) \{y_i(t_{ij}) - \cdots \} \right\}
$$

**Baseline est.**: $\hat{\alpha}(t; \beta) = \bar{y}(t) - \beta^T \bar{\mathbf{x}}(t)$ with $\xi_i(t) = I(c_i \geq t)$

$$
\bar{\mathbf{x}}(t) = \frac{\sum_{i=1}^{n} \xi_i(t) \mathbf{x}_i(t)}{\sum_{i=1}^{n} \xi_i(t)}, \quad \bar{y}(t) = \frac{\sum_{i=1}^{n} \xi_i(t) y_i(t)}{\sum_{i=1}^{n} \xi_i(t)}
$$

**Parameter estimate**: Minimize

$$
\sum_{i=1}^{n} \int_{0}^{+\infty} w(t) \{y_i(t) - \bar{y}(t)\} - \beta^T \{\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\} \}^2 dN_i(t).
$$
LY method — dependence

Intensity: \( E\{dN_i^*(t)\mid x_i(t), y_i(t), c_i \geq t\} = \lambda(t) \exp\{\gamma' x_i(t)\} \) (Cox model). \( \gamma = 0 \) corresponds to the independence case.

Baseline estimation: \( \hat{\alpha}(t; \beta) = \bar{y}(t; \gamma) - \beta^T \bar{x}(t; \gamma) \) with

\[
\bar{x}(t, \gamma) = \frac{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^T x_i(t)\} x_i(t)}{\sum_{i=1}^{n} \xi_i(t) \exp\{\gamma^T x_i(t)\}},
\]

Parameter estimation: Minimize

\[
\sum_{i=1}^{n} \int_{0}^{+\infty} w(t)\{y_i(t) - \bar{y}(t; \gamma)\} - \beta^T \{x_i(t; \gamma) - \bar{x}(t; \gamma)\} \}^2 dN_i(t).
\]

Estimation of nuisance parameter: Solving the partial likelihood equation (see Pepe and Cai, 1993):

\[
\sum_{i=1}^{n} \int_{0}^{+\infty} \{x_i(t) - \bar{x}(t, \gamma)\} dN_i(t) = 0.
\]
Remarks

— LY method is simple and does not involve any smoothing.

— No smoothness is used in estimation. Hence, the efficiency can be significantly improved.

— Interpolation can create large biases, making procedures inconsistency.

— Counting process formulation and its nuisance parameters estimation are cumbersome and artificial.

— The difference based estimator overcome all disadvantages and does not involve any smoothing. It is useful for smoothing in the profile least-squares estimator.
**Difference based method**

**Idea**: From Fan and Huang (2001), Yatchew (1997). Order

\[
\{(t_{ij}, x(t_{ij})^T, y(t_{ij})), j = 1, \cdots, J_i, i = 1, \cdots, n\}
\]

according to \(\{t_{ij}\}\) as \(\{(t_i, x_i^T, y_i), i = 1, \cdots, n^*\}\), with \(n^* = \sum_{i=1}^{n} J_i\).

**Marginal model**: \(y_i = \alpha(t_i) + \beta^T x_i + \varepsilon_i\).

**Differencing**: The nuisance function eliminated by differencing:

\[
y_{i+1} - y_i = \alpha(t_{i+1}) - \alpha(t_i) + \beta^T (x_{i+1} - x_i) + e_i \approx \beta^T (x_{i+1} - x_i) + e_i.
\]

**Parameter estimation**: Applied weighted LS to

\[
y_{i+1} - y_i = \alpha_0 + \alpha_1 (t_{i+1} - t_i) + \beta^T (x_{i+1} - x_i) + e_i.
\]

**Limited lost of efficiency**: among working independence estimators. Pretend independence. The data \(\{y_{2i+1} - y_{2i}\}\) are independ.

Lose only \(\{y_{2i+1} + y_{2i}\}\), containing less information about \(\beta\).
Profile least-squares (I)

**Nonparametric regression:** Letting $y^*(t) \equiv y(t) - \beta^T x(t)$, we have

$$y^*(t) = \alpha(t) + \varepsilon(t).$$

**Local linear fit:** For each given $t_0$, approximate

$$\alpha(t) \approx \alpha(t_0) + \alpha'(t_0)(t - t_0) \equiv a + b(t - t_0).$$

Minimize with respect to $a$ and $b$

$$\sum_{i=1}^{n} \sum_{j=1}^{J_i} \{y_i^*(t_{ij}) - a - b(t_{ij} - t_0)\}^2 w(t_{ij}) K_h(t_{ij} - t_0),$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K$ is a kernel and $h$ is a bandwidth, resulting in $\hat{\alpha}(t_0; \beta) = \hat{a}$.

**Notation.** $y = (y_1^T, \cdots, y_n^T)$, $X = (X_1^T, \cdots, X_n^T)^T$ and $\alpha = (\alpha_1^T, \cdots, \alpha_n^T)^T$. 


Profile least-squares (II)

**Marginal model:** \( y = \alpha + X\beta + \varepsilon. \)

**Local linear estimator:** linear in response \( y - X\beta. \) Hence, \( \hat{\alpha} = S(y - X\beta), \) where \( S \) is a smoothing matrix.

**Synthetic model:** \( (I - S)y = (I - S)X\beta + \varepsilon. \)

**PLS:** \( \hat{\beta} = \{X^T(I - S)^T W(I - S)X\}^{-1}X^T(I - S)^T W(I - S)y \)

**Estimated covariance matrix:** \( \text{cov}\{\hat{\beta}|t_{ij}, x_i(t_{ij})\} = D^{-1}VD^{-1}, \)

where \( D = X^T(I - S)^T W(I - S)X \) and \( V = \text{cov}\{X^T(I - S)^T W\varepsilon\}. \)

The matrix \( V \) can be estimated as \( \hat{V} = X^T(I - S)^T WCW^T(I - S)X \)

with \( C = \text{diag}\{\hat{\varepsilon}_1\hat{\varepsilon}_1^T, \ldots, \hat{\varepsilon}_n\hat{\varepsilon}_n^T\}. \)
Implementations

**Sparsity**: The function $\alpha(\cdot)$ can not be estimated well at both tails. We only estimate function at inter 90% of the data.

**Bandwidth selection**: Use DBE to get an estimate $\hat{\boldsymbol{\beta}}_{DBE}$. Letting $y^*(t) = y(t) - \hat{\boldsymbol{\beta}}_{DBE}^T \mathbf{x}(t)$, we have $y^*(t) \approx \alpha(t) + \varepsilon(t)$. Apply a bandwidth selection method to get an $\hat{h}$. Using this $\hat{h}$ to obtain the profile least-squares estimator and the estimator of $\alpha(\cdot)$.

**Methods**: Cross-validation, pre-asymptotic substitution method (Fan and Gijbels, 1995), asymptotic substitution method (Ruppert *et al*, 1995), empirical bias method (Ruppert, 1997), among others.
Simulation Models

Model: \( y(t) = \alpha(t) + \beta^T x(t) + \varepsilon(t) \), where \( \alpha(t) = \tau \sqrt{t/\tau} \) or \( \tau \sin(2\pi t/\tau) \), \( \beta = (3, 1.5, 0, 0, 2, 0, 0, 0)^T \)

\( \varepsilon(t) \) is Gaussian proc. with cov. \( E\{\varepsilon(s)\varepsilon(t)\} = \exp(-2|t - s|) \)

\( x \sim N(0, \Sigma) \) with cov\( (x_i, x_j) = 0.5|i-j| \).

Design of observation time

Case I: LY design — independence. The process \( N^*(t) \) was generated from a random effect Poisson process with intensity rate \( \eta \), where \( \eta \sim \text{Gamma}(1, 0.5) \). The censoring \( c \sim \text{unif}(0, \tau) \) with \( \tau = 4 \) or 20, resulting in, on average, 3 or 11 observations per subject.

Case II: LY design — dependence. The intensity function is of form

\[ \eta \exp(0.5x_1), \text{ instead of } \eta. \]
**case III**: Observation time to be $t_{ij} = 0, 1, \cdots, [c_i]$, with $c_i \sim \text{unif}(0, \tau)$, where $\tau = 10$ or $20$. On average, 6 or 11 obs. per subject.

**case IV** Huang, Hu and Zhou (2002) design: Each subject has a set of ‘scheduled’ time $\{0,1,3,\ldots,29\}$, and each scheduled time has a probability of being skipped 60%. For non-skipped time, a uniform $[-1, 1]$ is added to each scheduled time.

**Objective**: Performance comparison and accuracy of the Sandwich formula.

**Performance measure**: $\text{MSE} = E\|\hat{\beta} - \beta\|^2$. 
Numerical Results — performance comparison

Table 1: Relative Efficiency — Ratios of MSEs with respect to LY estimator

<table>
<thead>
<tr>
<th>Case I</th>
<th></th>
<th>n = 50</th>
<th>n = 75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha(t)$</td>
<td>$\tau$ DBE</td>
<td>$\tau$ Profile LSE</td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>4</td>
<td>0.8481</td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>20</td>
<td>0.3450</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>4</td>
<td>0.6632</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>20</td>
<td>0.2798</td>
</tr>
<tr>
<td>Case II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>4</td>
<td>0.7868</td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>20</td>
<td>0.2518</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>4</td>
<td>0.5627</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>20</td>
<td>0.1705</td>
</tr>
<tr>
<td>Case III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>10</td>
<td>1.0785</td>
</tr>
<tr>
<td></td>
<td>$\tau \sqrt{t/\tau}$</td>
<td>20</td>
<td>0.7748</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>10</td>
<td>1.2188</td>
</tr>
<tr>
<td></td>
<td>$\tau \sin(2\pi t/\tau)$</td>
<td>20</td>
<td>0.9868</td>
</tr>
<tr>
<td>Case IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$30\sqrt{t/30}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$30 \sin(2\pi t/30)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of simulations: 400.

Remark: The deteriorated performance in case 3 is due to wide gap between observation times.
Numerical results— baseline estimation

Remarks. Typical estimated baseline curves with \( n = 50 \) and \( \tau = 20 \). Right panel \( \alpha(t) = \tau \sqrt{t/\tau} \) for case I and case II, respectively. Left panel for \( \alpha(t) = \tau \sin(t/\tau) \) for Cases I and II, respectively.
Numerical results — accuracy of SE

Table 2: Stds and SEs of Profile LSE for Case I with $\alpha(t) = \tau \sqrt{t/\tau}$

<table>
<thead>
<tr>
<th>$(n, \tau)$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD  se (SD(se))</td>
<td>SD  se (SD(se))</td>
<td>SD  se (SD(se))</td>
</tr>
<tr>
<td>(50, 4)</td>
<td>0.1512 0.1377(0.0327)</td>
<td>0.1683 0.1579(0.0369)</td>
<td>0.1664 0.1543(0.0381)</td>
</tr>
<tr>
<td>(75, 4)</td>
<td>0.1200 0.1148(0.0211)</td>
<td>0.1262 0.1273(0.0243)</td>
<td>0.1274 0.1287(0.0240)</td>
</tr>
<tr>
<td>(50, 20)</td>
<td>0.0854 0.0820(0.0182)</td>
<td>0.1004 0.0910(0.0211)</td>
<td>0.1012 0.0933(0.0203)</td>
</tr>
<tr>
<td>(75, 20)</td>
<td>0.0651 0.0675(0.0130)</td>
<td>0.0718 0.0748(0.0144)</td>
<td>0.0708 0.0749(0.0149)</td>
</tr>
</tbody>
</table>

Note. — The relative Monte Carlo error is of size $1/\sqrt{800}$ for the sample SD. It is negligible.

— The discrepancy is less than half of SD(se).

— Results for other cases are similar.
Asymptotic results

Formulation: The asymptotic result depends on the formulation.

Follow Lin and Ying’s formulation for comparison.

Notation. Let \( a^{\otimes 2} = aa^T \) and

\[
A = E \int_0^\infty \{x(t) - E x(t)\}^{\otimes 2} w(t)dN(t)
\]

\[
B = E \left\{ \int_0^\infty \{x(t) - E x(t)\} \varepsilon(t)w(t)dN(t) \right\}^{\otimes 2}
\]

Results: If \( h_n = bn^{-a} \), for \( 1/8 < a < 1/2 \), then as \( n \to \infty \),

\[
\sqrt{n}(\hat{\beta} - \beta_0) \overset{L}{\to} N(0, A^{-1}BA^{-1}).
\]

Consistency of Sandwich Formula: \( D^{-1} \hat{V} D^{-1} \) consistently estimates \( A^{-1}BA^{-1} \).
Variable selection

**Background**: To reduce modeling biases, large parametric models are introduced.

**Questions**: How to

— automatically select significant variables?

— construct confidence intervals?

— verify properties of data-driven procedures?

**Challenge**: Not yet studied neither in semiparametric models nor longitudinal data.

**Traditional Approaches**: stepwise procedure, best subset, · · ·

**Drawbacks**: — hard to establish sampling properties;

— expensive in computation
A viable solution

Setting: i.i.d samples first (see Fan and Li, 2001; 2002).

— Penalized likelihood: \( \ell(X\beta, Y) - \sum_j p_\lambda(|\beta_j|) \) where e.g.
\[
p_\lambda(|\beta_j|) = \lambda |\beta_j|.
\]

— Sandwich formula based on the penalized likelihood.

— Oracle property: Suppose that
\[
X\beta = X_1^T\beta_1 + X_2^T\beta_2 \text{ with } \beta_2 = 0.
\]
The oracle estimator: \( \hat{\beta}_2 = 0 \) and \( \hat{\beta}_1 \) being the MLE of submodel.

— What kind of \( p_\lambda(\cdot) \)?
Penalty functions

**Assumption:** Normal likelihood with orthonormal $\mathbf{X}$. Then, PLK is equivalent to

$$\|y - x\beta\|^2 + \sum_j p_{\lambda}(|\beta_j|) = c + \sum_j \{(z_i - \beta_j)^2 + p_{\lambda}(|\beta_j|)\},$$

where $\mathbf{z} = \mathbf{x}^T \mathbf{y}$ and $c = \|\mathbf{y} - \mathbf{xz}\|^2$.

**Componentwise minimization:** $(z_i - \theta)^2 + p_{\lambda}(|\theta|)$.

**$L_2$ penalty:** $p_{\lambda}(|\theta|) = \lambda|\theta|^2 \implies$ ridge regression

**Entropy penalty:** $p_{\lambda}(|\theta|) = \lambda I(|\theta| \neq 0) \implies$ Best subset

**Hard-thresholding penalty:** $p_{\lambda}(|\theta|) = \lambda^2 - (|\theta| - \lambda)^2 I(|\theta| < \lambda) \implies$ Best subset.

**$L_1$-penalty:** $p_{\lambda}(|\theta|) = \lambda|\theta| \implies$ soft-thresholding (Dohono and Johnstone, 1994) and LASSO (Tibshirani 1996, 97, Knight and Fu, 2000).
Other penalty functions

**SCAD**: $p'_\lambda(\theta) = I(\theta \leq \lambda) + \frac{(a\lambda-\theta)_+}{(a-1)\lambda} I(\theta > \lambda)$,

**Transformed $L_1$**: $p_\lambda = \lambda a|x|/(1+a|x|)$

**$L_p$-penalty**: $p_\lambda = \lambda|x|^p$
Desired Properties

— **continuity**: to avoid instability in model prediction

\[ \iff \arg\min_\theta \{ |\theta| + p'_\lambda(|\theta|) \} = 0 \]

— **sparsity**: to reduce model complexity \( \iff p'_\lambda(0+) > 0 \)

— **Unbiasedness**: to avoid unnecessary modeling bias \( \iff \lim_{|\theta| \to \infty} p_\lambda(|\theta|) = 0 \)

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Ridge</th>
<th>LASSO</th>
<th>SCAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Sparsity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Unbiasedness</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
Variable Selection for Semiparametric Model

**Weighted profile least-squares:** Minimize

\[
\ell(\beta) = \frac{1}{2}(y - X\beta)^T(I - S)^T W(I - S)(y - X\beta)
\]

**Variable selection:** Penalized WPLS \( \mathcal{L}(\beta) = \ell(\beta) + n \sum_{j=1}^{d} p_\lambda(|\beta_j|) \)

It can be obtained from the penalized likelihood via profiling.

**Computation:** Use the iterated ridge regression (modified Newton-Raphson method)

**Covariance matrix** can be estimated from a modified sandwich formula.

**Choice of regularization parameter** can be obtained by a form of GCV.
Sampling Properties

**Notation:** $a_n = \max_j \{|p'_{\lambda_n}(|\beta_{j0}|)| : \beta_{j0} \neq 0\}$

and $b_n = \max_j \{|p''_{\lambda_n}(|\beta_{j0}|)| : \beta_{j0} \neq 0\}$. Note that $a_n = b_n = 0$ for SCAD and hard-thresholding penalty and $a_n = \lambda_n, b_n = 0$ for $L_1$.

**Rate of convergence:** If $a_n \to 0$ and $b_n \to 0$ then there exists a local minimizer $\hat{\beta}$ of $\mathcal{L}(\beta)$ such that $||\hat{\beta} - \beta_0|| = O_P(n^{-1/2} + a_n)$.

**Significant Variables:** WOLG $\beta_0 = (\beta_{10}, 0)$.

**Oracle properties:** If $\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to \infty$, then the above root-n consistent estimator satisfies

— (Sparsity) $\hat{\beta}_2 = 0$;

— (Normality) $\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \to N_s(0, A_{11}^{-1}B_{11}A_{11}^{-1})$

**Remark.** No oracle property of $L_1$-penalty.
**Evaluation of the Procedure**

**Criterion** — Prediction error. Letting \( \{\tilde{x}(t), \tilde{y}(t), \tilde{N}(t)\} \) be a new observation,

\[
PE(\hat{\alpha}, \hat{\beta}) = E \int_0^\infty \{\tilde{y}(t) - \hat{\alpha}(t) - \hat{\beta}^T \tilde{x}(t)\}^2 d\tilde{N}(t).
\]

\[PE = \text{Noise error} + \text{error of } \hat{\alpha} + \text{error of } \hat{\beta} + \ldots\]

**Generalize MSE**: The effectiveness of \( \hat{\beta} \) is assessed via

\[
\text{GMSE} = (\hat{\beta} - \beta_0)^T \left\{ \int_0^\infty E\tilde{x}(t)^\otimes 2 \exp\{\gamma^T \tilde{x}(t)\} \xi(t) d\Lambda(t) \right\} (\hat{\beta} - \beta_0).
\]

The blacket term can be evaluated by the Monte Carlo method.

When \( x(t) \) is Gaussian, it has an analytic form.
Effectiveness of PPLS

Table 3: Comparison of Variable Selection Procedures: Ratio of PPLS versus PLS based on 400 simulations

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha(t) = \tau \sqrt{t/\tau}$</th>
<th>$\alpha(t) = \tau \sin(2\pi t/\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RGMSE Zero Coefficient</td>
<td>RGMSE Zero Coefficient</td>
</tr>
<tr>
<td>Case I: $n = 50$, $\tau = 20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>mean (std) C I</td>
<td>mean (std) C I</td>
</tr>
<tr>
<td>$L_1$</td>
<td><strong>0.3936</strong> (0.2966) 4.9950 0</td>
<td>0.3923 (0.2863) 4.9900 0</td>
</tr>
<tr>
<td>SCAD</td>
<td><strong>0.3549</strong> (0.2453) 4.9950 0</td>
<td>0.3533 (0.2453) 4.9925 0</td>
</tr>
<tr>
<td>Oracle</td>
<td><strong>0.3502</strong> (0.2412) 5.0000 0</td>
<td>0.3480 (0.2425) 5.0000 0</td>
</tr>
<tr>
<td>Case II: $n = 75$, $\tau = 4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_1$</td>
<td><strong>0.5772</strong> (0.2614) 4.3325 0</td>
<td>0.5733 (0.2648) 4.3500 0</td>
</tr>
<tr>
<td>SCAD</td>
<td><strong>0.5127</strong> (0.2101) 4.4275 0</td>
<td>0.5115 (0.2107) 4.4250 0</td>
</tr>
<tr>
<td>Oracle</td>
<td><strong>0.3939</strong> (0.2326) 5.0000 0</td>
<td>0.3915 (0.2318) 5.0000 0</td>
</tr>
<tr>
<td>Case III: $n = 50$, $\tau = 20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_1$</td>
<td><strong>0.3975</strong> (0.2843) 4.9950 0</td>
<td>0.4002 (0.2860) 4.9975 0</td>
</tr>
<tr>
<td>SCAD</td>
<td><strong>0.3450</strong> (0.2278) 4.9975 0</td>
<td>0.3460 (0.2279) 4.9975 0</td>
</tr>
<tr>
<td>Oracle</td>
<td><strong>0.3438</strong> (0.2269) 5.0000 0</td>
<td>0.3450 (0.2271) 5.0000 0</td>
</tr>
<tr>
<td>Case IV: $n = 50$, $\tau = 30$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_1$</td>
<td><strong>0.4091</strong> (0.2716) 4.9975 0</td>
<td>0.4074 (0.2717) 5.0000 0</td>
</tr>
<tr>
<td>SCAD</td>
<td><strong>0.3554</strong> (0.2210) 5.0000 0</td>
<td>0.3546 (0.2205) 5.0000 0</td>
</tr>
<tr>
<td>Oracle</td>
<td><strong>0.3549</strong> (0.2200) 5.0000 0</td>
<td>0.3542 (0.2199) 5.0000 0</td>
</tr>
</tbody>
</table>

Table 4: Accuracy of Standard Errors formula

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>std se (std(se))</td>
<td>std se (std(se))</td>
<td>std se (std(se))</td>
</tr>
<tr>
<td>$L_1$</td>
<td><strong>0.0823</strong> 0.0798 (0.0176)</td>
<td>0.0826 0.0775 (0.0177)</td>
<td>0.0735 0.0702 (0.0166)</td>
</tr>
<tr>
<td>SCAD</td>
<td><strong>0.0810</strong> 0.0808 (0.0180)</td>
<td>0.0821 0.0793 (0.0187)</td>
<td>0.0738 0.0708 (0.0169)</td>
</tr>
<tr>
<td>Oracle</td>
<td><strong>0.0808</strong> 0.0808 (0.0181)</td>
<td>0.0810 0.0794 (0.0188)</td>
<td>0.0737 0.0709 (0.0169)</td>
</tr>
</tbody>
</table>

Based on case I design with $n = 50$, $\alpha(t) = \tau \sqrt{t/\tau}$ with $\tau = 20$.  

29
Application to the CD4 Data

**Previous Analysis:** Wu and Chiang (2000) and Fan and Zhang (2000) used the model

\[ y(t) = \beta_0(t) + \beta_1(t)\text{Smoking} + \beta_2(t)\text{Age}(t) + \beta_3(t)\text{PreCD4}(t) + \varepsilon(t). \]

Results of Huang, Wu and Zhou (2002) suggests that functions \( \beta_1(\cdot) \), \( \beta_3(\cdot) \), and \( \beta_3(\cdot) \) are constant.

**Our Analysis:** Involve interactions

\[
\begin{align*}
y(t) &= \alpha(t) + \beta_1 x_1 + \beta_2 x_2(t) + \beta_3 x_3(t) + \beta_4 x_2^2(t) + \beta_5 x_3^2(t) \\
&\quad + \beta_6 x_1 x_2(t) + \beta_7 x_1 x_3(t) + \beta_8 x_2(t) x_3(t) + \varepsilon(t).
\end{align*}
\]

\( x_2 \) and \( x_3 \) are standardized age and PreCD4, resp.
Results

Table 5: Estimated Coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Profile LS $\hat{\beta}(\text{se(}\hat{\beta})\text{))}$</th>
<th>$L_1$ $\hat{\beta}(\text{se(}\hat{\beta})\text{))}$</th>
<th>SCAD $\hat{\beta}(\text{se(}\hat{\beta})\text{))}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>0.5333(1.0972)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.1010(0.9167)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>PreCD4</td>
<td>2.8252(0.8244)</td>
<td>3.0932(0.5500)</td>
<td>3.1993(0.5699)</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>0.1171(0.4558)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>PreCD4$^2$</td>
<td>-0.0333(0.3269)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>Smoking*Age</td>
<td>-1.7084(1.1192)</td>
<td>-0.9684(0.4904)</td>
<td>-1.0581(0.5221)</td>
</tr>
<tr>
<td>Smoking*PreCD4</td>
<td>1.3277(1.3125)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>Age*PreCD4</td>
<td>-0.1360(0.5413)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
</tbody>
</table>