UW Biostatistics Working Paper Series

6-28-2005

On Additive Regression of Expectancy

Ying Qing Chen

Fred Hutchinson Cancer Research Center, yqchen@u.washington.edu

1 INTRODUCTION

Regression models have proven to be powerful tools in studying the association between the outcome variables and their covariates, and further predicting an expected outcome. The conventional linear regression models usually focus on the summary statistics, such as the expectations. For some of these outcomes, we are interested in their expectancies. The expectancy function of an outcome, Y, is defined as

$$e(y) = E(Y - y \mid Y \geq y),$$

for $y \in R$, which has been an important function in many scientific areas, such as actuarial sciences, reliability and demography, although its use mainly focused on studying the distributions of the time-to-event outcomes. In reality, it can be useful for other time-likewise outcomes as well, for examples, the height-and-weight study of children growth, the cumulative medical cost of treating a chronic disease, and the house price in a highly sought-after area. For such outcomes, the expectancy is meaningful in the sense that it characterizes the additional expectation on an outcome given its present value. Moreover, the expectancy function itself fully determines the distribution function with finite expectation when Y is continuous (Kalbfleisch and Prentice, 2002, p. 7):

$$\bar{F}(y) = 1 - F(y) = \frac{e(0)}{e(y)} \exp\left\{ - \int_0^y \frac{1}{e(u)} du \right\}, \tag{1}$$

where $F(\cdot)$ denotes the cumulative distribution function.

Consider a data example in Weisberg (1985, p.55) to study the fatness of 26 boys and 32 girls. Recommended by the Centers for Disease Control and Prevention (CDC), the Body Mass Index, defined as the ratio of weight (in kilograms) and squared height (in meters), is a useful tool to assess underweight, overweight and at-risk of overweight in children and teens. The BMIs are calculated at age 18 for all the boys and girls. Their histograms are shown in Figure (1), respectively. Both of them are positively skewed. Their sample means are 22.12 ($s.e. = 0.73$) for the boys and 21.82 ($s.e. = 0.50$) for the girls, respectively. According to the CDC cutoffs, two (one) of the boys (girls) are underweight, three (four) are at-risk of overweight and two (one) are overweight. Their empirical BMI expectancy functions (Yang, 1978) are also plotted in Figure (1), respectively. If the CDC cutoffs were applied to the expectancies, given their current BMIs, six (three) of the boys (girls) would be expected to be overweight, while ten (thirteen) of the boys (girls) would be expected to be at-risk of overweight. This means, even though the current BMIs for some boys and girls are still normal or at-risk of overweight, their expected increment may still result in the at-risk or the overweight category. In this example, the BMI expectancy functions can help with the early overweight detection and timely prevention initiation.

[Figure (1) about here]

Also shown in Figure (1), the boys and girls do not appear to have identical BMI expectancy functions, although the two-sample t-test does not show any significant difference ($p = 0.73$). It is thus of interest to design appropriate regression models to quantify and identify the discrepancy in expectancy functions due to important covariates. Most of the traditional regression models, however, usually focus on modeling the expectations of the outcome variables by some link functions to the covariates and parameters. For example, in the generalized linear models (McCullagh and Nelder, 1989, p. 27), the expectation of Y, $\mu = EY$, is linked to the linear predictor of $\eta = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}$ with a link function of $g(\cdot)$:

$$g(\mu) = \eta,$$

where $\boldsymbol{x}$ is the associated $p-$dimensional covariate and $\boldsymbol{\beta}$ is the parameter of the same dimension. With different distributional assumptions and link functions, a wide spectrum of the regression models, including the linear regression model, the logistic regression model and the log-linear model, are embraced. The parameters in these expectation-based models are meaningful and useful in characterizing the covariate effect on the outcome expectations.

However, the parameters of the traditional expectation regression models usually do not lead to the same straightforward interpretation in expectancy functions except for certain special distributions. For example, in the typical linear regression model assumed as $Y = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon$, where ϵ is zero-mean normal with variance σ^2, its expectancy is

$$e(y \mid \boldsymbol{x}) = \left\{1 - \Phi\left(\frac{y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}\right)\right\}^{-1}\left[\int_{(y-x^{\mathrm{T}}\beta)/\sigma}^{\infty} \{1 - \Phi(u)\}\, du\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard $N(0,1)$. Assume $\boldsymbol{x}$ takes value of 0 or 1 in a two-sample setting. Although $\boldsymbol{\beta} = E(y \mid \boldsymbol{x} = 1) - E(y \mid \boldsymbol{x} = 0)$ can be interpreted as the average change in outcome due to the sample difference, it is unclear from this form what direct interpretation of $\boldsymbol{\beta}$ would entail on the expectancies. In Figure (2), the expectancy functions are also plotted when $\boldsymbol{\beta}$ and σ are assumed to be 1. Neither is it clear from the plot if the expectancy functions would hold any simple relationship, either additive or multiplicative.

[Figure (2) about here]

In the rest of this article, we thus focus on the methodology development of regression models for the expectancy functions. First, we propose an additive expectancy regression model and develop its parametric and semiparametric estimation procedures for the regression parameters. Then we study related issues, such as estimation efficiency, model adequacy assessment and alternative modeling strategies. Simulation studies are conducted to evaluate the validity and efficiency of the proposed estimators. The weight-and-height example is further used to demonstrate the methodology and theory to be developed.

2 EXPECTANCY REGRESSION ANALYSIS

2.1 Additive Expectancy Models

The expectancy function itself has been studied thoroughly in the literature. The article of Guess and Proschan (1988) gives a comprehensive review on the theory of expectancy functions. The regression models for the expectancy functions, however, have been underdeveloped. One of the early works was by Oakes and Dasu (1990) to study relationship between the expectancy function and its associated covariates in the following multiplicative form:

$$e(y \mid \boldsymbol{x}) = e_0(y) \exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}), \tag{2}$$

where $e_0(y)$ is some baseline expectancy function and $\boldsymbol{\beta}$ is the parameter. In this model, the parameter characterizes the multiplicative effect of covariate $\boldsymbol{x}$ on the expectancy functions. This model is appealing that the expectancy functions are explicitly modeled by the interpretable parameter which may be practically meaningful in, for example, assessing treatment efficacy in therapeutic settings. One of the restrictions on this model is, however, the embedded monotonicity on $e(y) + y$. It is not clear if a parameter estimator, $\widehat{\boldsymbol{\beta}}$, would still maintain the monotonicity of $\widehat{e}(y \mid \boldsymbol{x}) + y = \widehat{e}_0(y) \exp(\boldsymbol{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}})$, even though $\widehat{e}_0(y) + y$ is managed to be monotonically increasing.

Instead, we consider an additive expectancy regression model, which assumes that

$$e(y \mid \boldsymbol{x}) = e_0(y) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}. \tag{3}$$

Without loss of generality, we assume that $Y > 0$. As shown in the later development, all the results can be easily extended onto the whole real line. The parameter in model (3) thus characterizes an additive covariate effect on the expectancy functions, when they are assumed of similar shapes as shown in Figure (1) for boys and girls. For example, if $\boldsymbol{\beta} = 2$, it means the BMI expectancy of the girls ($\boldsymbol{x} = 2$) is two points more than that of the boys ($\boldsymbol{x} = 1$). If $\boldsymbol{\beta} = -2$, it means the BMI expectancy of the girls is two points less than that of the boys. If $\boldsymbol{\beta} = 0$, it means there is no gender difference in the BMIs. Thus the parameter $\boldsymbol{\beta}$ quantifies and identifies the discrepancy in the expectancy functions if its estimator is tested to be significantly different from zero. Since the additive expectancy model (3) implies that $E(Y \mid \boldsymbol{x}) = e_0(0) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}$, which shares the same expectation assumption as the usual linear regression model, it may be considered as a generalized type of the linear regression model. Compared with the multiplicative expectancy model (2), the embedded monotonicity is well preserved in the model (3) as long as $e_0(y)$ satisfies such a constraint. A few properties of the additive expectancy model are summarized as follows.

Property 1 Under the additive expectancy model (3),

1. If the baseline expectancy function $e_0(\cdot)$ is properly defined, then the additive expectancy model is also properly defined given $e(y \mid \boldsymbol{x}) \geq 0$;

2. The sign of β determines the relative ordering in both the expectancy functions and the cumulative distribution functions, respectively;

3. $e(y \mid \boldsymbol{x})/e_0(y) = \{\lambda(y \mid \boldsymbol{x})/\lambda_0(y)\}^{-1}$, where $\lambda(\cdot)$ are the hazard functions.

The proofs of these properties are straightforward. Among these properties, the first one offers a characterization of the additive model. The second property suggests that the parameter can be used for assessment of treatment efficacy. The third property implies that the additive model and the usual Cox proportional hazards model are identical if and only if the baseline expectancy function is constant, i.e., exponential. One special class of the models is of the Hall-Wellner linear type (Hall and Wellner, 1984), i.e., $e_0(y) = \alpha_0 + \alpha_1 y$ ($\alpha_0 > 0$, $\alpha_1 > -1$). It is clear that the Hall-Wellner class of distributions satisfies the additive expectancy model. That is, $e(y \mid \boldsymbol{x}) = \alpha_0 + \alpha_1 y + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}$ is also of the Hall-Wellner type, with

$$\bar{F}(y \mid \boldsymbol{x}) = \left(\frac{\alpha_0 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}}{\alpha_0 + \alpha_1 y + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}}\right)^{1+1/\alpha_1}.$$

Another example is when $e_0(y) = \alpha_0 \exp(-\alpha_1 y)$ of Gompertz distribution. Then $e(y \mid \boldsymbol{x}) = \alpha_0 \exp(-\alpha_1 y) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}$ with the cumulative distribution function

$$\bar{F}(y \mid \boldsymbol{x}) = \frac{\alpha_0 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}}{\alpha_0 \exp(-\alpha_1 y) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}} \left\{ \frac{\alpha_0 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} \exp(\alpha_1 y)}{\alpha_0 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}} \right\}^{-1/(\alpha_1 x^{\mathrm{T}}\beta)}.$$

2.2 Parametric and semiparametric inferences

Suppose that the observed data consist of n iid copies of $(y_i, \boldsymbol{x}_i)$, $i = 1, 2, \ldots, n$ with the same distribution as that of the random variable $(Y, \boldsymbol{X})$. When $e_0(\cdot)$ is known or characterized by a finite-dimensional parameter $\boldsymbol{\alpha} \in R^q$ as $e_0(y; \boldsymbol{\alpha})$, the usual maximum likelihood inference procedure can be applied to estimate the parameter $\boldsymbol{\beta}$ by maximizing the loglikelihood function of $l(\boldsymbol{\alpha}, \boldsymbol{\beta})$:

$$\sum_{i=1}^{n} \left[\log\left\{1 + e_0'(y_i; \boldsymbol{\alpha})\right\} + \log\{e_0(0; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - 2\log\{e_0(y_i; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \int_0^y \frac{du}{e_0(u; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} \right]$$

with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Straightforward calculation leads to

$$l_\alpha = \sum_{i=1}^{n} \left[\frac{e_{0,\alpha}'(y_i; \boldsymbol{\alpha})}{1 + e_0'(y_i; \boldsymbol{\alpha})} + \frac{e_{0,\alpha}(0; \boldsymbol{\alpha})}{e_0(0; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} - \frac{2e_{0,\alpha}(y_i; \boldsymbol{\alpha})}{e_0(y_i; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} + \int_0^y \frac{e_{0,\alpha}(u; \boldsymbol{\alpha}) du}{\{e_0(u; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}^2} \right],$$
$$l_\beta = \sum_{i=1}^{n} \boldsymbol{x}_i^{\mathrm{T}} \left[\int_0^y \frac{du}{\{e_0(u; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}^2} - \frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + 2e_0(0; \boldsymbol{\alpha}) - e_0(y_i; \boldsymbol{\alpha})}{\{e_0(0; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\{e_0(y_i; \boldsymbol{\alpha}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}} \right],$$

where the subscripts of α and β represent the partial derivatives with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Then the maximum likelihood estimates of $(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ can be obtained by solving the equations of $l_\alpha = l_\beta = 0$. The solutions are denoted $\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{mle}}$, respectively. Let the true values of the parameters be $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}_*$, respectively. Then by the theory of maximum likelihood methods, $\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{mle}}$ are consistent estimators of $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}_*$, respectively, and $n^{1/2}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}^{\mathrm{T}} - \boldsymbol{\alpha}_*^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}^{\mathrm{T}} - \boldsymbol{\beta}_*^{\mathrm{T}})^{\mathrm{T}}$ is asymptotically zero-mean normal with the variance of $I^{-1}(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*)$, where I is the Fisher information matrix estimated by its observed value

$$\widehat{I}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) = \begin{bmatrix} l_{\alpha\alpha}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) & l_{\alpha\beta}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) \\ l_{\beta\alpha}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) & l_{\beta\beta}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) \end{bmatrix}.$$

As a result, we can use $\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{mle}}$ to make inferences on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and further estimate the baseline frequency functions by $\widehat{e}_0(y; \widehat{\boldsymbol{\alpha}}_{\mathrm{mle}})$.

In practice, it is often desirable of $e_0(\cdot)$ being unspecified. Consider $N_i(y) = I(Y_i \le y)$ and $\Delta_i(y) = I(Y_i \ge y)$, for $y > 0$, respectively. Let $\mathcal{F}_y = \sigma\{N_i(u), Y_i(u), \boldsymbol{X}_i; u \le y, i = 1, 2, \ldots, n\}$, which is the σ-algebra generated by the collection of observations of $(N_i(y), Y_i(y), \boldsymbol{X}_i)$. Straightforward calculation leads to that

$$E\{dN_i(y) \mid \mathcal{F}_{y-}\} = \Delta_i(y)\lambda(y \mid \boldsymbol{x}_i)dy,$$

where $dN_i(y) = N_i(y + dy) - N_i(y)$ and $\lambda(y \mid \boldsymbol{x}_i) = \{1 + e_0'(y)\}/\{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$. Denote the true values of $e_0(y)$ and $\boldsymbol{\beta}$ by $e_*(\cdot)$ and $\boldsymbol{\beta}_*$, respectively. Then the following estimating equations are unbiased for the true values, $y > 0$,

$$U_0\{\boldsymbol{\beta}, e_0(\cdot)\} = \sum_{i=1}^{n} \left[\{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\, dN_i(y) - \Delta_i(y) d\left\{y + e_0(y)\right\}\right]. \tag{4}$$

Then it is natural to estimate $e_0(\cdot)$ by the estimating equations of $U_0\{\boldsymbol{\beta}, \widehat{e}_0(\boldsymbol{\beta})\} = \sum_i [\{\widehat{e}_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}dN_i(y) - \Delta_i(y)d\{y + \widehat{e}_0(y)\}] = 0$, as if $\boldsymbol{\beta}$ were known. Let $A_n(y)$ be the right continuous version of $\exp\{-\int_0^y \sum_i dN_i(u)/\sum_i \Delta_i(u)\}$ and $B_n(y;\boldsymbol{\beta})dy$ be $\sum_i\{\Delta_i(y)dy - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} dN_i(y)\}/\sum_i \Delta_i(y)$, respectively. Straightforward algebra on $U_0\{\beta, \widehat{e}_0(\beta)\} = 0$ thus leads to

$$-\widehat{e}_0(y;\boldsymbol{\beta})d\log A_n(y) - d\widehat{e}_0(y;\boldsymbol{\beta}) = B_n(y;\boldsymbol{\beta})dy,$$

which is a first-order nonhomogeneous ordinary differential equation. As a result, there is a closed-form solution for $e_0(y)$,

$$\widehat{e}_0(y;\boldsymbol{\beta}) = A_n(y)^{-1}\int_y^\infty A_n(u)B_n(u;\boldsymbol{\beta})du.$$

Given such an estimator, we have

Lemma 2 For a fixed constant $\tau \in [0, \infty)$, as $n \to \infty$,

1. $\widehat{e}_0(\cdot;\boldsymbol{\beta}_*)$ is consistent almost surely, i.e., $\|\widehat{e}_0(\boldsymbol{\beta}_*) - e_0\| = \sup_{y\in[0,\tau]} |\widehat{e}_0(y;\boldsymbol{\beta}_*) - e_*(y)|$ converges to 0 almost surely;

2. $n^{1/2}\{\widehat{e}_0(y;\boldsymbol{\beta}_*) - e_*(y)\}$ is asymptotically zero-mean normal with the variance of $[E\{\bar{F}(y \mid \boldsymbol{X})\}]^{-2}\int_y^\infty E[\{1 + m'(y \mid \boldsymbol{X})\}\bar{F}(y \mid \boldsymbol{X})]dy$.

3. $\widehat{e}_{0,\beta}(y;\boldsymbol{\beta}_*) = \partial\widehat{e}_0(y;\boldsymbol{\beta}_*)/\partial\boldsymbol{\beta}$ converges to $-\boldsymbol{\mu}_*(y) = -E\{\boldsymbol{X}\bar{F}(y \mid \boldsymbol{X})\}/E\{\bar{F}(y \mid \boldsymbol{X})\}$, $y \in [0, \tau]$.

When $\boldsymbol{\beta}_* = 0$, it reduces to the one-sample problem with $B_n(u;\boldsymbol{\beta}_*) \equiv 1$. Then $\widehat{e}_0(\cdot)$ becomes an empirical estimator of the baseline expectancy function $e_0(y) = \int_y^\infty \bar{F}_0(u)du/\bar{F}_0(y)$, where $\bar{F}_0(\cdot)$ is estimated by $A_n(\cdot)$.

To estimate the regression parameter $\boldsymbol{\beta}$, the following estimating functions are considered since they are also unbiased for the true parameters:

$$U(\boldsymbol{\beta}; e_0) = \sum_{i=1}^{n} \int_0^{\infty} \boldsymbol{x}_i \left[\{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\, dN_i(y) - \Delta_i(y) d\{y + e_0(y)\}\right].$$

By replacing $e_0(\cdot)$ with $\widehat{e}_0(\cdot;\boldsymbol{\beta})$, we thus obtain the estimating equations of

$$U\{\boldsymbol{\beta}; \widehat{e}_0(\boldsymbol{\beta})\} = \sum_{i=1}^{n} \int_0^{\infty} \boldsymbol{x}_i[\{\widehat{e}_0(y;\boldsymbol{\beta}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} dN_i(y) - \Delta_i(y) d\{y + \widehat{e}_0(y;\boldsymbol{\beta})\}]$$

$$= \sum_{i=1}^{n} \int_0^{\infty} \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}\{\widehat{e}_0(y;\boldsymbol{\beta}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\, dN_i(y) = \sum_{i=1}^{n} \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y_i)\}\{\widehat{e}_0(y_i;\boldsymbol{\beta}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} = 0,$$

where $\bar{\boldsymbol{x}}(y) = \sum_i \boldsymbol{x}_i \Delta_i(y)/\sum_i \Delta_i(y)$. Denote $\widehat{\boldsymbol{\beta}}$ the solution to the equation. Then the following theorem can be used to make inference for $\boldsymbol{\beta}_*$:

Theorem 3 Assume that there exists some constant $\Gamma > 0$ such that $\mathrm{pr}\{\|\boldsymbol{X}\| > \Gamma\} = 0$, and $e_*(\cdot)$ is continuously differentiable on $[0, \tau]$. For any individual term of $\boldsymbol{\mu}_*(\cdot)$, $\mu_{*,i}(\cdot)$, say, there exists $y > 0$ such that $|\mu'_{*,i}(y)| > 0$, $i = 1, 2, \ldots, n$. Let

$$\boldsymbol{\mu}^*(y) = \boldsymbol{\mu}_*(y) - \int_0^y \frac{E[\{\boldsymbol{X} - \boldsymbol{\mu}_*(u)\} d\bar{F}(u \mid \boldsymbol{X})]}{E\{\bar{F}(u \mid \boldsymbol{X})\}}.$$

As $n \to \infty$,

1. $\widehat{\beta}$ converges consistently to $\boldsymbol{\beta}_*$;

2. $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)$ converges weakly to a zero-mean normal variate with the variance-covariance $D^{-1}VD^{-1}$, where $D = E[\int_0^{\infty}\{\boldsymbol{X} - \boldsymbol{\mu}_*(y)\}^{\otimes 2} dF(y \mid \boldsymbol{X})]$ and $V = E[\int_0^{\infty}\{\boldsymbol{X} - \boldsymbol{\mu}^*(y)\}^{\otimes 2} e(y \mid X)^2 dF(y \mid \boldsymbol{X})]$, respectively;

3. D and V can be consistently estimated by $\widehat{D} = n^{-1}\sum_{i=1}^{n}\int_0^{\infty}\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}^{\otimes 2} dN_i(y)$, and $\widehat{V} = n^{-1}\sum_{i=1}^{n}\int_0^{\infty}\{\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}^*(y)\}^{\otimes 2}\widehat{e}(y \mid \boldsymbol{x}_i)^2 dN_i(y)$, respectively., where

$$\widehat{\boldsymbol{\mu}}^*(y) = \bar{\boldsymbol{x}}(y) + \int_0^y \frac{n^{-1}\sum_i\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(u)\} dN_i(u)}{A_n(u)}.$$

Here $\boldsymbol{v}^{\otimes 2}$ defines $\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$. In general, $\boldsymbol{\mu}_*(\cdot)$ and $\boldsymbol{\mu}^*(\cdot)$ are not necessarily equal. When $E\{\boldsymbol{X}\bar{F}(y \mid \boldsymbol{X})\}/E\{\bar{F}(y \mid \boldsymbol{X})\} = \int_0^y E\{\boldsymbol{X} d\bar{F}(u \mid \boldsymbol{X})\}/E\{\bar{F}(u \mid \boldsymbol{X})\}$, however, $\boldsymbol{\mu}_*(y) \equiv \boldsymbol{\mu}^*(y)$ for any y.

2.3 Weighted estimation and semiparametric efficiency

In the maximum likelihood estimation, the estimators of $\widehat{\boldsymbol{\alpha}}_{\text{mle}}$ and $\widehat{\boldsymbol{\beta}}_{\text{mle}}$ are usually fully efficient when the baseline expectancy function is known. It is however unclear how efficient the semiparametric estimators of $\widehat{\boldsymbol{\beta}}$ is given the ad hoc nature of the estimating functions used. One common approach to potentially improve its efficiency is by way of weighted estimation, i.e., calculating the estimators by the weighted estimating equations as:

$$U_w(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} w(y) \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\} \{\widehat{e}_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} \, dN_i(y), \tag{5}$$

where $w(\cdot)$ is some weight function converging to a deterministic function of $w_*(\cdot)$ almost surely. Denote the solution to the above equation as $\widehat{\boldsymbol{\beta}}_w$. Then parallel to Theorem 3, we have this corollary:

Corollary 4 Given the conditions specified in Theorem 3, as $n \to \infty$,

1. $\widehat{\boldsymbol{\beta}}_w$ converges consistently to $\boldsymbol{\beta}_*$;
2. $n^{1/2}(\widehat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_*)$ converges weakly to a zero-mean normal variate with the variance-covariance $D_w^{-1} V_w D_w^{-1}$, where $D_w = E[\int_0^{\infty} w_*(y)\{\boldsymbol{X} - \boldsymbol{\mu}_*(y)\}^{\otimes 2} dN(y \mid \boldsymbol{X})]$ and $V_w = E[\int_0^{\infty} w_*(y)^2\{\boldsymbol{X} - \boldsymbol{\mu}^*(y)\}^{\otimes 2} e(y \mid X)\{1 + e'(y \mid \boldsymbol{X})\} dy]$, respectively;
3. D and V can be consistently estimated by $\widehat{D} = n^{-1} \sum_{i=1}^{n} \int_0^{\infty} w(y)\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}^{\otimes 2} dN_i(y)$, and $\widehat{V} = n^{-1} \sum_{i=1}^{n} \int_0^{\infty} w(y)^2\{\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}^*(y)\}^{\otimes 2} \widehat{e}(y \mid \boldsymbol{x}_i)^2 dN_i(y)$, respectively.

By an application of the Cauchy-Schwarz inequality, the optimal weight function for the weighted estimating equations in (5) should be thus proportional to $e(y \mid \boldsymbol{X}_1)^{-2}$. incidentally, these optimal weighted estimating functions possess the similar weight coefficients as those in the maximum likelihood score function of l_β. This fact may imply that the weighted estimating functions for the semiparametric estimation would substantially improve the efficiency of its estimators.

Additional efficiency consideration is by way of the semiparametric efficiency bound calculation assuming that the baseline expectancy functions are unknown. Consider the parametric submodels in the form of $e(y \mid \boldsymbol{x}) = e_0(y) + \gamma e_1(y) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}$. Here $e_0(\cdot)$ and $e_1(\cdot)$ are both known fixed functions, and $(\boldsymbol{\beta}^{\mathrm{T}},\gamma)^{\mathrm{T}}$ are unknown parameters. Then its associated loglikelihood function of $(\boldsymbol{\beta}^{\mathrm{T}},\gamma)^{\mathrm{T}}$ is

$$l(\boldsymbol{\beta},\gamma)=\sum_{i=1}^{n}\left[\int_0^\infty \log\left\{\frac{e_0'(y)+\gamma e_1'(y)}{e_0(y)+\gamma e_1(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}\right\}dN_i(y)-\Delta_i(y)\left\{\frac{1+e_0'(y)+\gamma e_1'(y)}{e_0(y)+\gamma e_1(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}\right\}dy\right],$$

and

$$\frac{\partial l(\boldsymbol{\beta}_*,0)}{\partial\boldsymbol{\beta}}=-\sum_{i=1}^{n}\int_0^\infty \frac{\boldsymbol{x}_i}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\left[dN_i(y)-\frac{\Delta_i(y)\{1+e_0'(y)\}dy}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\right]$$

$$\frac{\partial l(\boldsymbol{\beta}_*,0)}{\partial\gamma}=\sum_{i=1}^{n}\int_0^\infty \left\{\frac{e_1'(y)}{1+e_0'(y)}-\frac{e_1(y)}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\right\}\left[dN_i(y)-\frac{\Delta_i(y)\{1+e_0'(y)\}dy}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\right].$$

Consider the Fisher information at $\boldsymbol{\beta}_*$ and $\gamma=0$, which is denoted by the matrix

$$I(e_1)=\begin{bmatrix} I_{\beta\beta}(e_1) & I_{\beta\gamma}(e_1)\\ I_{\gamma\beta}(e_1) & I_{\gamma\gamma}(e_1)\end{bmatrix},$$

with $I_{\beta\beta}=E(\partial^2 l/\partial\boldsymbol{\beta}^2)$, $I_{\beta\gamma}=E(\partial^2 l/\partial\boldsymbol{\beta}\partial\gamma)$ and $I_{\beta\beta}=E(\partial^2 l/\partial\gamma^2)$, respectively. Then by an application of the Cauchy-Schwarz inequality, the variance-covariance matrix of any regular semi-parametric estimator $\widetilde{\boldsymbol{\beta}}$ in the linear model, if $n^{1/2}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_*)$ converges to a zero-mean normal, would be larger than $(I_{\beta\beta}-I_{\beta\gamma}I_{\gamma\gamma}^{-1}I_{\beta\gamma}^{\mathrm{T}})^{-1}$ for any e_1. Here matrix M_1 is 'larger' than matrix M_2 if M_1-M_2 is nonnegative definite. Since

$$I_{\beta\beta}(e_1)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_0^\infty E\left[\frac{\Delta_i(y)\{1+e_0'(y)\}\boldsymbol{x}_i^{\otimes 2}}{\{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}^3}\right]dy,$$

$$I_{\beta\gamma}(e_1)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_0^\infty E\left[\frac{\Delta_i(y)\{1+e_0'(y)\}\boldsymbol{x}_i}{\{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}^2}\left\{\frac{e_1'(y)}{1+e_0'(y)}-\frac{e_1(y)}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\right\}^{\mathrm{T}}\right]dy,\text{ and}$$

$$I_{\gamma\gamma}(e_1)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_0^\infty E\left[\frac{\Delta_i(y)\{1+e_0'(y)\}}{\{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}}\left\{\frac{e_1'(y)}{1+e_0'(y)}-\frac{e_1(y)}{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*}\right\}^{\otimes 2}\right]dy,$$

$(I_{\beta\beta}-I_{\beta\gamma}I_{\gamma\gamma}^{-1}I_{\beta\gamma}^{\mathrm{T}})^{-1}$ thus reaches its maximum at the $e_1(y)$ such that

$$e_1'(y)E\left\{\frac{\Delta(y)}{1+e_0'(y)}\right\}-e_1(y)E\left\{\frac{\Delta(y)}{e_0(y)+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_*}\right\}=E\left\{\frac{\Delta(y)\boldsymbol{X}}{e_0(y)+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_*}\right\},$$

and yields a closed-form solution in $e_1(\cdot)=P(y)^{-1}\int_y^\infty P(u)Q(u)du$, where

$$P(y)=\exp\left[-\int_0^y E\left\{\frac{\Delta(u)}{e_0(u)+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_*}\right\}\Big/E\left\{\frac{\Delta(u)}{1+e_0'(u)}\right\}du\right]$$

and $Q(y)=E\left[\Delta(y)\boldsymbol{X}/\{e_0(u)+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_*\}\right]/E[\Delta(y)/\{1+e_0'(y)\}]$, respectively. Therefore, the semi-parametric information bound for $\boldsymbol{\beta}$ at $\boldsymbol{\beta}_*$ is the supremum parametric information bound at $\boldsymbol{\beta}_*$ given any choice of $e_0(\cdot)$, which is

$$\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_0^\infty E\left[\frac{\Delta_i(y)\{\boldsymbol{x}_i-\bar{\boldsymbol{x}}_0(y)\}^{\otimes 2}}{\{e_0(y)+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}^2}\right]dy.$$

Here

$$\bar{\boldsymbol{x}}_0(y) = \lim_{n\to} \widetilde{\boldsymbol{x}}(y) = \lim_{n\to\infty} \frac{\sum_i \Delta_i(y)\boldsymbol{x}_i/\{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}}{\sum_i \Delta_i(y)/\{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}}.$$

Therefore, when $e_0(\cdot)$ is known, the optimal estimating function for $\boldsymbol{\beta}$ in the model (3) is

$$U_{\mathrm{opt}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\infty \frac{\boldsymbol{x}_i - \widetilde{\boldsymbol{x}}(y)}{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} \left[dN_i(y) - \frac{\Delta_i(y)\{1 + e_0'(y)\}dy}{e_0(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}} \right].$$

If comparing $U_{\mathrm{opt}}(\cdot)$ with the weighted estimating equations $U_w(\cdot)$, it is straightforward to find that they are similar except in how to estimate the expectation of $\boldsymbol{X}$. In addition, either the use of the optimal weighted estimating equations or the use of the semiparametric efficient estimating equations is feasible in practice given the closed-form of $\widehat{e}(\cdot)$.

2.4 Model-based expectancy prediction

For a specific covariate $\boldsymbol{x}_0$, the prediction of its associated expectancy function is also of practical interest. A straightforward prediction can be based on the model (3) with its maximum likelihood estimates, which is $\widehat{e}(y \mid \boldsymbol{x}_0) = \widehat{e}_0(y; \widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}) + \boldsymbol{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{mle}}$. Its p-th percentile pointwise confidence interval can be further constructed with an application of the delta-method for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as:

$$\widehat{e}(y \mid \boldsymbol{x}_0; \widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) \mp Z_{(p+1)/2} \cdot \widehat{\mathrm{se}}\{\widehat{e}(y \mid \boldsymbol{x}_0; \widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}})\},$$

where the estimated standard error $\widehat{\mathrm{se}}$ is computed as the square-root of

$$\left\{ \left(\frac{\partial e}{\partial \boldsymbol{\alpha}}\right)^{\mathrm{T}}, \left(\frac{\partial e}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} \right\} \widehat{I}^{-1}(\widehat{\boldsymbol{\alpha}}_{\mathrm{mle}}, \widehat{\boldsymbol{\beta}}_{\mathrm{mle}}) \left\{ \left(\frac{\partial e}{\partial \boldsymbol{\alpha}}\right)^{\mathrm{T}}, \left(\frac{\partial e}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} \right\}^{\mathrm{T}},$$

and $Z_{(p+1)/2}$ is the $(p+1)/2$-th normal percentile.

For the regression parameter estimates obtained in the semiparametric estimation, the model-based prediction of expectancy for a given covariate $\boldsymbol{x}_0$ is naturally $\widehat{e}(y \mid \boldsymbol{x}_0) = \widehat{e}_0(y; \widehat{\boldsymbol{\beta}}) + \boldsymbol{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$. Here, $\widehat{e}_0(y; \widehat{\boldsymbol{\beta}}) = A_n(y)^{-1} \int_y^\infty A_n(u) B_n(u; \widehat{\boldsymbol{\beta}}) du$. Given the consistency of $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_*$, straightforward algebra shows that $\widehat{e}(y \mid \boldsymbol{x}_0; \widehat{\boldsymbol{\beta}})$ is also consistent of $e(y \mid \boldsymbol{x}_0)$. It is further shown that $n^{1/2}\{\widehat{e}(y \mid \boldsymbol{x}_0; \widehat{\boldsymbol{\beta}}) - e(y \mid \boldsymbol{x}_0)\} = n^{1/2}\{\widehat{e}_0(y; \widehat{\boldsymbol{\beta}}) - e_*(y) + \boldsymbol{x}_0^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)\}$ is asymptotically equivalent to the following process:

$$\mathcal{E}(y) = n^{1/2} \{\boldsymbol{x}_0 - \boldsymbol{\mu}_*(y)\}^{\mathrm{T}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*) + n^{1/2} \{\widehat{e}_0(y, \boldsymbol{\beta}_*) - e_*(y)\},$$

where $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)$ and $\widehat{e}_0(y, \boldsymbol{\beta}_*) - e_*(y)$ are further asymptotically equivalent to $D^{-1}n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}$, and $n^{-1/2}\sum_i \int_y^\infty \{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} dM_i(u) / \sum_i \Delta_i(u)$, respectively. Here, $dM_i(y) = dN_i(y) - \Delta_i(y)\lambda(y \mid$

$\boldsymbol{x}_i)dy$, as shown in the Appendix. This process in fact converges weakly to a Gaussian process with mean zero and the covariance function $\sigma(y_1, y_2; \boldsymbol{x}_0)$ that can be estimated by its empirical counterparts. Thus the pth-percentile pointwise confidence interval for $e(y \mid \boldsymbol{x}_0)$ can be constructed as

$$\widehat{e}(y \mid \boldsymbol{x}_0) \mp Z_{(p+1)/2} \cdot \sqrt{\widehat{\sigma}(y, y; \boldsymbol{x}_0)/n}.$$

Simultaneous confidence bands can be also constructed for the expectancy as a function. Given the complex form of $\sigma(y_1, y_2)$, however, it is usually less straightforward to obtain a closed form for the confidence bands. A more straightforward approach is to apply the bootstrap method. An alternative to the bootstrap is to adapt the simulation approach due to Lin, Fleming and Wei (1994). Consider $\psi_1, \psi_2, \ldots, \psi_n$ randomly generated from a standard normal distribution. They are multiplied to 'disturb' $N_i(\cdot)$ as in the process of $\widetilde{\mathcal{E}}(y)$, which is

$$n^{-1/2} \sum_{i=1}^{n} \left[\{\boldsymbol{x}_0 - \widehat{\boldsymbol{\mu}}_*(y)\}^{\mathrm{T}} \widehat{D}^{-1} \int_0^\infty \{\boldsymbol{x}_0 - \widehat{\boldsymbol{\mu}}^*(y)\} \{\widehat{e}_0(y) + \boldsymbol{x}_i \widehat{\boldsymbol{\beta}}\} d\{\psi_i N_i(y)\} + \int_y^\infty \frac{\{\widehat{e}_0(u) + \boldsymbol{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}\} d\{\psi_i N_i(u)\}}{\sum_j \Delta_j(u)} \right].$$

Conditional on the observed data, $\widetilde{\mathcal{E}}(y)$ is zero-mean Gaussian process. In fact, it has the same limiting distribution as that of $\mathcal{E}(y)$. Thus, the distribution of $\mathcal{E}(\cdot)$ can be simulated by repeatedly generating normal batches of $\{\psi_i\}$'s. To determine the values for the confidence bands, large amount of $\widetilde{\mathcal{E}}(\cdot)$ can be simulated to calculate the value of h_p such that $\mathrm{pr}\{\max_{Y_i} |\widetilde{\mathcal{E}}(Y_i)| > h_p\} = 1 - p$. Thus the confidence bands at pth level is approximately $\widehat{e}(y \mid \boldsymbol{x}_0) \mp h_p/\sqrt{n}$. In practice, to avoid possible negative values for the confidence bands, appropriate transformation such as the log-transformation can be used for confidence band construction. As demonstrated in Lin, Fleming and Wei (1994), additional weight functions can be also incorporated to change the relative widths given possible differential influence at the different y-values.

2.5 Outcome-dependent coefficients and Goodness-of-fit

Similar as for the linear regression model, model adequacy assessment is necessary for the additive expectancy regression model (3) to evaluate its proper use and interpretation. Unlike the linear regression model, the expectancy regression model has a key assumption of constant additivity on the expectancy functions for the covariates. This assumption is powerful in summarizing the observed difference of the outcomes due to that of the covariates, yet it requires strong 'overall' additivity on the expectancies. One straightforward way to relax this seemly stringent assumption is to assume this model:

$$e(y \mid \boldsymbol{x}) = e_0(y) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}(y), \tag{6}$$

where $\boldsymbol{\beta}(\cdot)$ may not be constant but outcome-dependent. This model would be practically useful when the covariate effect on the outcome's expectancy varies due to its magnitude. For example, the covariate effect on expectancy may gradually disappear as the outcome gets larger. One direct application of this model is, however, when $\boldsymbol{\beta}(y) \equiv \boldsymbol{\beta}$, it reduces to the additive expectancy regression model. Hence the visual assessment on the constant additivity can be done by plotting its estimate against the outcomes. An approximately horizontal line may suggest the goodness-of-fit of model (3).

In the parametric models when $\boldsymbol{\beta}(\cdot)$ can be characterized by certain parameter $\boldsymbol{\theta}$ as $\boldsymbol{\beta}(\cdot, \boldsymbol{\theta})$, it is relatively straightforward to use the maximum likelihood method to compute $\widehat{\boldsymbol{\theta}}$ and test the hypothesis of constant additivity. When $\boldsymbol{\beta}(\cdot)$ is unknown, however, some special tool is needed. Consider the individual terms in $U(\boldsymbol{\beta})$,

$$U_i(\boldsymbol{\beta}) = \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y_i)\}\left\{\widehat{e}_0(y_i; \boldsymbol{\beta}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\}, \tag{7}$$

$i = 1, 2, \ldots, n$, which are indeed the Schoenfeld residuals (Schoenfeld, 1981),

$$\int_{y_{i-1}}^{y_i} \sum_{i=1}^{n} \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}\left\{\widehat{e}_0(y; \boldsymbol{\beta}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\} dN_i(y),$$

if we do not distinguish between y_i's and their order statistics. Let $V(y; \boldsymbol{\beta}) = \sum_i \Delta_i(y)\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}^{\mathrm{T}}\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}/\sum_i \Delta_i(y)$. Then similar to that in Grambsch and Therneau (1994), the jth element of $\boldsymbol{\beta}(y)$ can be approximated by that of $\widehat{\boldsymbol{\beta}} + U_i(\widehat{\boldsymbol{\beta}})V^{-1}(y_i, \widehat{\boldsymbol{\beta}})$. Thus a scatter plot of the approximating terms against time would yield a visual tool for checking constant $\boldsymbol{\beta}(\cdot)$: a significant deviation from zero would suggest that $\boldsymbol{\beta}(y)$ is outcome-dependent. A formal hypothesis testing for zero slope in the scatter plot would give more guidance as well.

As suggested by Grambsch and Therneau (1994), different transformation of the outcomes would lead to a variety of goodness-of-fit tests. Consider a known transformation of $\phi(\cdot)$ for the outcome y. Instead of fitting a linear regression model of the jth element of the approximated $\boldsymbol{\beta}(y)$ against y, it can be fitted against $\phi(y) - \bar{\phi}$, where $\bar{\phi}$ is the average of the $\phi(\cdot)$ over the observed outcomes. Then the regression coefficients, $\widehat{\boldsymbol{\eta}}$, say, and its standardized quadratic form of $\widehat{\boldsymbol{\eta}}^{\mathrm{T}}\widehat{\Sigma}_{\eta}^{-1}\widehat{\boldsymbol{\eta}}$ would serve an asymptotically χ^2-test statistic to test the constant $\boldsymbol{\beta}(\cdot)$. Specific choices for $\phi(\cdot)$ include, for examples, step functions of y with known jumps, or simply $N(y-)$.

3 Examples and numerical studies

To understand the implication of the proposed additive expectancy regression model, we consider the following special baseline expectancy functions: (1) $e_0(y) = y + 1$; (2) $e_0(y) = 1$, i.e., exponentially distributions; (3) $e_0(y) = 1/(y + 1)$. Assume that there is the only covariate x of the binary indicator, and $\beta = 1$, respectively. Their corresponding expectancy functions, hazard functions, cumulative distribution functions and density functions are plotted in Figures (3), (4) and (5), respectively. In the first example, the baseline expectancy is monotonically increasing, which implies the monotonically decreasing hazard functions. Their density functions are highly right-skewed, thus a naive use of the normal-based regression methods may not be appropriate for such distributions. In the second example, the density functions are similarly right-skewed, when the expectancy functions are constant. In this example, the hazard functions are also constant, which means that the Cox proportional hazards model would apply. In the third example, the Cox model, however, may not well fit, since the hazard functions are apparently identical at $y = 0$.

[Figures (3)-(5) about here]

Since the theory of maximum likelihood estimation has been well established for the parametric models, moderate simulations focus on assessing the validity of the proposed semiparametric estimation procedure of the regression parameter in the additive expectancy regression model. In the actual simulations, one more continuous covariate z is also simulated according to the uniform distribution $U(0, 1)$, in addition to the simulated binary indicator x. The observations of y are simulated under the model of

$$e(y \mid x, z) = e_0(y) + x\beta + z\gamma,$$

where γ is also parameter. The distribution functions of y are specified by $\beta = \gamma = 0$, 0.5 and 1, respectively, along with the aforementioned baseline expectancy function examples. For each simulated data set, the observations of (y, x, z)'s are generated n times, where n is the sample size of 50, 100 or 200. Simulation studies are summarized in Table (1). Each entry in the table is based on 1000 simulated data sets. In the table, the bias is defined as the difference between the average of 1000 estimated coefficients and their true value, and the coverage probability is defined as the percentage of the nominal 95% confidence intervals containing the true value. As shown in the table, the estimates are virtually unbiased and the coverage probabilities are mostly close to the nominal level, especially when the sample sizes are relatively large.

[Table (1) about here]

Additional analysis is also done with the BMI data set mentioned earlier in this article. As displayed in the histograms, the distributions of both boys and girls in their BMI are right skewed. The computed skewnesses are 1.68 and 1.63 for the boys and girls, respectively. In addition, their kutosis are computed as 4.07 and 3.55, respectively. These summary statistics may suggest that a normal-based model may not fit well. When the log-transformation apply to the BMI, the distributions are 'normalized' to certain degree, with the skewness of 1.08 and 1.21, respectively, although the interpretation of differences in the log scale would result in different meanings from that in the original scale. A simple linear regression model of $y = x\beta + \epsilon$ with normal random error is fitted. The regression parameter is estimated as -0.298, which means the girls would have on average 0.298 less in the BMI than the boys. Its standard error is 0.840 and leads to a 95% confidence interval of $(-1.944, 1.349)$ with p-value of 0.72. Thus the difference in the mean BMI is not significant between the boys and the girls in this study. However, a normal probability plot of the Pearson residuals in Figure (6) may not suggest that this linear regression model fits well. The linear regression model also applies to the log-transformed BMIs, which yields a regression parameter estimate of -0.009 with the standard error of 0.035 (p-value= 0.8). The normal probability plot of residuals for the log-transformed outcomes seemly fits the normal assumption better, although the interpretation of the regression coefficient is quite different from that in the original scale. The additive expectancy model of $e(y \mid x) = e_0(y) + x\beta$ is also fitted. For its parametric version, the baseline expectancy function is chosen to be the $\alpha_0/(\alpha_1 + \alpha_2 y)$. The maximum likelihood estimate of the regression parameter is obtained as -0.624 with the standard error of 0.524. Therefore the difference in BMI expectancy is not significant between the boys and the girls, either. In the semiparametric additive expectancy regression model when the baseline is unknown, the regression parameter is estimated as -0.511 with the standard error of 0.831. Again, this result suggests that the girls tend to lower expectancy in their BMIs than the boys. This difference, however, may need to be confirmed by more samples included.

[Figure (6) about here]

4 DISCUSSION

There are two features of the expectancy regression models such as the one studied in this article. The first feature is parallel to that of the usual linear regression model, which usually fits mostly for the normally distributed outcomes. The expectancy regression model, however, may fit mostly for the exponentially distributed outcomes. Therefore the exponential distribution seems to have same role in the expectancy regression model as that of the normal distribution in the usual linear regression model. As a result of such feature, the second feature of the expectancy regression model is that, it may fit more appropriately with the time wise outcomes, such as height and weight, whose expectancy has clear interpretation in their stochastic ordering. In fact, due to the consideration of the stochastic ordering of the outcomes, the expectancy regression modeling is developed with the same spirit of the Cox hazards model, that is, modeling the outcome distributions rather that their summary statistics.

The significance of expectancy regression models is not to replace the role of usual linear regression model in actual data analysis, but rather provide an alternative when the expectancy of the outcome is of great interest, such as in the real estate when the expectancy of a property value needs to be evaluated with the information from those recently sold in the market. Thus it may have more value in short-term forecast in, for examples, resource planning, market research or clinical consultation.

APPENDIX: PROOFS

In this section, we will establish asymptotic results mainly of Lemma 2 and Theorem 3. Proofs of Property 1 and Corollary 4 are straightforward and hence omitted.

Proof of Lemma 2.

Denote $M_i\{y; \boldsymbol{\beta}_*, e_*\} = N_i(y) - \int_0^y \Delta_i(y)\lambda(y \mid \boldsymbol{x}_i; \boldsymbol{\beta}_*)dy$, $i = 1, 2, \ldots, n$. Then $\{M_i(\cdot; \boldsymbol{\beta}_*, e_*); i = 1, 2, \ldots, n\}$ form martingales. As a result,

$$\sum_{i=1}^{n}[\{e_*(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}dN_i(y) - \Delta_i(y)d\{y + e_*(y)\}] = \sum_{i=1}^{n}\{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_i(y; \beta_*, e_*). \qquad \text{(A·1)}$$

Comparing it with $\sum_i [\{\widehat{e}(y;\boldsymbol{\beta}_*) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\} dN_i(y) - \Delta_i(y) d\{y + \widehat{e}(y;\boldsymbol{\beta}_*)\}] = 0$, we obtain that

$$\{\widehat{e}_0(y;\boldsymbol{\beta}_*) - e_*(y)\} \sum_{i=1}^n dN_i(y) - \left\{\sum_{i=1}^n \Delta_i(y)\right\} d\left\{\widehat{e}_0(y;\beta_*) - e_*(y)\right\} = -\sum_{i=1}^n \{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\} dM_i(y;\beta_*, e_*).$$

By solving this equation with respect to $\widehat{e}_0(y;\boldsymbol{\beta}_*) - e_*(y)$, there is the following martingale representation for $\widehat{e}(\cdot;\beta_*)$ such that

$$\widehat{e}_0(y;\boldsymbol{\beta}_*) - e_*(y) = -A_n(y)^{-1} \int_y^\infty \frac{A_n(u)}{C_n(u)} \left[n^{-1} \sum_{i=1}^n \{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\} dM_i(u;\boldsymbol{\beta}_*, e_*) \right], \qquad (\mathrm{A}{\cdot}2)$$

where $C_n(y) = n^{-1}\sum_i \Delta_i(y)$ with $\lim_n C_n(y) = \lim_n A_n(y) = E\{\bar{F}(y \mid \boldsymbol{X})\}$. Then by standard martingale theory for counting processes, it is straightforward that $\widehat{e}_0(\boldsymbol{\beta}_*)$ is consistent and asymptotically normal with the specified variance computed in Lemma 2.

Since $\partial M_i(y;\boldsymbol{\beta}_*, e_*)/\partial\boldsymbol{\beta} = \int_0^y \Delta_i(u)\{1 + e_*'(u)\}\boldsymbol{x}_i / \{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_* + e_*(u)\}^2 du$, therefore by (A·2),

$$\frac{\partial\widehat{e}_0(y;\boldsymbol{\beta}_*)}{\partial\beta} = -A_n(y)^{-1} \int_y^\infty \frac{A_n(u)}{C_n(u)} n^{-1} \left[\sum_{i=1}^n \boldsymbol{x}_i dM_i(u;\boldsymbol{\beta}_*, e_*) + \sum_{i=1}^n \{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\} d\left\{\frac{\partial M_i(u;\boldsymbol{\beta}_*, e_*)}{\partial\boldsymbol{\beta}}\right\} \right]$$

$$= -A_n(y)^{-1} \int_y^\infty \frac{A_n(u)}{C_n(u)} \left[n^{-1} \sum_{i=1}^n \boldsymbol{x}_i dM_i(u;\boldsymbol{\beta}_*, e_*) \right] - A_n(y)^{-1} \int_y^\infty \frac{A_n(u)}{C_n(u)} \left[n^{-1} \sum_{i=1}^n \frac{\Delta_i(u)\{1 + e_*'(u)\}\boldsymbol{x}_i}{e_*(u) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*} du \right]$$

$$= -E\{\bar{F}(y \mid \boldsymbol{X})\}^{-1} \int_y^\infty E\{\boldsymbol{X}\bar{F}(y \mid \boldsymbol{X})\lambda(y \mid \boldsymbol{X})\} du + o_p(1)$$

$$= -\boldsymbol{\mu}_*(y) + o_p(1)$$

as stated in Lemma 2.

Proof of Theorem 3.

Since

$$n^{-1}\frac{\partial U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}}{\partial\boldsymbol{\beta}} = n^{-1} \sum_{i=1}^n \int_0^\infty \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\} \{\widehat{e}_{0,\beta}(y;\boldsymbol{\beta}) + \boldsymbol{x}_i\} dN_i(y)$$
$$= n^{-1} \sum_{i=1}^n \int_0^\infty \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\} \{\boldsymbol{x}_i - \boldsymbol{\mu}_*(y)\}^{\mathrm{T}} dN_i(y) + o_p(1),$$

it is then seen that $n^{-1}\partial U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}/\partial\boldsymbol{\beta}$ converges to D almost surely as in Theorem 3. In addition, consider a decomposition of $n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}$ as

$$n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\} = n^{-1/2}U(\boldsymbol{\beta}_*, e_*) + n^{-1/2}\left[U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\} - U(\boldsymbol{\beta}_*, e_*)\right].$$

The first term is indeed $n^{-1/2}\sum_i \int_0^\infty \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}\{e_*(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_i(y)$. The second term can be further written as

$$n^{-1/2}\left[U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\} - U(\boldsymbol{\beta}_*, e_*)\right] = n^{-1/2}\sum_{i=1}^n \int_0^\infty \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}\{\widehat{e}_0(\boldsymbol{\beta}_*) - e_*(y)\}dN_i(y)$$

$$= - n^{-1/2}\sum_{i=1}^n \int_0^\infty \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\} A_n(y)^{-1}\int_y^\infty \frac{A_n(u)}{C_n(u)}\left[n^{-1}\sum_{j=1}^n \{e_*(u) + \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_j(u)\right]dN_i(y)$$

$$= - n^{-1/2}\sum_{j=1}^n \int_0^\infty \frac{A_n(u)}{C_n(u)}\left[\int_0^u \frac{n^{-1}\sum_i \{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\}dN_i(y)}{A_n(y)}\right]\{e_*(u) + \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_j(u)$$

$$= - n^{-1/2}\sum_{j=1}^n \int_0^\infty \frac{A_n(u)}{C_n(u)}\{\widehat{\boldsymbol{\mu}}^*(u) - \bar{\boldsymbol{x}}(y)\}\{e_*(u) + \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_j(u).$$

By summing over these two terms,

$$n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\} = n^{-1/2}\sum_{i=1}^n \int_0^\infty \left[\{\boldsymbol{x}_i - \bar{\boldsymbol{x}}(y)\} - \frac{A_n(u)}{C_n(u)}\{\widehat{\boldsymbol{\mu}}^*(u) - \bar{\boldsymbol{x}}(y)\}\right]\{e_*(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_i(y)$$

$$= n^{-1/2}\sum_{i=1}^n \int_0^\infty \{\boldsymbol{x}_i - \boldsymbol{\mu}^*(u)\}\{e_*(y) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*\}dM_i(y) + o_p(1).$$

As a result, $n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}$ converges in distribution to zero-mean normal distribution with asymptotic variance of V as specified in Theorem 3. Furthermore, a straightforward Taylor expansion of $U\{\widehat{\boldsymbol{\beta}}, \widehat{e}_0(\widehat{\boldsymbol{\beta}})\}$ at $\boldsymbol{\beta} = \boldsymbol{\beta}_*$ would lead to

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*) = \left[n^{-1}\frac{\partial U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}}{\partial \boldsymbol{\beta}}\right]^{-1} \cdot \left[-n^{-1/2}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}\right] + o_p(1)$$

and hence its asymptotic normality in Theorem (3) as well.

Given the condition on $\boldsymbol{\mu}_*(\cdot)$, D is nonsingular. Since $n^{-1}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}$ converges to zero almost surely, there exists a neighborhood of $\boldsymbol{\beta}_*$ such that $-[n^{-1}\partial U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}/\partial\boldsymbol{\beta}]^{-1}\cdot[n^{-1}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}]$ and $-D^{-1}[n^{-1}U\{\boldsymbol{\beta}_*, \widehat{e}_0(\boldsymbol{\beta}_*)\}]$ are as well within the same neighborhood. Hence the consistency holds for $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*$ at 0 given the arbitrarily small size of such a neighborhood. Straightforward calculation with the Taylor expansion would lead to the consistency of the variance estimators as well.

REFERENCES

Grambsch, P. M. and Therneau, T. M. (1994). "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, 81, 515-526.

Guess, F. and Proschan, F. (1988), "Mean residual life: theory and application," in *Handbook of Statistics, vol 7*, eds. P. R. Krishnaiah and C. R. Rao, 215-224, New York: North-Holland.

Gupta, R. C. and Kirmani, S. N. U. A. (1998), "On the proportional mean residual life model and its implication," *Statistics*, 32, 175-187.

Hall, W. J. and Wellner, J. A. (1984), "Mean residual life," in *Proceedings of the International Symposium on Statistics and Related Topics*, eds. M. Csorgo, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, 169-184, Amsterdam: North-Holland.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data, 2nd Ed.*, Hoboken: John Wiley.

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.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models, 2nd Ed.*, London: Chapman and Hall.

Oakes, D. and Dasu, T. (1990), "A note on residual life," *Biometrika*, 77, 409-410.

Schoenfeld, D. (1980). "Chi-squared goodness-of-fit tests for the proportional hazards regression model," *Biometrika*, 67, 145-153.

Yang, G. L. (1978), "Estimation of a biometric function," *Journal of American Statistical Association*, 6, 112-116.

Last revised: June 27, 2005

Table 1: Summary of Simulation Studies for Model $e(y \mid x, z) = e_*(y) + x\beta_* + z\gamma_*$ with 1000 Replicates

		$\beta_* = \gamma_* = 0$				$\beta_* = \gamma_* = 0.5$				$\beta_* = \gamma_* = 1$			
		x		z		x		z		x		z	
$e_*(t)$	n	Bias	Cov. Prob.	Bias	Cov. Prob.	Bias	Cov. Prob.	Bias	Cov. Prob.	Bias	Cov. Prob.	Bias	Cov. Prob.
$y+1$	50	0.0608	0.958	0.175	0.916	0.2063	0.951	0.1818	0.963	-0.2161	0.955	0.0746	0.941
	100	0.0442	0.947	-0.0128	0.967	-0.2354	0.958	-0.0031	0.937	0.1223	0.947	-0.0069	0.963
	500	-0.0797	0.941	-0.1851	0.936	-0.0057	0.946	-0.0977	0.951	0.0063	0.957	-0.0159	0.944
1	50	0.0311	0.921	-0.1554	0.949	-0.1884	0.957	-0.0724	0.934	-0.0331	0.938	0.1211	0.931
	100	0.1103	0.962	0.1329	0.945	-0.1007	0.963	-0.1504	0.957	0.2534	0.958	0.0549	0.935
	500	0.0793	0.952	-0.1035	0.959	-0.0197	0.975	0.0177	0.955	0.051	0.965	-0.0325	0.946
$1/(y+1)$	50	0.1940	0.935	-0.0452	0.942	-0.0441	0.937	-0.1161	0.949	0.2443	0.951	0.1375	0.968
	100	-0.1013	0.953	-0.1172	0.941	0.0329	0.947	-0.0795	0.947	0.0804	0.943	0.0421	0.957
	500	0.0172	0.963	-0.0307	0.954	-0.0321	0.962	-0.0358	0.966	-0.0786	0.941	0.0317	0.953

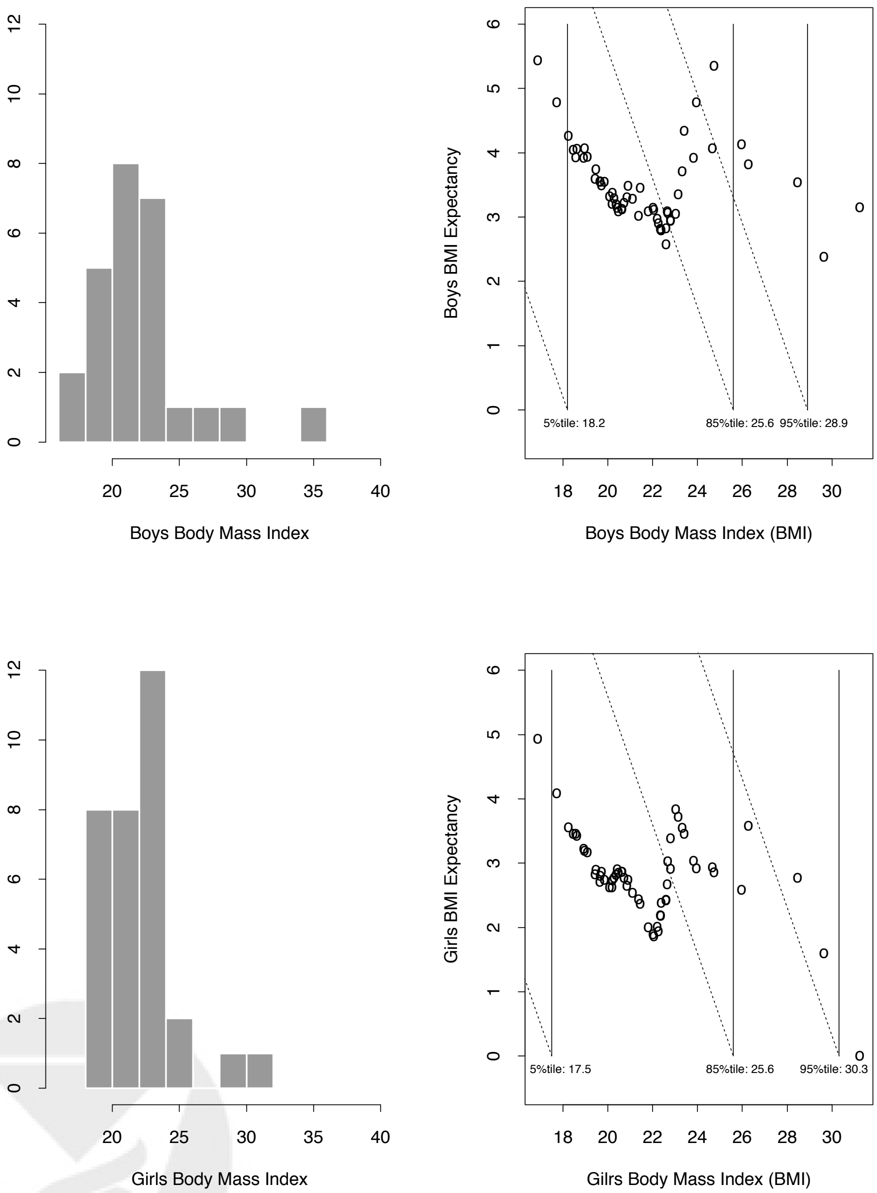

Figure 1: Histograms and estimated BMI expectancies for boys and girls of age 18, respectively. Teens are considered underweight if less than 5th percentile, normal if between 5th and 85th percentiles, overweight at-risk if between 85th and 95th percentiles and overweight if more than 95th percentile. Solid lines represent the cutoffs based on current BMI values. Dotted lines represent the cutoffs based on BMI expectancies.

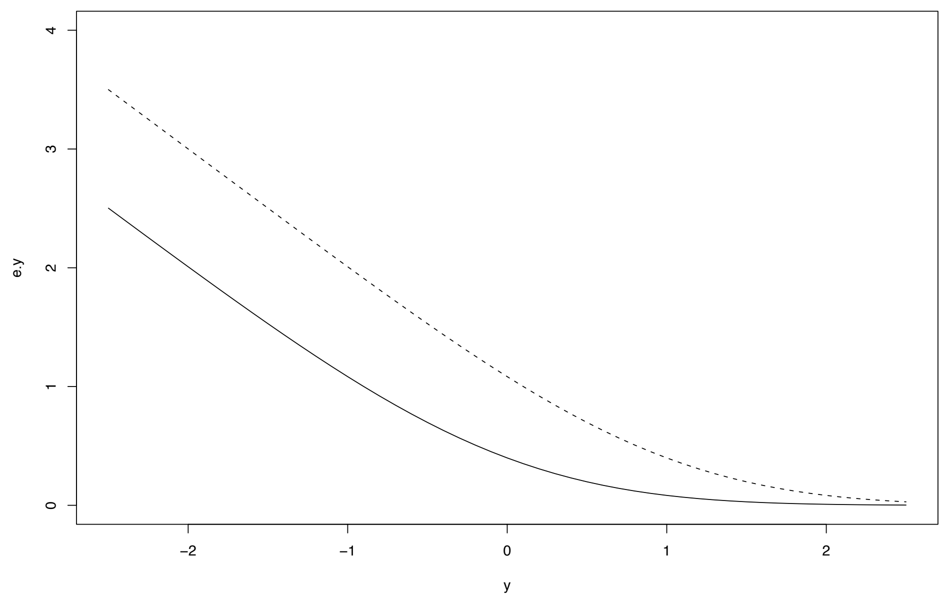

Figure 2: Expectancy functions in the normal linear regression model. Solid line represents that of $\boldsymbol{x} = 0$. Dotted line represents that of $\boldsymbol{x} = 1$.

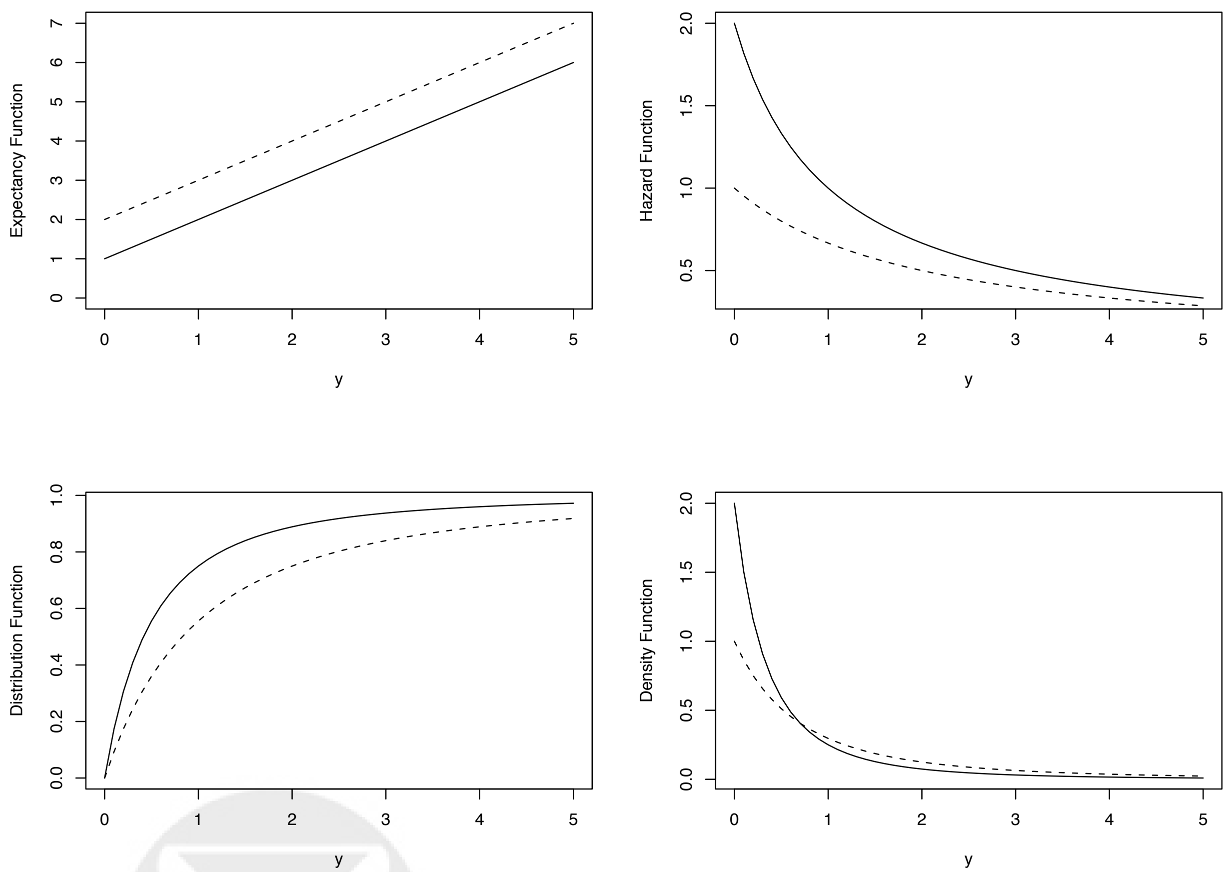

Figure 3: Additive expectancy model with $e_0(y) = y + 1$. Solid lines are of $\boldsymbol{x} = 0$ and dotted lines of $\boldsymbol{x} = 1$, respectively.

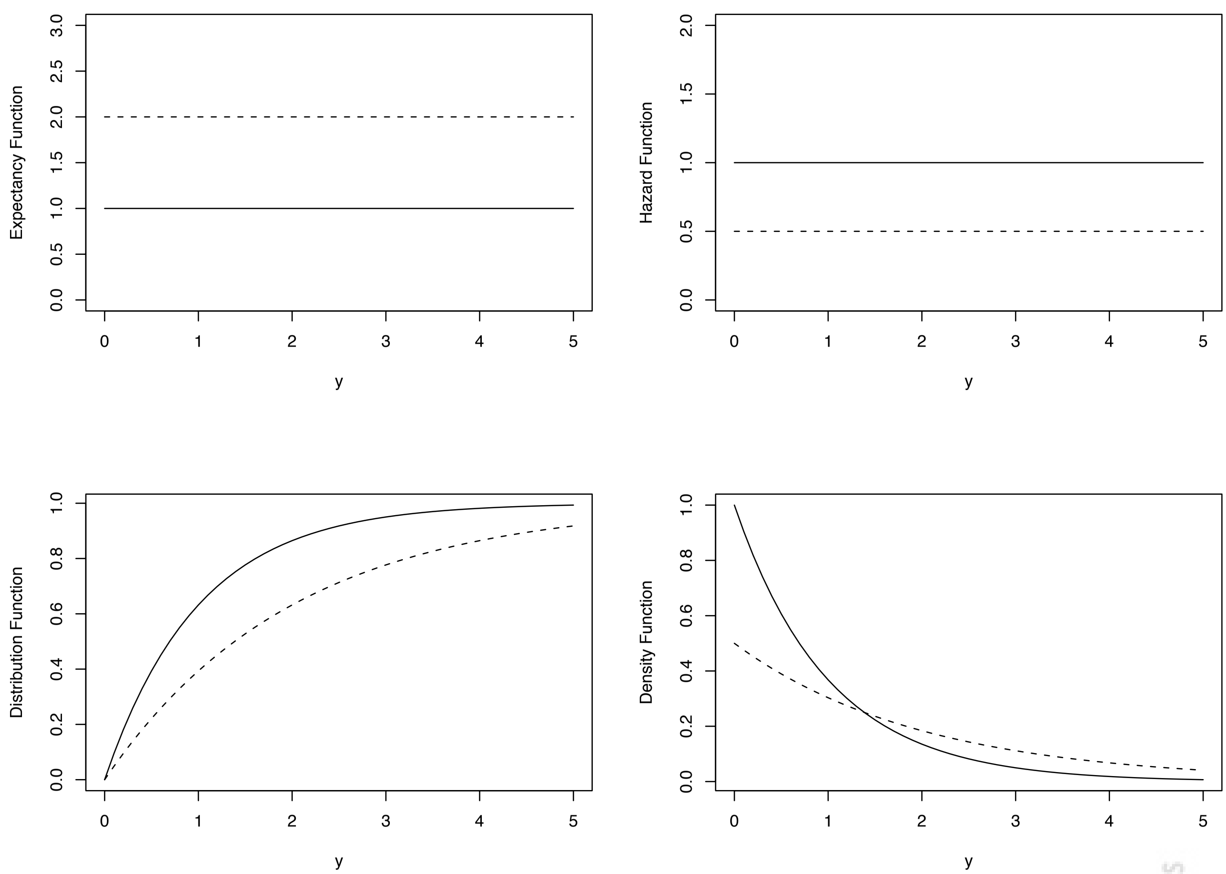

Figure 4: Additive expectancy model with $e_0(y) = 1$. Solid lines are of $\boldsymbol{x} = 0$ and dotted lines of $\boldsymbol{x} = 1$, respectively.

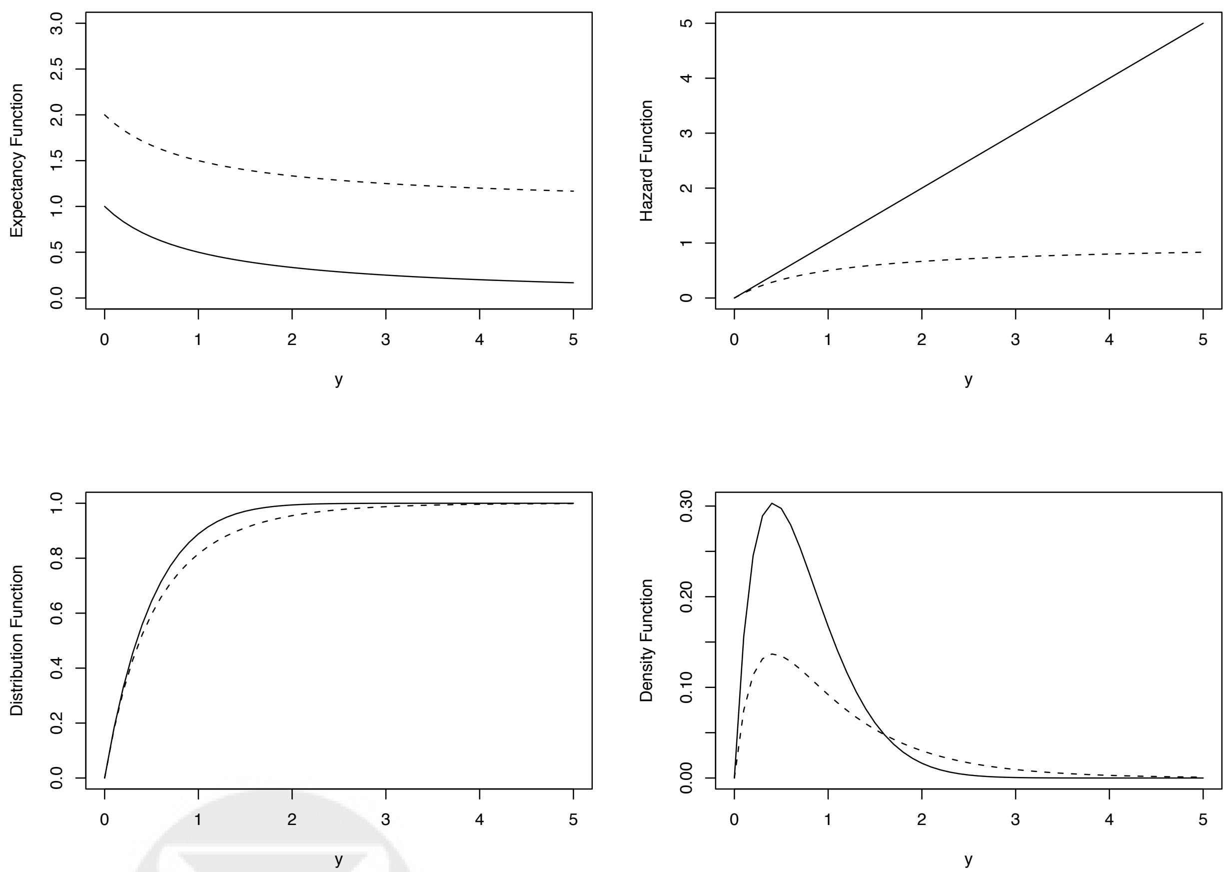

Figure 5: Additive expectancy model with $e_0(y) = 1/(y+1)$. Solid lines are of $\boldsymbol{x} = 0$ and dotted lines of $\boldsymbol{x} = 1$, respectively.

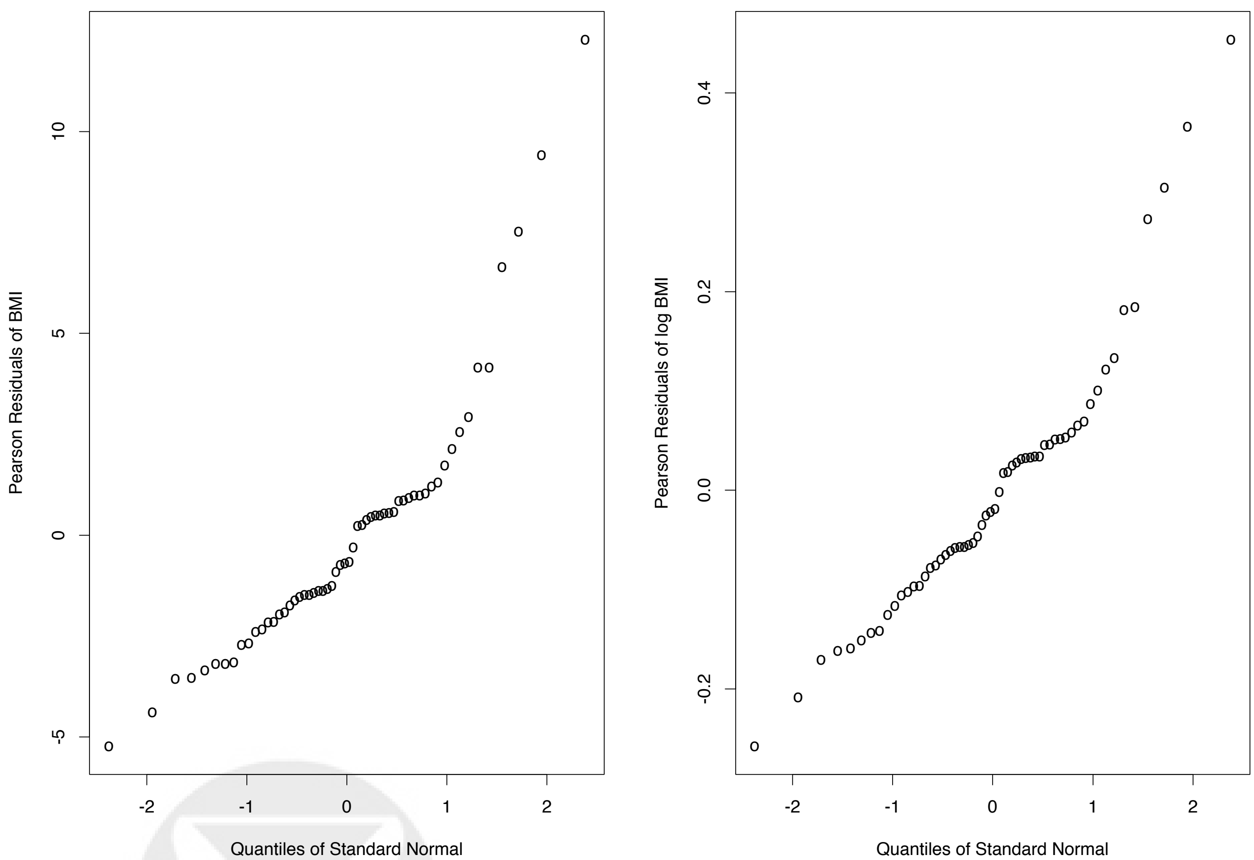

Figure 6: Normal probability plot for the Pearson residuals of BMI in model $y = x\beta + \epsilon$.