Statistical Science
1997, Vol. 12, No. 1, 20–34

Effects of Unobserved and Partially Observed Covariate Processes on System Failure: A Review of Models and Estimation Strategies

Anatoli I. Yashin and Kenneth G. Manton

Abstract. Stochastically changing covariates may influence survival. They may be observed, unobserved or partly observed. We review the properties of hazard models explicitly representing the effects of unobserved, and partially observed, stochastic covariates. Such models will increase in importance as new longitudinal population studies, and longitudinal surveys of high dimensional failure processes in humans, become available—many are now in progress. It is shown that marginal survival distributions and likelihoods generated in analytically closed form make such parametrically detailed models computationally tractable. Several ways of defining the marginal distribution of the data for constructing a likelihood function are considered. The most complete models can handle both continuously and discretely evolving covariates. Parameters can be estimated from multiple data sets to retrospectively and prospectively evaluate covariate trajectories. Such methods will both extract more information from a longitudinal study and use it in a parametric structure that is logically consistent with the behavior of the underlying processes of substantive interest.

Key words and phrases: Stochastic hazards, random covariates, Cameron–Martin, martingales, conditional Gaussian processes, Wiener processes, Kalman filters, quadratic hazard functions.

1. INTRODUCTION

The goal of survival studies is to evaluate the failure of stochastic dynamic systems. However, standard survival models often ignore both the dynamics of unobserved, or partly observed, stochastic covariates—and information about them available from prior empirical studies or theory. We discuss the statistical and mathematical background necessary to develop a stochastic process model for survival analysis which can be estimated from longitudinal study and survey data, and the potential use of ancillary information about the structure of the process. We present the likelihood for, and illustrate, a specific stochastic process model, referred to as the MWY approach (Woodbury and Manton, 1977; Yashin, 1980, 1985), which has the necessary properties to estimate the risk of failure conditional on the past trajectory of the system's state from longitudinal data—possibly augmented with information (e.g., estimates of specific system parameters) from prior studies of a system's structure (e.g., Singpurwalla, 1995).

In developing such a model, when trajectories are unobserved or partly observed, the conditional risk of failure must be averaged over influential, unobserved variables. The difficulty and burden of computation depends on whether that averaging can be expressed in an analytic form. Several methods which produce analytic averaging functions like Cameron–Martin (Cameron and Martin, 1944) and its generalizations (e.g., Myers, 1981) only evaluate marginal survival distributions. The general conditional hazard form in MWY and models for

Anatoli I. Yashin and Kenneth G. Manton are Professors, Center for Demographic Studies, Duke University, 2117 Campus Drive, Durham, North Carolina 27708 (e-mail: kgm@cds.duke.edu).

more complicated observational plans require the application of martingale techniques. Thus, MWY significantly extends survival analysis procedures used in demography, epidemiology and biostatistics where models of fixed unobserved heterogeneity (e.g., frailty) are used to approximate the influence of latent multidimensional stochastic processes (e.g., Vaupel, Manton and Stallard, 1979). In MWY, influential partly or fully observed process components may also be used; this makes the approach especially useful for analyzing longitudinal studies.

We first examine the structure of likelihood functions which use ancillary information on the conditional (random) hazard and on the probabilistic properties of unobserved or partly observed covariates. We examine assumptions guaranteeing the conditional survival function is exponential in form for stochastic covariates. Cameron–Martin methods are reviewed, as are martingale techniques generalizing Kalman filters to deal with failure processes. These latter techniques lead to the MWY approach for analyzing longitudinal data. Extensions of the MWY approach, using Gaussian semimartingale methods to analyze unobserved covariates with piecewise continuous trajectories, are also discussed.

We assume all random variables and processes are defined for a probability space with stochastic basis $(\Omega, F, \Im, P)$ (e.g., Liptser and Shiryayev, 1988), so the failure time T is an F-stopping time and covariate processes Y_t or Z_t are F-adapted stochastic processes. Since our goal is to examine the applied statistical aspects of models of stochastic processes, we do not present all of the formal mathematical rationale for each model. Those are found in detail in various references.

2. MOTIVATION

Two situations illustrate most of the methodological issues of concern here: analyses of failure times with only ancillary or indirect data on latent, influential covariate processes; and longitudinal studies of survival where covariates are observed at multiple points in time.

2.1 Analysis of Failure Times Generated by a Latent Process

Let T be the failure time, and let $Y = (Y_t)_{t\geq 0}$ be an unobserved process, defining the randomly changing state of a system. Assume the conditional survival function for $Y_0^t = \{Y_u,\ 0 \leq u \leq t\}$ is

$$P(T > t|Y_0^t) = \exp\left[-\int_0^t \mu(Y_u, u)\, du\right],$$

$\mu(Y_u, u)$ is a nonnegative function where $E\,\mu(Y_u, u) < \infty$ for any $u > 0$. The probability distribution of Y and the structure of $\mu(Y_u, u)$ are assumed known—but not the values of parameters. Trajectories of Y are unobserved. T is measured for each system being assessed. The problem is to estimate parameters of the probability distribution of Y and the conditional hazard $\mu(Y_u, u)$ when only the failure times for N systems are observed. The marginal survival function, averaged over Y_u, is

$$P(T > t) = \exp\left[-\int_0^t \bar{\mu}(u)\, du\right].$$

The marginal hazard $\bar{\mu}(u)$ [in a form consistent with the distribution of Y, and the structure of $\mu(Y_u, u)$] is needed to form a likelihood function. Both Cameron–Martin and the martingale versions of MWY can be used to define a form for $\bar{\mu}(u)$, estimate parameters of $\mu(Y_u, u)$ and of Y's distribution, from failure times alone, that is, when the trajectories of Y are not observed. In longitudinal data with covariates measured, however, the use of the Cameron–Martin approach is complicated by boundary conditions on differential equations describing state changes. In contrast, in the martingale version of MWY, the Kalman filter is generalized by adding terms reflecting systematic attrition, transforming the problem of finding the form of $\bar{\mu}(u)$ to one of solving ordinary differential equations from initial conditions.

2.2 Longitudinal Study Data Where Covariates Are Measured

We define T, $Y = (Y_t)_{t\geq 0}$ and $P(T > t|Y_0^t)$ in Section 2.1.

Observations of the state of the system at $t_1, t_2, \ldots$ are described by X_{t_j}, $j > 0$, which are related to an underlying process by

$$X_{t_j} = A(t_j) + A_1(t_j)Y_{t_j} + B(t_j)\varepsilon_j,$$

where A, A_1 and B are fixed functions of time, and $\{\varepsilon_j\}_{j\geq 0}$ are independent normally distributed random variables. The conditional survival function for $X_0^t = \{X_{t_j},\ 0 \leq t_j \leq t\}$ is

$$P(T > t|X_0^t) = \exp\left[-\int_0^t \bar{\mu}(X_0^u, u)du\right].$$

The marginal hazard $\bar{\mu}(X_0^u, u)$ [in a form consistent with the distribution of Y given X_0^u and the structure of $\mu(Y_u, u)$] is needed for the likelihood. We will show how MWY is used to estimate $\bar{\mu}(X_0^u, u)$, the parameters of the conditional hazard $\mu(Y_u, u)$ and the probability distribution of Y from data on X_{t_j} and T for N systems. Before examining the relation between $\bar{\mu}(u)$ and $\mu(Y_u, u)$ in Section 2.1,

and $\bar{\mu}(X_0^u, u)$ and $\mu(Y_u, u)$ in Section 2.2, we will briefly review the properties of conditional survival functions.

2.3 Randomly Changing Hazards and Conditional Survival Function

For T, if $P(T > t)$ is absolutely continuous, then

$$P(T > t) = \exp\left(-\int_0^t \mu(u)\,du\right), \tag{1}$$

where $\mu(u)$, $u \geq 0$, is the hazard function

$$\mu(t) = -P(T > t)^{-1}\,\frac{dP(T > t)}{dt}. \tag{2}$$

Different forms of (2) may be used in martingale representations (e.g., Andersen, Borgan, Gill and Keiding, 1993; Arjas, 1989). If Y is a latent random variable influencing T, the conditional survival function is

$$S(t|Y) = P(T > t|Y) = \exp\left(-\int_0^t \mu(Y, u)\,du\right), \tag{3}$$

where $\mu(Y, u)$ is a *random* hazard. In proportional hazard, frailty models $\mu(Y_u, u)$ is a linear function of Y; for example, $\mu(Y_u, u) = \mu_0(u)Y$. When Y is a stochastic process, as for MWY, and $\mu(Y, u) = \mu(Y_u, u)$, one can use the conditional survival function $S(t|Y_0^t) = P(T > t|Y_u,\ 0 \leq u \leq t)$; that is, the probability of living to age t when Y's trajectory is observed for the interval $[0, t]$. This survival function is written

$$\begin{aligned} S(t|Y_0^t) &= P(T > t|Y_u,\ 0 \leq u \leq t) \\ &= \exp\left(-\int_0^t \mu(Y_u, u)\,du\right). \end{aligned} \tag{4}$$

Sufficient conditions for (4) to be exponential in form are discussed by Yashin and Arjas (1988). For example, one such condition is that $S(t|Y_0^t)$ is almost surely absolutely continuous. Since Y_0^t is unobserved, ancillary information is required to average $S(t|Y_0^t)$ over Y_0^t to define a marginal survival distribution, and then a marginal probability density function to construct a likelihood function.

3. STRATEGIES FOR EVALUATING SURVIVAL FUNCTIONS

In some studies, covariates are measured from the start. In others, covariates are also observed at fixed times during the study. For example, the National Long Term Care Surveys (NLTCS) examine persons age 65+ up to 4 times (1982, 1984, 1989 and 1994) over 12 years. Individuals who enter the sample, say, in 1989, are either reassessed in 1994, drop out of the study before 1994 or die before 1994 (i.e., the next measurement). All three situations occur in the NLTCS. When covariates are not measured during a study (e.g., between 1989 and 1994 in the NLTCS), one may assess the properties of Y and evaluate $\bar{\mu}(u)$, using, under certain assumptions, the Cameron–Martin approach.

3.1 Cameron–Martin

The Cameron–Martin (Cameron and Martin, 1944) approach can be used to define a marginal life-span distribution with unobserved randomly changing covariates for likelihood estimation. Although we show how this is done below, the original Cameron–Martin paper (Cameron and Martin, 1944) did not explicitly adapt the procedure for survival analysis. Let $w = (W_t)_{t\geq 0}$, $W_t \in R^k$, be a k-dimensional Wiener process, and let $Q(t) = \{q_{ij}(t)\}_{ij} = 1, 2, \ldots, k$, be a nonnegative-definite symmetric $k \times k$ matrix where

$$\int_0^t \sum_{i,j=1}^{k} q_{ij}(s)\,ds > 0. \tag{5}$$

PROPOSITION 1. *With W and Q defined above (an asterisk indicates transposition),*

$$\begin{aligned} &E\left[\exp\left(-\int_0^t W_u^* Q(u) W_u\,du\right)\right] \\ &\quad = \exp\left(\tfrac{1}{2}\int_0^t \operatorname{tr}\Gamma(u)\,du\right), \end{aligned} \tag{6}$$

where $\Gamma(u)$ is a symmetric nonpositive-definite matrix which is the unique solution of the matrix Riccati equation,

$$\frac{d\Gamma(u)}{du} = 2Q(u) - \Gamma^2(u), \tag{7}$$

with boundary condition $\Gamma(t) = 0$.

The proof uses Radon–Nikodym derivatives of two Wiener measures (Liptser and Shiryayev, 1977). For one dimension, (7) can be solved

$$E\left\{\exp\left(-q\int_0^t W_u^2\,du\right)\right\} = \left[\cos h\left(\sqrt{2\,qt}\right)\right]^{1/2},$$

where $q > 0$.

3.2 Generalization of Cameron–Martin

The Cameron–Martin approach can be used in survival analyses by interpreting (6) as an average of a conditional survival function with a Wiener process as an unobserved covariate. For a quadratic hazard this is

$$\begin{aligned} &P(T > t|W_u,\ 0 \leq u \leq t) \\ &\quad = \exp\left(-\int_0^t W_u^* Q(u) W_u\,du\right). \end{aligned}$$

However, using Wiener processes as covariates is unrealistic because real covariates start from nonzero initial values and their dynamics reflect not only stochasticity, but also the influence of external factors, deterministic trends and homeostatic forces. Such dynamics can be represented by stochastic differential equations *driven* by Wiener processes. Woodbury and Manton (1977) suggest using stochastic processes as covariates, for example, by a k-dimensional diffusion process $Y = (Y_t)_{t\geq 0}$, satisfying the stochastic Itô differential equation,

$$dY_t = a(t)Y_t\,dt + b(t)\,dW_t, \tag{8}$$

with initial conditions Y_0, Wiener process $W = (W_t)_{t\geq 0}$ and matrices $a(t)$ and $b(t)$ whose elements are continuous functions of time. If $a(t) \equiv b(t) \equiv 0$, the process reduces to a random variable Y_0. The hazard is a time-dependent, quadratic function (Woodbury and Manton, 1977),

$$\mu(Y_u, u) = \mu_0(u) + Y_u^* Q(u) Y_u \tag{9}$$

with conditional survival function, where $Q(u)$ is positive definite

$$P(T > t|Y_0^t) = \exp\left(-\int_0^t \mu_0(u)\,du - \int_0^t Y_u^* Q(u) Y_u\,du\right), \tag{10}$$

$Y_u \equiv 0$ is the covariate value with the lowest failure rate. The point with the lowest failure rate may not be equal to the value of an observed covariate when it is zero. To deal with the problem of identifying location parameters for Y_u, we include sets of linear terms in the quadratic function actually estimated in empirical analyses which we illustrate below.

We will show how (10) can be applied to a survival analysis when measurements X_t are made of the underlying process Y. The use of a quadratic hazard like (10) (with location parameters) can be justified in several ways. Empirically, the relation of risk-to-risk factors (or their transforms) is often J- or U-shaped—especially in elderly populations (e.g., Witteman et al., 1994). Those observations are consistent with theoretical arguments that a complex system such as the human organism has homeostatic mechanisms that keep it from straying to either extreme positive or negative physiological state variable values. This also implies that an optimal point exists which minimizes risk in the interior of the state variable space. That point is identified by location parameters added to Q.

A second biological argument for a U- or J-shaped hazard function, with a minimum point in the interior of the state variable space, is the principle of *hormesis* (Stebbing, 1987), that is, very low exposure to stresses stimulate the organism to increase its "fitness"—the ability to resist environmental stress. This suggests the lowest levels of risk for a biological system may not occur precisely at zero exposure, even for noxious stimuli, but at a positive, *low* level of exposure. When this model is valid the quadratic hazard represents the effects of covariate interactions on risk conditional on the location parameters. In contrast, including quadratic and higher order interactions in a Cox model, with its exponential relation to the hazard, produces location-dependent quadratic coefficients, that is, location parameters interact with the quadratic terms (Pekkanen et al., 1992). Because of their location independence, coefficients in (9) can be more easily compared across study populations. Fundamentally, since any physiological process under homeostatic control operates within specific state space regions, U-shaped hazard functions are needed to describe the state evolution of systems with well-developed internal feedback control mechanisms. For the hazard in (9) and the process in (8) the following holds (Myers, 1981).

PROPOSITION 2. *The survival function given Y_0 is*

$$P(T > t|Y_0) = \exp\left(-\int_0^t \mu_0(u)\,du + Y_0^* U_t(0) Y_0 + \operatorname{tr}\int_0^t b(u)b^*(u)U_t(u)\,du\right), \tag{11}$$

where $U_t(u)$ is a solution of the matrix differential equation

$$\frac{dU_t^t(u)}{du} = Q(u) - a(u)\big(U_t(u) + U_t^*(u)\big) - \frac{1}{2}\big(U_t(u) + U_t^*(u)\big) \cdot b(u)b^*(u)\big(U_t(u) + U_t^*(u)\big), \tag{12}$$

with terminal condition $U_t(t) = 0$.

REMARK. If $a(t) \equiv 0$ and $b^*(t)b(t) \equiv I$ in (11) the result is equivalent to the Cameron–Martin approach. For consistency we need a function $\tilde{\gamma}(u) = U_t(u) + U_t^*(u)$. The differential equation for $U_t^*(u)$ is

$$\frac{dU_t^*(u)}{du} = Q(u) - \big(U_t(u) + U_t^*(u)\big)a^*(u) - \frac{1}{2}\big(U_t(u) + U_t^*(u)\big) \cdot b(u)b^*(u)\big(U_t(u) + U_t^*(u)\big), \tag{13}$$

For $\tilde{\gamma}(u)$ the Riccati equation is

$$\frac{d\tilde{\gamma}(u)}{du} = 2Q(u) - a(u)\tilde{\gamma}(u) - \tilde{\gamma}(u)a^*(u) - \tilde{\gamma}(u)b(u)b^*(u)\tilde{\gamma}(u) \tag{14}$$

with boundary condition $\tilde{\gamma}(t) = 0$. Now Myers's result can be written

$$E\left[\exp\left(-\int_0^t \left[\mu_0(u) + Y_u^* Q(u) Y_u\right] du\right) \middle| Y_0\right] = \exp\left[-\int_0^t \mu_0(u)\, du + \tfrac{1}{2} Y_0^* \tilde{\gamma}(0) Y_0 + \tfrac{1}{2}\mathrm{tr}\int_0^t \tilde{\gamma}(u)b(u)b^*(u)\, du\right]. \tag{15}$$

Equations (14) and (15) generalize (7) and (6) and more realistically describe hazard, and conditional survival, functions.

4. GENERAL RESULTS

A different approach is required for nonquadratic hazards calculating a marginal survival function.

4.1 Likelihoods for General Hazards When Only Survival Data Is Observed

Yashin (1985) gives a general formula for marginal survival functions.

PROPOSITION 3. *If $Y = (Y_u),\ u \geq 0$, is a random process and $\mu(Y, u)$ is a nonnegative functional satisfying measurability conditions such that, for $t \geq 0$,*

$$E\int_0^t \mu(Y, u) du < \infty, \tag{16}$$

then

$$E\left[\exp\left(-\int_0^t \mu(Y, u)\, du\right)\right] = \exp\left(-\int_0^t E[\mu(Y, u)|T > u]\, du\right), \tag{17}$$

where T is related to Y_u by

$$P(T > t|Y_s,\ s \leq t) = \exp\left(-\int_0^t \mu(Y, u)\, du\right). \tag{18}$$

The relation of the observed and conditional hazards is

$$\bar{\mu}(t) = E\big(\mu(Y, t)|T > t\big). \tag{19}$$

The random hazard $\mu(Y, t)$ may depend on either the current value of Y [e.g., $\mu(Y_t, t)$, or the trajectory of Y up to t [e.g., $\mu(Y_0^t, t)$, where $Y_0^t = \{Y_s,\ 0 \leq s \leq t\}$].

REMARK. Equations similar to (17) and (19) are used to analyze populations where risk heterogeneity is "fixed" (e.g., Vaupel and Yashin, 1985). Proposition 3 shows the relation of observed and conditional hazards holds for both fixed *and* stochastically changing "frailty." Formulas (17) and (19) are "natural" parameterizations of a marginal failure distribution if the probabilistic properties of Y are known.

4.2 Survival Functions for Data with Observed and Unobserved Covariates

When some covariates are observed (X_t) and others are not (Y_t), a likelihood can be constructed if the form of the survival function, conditional on X_t, is known. A fundamental result for analyzing longitudinal studies where both X_t and Y_t are influential is the following.

PROPOSITION 4. *Let X_t and Y_t be components of a stochastic process influencing a hazard $\mu(X, Y, t)$, satisfying measurability conditions, and let*

$$E\int_0^t \mu(X, Y, u)\, du < \infty,$$

where T is associated with X_u and Y_u by

$$P(T > t|X_s, Y_s,\ s \leq t) = \exp\left(-\int_0^t \mu(X, Y, u)\, du\right).$$

If the trajectories of X are observed to t, then

$$P(T > t|X_0^t) = \exp\left(-\int_0^t \bar{\mu}(X_0^u, u)\, du\right), \tag{20}$$

where

$$\bar{\mu}\big(X_0^t, t\big) = E\big(\mu(X, Y, t)|X_0^t,\ T > t\big). \tag{21}$$

To prove that (20) provides the correct formula for averaged with respect to $\mu(Y_u, u)$, define

$$\bar{\mu}\big(X_0^t, t\big) = \frac{f\big(t|X_0^t\big)}{P\big(T > t|X_0^t\big)} = \frac{E\big[f\big(t|X_0^t, Y_0^t\big)|X_0^t\big]}{P\big(T > t|X_0^t\big)},$$

where $f(t|X_0^t)$ and $f(t|X_0^t, Y_0^t)$ are probability density functions of T given X_0^t and X_0^t, Y_0^t ($X_0^t = \{X_s,\ 0 \leq s \leq t\}, \{Y_0^t,\ 0 \leq s \leq t\}$). Representing $f(t|X_0^t, Y_0^t)$ in terms of $\mu(X, Y, t)$,

$$\begin{aligned}\bar{\mu}(X_0^t, t) &= \frac{E\big(\mu(X, Y, t)\, P\big(T > t|X_0^t, Y_0^t\big)\big|X_0^t\big)}{P\big(T > t|X_0^t\big)} \\ &= \frac{E\big(\mu(X, Y, t)\, I(T > t)|X_0^t\big)}{P\big(T > t|X_0^t\big)} \\ &= E\big(\mu(X, Y, t)|X_0^t,\ T > t\big),\end{aligned}$$

Q.E.D. Equation (17) is a special case of (20) when there are no X_t.

5. THE MARTINGALE FORM OF THE MWY APPROACH

The Cameron–Martin approach is limited by the terminal boundary conditions in equations (7) and (12). Woodbury and Manton's (1977) diffusion process model with randomly changing covariates and a quadratic hazard used *initial* conditions and martingale techniques. Yashin (1985) proved the Gaussian property of $P(Y \leq y|T > t)$ and the following proposition.

PROPOSITION 5. *Let the k-dimensional process* Y_t *satisfy*

$$dY_t = \left[a_0(t) + a_1(t)Y_t\right]dt + b(t)\,dW_t, \tag{22}$$

where Y_0 *is a vector of Gaussian random variables with a k-element vector of means* m_0 *and a* $k \times k$ *variance–covariance matrix* γ_0. *The matrix* $Q(u)$ *is a symmetric nonnegative-definite matrix whose elements satisfy* (5). *Then*

$$E\left[\exp\left\{-\int_0^t [Y_u^* Q(u) Y_u]\,du\right\}\right] = \exp\left\{-\int_0^t \left(m_u^* Q(u) m_u + \operatorname{tr}\left[Q(u)\gamma_u\right]\right)du\right\}, \tag{23}$$

where m_u *and* γ_u *are solutions of ordinary nonlinear differential equations*:

$$\frac{dm_t}{dt} = a_0(t) + a_1(t)m_t - 2\gamma_t Q(t)m_t, \tag{24}$$

$$\frac{d\gamma_t}{dt} = a_1(t)\gamma_t + \gamma_t a_1^*(t) + b(t)b^*(t) - 2\gamma_t Q(t)\gamma_t, \tag{25}$$

with initial conditions m_0 *and* γ_0, *and where* m_t *is a k-element vector and* γ_t *is a* $k \times k$ *matrix.*

Equation (22) includes random, deterministically changing and fixed covariates as special cases. Equations (24) and (25) can be solved by numerical methods for ordinary differential equations if $a_0(t)$, $a_1(t)$, $b(t)$ and $Q(t)$ are known (e.g., Runge–Kutta). Parameter estimation is discussed in Section 7. These equations resemble Kalman-filter extrapolation equations (Liptser and Shiryayev, 1977) except for terms representing selection, that is, a probability of loss from the population that is a function of state variable values (Yashin, 1985). Note (23)–(25) can be used to calculate marginal survival functions if T and Y_0 are known (Section 3.2).

6. RETROSPECTIVE PROJECTIONS OF COVARIATE TRAJECTORIES

The evaluation of past covariate values is important when events did not occur at the expected rate or when covariates were unobserved. For example, asbestos exposure and mesothelioma risks measured for shipyard workers and heating and insulation workers (Selikoff, 1981) can be used to evaluate exposure for occupations where it was not measured (e.g., construction workers). This problem arises for agents with effects evident long after exposure and for left censoring, for example, when a supplementary sample is drawn to represent the population reaching a criterion age in an inter-survey period.

Unobserved covariate trajectories can be retrospectively assessed to see if they affected events if the hazard function, its parameter values and probabilistic properties of covariates are known from ancillary information. Retrospective analyses require using "smoothing" (conditional expectations for different order moments) equations. Liptser and Shiryayev (1977) discuss smoothing equations for continuous diffusion processes; Khametov and Yashin (1983) used them for multivariate point processes with observed trajectories.

"Filter" equations (24) and (25) are used to calculate smoothing equations, that is, conditional expectations (Yashin and Manton, 1994),

$$\begin{aligned} m(s,t) &= E(Y_s|T > t), \\ \Gamma_{12}(t,s) &= E\big((Y_t - m_t) \cdot \left[Y_s - m(s,t)\right]^*|T > t\big), \\ \Gamma_{22}(s,t) &= E\big(\left[Y_s - m(s,t)\right] \cdot \left[Y_s - m(s,t)\right]^*|T > t\big), \\ \Gamma_{21}(t)(s,t) &= E\big(\left[Y_s - m(s,t)\right] \cdot (Y_t - m_t)^*|T > t\big). \end{aligned} \tag{26}$$

PROPOSITION 6. *For* T *and* $Y = (Y_t)_{t\geq 0}$ *defined by* (5) *and* (22), *the forward smoothing equations* (*i.e., fixed s, increasing t*) *satisfy*

$$m(s,t) = m_s - 2\int_s^t \Gamma_{21}(s,t)Q(u)m_s\,du, \tag{27}$$

$$\Gamma_{12}(t,s) = \gamma_s + \int_s^t a(u)\Gamma_{12}(u,s)\,du - 2\int_s^t \gamma_u Q(u)\Gamma_{12}(s,u)\,du, \tag{28}$$

$$\Gamma_{21}(s,t) = \gamma_s + \int_s^t \Gamma_{21}(s,u)a^*(u)\,du - 2\int_s^t \Gamma_{21}(s,u)Q(u)\gamma_u\,du, \tag{29}$$

$$\Gamma_{22}(s,t) = \gamma_s - 2\int_0^t \Gamma_{12}(s,u)Q(u)\Gamma_{21}(u,s)\,du, \tag{30}$$

where m_u *and* γ_u *satisfy* (24) *and* (25).

PROPOSITION 7. *For* T *and* $Y = (Y_t)_{t\geq 0}$, *the backward smoothing equations* (t *fixed*, *decreasing* s) *satisfy the following*:

$$\frac{d}{ds}m(s,t) = a_0(s) + a(s)m(s,t) + b(s)b^*(s)\gamma_s^{-1}\big(m(s,t) - m(s)\big), \quad m(t,t) = m_t; \tag{31}$$

$$\frac{d}{ds}\Gamma_{21}(s,t) = a(s)\Gamma_{21}(s,t) + b(s)b^*(s)\gamma_s^{-1}\Gamma_{21}(s,t), \quad \Gamma_{21}(t,t) = \gamma_t; \tag{32}$$

$$\frac{d}{ds}\Gamma_{12}(t,s) = \Gamma_{12}(t,s)a^*(s) + \Gamma_{12}(t,s)\gamma_s^{-1}b(s)b^*(s), \quad \Gamma_{12}(t,t) = \gamma_t; \tag{33}$$

$$\frac{d}{ds}\Gamma_{22}(s,t) = a(s)\Gamma_{22}(s,t) + \Gamma_{22}(s,t)a^*(s) - b(s)b^*(s) + b(s)b^*(s)\gamma_s^{-1}\Gamma_{22}(s,t) + \Gamma_{22}(s,t)\gamma_s^{-1}b(s)b^*(s), \quad \Gamma_{22}(t,t) = \gamma_t. \tag{34}$$

"Backward" equations estimate past covariate values generated by a stochastic process. "Forward" equations update parameters as new data becomes available.

7. MWY APPLICATIONS TO PARTIALLY OBSERVED COVARIATES

There are several ways a process in a longitudinal study may be imperfectly or incompletely observed. Measurements may be right censored by loss to follow-up, end of study or mortality. In left censoring, there are differences between censored and uncensored cases not described by covariates measured at study entry. To adjust for such information loss, a model relating the event of censoring to the available data is needed. Wu and Carroll (1988) modeled right censoring as a function of the initial value and slope of a latent covariate in a linear random effects model. For longitudinal data, Diggle and Kenward (1994) combined a multivariate linear model of observed covariate changes with a logistic regression describing the dependence of censoring on Y_t. The quadratic hazard model not only deals with censoring but also with information lost as covariates change between measurements. To represent covariates whose changes are partly observed, we define $Z(t)$, an n-dimensional process measured a finite number (say K) of times.

7.1 Continuously Changing Stochastic Covariates Observed at Discrete Times

Suppose mortality is a quadratic function of $Z(t)$, that is,

$$\mu\big(Z(t),t\big) = \mu_0(t) + Z^*(t)Q(t)Z(t), \tag{35}$$

where $Q(t)$ satisfies (5) and $Z(t)$ satisfies

$$dZ(t) = \big(a_0(t) + a_1(t)Z(t)\big)\,dt + b(t)\,dW_t, \tag{36}$$

where $a_0(t)$ is an n-dimensional vector function of t with bounded elements for any $t \geq 0$, $b(t)$ is a bounded $n \times r$ matrix and W_t is an r-dimensional Wiener process independent of $Z(0)$. For $Z(t)$ measured at $t_1, t_2, \ldots, t_k$, Yashin, Manton and Stallard (1986a, b) examined the conditional survival function

$$S\big(t|\hat{z}(t)\big) = P\big(T > t|\,\hat{z}(t)\big), \tag{37}$$

where

$$\hat{z}(t) = \big(z(t_1), z(t_2), \ldots, z[t_p(t)]\big), \quad t_p(t) = \sup\big\{t_j\colon t_j < t\big\}, \tag{38}$$

with $z(t_i)$ the value of $Z(t)$ observed at time t_i. *Between* observations

$$\hat{\mu}\big(\hat{z}(t)t\big) = -\frac{\partial}{\partial t}\,\ln\, S\big(t|\hat{z}(t)\big), \tag{39}$$

where $\hat{\mu}[\hat{z}(t_i),t]$ is the right-continuous mortality rate for $S[t|\hat{z}(t)]$. Yashin, Manton and Stallard (1986a,b) showed that the relation of $\hat{\mu}[\hat{z}(t),t]$ to $\hat{z}(t)$ can be expressed as functions of the conditional means $m(t)$ and covariances $\gamma(t)$, that is,

$$\hat{\mu}\big(\hat{z}(t)t\big) = m^*(t)Q(t)m(t) + \operatorname{tr}\big(Q(t)\gamma(t)\big) + \mu_0(t), \tag{40}$$

where equations for $m(t)$ and $\gamma(t)$ for intervals $t_j \leq t < t_{j+1}$ are

$$\frac{dm(t)}{dt} = a_0(t) + a_1(t)m(t) - 2\gamma(t)Q(t)m(t), \tag{41}$$

and

$$\frac{d\gamma(t)}{dt} = a_1(t)\gamma(t) + \gamma(t)a_1^*(t) + b(t)b^*(t) - 2\gamma(t)Q(t)\gamma(t), \tag{42}$$

with initial conditions $m(t_j) = z(t_j)$, $\gamma(t_j) = 0$; (40) is a special case of (21) where $X_0^t = \hat{z}(t)$.

The hazard $\hat{\mu}(\hat{z}(t), t)$ is used in the likelihood

$$L = \prod_{i=1}^{N} \hat{\mu}\big(\tau_i, \hat{z}_i(\tau_i)\big)^{\delta_i} \exp\Big(- \int_0^{\tau_i} \hat{\mu}\big(u, \hat{z}_i(u)\big)\, du\Big) \cdot \prod_{j=1}^{k_i} f\big(z_i(t_j)|\hat{z}_i(t_{j-1})\big), \tag{43}$$

where $f(z_i(t_j)|\hat{z}_i(t_{j-1})$ is the Gaussian density of $z_i(t_j)$ conditional on prior observations, $\hat{z}(t_{j-1})$ and τ_i, $i = 1, 2, \ldots, N$, death times; $\delta_i = \{0, 1\}$ indicates right censoring.

Equations (40), (41) and (42) permit (43) to be written

$$\begin{aligned} L = \prod_{i=1}^{u} & \hat{\mu}\big[\tau_i, m_i\big(\tau_i, \beta, \hat{z}_i(\tau_i)\big), \gamma(\tau_i, \beta),\ Q(\tau_i, \beta)\big]^{\delta_i} \\ & \cdot \exp\Big\{ - \int_0^{\tau_i} \hat{\mu}\big(u, m_i(u, \beta, \hat{z}_i(u)\big), \\ & \qquad \gamma\big((u, \beta), Q(u, \beta)\big)\, du\Big\} \\ & \cdot \prod_{j=1}^{k_i} (2\pi)^{-n/2} \big|\, \gamma_i(t_{j-}, \beta)\big|^{-\frac{1}{2}} \\ & \cdot \exp\Big\{ - \tfrac{1}{2}\big[z_i(t_j) - m_i(t_{j^-}, \beta)\big] * \\ & \qquad \gamma^{-1}(t_{j^-}, \beta)\big[z_i(t_j) - m_i(t_{j^-}, \beta)\big]\Big\}, \end{aligned} \tag{44}$$

where β is a vector of unknown parameters in (41)–(43), and $m(t_{j^-}, \beta)$, and $\gamma(t_{j^-}, \beta)$ are left-hand limits of $m(t, \beta)$ and $\gamma(t, \beta)$ when $t \uparrow t_j$. In (44), τ_i is known—or adjusted for censoring.

8. EMPIRICAL EXAMPLES: LONGITUDINAL SURVEYS AND POPULATION STUDIES

The likelihood in (44) cannot be optimized using standard algorithms since $m(t, b)$ and $\gamma(t, \beta)$ are not explicit functions of β. However, modified procedures (e.g., based on Newton's method) can be used. Below we use an iterative maximization procedure where the differential equations (41) and (42) are solved at each iteration with parameter values taken equal to the current value of parameter estimates. The calculation of the information matrix is complicated because it involves the partial derivatives of $m(t, \beta)$ and $\beta(t, \beta)$ with respect to components of β. The derivatives are solutions to differential equations produced from (41) and (42) by differentiating both parts with respect to β_i with zero initial conditions at each observation interval. For each individual these equations are solved once for the optimal parameter values determined by (44).

To illustrate the application of the MWY approach to data, where both covariate dynamics and the times to failure, T, are measured, we review analyses of (a) the 34-year Framingham Heart Study follow-up and (b) the 1982, 1984 and 1989 NLTCS. In both, survival, conditional on covariates, is analyzed to ages by which much of the population has died. Although the measured variables $Z(t)$ are information rich, unobserved variables still may have important age-related effects on mortality. The age dependence of the hazard is represented by (35), where time (age) effects factor into an exponential term

$$\mu(Z^+(t), t) = \big(Z^{+*}(t)Q^+Z^+(t)\big)e^{\theta t}.$$

where Q is a constant matrix of hazard coefficient and μ_0 is a constant mortality rate.

Equation (45) can be viewed as generalizing a Gompertz hazard, $\eta\, e^{\theta t}$, where η depends on $Z(t)$. The exponential term $e^{\theta t}$ is assumed to represent the age-related influence of unobserved [i.e., other than $Z(t)$] variables. In this form the proportionality, or scale, factor η is generalized to be a quadratic function of a multivariate stochastic process $Z(t)$ with both deterministic and random components. Since $Z(t)$ represents age-related changes in the observed covariate processes, the value of θ in (45) characterizes how much remains to be learned about the age dependence of unobserved risk factor dynamics on survival, that is, estimates of θ in (45) are generally expected to be smaller than when θ is conditionally estimated on the information in $Z(t)$. In theory, this is because if all of the age-related covariates could be measured, and at frequent intervals, the estimate of θ could go to 0.0. In practice, in most studies of such complex systems as humans, only a limited number of factors associated with aging can be measured. Consequently, the θ will likely retain a significant effect. If θ is significant, then it implies that the hazard coefficients in (45) are age dependent, that is, that $Q \cdot e^{\theta t} = Q(t)$. Also, in interpreting the results of the analysis the coefficients of the stochastic dynamic equation describing the changes in $Z(t)$ can be of substantive interest.

In addition, one might ask whether an alternate specification of the age effect is possible. That is, could the unobserved variable effect be generated

by a failure process like the Weibull. The selection of the exponential form here was based on theoretical arguments that suggest that overall mortality at late ages, especially if adjusted for stochastic risk differentials, can be explained by a Gompertz-type process (e.g., Strehler, 1977; Strehler and Mildvan, 1960; Sacher and Trucco, 1962). Although others have argued for a Weibull-type failure model (e.g., Rosenberg et al., 1973) the Gompertz model has received more empirical support, for example, it s ability to describe adult mortality across human populations and over animal species (e.g., Finch 1990; Finch and Pike, 1996). Thus, this is one area where theoretical arguments and ancillary information about latent processes are important.

8.1 The Framingham Heart Study

We first apply the MWY model to data where observations are made at fixed intervals of the same length. If intervals are short, and the population not elderly, logistic or Cox regression can be applied assuming events in each interval are generated conditionally independently of events in other intervals (e.g., Wu and Ware, 1979), that is, when covariate dynamics between measurements do not strongly affect outcomes. Those procedures are not applicable if time- or age-dependent unobserved covariates influence the time to failure.

In the 34-year Framingham follow-up, 10 risk factors [i.e., pulse pressure (in millimeters of mercury), the difference of systolic and diastolic blood pressure (BP), diastolic BP (in millimeters of mercury), body mass index (weight in kilograms divided by height in meters squared), cholesterol (milligrams per deciliter), blood glucose (mg%), hematocrit (%), vital capacity index (VCI, centiliters divided by height in meters squared), smoking (cigarettes per day), left ventricular hypertrophy, ventricular rate] were measured biennially—in addition to age, sex and date of death. In the analysis of dynamics, a linear autoregressive process was estimated where each of the 10 risk factors at time $t+1$ was assumed to be a function of their values at t, a constant and a term representing age at t. These 10 equations were estimated from 16 pairs of measurements of the risk factors made over 34 years. Some risk factors (e.g., body mass index) were quite stable (i.e., dominated by the autoregressive effect); others (e.g., ventricular rate) had high degrees of variability. Certain covariates show strong age trends (e.g., vital capacity index declined strongly with age) while others showed strong gender differences. Other factors were strongly influenced by other risk factors over time (e.g., blood glucose at $t+1$ was positively related to body mass index at t).

Risk factors in epidemiological studies often have U- or J-shaped relations to mortality, for example, both very low and high blood pressures are positively related to the risk of death (e.g., Witteman et al., 1994; Manton, Stallard, Woodbury and Dowd, 1994). Thus, it seemed appropriate to use the quadratic hazard in (45) to approximate the conditional hazard function.

In estimation, it is convenient to augment the vector of risk factors $Z(t)$ with a 1.0. Thus the $(k+1)$-element vector $Z^+(t)$ can then be used to define the hazard matrix,

$$Q^+ = \begin{bmatrix} \mu_0 & \frac{c}{2} \\ \frac{c^*}{2} & Q \end{bmatrix},$$

where μ_0 is the constant in (45), Q is the matrix of quadratic coefficients in (45) and c represents linear coefficients adjusting the location of $Z(t)$ to reflect the location of the point of minimum risk in the risk factor space. Thus, the matrix Q^+ has dimension $(k+1)\times(k+1)$, where μ_0 represents the null hypothesis of constant mortality, c is a k-element vector of linear terms and Q is $k \times k$. With this generalization of Q the quadratic hazard can be expressed

$$\mu\big(Z^+(t), t\big) = \big(Z^{+*}(t) Q^+ Z^+(t)\big) e^{\theta t}.$$

In the example, the likelihood in (44) is used. Parameters estimated include those of a linear process describing temporal and age changes in the risk factors, diffusion, hazard coefficients and θ—the age-related effect of unobserved variables. The vector and matrix coefficients $a_0(t)$, $a_1(t)$ and $b(t)$ are assumed constant, that is, a_0, a_1, b; Q^+ is assumed positive definite and symmetric with constant entries. Further details of estimation and interpretation of results are discussed in Manton and Stallard (1988).

The model's fit to mortality outcomes [conditional on the realized outcome of $Z(t)$], the size of θ, its standard error and the reduction of the size of θ due to the introduction of observed risk factors into the hazard function are presented in Table 1.

In Table 1 we present log likelihood ratio approximations to χ^2. The null hypothesis against which χ^2 is initially assessed is that mortality is constant (i.e., mortality can be described by μ_0). First, we tested the improvement in χ^2 when a Gompertz function is estimated, that is, one additional parameter, θ, is entered in the mortality equation. The Gompertz improved the fit of male mortality by 1350.4 χ^2 points with one degree of freedom. This is 62.1% of the maximum χ^2 achievable (line 4 in

TABLE 1
Chi-squared values associated with three different models of total mortality functions for the 34-year follow-up of the Framingham Heart Study

Model	Male χ^2	χ^2/χ^2_4	Females χ^2	χ^2/χ^2_4
1. Standard Gompertz	1350.4 ($\theta = 9.4\% \pm 0.18\%$)	62.1%	1406.1 ($\theta = 10.0\% \pm 0.19\%$)	68.7%
2. Effects of covariate ($\hat{z}_t$) ($\theta = 0.0$)	1518.8	69.8%	1445.5	70.6%
3. Effect of q, net of $\hat{z}_t$	656.0 ($\theta = 8.1\% \pm 0.22\%$)	30.2%	602.3 ($\theta = 8.1\% \pm 0.23\%$)	29.4%
4. Full process (representing effects of both θ and $\hat{z}_t$)	2174.8	100.0%	2047.8%	100.0%

Table 1) and is highly significant. For females the χ^2 for the same model was 1406.1 for one degree of freedom, or 68.7% of the χ^2 for the complete model. The percent increase in mortality per year of life (θ) is 9.4% for males and 10.0% for females; both estimates have a high degree of precision. Thus θ has a highly significant effect for both genders.

If the effects represented by θ can be described by the $Z(t)$, then covariate interactions over time, which covariates best predict risk, which change rapidly or how optimal covariate interventions might be designed, can be explicitly examined. Consequently, we conducted a second test where we estimate the coefficients in the quadratic function for the 10 risk factors with θ = 0.0, that is, an assumption that there is no age effect independent of $Z(t)$. The change over the null hypothesis is 1518.8 (69.8% of the total) χ^2 points for 65 degrees of freedom (i.e., the upper triangle of the 11 $\times$ 11, Q^+ matrix; there are 10 risk factors and a term representing the constant) for males; and 1445.5 χ^2 points for females for 65 degrees of freedom (70.6% of the total). Thus the covariates, and their dynamics, significantly predict survival by themselves. Note that in this example a_0, a_1 and b describing the age dynamics of Z were fixed at the levels estimated by the maximum likelihood procedures.

Third, we determined if there is a significant effect of θ on mortality beyond that explained by the quadratic function of the 10 risk factors and their dynamics. This could be viewed as a test of whether the hazard function can be described by Q^+—or whether the hazard coefficients are age dependent [i.e., $Q^+(t) = e^{\theta t}$.]

Line 3 in Table 1 shows θ improved the fit over using only the 10 risk factors in the quadratic hazard function by 656.0 χ^2 points with one degree of freedom—or 30.2% of the total. For females the effect attributable to θ, net of the 10 risk factors, is 602.3 χ^2 points with one degree of freedom—or 29.4% of the total. Ideally (i.e., all observed covariates are informative about the aging process) all age variation in mortality would be accounted for by the 10 risk factors so that θ's net effect would be negligible. However, with χ^2 changes of 656.0 and 602.3 (one degree of freedom), the effects of unobserved processes associated with age are potent even after controlling for the dynamics of the observed covariate processes. Thus, the hazard matrix is age dependent [i.e., $Q^+(t)$]. Estimation procedures which do not represent the effect of those age-related unobserved variables will produce coefficient estimates in Q^+ which can be biased in complex ways.

The 10 risk factors did reduce the effect of the unobserved variables represented by θ from 9.4% to 8.1% for males; and from 10.0% to 8.1% for females. Thus, the mortality doubling time, conditional on the 10 risk factors, increased from 7.7 to 8.9 years (13.5%) for males and from 7.3 to 8.9 years (18.0%) for females.

In evaluating the effects of specific observed covariates on mortality, one can start by examining specific sets of coefficients in $Q^+(t)$. In $Q^+(t)$, the diagonals represent the quadratic effects of each observed variable. The product of the constant times each covariate represents the linear effects of the covariates. The product of each risk factor with each other represents the pairwise interactions of the covariates. In this example, the dominant effects in the $Q^+(t)$ are the 10 quadratic terms. In addition, certain interaction terms (e.g., ventricular rate $\times$ left ventricular hypertrophy) had large effects on mortality. Some of the interaction effects were negative, suggesting that the effect of one risk factor on mortality varied over the level of a second risk factor (blood glucose in mg% and hematocrit). There were also significant gender differences in the effects of specific coefficients; for example, the quadratic effect of hematocrit was larger for males than females. The age increase in the effects of some risk factors was large enough that mortality selection could cause the population mean and variance of some risk factors (e.g., cholesterol) to decline at advanced ages (Manton, Woodbury, and Tolley, 1994).

8.2 Application to the 1982, 1984 and 1989 NLTCS

The NLTCS is a series of large longitudinal surveys done in 1982, 1984 and 1989 (overall N = 30,308) designed to assess temporal changes in chronic disability in the U.S. elderly population. Chronic disability was assessed using the same set of questions in 1982, 1984 and 1989. The sample in each survey was drawn from lists of Medicare enrolled persons. Details of the sampling procedure are presented in Manton, Corder and Stallard (1993).

The NLTCS data poses some analytic complications not found in the Framingham data. First, the measures of disability are discrete responses to questions about the ability to perform 27 activities (e.g., eating, dressing, bathing, moving around inside). To translate the 27 discrete activity measures into a smaller set of continuous dimensions of function/dysfunction a multivariate procedure designed for discrete variables was used (Woodbury, Manton and Tolley, 1996). This procedure was used to reduce the dimensionality of the measurements, rather than principal components, because it was explicitly designed for discrete measures. After the multivariate analysis, the effects of the 27 measurements of activity on mortality could be represented by continuously scaled scores on the 7 disability dimensions defined by the procedure which were constrained to a convex space in estimation (Manton et al., 1994). In addition, the dynamic equations described the changes in those scores over time.

One of the seven disability scores represents the level of complete functioning. The other six scores represent the degree and type of disability. If the loss of function is related to the risk of death, then a nonzero score on any of the six disability dimensions indicates an increased risk of death. Thus the scores influence a multidimensional convex hazard function which we approximated by a time-dependent quadratic hazard function, $Q^+(t)$, except that a separate constant variable is not introduced because of the convexity constraints imposed by the multivariate analysis (Manton, Woodbury and Tolley, 1994). The quadratic hazard for the survival analysis therefore has a well-defined origin (i.e., point of minimum mortality) corresponding precisely to the value of 1.0 on the first nondisabled dimension or score.

To adjust for the truncation of the variation of scores by mortality and the convexity constraints when using the parameter estimates to calculate life table functions, a special matrix rescaling procedure was used (Manton, Stallard and Singer, 1994). The rescaling procedure is a dynamic, multidimensional generalization of normalizing functions applied to the error distribution for a discrete dependent variable (Woodbury, Manton and Tolley, 1996).

The second score (the first disability dimension) represents the degree to which a person has moderate physical but no cognitive impairments. The score on the second disability dimension represents mild cognitive but no physical impairment. The score on the third disability dimension indicates loss of the physical ability to live independently (e.g., the ability to do laundry, grocery shopping etc. without help). The score on the fourth disability dimension indicates loss of the ability to maintain basic physical functions (e.g., the ability to bath, eat, dress without help). The score on the fifth disability dimension represents the degree of physical frailty (e.g., limitations on the ability to perform basic physical movements; use of hands, legs). The final score indicates disability so severe and complete that institutional care is needed (Manton et al., 1994).

A second difference from the Framingham example was that modeling the rate at which the processes change, and were measured, is more complex in the NLTCS. Since the intervals between NLTCS's are not equal (i.e., 1982 to 1984 is two years; 1984 to 1989 is five years) the time dependence of disability score changes must be parameterized to combine data on covariate dynamics from different length intervals. This is difficult to do in logistic or Cox regression because of the location dependence of the hazard function (Pekkanen et al., 1992). The NLTCS population is elderly (i.e., all persons are 65+) and mortality risks, and the number of dimensions and degree of disability, tend to increase with age. The Cox model does not have parameters describing covariate's age and time-dependent rates of change, subject to random perturbation (Law, Wald and Thompson, 1994; Law et al., 1994), within the long (relative to the rate of occurrence of health events at late ages) and unequal intersurvey periods. Although the age rate of change of many traits is known from physiological (e.g., age dependence of cardiac function; Kasch et al., 1993), clinical and epidemiological studies, the trajectory for an individual is unique because of genetic, environmental, and other influences. Longitudinal studies show that the dynamics of health variables are a mix of deterministic trends and stochastic influences (e.g., Law, Wald and Thompson, 1994; Law et al., 1994; Manton et al., 1994). Therefore, modeling individual covariate trajectories and their influences on survival is important in longitudinal studies when those trajectories are only partly observed.

Thus, a stochastic process model of mortality with observed and unobserved dynamic covariates was used to analyze the NLTCS data. We assume that the seven scores representing an individual's level of impairment on six sets of activities satisfy stochastic differential equations like (36) where coefficients $a_0(t)$, $a_1(t)$ and $b(t)$ are constant parameters, with the conditional hazard given by (45), where θ, elements of the matrix Q^+ and a_0, a_1 and b are estimated using a likelihood like (44).

Estimation, however, is not done with traditional optimization procedures. Optimization had to be constrained: to maximize (44) with unequal intervals the ordinary nonlinear differential equations (41) and (42) are solved for each iteration. If (41) and (42) have analytic solutions, then $\bar{\mu}(t)$ is an explicit function of unknown parameters, and maximization is straightforward. In the most general situation, methods such as Runge–Kutta are needed to solve the differential equations for m and γ between measurements. For individuals entering the study, say, in 1989, the values of μ_0 are taken equal to the values of covariates measured for these individuals in 1989. Values of γ_0 are taken equal to zero. For those who become 65 between 1989 and 1994, μ_0 and γ_0 are assumed equal to the mean and covariance matrix of covariates values calculated for 65-year-old individuals measured in 1989. This allows us to include, in the likelihood, information on, for example, persons who became 65 between 1989 and 1994 and died before 1994.

The effect of age and the covariates on mortality is shown in Table 2.

The same set of hypotheses about model structure was examined for the NLTCS as was examined for the Framingham data (Table 1). The first involved adding the Gompertz parameter θ. The parameter θ is compared against the null hypothesis that mortality is constant. The θ parameter increased the likelihood χ^2 by 2622.2 points with one degree of freedom for males; χ^2 increased 4966.0 points (line 1 in Table 2) with one degree of freedom for females. The θ-estimates were 8.2% for males, 9.1% for females.

With θ set to 0.0 and the disability covariates entered in a quadratic hazard function, a χ^2 of 3821.5 points with 27 degrees of freedom (a Q^+ matrix with the scores as seven covariates) was produced for males; a χ^2 of 6550.5 points with 27 degrees of freedom was produced for females. The first and second models are not nested, but the inclusion of the disability scores produce a much higher χ^2 than using only θ. Note that when testing this hypothesis we kept parameters a_0, a_1 and b describing the dynamics of Z^+ at the levels estimated by maximum likelihood procedures.

We examined the effect of θ, net of the disability scores, by examining differences between the model with risk factors with $\theta = 0.0$ and the model with both covariates and θ. The net χ^2 increment due to the addition of the single parameter θ is large (i.e., 1016.4 χ^2 points with one degree of freedom for males; 943.3 χ^2 points with one degree of freedom for females), although proportionately smaller than in Framingham. The declines in θ due to the introduction of the disability scores were larger, that is, from 8.2 to 5.3% for males and from 9.1 to 4.8% for females. Thus, the seven disability scores in the NLTCS data are more informative about age-related mortality changes than the 10 risk factors in the Framingham data even though the length of time between risk factor measurements was larger.

In the matrix of hazard coefficients there were large differences in the mortality risk of specific disability dimensions. The nondisabled dimension always had the lowest mortality risk. The highly frail group had the highest mortality. Some of the disabled groups with intermediate levels of mortality represented relatively young groups with high mortality associated with acute cardiopulmonary problems. These groups tended either to recover without lasting disability or to die rapidly. The highest mortality was found for the fifth disability dimension, a frail, but not institutionalized,

TABLE 2
Chi-squared values associated with three different models of total mortality for 1982, 1984 and 1989 NLTCS

Model	Male χ^2	χ^2 / χ^2_4	Females χ^2	χ^2 / χ^2_4
1. Standard Gompertz	2622.2	54.2%	4966.0	59.6%
	($\theta = 8.2\% \pm 0.11\%$)		($\theta = 9.1\% \pm 0.09\%$)	
2. Effects of covariate ($\hat{z}_t$) ($\theta = 0.0$)	3821.5	79.0%	6550.5	87.4%
3. Effect of θ, net of $\hat{z}_t$	1016.4	21.0%	943.3	12.6%
	($\theta = 5.3\% \pm 0.12\%$)		($\theta = 4.8\% \pm 0.11\%$)	
4. Full process (representing effects of both θ and $\hat{z}_t$)	4837.9	100.0%	7493.8%	100.0%

group. The ratio of the size of the coefficients for this group to those for the nondisabled group was about 8 to 1 for both males and females, although females had baseline mortality about half that of males. Also of interest was the fact that females tended to tolerate disability better than males, that is, the mortality coefficients for scores representing greater disability were smaller for females than males.

8.3 General Properties of the Age Dependent Quadratic Hazard

The dynamic properties of risk in the quadratic hazard (44) are illustrated in Figure 1.

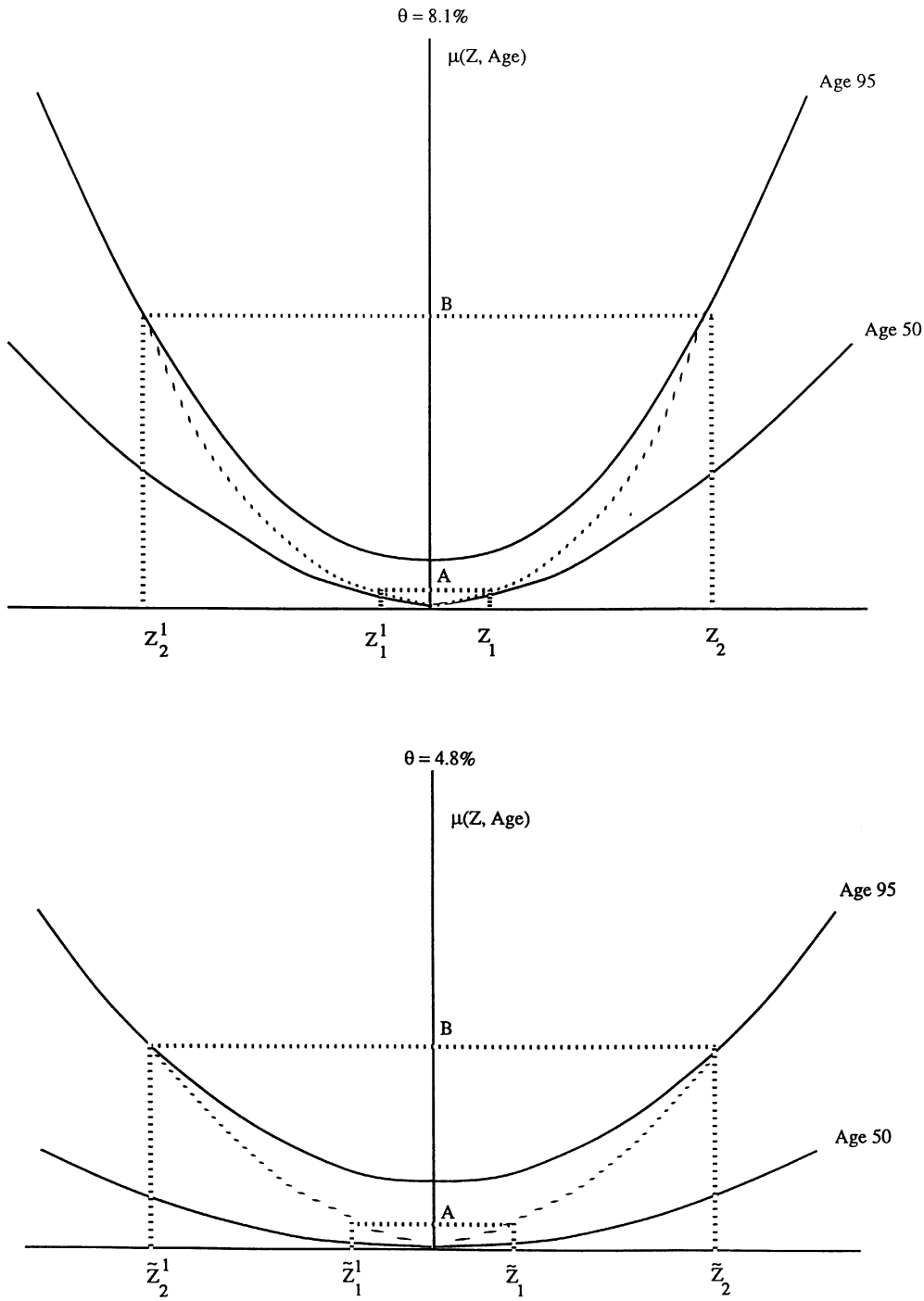

FIG. 1. *Description of age changes in quadratic hazard functions when covariate information is informative* ($\theta = 8.1\%$) *and highly informative* ($\theta = 4.8\%$).

In Figure 1 we show the hypothetical mortality for the quadratic hazard with a one-dimensional covariate process $Z(t)$ for ages 50 and 95. The parameter θ is 8.1% in the top figure (from the Framingham Study) and 4.8% in the bottom figure (as estimated in the NLTCS). The hazard becomes more sensitive to changes in covariate values with age. The "real" age trajectory of the hazard is represented by the dashed line in Figure 1: when covariate Z changes from Z_1 at age 50 to Z_2 at age 95, mortality changes from level A to level B. These changes in mortality are higher than for the quadratic hazard without dynamics represented. Not only do increases in the covariate matter but the deviation of covariates from the optimal point with age increases risk more than for the conditional Gompertz—or a quadratic hazard with constant parameters.

The disability scores reduced age-related uncertainty more than risk factors in Framingham, although the NLTCS intersurvey intervals are longer. This may be because disability has temporally more proximate effects on mortality than risk factors. Disability also predicts risk factor changes very well. For example, declines in physical activity predict changes in blood pressure, cholesterol and vascular tone—as well as end points like stroke (Colantonio, Kasl and Ostfeld, 1992) and cardiovascular disease. Even so, the effects of age-related unobserved variables remain significant in the NLTCS so that methods not adjusting for their effects on mortality produce biased coefficient estimates for observed covariates.

The MWY approach can be generalized to the case where the trajectories of the covariate processes may have discrete changes (jumps) in value. One way to generalize the model to account for such covariate jumps is described in Yashin (1993). He proposed modeling the covariate process as a Gaussian martingale with piecewise continuous trajectories. An alternate strategy is to extend filtration methods for piecewise continuous stochastic processes to deal with observations of jumps in covariate values (Yashin, 1980). Both approaches can be used to extend the application of Kalman filters to stochastic dynamic systems with catastrophic failures.

9. CONCLUSION

In longitudinal studies, the effects of observed and unobserved covariates's evolution over age cannot be ignored. Logistic or Cox regressions do not explicitly represent these influences. A stochastic process model based on a parametric specification of the conditional hazard, and randomly changing covariates, is needed. When the conditional hazard is a quadratic function of covariates, both the Cameron–Martin procedure (Myers, 1981; Yashin, 1985, 1993) and the martingale version of the MWY (Woodbury and Manton, 1977; Yashin, 1980, 1985) procedure can be used to calculate marginal survival functions.

The martingale version of MWY is a preferable procedure for several reasons. First, parameter estimates can be recursively updated when new measurements are made. Second, covariate dynamics are better represented (e.g., Woodbury and Manton, 1977; Yashin, 1985; Yashin, Manton and Stallard, 1986a, b) because all data on the covariate's evolution in a longitudinal study is used. Third, knowledge from prior studies can be used to define probabilistic regularities of the stochastic evolution of unobserved covariates, for example, by specifying the conditional survival function from other longitudinal studies where hazards are estimated as functions of measured covariates—important because the marginal survival function is calculated by averaging the conditional one. Influential variables can satisfy stochastic differential equations representing random environmental influences and describe forces preserving the equilibrium (homeostasis) of multidimensional, physiological processes. Thus, knowledge of the structure of the conditional hazard can be applied when covariates are unobserved—or partly observed. Thus (a) a probabilistic description of covariate dynamics with possibly free parameters and (b) specification of the conditional survival function given covariates's trajectories allow the MWY approach to be used in many longitudinal studies.

This was illustrated by estimating parameters for two longitudinal studies with different data structures (i.e., NLTCS and Framingham). The model performed well in each and demonstrated the presence of information in the dynamics of observed and unobserved covariates that standard survival analyses do not utilize. Thus, modeling and estimation strategies for failure processes with time-varying covariates better describe the state dynamics of multivariate systems, and their failure, and can improve the understanding of complex system failure, interventions in failure processes and forecasts of distributions of times to system failure.

ACKNOWLEDGMENTS

Professor Manton's efforts in this research were supported by National Institute on Aging Grants AG01159 and AG07025. Professor Yashin's efforts were supported by National Institute on Aging Grant AG08767. The authors thank anonymous reviewers for their valuable comments.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Continuing Processes*. Springer, New York.

ARJAS, E. (1989). Survival models and martingale dynamics. *Scand. J. Statist.* **16** 177–225.

CAMERON, R. and MARTIN, W. (1944). The Wiener measure of Hilbert neighborhoods in the space of real continuous functions. *J. Math. Phys.* **23** 195–209.

COLANTONIO, A., KASL, S. W. and OSTFELD, A. M. (1992). Level of function predicts first stroke in the elderly. *Stroke* **23** 1355–1357.

DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis. *J. Roy. Statist. Soc. Ser. C* **41** 49–93.

FINCH, C. E. (1990). *Longevity, Senescence, and the Genome.* Univ. Chicago Press.

FINCH, C. and PIKE, M. (1996). Maximum life span predictions from the Gompertz mortality model. *Journal of Gerontology: Biological Sciences* **51** B183–B194.

KASCH, F. W., BOYER, J. L., VAN CAMP, S. P., VERITY, L. S. and WALLACE, J. P. (1993). Effect of exercise on cardiovascular ageing. *Age and Ageing* **22** 5–10.

KHAMETOV, V. M. and YASHIN, A. I. (1983). Efficient solution to the interpolation problem on the basis of observations of jump processes. *Problems Inform. Transmission* **19** 38–51.

LAW, M. R., WALD, N. J. and THOMPSON, S. G. (1994). By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *British Medical Journal* **308** 367–373.

LAW, M. R., WALD, N. J., WU, T., HACKSHAW, A. and BAILEY, A. (1994). Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study. *British Medical Journal* **308** 363–366.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1977). *Statistics of Random Processes* **1**. *General Theory.* Springer, New York.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1988). *The Theory of Martingales.* Springer, New York.

MANTON, K. G., CORDER, L. S. and STALLARD, E. (1993). Estimates of change in chronic disability and institutional incidence and prevalence rates in the U.S. elderly population from the 1982, 1984, and 1989 National Long Term Care Survey. *Journal of Gerontology: Social Sciences* **47** S153–S166.

MANTON, K. G. and STALLARD, E. (1988). *Chronic Disease Modeling: Measurement and Evaluation of the Risks of Chronic Disease Processes*. Griffin, London.

MANTON, K. G., STALLARD, E. and SINGER, B. H. (1994). Methods for projecting the future size and health status of the U.S. elderly population. In *Studies of the Economics of Aging* (D. Wise, ed.) 41–77. National Bureau of Economic Research, Univ. Chicago Press.

MANTON, K. G., STALLARD, E., WOODBURY, M. A. and DOWD, J. (1994). Time-varying covariates in models of human mortality and aging: multidimensional generalization of the Gompertz. *Journal of Gerontology: Biological Sciences* **49** B169–B190.

MANTON, K. G., WOODBURY, M. A. and TOLLEY, H. (1994). *Statistical Applications Using Fuzzy Sets.* Wiley, New York.

MYERS, L. E. (1981). Survival functions induced by stochastic covariate processes. *J. Appl. Probab.* **18** 523–529.

PEKKANEN, J., MANTON, K., STALLARD, E., NISSINEN, A. and KARVONEN, M. (1992). Risk factor dynamics, mortality and life expectancy differences between eastern and western Finland: the Finnish cohorts of the seven countries studies. *International Journal Epidemiology* **21** 406–419.

ROSENBERG, A., KEMENY, G., SMITH, L. G., SKURNICK, I. D. and BANDURSKI, M. J. (1973). The kinetics and thermodynamics of death in multicellular organisms. *Mechanisms of Aging and Development* **2** 275–293.

SACHER, G. A. and TRUCCO, E. (1962). The stochastic theory of mortality. *Ann. New York Acad. Sci.* **96** 985.

SELIKOFF, I. J. (1981). Disability compensation for asbestos-associated disease in the United States. Report to the U.S. Dept. Labor Environmental Sciences Laboratory, Mount Sinai School of Medicine, New York. (Contract No. J-9-M-8-0165.)

SINGPURWALLA, N. D. (1995). Survival in dynamic environments. *Statist. Sci.* **10** 86–103.

STEBBING, A. D. (1987). Growth hormesis: a by-product of control. *Health Physics* **52** 543–547.

STREHLER, B. L. (1977). *Time, Cells and Aging.* Academic Press, New York.

STREHLER, B. L. and MILDVAN, A. S. (1960). General theory of mortality and aging. *Science* **132** 14–21.

VAUPEL, J., MANTON, K. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.

VAUPEL, J. and YASHIN, A. I. (1985). Heterogeneity ruses: some surprising effects of selection on population dynamics. *American Statistician* **39** 176–185.

WITTEMAN, J. M., GROBBEE, D. E., VALKENBURG, H. A., VAN HEMERT, A. M., STIJNEN, T., BURGER, H. and HOFMAN, A. (1994). *J*-shaped relation between change in diastolic blood pressure and progression of aortic atherosclerosis. *The Lancet* **343** 504–507.

WOODBURY, M. A. and MANTON, K. G. (1977). A random-walk model of human mortality and aging. *Theoret. Population Biol.* **11** 37–48.

WOODBURY, M. A., MANTON, K. G. and TOLLEY, H. D. (1996). Convex models of high dimensional discrete data. *Ann. Inst. Statist. Math.* To appear.

WU, M. C. and BAILEY, K. R. (1988). Analyzing changes in the presence of right censoring caused by death and withdrawal. *Statistics in Medicine* **7** 337–346.

WU, M. C. and CARROLL, R. (1988). Estimation and comparison of changes in the presence of right censoring by modeling the censoring process. *Biometrics* **44** 175–188.

WU, M. and WARE, J. (1979). On the use of repeated measurements of regression analysis with dichotomous responses. *Biometrics* **35** 513–521.

YASHIN, A. I. (1980). Conditional Gaussian estimation of dynamic systems under jumping observations. *Automat. Remote Control* **5** 618–626. (Translated from Russian.)

YASHIN, A. I. (1985). Dynamics in survival analysis: conditional Gaussian property versus Cameron–Martin formula. In *Statistics and Control of Stochastic Processes* (N. V. Krylov, R. Sh. Liptser and A. A. Novikov, eds.) 466–475. Springer, New York.

YASHIN, A. I. (1993). An extension of the Cameron–Martin result. *J. Appl. Probab.* **30** 247–251.

YASHIN, A. I. and ARJAS, E. (1988). A note on random intensities and conditional survival functions. *J. Appl. Probab.* **25** 630–635.

YASHIN, A. I. and MANTON, K. G. (1994). Modifications of the EM algorithm for survival influenced by an unobserved stochastic process. *Stochastic Process. Appl.* **54** 257–274.

YASHIN, A. I., MANTON, K. G. and STALLARD, E. (1986a). Dependent competing risks: a stochastic process model. *J. Math. Biol.* **24** 119–140.

YASHIN, A. I., MANTON, K. G. and STALLARD, E. (1986b). Evaluating the effects of observed and unobserved diffusion processes. *Survival Analysis Modelling* **7** 1353–1363.