Exponential tilt models for two-group comparison with censored data

Chi Wang a,*, Zhiqiang Tan b, Thomas A. Louis c

a Markey Cancer Center, University of Kentucky, 800 Rose St., Lexington, KY 40536, USA
b Department of Statistics, Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08854, USA
c Department of Biostatistics, Johns Hopkins University, 615 N.Wolfe St., Baltimore, MD 21205, USA

A R T I C L E   I N F O

Article history:
Received 17 March 2010
Received in revised form 8 September 2010
Accepted 11 September 2010
Available online 29 September 2010

Keywords:
Censored data
Exponential tilt model
Non-parametric likelihood
Proportional hazards model

A B S T R A C T

We study application of the Exponential Tilt Model (ETM) to compare survival distributions in two groups. The ETM assumes a parametric form for the density ratio of the two distributions. It accommodates a broad array of parametric models such as the log-normal and gamma models and can be sufficiently flexible to allow for crossing hazard and crossing survival functions. We develop a nonparametric likelihood approach to estimate ETM parameters in the presence of censoring and establish related asymptotic results. We compare the ETM to the Proportional Hazards Model (PHM) in simulation studies. When the proportional hazards assumption is not satisfied but the ETM assumption is, the ETM has better power for testing the hypothesis of no difference between the two groups. And, importantly, when the ETM relation is not satisfied but the PHM assumption is, the ETM can still have power reasonably close to that of the PHM. Application of the ETM is illustrated by a gastrointestinal tumor study.

1. Introduction

Randomized trials are often used to evaluate treatments in clinical and public health researches. Participants are randomly assigned to one of two (or several) groups and treatment effectiveness is assessed by comparing the event times (e.g., survival times) of the two groups. Parametric models such as Weibull, gamma or log-normal can be used to estimate the treatment effect. However, in many situations it is difficult to choose a parametric model, which makes semiparametric models become attractive. These models combine a parametric form for the treatment effect with a nonparametric "baseline model". The most popular semiparametric model is the Proportional Hazards Model (PHM) (Cox, 1972). It has the following form for the two-group comparison:

$$\lambda_1(t) = \beta \lambda_0(t),$$

where $\lambda_0(t)$ is the unspecified hazard function of the placebo group, $\lambda_1(t)$ is the hazard function of the treatment group, and $\beta$ quantifies the treatment effect. This model assumes that the ratio of hazard functions, $\beta$, is time-constant, with $\beta < 1$ producing a uniformly smaller hazard for the treatment group.

In addition to the PHM, there are several other semiparametric survival models such as the accelerated failure (Louis, 1981; Tsiatis, 1990), accelerated hazards (Chen and Wang, 2000) and transformation models (Cheng et al., 1995; Zeng and Lin, 2007).

* Corresponding author.
E-mail addresses: chi.wang@uky.edu (C. Wang), ztan@stat.rutgers.edu (Z. Tan), tlouis@jhsph.edu (T.A. Louis).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.09.009
The constant hazard ratio assumption in the PHM is quite strong. For example, as pointed out by Chen and Wang (2000), in randomized trials, there may be a lag period for the treatment to be effective. So the hazard ratio at the beginning of the study should be close to one, which will be different from the hazard ratio at the time when the treatment has taken effect. As another example, certain aggressive treatments may increase risks of short-term death but may improve long-term survival, which yield crossing hazard and survival functions (Stablein and Koutrouvelis, 1985; Zeng and Lin, 2007). To relax this assumption, one may consider adding time as a covariate in the log-linear model. However, choice of form for this covariate term is difficult. Guidance is provided by the time-dependent hazard ratio functions induced by parametric models such as the log-normal and gamma. But the hazard ratio functions for candidate parametric models are usually very complicated.

As an alternative to building hazard ratio models, we consider the Exponential Tilt Model (ETM). Rather than comparing hazards, the ETM uses a log-linear (i.e., exponential multiplier) model for the density ratio, leaving the baseline density unspecified. In contrast to hazard ratios, for a broad class of parametric models, density ratios have very simple forms. Thus, relatively parsimonious, log-linear models for the density ratio accommodate a broad class of underlying distributions. Following this idea, we specifically consider an ETM with three functions of follow-up time, referring to it as a "general ETM". This general ETM is exact for exponential, log-normal and gamma models. Another appealing feature of ETMs is that they can be sufficiently flexible to allow for crossing hazards and crossing survival functions.

For completely observed data, ETMs have been extensively studied in connection with case-control studies (Qin, 1998) and biased sampling (Gilbert et al., 1999). Shen et al. (2007) extended application of such models to censored event time data. They proposed an estimating equation approach to estimate model parameters. Alternatively, Li and Lin (2009) considered an EM algorithm for parameter estimation. But the asymptotic properties of these estimators have not been studied.

In this article, we develop and implement a nonparametric maximum likelihood approach for estimating ETM parameters and via simulation compare the ETM to the PHM. In Section 2, we introduce the ETM and discuss its connection to logistic regression. In Section 3, we define and evaluate the proposed inference procedures. In Section 4, we present simulation results that evaluate performance of the ETM and compare it with the PHM. In Section 5, we illustrate our methods by comparing chemotherapy and combined chemotherapy/radiotherapy in treating gastrointestinal tumors. In Section 6 we discuss our findings and identify future research.

2. Exponential tilt models

We focus on the two-group comparison with no covariate adjustment. Let Z be the treatment assignment indicator, taking value 0 for the placebo group and 1 for the treatment group. Let T be the continuous or discrete potential event time and C be the potential censoring time, independent of T given Z. Suppose the clinical trial ends at time t, the observed event (survival) time is X = min(T, C). Let Δ = I(T ≤ min(C, t)) be the censoring indicator.

2.1. Framework

The Exponential Tilt Model (ETM) provides an attractive alternative to the PHM. It assumes that
\[
dF_1(t) = \frac{dF_0(t) \exp(h(t, β))}{dF_0(t)}, \quad t \leq t.
\]
Here, \(dF_0(t)\) is the unspecified, generalized density function of the placebo group and \(dF_1(t)\) is the density function of the treatment group. These are related via a specified parametric function \(h\) with a \(r\)-dimensional vector of unknown parameters \(β\). To see the connection to parametric models, consider the following example:

Example 1. Consider a log-normal model:
\[
F_0(\cdot) \sim LN(μ_0, σ_0^2), \quad F_1(\cdot) \sim LN(μ_1, σ_1^2).
\]
The density ratio function is
\[
\frac{dF_1}{dF_0} = \exp(β_1 \log(t) + β_2 [\log(t)]^2),
\]
where \(β = (β_1, β_2)^T\) with \(β_1 = (μ_1/σ_1^2 − μ_0/σ_0^2)\) and \(β_2 = 1/(2σ_0^2) − 1/(2σ_1^2)\). The density ratio function satisfies (1) with \(h(t, β) = β_1 \log(t) + β_2 [\log(t)]^2\).

Table 1 lists \(h\) for commonly used parametric models. Within \(h\), each function of \(t\) is called a basis and many parametric models share basis terms. For example,
\[
h^*(t, β) = β_1 t + β_2 \log(t) + β_3 [\log(t)]^2,
\]
includes three basis functions \(t, \log(t)\) and \([\log(t)]^2\), which are all the basis functions for the exponential, log-normal and gamma models. We refer to \(h^*\) as the "general" model (it is our working model) and study performance relative to the true \(h\). Our \(h^*\) is but one example of a working model. For example, it can be expanded to include the Weibull model by adding the basis \(t^{γ}\). But such expansion may cause unstableness in parameter estimation based on some numerical studies. So we will focus on \(h^*\) in this paper.
2.2. Interpretation of model parameters

The interpretation of model parameters is best illustrated in a simple case where \( h(t, \beta) = \beta_t t \). From

\[
\frac{P(T = t | Z = 1)}{P(T = t | Z = 0)} = \frac{dF_1(t)}{dF_0(t)} = \exp(\alpha + \beta_t t), \quad \alpha = -\log \left( \int \exp(\beta_t t) \, dF_0(t) \right),
\]

we have,

\[
\log \left( \frac{P(Z = 1 | T = t)}{P(Z = 0 | T = t)} \right) = \log \left( \frac{P(T = t | Z = 1) \times P(Z = 1)}{P(T = t | Z = 0) \times P(Z = 0)} \right) = \alpha^* + \beta_t t,
\]

(3)

where \( \alpha^* = \alpha + \log[P(Z = 1)/P(Z = 0)] \). Eq. (3) is a logistic regression and implies that conditional on an event at time \( t \), the probability that it is associated with the treatment group is \( e^{\alpha^* + \beta_t t}/(1 + e^{\alpha^* + \beta_t t}) \). If \( \beta_t > 0 \), as \( t \) increases, the proportion of observations coming from the treatment group increases. Equivalently, the treatment tends to produce longer survival times. In Eq. (3), only the intercept \( \alpha^* \) depends on the sample size ratio of the two groups. The interpretation of \( \beta_t \) does not depend on the sample size ratio.

3. Inferences

We assume censoring time \( \min(C, \tau) \) to be discrete with a finite number of values \( c_0, \ldots, c_{d_0}, c_{d_0} + 1 = \tau \) for the placebo group and \( c_1, \ldots, c_{d_1}, c_{d_1} + 1 = \tau \) for the treatment group. Let \( c_z = (c_1, \ldots, c_{d_z} + 1)^T, z = 0, 1 \). The data consist of \( N \) independent and identically distributed random triplets \( (X_i, \delta_i, Z_i), i = 1, \ldots, N \). Their realizations are denoted by \( (x_i, \delta_i, z_i) \). For convenience, suppose the first \( n \) observations are uncensored, where the first \( n_0 \) observations come from the placebo group and have density \( dF_0(.) \) and the next \( n_1 \) come from the treatment group with density \( dF_1(.) \). The data also contain \( m_0 \) censored observations from the placebo group and \( m_1 \) censored observations from the treatment group, with \( m_{i0} \) observations censored at time \( c_{j_0} (j = 1, \ldots, d_0 + 1) \) and \( m_{i1} \) censored at time \( c_{j_1} (j = 1, \ldots, d_1 + 1) \). Let \( N_0 = n_0 + m_{i0} \) and \( N_1 = n_1 + m_{i1} \) be the numbers of observations in the two groups. Furthermore, let \( \rho_0 = N_0/N, \rho_1 = N_1/N \).

Consider discrete distributions which, before \( \tau \), have point masses only at the observed failure times. Let \( p_i = F_0(x_i) \) and \( s_i = S_0(x_i) \), where \( S_0(t) = 1 - F_0(t) \), \( z = 0, 1 \), are survival functions. The nonparametric log-likelihood function (Owen, 2001; Vardi, 1985) is

\[
\log l(\beta, F_0) = \sum_{i=1}^{n} \log p_i + \sum_{i=n+1}^{n+m_0} \left[ h(x_i, \beta) + \log(1-s_1) - \log \left( \sum_{k=1}^{n} p_k \exp[h(x_k, \beta)] \right) \right] + \sum_{j=1}^{d_0+1} m_{i0} \log \left( \sum_{j=1}^{n} p_j I(x_j > c_{j0}) + s_0 \right) + \sum_{j=1}^{d_1+1} m_{i1} \log \left( \sum_{k=1}^{n} p_k \exp[h(x_k, \beta)] \right),
\]

subject to the constraint, \( \sum_{i=1}^{n} p_i + s_0 = 1 \).

The log profile likelihood of \( \beta \) is computed by maximizing \( \log l(\beta, F_0) \) with respect to \( F_0 \), that is, \( pl_0(\beta) = \max_{F_0} l(\beta, F_0) \). The profile MLE of \( \beta \) is obtained by maximizing \( pl_0(\beta) \), that is, \( \beta = \arg \max pl_0(\beta) \). For fixed \( \beta_0 \) and \( s_1 \), the \( p_i, i = 1, \ldots, n \), that maximize the log-likelihood (4) satisfy

\[
p_i = \frac{1}{N} [\omega_0(x_i, Q_0) + \omega_1(x_i, Q_1) \exp[\alpha + h(x_i, \beta)]]^{-1},
\]

(5)

where \( \omega_0(x_i, Q_0) = \rho_x - \sum_{j=1}^{d_0} m_{ij} I(x_i > c_{j0})/N/Q_{ij} + s_z/(1 - s_z) (1 - \sum_{j=1}^{d_0+1} m_{ij})/N/Q_{ij} \), \( z = 0, 1 \), and \( x, Q_0 = (Q_1, \ldots, Q_{d_z+1})^T \) are functions of \( \beta \) and \( F_0 \) defined as

\[
\alpha = \log(1-s_1) - \log \left( \sum_{i=1}^{n} p_i \exp[h(x_i, \beta)] \right),
\]

\[
Q_{ij} = \sum_{i=1}^{n} p_i I(x_i > c_{ij}) + s_0, \quad j = 1, \ldots, d_0 + 1,
\]

\[
Q_{ij} = \sum_{i=1}^{n} p_i I(x_i > c_{ij}) \exp[h(x_i, \beta)] + s_1, \quad j = 1, \ldots, d_1 + 1.
\]
The following is a heuristic interpretation of Eq. (5). We consider the situation that $T$ is discrete. The equation can be re-written as

$$p_t c o(x, Q_0) + p_t \exp (x + h(x, \hat{\beta})) c o(x, Q_1) = \frac{1}{N}.$$  \hspace{1cm} (7)

Notice that $p_t = p(T = x | Z = 0)$ and $p_t \exp (x + h(x, \hat{\beta})) = p(T = x | Z = 1)$. And it can be shown that, for $x = 0, 1$, $\omega_j(x, Q_j)$ converge to $P(C \geq x, Z = z)$ when evaluating at the true values of $Q_y, S_y(t)$. So the left hand side of Eq. (7) is a model based estimate of the joint density of $(X, A)$ at $X = x$ and $A = 1$. This estimate is naturally set to equal $1/N$, the empirical estimate of the joint density of $(X, A)$.

In practice, $p_l(x, \hat{\beta})$ can be obtained via the EM algorithm (Dempster et al., 1977). We define $Y = \min(T, \tau)$ and regard the $Y$ s for observations censored before $\tau$ as “missing data”. Then the EM algorithm is used to maximize the log-likelihood over $p_l s_0$ and $s_1$ while holding $\hat{\beta}$ fixed. In this algorithm, we regard $\tau$ as part of the complete data and work on $Y$ instead of $T$ because the expected value of survival time for a censored observation can hardly be derived without additional assumptions on the distribution of $T$ beyond $\tau$. In contrast, by regarding $Y$ as the complete data, we only need the distribution of $T$ up to $\tau$, which can be estimated from observed data. The details of the EM algorithm to obtain $p_l(x, \hat{\beta})$ are given in Appendix A. Then, $\hat{\beta}$ can be obtained via a Newton–Raphson method by maximizing $p_l(x, \hat{\beta})$.

Proposition 1 below studies the asymptotic properties of $\hat{\beta}$. We define $U$ as the Fisher information matrix, $U = -\lim_{N \to \infty} \partial^2 p_l(x, \hat{\beta}) / \partial \beta^2 / N$, where the limit can be shown to exist under the assumptions of the proposition. Proposition 1 shows that the profile MLE obtained from the semiparametric model (1) behaves like the profile MLE from a parametric model. Its asymptotic variance equals the inverse of $U$.

**Proposition 1.** Assume that all the elements in $\partial h(x, \beta) / \partial \beta$, $\partial^2 h(x, \beta) / \partial \beta^2$, and $(\partial h(x, \beta) / \partial \beta)(\partial h(x, \beta) / \partial \beta)^T$ are continuous and bounded by some function $\kappa(x)$ for $x \in (0, \tau)$ in a neighborhood of the true value of $\beta$ satisfying $\int_0^\tau \kappa(x) dF_0 < \infty$ and $\int_0^\tau \kappa(x) dF_1 < \infty$.

(i) $\hat{\beta}$ is a consistent estimator of $\beta$.

(ii) $\sqrt{N}(\hat{\beta} - \beta) \to N(0, U^{-1})$.

We next study the likelihood-ratio test statistic constructed using the semiparametric profile likelihood. The following proposition shows that it behaves like the likelihood-ratio test statistic from a parametric model.

**Proposition 2.** Under the same conditions as in Proposition 1, the likelihood-ratio test statistic $LRT = -2[p_l(x, \hat{\beta}) - p_l(x, \hat{\beta})]$ converges to a chi-square distribution with $r$ degrees of freedom.

To estimate survival functions at a given time point $t_0 \leq \tau$, it is natural to consider the following estimators

$$\begin{pmatrix} \hat{S}_0(t_0) \\ \hat{S}_1(t_0) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \hat{p}_i I(X_i > t_0) + \hat{s}_0 \\ \sum_{i=1}^n \hat{p}_i \exp (\hat{z} + h(X_i, \hat{\beta})) I(X_i > t_0) + \hat{s}_1 \end{pmatrix},$$

where $\hat{p}_i s_0$ and $\hat{s}_1$ are nonparametric maximum likelihood estimators based on (4), and $\hat{z}$ is calculated from (6) by plugging in these estimators. Proposition 3 shows that the survival function estimators are asymptotically linear.

**Proposition 3.** Under the same conditions as in Proposition 1,

$$\sqrt{N} \left( \begin{pmatrix} \hat{S}_0(t_0) \\ \hat{S}_1(t_0) \end{pmatrix} - \left( \begin{pmatrix} S_0(t_0) \\ S_1(t_0) \end{pmatrix} \right) \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi(t_0, X_i, A_i, Z_i, \beta, \eta) + o_p(1),$$

where $\beta$ and $\eta = (Q_{1i}, Q_{1i}^T, x)$ are evaluated at their true values, and the expression of $\phi(t_0, X_i, A_i, Z_i, \beta, \eta)$ is given in (10).

To obtain confidence intervals for survival functions, we first calculate the Wald confidence intervals for $(\log(S_0(t_0)/(1 - \hat{S}_0(t_0))), \log(S_1(t_0)/(1 - \hat{S}_1(t_0))))^T$, then transform them back into the original scale. The variance matrix of $(\log(S_0(t_0)/(1 - \hat{S}_0(t_0))), \log(S_1(t_0)/(1 - \hat{S}_1(t_0))))^T$ can be estimated by

$$\frac{1}{N^2} \sum_{i=1}^N \Phi(t_0, X_i, A_i, Z_i, \hat{\beta}, \hat{\eta}(\hat{\beta})) \phi(t_0, X_i, A_i, Z_i, \hat{\beta}, \hat{\eta}(\hat{\beta}))^T \Phi^T,$$

where $\Phi$ is a diagonal matrix with diagonal elements equal to $(1/\hat{S}_0(t_0)/(1 - \hat{S}_0(t_0)), 1/\hat{S}_1(t_0)/(1 - \hat{S}_1(t_0)))$, $\hat{\eta}(\hat{\beta})$ is defined in the proof of Proposition 1(i) and $\phi$ is defined in Remark 1 after the proof of Proposition 3 in Appendix B.

4. Simulations

Simulation studies were conducted to evaluate the performance of the ETM and to compare to the PHM.

4.1. Performance of the ETM

The survival times of the $N_0$ observations in the placebo group and the $N_1$ observations in the treatment group were independently generated from exponential distributions $F_0(t) \sim \text{Exp}(\lambda_0)$ and $F_1(t) \sim \text{Exp}(\lambda_1)$, respectively. The density ratio
of these two distributions is \(dF_1(t)/dF_0(t) \propto \exp[\beta_1 t]\), where \(\beta_1 = 1/\lambda_0 - 1/\lambda_1\). We compared the performance of the ETM to that of parametric (exponential) model, considering ETMs with two different \(h\) functions. The first uses \(h(t) = \beta_1 t\), the "true ETM"; the second uses the general \(h\) function (2), the "general ETM".

We assumed censoring to occur at six fixed time points, the 30%, 40%, 50%, 60%, 70%, 80% quantiles of \(F_0\) or \(F_1\) for the two groups, respectively, and the follow-up time \(\tau = 4.61\), the 90% quantile of \(F_1\). The censoring probability was 0.135. The true value of \(\beta_1\) was 0.125 (\(\lambda_0 = 1.6, \lambda_1 = 2\)), representing an increasing density ratio. Separate simulations were run for \(N_0=N_1=50\), 150 or 500. Under each setting, 1000 independent datasets were simulated. Results for estimating \(\beta_1\) are summarized in Table 2.

Note that the bias of the ETM estimate decreases as sample size increases. For example, for the general ETM, the bias decreases from 0.090 to 0.012, when the sample size for each group increases from 50 to 500. All the coverage probabilities of ETMs are close to the desired value 0.95. Compared to the true ETM, the standard error of the general ETM is larger. For example, for the \(N_0=N_1=500\) case, the standard error of the general ETM is four times larger than that of the true ETM. That is the price one needs to pay to gain robustness. The standard error of the true ETM is about 50% larger than that of the parametric model.

Results for estimating survival functions are summarized in Table 2. We only report results for estimating \(S_0\). Results for estimating \(S_1\) show similar patterns. The estimates are reported at time points corresponding to survival probabilities equal to 0.9, 0.5 and 0.1. The biases of survival function estimators from both the general ETM and the true ETM are negligible. The standard errors of the general ETM are larger compared to the true ETM. The coverage probabilities at the tail \((S_0(t) = 0.1)\) are higher than the desired value 0.95 when sample sizes are small \((N_0=N_1=50)\). As sample sizes increase, the probabilities become close to the desired value.

We next investigated the performance of the ETM when censoring time is continuous. Although we have not obtained theoretical results, it is of interest to numerically explore whether the proposed methods can be applied to this situation. The simulation settings were similar as before, except that the censoring time \(C\) was generated from a uniform distribution on \((0, 2\tau)\). The simulation results are summarized in Table 3. The estimates of both model coefficient and baseline survival function perform sufficiently well, indicating that the ETM may still be applicable under this situation.

### 4.2. Comparison of the ETM to the PHM

The second set of simulations compared the performance of the ETM to the Proportional Hazards Model (PHM). Three different situations were considered: (a) the assumptions of both the general ETM and the PHM are satisfied; (b) only the
assumption of the general ETM is satisfied; (c) only the assumption of the PHM is satisfied. We first illustrated the performance of the ETM and the PHM in parameter estimation. One dataset with 500 observations in each group was generated from each of the following three distributions: exponential, log-normal and Weibull (same shape parameter for both groups). We compared results based on the true ETM, the general ETM, the PHM, and the Kaplan–Meier method (K–M) (Kaplan and Meier, 1958). For the true ETM, the $h$ function was defined as the true $h$ function corresponding to the distribution from which the data were generated. Specifically, $h(t) = \beta_1 t$ for exponential distribution, $h(t) = \beta_1 t + \beta_2 \log(t)$ for gamma distribution and $h(t) = \beta_1 t^{\beta_2}$ for Weibull distribution. In our simulations, for the $h$ corresponding to Weibull distribution, we additionally assumed that $\beta_2$ was known in order to make the algorithm more stable.

We report results for estimating the survival odds ratio function using the two ETMs, PHM and K–M. Similar relations hold for estimates of the density ratio function, survival functions and the hazard ratio function. The survival odds ratio function was defined as the true $\beta_1$ $h$ function corresponding to the distribution $\times$ with $\beta_1$ corresponding to Weibull $\times$ and $\beta_2$ equal to 150 or $\times$.

In the middle is the estimated curve; the dashed curves on each side are the 95% CI limits. One can see that the estimated curves based on the true ETM (red curves) are always very close to the true functions of the two ETMs and the PHM. In the middle is the estimated curve; the dashed curves on each side are the 95% CI limits. When the model assumption for both the general ETM and the PHM are satisfied (situation a where data were generated from $\times$), estimated curves based on the K–M (black curves) are always close to the true curves in all the three situations. In our simulations, for the $h$ corresponding to Weibull distribution, we additionally assumed that $\beta_2$ was known in order to make the algorithm more stable.

We next compared the hypothesis test power of ETMs to PHMs. As for the estimation comparisons, we considered three situations and simulated data based on the three parametric distributions. For each situation, we simulated a series of data sets, making the survival time distribution of the treatment group progressively departing from that of the placebo group. For example, for situation a, the observations in the placebo group were generated from $\times$ and those in the treatment group were generated from $\times$, with $b=(0.00, 0.02, 0.04, \ldots, 0.40)$. For each $b$, we chose sample sizes $N_0=N_1$ equal to 150 or 500.

### Table 3
Comparison of parameter estimation based on ETMs and parametric model when censoring distribution is continuous.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Sample size per group</th>
<th>Method</th>
<th>Bias</th>
<th>RMSE</th>
<th>SSE</th>
<th>SEE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.125</td>
<td>50</td>
<td>True ETM</td>
<td>0.002</td>
<td>0.219</td>
<td>0.219</td>
<td>0.213</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.128</td>
<td>1.247</td>
<td>1.241</td>
<td>1.186</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parametric</td>
<td>0.003</td>
<td>0.136</td>
<td>0.136</td>
<td>0.130</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.003</td>
<td>0.114</td>
<td>0.114</td>
<td>0.118</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.000</td>
<td>0.537</td>
<td>0.537</td>
<td>0.508</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parametric</td>
<td>−0.003</td>
<td>0.073</td>
<td>0.073</td>
<td>0.074</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>0.003</td>
<td>0.062</td>
<td>0.062</td>
<td>0.064</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.021</td>
<td>0.248</td>
<td>0.248</td>
<td>0.242</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parametric</td>
<td>0.001</td>
<td>0.040</td>
<td>0.040</td>
<td>0.041</td>
<td>0.96</td>
</tr>
<tr>
<td>$S_0$</td>
<td>0.9</td>
<td>50</td>
<td>True ETM</td>
<td>0.001</td>
<td>0.035</td>
<td>0.035</td>
<td>0.035</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.001</td>
<td>0.041</td>
<td>0.041</td>
<td>0.041</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.000</td>
<td>0.019</td>
<td>0.020</td>
<td>0.019</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.000</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.000</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.000</td>
<td>0.013</td>
<td>0.012</td>
<td>0.013</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>0.002</td>
<td>0.068</td>
<td>0.069</td>
<td>0.068</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.003</td>
<td>0.070</td>
<td>0.070</td>
<td>0.070</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>0.000</td>
<td>0.038</td>
<td>0.039</td>
<td>0.038</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.000</td>
<td>0.038</td>
<td>0.040</td>
<td>0.038</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.002</td>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.001</td>
<td>0.021</td>
<td>0.022</td>
<td>0.022</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.001</td>
<td>0.045</td>
<td>0.053</td>
<td>0.045</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.001</td>
<td>0.046</td>
<td>0.054</td>
<td>0.046</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>0.001</td>
<td>0.025</td>
<td>0.026</td>
<td>0.025</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>0.001</td>
<td>0.026</td>
<td>0.027</td>
<td>0.026</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True ETM</td>
<td>−0.001</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>General ETM</td>
<td>−0.001</td>
<td>0.014</td>
<td>0.015</td>
<td>0.014</td>
<td>0.96</td>
</tr>
</tbody>
</table>
500 and simulated 1000 independent data sets under each setting. The power of testing for no difference between the two groups was calculated based on the true ETM, the general ETM and the PHM. Results are displayed in Fig. 2.

In each of the three situations, the power curve for the general ETM is always lower than that based on the true ETM, the cost of obtaining robustness. But, the difference between them is relatively small and reduces as sample size increases from 300 to 1000. The power curve for the PHM is close to that for the true ETM, when the PHM assumption is satisfied (situation a and c). However, when the PHM assumption is not satisfied (situation b), the PHM loses power relative to the ETM. As shown in the two panels in the second row of Fig. 2, the power curve for the PHM increases very slowly. For example, when the survival times of the treatment group were generated from Gamma(1.1, 1.9) and the sample sizes were $N_0=N_1=500$, the power for the PHM is only 0.24 compared to 0.94 for the general ETM. On the other hand, when the general ETM assumption is not satisfied (situation c), the power curve for the general ETM is still reasonably close to that for the PHM, indicating the general ETM is relatively robust in this situation.

5. The gastrointestinal tumor study

We apply ETMs to the survival data of locally unresectable gastric cancer studied by Stablein and Koutrouvelis (1985). The trial compared chemotherapy and combined chemotherapy and radiotherapy and had 45 patients on each treatment.
The survival times were recorded in days and there were six censored observations. Fig. 3 shows the K–M estimates of survival curves. The crossing of survival curves indicates that the proportional hazards assumption is not valid.

We consider two ETMs: one is the simplest ETM having the $h$ function associated with the exponential, $h(t) = \beta_1 t$, and the second is the general ETM using the general $h$ function (2). The estimated model parameters are listed in Table 4. The likelihood-ratio test comparing these two ETMs is highly significant ($p$-value 0.001).

The estimated survival curves based on these two ETMs and the PHM are shown in Fig. 3. Both ETMs capture the observed pattern of crossing survival curves and the general ETM appears to fit the data better. The estimated curves based on the PHM show poor fitting, because the PHM does not accommodate crossings. We also compare the estimated survival odds ratio function based on the exponential ETM, the general ETM, the PHM, and the K–M. The results are shown in Fig. 4. The estimate based on the general ETM (orange curve) is very close to that based on the K–M (black curve) and is much smoother. The estimate based on the exponential ETM (green curve) is close to the K–M estimate after the first year but deviates from the K–M in the first year. The estimate based on the PHM (blue curve) deviates considerably from the K–M.

![Fig. 2. Power comparison. The survival times in the placebo group were generated from $\text{Exp}(1.5)$ (situation a), $\Gamma(1.5,1.5)$ (situation b), and Weibull$(2,3)$ (situation c). Those in the treatment group were generated from $\text{Exp}(1.5+\beta)$, $\Gamma(1.5+\beta,1.5+\beta)$, and Weibull$(2.3+\beta)$, where $\beta$ changed from 0.0 to 0.4.](image-url)
To test whether there is significant difference in survival time between the two groups, we perform the likelihood-ratio test for $\beta = 0$. Both the exponential ETM and the general ETM provide significant results ($p$-values 0.017 and 0.001, respectively). But for the PHM, the test is not significant ($p$-value 0.638).

### 6. Conclusions

In this paper, we studied the application of the Exponential Tilt Model (ETM) in survival analysis. We extended the theoretical results for the ETM to handle censored data. The connection between the ETM and logistic regression was also emphasized. Based on simulation studies, we showed that the ETM has a better performance than the PHM in both estimation and hypothesis testing problems when the model assumption for the PHM is not satisfied but the model assumption for the ETM is satisfied. On the other hand, the ETM is relatively robust when its model assumption is not satisfied.

The ETM has several desirable features. First, it has clear connection to parametric models. The general ETM includes a broad array of parametric models. Second, it allows crossing in both hazard and survival functions. Third, model coefficients in the ETM have clear interpretation because of the nice connection of the ETM to logistic regression models.

The general ETM can be expanded to include even more types of parametric models. For example, it can include the Weibull model by adding a power function of survival time as a basis in the $h^*$ function. But this approach is not pursued in this paper because of two reasons: First, adding a power term of survival time makes $h^*$ complicated and may cause the unsteadiness in parameter estimation; Second, based on our simulation studies, the current $h^*$ function is robust for data simulated from Weibull distribution.

---

**Table 4**

Model coefficients estimates based on the exponential ETM and the general ETM for the gastrointestinal tumor study.

<table>
<thead>
<tr>
<th></th>
<th>Exponential ETM</th>
<th>General ETM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-0.457</td>
<td>1.359</td>
</tr>
<tr>
<td>SE</td>
<td>0.215</td>
<td>0.628</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_2$</th>
<th>$\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-2.752</td>
<td>-0.509</td>
</tr>
<tr>
<td>SE</td>
<td>0.945</td>
<td>0.235</td>
</tr>
</tbody>
</table>

Note: SE is the standard error estimate.
For clarity and focus, we have considered the two-group comparison with no covariate adjustment. Generalization of the ETM to include discrete covariates with a finite range is theoretically feasible. One can divide the data into several subgroups, analogous to the treatment and control groups in the basic, two-group comparison model, compute within subgroup comparisons and combine over subgroups. In this situation, the theoretical argumentation and practical implementation will be quite similar to the two-group comparison problem, though the formulas will become more complicated. More challenging will be to include continuous covariates.

A rigorous investigation of statistical properties in the context of continuous censoring times is challenging. We conjecture results similar to those for the discrete case will hold. The conjecture is supported by simulation results in Section 4, but a more sophisticated mathematical approach (possibly coupled with compactness or other assumptions) will be needed. This situation is primarily of mathematical interest, since in practice censoring times can be approximated by a sufficiently fine grained set of discrete points.

Acknowledgements

The authors wish to thank a reviewer for helpful comments that have greatly improved the article. This research was supported by the U.S. National Science Foundation for Zhiqiang Tan.

Appendix A. EM algorithm for estimating $p_i$

Consider $Y_i = \min(T_i, \tau), i = 1, \ldots, N$ as the complete data, and assume censoring occurs at $\{c_{z1}, \ldots, c_{zd}\}, z = 0, 1$.

$E$ Step: for a censored observation from the placebo group, say $Y_{mis}$ censored at time $c_{0i}$, its density function given the current parameter values $p_i^{(k)}, s_0^{(k)}$ at the $k$th step is

$$P(Y_{mis} = x_i | s_0^{(k)}, \phi_0^{(k)}, x_i, u = 1, \ldots, n) = \frac{Q_{0i}^{(k)}}{\sum_{u = 1}^{n} Q_{0i}^{(k)} + s_0^{(k)}}.$$ 

$$P(Y_{mis} = z | s_0^{(k)}, \phi_0^{(k)}, x_i, u = 1, \ldots, n) = \frac{s_0^{(k)}}{\sum_{u = 1}^{n} Q_{0i}^{(k)} + s_0^{(k)}}.$$ 

**Fig. 4.** Estimation of the survival odds ratio function for the gastrointestinal tumor study. The 95% confidence limits for the PHM are the two dashed blue curves on each side of the estimated curves. The plot is in the log scale. The time interval is from zero to the largest uncensored failure time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
where $Q_{ij}^{(k)} = p_{ij}^{(k)} I(x_i > c_{ij})$. Similarly, for a censored observation $Y_{mis}$ from the treatment group censored at time $c_{ij}$, given the current parameter estimates $p_{ij}^{(k)}$, $s_{1i}^{(k)}$, at the $k$th step,

$$ P(Y_{mis} = x_i | s_{1i}^{(k)}, p_{ij}^{(k)}, x_{ui}, u = 1, \ldots, n) = \frac{Q_{ij}^{(k)}}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}}, $$

$$ P(Y_{mis} = s_{1i}^{(k)} | p_{ij}^{(k)}, x_{ui}, u = 1, \ldots, n) = \frac{s_{1i}^{(k)}}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}}, $$

where $Q_{ij}^{(k)} = P_{ij}^{(k)}(1 - s_{ij}^{(k)} \exp[h(x_i, \beta)] I(x_i > c_{ij}) / \sum_{n} P_{ij}^{(k)} \exp[h(x_i, \beta)])$.

The expectation of the complete data log-likelihood over the censored observations given $p_{ij}^{(k)}$ is

$$ \mathcal{E}^{(k+1)} = E_{(Y_{mis})} \log \mathcal{L}_{p_{ij}^{(k)}, x_{ui}, i = 1, \ldots, n} $$

$$ = \sum_{i=1}^{n} \log p_{ij} + \sum_{i=1}^{n} \left[ \log \left(\frac{n}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}} \right) \right] $$

$$ + \sum_{j=1}^{d_0} \sum_{u=1}^{m_0} \left[ \log \left(\frac{n}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}} \right) \right] $$

$$ + \sum_{j=1}^{d_1} \sum_{u=1}^{m_1} \left[ \log \left(\frac{n}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}} \right) \right] $$

$$ + \frac{m_{bd_0} + 1}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}} + \frac{m_{bd_1} + 1}{\sum_{n} Q_{ij}^{(k)} + s_{1i}^{(k)}}. $$

**M Step:** maximize $\mathcal{E}^{(k+1)}$ under the constraint:

$$ \sum_{i=1}^{n} p_{ij} = 1. $$

Consider the lagrange multiplier:

$$ l_{m}^{(k+1)} = \mathcal{E}^{(k+1)} - \lambda \left( \sum_{i=1}^{n} p_{ij} - 1 \right). $$

We can get

$$ \lambda = n_0 + m_0, $$

and the parameter estimates at step $k+1$ satisfy

$$ p_{ij}^{(k+1)} = \frac{1 + \sum_{j=1}^{d_0} \sum_{u=1}^{m_0} Q_{ij}^{(k)} + \sum_{j=1}^{d_1} \sum_{u=1}^{m_1} Q_{ij}^{(k)} + s_{1i}^{(k)}}{n_0 + m_0 + \left( n_1 + \sum_{j=1}^{d_1} m_1 + \sum_{j=1}^{d_0} m_0 + \sum_{j=1}^{n} Q_{ij}^{(k)} + s_{1i}^{(k)} \right) \exp[h(x_i, \beta)]}, $$

$$ s_{0i}^{(k+1)} = \frac{\sum_{j=1}^{d_0} m_0 (Q_{ij}^{(k)} + s_{0i}^{(k)} + m_{bd_0} + 1)}{n_0 + m_0}, $$

$$ s_{1i}^{(k+1)} = \frac{\sum_{j=1}^{d_1} m_1 (Q_{ij}^{(k)} + s_{1i}^{(k)} + m_{bd_1} + 1)}{n_1 + m_1}. $$

**Appendix B. Proofs**

**Proof of Proposition 1.** (i) We treat $\beta$ and $\eta$ as free parameters. Plugging (5) into (4) and define

$$ l_n(\beta, \eta) = -n \log N - \sum_{i=1}^{n} \log \left( \omega_0 (x_i, Q_0) + \omega_1 (x_i, Q_1) \exp[z + h(x_i, \beta)] \right) + \sum_{i=n_0+1}^{n} h(x_i, \beta) + n_1 z + \sum_{j=1}^{d_0} m_0 \log Q_{ij} + \sum_{j=1}^{d_1} m_1 \log Q_{ij} $$

$$ + m_{bd_0} + 1 \log s_0 + m_{bd_1} + 1 \log s_1. $$
The derivatives of \( l_n \) are

\[
\begin{align*}
\frac{\partial l_n}{\partial \beta} &= -\frac{1}{N} \sum_{i=1}^{N} [N_p \hat{\rho}_i (1-z_i) \xi_0 (x_i, Q_0) + z_i \xi_1 (x_i, Q_1) \exp(x + h(x_i, \beta))] h'(x_i, \beta), \\
\frac{\partial l_n}{\partial Q_0} &= -\frac{1}{N} \sum_{i=1}^{N} \left[ N_p \hat{\rho}_i \left\{ \frac{s_0}{1 - s_0} \frac{m_{o_j}}{NQ_{0j}} + \frac{m_{o_j}}{NQ_{0j}} \right\} \right], \quad j = 1, \ldots, d_0, \\
\frac{\partial l_n}{\partial S_0} &= -\frac{1}{N} \sum_{i=1}^{N} \left[ N_p \hat{\rho}_i \left\{ \frac{\rho_0 - \sum_{j=1}^{d_0} \frac{m_{o_j}}{NQ_{0j}} - \frac{m_{o_{d_0}}}{N} + \frac{m_{o_{d_0}}}{N} \right\} \right], \\
\frac{\partial l_n}{\partial Q_1} &= -\frac{1}{N} \sum_{i=1}^{N} \left[ N_p \hat{\rho}_i \left\{ \frac{s_1}{1 - s_1} \frac{m_{i_j}}{NQ_{1j}} + \frac{m_{i_j}}{NQ_{1j}} \right\} \exp(x + h(x_i, \beta)) + \frac{m_{i_1}}{NQ_{1j}} \right], \quad j = 1, \ldots, d_1, \\
\frac{\partial l_n}{\partial S_1} &= -\frac{1}{N} \sum_{i=1}^{N} \left[ N_p \hat{\rho}_i \left\{ \frac{\rho_1 - \sum_{j=1}^{d_1} \frac{m_{i_j}}{NQ_{1j}} - \frac{m_{i_{d_1}}}{N} + \frac{m_{i_{d_1}}}{N} \right\} \exp(x + h(x_i, \beta)) + \frac{m_{i_{d_1}}}{NQ_{1j}} \right], \\
\frac{\partial l_n}{\partial \varepsilon} &= -\frac{1}{N} \sum_{i=1}^{N} [N_p \hat{\rho}_i (1-z_i) \xi_0 (x_i, Q_0) + z_i \xi_1 (x_i, Q_1) \exp(x + h(x_i, \beta))].
\end{align*}
\]

For fixed \( \beta \), maximizing \( l(\beta, F_0) \) over \( p, s_0, \) and \( s_1 \) is equivalent to solving \( \frac{\partial l_n(\beta, \eta)}{\partial \eta} = 0 \). Thus, the profile log-likelihood of \( \beta \) is

\[
pl_n(\beta) = l_n(\beta, \tilde{\eta}(\beta)),
\]

where \( \tilde{\eta}(\beta) \) is the solution of \( \frac{\partial l_n(\beta, \eta)}{\partial \eta} = 0 \). Therefore, \( \tilde{\beta} \) satisfies 0 = \( \frac{\partial pl_n(\beta)}{\partial \beta} \) if and only if \((\tilde{\beta}, \tilde{\eta}(\tilde{\beta}))\) satisfies 0 = \( \frac{\partial l_n}{\partial \beta} \) and 0 = \( \frac{\partial l_n}{\partial \eta} \). Since \( m_{o_j}/N \) and \( m_{i_j}/N \) converge to constants and the individual terms in \( \frac{\partial l_n}{\partial \beta} \) and \( \frac{\partial l_n}{\partial \eta} \) are uniformly bounded by some linear function of \( \kappa(x) \) in a neighborhood of the true values of \((\beta, \eta)\), \((\tilde{\beta}, \tilde{\eta}(\tilde{\beta}))\) are consistent estimators based on the asymptotic theory of M-estimators.

(ii) The second-order derivatives of \( l_n(\beta, \eta) \) are uniformly bounded by some linear function of \( \kappa(x) \) in a neighborhood of the true values of \((\beta, \eta)\). Based on the asymptotic theory of M-estimators, \((\tilde{\beta}, \tilde{\eta})\) are asymptotically normal. Next, we show the asymptotic variance of \( \tilde{\beta} \) equal to \( U^{-1} \) in four steps. We need the following notations. For \( z = 0, 1 \), define \( \gamma_{2j}(X, A, Z) = (I(X = c_0, A = 0, Z = z)/Q_{z0}, \ldots, I(X = c_{d_0+1}, A = 0, Z = z)/Q_{z_{d_0+1}}) \) and \( \gamma_2 = \sum_{j=1}^{N} \gamma_{2j}(X_j, A_j, Z_j)/N = (m_{z0}/Q_{z0}, \ldots, m_{z_{d_0}}/Q_{z_{d_0+1}})^T \). Evaluated at the true values of \( \xi_{0j}, \xi_{1j}, 1 \leq j \leq d_0, S_0 = (c_0), \gamma_{2j} \gamma_{2j}^T = \gamma_2^2 \). Let \( \gamma = (\gamma_2^T, \gamma_2^T) \) and \( \gamma = (\gamma_2^T, \gamma_2^T) \). Let \( q = \rho_0 - \gamma_{20}^T Z_0 + (\rho_1 - \gamma_{21}^T I_1)e^{x+h} \), where \( I_2 = (I(X > c_0), \ldots, I(X > c_{d_0+1})) \) and \( h \) is the abbreviation of \( h(X, \beta) \). Let \( d = d_0 + d_1 \). For a function \( g(X) \), we will use \( E[g(X)] \) to denote \( \int_0^2 g(x) dF_0(x) \).

1. By direct calculation,

\[
\frac{1}{N} \begin{pmatrix}
\frac{\partial^2 l_n}{\partial \beta^2} & \frac{\partial^2 l_n}{\partial \beta \partial \eta^T} \\
\frac{\partial^2 l_n}{\partial \beta \partial \eta^T} & \frac{\partial^2 l_n}{\partial \eta^2}
\end{pmatrix}
= \begin{pmatrix}
-B & DA \\
AD^T & A + ARA
\end{pmatrix}
= H,
\]

where \( A \) is a \((d+3) \times (d+3)\) diagonal matrix with the diagonal elements \((\gamma_{00}/Q_{00}, \ldots, \gamma_0/d_0+1/Q_{d_0+1}; \gamma_{10}/Q_{10}, \ldots, \gamma_{1d_1+1}/Q_{2d_1+1}, 1)\) and

\[
B = E(\nu \eta \eta^T)_{r \times r},
\]

\[
D = E \left( \psi_0 h \left( \frac{s_0}{1 - s_0} + I_0 \right) \right) - E \left( \psi_1 h \left( \frac{s_1}{1 - s_1} + I_1 \right) \right) - E(\nu \eta^T)_{r \times (d+3)}.
\]
Based on central limit theorem, it can be shown that

\[
\sqrt{N} \begin{pmatrix} \frac{\partial l_n}{\partial \beta} \\ \frac{\partial l_n}{\partial \eta} \end{pmatrix} \rightarrow N(0, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}),
\]

where

\[
\Sigma_{11} = \begin{pmatrix} B & 0 \\ 0 & A(R - Q \otimes^2 A) \end{pmatrix} + P + P^T - \frac{P_0}{\rho_1} \mu \mu^T - J(0) A - \begin{pmatrix} 0 \\ A \end{pmatrix}^T,
\]

\[
\Sigma_{12} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \frac{P_0}{\rho_1} \mu \zeta^T, \quad \Sigma_{22} = II - \frac{P_0}{\rho_1} \zeta^T \zeta^T,
\]

and

\[
\zeta^* = \begin{pmatrix} -\rho_1 / \rho_0 \times \gamma_0 \\ \gamma_1 \end{pmatrix}, \quad F = \begin{pmatrix} \gamma_0^T \\ 0_{1 \times d} \end{pmatrix}_{(d+3) \times (d+2)}, \quad J = \begin{pmatrix} 0 \\ A M \end{pmatrix},
\]

\[
Q \otimes^2 = \begin{pmatrix} Q_0 Q_0^T \\ Q_0 Q_1^T \\ Q_1 Q_1^T \end{pmatrix}_{(d+3) \times (d+3)}, \quad \mu = \begin{pmatrix} \zeta \end{pmatrix} + \frac{P_0}{\rho_1} \begin{pmatrix} D_2 \\ A R_2 \end{pmatrix},
\]

\[
II = A^* - \gamma_{zz}^T, \quad P = \begin{pmatrix} 0_{r \times (d+r+2)} & -D_2 \\ 0_{(d+3) \times (d+r+2)} & -A R_2 - v \end{pmatrix}_{(r+d+3) \times (r+d+3)},
\]

with \( \zeta = (\zeta^T, 1 / \rho_0)^T, v = (0_{1 \times (d+1)})^T, D_2 \) being the last column of \( D, R_2 \) being the last column of \( R_\), and \( A^* \) being the upper left \((d+2) \times (d+2)\) submatrix of \( A \). Let \( R_1 \) be the first \( d+2 \) columns of \( R, D_1 \) be the first \( d+2 \) columns of \( D_\), and \( M_1 \) be
the first \(d+2\) columns of \(M\). Since

\[
\begin{pmatrix}
\frac{1}{N} \partial^2 L_n^T \\
\frac{1}{N} \partial^2 L_n^T
\end{pmatrix} = \begin{pmatrix}
-D_1 \\
-AR_1
\end{pmatrix} + \begin{pmatrix}
0 \\
AM_1
\end{pmatrix} = K + J_1,
\]

we have \(N^{-1/2}(\partial \hat{L}_n/\partial \beta^T, \partial \hat{L}_n/\partial \eta^T)^T\) converges to a normal distribution with mean zero and variance \(V^+ = V_1^+ + V_2^+\), where

\[
V_1^+ = \begin{pmatrix}
B & 0 \\
0 & A(R - Q \otimes^2 A)
\end{pmatrix} + KIJK + \begin{pmatrix}
0 & -D_1 I^T \\
-1 & -F R_1^T (A - AR_1) I^T - F R_1^T (A - AR_1) I^T
\end{pmatrix},
\]

\[
V_2^+ = -J_0 \left( \begin{array}{c}
A \\
\rho_1
\end{array} \right)^T + J_1 (\begin{array}{c}
1 \\
\rho_1
\end{array}) I^T + \begin{pmatrix}
0 \\
\rho_1
\end{pmatrix} F P + F K I^T
\]

\[
= \frac{\rho_1}{\rho_1} \begin{pmatrix}
(L \zeta)(L \zeta)^T + P + A^T - (A^T - O) Y^T - L(A^T - O) Y^T - (A^T - O) Y^T
\end{pmatrix},
\]

and

\[
L = \begin{pmatrix}
D \\
I + AR
\end{pmatrix},
\]

\[
O = \begin{pmatrix}
1 & \gamma_1 \gamma_0^T & \gamma_1 \\
0 & \gamma_1 \gamma_0^T & \gamma_1 \\
-1 & \gamma_1 \gamma_0^T & \gamma_1
\end{pmatrix},
\]

\[
A_0 = \begin{pmatrix}
A^T & 0 \\
0 & 0
\end{pmatrix}.
\]

3. Based on Taylor expansion,

\[
\frac{1}{N} \partial p_N^T
\]

\[
= \frac{1}{N} \left[ \partial \eta / \partial \beta - \frac{\partial^2 L_n}{\partial \beta^T} \left( \frac{\partial^2 L_n}{\partial \eta^T} \right) \right] + o_p(N^{-1/2}).
\]

So \(N^{-1/2}(\partial \hat{p}_n/\partial \beta)\) converges to a normal distribution with mean zero and variance

\[
V = (L, -DA(A + AR)^{-1}) V^+ \begin{pmatrix}
L \\
-(A + AR)^{-1} AD^T
\end{pmatrix},
\]

where \(L\) is the \(r \times r\) identity matrix.

For the \(V_2^+\) part in Eq. (8), one can verify that

\[
(L, -DA(A + AR)^{-1}) L = 0_{r \times (d+3)},
\]

\[
(L, -DA(A + AR)^{-1}) P = 0_{r \times (d+r+3)},
\]

\[
(L, -DA(A + AR)^{-1}) J = 0_{r \times (d+3)}.
\]

To show the last equation in (9) holds, let

\[
\begin{pmatrix}
0_{d_0 \times d_0} & -1_{a_0 b_0} Q_0^* & 0_{d_0 \times d_1} & 0_{d_0 \times 1} \\
1_{a_0 b_0} & 0_{d_0 \times d_1} & 0_{d_0 \times 1} & 0
\end{pmatrix}
\]

\[
A = \begin{pmatrix}
0_{d_1 \times d_0} & 0_{d_1 \times d_1} & 0_{d_1 \times 1} & -1_{a_1 b_1} Q_1^* \\
1_{a_1 b_1} & 0_{d_1 \times d_1} & 0_{d_1 \times 1} & 0
\end{pmatrix},
\]

where \(Q_0^*\) is a vector of the first \(d_0\) elements of \(Q_0\), \(Q_1^*\) is a vector of the first \(d_1\) elements of \(Q_1\), \(a_0 = \gamma_0 d_0 + 1/Q_0 d_0 + 1\), \(a_1 = \gamma_1 d_1 + 1/Q_1 d_1 + 1\), \(b_0 = (\rho_0(1 - s_0) + s_0 \sum_{j=1}^{d_0} \gamma_j)/\gamma_0 d_0 - s_0\), \(b_1 = (\rho_1(1 - s_1) + s_1 \sum_{j=1}^{d_1} \gamma_j)/\gamma_1 d_1 - s_1\). It
can be verified that 
\[(I_r - DA(A + ARA)^{-1}A)M = -DA(A + ARA)^{-1}A = 0.\]

Based on (9), we have 
\[(I_r - DA(A + ARA)^{-1}A)^{+}V_2^+ \begin{pmatrix} I_r \\ -(A + ARA)^{-1}AD^T \end{pmatrix} = 0_{r \times r}.\]

For the \(V_1^+\) part in Eq. (8), it can be shown that 
\[V_1^+ = \begin{pmatrix} B & 0 \\ 0 & A(RQ^{-2})A \end{pmatrix} + \begin{pmatrix} -D \\ -AD \end{pmatrix}I_r(-D^T - RA) + \begin{pmatrix} 0 & -D\Gamma_e^T \\ -\Gamma_eD^T & -\Gamma_eR^T(A - AR\Gamma_e^T), \end{pmatrix}\]

where 
\[\Pi_e = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}_{(d+3) \times (d+3)}, \Gamma_e = (\Gamma^0)_{(d+3) \times (d+3)}.\]

Since 
\[\Pi_e + 2\Gamma_e - AQ^{-2}A = A,\]

based on the same derivation as in Tan (2010), 
\[(I_r - DA(A + ARA)^{-1}A)^{+}V_1^+ \begin{pmatrix} I_r \\ -(A + ARA)^{-1}AD^T \end{pmatrix} \]
\[= B + D(I_{d+3} + AR)^{-1}A(RQ^{-2})(I_{d+3} + RA)^{-1}D^T + D(I_{d+3} + AR)^{-1}(\Pi_e + 2\Gamma_e)(I_{d+3} + RA)^{-1}D^T \]
\[= B + D(I_{d+3} + AR)^{-1}D^T = U,\]

where \(I_{d+3}\) is the \((d+3) \times (d+3)\) identity matrix. Therefore, 
\[V = U.\]

4. Based on Taylor expansion, 
\[\hat{\beta} - \beta = - \left( \frac{\partial^2 p_{N}(\beta)}{\partial^2 \beta} \right)^{-1} \frac{\partial p_{N}}{\partial \beta} + o_p(N^{-1/2}).\]

So \(\sqrt{N}(\hat{\beta} - \beta)\) is asymptotically normal with mean zero and variance 
\[U^{-1}VU^{-1} = U^{-1}.\]

**Proof of Proposition 2.** A second-order Taylor expansion of \(p_{N}(\hat{\beta})\) yields 
\[-2p_{N}(\beta) - p_{N}(\hat{\beta}) = (\hat{\beta} - \beta)^T \left( \frac{\partial^2 p_{N}(\beta)}{\partial^2 \beta} \right) (\hat{\beta} - \beta) + o_p(1).\]

Proposition 1 produces the result.

**Proof of Proposition 3.** Let \(\theta = (\beta^T, \eta^T)^T\), based on Taylor expansion, 
\[\begin{pmatrix} \tilde{S}_0(t_0) \\ \tilde{S}_1(t_0) \end{pmatrix} - \begin{pmatrix} S_0(t_0) \\ S_1(t_0) \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} G(X_i, t_0) - \left( \frac{1}{N} \sum_{i=1}^{N} \frac{\partial G(X_i, t_0)}{\partial \theta} \right) (\hat{\theta} - \theta) + o_p(N^{-1/2}),\]

where 
\[G(X_i, t_0) = \left( N p_i I(X_i > t_0) I_1 + s_0 \right) - \left( \begin{pmatrix} S_0(t_0) \\ S_1(t_0) \end{pmatrix} \right).\]

Since 
\[(\hat{\theta} - \theta) = \left( \frac{1}{N} \frac{\partial^2 \ln}{\partial \theta^2} \right)^{-1} \frac{1}{N} \frac{\partial \ln}{\partial \theta} + o_p(N^{-1/2}),\]

we have 
\[\begin{pmatrix} \tilde{S}_0(t_0) \\ \tilde{S}_1(t_0) \end{pmatrix} - \begin{pmatrix} S_0(t_0) \\ S_1(t_0) \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} G(X_i, t_0) - \left( \frac{1}{N} \sum_{i=1}^{N} \frac{\partial G(X_i, t_0)}{\partial \theta} \right) \left( \frac{1}{N} \frac{\partial^2 \ln}{\partial \theta^2} \right)^{-1} \frac{1}{N} \frac{\partial \ln}{\partial \theta} + o_p(N^{-1/2}).\]
From Taylor expansion,
\[
\frac{1}{N} \sum_{i=1}^{N} \frac{G(X_i, t_0)}{\partial x} \bigg|_{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial G(X_i, t_0)}{\partial x} \bigg|_{\gamma} + \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2 G(X_i, t_0)}{\partial x^2} \bigg|_{\gamma} (\gamma - \gamma) + o_p(N^{-1/2}).
\]
\[
\frac{1}{N} \frac{\partial G}{\partial t} \bigg|_{\gamma} = \frac{1}{N} \frac{\partial G}{\partial t} \bigg|_{\gamma} + \frac{1}{N} \frac{\partial^2 G}{\partial x \partial t} \bigg|_{\gamma} (\gamma - \gamma) + o_p(N^{-1/2}).
\]
So
\[
\begin{pmatrix}
\hat{S}_0(t_0) \\
\hat{S}_1(t_0)
\end{pmatrix} = \begin{pmatrix}
S_0(t_0) \\
S_1(t_0)
\end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} \varphi(t_0, X_i, A_i, Z_i, \beta, \eta) + o_p(N^{-1/2}),
\]
where
\[
\varphi(t_0, X_i, A_i, Z_i, \beta, \eta) = \frac{1}{N} W_1 W_2 \sum_{i=1}^{N} \bar{\zeta}(t_0, X_i, A_i, Z_i, \beta, \eta),
\]
\[
W_1 = \begin{pmatrix}
I_2 & -E \left( \frac{\partial^2 G(X_i, t_0)}{\partial x} \bigg|_{\gamma} g \right) H^{-1}
\end{pmatrix},
\]
\[
W_2 = \begin{pmatrix}
I_2 & 0 & E \left( \frac{\partial^2 G(X_i, t_0)}{\partial x} \bigg|_{\gamma} \right) q
0 & I_{d+r+3} & K
\end{pmatrix},
\]
\[
\bar{\zeta}(t_0, X, A, Z, \beta, \eta) = (\bar{\zeta}_1, \bar{\zeta}_2, \bar{\zeta}_3)^T,
\]
with \( \bar{\zeta}_1 = G(X, t_0) \), \( \bar{\zeta}_3 = \gamma(X, A, Z) - \gamma \), and \( \bar{\zeta}_2 \) is the individual term of the derivative of \( l_n \) with respect to \( \theta \), \( \bar{\zeta}_2 = \left( (1 + (p_2 - \gamma^2_2 I_2)e^{\alpha + hZ}/q(h)^T A)^T, \right. \)
\[
\left. \{y_0 - (1 - s_0) + t_0 A/q, Q_0 \}, \{y_1 - (1 - s_1) + t_1 A/q, Q_1 \}, \{1 - (p_2 - \gamma^2_2 I_2)e^{\alpha + hZ}/q, A \}^T \right). \]

**Remark 1.** In practice, \( \varphi \) can be estimated by \( \hat{\varphi} \), wherein integrals in \( W_1 \) and \( W_2 \) are replaced by sample averages, and \( (\beta, \eta) \) are replaced by \( (\hat{\beta}, \hat{\eta}) \).

References:


