Partly Functional Temporal Process Regression with Semiparametric Profile Estimating Functions

Jun Yan Jian Huang *

Abstract

Marginal mean models of temporal processes in event time data analysis are gaining more attention for its milder assumptions than the traditional intensity models. Recent work of fully functional temporal process regression offers great flexibility by allowing all the regression coefficients to be nonparametrically time-varying. The estimation procedure, however, prevents successive goodness-of-fit test for covariate coefficient in comparing a sequence of nested models. This article proposes a partly functional temporal process regression model in the line of marginal mean models. Some covariate effects are time-independent while the others are completely unspecified in time. This class of models is very rich, including the fully functional model and semiparametric model as special cases. To estimate the parameters, we propose semiparametric profile estimating functions which are solved via an iterative algorithm, starting with a consistent parameter estimate from a fully functional model in the existing work. No smoothing is needed, in contrast to other varying-coefficient methods. The weak convergence of the resultant estimators are developed using the empirical process theory. Successive tests of time-varying effects and backward model selection procedure

*Jun Yan is Assistant Professor, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA (E-mail: jyan@stat.uiowa.edu). Jian Huang is Professor, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA (E-mail: jian@stat.uiowa.edu).

can then be carried out. The practical usefulness of the methodology is demonstrated through a simulation study and a real example of recurrent exacerbations among cystic fibrosis patients.

KEY WORDS: Cystic fibrosis; Estimating function; Functional regression; Generalized linear model; Recurrent event; Semiparametric method.

1. INTRODUCTION

Event time data arise in a variety of fields such as engineering, social sciences, and biomedical research. The quantity of interest is the time until some event occurs. In many applications, the event can be recurrent, where an individual subject may experience multiple events. Some examples of recurrent event data in the literature are times to system repairs (Lawless and Nadeau 1995), times to product warranty claims (Kalbfleisch, Lawless, and Robinson 1991), times to unemployment (Lancaster 1990), and times to recurrent disease episodes (Pepe and Cai 1993; Therneau and Hamilton 1997).

Event times have traditionally been analyzed based on intensities which model the instantaneous risk for an event to occur (e.g. Kalbfleisch and Prentice 2002). For illustration, consider a recurrent event setting, with $Y(t)$ being the number of events occurred by time t. The intensity function is defined as

$$\lambda(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{Y(t + \Delta) - Y(t) = 1 | \mathcal{H}(t)\}, \tag{1}$$

where $\mathcal{H}(t)$ is the history of the process up to time t. Covariate effects are incorporated into the intensity either multiplicatively (Cox 1972) or additively (Aalen 1980). Intensity models fully specify the stochastic nature of the event times, which may be too strong to assume in many applications.

With milder assumptions, marginal mean models have been advocated recently (Pepe and Cai 1993; Lawless and Nadeau 1995; Lin, Wei, Yang, and Ying 2000; Lin, Wei, and Ying 2001). These models study the temporal processes associated with the event times, such as point processes or counting processes, without fully specifying the stochastic structure of the processes as intensity function (1) does. The focus is the mean of a temporal process $Y(t)$ marginalized at each time point in a time period of interest: $\mu(t) = E\{Y(t)\}$. As Lin et al. (2001) pointed out, marginal mean models have several advantages over intensity models: mean function is more intuitive than intensity function; mean models do not require the Poisson structure; and mean models can be used for point processes with positive jumps of arbitrary size. The common form of these models with covariates has an unspecified baseline mean function and time-independent covariate coefficients.

Extending the line of marginal mean models, Fine, Yan, and Kosorok (2004) proposed a fully functional temporal process regression (TPR) model, where all the covariate coefficients, not just the baseline, are completely unspecified over time as in Aalen's (1980) model. Let $X(t)$ be a $p \times 1$ vector of covariate, which may include an intercept. A TPR model specifies the conditional mean function of $Y(t)$ given $X(t)$ by

$$g\{\mu(t|X)\} = X^\top(t)\alpha(t), \tag{2}$$

where $\mu(t|X) = E\{Y(t)|X(t)\}$, g is a known link function, and $\alpha(t)$ is a $p \times 1$ vector of time-varying regression coefficients. At any fixed t, model (2) is a generalized linear model (GLM). Considered over a time period, model (2) is a fully functional model in that all components of $\alpha(t)$ are completely unspecified in time t, allowing robust exploration of the functional forms of the coefficients. The time-varying coefficients at each time are estimated separately from estimating functions using the cross-sectional data at that time. This model is a fully functional extension of the marginal mean models advocated by Lawless and Nadeau (1995), Lin et al. (2000), and Lin et al. (2001). In particular, the semiparametric transformation model of Lin et al. (2001) has the form

$$g\{\mu(t|Z)\} = \alpha(t) + Z^\top(t)\theta, \tag{3}$$

where $\alpha(t)$ is a scalar function of t, $Z(t)$ is a $q \times 1$ vector of covariate, and θ is a $q \times 1$ vector of time-independent covariate coefficients. Note that $g^{-1}\{\alpha(t)\}$ is interpreted as the baseline mean function with $Z(t) = 0$. Model (3) is much simpler than model (2) in that only the intercept $\alpha(t)$ is functional while other covariate coefficient are time-independent.

The great flexibility of model (2) comes with some limitations. In estimating the fully functional parameters, the variance-bias tradeoff may be critical in small to medium samples and the number of covariates it can handle is therefore limited. Furthermore, it may be desirable that some of the coefficients be parametric in many applications. This need arises, for example, when there exists strong prior knowledge about the forms of particular coefficients, or when the goodness-of-fit test after fitting model (2) suggests that a regression coefficient is constant over time. Researchers are often interested in examining the time-varying coefficients in model (2) with a number of successive tests that investigate the covariates one at a time. The approach of Fine et al. (2004), however, only allows limited parametric submodels fitting at a second stage after nonparametric coefficient estimates have been obtained from the fully functional model (2). Each covariate coefficient is tested against the full model instead of the model accepted from a previous test. Successive testing necessitates the study of partly functional models.

In this article, we propose a partly functional TPR (PFTPR) model

$$g\{\mu(t|X, Z)\} = X^\top(t)\alpha(t) + Z^\top(t)\theta, \tag{4}$$

where g is a twice continuously differentiable link function, $\mu(t|X, Z) = E\{Y(t)|X(t), Z(t)\}$, $X(t)$ is a $p\times 1$ covariate vector with time-varying coefficient $\alpha(t)$, and $Z(t)$ is a $q \times 1$ covariate vector with time-independent coefficient θ. This is a rich family of models. It reduces to the fully functional TPR model of Fine et al. (2004) when Z is absent. It reduces to the semiparametric transformation model of Lin et al. (2001) when X contains only an intercept. The models of both Fine et al. (2004) and Lin et al. (2001) are rich families themselves.

Partly functional forms have been proposed for intensities models. Let $\lambda(t|X, Z)$ be the intensity conditional on $X(t)$ and $Z(t)$. A partly functional intensity model analogous to (4) has the form

$$g\{\lambda(t|X, Z)\} = X^\top(t)\alpha(t) + Z^\top(t)\theta. \tag{5}$$

McKeague and Sasieni (1994) considered additive intensity models with $g(u) = u$. Martinussen, Scheike, and Skovgaard (2002) considered multiplicative intensity models with $g(u) = \log u$. These approaches are likelihood based and the estimation may involve smoothing.

The goal of this paper is to develop inferences for the PFTPR model (4), the counterpart of intensity model (5) in the line of marginal mean models. Estimation of the parameters in the PFTPR model is through an approach called semiparametric profile estimating functions, which reduces to the separate estimation method of Fine et al. (2004) when θ is given. Unlike the likelihood based approaches, only the mean, instead of the complete stochastic structure, of the temporal processes needs to be correctly specified to obtain consistent estimators. Similar to the method of Fine et al. (2004), no smoothing is needed for the functional coefficient estimation.

This paper is organized as follows. Estimation of the parameters θ and $\alpha(t)$ with semiparametric profile estimating functions is proposed in Section 2. The asymptotic properties of the estimator are presented in Section 3. Hypothesis tests for parameter significance and goodness-of-fit are developed in Section 4. A simulation study in a recurrent event setting is reported in Section 5. The methodology is applied to a clinical trial in Section 6 to assess the efficacy of a treatment for cystic fibrosis patients. A discussion concludes in Section 7.

2. ESTIMATION

Let $Y(t)$ be a response temporal process, $X(t)$ be a $p \times 1$ vector of covariates with time-varying coefficients, $Z(t)$ be a $q \times 1$ vector of covariates with time-independent coefficients, and C be a censoring time. The information from the censoring time C is equivalent to a data availability indicator process constructed as $\delta(t) = I(t < C)$. Suppose that, in a time window $[l, u]$, we observe n independent copies of $\{Y(t), X(t), Z(t), \delta(t)\}$, denoted by

$$\{Y_i(t), X_i(t), Z_i(t), \delta_i(t)\},$$

$i = 1, \ldots, n$. These data will be used to for inferences about the PFTPR model (4).

Similar to Fine et al. (2004), two assumptions associated with the data availability $\delta(t)$ are required in order to carry out the methodology. The first assumption is

$$E\{Y(t)|X(t), Z(t)\} = E\{Y(t)|X(t), Z(t), \delta(t) = 1\},$$

for each $t \in [l, u]$. That is, conditioning on covariate $X(t)$ and $Z(t)$, the mean of the response $Y(t)$ is the same whether or not the data is observed. This is analogous to the "non-informative censoring" in the standard survival analysis, and to "missing at random" in longitudinal data analysis (Little and Rubin 2002). The second assumption is positive probability of complete data, that is,

$$\Pr\{\delta(t) = 1|X(t), Z(t)\} > 0$$

for each $t \in [l, u]$. This assumption ensures at least some information for the cross-sectional data at each time point.

Estimation of the parameters in model (4) is challenging. Due to the presence of time-independent coefficients θ, the estimate of the time-varying coefficients $\alpha(t)$ cannot be obtained from separate estimating functions at each t as done for the fully functional model (2) in Fine et al. (2004). The proposed estimation procedure obtains the estimate $\check{\alpha}(t, \theta)$ for a given θ from one estimating function and substitute $\alpha(t)$ with it to construct another estimating function for θ.

We now introduce the notations used in the estimating functions. Let h be the inverse function of the link function g. Let $\mu_i\{\alpha(t), \theta\} = h\{X_i^\top(t)\alpha(t) + Z_i^\top(t)\theta\}$, $D_i\{\alpha(t), \theta\} = \mathrm{d}h(v)/\,\mathrm{d}v|_{v=\mu_i\{\alpha(t),\theta\}}$, and $V_i\{\alpha(t), \theta\} = V[\mu_i\{\alpha(t), \theta\}]$, where V is the variance function in a GLM setting.

For a given θ and $t \in [l, u]$, we estimate $\alpha(t)$ with the root of the estimating function

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i(t) D_i\{\alpha(t), \theta\} V_i^{-1}\{\alpha(t), \theta\} X_i(t)[Y_i(t) - \mu_i\{\alpha(t), \theta\}] = 0. \tag{6}$$

The nonparametric estimator of $\alpha(t)$ for a given θ, $\check{\alpha}(t;\theta)$, is the quasi-likelihood estimator with response $Y_i(t)$, covariate $X_i(t)$, and an offset $Z_i^\top(t)\theta$, $i = 1, \ldots, n$. Let $l = \tau_0 < \tau_1 < \cdots < \tau_M = u$ be all the jump points observed in the data. Then $\alpha(t)$ only needs to be estimated at each of these jump points. This step is similar to the estimation method used in Fine et al. (2004), except for an offset $Z_i^\top(t)\theta$.

To estimate θ, we propose the estimating function

$$\frac{1}{n}\sum_{i=1}^{n}\int_l^u \delta_i(t)D_i\{\check{\alpha}(t;\theta),\theta\}V_i^{-1}\{\check{\alpha}(t;\theta),\theta\}Z_i(t)[Y_i(t)-\mu_i\{\check{\alpha}(t;\theta),\theta\}]\,\mathrm{d}H_n(t) = 0, \qquad (7)$$

where H_n is an increasing weight function over $[l, u]$, possibly random and converging uniformly to a fixed $H(t)$ with bounded variation. Denote the solution to (7) by $\hat{\theta}$. The corresponding estimator for $\alpha(t)$ is then $\hat{\alpha}(t) = \check{\alpha}(t;\hat{\theta})$.

The above procedure is similar to the profile likelihood method. The difference is that here we apply the idea of profiling to semiparametric estimating functions. Therefore we call this procedure the semiparametric profile estimating function (SPEF) method. Numerically, the SPEF estimate of $\alpha(t)$ and θ are obtained by alternate one-step-updating of the current estimate in solving (6) and (7) until convergence. The initial values of the estimate are hence important. We suggest getting initial values which are consistent from fitting the fully functional model (2) of Fine et al. (2004) separately on a grid including the end points of $[l, u]$. Initial values of $\alpha(t)$ are obtained by linear interpolation using the estimates at the grid points; initial values of θ are obtained as the average of the estimates at the grid points. In the simulation study in Section 5, the algorithm with these initial values converges very fast. For example, 915 out of 1000 replications converged within 5 iterations in one of the configurations with sample size $n = 100$.

3. ASYMPTOTIC PROPERTIES

The presence of the time-independent parameter θ also makes it challenging to establish the asymptotic properties of the SPEF estimators $\hat{\theta}$ and $\hat{\alpha}(t)$. Let θ_0 and $\alpha_0(t)$ be the true coefficients in the PFTPR model (4). We now discuss convergence properties of $\hat{\theta}$ to θ_0 and $\hat{\alpha}(t)$ to $\alpha_0(t)$. The empirical process theory will be used in the derivation, extending the work of Fine et al. (2004) and Lin et al. (2001).

When computing the estimator $\hat{\theta}$ and $\hat{\alpha}$, we start with an initial estimator constructed from the full functional model (2). This initial estimator is consistent by Theorem 1 of Fine et al. (2004). Thus for the asymptotic results of $\hat{\theta}$, we can focus on a local neighborhood of θ_0 and develop the weak convergence results only. We show that, with probability converging to one, there exists an solution $\hat{\theta}$ to the estimating function (7) that is asymptotically normal. The weak convergence of $\hat{\alpha}$ follows from result on $\hat{\theta}$. These results are formally summarized in Theorem 1 and Theorem 2, with detailed regularity conditions R1–R5 and proofs presented in the Appendix. A notation needed to present the theorems is $\|x\|$, the L_2 norm of any vector or matrix x.

Theorem 1. *For a sequence of positive numbers satisfying $\delta_n \searrow 0$, let $\mathcal{N}_{\theta_0,\delta_n} = \{\theta : \|\theta - \theta_0\| \le \delta_n\}$. Let $\bar{z}$ be the almost sure limit of $\bar{Z}$ given in* (A.9). *Suppose that R1–R5 hold and that matrix A given in* (A.12) *is nonsingular. Then, with probability converging to one, there exists a $\hat{\theta}$ in $\mathcal{N}_{\theta_0,\delta_n}$ that is a solution to the estimating function* (7). *Furthermore,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_D N(0, A^{-1}BA^{-1}),$$

with matrix A and B defined in (A.12) *and* (A.17), *respectively.*

Theorem 2. *Suppose that R1–R5 hold and that A is nonsingular. Then, $\sqrt{n}\{\hat{\alpha}(t) - \alpha_0(t)\}$ converges in distribution to a Gaussian process with mean zero and covariance function $\Sigma(s,t) = E\{\jmath_1(s)\jmath_1^\top(t)\}$, where $\jmath_1$ is defined in* (A.19).

The asymptotic covariance matrix $A^{-1}BA^{-1}$ for $\hat{\theta}$ and covariance function $\Sigma(s,t)$ for $\hat{\alpha}$ can be consistently estimated by their empirical version with the true parameters replaced by their estimates. These results are summarized in Theorem 3.

Theorem 3. *Suppose that R1–R5 hold and that A is nonsingular.*

(a) The asymptotic variance-covariance matrix $A^{-1}BA^{-1}$ *in Theorem 1 can be consistently estimated by* $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$*, with* $\hat{A}$ *and* $\hat{B}$ *defined in* (A.20) *and* (A.21).

(b) The asymptotic covariance function Σ *of* $\hat{\alpha}$ *in Theorem 2 can be consistently estimated by* $\hat{\Sigma}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \hat{\jmath}_i(t;\hat{\theta},\hat{\alpha})\hat{\jmath}_i^{\top}(s;\hat{\theta},\hat{\alpha})$*, with* $\hat{\jmath}_i$ *defined in* (A.22)*. In fact,*

$$\sup_{(s,t)\in[l,u]\times[l,u]} \|\hat{\Sigma}(s,t) - \Sigma(s,t)\| \to_p 0.$$

These theoretical properties of the parameter estimates form the foundation for significance test and goodness-of-fit test in Section 4. In addition, the uniform results on $\hat{\alpha}$ can be used to construct confidence bands for the time-varying coefficients with the resampling method of Lin, Fleming, and Wei (1994) and van der Vaart and Wellner (1996).

4. HYPOTHESIS TESTING

Two important hypotheses in models with a component of time-varying regression coefficient are that a regression coefficient is not significant

$$H_0 : \alpha_j(t) = 0 \tag{8}$$

and that a regression coefficient is constant

$$H_c : \alpha_j(t) = \alpha_j. \tag{9}$$

Test statistics can be conveniently constructed using the estimate of $\alpha_j(t)$: time-varying estimate $\hat{\alpha}_j(t)$ and/or time-independent estimate $\hat{\alpha}_j$. The asymptotic distribution of these estimates given by the theorems in Section 3 can then be used to obtain the null distribution of the test statistics. The goodness-of-fit testing problem (9) is more difficult than the significance test problem (8), because it demands the joint distribution of both the time-varying estimate $\hat{\alpha}_j(t)$ and time-independent estimate $\hat{\alpha}_j$. We only study the goodness-of-fit testing problem closely in this section, as the significance testing procedures can be obtained in a similar but easier fashion.

To examine the time-varying effects in model (2), Fine et al. (2004) use the nonparametric estimate from the fully functional model to construct parametric estimate of α_j at a second stage. In particular, their parametric estimator of α_j is

$$\hat{\alpha}_j = \arg\min_{\eta} \int_l^u \{\hat{\alpha}_j(t) - \eta\}^2 W(t)\,\mathrm{d}t, \tag{10}$$

where $W(t)$ is a weight function, possibly data dependent with a limit, and $\hat{\alpha}_j(t)$ is the nonparametric estimate from (2). Different covariate effects are investigated separately, all based on the fully functional model without taking any account of that some covariate effects may have been confirmed to be time-independent.

Model (4) makes successive tests of time-varying effects and backward model selection procedure possible. To test the goodness-of-fit for a constant fit $\alpha_j(t) = \alpha_j$, two partly functional models can be fit, one nested to the other. They differ only in the form of α_j: the full model has α_j time-varying and the restricted model has α_j time-independent. Let $\hat{\alpha}_j^{\mathrm{F}}(t)$ and $\hat{\alpha}_j^{\mathrm{R}}$ be the estimate of α_j from full model and restricted model, respectively. Goodness-of-fit test can be constructed based on their discrepancy

$$\Delta(t) = \hat{\alpha}_j^{\mathrm{F}}(t) - \hat{\alpha}_j^{\mathrm{R}}. \tag{11}$$

Examples test statistics are

$$T_1\left(\hat{\alpha}_j^{\mathrm{F}}(\cdot), \hat{\alpha}_j^{\mathrm{R}}\right) = \sup_{t\in[l,u]} |\Delta(t)W(t)|, \tag{12}$$

$$T_2\left(\hat{\alpha}_j^{\mathrm{F}}(\cdot), \hat{\alpha}_j^{\mathrm{R}}\right) = \int_l^u |\Delta(t)|W(t)\,\mathrm{d}t, \tag{13}$$

$$T_3\left(\hat{\alpha}_j^{\mathrm{F}}(\cdot), \hat{\alpha}_j^{\mathrm{R}}\right) = \int_l^u \Delta^2(t)W(t)\,\mathrm{d}t, \tag{14}$$

where $W(t)$ is a weight function, possibly random with a limit. The weight $W(t)$ may be chosen to inversely weight the variation of $\Delta(t)$. It can also be chosen with higher values on some segment in $[l, u]$ than others to accentuate anticipated departures from H_c. Similar test statistics have been used by Scheike (2004) in a partly functional model for intensities. The limiting distribution of these test statistics depend on the influence functions from the two models, which can be used to approximate the limiting Gaussian process of $\Delta(t)$. The limiting distribution of these test statistics need to be obtained via bootstrapping (Lin et al. 1994; van der Vaart and Wellner 1996). The size and power of these tests are studied via simulation in the next section.

5. SIMULATION STUDIES

To investigate the performance of the SPEF estimators and test procedure, we conduct simulation studies in a recurrent event setup. Let $Y(t)$ be the number of events that have occurred by time t. We include two time-independent covariates in the model: X is a binary covariate, with probability 0.5 being 1, and Z is a truncated standard normal covariate in $(-2, 2)$. The true model is

$$E\{Y(t)|\psi, X, Z\} = \psi \exp\{\alpha_0(t) + \alpha_1(t)X + \theta_1 Z\}, \tag{15}$$

where ψ is an independent gamma variate with mean 1 and variance σ^2, $\alpha_0(t) = \log(t)$ is a time-varying intercept, $\alpha_1(t)$ is the to-be-configured coefficient of X, and $\theta_1 = 0.5$ is the time-independent coefficient of Z. Censoring times are independently generated from a uniform distribution over $(0, 1/p)$, resulting $100p$ percent of censoring by time $t = 1$. Due to numerical instability of $\log(t)$ as $t \searrow 0$, the observation window $[l, u]$ is chosen to be $[0.15, 1]$.

There are four factors to be configured: time-varying coefficient $\alpha_1(t)$, variance σ^2 of the random effect ψ, censoring proportion p, and sample size n. For $\alpha_1(t)$, we consider three choices: $f_1(t) = 1$, $f_2(t) = 8t/3$ if $t < 0.5$, $4/3$ otherwise, and $f_3(t) = 1 - \cos(\pi t)$. These functions are plotted in Figure 1. Note that the average of these functions are the same over $(0, 1)$. That is, all three situations would yield the same estimate if the coefficient is specified to be time-independent over $(0, 1)$. Four values of σ^2 are considered: 0, 0.25, 0.5, and 1. This simulation setup is a modification of those used by Lin et al. (2001) and Fine et al. (2004). Two values of p are considered: 0.3 and 0.6, corresponding to moderate and heavy censoring. Three values of n are considered: 100, 200, and 400, in order to find how big n is needed for the large sample results. For each configuration, we generate 1,000 datasets.

[Figure 1 about here.]

We first study the performance of the time-independent estimator and time-varying estimator when the coefficient $\alpha_1(t)$ is specified as time-varying and θ_1 is specified as time-independent. Therefore, the model is correctly specified when $\alpha_1(t)$ is $f_2(t)$ or $f_3(t)$, but is misspecified when $\alpha_1(t) = f_1(t) = 1$. Table 1 summarizes the results for the estimation of θ_1. The estimator $\hat{\theta}_1$ is virtually unbiased under all configurations. The sampling mean of the standard error estimator in general agrees with the sampling standard error. For sample size $n = 100$, this agreement is weakened as σ^2 increases or as p increases. As a consequence, the empirical coverage probability of 95% confidence intervals are slightly lower than the nominal level for larger values σ^2 and p. When the sample size is increased to $n = 400$, the proposed standard error estimator reflects well the true variability of the estimators and the the 95% confidence intervals have coverage probabilities close to the nominal level.

[Table 1 about here.]

Summary of the estimator of $\alpha_1(t)$ is done graphically in Figure 2. To suppress the result, only the most difficult situation for each choice of $\alpha_1(t)$ is presented: $\sigma^2 = 1$ and $p = 0.6$. The sample size is $n = 400$. As evident from the Figure, the proposed estimators are practically unbiased, and the proposed standard error estimator agrees well with the empirical standard error.

[Figure 2 about here.]

We then study the efficiency gain of in the SPEF estimation of the coefficient of Z, $\theta_1(t)$, by constraining it to be constant. The estimator of $\theta_1(t)$ obtained from two models are compared. In the first model, all covariate coefficients, including $\theta_1(t)$, are specified as time-varying. The second model is the same as the first model, except that the $\theta_1(t)$ is specified as a constant θ_1. This may be necessary when there is strong belief from substantive expert or after a goodness-of-fit test. Figure 3 presents the bias $\hat{\theta}_1(t) - \theta_1$ as well as its pointwise 95% confidence interval for the most difficult case for each choice of $\alpha_1(t)$ with $\sigma^2 = 1$ and $p = 0.6$ at sample size $n = 400$. The dark lines are from the unconstrained specification and gray lines are from the constrained specification. The variation level of the estimator $\hat{\theta}_1$ under constant constraint is smaller than that of the nonparametric estimator $\hat{\theta}_1(t)$ for all $t \in [l, u]$. This implies that the prediction interval from a correctly specified partly functional model is narrower than that from a fully functional model.

[Figure 3 about here.]

We finally investigate the performance of the goodness-of-fit test. This test essentially compares two models: with or without constraining $\alpha_1(t)$ to be constant, assuming that the time-independent coefficient θ_1 has already been correctly specified. The null hypothesis $H_0 : \alpha_1(t) = \gamma$. Table 2 reports the rejection rate of the goodness-of-fit test with size 0.05 under various configurations. The test statistics are T_1, T_2, and T_3 with unit weight. When $\alpha_1(t) = f_1(t)$, these rejection rates are the empirical size of the test. When $\alpha_1(t) = f_2(t)$ or $f_3(t)$, these rejection rates are the empirical power of the test. The results in Table 2 shows that the empirical size of the test is slightly over the nominal size when $n = 100$ but their disagreement drops to a reasonable level when $n = 400$. All three test statistics work well, with T_1 having the highest power but highest discrepancy in nominal and empirical size. The power of the test is shown to be substantial. The time-varying nature of $f_2(t)$ is much weaker than that of $f_3(t)$. Therefore, the power of test when $\alpha_1(t) = f_2(t)$ is lower than that when $\alpha_1(t) = f_3(t)$. Even for the more difficult case of $\alpha_1(t) = f_2(t)$ with $\sigma^2 = 1$, $p = 0.6$ and $n = 100$, the test statistic T_1 rejects the constant coefficient H_0 49.4% of the the time. In the case of $\alpha_1(t) = f_3(t)$, all the test rejects the constant coefficient H_0 almost all the time even for $n = 100$.

[Table 2 about here.]

6. EXAMPLE

We illustrate through an example the SPEF estimation of partly functional temporal process regression and its usage in successive goodness-of-fit testing for the constancy of covariate coefficients one at a time. Consider the rhDNase data (Therneau and Hamilton 1997) which has been used by Lin et al. (2001) to motivate their semiparametric transformation model. This data is from a double blind clinical trial to assess the efficacy of rhDNase, a highly purified recombinant enzyme, in treating cystic fibrosis patients, who often suffer from repeated pulmonary exacerbation. There were 321 patients on rhDNase and 324 on placebo. Most of the patients were followed for 24 weeks. Many of them experienced multiple episodes of exacerbations. The most important covariate is FEV, the baseline record of forced expiratory volume in the first second. Researchers are interested in assessing the efficacy of rhDNase on reducing the number of exacerbations $Y(t)$ after taking into account other significant risk factors. In this section, we compare temporal process regression models with two covariates: treatment indicator TRT (rhDNase = 1) and FEV. Existing analyses (Lin et al. 2001) have indicated that the treatment effect may not be adequately captured by a proportional means model. Using the partly temporal process regression methodology, we will find marginal evidence that the coefficient of treatment is time-varying, given that the coefficient of FEV is specified as constant over time.

The fully functional temporal process regression model (Fine et al. 2004) for $Y(t)$ is

$$E\{Y(t)|\mathrm{TRT},\mathrm{FEV}\} = \exp\{\alpha_0(\mathrm{t}) + \alpha_1(\mathrm{t})\mathrm{TRT} + \theta_1(\mathrm{t})\mathrm{FEV}\}, \qquad (16)$$

where $\alpha_0(t)$ is a nonparametric intercept, and $\alpha_1(t)$ and $\theta_1(t)$ are time-varying covariate coefficients for TRT and FEV, respectively. The link function used here is the logarithm function, which is common for count data in a GLM setting. A proportional means model (Lin et al. 2000) constrains $\alpha_1(t) = \alpha_1$ and $\theta_1(t) = \theta$ in model (16). Denote model (16) by M0 and the constrained model by M3. There are two intermediate partly functional models between M0 and M3. Model M1 constrains $\alpha_1(t) = \alpha_1$ in M0 and model M2 constrains

$\theta_1(t) = \theta_1$ in M0. Both model M1 and model M2 nest M3. All models are fitted over time window $[10, 160]$ days.

The coefficient estimates and their pointwise 95% confidence intervals for the fully functional model M0 are displayed as gray lines in Figure 4. The same quantities for model M2, where the coefficient of FEV is constrained to be time-independent, are overlayed as dark lines in Figure 4. The time-independent estimate $\hat{\theta}_1$ in model M2 is close to the time-varying estimate $\hat{\theta}_1(t)$ in model M0, but with a smaller variation, particularly for the early time period. The estimate $\hat{\alpha}_1(t)$ from the two models are virtually the same. The estimate $\hat{\alpha}_0(t)$ from the two models are only slightly different during the early time period, which is not surprising given the difference in $\hat{\theta}_1(t)$ in the same time period. We have also overlayed the results of model M0 and model M1, which gives similar observations and hence is not reported.

[Figure 4 about here.]

The estimation results from model M2 and model M3 are overlayed in Figure 5, with gray lines representing model M2 and dark lines representing model M3. Both models constrain the effect of FEV to be time-independent and yield virtually the same point estimate and confidence interval. The time-independent estimate $\hat{\alpha}_1$ in model M3 seems to be averaging the time-varying $\hat{\alpha}_1(t)$ in model M2. The intercept estimate $\hat{\alpha}_0(t)$ from the two models are also very close.

[Figure 5 about here.]

Both covariate coefficients have been shown to be significantly different from zero in existing analyses. Therefore, we now focus on the goodness-of-fit of constant coefficient for each covariate. Figure 4 suggests a constant model for the FEV coefficient. A formal goodness-of-fit test for $H_0 : \theta_1(t) = \theta_1$ in model M0 can be carried out by comparing the full model M0 and the constrained model M2. The test statistics in Section 4 are obtained with 2,000 bootstrapping samples. The observed test statistics have p-values 0.470, 0.707, and 0.677, suggesting that a constant FEV coefficient may be appropriate. The goodness-of-fit test of constant TRT effect in model M0 yields p-values 0.055, 0.061, and 0.067, which provides marginal evidence against constant TRT coefficient. From the graphical display and these test results, we proceed with the time-independent specification of the FEV effect and conditioning on this specification, test the goodness-of-fit of constant TRT effect again. This time, the models under comparison are the full model M2 and the constrained model M3 presented in Figure 5. The goodness-of-fit test results p-values 0.042, 0.048, and 0.052, still marginal but slightly stronger evidence against the constant TRT coefficient. Note that this comparison of M3 and M2 can not be obtained from the fully functional temporal process regression methodology developed in Fine et al. (2004). The successive goodness-of-fit test is made possible by the proposed partly functional model.

As an illustration, Figure 6 presents the observed and the first 50 simulated Gaussian processes used in the goodness-of-fit test above for constant covariate coefficient. We observe that the most extreme deviation in the goodness-of-fit testing for constant TRT effect occurs in the latter time period in Figure 6, indicating that the treatment efficacy of rhDNase may be diminishing after about 100 days. This finding is consistent with those in Therneau and Hamilton (1997) and Lin et al. (2001).

[Figure 6 about here.]

7. DISCUSSION

The proposed PFTPR model is in the line of marginal mean models. It has the most general form and covers as special cases the wide model classes, for example, Fine et al. (2004), Lin et al. (2001), and subclasses therein. The proposed model advances the mean modeling of temporal processes to a new level of flexibility comparable to that of intensity models. The significance of covariate effects and the goodness-of-fit of their parametric forms can be tested sequentially and used in model selection.

The model parameters are interpreted marginally for each t. In the presence of time-varying covariates, the regression coefficient should be interpreted with care. As distinguished by Kalbfleisch and Prentice (2002), time-varying covariates can be internal or external. Marginal mean models are easier to interpret with external time-varying covariates. The methodology can be of particularly help in understanding the natural history of disease progression with only mild assumptions. Use of internal time-varying covariates, on the other hand, can be misleading, especially in assessing treatment effect in randomized studies.

Straightforward application of the standard estimating function method is not appropriate for the PFTPR model, because it contains a finite dimensional parameter and a nonparametric parameter. To deal with the presence of two different kinds of parameters, we proposed the SPEF approach, which extends the separate estimating function approach and keeps the advantage of no smoothing from Fine et al. (2004). While the SPEF method is designed specifically for the PFTPR, it can be used as a general procedure for solving semiparametric estimating function problems with both finite- and infinite-dimensional parameters. Some related works on profile estimating functions are developed under the presence of nuisance parameters (Wang and Hanfelt 2003; Jørgensen and Knudsen 2004). In our case, both kinds of parameters are of interests. The plug-in of $\check{\alpha}$ in the SPEF approach has consequences in the asymptotic properties of the resulting estimator in terms of both bias and variance. They are accounted for in our derivation of the weak convergence of $\hat{\theta}$ in Theorem 1. A similar treatment was used by Lin et al. (2001).

The efficiency of the SPEF estimator from the alternate updating algorithm may be improved by accounting for the temporal correlation of the observations from the same temporal process. The estimation of the fully functional regression coefficients in Fine et al. (2004) is a functional extension of the generalized estimating function (Liang and Zeger 1986) under "working independence". It has been criticized for its low efficiency, particularly for time points where censoring is heavier. To improve the efficiency, Fine et al. (2004) proposed a one-step update of the "working independence" estimator by the generalized estimating function approach applied to the discretized temporal processes on a fine grid. This method is not practical for two reasons. First, a dense grid yields extremely high correlation between neighboring points, which may cause numerical instability if a working correlation matrix as close as possible to the true temporal correlation is to be used. Second, even if the inversion of the working correlation matrix were not a problem, the inversion is expensive when the grid is dense. This efficiency problem still exists with the proposed estimation procedure for the PFTPR model. It is an interesting topic of future research how to increase the efficiency of the estimator by better accounting for the temporal correlation with methods that are computationally tractable and numerically stable.

A. PROOFS OF THE THEOREMS

We now prove the results state in Section 3. We assume the following regularity conditions, which are similar to those in Fine et al. (2004) with some modifications.

R1: The time-varying coefficient $\alpha_0(t)$ is right-continuous with left-hand limits, or cadlag. In addition, $\sup_{t\in[l,u]} \|\alpha_0(t)\| < \infty$.

R2: The inverse link function h and its derivative $\dot{h}$ are Lipschitz continuous and bounded on compact sets, and at θ_0, the variance function V_1 is differentiable.

R3: $\{Y_i(t), X_i(t), Z_i(t), \delta_i(t) : i = 1, \ldots, n\}$ are independent and identically distributed, $\{X_i^\top(t), Z_i^\top(t)\}^\top$ are cadlag and have total variation on $[l, u]$ bounded by some constant $c < \infty$, almost surely, and the total variation of $Y_i(t)$ has bounded second moment.

R4: $\inf_{t\in[l,u]} \lambda_{\min}[E\{\delta_1(t)W_1(t)W_1^\top(t)\}] > 0$, where $\lambda_{\min}$ is the smallest eigenvalue of a matrix and $W_1(t) = \{X_1^\top(t), Z_1^\top(t)\}^\top$.

R5: Write $V_1(t; \alpha, \theta) = V_1\{\alpha(t), \theta\}$. For all bounded functions $\mathcal{A}$ in R^q and $\theta \in R^p$, the class of random functions $\{V_1(t; \alpha, \theta) : \alpha \in \mathcal{A}, t \in [l, u]\}$ is bounded above and below by positive constants, is bounded in uniform entropy integral (van der Vaart and Wellner 1996, p.58) with bounded envelope, and is pointwise measurable (van der Vaart and Wellner 1996, p.111).

We note that R5 is satisfied if $V_1(t;\alpha,\theta) = V[X^\top(t)\alpha(t) + Z^\top(t)\theta]$ for a nonrandom Lipschitz continuous function V.

For simplicity, we only prove the results for a fixed $H_n(t) \equiv H(t)$. Only minor modifications are needed for H_n that converges uniformly to H with bounded variation over $[l,u]$. Denote

$$\begin{aligned} w_{1i}(t;\theta,\alpha) &= \delta_i(t) D_i\{\alpha(t),\theta\} V_i^{-1}\{\alpha(t),\theta\} X_i(t), \\ w_{2i}(t;\theta,\alpha) &= \delta_i(t) D_i\{\alpha(t;\theta),\theta\} V_i^{-1}\{\alpha(t;\theta),\theta\} Z_i(t). \end{aligned}$$

Let

$$\begin{aligned} U_{1n}(\alpha;t,\theta) &= \frac{1}{n}\sum_{i=1}^n w_{1i}(t;\theta,\alpha)[Y_i(t) - \mu_i\{\alpha(t),\theta\}], \\ U_{2n}(\theta;\alpha) &= \frac{1}{n}\sum_{i=1}^n \int_l^u w_{2i}(t;\theta,\alpha)[Y_i(t) - \mu_i\{\alpha(t;\theta),\theta\}]\,\mathrm{d}H(t). \end{aligned}$$

Then the estimating functions (6) and (7) can be written as

$$U_{1n}(\alpha;t,\theta) = 0, \quad \text{and} \quad U_{2n}(\theta;\check{\alpha}) = 0. \tag{A.1}$$

The following Lemmas are needed for the proofs.

Lemma 1. *Under the regularity conditions R1–R5,*

$$U_{2n}(\theta_0;\check{\alpha}) = U_{2n}(\theta_0,\alpha_0) + \frac{1}{n}\sum_{i=1}^n \int_l^u w_{2i}(t;\theta_0,\alpha_0) D_i\{\alpha_0(t),\theta_0\}[\check{\alpha}(t;\theta_0) - \alpha_0(t)]\,\mathrm{d}H(t) + o_p(n^{-1/2}).$$

Proof. With slight abuse of notation in this proof, we write $w_{2i}(t;\alpha) = w_{2i}(t;\theta_0,\alpha)$ and $\check{\alpha}(t;\theta_0) = \check{\alpha}(t)$. By the definition of U_{2n} and some algebra, we have

$$U_{2n}(\theta_0,\check{\alpha}) - U_{2n}(\theta_0,\alpha_0) = I_{1n} + I_{2n}, \tag{A.2}$$

where

$$I_{1n} = \frac{1}{n}\sum_{i=1}^n \int_l^u w_{2i}(t;\check{\alpha})[-\mu_i\{\check{\alpha}(t),\theta_0\} + \mu_i\{\alpha_0(t),\theta_0\}]\,\mathrm{d}H(t),$$

and

$$I_{2n} = \frac{1}{n}\sum_{i=1}^n \int_l^u [w_{2i}(t;\check{\alpha}) - w_{2i}(t;\alpha_0)][Y_i(t) - \mu_i\{\alpha_0(t),\theta_0\}]\,\mathrm{d}H(t).$$

First consider I_{2n}. By Theorem A1 in Fine et al. (2004), $\sup_{t\in[l,u]} \|\breve{\alpha}(t;\theta_0) - \alpha_0(t)\| = o_p(1)$. Therefore, under conditions R2 and R4,

$$E\|w_{2i}(t;\breve{\alpha}) - w_{2i}(t;\alpha_0)\|^2 \to_p 0.$$

By the definition of the model, $E[Y_i(t) - \mu_i(t;\alpha_0)] = 0$. Furthermore, the class

$$\{[w_{2i}(t;\alpha) - w_{2i}(t;\alpha_0)][Y_i(t) - \mu_i\{\alpha_0(t),\theta_0\}] : \alpha \text{ is cadlag}, t \in [l,u]\}$$

is Donsker by Theorem 2.5.2 of van der Vaart and Wellner (1996). Therefore, under R1–R5, by Lemma 19.24 of van der Vaart (1998), we have $I_{2n} = o_p(n^{-1/2})$.

Now consider I_{1n}. Under R2, by Taylor expansion and $\sup_{t\in[l,u]} \|\breve{\alpha}(t;\theta_0) - \alpha_0(t)\| = O_p(n^{-1/2})$ (Fine et al. 2004)

$$-\mu_i\{\breve{\alpha}(t),\theta_0\} + \mu_i\{\alpha_0(t),\theta_0\} = -D_i\{\alpha_0(t)\}X_i^\top(t)[\breve{\alpha}(t) - \alpha_0(t)] + o_p(n^{-1/2}). \tag{A.3}$$

Thus

$$\sup_{t\in[l,u]} |\mu_i\{\breve{\alpha}(t),\theta_0\} - \mu_i\{\alpha_0(t),\theta_0\}| = O_p(n^{-1/2}).$$

Similarly,

$$\sup_{t\in[l,u]} \|w_{2i}(t;\breve{\alpha}) - w_{2i}(t;\alpha_0)\| = O_p(n^{-1/2}).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\int_l^u [w_{2i}(t;\breve{\alpha}) - w_{2i}(t;\alpha_0)][-\mu_i\{\breve{\alpha}(t),\theta_0\} + \mu_i\{\alpha_0(t),\theta_0\}]\,\mathrm{d}H(t) = O_p(n^{-1}).$$

It follows that

$$I_{1n} = \frac{1}{n}\sum_{i=1}^{n}\int_l^u w_{2i}(t;\alpha_0)[-\mu_i\{\breve{\alpha}(t),\theta_0\} + \mu_i\{\alpha_0(t),\theta_0\}]\,\mathrm{d}H(t) + o_p(n^{-1/2}). \tag{A.4}$$

Combining (A.2), (A.3), and (A.4), Lemma 1 follows. □

Lemma 2. *Let $\dot{U}_{2n}$ be the derivative of U_{2n} with respect to θ. Suppose that R1–R5 hold. For any $\theta \in \mathcal{N}_{\theta_0,\delta_n}$, we have*

$$U_{2n}(\theta;\breve{\alpha}) = U_{2n}(\theta_0,\breve{\alpha}) + \dot{U}_{2n}(\theta_0;\breve{\alpha})(\theta - \theta_0) + o_p(\|\theta - \theta_0\|).$$

Proof. Let $\dot{U}_{2n}$ be the derivative of U_{2n} with respect to θ. By Taylor expansion,

$$U_{2n}(\theta;\breve{\alpha}) = U_{2n}(\theta_0,\breve{\alpha}) + \dot{U}_{2n}(\theta_0;\breve{\alpha})(\theta-\theta_0) + [\dot{U}_{2n}(\tilde{\theta};\breve{\alpha}) - \dot{U}_{2n}(\theta_0;\breve{\alpha})](\theta-\theta_0), \tag{A.5}$$

for some $\tilde{\theta} \in \mathcal{N}_{\theta_0,\delta_n}$. So to prove the lemma, it suffices to show that

$$\sup_{\theta\in\mathcal{N}_{\theta_0,\delta_n}} \|\dot{U}_{2n}(\theta;\breve{\alpha}) - \dot{U}_{2n}(\theta_0;\breve{\alpha})\| = o_p(1). \tag{A.6}$$

Direct calculation shows

$$\dot{U}_{2n}(\theta;\breve{\alpha}) = n^{-1}\sum_{i=1}^{n}\int_l^u \delta_i(t)D_i^2\{\breve{\alpha}(t;\theta),\theta\}V_i^{-1}\{\breve{\alpha}(t;\theta),\theta\}Z_i(t)\left\{Z_i^\top(t) + X_i^\top(t)\frac{\partial\breve{\alpha}^\top(t;\theta)}{\partial\theta}\right\}\mathrm{d}H(t). \tag{A.7}$$

Differentiation of the first equation in (A.1) with respect to θ yields

$$\frac{\partial\breve{\alpha}^\top(t;\theta)}{\partial\theta} = -\bar{Z}(t,\theta,\breve{\alpha}), \tag{A.8}$$

where

$$\begin{aligned}\bar{Z}(t,\theta,\alpha) &= \left[\sum_{i=1}^{n}\delta_i(t)D_i^2\{\alpha(t),\theta\}V_i^{-1}\{\alpha(t),\theta\}X_i(t)X_i^\top(t)\right]^{-1}\times \\ &\quad \left[\sum_{i=1}^{n}\delta_i(t)D_i^2\{\alpha(t),\theta\}V_i^{-1}\{\alpha(t),\theta\}X_i(t)Z_i^\top(t)\right].\end{aligned} \tag{A.9}$$

Expressions (A.7), (A.8), and (A.9) show that $\dot{U}_{2n}$ can be expressed in terms of $\breve{\alpha}$ and does not depend on the derivative of $\breve{\alpha}$ with respect to θ. Now, by the triangle inequality,

$$\begin{aligned}\|\breve{\alpha}(t;\theta)-\alpha_0(t)\| &\le \|\breve{\alpha}(t;\theta)-\breve{\alpha}(t;\theta_0)\| + \|\breve{\alpha}(t;\theta_0)-\alpha_0(t)\|. \\ &= \|\bar{Z}^\top(t,\tilde{\theta},\breve{\alpha})(\theta-\theta_0)\| + \|\breve{\alpha}(t;\theta_0)-\alpha_0(t)\|,\end{aligned} \tag{A.10}$$

for some $\tilde{\theta} \in \mathcal{N}_{\theta_0,\delta_n}$. Under R1–R5, $\bar{Z}^\top(t,\tilde{\theta},\breve{\alpha})$ is bounded in probability and $\sup_{t\in[l,u]}\|\breve{\alpha}(t;\theta_0)-\alpha_0(t)\| = o_p(1)$ (Fine et al. 2004). It follows that

$$\sup_{\|\theta-\theta_0\|\le\delta_n, t\in[l,u]} \|\breve{\alpha}(t;\theta)-\alpha_0(t)\| = o_p(1).$$

Therefore, (A.6) follows from R2, R3, and R5. Now Lemma 2 follows from (A.5) and (A.6). □

Proof of Theorem 1. The existence of a solution to (7) with high probability follows from the asymptotic linearity of $U_{2n}(\theta, \check{\alpha})$ in $\mathcal{N}_{\theta_0,\delta_n}$ in Lemma 2. Furthermore, by Lemma 2,

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\dot{U}_{2n}^{-1}(\theta_0, \check{\alpha})\sqrt{n}\, U_{2n}(\theta_0, \check{\alpha}) + o_p(1). \tag{A.11}$$

By (A.7), (A.8), (A.9) and some algebra, we have

$$\dot{U}_{2n}(\theta_0, \check{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\int_l^u \delta_i(t)D_i^2\{\check{\alpha}(t), \theta_0\}V_i^{-1}\{\check{\alpha}(t), \theta\}Z_i(t)\{Z_i^\top(t) - X_i^\top(t)\bar{Z}(t; \theta_0, \check{\alpha})\}\, \mathrm{d}H(t) + o_p(1).$$

By the uniform convergence of $\check{\alpha}(\cdot; \theta_0)$ to α_0 (Fine et al. 2004), we have

$$\dot{U}_{2n}(\theta_0, \check{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\int_l^u \delta_i(t)D_i^2\{\alpha_0(t), \theta_0\}V_i^{-1}\{\alpha_0(t), \theta\}Z_i(t)\{Z_i^\top(t) - X_i^\top(t)\bar{Z}(t; \theta_0, \alpha_0)\}\, \mathrm{d}H(t) + o_p(1),$$

By the uniform strong law of large number, $\bar{Z}(t; \theta_0, \alpha_0)$ converges almost surely to a nonrandom function $\bar{z}(t; \theta_0, \alpha_0)$ uniformly for $t \in [l, u]$. Therefore, $\dot{U}_{2n}(\theta_0, \check{\alpha})$ converges almost surely to a nonrandom function $A(\theta_0)$. Let $A = A(\theta_0)$ and $\bar{z}(t) = \bar{z}(t; \theta_0, \alpha_0)$. Then

$$A = E\int_l^u \delta_i(t)D_i^2\{\alpha_0(t), \theta\}V_i^{-1}\{\alpha_0(t), \theta\}Z_i(t)\{Z_i^\top(t) - X_i^\top(t)\bar{z}(t)\}\, \mathrm{d}H(t). \tag{A.12}$$

By Lemma 2,

$$U_{2n}(\theta_0; \check{\alpha}) = U_{2n}(\theta_0, \alpha_0) + \frac{1}{n}\sum_{i=1}^{n}\int_l^u w_{2i}(t; \theta_0, \alpha_0)D_i\{\alpha_0(t), \theta_0\}[\check{\alpha}(t; \theta_0) - \alpha_0(t)]\, \mathrm{d}H(t). \tag{A.13}$$

By Theorem 2 of Fine et al. (2004),

$$\begin{aligned}\check{\alpha}(t; \theta_0) - \alpha_0(t) \quad &= \quad \left[\sum_{i=1}^{n}\delta_i(t)D_i^2\{\alpha_0(t), \theta_0\}V_i^{-1}\{\alpha_0(t), \theta_0\}X_i(t)X_i^\top(t)\right]^{-1} \times \\ &\quad\quad \left[\sum_{i=1}^{n}X_i(t)M_i(t; \theta_0, \alpha_0)\right] + o_p(n^{-1/2}),\end{aligned} \tag{A.14}$$

where

$$M_i(t; \theta, \alpha) = \delta_i(t)D_i\{\alpha(t), \theta\}V_i\{\alpha(t), \theta\}\left[Y_i(t) - \mu_i\{\alpha(t), \theta\}\right]. \tag{A.15}$$

By putting (A.14) into (A.13) and switching the order of the obtained double summation, we have

$$U_{2n}(\theta_0; \check{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\kappa_i + o_p(1), \tag{A.16}$$

where

$$\kappa_i = \int_l^u M_i(t;\theta_0,\alpha_0)\{Z_i(t) - X_i(t)\bar{z}(t)\}\,\mathrm{d}H(t).$$

Multivariate central limit theorem implies that $n^{1/2}U_{2n}(\theta_0;\breve{\alpha})$ converges in distribution to $N(0,B)$, where

$$B = E\kappa_1\kappa_1^\top. \tag{A.17}$$

This result and (A.11) then complete the proof. □

Proof of Theorem 2. First note that $\hat{\alpha}(t) = \breve{\alpha}(t;\hat{\theta})$ and

$$\begin{aligned}\sqrt{n}\{\hat{\alpha}(t) - \alpha_0(t)\} &= \sqrt{n}\{\breve{\alpha}(t;\theta_0) - \alpha_0(t)\} + \sqrt{n}\{\breve{\alpha}(t;\hat{\theta}) - \breve{\alpha}(t;\theta_0)\} \\ &= \sqrt{n}\{\breve{\alpha}(t;\theta_0) - \alpha_0(t)\} - \bar{z}(t)\sqrt{n}(\hat{\theta} - \theta_0) + o_p(1),\end{aligned}$$

where the second equality follows from the Taylor expansion of $\hat{\alpha}$, $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$, expression (A.8), and the convergence of $\bar{Z}(t)$. After some lengthy calculation, we have

$$\sqrt{n}\{\hat{\alpha}(t) - \alpha_0(t)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \jmath_i(t) + o_p(1), \tag{A.18}$$

where

$$\jmath_i(t) = \left[E\{\delta_i(t)D_i^2\{\alpha_0(t),\theta\}V_i^{-1}\{\alpha_0(t),\theta\}X_i(t)X_i^\top(t)\}\right]^{-1} X_i(t)M_i(t) - A^{-1}\bar{z}(t)\kappa_i. \tag{A.19}$$

Now the class $\{\jmath_1 : t \in [l,u]\}$ is a Donsker class. From the functional central limit theorem, $\sqrt{n}\{\hat{\alpha}(t) - \alpha_0(t)\}$ converge to a zero-mean Gaussian process with covariance function $\Sigma(s,t) = E\{\jmath_1(s)\jmath_1(t)\}$ at $(s,t) \in [l,u]^2$. □

To obtain consistent estimator for variance matrix $A^{-1}BA^{-1}$ we consider the sample versions of matrix A and B. Let

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \int_l^u \delta_i D_i^2(\hat{\alpha}(t),\hat{\theta})V_i^{-1}\{\hat{\alpha}(t),\hat{\theta})Z_i(t)\{Z_i(t)^\top - X_i^\top \bar{Z}(t;\hat{\theta},\hat{\alpha})\}\,\mathrm{d}H_n(t), \tag{A.20}$$

and

$$\hat{B} = \frac{1}{n}\sum_{i=1}^n \hat{\kappa}_i\hat{\kappa}_i^\top, \tag{A.21}$$

where

$$\hat{\kappa}_i = \int_l^u M_i(t; \hat{\theta}, \hat{\alpha})\{Z_i(t) - X_i(t)\bar{Z}(t; \hat{\theta}, \hat{\alpha})\}\, \mathrm{d}H_n(t).$$

Our estimator of $A^{-1}BA^{-1}$ is then $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$. Similarly, to estimate $\Sigma(s,t)$, we consider sample versions of the influence functions $\jmath_i$'s. Let

$$\hat{\jmath}_i(t; \theta, \alpha) = w_n^{-1}(t; \theta, \alpha) X_i(t) M_i(t; \theta, \alpha) - \hat{A}^{-1}\bar{Z}(t; \theta, \alpha)\kappa_i(t; \theta, \alpha), \tag{A.22}$$

where

$$w_n(t; \theta, \alpha) = \frac{1}{n}\sum_{i=1}^n \delta_i(t) D_i^2\{\alpha(t), \theta\} V_i^{-1}\{\alpha(t), \theta\} X_i(t) X_i^\top(t).$$

Out estimator of $\Sigma(s,t)$ is $\hat{\Sigma}(s,t) = \frac{1}{n}\sum_{i=1}^n \hat{\jmath}_i(s; \hat{\theta}, \hat{\alpha})\hat{\jmath}_i^\top(t; \hat{\theta}, \hat{\alpha})$.

Proof of Theorem 3. Write

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \psi_i(\hat{\theta}, \hat{\alpha}),$$

where

$$\psi_i(\theta, \alpha) = \int_l^u \delta_i D_i^2\{\alpha(t), \theta\} V_i^{-1}\{\alpha(t), \theta\} Z_i(t)\{Z_i(t)^\top - X_i^\top \bar{Z}(t; \theta, \alpha)\}\, \mathrm{d}H_n(t).$$

Under R1–R3 and R5, first applying Glivenko-Cantelli to $\bar{Z}$, we have

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \psi_i^*(\hat{\theta}, \hat{\alpha}) + o_p(1),$$

where

$$\psi_i^*(\theta, \alpha) = \int_l^u \delta_i D_i^2\{\alpha(t), \theta\} V_i^{-1}\{\alpha(t), \theta) Z_i(t)\{Z_i(t)^\top - X_i^\top \bar{z}(t)\}\, \mathrm{d}H_n(t).$$

For a finite $c > 0$, let $\mathcal{A} = \{\alpha : \alpha \text{ is cadlag and } \sup_{t\in[l,u]} |\alpha(t)| \le c\}$. Since $\mathcal{A}$ is bounded in uniform entropy integral (van der Vaart and Wellner 1996, (2.5.1), page 127), under R1, R2, and R5, the class $\{\psi_1^*(\theta, \alpha) : \theta \in \Theta, \alpha \in \mathcal{A}\}$ is Donsker (van der Vaart and Wellner 1996, Theorem 2.5.2), and hence Glivenko-Cantelli. It follows that, under R1–R5,

$$\hat{A} = E\psi_1^*(\hat{\theta}, \hat{\alpha}) + o_p(1) = E\psi_1^*(\theta_0, \alpha_0) + o_p(1) = A + o_p(1).$$

Similarly, it can be shown that $\hat{B} = B + o_p(1)$. So part (a) follows from Slutzky's lemma.

Part (b) can be proved similarly by using the Glivenko-Cantelli Theorem. □

REFERENCES

Aalen, O. O. (1980), "A model for non-parametric regression analysis of counting processes," in *Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory*, eds. Klonechi, W. and Rosinski, J., pp. 1–25.

Cox, D. R. (1972), "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, Series B, Methodological*, 34, 187–220.

Fine, J., Yan, J., and Kosorok, M. (2004), "Temporal Process Regression," *Biometrika*, 91, 683–703.

Jørgensen, B. and Knudsen, S. J. (2004), "Parameter Orthogonality and Bias Adjustment for Estimating Functions," *Scandinavian Journal of Statistics*, 31, 93–114.

Kalbfleisch, J. D., Lawless, J. F., and Robinson, J. A. (1991), "Methods for the analysis and prediction of warranty claims," *Technometrics*, 33, 273–285.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The statistical analysis of failure time data*, John Wiley & Sons.

Lancaster, T. (1990), *The econometric analysis of transition data*, Cambridge University Press.

Lawless, J. F. and Nadeau, C. (1995), "Some simple robust methods for the analysis of recurrent events," *Technometrics*, 37, 158–168.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994), "Confidence bands for survival curves under the proportional hazards model," *Biometrika*, 81, 73–81.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000), "Semiparametric regression for the mean and rate functions of recurrent events," *Journal of the Royal Statistical Society, Series B, Methodological*, 62, 711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (2001), "Semiparametric transformation models for point processes," *Journal of the American Statistical Association*, 96, 620–628.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, John Wiley & Sons.

Martinussen, T., Scheike, T. H., and Skovgaard, I. (2002), "Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models," *Scandinavian Journal of Statistics*, 29, 58–74.

McKeague, I. W. and Sasieni, P. D. (1994), "A partly parametric additive risk model," *Biometrika*, 81, 501–514.

Pepe, M. S. and Cai, J. (1993), "Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates," *Journal of the American Statistical Association*, 88, 811–820.

Scheike, T. H. (2004), "Time-Varying Effects in Survival Analysis," in *Handbook of Statistics 23: Advances in Survival Analysis Industry*, eds. Balakrishnan, N. and Rao, C., pp. 61–85.

Therneau, T. M. and Hamilton, S. A. (1997), "rhDNase as an example of recurrent event analysis," *Statistics in Medicine*, 16, 2029–2047.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes: With applications to statistics (ISBN 0387946403)*, Springer-Verlag Inc.

Wang, M. and Hanfelt, J. J. (2003), "Adjusted Profile Estimating Function," *Biometrika*, 90, 845–858.

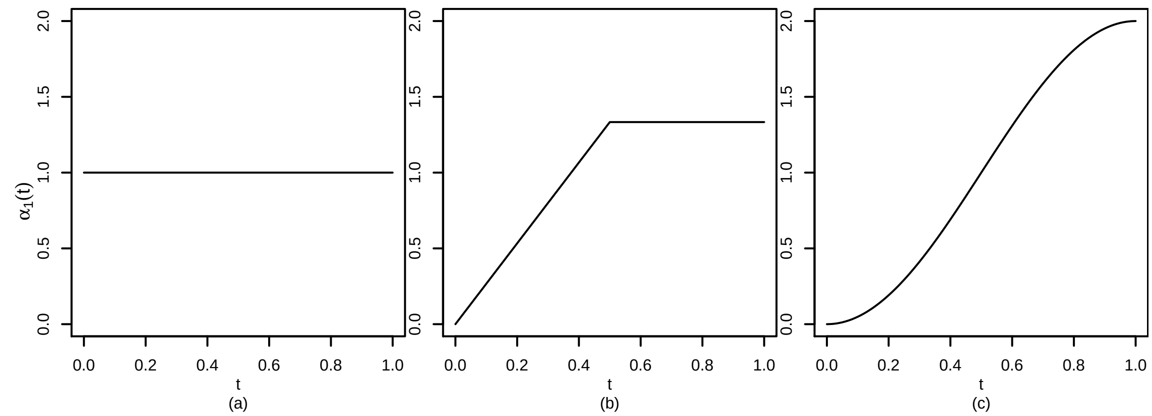

Figure 1: Three choices of true $\alpha_1(t)$.

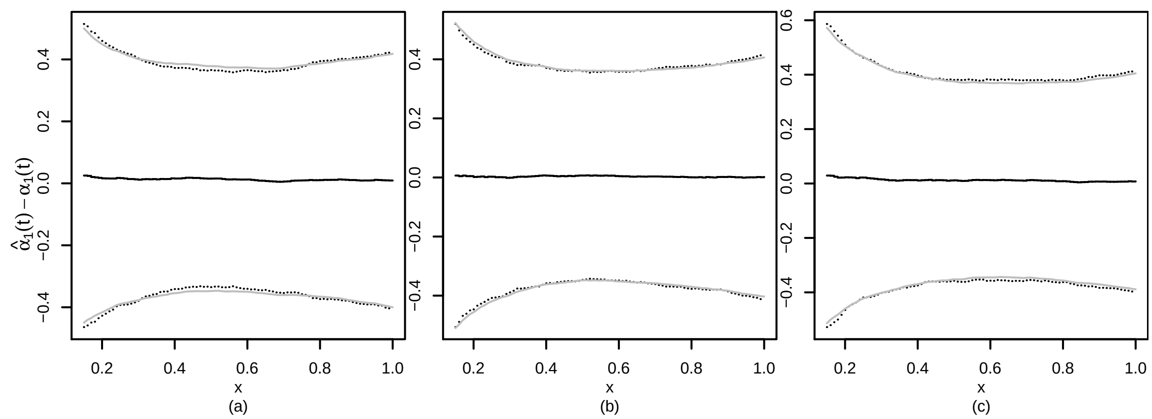

Figure 2: Summary of the simulation result for time-varying coefficient estimator $\hat{\alpha}_1(t)$. The dark curves are the sampling bias and pointwise 95% confidence interval. The gray lines are the pointwise 95% confidence interval constructed using the proposed standard error estimator.

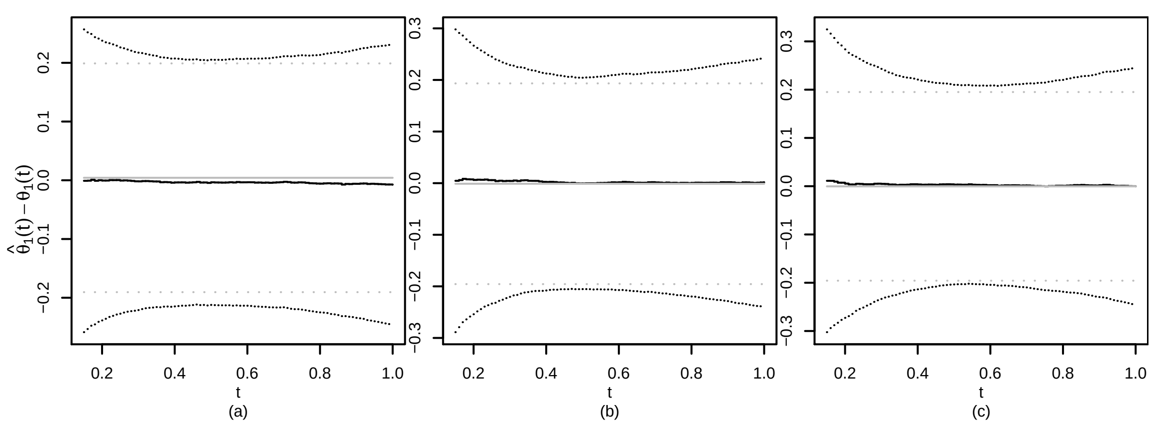

Figure 3: Efficiency comparison of the time-independent coefficient estimator $\hat{\theta}_1$.

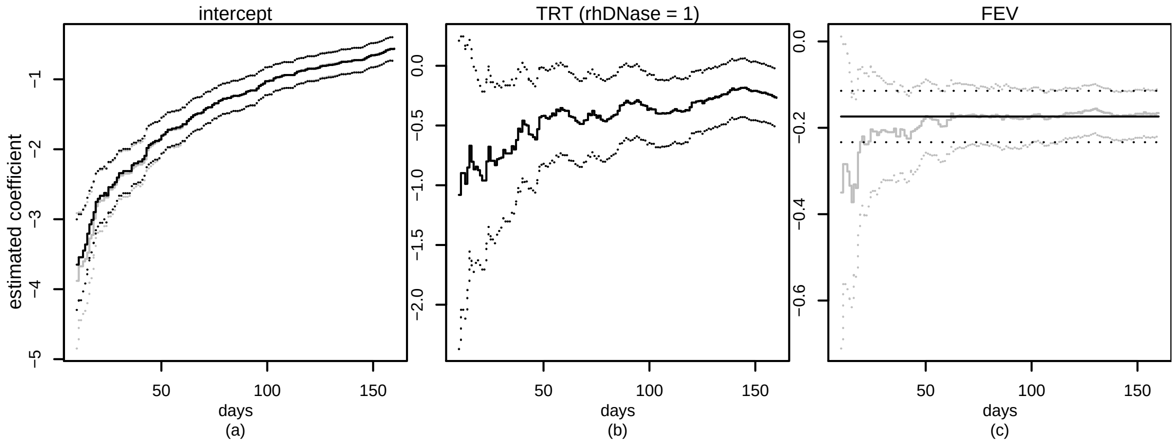

Figure 4: Overlayed coefficient estimates and pointwise 95% confidence intervals: model M0 (gray lines) and model M2 (dark lines).

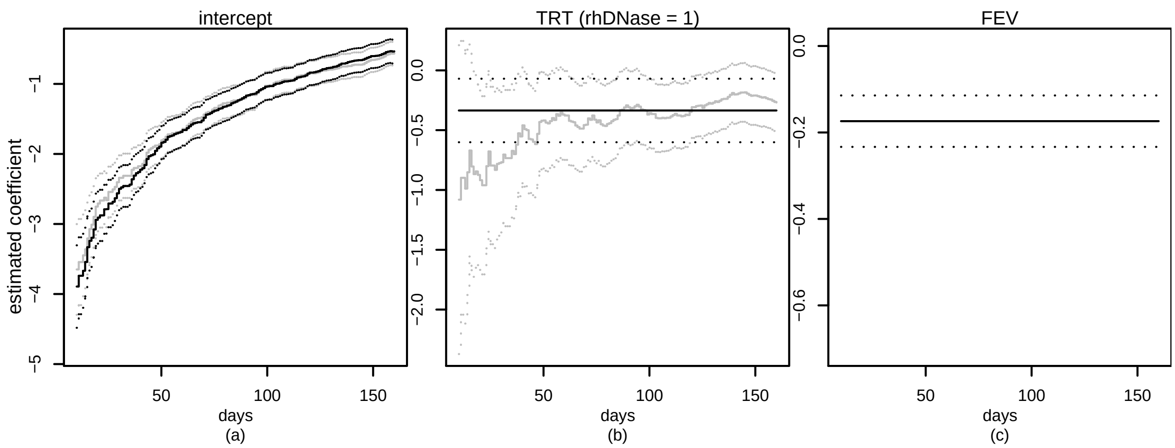

Figure 5: Overlayed coefficient estimates and pointwise 95% confidence intervals: model M2 (gray lines) and model M3 (dark lines).

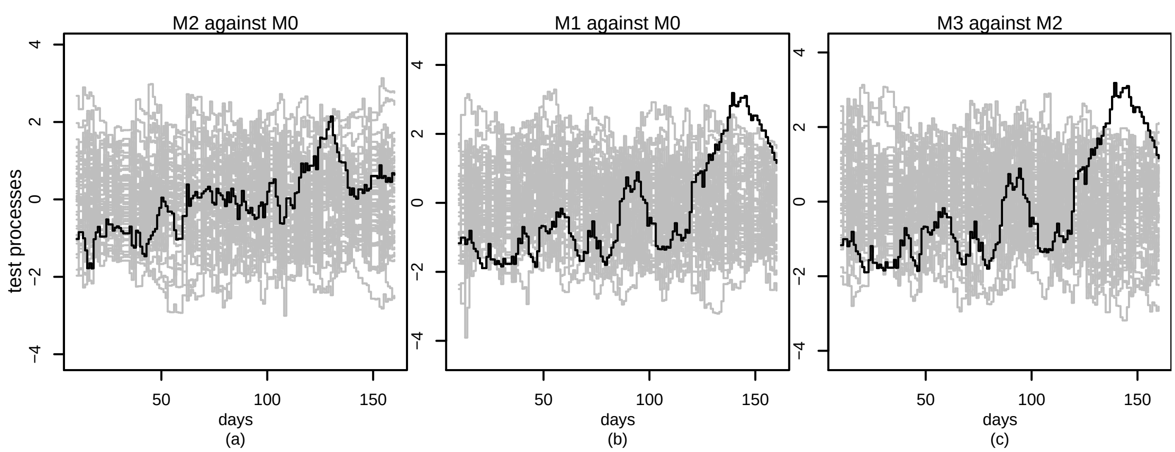

Figure 6: Observed (dark lines) and simulated Gaussian processes (gray lines) used in goodness-of-fit test for constant covariate coefficient. (a) Model M2 against model M0. (b) Model M1 against model M0. (c) Model M3 against model M2.

Table 1: Summary of the simulation result for time-independent coefficient estimator $\hat{\theta}_1$: p: SE: empirical standard error; SEE: mean of standard error estimate; CP: coverage probability.

			$n = 100$				$n = 400$			
$\alpha_1(t)$	p	σ^2	Bias	SE	SEE	CP	Bias	SE	SEE	CP
$f_1(t)$	0.3	0.00	−0.004	0.105	0.102	0.933	0.002	0.052	0.051	0.943
		0.25	0.005	0.131	0.122	0.927	0.001	0.067	0.064	0.942
		0.50	−0.005	0.157	0.140	0.908	0.002	0.078	0.074	0.932
		1.00	0.003	0.191	0.170	0.902	−0.000	0.093	0.090	0.951
	0.6	0.00	0.001	0.123	0.112	0.918	0.003	0.057	0.058	0.943
		0.25	0.000	0.151	0.136	0.923	0.000	0.072	0.071	0.951
		0.50	0.003	0.173	0.154	0.901	−0.003	0.084	0.081	0.933
		1.00	−0.010	0.207	0.181	0.898	−0.005	0.103	0.098	0.921
$f_2(t)$	0.3	0.00	0.004	0.096	0.089	0.936	0.002	0.045	0.045	0.944
		0.25	−0.003	0.121	0.115	0.925	0.001	0.060	0.060	0.951
		0.50	−0.001	0.153	0.136	0.908	−0.001	0.076	0.071	0.918
		1.00	−0.004	0.183	0.164	0.912	−0.001	0.096	0.090	0.923
	0.6	0.00	0.003	0.111	0.100	0.912	0.000	0.054	0.051	0.934
		0.25	−0.002	0.139	0.128	0.920	0.000	0.071	0.067	0.928
		0.50	0.005	0.164	0.148	0.903	0.000	0.082	0.079	0.926
		1.00	−0.004	0.219	0.181	0.884	−0.004	0.104	0.100	0.931
$f_3(t)$	0.3	0.00	0.001	0.076	0.073	0.936	0.001	0.037	0.037	0.952
		0.25	0.005	0.114	0.105	0.913	−0.001	0.055	0.055	0.946
		0.50	−0.003	0.141	0.128	0.916	−0.003	0.071	0.069	0.942
		1.00	0.008	0.187	0.164	0.908	−0.002	0.093	0.089	0.933
	0.6	0.00	0.000	0.090	0.083	0.926	0.002	0.043	0.043	0.940
		0.25	−0.003	0.130	0.118	0.914	−0.003	0.067	0.063	0.929
		0.50	−0.003	0.169	0.145	0.893	−0.002	0.083	0.078	0.931
		1.00	−0.011	0.213	0.180	0.887	−0.003	0.106	0.102	0.928

Table 2: Rejection rate of goodness-of-fit testing for $\alpha_1(t) = \gamma$. For sample size 100 and 200, the results are based on 1000 replicates, each using 1000 bootstraps. For sample size 400, the results are based on 400 replicates, each using 500 bootstraps.

			$n = 100$			$n = 200$			$n = 400$		
$\alpha_1(t)$	p	σ^2	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$f_1(t)$	0.3	0.00	0.075	0.058	0.057	0.058	0.048	0.044	0.052	0.048	0.043
		0.25	0.093	0.064	0.067	0.073	0.061	0.067	0.048	0.037	0.040
		0.50	0.081	0.056	0.062	0.056	0.048	0.045	0.045	0.037	0.035
		1.00	0.107	0.077	0.080	0.074	0.061	0.059	0.060	0.050	0.043
	0.6	0.00	0.084	0.066	0.062	0.062	0.043	0.044	0.062	0.040	0.048
		0.25	0.084	0.049	0.051	0.046	0.047	0.045	0.050	0.060	0.060
		0.50	0.063	0.053	0.052	0.056	0.045	0.047	0.060	0.060	0.060
		1.00	0.091	0.066	0.069	0.073	0.057	0.054	0.058	0.058	0.062
$f_2(t)$	0.3	0.00	0.588	0.483	0.527	0.848	0.796	0.832	0.990	0.990	0.993
		0.25	0.602	0.461	0.495	0.841	0.772	0.805	0.990	0.978	0.988
		0.50	0.614	0.526	0.566	0.826	0.760	0.793	0.990	0.975	0.980
		1.00	0.580	0.456	0.496	0.805	0.727	0.758	0.985	0.968	0.983
	0.6	0.00	0.538	0.455	0.484	0.752	0.694	0.733	0.975	0.965	0.973
		0.25	0.509	0.410	0.437	0.753	0.685	0.729	0.970	0.963	0.970
		0.50	0.503	0.374	0.425	0.756	0.652	0.712	0.948	0.940	0.953
		1.00	0.494	0.332	0.383	0.729	0.601	0.661	0.950	0.907	0.943
$f_3(t)$	0.3	0.00	0.997	0.997	0.997	1.000	1.000	1.000	1.000	1.000	1.000
		0.25	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.50	0.993	0.992	0.993	1.000	1.000	1.000	1.000	1.000	1.000
		1.00	0.996	0.993	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	0.6	0.00	0.990	0.994	0.994	1.000	1.000	1.000	1.000	1.000	1.000
		0.25	0.988	0.986	0.988	1.000	1.000	1.000	1.000	1.000	1.000
		0.50	0.977	0.984	0.984	0.999	0.999	0.999	1.000	1.000	1.000
		1.00	0.964	0.997	0.997	1.000	1.000	1.000	1.000	1.000	1.000