

Overview

Temporal Processes Both response and covariates are continuously observed temporal processes, which can be viewed as functional data.

Functional Regression The process means are modeled marginally using a generalized linear model with covariate coefficients unspecified over time. No Markov assumption is needed.

Moment Based Estimators Time-varying coefficients are estimated by applying the easy-to-implement moment methods on the abundant cross-sectional data. Smoothing is not required, unlike most methods for varying-coefficient models.

Statistical Inferences With the nonparametric time-varying estimators, we can do

- hypotheses tests of the covariate effects
- estimation and goodness-of-fit testing of parametric submodels
- predictions involving estimated components from both the functional model and the submodels.

Functional Regression Models and Temporal Processes

Jun Yan

Department of Statistics
University of Wisconsin–Madison

Multistate Motivating Example: Prevalence of Chronic Graft Versus Host Disease (C-GVHD)

Data: 147 bone marrow transplant patients, followed up to 36 months for C-GVHD onset time, C-GVHD remission time, relapse time, death time, last contact time (all times may be censored), as well as treatment arm (MC and PMC) and other prognosis factors, such as patient gender, patient age, donor gender, donor age, and phenotype mismatch between marrow donor and recipient.

Research Objective: Evaluate the effects of treatment and other prognosis factors on the prevalence of C-GVHD in bone marrow transplantees, alive and relapse free.

Existing Methods: Intensity based methods; Marginal means models.

Functional View: C-GVHD indicator $Y_i(t)$, $p \times 1$ covariates $X_i(t)$, alive and relapse free indicator $S_i(t)$, and data availability indicator $\delta_i(t)$, which accommodates loss to follow-up and missing covariates.

Outline

1. Introduction and Motivation
2. Review of Previous Work for Multistate Processes
3. Functional Estimating Equations (FEE) for Mean Models
4. Extension to Functional Association Models for Multivariate Processes
5. Summaries and Future Work

Previous Work: Marginal Models for Means

Point Processes:

Proportional Rate Models Pepe and Cai 1993, Lin et. al. 2000

$$d\mu_X(t) = d\mu_0(t) \exp\{X^T(t)\beta\}$$

Proportional Means Models Lawless and Nadeau 1995, Lin et. al. 2000

$$\mu_X(t) = \mu_0(t) \exp\{X^T(t)\beta\}$$

Discretely Observed Longitudinal Data:

Time-varying Coefficient Model Hoover et. al. 1998, Wu et. al. 1998,

Martinussen and Scheike 1999, Fan and Zhang 2000, Lin and Ying 2001

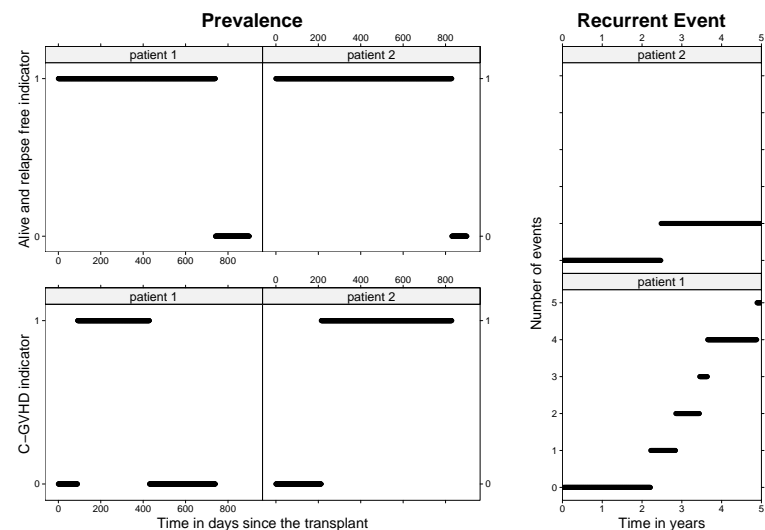
$$\mu_X(t) = X^T(t)\beta(t)$$

General Functional Regression Models Ramsay and Silverman 1997

Summary:

- Weaker assumptions
- More meaningful parameter interpretations for some applications
- Computationally intensive, may require smoothing.

Graphical Display of Temporal Processes



Existing Methods Applied to the Examples

Recurrent Event

Transition intensity The intensity of interest is the instantaneous risk of a jump conditioning on the entire history.

Andersen and Gill 1982

Marginal means model Concerns about the mean of the processes marginally.

Pepe and Cai 1993, Lawless and Nadeau 1995, Lin, Wei, and Ying 2001.

Prevalence

Transition intensity Prevalence is estimated indirectly from transition intensity estimates.

Tempkin 1978, Begg and Larson 1982

Marginal means model Prevalence is directly modeled with easier interpretation.

Pepe and Couper 1997 used a logistic model for prevalence, with time-varying intercept specified by cubic splines, and time-independent covariates.

Previous Work: Transition Intensity Models

Multiplicative Intensities Proportional intensity; Cox 1972, Andersen and Gill 1982.

$$\lambda_X(t) = \lambda_0(t) \exp\{X^T(t)\beta\}$$

Additive Intensities Aalen 1980, Huffer and McKeague 1991

$$\lambda_X(t) = X^T(t)\beta(t)$$

Partly Parametric Models Zucker and Karr 1990, Lin and Ying 1994

Additive-Multiplicative Model Lin and Ying 1995, Martinussen and Scheike 2002,

Generalized Additive Models (GAM) Hastie and Tibshirani 1990, 1995

Summary:

- Strong assumption: Poisson property
- Likelihood based estimation, high efficiency if correctly specified
- Lack of robustness, computationally intensive, smoothing may be required.

Asymptotic Result

Uniform Consistency $\sup_{t \in [l, u]} \|\hat{\beta}(t) - \beta_0(t)\| \xrightarrow{a.s.} 0$

Weak Convergence $N^{1/2}\{\hat{\beta}(t) - \beta_0(t)\}$ converges weakly in $\{\ell^\infty([l, u])\}^p$ to a tight, mean zero Gaussian process $\mathcal{Z}(t)$ with covariance

$$\Sigma(s, t) = E\{v_1(s)v_1^T(t)\},$$

where $v_i(t) = H_1^{-1}(t)U_{1i}\{\beta_0(t), t\}$, and

$$H_1(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \{S_i(t)\}\{\delta_i(t)\}D_{1i}^T\{\beta_0(t)\}V_{1i}^{-1}\{\beta_0(t)\}D_{1i}\{\beta_0(t)\}$$

Covariance Function Estimation

$$\hat{\Sigma}(s, t) = N^{-1}\{\hat{H}_1^{-1}(s)\} \sum_{i=1}^N \{\hat{U}_{1i}(s)\hat{U}_{1i}^T(t)\}\{\hat{H}_1^{-1}(t)\}^T,$$

where a hat ($\hat{\cdot}$) denotes evaluation at $\hat{\beta}(t)$.

Functional Means Model

Functional Data: $\{Y_i(t), X_i(t), S_i(t) : \delta_i(t) = 1, i = 1, \dots, N\}$

Varying-Coefficient Generalized Linear Model

$$g\{\mu_i(t)\} = X_i^T(t)\beta(t),$$

where $\mu_i = E\{Y_i(t)|X_i(t), S_i(t) = 1\}$, g is a link function, and $\beta(t)$ is the $p \times 1$ vector of regression coefficients.

Conditionally Independent Censoring Assumption

At each t , $\{Y_i(t)\} \perp \{\delta_i(t)\} | \{X_i(t), S_i(t) = 1\}$

Applications: Prevalence, Recurrent Event, Medical Cost, Survival.

Connections: This is in the framework of marginal mean modeling, unlike intensity based models, which completely specifies the processes via Markov assumption. The model generalizes the discrete time model in Pepe and Couper (1997) and allows the covariate effects to be completely unspecified over time.

Hypothesis Testing I

Null Hypothesis: Let $C(t)$ be a $r \times p$ matrix and $c(t)$ be a $r \times 1$ vector.

$$H_0 : C(t)\beta(t) = c(t),$$

Test Statistics:

Finite Time Points Test: Take a finite number of time points s_1, \dots, s_Q , each from a different segment in the grid formed by the M jump points in $[l, u]$. Let

$$\hat{\beta}^* = \begin{bmatrix} \hat{\beta}(s_1) \\ \vdots \\ \hat{\beta}(s_Q) \end{bmatrix}, C^* = \begin{bmatrix} C(s_1) & & \\ & \ddots & \\ & & C(s_Q) \end{bmatrix}, \text{ and } c^* = \begin{bmatrix} c(s_1) \\ \vdots \\ c(s_Q) \end{bmatrix}.$$

Define the test statistic

$$T_1 = (C^*\hat{\beta}^* - c^*)^T (C^*\hat{\Sigma}^*C^{*T})^{-1} (C^*\hat{\beta}^* - c^*),$$

where $\hat{\Sigma}^*$ is the estimated covariance matrix of $\hat{\beta}^*$. T_1 has limiting distribution which is Chi-squared with degrees of freedom rQ .

Nonparametric Estimation

Set Up For the temporal processes we are considering, each subject has only finitely many jump points in the observed quantities, which are step functions. Let $l < t_1 < \dots < t_M < u$ be the jump points in all the observed processes in $[l, u]$. The estimated time-varying coefficients $\hat{\beta}(t)$ would be piecewise constant, with jumps only at those jump points in the observed data.

Estimating Function For each t , let $\hat{\beta}(t)$ be the solution to

$$\sum_{i=1}^N U_{1i}\{\beta(t), t\} = 0,$$

where

$$U_{1i}\{\beta(t), t\} = D_{1i}^T\{\beta(t), t\}V_{1i}^{-1}\{\beta(t), t\}\{S_i(t)\}\{\delta_i(t)\}\{Y_i(t) - \mu_i(t)\},$$

$$D_{1i}\{\beta(t), t\} = \partial\mu_i(t)/\partial\beta(t), \text{ and } V_{1i}\{\beta(t), t\} \text{ is a weight matrix, possibly random.}$$

Key Features No smoothing, separate estimation at each time point.

Goodness-of-Fit and Prediction

Goodness-of-Fit

- For a parametric submodel with parameter estimate $\hat{\eta}$, modify the nonparametric test statistics by replacing $C(t)\hat{\beta}(t) - c(t)$ with $\hat{\beta}_i(t) - f(\hat{\eta}, t)$.
- The limiting distributions of the modified T_1 and T_2 are Chi-squared with degrees of freedom 1 under the null $\beta_{i0}(t) = f(\eta, t)$.

Prediction

- Estimate $F(t) = F\{\beta(t), \eta_1, \dots, \eta_p, t\}$ using

$$\hat{F}(t) = F\{\hat{\beta}(t), \hat{\eta}_1, \dots, \hat{\eta}_p, t\}$$
- Example: $F(t) = g^{-1}\{X_0^T(t)\beta(t)\}$.
- $\hat{F}(t)$ is uniformly consistent for $F(t)$, and $N^{1/2}\{\hat{F}(t) - F(t)\}$ converges weakly to a zero mean Gaussian process with covariance function which is easily estimated with a plug-in formula.

Hypothesis Testing II

Integral Test: Let $\Delta = \int_t^u \{C(t)\hat{\beta}(t) - c(t)\} \tilde{W}(t) dt$, where $\tilde{W}(t)$ is a non-negative weight function, possibly random, with limit $W(t)$. Taking $\tilde{W}(t) = [NC(t)\hat{\Sigma}(t, t)C^T(t)]^{-1}$ downweights time points where $\hat{\beta}(t)$ is more variable. Define

$$T_2 = \Delta^T \hat{\Sigma}_{\Delta}^{-1} \Delta,$$

where $\hat{\Sigma}_{\Delta}$ is the estimated variance of Δ . T_2 has limiting distribution Chi-squared with degrees of freedom r .

When $r = 1$, test statistic $T_2^* = \hat{\Sigma}_{\Delta}^{-1/2} \Delta$ has standard normal distribution. The sign of T_2^* indicates the direction of the deviation from H_0 .

Sup Test:

$$T_3 = N \sup_{t \in [t, u]} |\{C(t)\hat{\beta}(t) - c(t)\}^T \{C(t)^T \hat{\Sigma}(t, t) C(t)\}^{-1} \{C(t)\hat{\beta}(t) - c(t)\}|$$

T_3 has complicated limiting distribution. Implementation may be based on bootstrapping $\hat{\beta}(t)$.

Analyzing the C-GVHD Data

Simple Model Treatment is the only covariate (1 for PMC, 0 for MC).

- Testing $H_0 : \beta_{\text{trt}}(t) = 0$, the statistic $T_1 = 10.23$ using $\hat{\beta}_{\text{trt}}(t)$ at two time points equal to the 15th and 85th percentiles of the jump points of $\{Y_i(t), S_i(t), \delta_i(t), i = 1, \dots, 147\}$ and $T_2^* = -2.35$ with $\tilde{W}(t) = [NC(t)\hat{\Sigma}(t, t)C^T(t)]^{-1}$, p -values 0.006 and 0.019
- The parametric model $\beta_{\text{trt}}(t) = \eta$ was fit with $\tilde{W}(t) = \widehat{\text{var}}^{-1}\{\hat{\beta}_{\text{trt}}(t)\}/N$. The estimate $\hat{\eta} = 0.91$, with standard error 0.38. The estimate of η in Pepe and Couper (1997) was 0.92, with standard error 0.39, a close match.

Complex Model Treatment, age, and gender are covariates. Treatment and gender coefficients are found to be better described by linear models in time, but age coefficient appears to be constant. The linear model is $\hat{\beta}_{\text{trt}} = -.4372 - .0013t$ for treatment, and $\hat{\beta}_{\text{gender}} = 0.0336 - 0.0020t$ for patient gender. The constant model for patient age is $\hat{\beta}_{\text{age}} = 0.0366$.

Inferences for Parametric Submodels

Modeling the Temporal Structure: Consider modeling the i th covariate effect by $\beta_i(t) = f(\eta, t)$, e.g., $f(\eta, t) = \sum_{j=1}^l \eta_j t^{j-1}$.

Estimation:

$$\hat{\eta} = \arg \min \int_t^u \{\hat{\beta}_i(t) - f(\eta, t)\}^2 \tilde{W}(t) dt,$$

where $\tilde{W}(t)$ is a non-negative weight function, possibly random, with limit $W(t)$.

Sampling Properties: $\hat{\eta} \xrightarrow{p} \eta_0$ and $N^{1/2}(\hat{\eta} - \eta_0)$ has limiting normal distribution with influence function

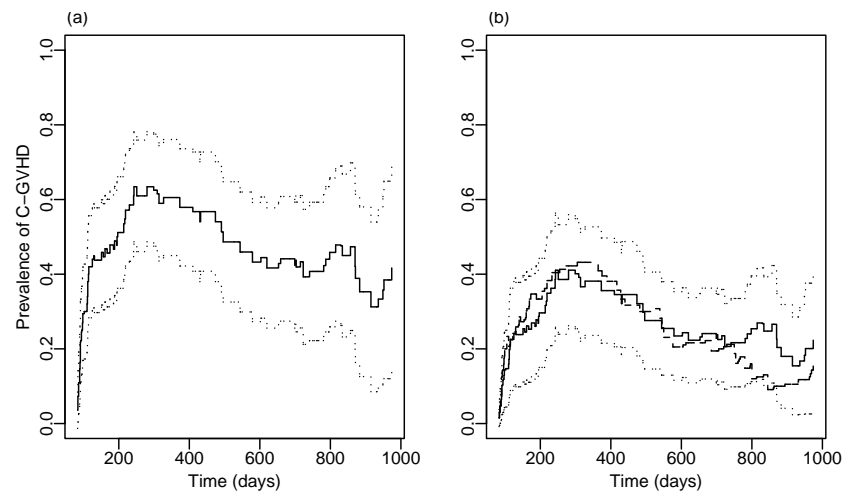
$$\ell_i\{\beta_0(t), \eta_0\} = A^{-1}(t) \int_t^u \dot{f}^T(\eta_0, t) v_i(t) W(t) dt,$$

where $\dot{f}(\eta, t) = \partial f(\eta, t) / \partial \eta$, and $A(t) = \int_t^u \dot{f}(\eta_0, t) W(t) \dot{f}^T(\eta_0, t) dt$. Asymptotic variance Γ can be consistently estimated by $\hat{\Gamma} = N^{-1} \sum_{i=1}^N \hat{\ell}_i^{\otimes 2}$, where $\hat{\ell}_i$ is ℓ_i evaluated at $\hat{\beta}(t)$, $\hat{\eta}$, $\hat{v}_i(t)$, $i = 1, \dots, N$, and \tilde{W} .

Hypothesis and Goodness-of-fit Testing for the Complex Model

- Test statistic T_1 is based on two time points s_1 and s_2 equal to the 15th and 85th percentiles of the jump points.
- Test statistic T_2^* is based on $\tilde{W}(t) = [N \widehat{\text{var}}\{\hat{\beta}_i(t)\}]^{-1}$. We computed T_{2a}^* , T_{2b}^* , and T_{2c}^* with $W(t) = 1, I\{t < (u-l)/2\}$, and $I\{t > (l-u)/2\}$, respectively, where $I(\cdot)$ is the indicator function.

Test	treatment		patient gender		patient age	
	Stat	p-value	Stat	p-value	Stat	p-value
Hypothesis testing of $H_0: \beta_i(t) = 0$						
T_1	9.899	0.007	4.683	0.096	4.502	0.105
T_2^*	-2.431	0.015	-2.005	0.045	1.970	0.049
Goodness-of-fit testing of constant submodel						
T_1	7.208	0.027	2.471	0.290	2.580	0.275
T_{2a}^*	-2.320	0.020	-1.752	0.080	0.720	0.483
T_{2b}^*	1.721	0.085	3.799	0.001	-1.712	0.087
T_{2c}^*	-2.767	0.006	-2.452	0.004	0.935	0.350
Goodness-of-fit testing of linear submodel						
T_1	4.456	0.108	0.688	0.709		
T_{2a}^*	-0.684	0.494	-0.517	0.605		
T_{2b}^*	0.399	0.690	0.053	0.958		
T_{2c}^*	-1.265	0.206	-0.726	0.468		



Prevalence estimates for C-GVHD from the simple model: nonparametric estimate (solid), semiparametric estimate (dashed), and 0.95 pointwise confidence intervals (dotted) from the semiparametric estimate. (a) MC, (b) PMC.

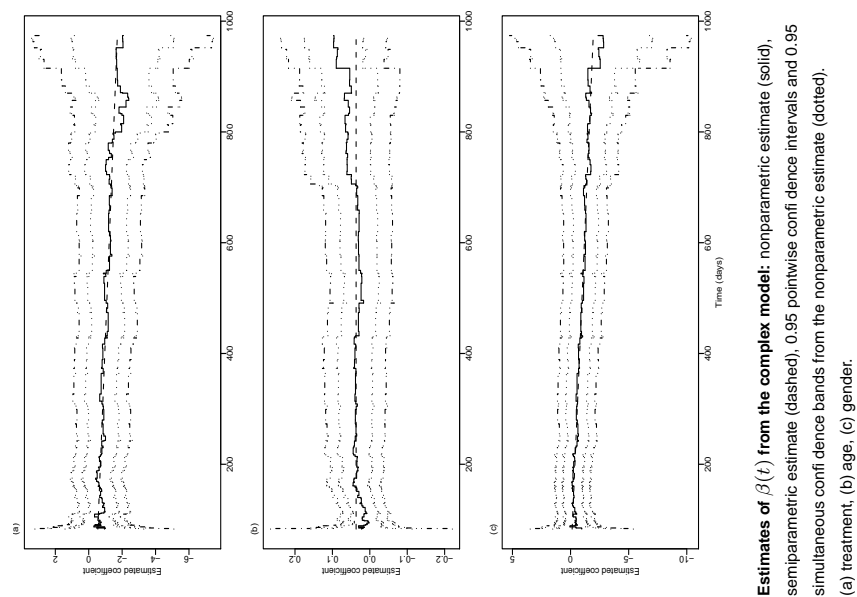
Multivariate Motivating Example: Familial Clustering of Alcoholism

Data: COGA, 105 high risk pedigrees, 1214 individuals interviewed; onset age of alcohol dependence according to ALDX1, covariates (such as gender and smoking status), and censoring time (interview age).

Research Objectives:

- Identify risk factors for alcohol dependence, and
- Investigate familial association in alcoholism, which may be time-dependent

Functional View: Alcoholism indicator $Y_{ij}(t)$, covariates $X_{ij}(t)$, data availability indicator $\delta_{ij}(t)$ for the mean model, covariate $Z_{i(j,k)}(t)$, data availability indicator $\zeta_{i(j,k)}(t)$ for the association model.



Estimates of $\beta(t)$ from the complex model: nonparametric estimate (solid), semiparametric estimate (dashed), 0.95 pointwise confidence intervals and 0.95 simultaneous confidence bands from the nonparametric estimate (dotted). (a) treatment, (b) age, (c) gender.

Analyzing the Familial Alcoholism Data

Data: Clustered processes. Cross-sectional data at each time t for each family are multivariate binary responses.

Mean Model: Functional logistic model,

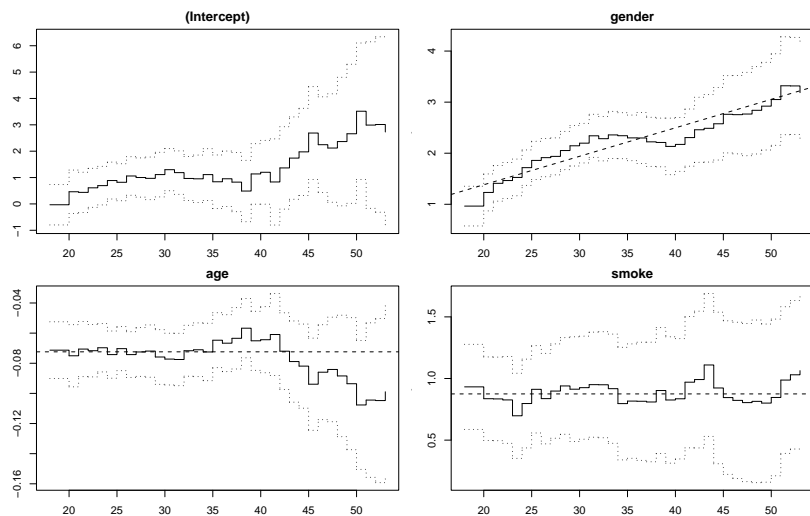
$$\text{logit}\{\mu_{ij}(t)\} = X_{ij}\beta(t),$$

where $\mu_{ij}(t) = E\{Y_{ij}(t)|X_{ij}, \delta(t) = 1\}$, covariate X_{ij} includes gender, age, and smoke. Note that age is included as a cohort effect.

Association Model: Odds ratio with log link, single time index for simplicity,

$$\log\{\psi_{i(j,k)}(t)\} = Z_{i(j,k)}(t)\alpha(t),$$

where $\psi_{i(j,k)}(t)$ is the odds ratio of the binary responses $Y_{ij}(t)$ and $Y_{ik}(t)$, covariate $Z_{i(j,k)}$ include relationship indicators sibs and other (neither sibs nor spouse pairs).



Estimates of $\beta(t)$: nonparametric estimate (solid), semiparametric estimate (dashed), 0.95 pointwise confidence intervals from the nonparametric estimate (dotted).

Previous Work: Multivariate Association

Global Association Models:

Parametric copula (Joe 1997, Nelsen 1999);

Random effects / frailty (Clayton 1978, Clayton and Cuzick 1985);

Moment based models (Prentice and Zhao 1991, Liang, Zeger, and Qaqish 1992)

Time Dependent Association Measure:

Oakes (1989) cross ratio,

$$\theta(t_1, t_2) = \frac{\lambda(t_1|T_2 = t_2)}{\lambda(t_1|T_2 > t_2)}.$$

Plackett (1965) odds ratio

$$\psi(t_1, t_2) = \frac{\Pr(T_1 \leq t_1, T_2 \leq t_2) \Pr(T_1 > t_1, T_2 > t_2)}{\Pr(T_1 \leq t_1, T_2 > t_2) \Pr(T_1 > t_1, T_2 \leq t_2)}.$$

Prentice and Cai (1992) covariance function of the marginal martingale.

FEE for Association Structure

Functional Association Model Let $t = (t_1, t_2)$, and $\rho_{i(j,k)}$ be an association measure, such as correlation, odds ratio, etc.

$$h\{\rho_{i(j,k)}(t)\} = Z_{i(j,k)}^T(t)\alpha(t)$$

Estimating Function Let $W_{i(j,k)}(s, t) = \{Y_{ij}(s) - \mu_{ij}(s)\}\{Y_{ik}(t) - \mu_{ik}(t)\}$ and $\eta_{i(j,k)}(s, t) = E\{W_{i(j,k)}(s, t)|\mathcal{D}_{i(j,k)}(s, t)\}$, where $\mathcal{D}_{i(j,k)}$ is all the covariate information for the paired responses. The estimating function is

$$\sum_{i=1}^N U_{2i}\{\alpha(t), \beta(t), t\} = 0,$$

where

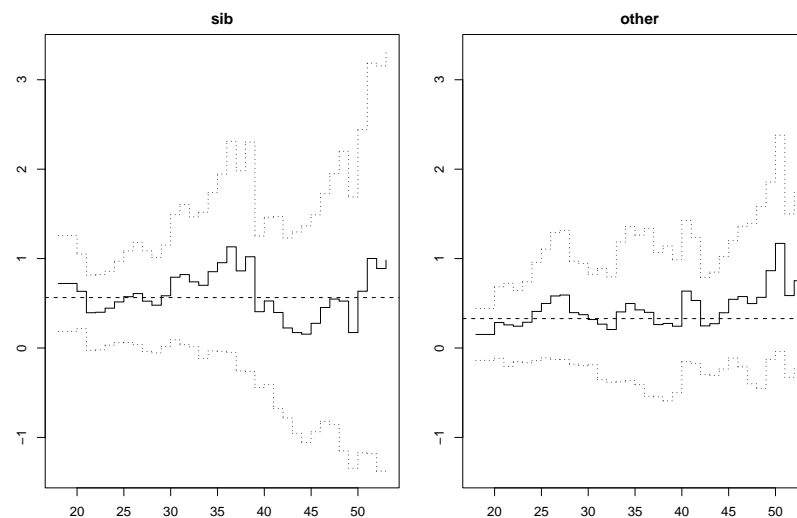
$$U_{2i}\{\alpha(t), \beta(t), t\} = D_{2i}^T\{\alpha(t), \beta(t), t\}V_{2i}^{-1}\{\alpha(t), t\}\text{diag}\{R_i(t)\}\text{diag}\{\zeta_i(t)\}\{W_i(t) - \eta_i(t)\},$$

$$D_{2i}\{\alpha(t), \beta(t), t\} = \partial\eta_i(t)/\partial\alpha(t),$$

and $V_{2i}\{\alpha(t), t\}$ is a weight matrix, possibly random, and may involve further estimated parameters.

Summaries

- Functional generalized linear model is proposed for continuously observed processes.
- Covariates are nonparametrically estimated using simple GEE approach on the abundant cross-sectional data.
- The estimates are uniformly consistent and converges weakly to a Gaussian processes, with easily estimated variances, which can be used to develop statistics for hypothesis testing and goodness-of-fit test.
- The method can be naturally extended for the multivariate outcomes when association is of research interest.
- The C-GVHD prevalence and familial alcoholism association examples illustrated the usefulness of the methodology.



Estimates of $\alpha(t)$: nonparametric estimate (solid), semiparametric estimate (dashed), 0.95 pointwise confidence intervals (dotted). The constant fit are 0.564 and 0.329, respectively.

Future Work

Partly Functional Models The full time-varying coefficient model may be used as a diagnostic tool for parsimonious models. It is desirable to be able to fit partly functional models.

Model Selection When there are many covariates, a model selection criterion is needed, which requires the development of model comparison procedures.

Informative Censoring The current setup assumes random censoring and it will be worthwhile to examine how dependent censoring can be introduced into this framework.

Computing There are two levels of computing: cross-sectional computing and functional computing. At the functional level, more efficient data representation need to be further investigated.

COGA: Hypothesis Testing and Goodness-of-fit Testing

	Mean Model		Association Model		
	gender	age	smoke	sib	other
Hypothesis testing of $H_0 : \theta_i(t) = 0$					
T_2^*	9.559	-9.861	4.955	2.003	1.426
Goodness-of-fit testing of constant submodel					
T_{2a}^*	-0.017	-0.022	0.001	0.928	
T_{2b}^*	-3.306	-0.334	-0.601	1.650	
T_{2c}^*	3.091	0.036	1.149	-1.088	
Goodness-of-fit testing of linear submodel					
T_{2a}^*	-0.134				
T_{2b}^*	0.955				
T_{2c}^*	-1.576				

Test statistic T_2^* is based on $\tilde{W}(t) = [N\widehat{\text{var}}\{\hat{\theta}_i(t)\}]^{-1}$. We computed T_{2a}^* , T_{2b}^* , and T_{2c}^* with $\tilde{W}(t)$, $\tilde{W}(t)I\{t < (u-l)/2\}$, and $\tilde{W}(t)I\{t > (l-u)/2\}$, respectively, where $I(\cdot)$ is the indicator function.