IMPROVED M-ESTIMATES IN CONVEX MINIMIZATION – AN EASY EMPIRICAL LIKELIHOOD APPROACH

By Fei Tan and Hanxiang Peng

Indiana University Purdue University Indianapolis

Let $m(z, \vartheta)$ be a criterion function convex in parameter ϑ for every z. For a random sample $Z_1, \dots, Z_n$, the M-estimate $\tilde{\vartheta}$ of ϑ minimizes the criterion function $\sum_{j=1}^n n^{-1} m(Z_j, \vartheta)$. Suppose side information is available given by $E(u(Z_1)) = 0$ for some square-integrable function u. In this article, we are concerned with the use of side information and propose to estimate ϑ by $\hat{\vartheta}$ which minimizes the criterion function $\sum_{j=1}^n \pi_{nj} m(Z_j, \vartheta)$ with $\pi_{nj} = n^{-1}(1 + \zeta_n^\top u(Z_j))^{-1}$ for some random variable ζ_n determined by $u(Z_j)$'s. We show $\hat{\vartheta}$ is asymptotically normal and more efficient than $\tilde{\vartheta}$. As applications of the results, we construct efficient estimates of quantitles, parameters in quantitle regression and in the Cox proportational hazard (PH) regression. A simulation study and real data application are performed to illustrate the use of side information in the Cox PH model to improve the efficiency of maximum partial likelihood estimates.

1. Introduction. Owen [9, 10] introduced empirical likelihood to construct confidence intervals in a nonparametric setting. Soon Qin and Lawless [14] used it to construct point estimates and studied maximum empirical likelihood estimates (MELEs). They proved many properties for MELEs such as MELEs are fully semiparametrically efficient in the sense of least dispersed regular estimators [1, 15]. The empirical likelihood approach is particularly convenient to incorporate side information. Just like parametric maximum likelihood estimates, nevertheless, MELEs involve highly nonlinear equations. Thus it is not a trivial task to find MELEs. Peng and Schick [12] explored MELEs in the case of constraint functions that may be discontinuous and/or depend on additional parameters and employed one-step estimates to construct MELEs. Peng [11] has identified a class of easy maximum empirical likelihood estimators, while the idea for determining the class was in fact already used by Zhang [17, 18] in M-estimation and quantile processes in the presence of auxiliary information. Hellerstein and Imbens [5] utilized this idea for the least squares estimators in a linear regression model and applied the results to analyze a real data. Relatively recently, Yuan *et al.* [16]

AMS 2000 subject classifications: Primary 62G05; secondary 62G10,62G20

Keywords and phrases: Cox hazards regression, depth function, maximum empirical likelihood estimator, quantile regression, semiparametric efficiency, side information

explored this idea in U-statistics with side information. Tang and Leng [3] utilized the idea to construct improved estimators of parameters in quantile regression. Bravo [2] introduced a class of M-estimators based on generalized empirical likelihood estimation (empirical likelihood is a special case) with side information and showed that the resulting class of estimators is efficient in the sense that it achieves the same asymptotic lower bound as that of the efficient GMM estimator with the same side information. These authors assumed that the available side information can be expressed in a finite number of expectation equations and does not depend on the parameters of interest. Under this setting Peng [11] demonstrated that these estimates are the MELEs which are semiparametrically efficient and mathematically simpler and computationally faster than the usual MELEs.

Let us now briefly detail the easy MELE. Let $(\mathscr{Z}, \mathscr{S})$ be a measurable space, $\mathscr{Q}$ be a family of probability measures on $\mathscr{S}$, and κ be a functional from $\mathscr{Q}$ onto an open subset Θ of $\mathbb{R}^k$. Let $Z_1, \dots, Z_n$ be independent and identically distributed (i.i.d.) copies of Z taking value in $\mathscr{Z}$ with an unknown distribution Q belonging to the model $\mathscr{Q}$. We are interested in statistical inference about the characteristic $\theta = \kappa(Q)$ when side information is available.

Suppose $w(z, \vartheta)$ is a measurable function such that $\int w(z, \vartheta)\, dQ(z) = 0$ for every $\vartheta \in \Theta$. To construct a confidence set for θ, Owen confronted the maximization problem:

$$\mathscr{R}_n(\vartheta) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n, \sum_{j=1}^n \pi_j w(Z_j, \vartheta) = 0 \Big\}, \quad \vartheta \in \Theta,$$

where $\mathscr{P}_n$ denotes the closed probability simplex in dimension n, i.e.,

$$\mathscr{P}_n = \big\{ \pi = (\pi_1, \dots, \pi_n)^\top \in [0,1]^n : \sum_{i=1}^n \pi_i = 1 \big\}.$$

Qin and Lawless [14] tackled point estimation for θ and studied the maximum empirical likelihood estimator:

$$\hat{\theta}_n = \arg\max_{\vartheta \in \Theta} \mathscr{R}_n(\vartheta). \tag{1.1}$$

Consider now $w(z, \vartheta) = (v^\top(z, \vartheta), u^\top(z))^\top, z \in \mathscr{Z}, \vartheta \in \Theta$. Suppose side information is available given by

(K) There is a measurable function $u : \mathscr{Z} \to \mathbb{R}^d$ such that $\int u\, dQ = 0$ and $W_u = \int u^{\otimes 2}\, dQ$ is positive definite.

Under (K), it is natural to look at the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j u(Z_j) = 0 \Big\}.$$

Following Owen, one uses Lagrange multipliers to derive the solution

$$\pi_{nj} = \frac{1}{n} \frac{1}{1+\zeta_n^\top u(Z_j)}, \quad j = 1, \ldots, n, \tag{1.2}$$

where ζ_n satisfies the equation

$$\frac{1}{n} \sum_{j=1}^{n} \frac{u(Z_j)}{1+\zeta^\top u(Z_j)} = 0. \tag{1.3}$$

Now the easy MELE $\hat{\vartheta}$ of θ studied by [11] is any solution to the equation

$$\sum_{j=1}^{n} \pi_{nj} v(Z_j, \vartheta) = \frac{1}{n} \sum_{j=1}^{n} \frac{v(Z_j, \vartheta)}{1+\zeta_n^\top u(Z_j)} = 0. \tag{1.4}$$

This is the case of an improved estimator of θ of the zero estimator $\tilde{\vartheta}$ to the sample equation

$$\frac{1}{n} \sum_{j=1}^{n} v(Z_j, \vartheta) = 0.$$

In this article, we extend the above method from estimating equations to minimization problems. Let $m : \mathscr{Z} \times \Theta \to \mathbb{R}^d$ be a measurable function such that $\int m(z, \kappa(R))\, dR(z)$ is finite for every $R \in \mathscr{Q}$. Based on a random sample $Z_1, \ldots, Z_n$ from Q, we are interested in estimating $\theta = \kappa(Q)$. A popular estimator of θ is the M-estimator which minimizes the sample criterion function,

$$M_n(\vartheta) = \frac{1}{n} \sum_{j=1}^{n} m(Z_j, \vartheta), \quad \vartheta \in \Theta.$$

Thus a natural estimator of θ is the M-estimator which minimizes

$$\mathbb{M}_n(\vartheta) = \sum_{j=1}^{n} \pi_{nj} m(Z_j, \vartheta) = \frac{1}{n} \sum_{j=1}^{n} \frac{m(Z_j, \vartheta)}{1+\zeta_n^\top u(Z_j)}, \quad \vartheta \in \Theta,$$

where π_{nj}'s are the EL-weights given in (1.2).

As pointed in [11], easy MELEs are mathematically tractible. Here is another application of the tractibility. Suppose $m(z, \vartheta)$ is *convex* in ϑ. Then we quickly claim that $\mathbb{M}_n(\vartheta)$ is also convex as $\pi_{nj}, j = 1, \dots, n$ are probability weights (at least for large sample size n). An further application of the tractibility is the concavity of $\ell(t, b)$ in the Cox hazard regression model, see (2.8). We can now use those nice properties of estimates defined by convex minimization which are well studied in the literature (see e.g. Hjort and Pollard [6]) to derive the asymptotic behaviors of the estimator defined as the minimizer of $M_n(\vartheta)$.

We shall refer the preceding $\mathbb{M}_n(\vartheta)$ to as the *empirical likelihood (EL)-weighted criterion function*. We shall apply the EL-weight method to derive efficient estimates for quantitles and parameters in quantile regression models when there is side information. We shall also use the method in the Cox proportional hazards regression to improve efficiency. It is well known that the maximum partial likelihood estimator is semiparametrically effcient in the proportional hazards model, see e.g. [1]. However, the result holds under the assumption that only information on time to event (possibly censored) and treatment assignment are available. In clinical-trial data, as remarked in [7], not only are survival and censoring times collected but also side information on variables that may be important prognostic factors which are correlated with time to event. The EL-weight method provides a convenient way to make use of side information to obtain improved estimators of parameters. We have run a small simulation in Section 3 to demonstrate the improvement.

The rest of this paper is organized as follows. In section 2, we show consistency and asymptotic normality for estimators defined by EL-weighted convex minimization. As applications, we derive the MELEs for quantiles and parameters in quantile regression and the Cox PH model. A small simulation and real data application are reported in Section 3. Section 4 contains some of the proofs and a useful theorem.

2. The main results.

In this section, we consider estimators defined by the minimizers of EL-weighted convex criterion functions. We shall use the convexity property to establish the asymptotic properties of the estimators.

As discussed in the Introduction, the EL-weighted version $\mathbb{M}_n(\vartheta)$ is convex (hence continuous). The convexity not only greatly simplifies the theoretical investigation of the estimator, but also reduces the computational burden. Here we present an asymptotic theory in the framework of Theorem 2.2 in [6]. It must be noted that the asymptotic normality results of the EL-weighted estimators hold under similar conditions to those for the asymptotic normality of the usual M-estimators. The proof is delayed to the Appendix.

THEOREM 2.1. *Let $m(z,\vartheta)$ be convex in ϑ. Assume there exists some function D from $\mathscr{Z}$ to $\mathbb{R}^k$ satisfying (K) with $W_u = \int D^{\otimes 2}\,dQ$ such that*

$$m(z,\theta+t) - m(z,\theta) = D^\top(z)t + R(z,t), \quad z \in \mathscr{Z}, t \in \mathbb{R}^k \tag{2.1}$$

for some measurable function $R(z,t)$ with $\operatorname{Var}(R(Z,t)) = o(\|t\|^2)$, and that

$$E\left(m(Z,\theta+t) - m(Z,\theta)\right) = E\left(R(Z,t)\right) = 1/2 t^\top H t + o(\|t\|^2),\ t \to 0 \tag{2.2}$$

for some positive definite matrix H. Then the estimator $\hat{\theta}_n$ which minimizes $\mathbb{M}_n(\vartheta)$ over Θ is $\sqrt{n}$-consistent for θ and satisfies the stochastic expansion,

$$\hat{\theta}_n = \theta - H^{-1}\frac{1}{n}\sum_{j=1}^{n}\big(D(Z_j) - C(D)W_u^{-1}u(Z_j)\big) + o_p(n^{-1/2}).$$

Hence $\sqrt{n}(\hat{\theta}_n - \theta) \implies \mathscr{N}(0,\Sigma)$ where $\Sigma = H^{-1}(K - C(D)W_u^{-1}C(D)^\top)H^{-\top}$.

QUANTILES. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables from a continuous density f positive in its support. The sample p-th quantile q_n is the value which minimizes the criterion function $M_n(\vartheta) = n^{-1}\sum_{j=1}^{n} m_p(Z_j,\vartheta)$, where $m_p(z,t)$ is the popular check function given by

$$m_p(z,t) = p\big((z-t)_+ - z_+\big) + (1-p)\big((t-z)_+ - (-z)_+\big), \quad z,t \in \mathbb{R}, \tag{2.3}$$

where $x_+ = \max(x,0)$ denotes the positive part of x. It is convex in t (hence continuous) and bounded by $|t|$, so that it is always integrable. Its expected value is minimized by $t = F^{-1}(p) := q$, the p-th quantile. One easily verifies

$$E\left(m_p(Z,t) - m_p(Z,q)\right) = 1/2 f(q)(t-q)^2 + o(|t-q|^2);$$

that (2.2) holds with $R(z,t) = (q+t-z)\mathbf{1}[q < z \le q+t]$ such that

$$E(R(Z,t)) = 1/2 t^2 f(q) + o(|t|^2), \quad E(R(Z,t)^2) = o(|t|^2);$$

and that (2.1) holds with

$$D(z) = (1-p)\mathbf{1}[z \le q] - p\mathbf{1}[z > q] = \mathbf{1}[z \le q] - p.$$

Suppose side information is available via $E(u(Z)) = 0$. By Theorem 2.1, the EL-weighted estimator $\hat{q}_n$ which minimizes the EL-weighted criterion function $\mathbb{M}_n(\vartheta) = \sum_{j=1}^{n}\pi_{nj} m_p(Z_j,\vartheta)$ is $\sqrt{n}$-consistent for q and satisfies the stochastic expansion,

$$\hat{q}_n = q - \frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{1}[Z_j \le q] - p - C_{\mathbf{1}}W_u^{-1}u(Z_j)}{f(q)} + o_p(n^{-1/2}),$$

where $C_{\mathbf{1}} = E(\mathbf{1}[Z \leq q]u(Z)^\top)$. Thus $\sqrt{n}(\hat{q}_n - q) \implies \mathscr{N}(0, \sigma^2)$, where $\sigma^2 = (p(1-p) - C_{\mathbf{1}} W_u^{-1} C_{\mathbf{1}}^\top)/f(q)^2$. This result was already obtained in [12, 17, 18]. What we have shown here is that $\hat{q}_n$ is in fact the MELE for q. In particular, if no side information is available, then $\hat{q}_n$ boils down to the usual sample quantile q_n, so the sample quantile is the MELE for the population quantile.

Quantile regression. Bassett and Koenker [**?**] considered the linear quantile regression model in which the response Y and covariate X satisfies

$$F_X^{-1}(p) = \beta^\top X, \tag{2.4}$$

where $p \in (0,1)$, β is a parameter, and F_x^{-1} is the inverse function of the conditional distribution function $F_x(y) = P(Y \leq y | X = x)$ of Y given $X = x$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. copies of $Z := (X, Y)$. When side information is available via $E(u(Z)) = 0$, by the principle of maximum empirical likelihood, the EL-weighted estimator $\hat{\beta}_n$ of β is defined by

$$\hat{\beta}_n = \arg\min_{b \in \mathbb{B}} \sum_{j=1}^{n} \pi_{nj} m_p(Y_j, b^\top X_j), \tag{2.5}$$

where m_p is the convex function given in (2.3) and $\mathbb{B}$ is some compact subset of $\mathbb{R}^k$. Clearly the above sum is convex in b (hence continuous). We will need the following regularity conditions to establish the asymptotic properties of the estimator.

(Q1) The conditional distribution function $F_x(y)$ of Y given $X = x$ is absolutely continuous with continuous density $f_x(y)$ such that it is bounded away from both zero and infinity for almost every $x \in \mathbb{R}^k$.

(Q2) The matrix $E(f_X(q(X))XX^\top)$ is finite and positive definite, where $q(x) = F_x^{-1}(p)$.

By applying Theorem 2.1, we obtain the following asymptotic result with the proof delayed to the Appendix.

Theorem 2.2. *Suppose (Q1)-(Q2) hold. Then the EL-weighted estimator $\hat{\beta}_n$ is $\sqrt{n}$-consistent for β and satisfies the stochastic expansion,*

$$\hat{\beta}_n = \beta - \frac{1}{n} \sum_{j=1}^{n} H^{-1} \left(D(Z_j) - C(D) W_u^{-1} u(Z_j) \right) + o_p(n^{-1/2}),$$

where $D(z) = x(\mathbf{1}[y \leq \beta^\top x] - p)$ and $H = E\left(X^{\otimes 2} f_X(\beta^\top X)\right)$. Thus $\sqrt{n}(\hat{\beta}_n - \beta) \implies \mathscr{N}(0, \Sigma)$, where $\Sigma = H^{-1}\big(K - C(D) W_u^{-1} C(D)^\top\big) H^{-\top}$ with $K = p(1-p)E(X^{\otimes 2})$.

Using Theorem 4.4 of [11], we can show $\hat{\beta}_n$ is the MELE for β as stated below with the proof delayed to the Appendix.

THEOREM 2.3. *Suppose the assumptions in Theorem 2.2 are met. In addition, assume X is bounded and $E\big(\big(X^\top(\mathbf{1}[Y \le \beta^\top X] - p),\, u(Z)^\top\big)^{\otimes 2}\big)$ is positive definite. Then $\hat{\beta}_n$ is the MELE for β in the model specified by the check-function-defined minimization.*

Quantile regression with side information was studied in [3] in a general setup in which the side information u is allowed to contain unknown parameters. There examples were given where side information is expressed via conditional moments, and the use of such information results in more efficient estimators of the parameters. What we have shown here is that the estimator is the MELE for the parameter in the model specified by $E(u(Z)) = 0$ and the check function.

COX REGRESSION. In this model, the hazard rate $h(t)$ for the survival time T of an individual with a p-dimensional covariate process $Z(t) \in \mathbb{Z}$ of time t for some compact $\mathbb{Z}$ is expressed as

$$h(t) = h_0(t)\exp(b^\top Z(t)), \quad t \in [0,\tau], b \in \mathbb{B},$$

where h_0 is an unspecified nonparametric baseline hazard function, $\mathbb{B}$ is some subset of $\mathbb{R}^k$, and τ is finite. Let U be a censoring time of a person. The data can be summarized as n i.i.d. realizations (X_i, δ_i, Z_i) of (X, δ, Z), where $X_i = \min(T_i, U_i)$, representing the observed time of person i; $\delta_i = 1[T_i \le U_i]$, indicating that the observed time is an event time not a censoring time. Let the counting process N_i have mass δ_i at T_i, i.e. $dN_i(t) = \mathbf{1}[T_i \in [t, t+dt], \delta_i = 1]$, and the at-risk process be $Y_i(t) = 1[X_i \ge t\}]$. The at-risk process is left continuous hence predictable. More discussions can be found in Fleming and Harrington [4]. The usual MPLE $\tilde{\beta}_n$ of β is the value which maximizes the log partial likelihood function:

$$l_n(b) = \sum_{j=1}^{n} \int_0^\tau (b^\top Z_j(t) - \log S_n(t,b))\, dN_j(t), \quad b \in \mathbb{B}, \tag{2.6}$$

where $S_n(t,b) = \sum_{j=1}^n Y_j(t)\exp(b^\top Z_j(t))$. Suppose additional information is available about the underlying model via $E(u(R)) = 0$, where $R = (X, \delta, r(Z), U)$ for some measurable function $r(Z)$ of the covariate process $Z(t), t \in [0,\tau]$. Here we shall focus on side information which does not vary with the time t in order to avoid lengthy presentation. Also we allow side information possibly depends on the censoring variable U when its observations are available, see Example 2.1.

By the principle of maximum empirical likelihood, a natural estimator $\hat{\beta}_n$ of β is the value which maximizes the EL-weighted log partial likelihood function:

$$\ell_n(b) = \sum_{j=1}^{n} \pi_{nj} \int_0^{\tau} (b^\top Z_j(t) - \log \mathbb{S}_n(t,b))\, dN_j(t), \quad b \in \mathbb{B}, \tag{2.7}$$

where $\mathbb{S}_n(t,b) = \sum_{j=1}^{n} (n\pi_{nj}) Y_i(t) \exp(b^\top Z_j(t))$ is the EL-weighted version of $S_n(t,b)$. Here the EL-weights π_{nj} are given in (1.2) with $u(Z_j) = u(R_j)$. It is well known that $l_n(b)$ is concave. The proof uses the urn model, see e.g. pages 148 – 151 in [4]. Using the same method, one can show $\ell_n(b)$ is also concave. In fact, similar to the first equation in page 151 one has

$$-\frac{\partial^2 \ell_n(b)}{\partial b \partial b^\top} = \int_0^{\tau} \mathbb{V}(t,b) \sum_{j=1}^{n} \pi_{nj}\, dN_j(t), \tag{2.8}$$

where analoguous to (3.23) in [4] it is easy to prove

$$\mathbb{V}(t,b) = \frac{\sum_{j=1} n\pi_{nj} (Z_j(t) - \mathbb{E}(b,t))^{\otimes 2} Y_j(t) \exp(b^\top Z_j(t))}{\mathbb{S}_n(t,b)}$$

with $\mathbb{E}(t,b) = \sum_{j=1}^{n} n\pi_{nj} Z_j(t) Y_j(t) \exp(b^\top Z_j(t)) / \mathbb{S}_n(t,b)$. This immediately yields the concavity of $\ell_n(b)$ at least for large n. Using the convex argument of [6], we can prove the following Theorem 2.4 with the proof delayed to the Appendix. Formally set $\mathscr{I}(Y,Z) = \int_0^{\tau} \big[Z(t) - e(t)\big] Y(t) e^{\beta^\top Z(t)} h_0(t)\, dt$,

$$\begin{aligned} &C(\mathscr{I}(Y,Z)) = E(\mathscr{I}(Y,Z) u(R)^\top), \\ &s_i(t) = E\big(Z^i(t) Y(t) \exp(\beta^\top Z(t))\big), \quad i = 0,1,2, \quad e = s_1/s_0, \end{aligned} \tag{2.9}$$

where $a^0 = 1$ and $a^2 = aa^\top$ for a vector a.

Theorem 2.4. *Assume $h_0(t)$ is a continuous baseline function. Assume $Y(t)$ is the at-risk process such that $P(Y(\tau) > 0) > 0$. Suppose the covariate processes $Z_j(t)$, $t \in [0,\tau]$ are predictable and uniformly bounded. Suppose $J = \int_0^{\tau} (s_2(t) - s_0(t)^{-1} s_1(t)^{\otimes 2}) h_0(t)\, dt$ is positive definite. Then $\hat{\beta}_n$ is $\sqrt{n}$-consistent for β and $\sqrt{n}(\hat{\beta}_n - \beta) \implies \mathscr{N}(0,\Sigma)$, where $\Sigma = J^{-1} - J^{-1} C(\mathscr{I}(Y,Z)) W_u^{-1}\, C(\mathscr{I}(Y,Z))^\top J^{-\top}$.*

To give an estimate of Σ, introduce

$$V_n(t,b) = \frac{\sum_{j=1}(Z_j(t) - E(t,b))^{\otimes 2} Y_j(t) \exp(b^\top Z_j(t))}{S_n(t,b)},$$

where $E(t,b) = \sum_{j=1}^n Z_j(t) Y_j(t) \exp(b^\top Z_j(t))/S_n(t,b)$. A consistent estimate of J in the literaure (e.g. Fleming and Harrington [4]) is given by

$$\hat{J} = \frac{1}{n} \sum_{j=1}^n \int_0^\tau V_n(t, \hat{\beta}_n)\, dN_j(t).$$

With a similar argument, a consistent estimate of $C(\mathscr{I}(Y,Z))$ is

$$\hat{C} = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \big(Z_j(t) - E(t, \hat{\beta}_n)\big)\, dN_j(t) u(R_j)^\top.$$

Thus one immediately obtains a consistent estimate of Σ as follows:

$$\hat{\Sigma} = \hat{J}^{-1} - \hat{J}^{-1} \hat{C} \hat{W}^{-1} \hat{C}^\top \hat{J}^{-\top},$$

where $\hat{W} = \frac{1}{n} \sum_{j=1}^n u(R_j) u(R_j)^\top$.

Below and the simulation in Section 3 are examples about side information in which the covariate processes are constant over time, i.e. $Z_j(t) = Z_j, t \in [0,\tau]$.

Example 2.1. One important situation in censoring data is that the censoring variable U is independent of the covariate variable Z. The independence implies $E(a(U) \otimes b(Z)) = 0$ for some known square-integrable vector functions a, b with mean zero. While the usual partial likelihood does not use this additional information, our EL-weighted partial likelihood can use this information by taking $u(u,z) = a(u) \otimes b(z)$. Choices of a, b can be obtained from basis functions as in the simulation in Section 3 for univariate continuous distributions.

3. A small simulation. As noted in the Introduction, in censored survival data there is usually available some additional information about covariate variables. While the partial likelihood does not use this information, the EL-weight method can use it to improve the efficiency of parameter estimates. To illustrate it, we run a small simulation based on a nice example given in [7].

Notice that the logrank test is commonly used for assessing treatment effects in survival analysis. It is well known that this test is equivalent to

TABLE 1
Simulated mean squared errors (multiplied by n) of the MPLE $\tilde{\beta}_n$ & EL-weighted MPLE $\hat{\beta}_n$ of β for $n = 150$, repetitions 2000 and number r of constraints.

Censoring %		25					50				
β	r	1	2	3	4	5	1	2	3	4	5
0	$\tilde{\beta}_n$	5.76	6.01	5.75	5.59	5.51	8.62	8.62	7.99	8.40	8.21
	$\hat{\beta}_n$	3.89	4.10	3.99	4.03	4.08	6.55	6.45	6.30	6.57	6.24
.25	$\tilde{\beta}_n$	5.84	6.01	5.56	5.96	5.55	8.77	8.63	8.78	8.28	8.65
	$\hat{\beta}_n$	4.05	4.12	4.01	4.19	4.10	6.47	6.65	6.54	6.59	7.11

the partial likelihood score test for the null hypothesis $b = 0$ in the Cox proportional hazards regression model, which postulates that the hazard rate $h(t)$ of the survival time T of an individual at time t and a $\{0,1\}$-valued covariate Z satisfy the relationship

$$h(t) = h_0(t)\exp(bZ), \quad t \geq 0,$$

where $h_0(t)$ is an unspecified nonparametric continuous baseline hazard function and b is a parameter. Let U be a censoring time of a person. As pointed out in [7], in clinical trails, in addition to data on T, U and treatment assignment, auxiliary (side) information is also collected on variable W such as age, gender and other health conditions that may be important prognostic factors which are correlated with T. Due to randomization, it is plausible to assume that the the randomization probability to treatment 1 is equal to a known π, i.e. $P(Z = 1) = \pi$, and the treatment indicator Z is independent of W. Independence of Z and W of course implies that

$$E\big(\mathbf{1}[Z = i]a_k(W)\big) = 0, \quad i = 0, 1,\ k = 1, \ldots, r,$$

for some measurable functions a_k such that $E(a_k(W)) = 0$ and $E(a_k(W)^2) < \infty$ for $k = 1, \ldots, r$. Let us assume W is univariate and has a continuous distribution G. Then $G(W)$ is uniformly distributed over $(0, 1)$ so we can choose $a_k(w) = \sqrt{2}\cos(k\pi G(w))$, the first r terms of the usual trigonometric basis. In this case, the side information can be expressed by taking $u(z, w) = \big(z - \pi,\ \tilde{a}^\top(w)\mathbf{1}[z = 0],\ \tilde{a}^\top(w)\mathbf{1}[z = 1]\big)^\top$ where $\tilde{a}^\top = (a_1, \ldots, a_r)$, so that $E(u(Z, W)) = 0$. The data available can be summarized as n realizations of i.i.d. random vectors $(X_i, \delta_i, Z_i, W_i)$ of (X, δ, Z, W), where $X_i = \min(T_i, U_i)$ and $\delta_i = \mathbf{1}[T_i \leq U_i]$. Suppose T and U are conditionally independent given Z, and $P(X \geq \tau) > 0$ for some finite $\tau > 0$. It then follows from Theorem 2.4 below that the EL-weighted maximum partial likelihood estimator

(MPLE) $\hat{\beta}_n$ is $\sqrt{n}$-consistent for β and asymptotic normal with mean zero and variance-covariance matrix Σ, i.e.,

$$\sqrt{n}(\hat{\beta}_n - \beta) \implies \mathscr{N}(0, \Sigma),$$

where $\Sigma = J^{-1} - J^{-1}C(\mathscr{I}(Y,Z))^{\otimes 2}J^{-\top}$ with J^{-1} the asymptotic variance-covariance matrix of the usual MPLE $\tilde{\beta}_n$ (assume it to be positive definite). Clearly, the EL-weighted MPLE $\hat{\beta}_n$ has a smaller variance-covariance matrix M than the MPLE $\tilde{\beta}_n$ in the sense of positive definiteness of matrices.

Following [7], the data are generated as follows. First generate (V, W) from the bivariate normal with mean zero, variance one and correlation $\rho = 0.7$; then generate Z from the Bernoulli distribution with probability of success $\pi = 0.5$; and then generate T by $T = -\exp(\beta Z)\log(1 - \Phi(V))$, where the null values of β are $\beta = 0$ and $\beta = 0.25$, and Φ is the cumulative distribution function (cdf) of the standard normal. This implies T has the exponential distribution with rate $\exp(\beta Z)$ as its conditional distribution given Z, i.e. $T|Z \sim \mathrm{Exp}(\exp(\beta Z))$, so that it follows the proportional hazards regression model $h(t) = \exp(\beta Z)$ with $h_0(t) \equiv 1$. For the censoring time U, we generate it from the conditional distribution given Z with density $c\exp(-cs)/(1 - \exp(-c\tau)), s \in [0, \tau]$. This is the truncated exponential distribution with truncation $\tau = 10$. Here c is chosen to take two values so that the censoring percentages are approximately 25% and 50%. Table 1 reports the mean squared errors (MSE's) multiplied by n for sample size $n = 150$, repetitions $M = 2000$, and the number of constraints $r = 1, \ldots, 5$. It can be seen that the MSE's of the EL-weighted MPLE $\hat{\beta}_n$ are about 20%-40% less than the MSE's of the usual MPLE $\tilde{\beta}_n$.

4. Details of the proofs.

In this section, we collect some of the proofs.

Let $\|A\|$ denote the euclidean norm and $\|A\|_o$ the operator (or spectral) norm of a matrix A which are defined by

$$\|A\|^2 = \mathrm{trace}(A^\top A) = \sum_{i,j} A_{ij}^2, \quad \|A\|_o = \sup_{\|u\|=1} \|Au\| = \sup_{\|u\|=1} (u^\top A^\top A u)^{1/2}.$$

Let $T_{n1}, \ldots, T_{nn}$ be m-dimensional random vectors. With these random vectors we associate the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n, \sum_{j=1}^n \pi_j T_{nj} = 0 \Big\}.$$

To study the asymptotic behavior of $\mathscr{R}_n$ we introduce

$$T_n^* = \max_{1\le j\le n} \|T_{nj}\|,\ \bar{T}_n = \frac{1}{n}\sum_{j=1}^n T_{nj},\ T_n^{(\nu)} = \sup_{\|u\|=1} \frac{1}{n}\sum_{j=1}^n (u^\top T_{nj})^\nu,\ \nu = 3, 4,$$

and let λ_n and Λ_n denote the smallest and largest eigen values of S_n,

$$S_n = \frac{1}{n}\sum_{j=1}^{n} T_{nj}T_{nj}^\top, \quad \lambda_n = \inf_{\|u\|=1} u^\top S_n u, \quad \Lambda_n = \sup_{\|u\|=1} u^\top S_n u.$$

We impose the following conditions on T_{nj}.

(A1) $T_n^* = o_p(n^{1/2})$.
(A2) $\|\bar{T}_n\| = O_p(n^{-1/2})$.
(A3) There is a sequence of positive definite $m \times m$ dispersion matrices W such that

$$\|S_n - W\|_o = o_p(1).$$

A sufficient condition for the preceding conditions is the square-integrability which is quoted from Proposition 6.1 of [11].

Proposition 4.1. *If $u : \mathscr{Z} \to \mathbb{R}^m$ fulfills (K), then $T_{nj} = u(Z_j), j = 1, \dots, n$ satisfy (A1) – (A3) with $W = W_u = E(u(Z)^{\otimes 2})$.*

The following is the first part of Theorem 6.1 of [11].

Theorem 4.1. *If (A1)-(A3) hold, then there exists a unique ζ_n such that*

$$1 + \zeta_n^\top T_{nj} > 0, \quad \frac{1}{n}\sum_{j=1}^{n} \frac{T_{nj}}{1 + \zeta_n^\top T_{nj}} = 0, \tag{4.1}$$

$$\|\zeta_n\| \le \frac{\|\bar{T}_n\|}{\lambda_n - \|\bar{T}_n\| T_n^*}, \quad \|\zeta_n - S_n^{-1}\bar{T}_n\|^2 \le 2\Big(\frac{1}{\lambda_n} + \frac{\Lambda_n}{9\lambda_n^2}\Big)\|\zeta_n\|^4 T_n^{(4)}, \tag{4.2}$$

and for arbitrary random vectors $R_{n1}, \dots, R_{nn}$ of the same dimension,

$$\Big\|\frac{1}{n}\sum_{j=1}^{n}\Big(\frac{R_{nj}}{1 + \zeta_n^\top T_{nj}} - R_{nj} + R_{nj}T_{nj}^\top \zeta_n\Big)\Big\|^2 \le 2\|\zeta_n\|^4 T_n^{(4)} \Big\|\frac{1}{n}\sum_{j=1}^{n} R_{nj}R_{nj}^\top\Big\|_o. \tag{4.3}$$

Proof of Theorem 2.1. To begin with, it is not difficult to verify that conditions (A1) – (A3) in Theorem 4.1 are met. Hence there exists a unique ζ_n such that

$$1 + \zeta_n^\top u(Z_j) > 0, \quad \frac{1}{n}\sum_{j=1}^{n} \frac{u(Z_j)}{1 + \zeta_n^\top u(Z_j)} = 0,$$

on an event whose probability converges to one as n tends to infinity. Thus, $\mathbb{M}_n(\vartheta)$ is well defined on this event and defined to be an arbitrary number on the complement of this event which has an vanishing probability as n tends to infinity.

The asymptotic behaviors of the estimator $\hat{\theta}_n$ follows from the investigation of the convex function $\Delta_n(s) := n(\mathbb{M}_n(\theta + n^{-1/2}s) - \mathbb{M}_n(\theta))$. It is minimized by $\sqrt{n}(\hat{\theta}_n - \theta)$. Note first that (2.2) implies $nE(R(Z, n^{-1/2}s) = 1/2s^\top Hs + r_{1n}(s)$, where $r_{1n}(s) = no(\|s\|^2/n) = o(1)$ for every fixed s. Thus, it follows from (2.1) that

$$\Delta_n(s) = n^{1/2}\sum_{j=1}^n \pi_{nj} D(Z_j)^\top s + 1/2s^\top Hs + r_{1n}(s) + r_{2n}(s), \tag{4.4}$$

where it is shown below that for every fixed s,

$$r_{2n}(s) = n\sum_{j=1}^n \pi_{nj}\Big(R(Z_j, n^{-1/2}s) - E(R(Z_j, n^{-1/2}s))\Big) = o_p(1). \tag{4.5}$$

Since both u and D satisfy (K1a), in view of Proposition 4.1, we can apply Theorem 4.1 with $R_{nj} = D(Z_j)$ in inequality (4.3) to conclude

$$\bar{\mathbb{D}}_n := \sum_{j=1}^n \pi_{nj} D(Z_j) = \frac{1}{n}\sum_{j=1}^n\big(D(Z_j) - C(D)W^{-1}u(Z_j)\big) + o_p(n^{-1/2}). \tag{4.6}$$

Substitution of (4.5) and (4.6) in (4.4) leads to

$$\Delta_n(s) = n^{1/2}\bar{\mathbb{D}}_n^\top s + 1/2s^\top Hs + o_p(1), \quad s \in \mathbb{R}^k.$$

The right side of the above equality is approximately a quadratic function in s and has the unique minimizer $\hat{s} = -H^{-1}n^{1/2}\mathbb{D}_n^\top$. This, (4.6) and the Basic Corollary of [6] now yield the desired result, i.e. $\hat{s} - \sqrt{n}(\hat{\theta}_n - \theta) = o_p(1)$.

We are now left to prove (4.5). To this end, fixed $s \in \mathbb{R}^k$, and apply inequality (4.3) with $R_{nj} = R_{nj}(s) := R(Z_j, n^{-1/2}s)$ to get

$$\|\frac{1}{n}\sum_{j=1}^n\big(\frac{R_{nj}(s)}{1+\zeta_n^\top u(Z_j)} - R_{nj}(s) + R_{nj}(s)u(Z_j)^\top\zeta_n\big)\|^2 \le 2\|\zeta_n\|^4 u^{(4)} V_n(s) \tag{4.7}$$

where $V_n(s) = n^{-1}\sum_{j=1}^n R(Z_j, n^{-1/2}s)^2$. Notice that

$$\begin{aligned} E(R(Z, n^{-1}s)^2) &= \mathrm{Var}(R(Z, n^{-1}s)) + \big(E(R(Z, n^{-1}s))\big)^2 \\ &= o(n^{-1}) + O(n^{-1}) = O(n^{-1}), \end{aligned}$$

where the second equality follows from the assumptions. Thus, by Markov inequality, for $A > 0$,

$$P(nV_n(s) > A) \le \frac{E(nV_n(s))}{A} \le \frac{nE\left(R(Z, n^{-1/2}s)^2\right)}{A} \le \frac{nO(n^{-1})}{A} \to 0,$$

as A tends to infinity. Therefore $V_n(s) = O_p(n^{-1})$. Besides from the first inequality of (4.2) and Proposition 4.1 it follows

$$\|\zeta_n\|^4 u_n^{(4)} = O_p(\|\bar{u}_n\|^4) O_p((u_n^*)^2) = o_p(n^{-1}). \tag{4.8}$$

Thus the right side of (4.7) is $o_p(n^{-2})$, so that

$$r_{2n} = \sum_{j=1}^{n} (R_{nj}(s) - E(R_{nj}(s)) - \sum_{j=1}^{n} R_{nj}(s) u(Z_j)^\top \zeta_n + o_p(1).$$

The second moment of the above first sum is equal to

$$n \operatorname{Var}(R(Z, n^{-1/2}s)) = no(\|s\|^2/n) = o(1).$$

Write the second sum as $nA_n(s) + nB_n(s)$, where

$$A_n(s) = \frac{1}{n} \sum_{j=1}^{n} \big(R_{nj}(s) - E(R_{nj}(s))\big) u(Z_j)^\top \zeta_n, \quad B_n(s) = E(R_{n1}(s)) \bar{u}_n^\top \zeta_n.$$

By Cauchy inequality,

$$\begin{aligned}
\|A_n(s)\|^2 &\le \frac{1}{n} \sum_{j=1}^{n} \big(R_{nj}(s) - E(R_{nj}(s))\big)^2 \|\zeta_n\|^2 \, \|\frac{1}{n} \sum_{j=1}^{n} u(Z_j)^{\otimes 2}\| \\
&= o(n^{-1}) O(n^{-1}) O(1) = o(n^{-2}),
\end{aligned}$$

so that $A_n(s) = o_p(n^{-1})$, while

$$\|B_n(s)\| \le \|E(R_{n1}(s))\| \, \|\bar{u}_n\| \, \|\zeta_n\| = O(n^{-1}) O_p(n^{-1/2}) O_p(n^{-1/2}) = O_p(n^{-2}).$$

Combining the above yields that the second sum is $o_p(1)$. This proves (4.5) and completes the proof. □

Proof of Theorem 2.2. We shall apply Theorem 2.1 to prove the results. To start with, one verifies that (2.1) holds with

$$D(z) = x(\mathbf{1}[y \le \beta^\top x] - p),$$

where clearly $E(D(Z)) = 0$ and $W = E(D(Z)^{\otimes 2}) = p(1-p)E(X^{\otimes 2})$ is positive definite implied by (Q1) and (Q2). Denote E_x the conditional expectation given $X = x$, and $\Delta_x(y,t) = m_p(y,(\beta+t)^\top x) - m_p(y,\beta^\top x)$, so that

$$\Delta_x(y,t) = pa_x(y,t) + (1-p)b_x(y,t), \tag{4.9}$$

where $a_x(y,t) = (y-(\beta+t)^\top x)_+ - (y-\beta^\top x)_+$ and $b_x(y,t) = ((\beta+t)^\top x - y)_+ - (\beta^\top x - y)_+$. With the above notation we now write

$$E_x(R(Z,t)) = E_x(\Delta_x(Y,t)) := -pA_x(t) + (1-p)B_x(t),$$

where

$$A_x(t) = -\int_{(\beta+t)^\top x}^{\infty} (y-(\beta+t)^\top x)f_x(y)\,dy + \int_{\beta^\top x}^{\infty} (y-\beta^\top x)f_x(y)\,dy.$$

$$B_x(t) = \int_{-\infty}^{(\beta+t)^\top x} ((\beta+t)^\top x - y)f_x(y)\,dy - \int_{-\infty}^{\beta^\top x} (\beta^\top x - y)f_x(y)\,dy,$$

Clearly,

$$A_x(t) = \int_{\beta^\top x}^{(\beta+t)^\top x} (y-\beta^\top x)f_x(y)\,dy + t^\top x(1 - F_x((\beta+t)^\top x)),$$

$$B_x(t) = \int_{\beta^\top x}^{(\beta+t)^\top x} (\beta^\top x - y)f_x(y)\,dy + t^\top x F_x((\beta+t)^\top x).$$

Hence,

$$E_x(R(Z,t)) = t^\top x\big(F_x((\beta+t)^\top x) - p\big) - \int_{\beta^\top x}^{(\beta+t)^\top x} (y-\beta^\top x)f_x(y)\,dy.$$

By (Q1) and Taylor expansion, for every x, we have

$$F_x((\beta+t)^\top x) - p = f(\beta^\top x)t^\top x + r_1(t,x),$$

where $r_1(t,x) = o(\|t\|)$. It is shown below that

$$\int_{\beta^\top x}^{(\beta+t)^\top x} (y-\beta^\top x)f_x(y)\,dy = 1/2t^\top xx^\top tf(\beta^\top x) + r_2(t,x), \tag{4.10}$$

where $r_2(t,x) = o(\|t\|^2)$ for every x. Thus, in view of the boundedness of $f_x(y)$ by (Q1) and the finite second moment of X implied by (Q2), we can apply the dominated convergence theorem to conclude

$$E(R(Z,t)) = E\big(E_X(R(Z,t))\big) = 1/2t^\top Ht + o(\|t\|^2). \tag{4.11}$$

This shows that (2.2) holds with $H = E\left((XX^\top)(f_X(\beta^\top X))\right)$. We now show

$$\mathrm{Var}(R(Z,t)) = o(\|t\|^2), \quad t \to 0. \tag{4.12}$$

To this end, recall $\Delta_x(y,t)$ in (4.9) and get

$$\begin{aligned} \mathrm{Var}(R(Z,t)) &= E\big([\Delta_X(Y,t) - E\left(\Delta_X(Y,t)\right) - D(Z)^\top t]^2\big) \\ &:= L + t^\top K t - 2Mt, \quad \text{say,} \end{aligned} \tag{4.13}$$

where

$$L = E\left(\Delta_X(Y,t) - E\left(\Delta_X(Y,t)\right)\right)^2,$$
$$M = E\big((\Delta_X(Y,t) - E\left(\Delta_X(Y,t)\right))D(Z)^\top\big).$$

Note first that $D(Z)$ has mean zero, so that

$$\begin{aligned} M = E\big(\Delta_X(Y,t)D(Z)^\top\big) &= E\big(X^\top \Delta_X(Y,t)(\mathbf{1}[Y \le \beta^\top X] - p)\big) \\ &= E\big(X^\top B(X)\big), \quad \text{say,} \end{aligned}$$

where

$$\begin{aligned} B(x) &= E_x\big(\Delta_x(Y,t)(\mathbf{1}[Y \le \beta^\top x] - p)\big) \\ &= E_x\big(\Delta_x(Y,t)\mathbf{1}[Y \le \beta^\top x]\big) - pE_x\big(\Delta_x(Y,t)\big). \end{aligned}$$

Since $E\left(E_X\big(\Delta_X(Y,t)\big)\right) = E\left(R(Z,t)\right)$, it follows from (4.11) that

$$E\big(\Delta_X(Y,t)\big) = 1/2 t^\top H t + o(\|t\|^2). \tag{4.14}$$

For $t^\top x \ge 0$ and in view of (4.9),

$$E_x\big(\Delta_x(Y,t)\mathbf{1}[Y \le \beta^\top x]\big) = (1-p)t^\top x F_x(\beta^\top x) = p(1-p)t^\top x, \tag{4.15}$$

whereas for $t^\top x < 0$, by (Q1) and apply Taylor expansion,

$$\begin{aligned} E_x\big(\Delta_x(Y,t)\mathbf{1}[Y \le \beta^\top x]\big) &= p\int_{(\beta+t)^\top x}^{\beta^\top x} (y - (\beta+t)^\top x) f(y)\,dy \\ &\quad + (1-p)(t^\top x)F_x((\beta+t)^\top x) - (1-p)\int_{(\beta+t)^\top x}^{\beta^\top x} (\beta^\top x - y) f(y)\,dy \\ &= p(t^\top x)(F_x(\beta^\top x) - F_x((\beta+t)^\top x)) + (1-p)(t^\top x)F_x((\beta+t)^\top x) \\ &\quad + \int_{(\beta+t)^\top x}^{\beta^\top x} (y - \beta^\top x) f(y)\,dy \\ &= p(t^\top x)\big(f_x(\beta^\top x)(t^\top x) + o(t)\big) + (1-p)(t^\top x)\big(F_x(\beta^\top x) \\ &\quad + f_x(\beta^\top x)(t^\top x) + o(t)\big) - 1/2 t^\top x x^\top t f_x(\beta^\top x) + o(\|t\|^2) \\ &= 1/2 t^\top x x^\top t f_x(\beta^\top x) + p(1-p)(t^\top x) + o(\|t\|^2) \\ &= p(1-p)(t^\top x) + O(\|t\|^2). \end{aligned}$$

Using the same argument as in proving (4.11) and (4.14)-(4.15), one has

$$M = p(1-p)t^\top E(XX^\top) + O(\|t\|^2). \tag{4.16}$$

To compute L, recall (4.9) and write the second moment in L as

$$\begin{aligned} E\left(\Delta_X(Y,t)^2\right) =& p^2 E(a_X(Y,t)^2) + (1-p)^2 E(b_X(Y,t)^2) \\ & - 2p(1-p)E(a_X(Y,t)b_X(Y,t)). \end{aligned}$$

Let us first deal with $c(x) := E_x(a_x(Y,t)b_x(Y,t))$. For $t^\top x \geq 0$, by Taylor formula and (Q1),

$$c(x) = -p(1-p)\int_{\beta^\top x}^{(\beta+t)^\top x} (y - \beta^\top x)((\beta+t)^\top x - y) f_x(y)\, dy = o(\|t\|^2).$$

It is easily seen that the above rate also holds for $t^\top x < 0$. Thus as argued above, one derives

$$E(a_X(Y,t)b_X(Y,t)) = o(\|t\|^2). \tag{4.17}$$

Now that for $t^\top x \geq 0$, by Taylor formula and (Q1), one has

$$\begin{aligned} E_x(a_x(Y,t)^2) =& \int_{\beta^\top x}^{(\beta+t)^\top x} (y - (\beta+t)^\top x)^2 f_x(y)\, dy \\ & + (t^\top x)^2(1 - F_x((\beta+t)^\top x)) = (1-p)(t^\top x)^2 + o_p(\|t\|^2). \end{aligned}$$

Similarly, for $t^\top x < 0$ the above also holds. Therefore,

$$E(a_X(Y,t)^2) = (1-p)t^\top Ht + o_p(\|t\|^2). \tag{4.18}$$

With the same method, one proves

$$E(b_X(Y,t)^2) = pt^\top Ht + o_p(\|t\|^2). \tag{4.19}$$

It thus follows from (4.17)-(4.19) that

$$E\left(\Delta_X(Y,t)^2\right) = p(1-p)t^\top Ht + o_p(\|t\|^2).$$

Hence, in view of (4.11), one arrives at

$$L = E\left(\Delta_X(Y,t)^2\right) - \left(E\left(\Delta_X(Y,t)\right)\right)^2 = p(1-p)t^\top Ht + o_p(\|t\|^2).$$

This, (4.13) and (4.16) immediately yield the desired (4.12).

We are now left to prove (4.10). Let us write $t = ru$, where u is a unit vector in the direction of t and $r = \|t\|$. Using L'hospital's rule the desired (4.10) now follows from

$$\begin{aligned}
&\lim_{r\to 0} r^{-2}\Big(\int_{\beta^\top x}^{(\beta+t)^\top x} (y-\beta^\top x) f_x(y)\,dy - 1/2 t^\top x x^\top t f(\beta^\top x)\Big)\\
&= \lim_{r\to 0} r^{-2}\int_0^{ru^\top x} s\big(f_x(s+\beta^\top x) - f(\beta^\top x)\big)\,ds\\
&= \lim_{r\to 0} 2^{-1} r^{-1} r u^\top x\big(f_x(ru^\top x+\beta^\top x) - f(\beta^\top x)\big) = 0,
\end{aligned}$$

where the continuity of f_x is used. □

Proof of Theorem 2.3. Let $\tilde\beta_n$ be the estimator under model (2.4) based on the check function approach. It is then the AGM-estimator of the quantile estimating equation

$$\bar v_n(b) := \frac{1}{n}\sum_{j=1}^n X_j(\mathbf{1}[Y_j \le b^\top X_j] - p) = o_p(n^{-1/2}).$$

For AGM-estimators, see e.g. [1] or [11]. This equation has more than one solutions because $\bar v_n(b)$ is a step function. But all of the solutions will converge to the unique solution of the equation $v(b) := E(X(F_X(b^\top X) - p)) = 0$ under assumptions (Q1)-(Q2). In fact, it has a unique solution β as $v'(\beta) = E(XX^\top f_X(\beta^\top X))$ is positive definite. See [8] for more discussions. Note that the EL-weighted AGM-estimator $\tilde\beta_n$ is the solution to the equation

$$\bar{\boldsymbol{v}}_n(b) := \frac{1}{n}\sum_{j=1}^n \frac{X_j(\mathbf{1}[Y_j \le b^\top X_j] - p)}{1+\zeta_n^\top u(Z_j)} = o_p(n^{-1/2}).$$

Thus it suffices to prove that $\tilde\beta_n$ has the identitical asymptotic distribution as $\hat\beta_n$ defined by the EL-weighted minimization problem (2.5). This can be done by applying Theorem 4.4 of [11], so we must verify its assumptions which are the combination of the assumptions of Theorems 4.1– 4.3 in [11]. Clearly the assumptions of Theorem 4.3 are satisfied in view of the fact that $\bar v_n(b)$ is upper semi-continuous. We now verify that assumptions (4.1)– (4.3) in Theorem 4.1 are met with $v(z,b) = x(\mathbf{1}[y \le b^\top x] - p)$, where $z = (x^\top, y)^\top$. Since X is square-integrable, it follows that (4.1) holds. To prove the uniform convergence in (4.2), let us assume $\mathbb{B} = \big\{t \in \mathbb{R}^k : \|t\| \le R\big\}$ for convenience. For $\delta > 0$, let $\mathbb{B}_\delta(a) = \big\{b \in \mathbb{R}^k : \|b-a\| \le \delta\big\}$ denote the closed ball of radius δ centered at $a \in \mathbb{R}^k$. Let $a_1, \dots, a_m$ be points in $\mathbb{B}$ such that $\mathbb{B} \subset \cup_{i=1}^m \mathbb{B}_\delta(a_i)$.

One can take these a_i's to be the elements of the grid $\{i\delta/\sqrt{k} : i \in \mathbb{Z}^k\}$ that fall into $\mathbb{B}$, in which case

$$m = [(1 + 2\sqrt{k}R/\delta)^k]. \tag{4.20}$$

Observe that for reals y_1, y_2 we have

$$|\mathbf{1}[Y \le y_2] - \mathbf{1}[Y \le y_1]| \le \mathbf{1}[y_1 \wedge y_1 \le Y \le y_1 \vee y_2].$$

For b in the ball $\mathbb{B}_\delta(a_i)$, there exist random vectors b_{ij}, B_{ij} in the ball and depending on X_j such that $X_j^\top b_{ij} \le X_j^\top b \le X_j^\top B_{ij}$. Consequently,

$$\sup_{b \in \mathbb{B}_\delta(a_i)} |\mathbf{1}[Y_j \le X_j^\top b] - \mathbf{1}[Y_j \le X_j^\top a_i]| \le \mathbf{1}[m_{ij} \le Y \le M_{ij}], \tag{4.21}$$

where $m_{ij} = X_j^\top a_i \wedge X_j^\top b_{ij}$ and $M_{ij} = X_j^\top a_i \vee X_j^\top B_{ij}$. Clearly, $M_{ij} - m_{ij} \le \|X_j\|\delta \le C\delta$, where $\|X_j\| \le C$. By (Q1), $f_x(y) \le B$ for all x, y, hence

$$E(\mathbf{1}[m_{ij} \le Y_j \le M_{ij}]) = E(F_X(M_{ij}) - F_X(m_{ij})) \le BC\delta. \tag{4.22}$$

Denote $V_n(b) = \bar{v}_n(b) - E(\bar{v}_n(b))$. Then using the usual chaining method,

$$\sup_{b \in \mathbb{B}} \|V_n(b)\| \le \max_i \|V_n(a_i)\| + \max_i \sup_{b \in \mathbb{B} \cap \mathbb{B}_\delta(a_i)} \|V_n(b) - V_n(a_i)\|.$$

Hence,

$$\begin{aligned} P\Big(\sup_{b \in \mathbb{B}} \|V_n(b)\| > 4\epsilon\Big) &\le P\Big(\max_i \|V_n(a_i)\| > 2\epsilon\Big) \\ &\quad + P\Big(\max_i \sup_{b \in \mathbb{B} \cap \mathbb{B}_\delta(a_i)} \|V_n(b) - V_n(a_i)\| > 2\epsilon\Big). \end{aligned} \tag{4.23}$$

Since $\mathrm{Var}(v(Z, a_i)) \le E(v^2(Z, a_i)) \le C^2$, it follows from Bernstein's inequality that for $\epsilon > 0$,

$$P(\max_i \|V_n(a_i)\| > 2\epsilon) \le 2m \exp\Big(-\frac{2n\epsilon^2}{C^2 + C\epsilon/3}\Big). \tag{4.24}$$

Notice that for $b \in \mathbb{B}_\delta(a_i)$, $E(\|v(Z, b) - v(Z, a_i\|) \le BC^2\delta$, hence

$$\|V_n(b) - V_n(a_i)\| \le \|\bar{v}_n(b) - \bar{v}_n(a_i)\| + BC^2\delta. \tag{4.25}$$

By (4.21),

$$\|\bar{v}_n(b) - \bar{v}_n(a_i)\| \le C\frac{1}{n}\sum_{j=1}^n \mathbf{1}[m_{ij} \le Y_j \le M_{ij}], \quad b \in \mathbb{B}_\delta(a_i).$$

Now use (4.22) and apply again Bernstein's inequality to get

$$
\begin{aligned}
&P\Big(C|\frac{1}{n}\sum_{j=1}^{n}\mathbf{1}[m_{ij}\le Y_j\le M_{ij}]-E(\mathbf{1}[m_{ij}\le Y_j\le M_{ij}])|>\epsilon\Big)\\
&\le 2\exp\Big(-\frac{n\epsilon^2}{2BC^3\delta+2C^2\epsilon/3}\Big).
\end{aligned}
\tag{4.26}
$$

Therefore, by (4.25), we derive

$$
\begin{aligned}
&P\Big(\max_i \sup_{b\in\mathbb{B}\cap\mathbb{B}_\delta(a_i)}\|V_n(b)-V_n(a_i)\|>4\epsilon\Big)\\
&\le \sum_{i=1}^{m}P\Big(\sup_{b\in\mathbb{B}\cap\mathbb{B}_\delta(a_i)}\|\bar{v}_n(b)-\bar{v}_n(a_i)\|>2\epsilon\Big)+m\mathbf{1}[BC^2\delta>2\epsilon]\\
&\le 2m\exp\Big(-\frac{n\epsilon^2}{2BC^3\delta+2C^2\epsilon/3}\Big)+m\mathbf{1}[BC^2\delta>\epsilon]+m\mathbf{1}[BC^2\delta>2\epsilon],
\end{aligned}
\tag{4.27}
$$

where the second last term is resulted from the centering in (4.24). In view of (4.23) and by (4.20), (4.24) and (4.27), we arrive at

$$
P\Big(\sup_{b\in\mathbb{B}}\|V_n(b)\|>4\epsilon\Big)\to 0 \tag{4.28}
$$

for every $\epsilon>0$ by choosing $m=m_n\to\infty$ such that $\log(m_n)/n\to 0$. This shows (4.2). To prove (4.3), notice that $E(v(Z,\beta))=0$ and by (Q1)-(Q2) we obtain

$$
E(v(Z,b))=E(v(Z,b)-v(Z,\beta))=E(X\otimes 2f(\beta^\top X)(b-\beta)+o(b-\beta).
$$

Accordingly $\|E(v(Z,b))\|=\|E(X^{\otimes 2}f(\beta^\top X)\|\|b-\beta\|+o(\|b-\beta\|)>0$ by the positive definiteness for $\|b-\beta\|>0$. This shows (4.3) in [11] hence proves Theorem 4.1 in [11].

We are left to verify assumptions (5.1) and (5.2) of Theorem 4.2 in [11]. Note first that the boundedness of the conditional density $f_x(y)$, $|X|\le C$ and the dominated convergence theorem imply that $E(v(Z,b))$ is differentiable with respect to b at $b=\beta$ with nonsingular gradient $E(X^{\otimes 2}f_X(X^\top\beta))$. To prove (5.1) in [11], recall the above V_n and decompose

$$
\sqrt{n}(V_n(b)-E(V_n(b)))=A_n(b)+B_n(b), \tag{4.29}
$$

where $A_n(b)=n^{-1/2}\sum_{j=1}^{n}s(Z_j,b)$ with

$$
s(Z_j,b)=X_j\big(\mathbf{1}[Y_j\le X_j^\top b]-\mathbf{1}[Y_j\le X_j^\top\beta]-F_{X_j}(X_j^\top b)+F_{X_j}(X_j^\top\beta)\big),
$$

and $B_n(b) = B_{1n}(b) + B_{2n}(b) - B_{3n}(b)$ with

$$B_{1n}(b) = n^{-1/2} \sum_{j=1}^{n} X_j \big(F_{X_j}(X_j^\top b) - F_{X_j}(X_j^\top \beta) - f_{X_j}(X_j^\top \beta) X_j^\top (b-\beta) \big),$$

$$B_{2n}(b) = n^{-1/2} \sum_{j=1}^{n} \big(X_j^{\otimes 2} f_{X_j}(X_j^\top \beta) - E(X^{\otimes 2} f_X(X^\top \beta)) \big)(b-\beta),$$

$$B_{3n}(b) = n^{1/2} E\big(X(F_X(X^\top b) - F_X(X^\top \beta)) - f_X(X^\top \beta) X^\top \big)(b-\beta).$$

Let $\epsilon_n \downarrow 0$. Since $f_x(y)$ is continuous and $X_j^\top b$ is bounded, it follows from the law of large numbers that

$$B_{12n} := \sup_{\|b-\beta\| \le \epsilon_n} \| \frac{1}{n} \sum_{j=1}^{n} \big(F_{X_j}(X_j^\top b) - F_{X_j}(X_j^\top \beta) - f_{X_j}(X_j^\top \beta) \big) X_j \| = o_p(1).$$

Denote $\tau_n(b) = 1 + \sqrt{n}\|b-\beta\|$. Then $\sqrt{n}\|b-\beta\|/\tau_n(b) \le 1$ and

$$\sup_{\|b-\beta\| \le \epsilon_n} \frac{B_{1n}(b)}{\tau_n(b)} = B_{12n} \sup_{\|b-\beta\| \le \epsilon_n} \frac{\sqrt{n}(b-\beta)}{\tau_n(b)} = o_p(1). \tag{4.30}$$

Observe

$$\frac{B_{2n}(b)}{\tau_n(b)} = \frac{1}{n} \sum_{j=1}^{n} \big(X_j^{\otimes 2} f_{X_j}(X_j^\top \beta) - E(X^{\otimes 2} f_X(X^\top \beta)) \big) \frac{\sqrt{n}(b-\beta)}{\tau_n(b)},$$

and use the law of large numbers to conclude,

$$\sup_{\|b-\beta\| \le \epsilon_n} \frac{B_{2n}(b)}{\tau_n(b)} = o_p(1). \tag{4.31}$$

By the dominated convergence theorem, one similarly has

$$\sup_{\|b-\beta\| \le \epsilon_n} \frac{B_{3n}(b)}{\tau_n(b)} = o_p(1). \tag{4.32}$$

Consequently the above (4.30) – (4.32) yield

$$\sup_{\|b-\beta\| \le \epsilon_n} \frac{B_n(b)}{\tau_n(b)} = o_p(1). \tag{4.33}$$

Let us now handle $A_n(b)$. Note first that b runs over the ball centered at β with radius ϵ_n, i.e. $\mathbb{B}_{\epsilon_n}(\beta)$. For any $b \in \mathbb{B}_{\epsilon_n}(\beta)$, by Bernstein's inequality,

$$P\Big(\frac{\|A_n(b)\|}{\tau_n(b)} > \epsilon \Big) \le 2 \exp\Big(- \frac{\epsilon^2 \tau_n^2(b)}{2\sigma^2(b) + (4C\epsilon/3) n^{-1/2} \tau_n(b)} \Big), \tag{4.34}$$

where $\sigma^2(b) = \mathrm{Var}(s(Z,b))$. To calculate this variance, set $t(Z,b) = \mathbf{1}[Y \leq X^\top b] - \mathbf{1}[Y \leq X^\top \beta]$. Then the conditional second moment given X is

$$E_X(t^2(Z,b)) = F_X(X^\top b) + F_X(X^\top \beta) - 2P_X(Y \leq X^\top b \wedge X^\top \beta).$$

By assumption, one can apply the mean value theorem to conclude that there is an η in between $X^\top b$ and $X^\top \beta$ such that $\mu_X(\eta) := E_X(t(Z,b)) = F_X(X^\top b) - F_X(X^\top \beta) = f_X(\eta)X^\top(b-\beta)$. Thus,

$$\begin{aligned} P_X(Y \leq X^\top b \wedge X^\top \beta) &= F_X(X^\top \beta) + \mathbf{1}[X^\top b < X^\top \beta](F_X(X^\top b) - F_X(X^\top \beta)) \\ &= F_X(X^\top \beta) + \mathbf{1}[X^\top b < X^\top \beta] f_X(\eta) X^\top (b-\beta). \end{aligned}$$

Hence,

$$\mathrm{Var}_X(t(Z,b)) = \big(1 - 2\mathbf{1}[X^\top b < X^\top \beta]\big)\mu_X(\eta) - \mu_X^2(\eta).$$

Since $f_x(y)$ is continuous and X is bounded, it immediately follows that $\mu_X(\eta) = f_X(X^\top \beta)X^\top(b-\beta) + o_p(b-\beta)$ and $\mu_X^2(\eta) = o_p(b-\beta)$ uniformly in $b \in \mathbb{B}_{\epsilon_n}(\beta)$ (e.g. $b \mapsto \mu_X^2(\eta)/\|b-\beta\|$ converges to zero in probability uniformly in $b \in \mathbb{B}_{\epsilon_n}(\beta)$ as $n \to \infty$), and hence the same uniformly holds for

$$\mathrm{Var}_X(t(Z,b)) = \big(1 - 2\mathbf{1}[X^\top b < X^\top \beta]\big) f_X(X^\top \beta) X^\top (b-\beta) + o_p(b-\beta).$$

Let $G(b) = E\big(X^{\otimes 2}(1 - 2\mathbf{1}[X^\top b < X^\top \beta]) f_X(X^\top \beta) X^\top\big)$. Then it holds uniformly in $b \in \mathbb{B}_{\epsilon_n}(\beta)$ as n tends to infinity that

$$\sigma^2(b) = E\big(X^{\otimes 2}\,\mathrm{Var}_X(t(Z,b))\big) = G(b)(b-\beta) + o(b-\beta). \tag{4.35}$$

Next, we partition $\mathbb{B}_{\epsilon_n}(\beta)$ as above, so the above R is equal to the ϵ_n here. In the present case, denote $\rho = \delta$, $l = m$ and $b_i = a_i, i = 1, \dots, l$, thus

$$l_n := l = [(1 + 2\sqrt{k}\epsilon_n/\rho)^k]. \tag{4.36}$$

Similar to (4.21), for b in the ball $\mathbb{B}_\rho(b_i)$, there exist random vectors a_{ij}, A_{ij} in the ball and depending on X_j such that $X_j^\top a_{ij} \leq X_j^\top b \leq X_j^\top A_{ij}$. Consequently,

$$\sup_{b \in \mathbb{B}_\rho(b_i)} |\mathbf{1}[Y_j \leq X_j^\top b] - \mathbf{1}[Y_j \leq X_j^\top b_i]| \leq \mathbf{1}[l_{ij} \leq Y_j \leq L_{ij}], \tag{4.37}$$

where $l_{ij} = X_j^\top b_i \wedge X_j^\top a_{ij}$ and $L_{ij} = X_j^\top b_i \vee X_j^\top A_{ij}$. Clearly $L_{ij} - l_{ij} \leq \|X_j\|\rho \leq C\rho$. Hence,

$$E(\mathbf{1}[l_{ij} \leq Y_j \leq L_{ij}]) = E(F_X(L_{ij}) - F_X(l_{ij})) \leq BC\rho. \tag{4.38}$$

Again using the usual chaining method, we push forward

$$
\begin{aligned}
P\Big(\sup_{b\in\mathbb{B}_{\epsilon_n}(\beta)} \frac{\|A_n(b)\|}{\tau_n(b)} > 4\epsilon\Big) &\le P\Big(\max_i \frac{\|A_n(b_i)\|}{\tau_n(b_i)} > 2\epsilon\Big) \\
&+ P\Big(\max_i \sup_{b\in\mathbb{B}_\rho(b_i)} \|\frac{A_n(b)}{\tau_n(b)} - \frac{A_n(b_i)}{\tau_n(b_i)}\| > 2\epsilon\Big) := C_n + DE_n, \text{ say.}
\end{aligned}
\tag{4.39}
$$

Applying (4.34) to the $b_i : i = 1, \ldots, l_n$, we obtain

$$
C_n \le \sum_{i=1}^{l_n} 2\exp\Big(-\frac{2\epsilon^2\tau_n^2(b_i)}{\sigma^2(b_i) + (2C\epsilon/3)n^{-1/2}\tau_n(b_i)}\Big).
$$

Notice that $\tau_n(b) \ge 1$, $\|\sqrt{n}(b-\beta)\|/\tau_n(b) \le 1$ and $\|G(b)\| \le BC^3$. Hence it follows from the uniform convergence in (4.35) that

$$
\frac{\sigma^2(b_i)}{\tau_n(b_i)} = G(b_i)\frac{b_i-\beta}{\tau_n(b_i)} + o_p\Big(\frac{b_i-\beta}{\tau_n(b_i)}\Big) \le n^{-1/2}BC^3 + o(n^{-1/2}), \quad i = 1, \ldots, l_n.
$$

Accordingly,

$$
C_n \le l_n \exp\Big(-\frac{2\epsilon^2 n^{1/2}}{BC^3 + o(1) + 2C\epsilon/3}\Big).
\tag{4.40}
$$

Let us now write

$$
\frac{A_n(b)}{\tau_n(b)} - \frac{A_n(b_i)}{\tau_n(b_i)} = A_n(b_i)\big(\frac{1}{\tau_n(b)} - \frac{1}{\tau_n(b_i)}\big) + \frac{A_n(b) - A_n(b_i)}{\tau_n(b)}.
$$

For $b \in \mathbb{B}_\rho(b_i)$, since

$$
\sup_{b\in\mathbb{B}_\delta(b_i)} |\frac{1}{\tau_n(b)} - \frac{1}{\tau_n(b_i)}| \le \frac{\sqrt{n}\rho}{\tau_n(b_i)}, \quad \tau_n(b) \ge 1,
$$

we derive

$$
\begin{aligned}
DE_n \le& \sum_{i=1}^{l_n} P\big(\frac{\|A_n(b_i)\|}{\tau_n(b_i)} > \frac{\epsilon}{\sqrt{n}\rho}\big) + \sum_{i=1}^{l_n} P\big(\sup_{b\in\mathbb{B}_\rho(b_i)} \|A_n(b) - A_n(b_i)\| > \epsilon\big) \\
&:= D_n + E_n, \quad \text{say.}
\end{aligned}
\tag{4.41}
$$

Use $\tau_n(b_i) \ge 1$ and apply (4.34) to get

$$
D_n \le 2l_n \exp\Big(\frac{-\epsilon^2}{2BC^3 n\rho^3 + o(n\rho^3) + (4C\epsilon/3)\rho}\Big).
\tag{4.42}
$$

For $b \in \mathbb{B}_\rho(b_i)$, from (4.37) it follows

$$\begin{aligned}\sqrt{n}\|A_n(b) - A_n(b_i)\| \le C \sum_{j=1}^{n} \big(\mathbf{1}[l_{ij} \le Y_j \le L_{ij}] - P(l_{i1} \le Y_1 \le L_{i1})\big) \\ + 2nCP(l_{i1} \le Y_1 \le L_{i1}).\end{aligned}$$

Thus, by (4.38) and (4.34), we obtain

$$E_n \le l_n \exp\Big(\frac{-\epsilon^2}{8BC^4\rho + (16C^3\epsilon/3)n^{-1/2}}\Big) + l_n \mathbf{1}[2\sqrt{n}BC^2\rho > \epsilon/2]. \tag{4.43}$$

In view of (4.36), it is easy to verify that $C_n + D_n + E_n = o(1)$ by choosing $l_n = O(n^{k/2})$. This, (4.29), (4.33) and (4.39) immediately yield the desired (5.1) in [11]. Finally, we shall prove the remaining uniform convergence in (5.2) of Theorem 4.2 in [11]. This can be done pretty much the same way as above. Here we give sketches. Note first that the desired (5.2) in [11] is implied by the following stronger result

$$\sup_{\|\vartheta-\theta\|\le\epsilon_n} \frac{1}{n}\sum_{j=1}^{n} |\mathbf{1}[Y_j \le X_j^\top b] - \mathbf{1}[Y_j \le X_j^\top \beta]| = o_p(1). \tag{4.44}$$

Denote $w(Z_j, b) = |\mathbf{1}[Y_j \le X_j^\top b] - \mathbf{1}[Y_j \le X_j^\top \beta]|$, $\bar{w}(Z_j, b) = w(Z_j, b) - E(w(Z, b))$ and $W_n(b) = n^{-1}\sum_{j=1}^{n} \bar{w}(Z_j, b)$. Then analoguously one derives

$$P\Big(\max_i \|W_n(b_i)\| > 2\epsilon\Big) \le 2l_n \exp\Big(-\frac{2n\epsilon^2}{BC^5 + 2C^2/3}\Big). \tag{4.45}$$

Since $\|\bar{w}(Z_j, b) - \bar{w}(Z_j, b_i)\| \le \mathbf{1}[l_{ij} < Y_j \le L_{ij}] - P(l_{ij} < Y_j \le L_{ij}) + 2BC\rho$, it follows

$$\begin{aligned}P\Big(\max_i \sup_{b\in\mathbb{B}_\delta(b_i)} \|W_n(b) - W_n(b_i)\| > 2\epsilon\Big) \le 2l_n \exp\Big(-\frac{n\epsilon^2}{2BC + 2/3}\Big) \\ + l_n\mathbf{1}[2BC\rho > \epsilon].\end{aligned}$$

This and (4.45) yield the desired (4.44) by choosing $l_n = n^{k/2}$ and finish the proof. □

Before proving Theorem 2.4, we need the following result. Recall a n-variate process $\{N_1, \ldots, N_n\}$ is called a *multivariate counting process* if (i) Each $N_i, i = 1, \ldots, n$ is a counting process, and (ii) No two component processes jump at the same time. The following result generalizes Corolloary 3.4.1. of Fleming and Harrington [4] from $i = 1$ to $i = n$.

LEMMA 4.1. *Let $\{N_1, \ldots, N_n\}$ be a locally bounded multivariate counting process. Let $\{\mathcal{F}_t : t \leq 0\}$ be a right-continuous filtration such that for each i, $M_i = N_i - A_i$ is the corresponding local square-integrable martiginale with A_i the compensator process, and H_i is a locally bounded $\mathcal{F}_t$-predictable process. Then for any stopping time T such that $P(T < \infty) = 1$, and any $\epsilon, \eta > 0$,*

$$P\Big(\sup_{t \leq T}\Big(\sum_{i=1}^n \int_0^t H_i(s)\, dM_i(s)\Big)^2 \geq \epsilon\Big) \leq \frac{\eta}{\epsilon} + P\Big(\sum_{i=1}^n \int_0^T H_i^2(s)\, d\langle M_i, M_i\rangle(s) \geq \eta\Big)$$

SKETCHES OF PROOF. This can be proved similar to the proof of Corolloary 3.4.1. of Fleming and Harrington [4]. Let $\{\tau_k : k = 1, 2, \ldots\}$ be a localizing sequence such that, for any k, $N_i(\cdot \wedge \tau_k)$, $A_i(\cdot \wedge \tau_k)$ and $H_i(\cdot \wedge \tau_k)$ for $i = 1, \ldots, n$ are processes bounded by k, and $M_i(\cdot \wedge \tau_k)$ is a square-integrable martingale. Let $U = \sum_{i=1}^n \int H_i\, dM_i$. Then it follows from their Theorem 2.4.5 that U is a local square-integrable martingale and satisfies $E(U(t)) = 0$ and $\mathrm{Var}(U(t)) = \sum_{i=1}^n E(\int_0^t H_i^2\, d\langle M_i, M_i\rangle)$. Using this and similar to their proof of Corolloary 3.4.1. one can prove the result by replacing their X_k and Y_k with the following

$$X_k(t) = \Big(\sum_{i=1}^n \int_0^{t\wedge\tau_k} H_i(x)\, dM_i(s)\Big)^2, \; Y_k(t) = \sum_{i=1}^n \int_0^{t\wedge\tau_k} H_i^2(x)\, d\langle M_i, M_i\rangle(s).$$

□

PROOF OF THEOREM 2.4. It follows from Proposition 4.1 that (K) implies (A1) – (A3) in Theorem 4.1, hence there exists a unique ζ_n such that

$$1 + \zeta_n^\top u_j > 0, \quad j = 1, \ldots, n, \quad \frac{1}{n}\sum_{j=1}^n \frac{u_j}{1 + \zeta_n^\top u_j} = 0,$$

$$\|\zeta_n\| \leq \frac{\|\bar{u}_n\|}{\lambda_n - \|\bar{u}_n\| u_*}, \quad \|\zeta_n - S_n^{-1}\bar{u}_n\|^2 \leq c_0 \|\zeta_n\|^4 u_n^{(4)}, \tag{4.46}$$

on an event whose probability converges to one as n tends to infinity, where c_0 is a constant, Accordingly, $\ell_n(b)$ is well defined on this event (and defined to be an arbitrary number on the complement of this event which has an vanishing probability as n tends to infinity).

Since $Z(t)$ is bounded (by c_2 say), it follows that the $s_i(t), i = 0, 1, 2$ given in (2.9) are well defined and bounded for every $t \in [0, \tau]$. Note that

$P(Y(\tau) > 0) > 0$ implies $\inf_{t\in[0,\tau]} s_0(t) > 0$. Indeed, if it were 0 then it follows from $s_0(t) \geq \exp(-c_2\|\beta\|)P(X \geq t)$ that $P(X \geq \tau) = 0$ which contradicts $P(Y(\tau) > 0) > 0$. Thus, the $e(t)$ defined in (2.9) is bounded by some c_3. This, the boundedness of $Z(t)$, the square-integrability of $u(R)$ and $\int_0^\tau h_0(t)\,dt < \infty$ imply that $C(\mathscr{I}(Y, Z))$ given in (2.8) is well defined and finite.

Introduce the following two predictable processes

$$\bar{Z}_n(t,b) = \sum_{j=1} p_{nj}(t,b)Z_j(t),\ V_n(t,b) = \sum_{j=1}^{n} p_{nj}(t,b)(Z_j(t) - \bar{Z}_n(t,b))^{\otimes 2},$$

where $p_{nj}(t,b) = Y_j(t)\exp(b^\top Z_j(t))/S_n(t,b)$. Write $\bar{Z}_n(t) = \bar{Z}_n(t,\beta)$ and $V_n(t) = V_n(t,\beta)$. By the weak law of large numbers, we have

$$S_n(t,\beta)/n \xrightarrow{p} s_0(t), \quad \bar{Z}_n(t) \xrightarrow{p} e(t), \quad V_n(t) \xrightarrow{p} V(t),$$

where $V(t) = s_0(t)^{-1}s_2(t) - e(t)^{\otimes 2}$. By the standard argument (see e.g. [4]), $V(t)$ can be viewed as a variance-covariance matrix so that it is positive definite. Introduce the EL-weighted versions of the above two processes:

$$\bar{\mathbb{Z}}_n(t) = \sum_{j=1} \mathbb{p}_{nj}(t)Z_j(t), \quad \mathbb{V}_n(t) = \sum_{j=1}^{n} \mathbb{p}_{nj}(t)(Z_j(t) - \bar{\mathbb{Z}}_n(t))^{\otimes 2},$$

where $\mathbb{p}_{nj}(t) = (n\pi_{nj})Y_j(t)\exp(\beta^\top Z_j(t))/\mathbb{S}_n(t,\beta)$. Clearly, they are also predictable. An application of Lemma A2 of [6] with $w_j = \mathbb{p}_{nj}(t)Y_j(t)\exp(\beta^\top Z_j(t))$ and $a_j = a^\top Z_j(t)$ allows us to have an expansion for $\log \mathbb{S}_n(t, \beta + a)$ with $a \in \mathbb{R}^k$. The result is

$$\log \mathbb{S}_n(t,\beta + a) - \log \mathbb{S}_n(t,\beta) = a^\top \bar{\mathbb{Z}}_n(t) + 1/2a^\top \mathbb{V}_n(t)a + r_n(t,a), \tag{4.47}$$

where the remainder $r_n(t,a)$ has the property

$$|r_n(t,a)| \leq 4/3|a|^3 \max_{1\leq j\leq n} \|Z_j(t) - \bar{Z}_n(t)\|^3. \tag{4.48}$$

The limit behavior of $\hat{\beta}_n$ can be derived from the study of the following ℓ_n^*. To this end, we use (4.47) to obtain its two-term Taylor expansion,

$$\begin{aligned}
\ell_n^*(a) &:= n\big(\ell_n(\beta + n^{-1/2}a) - \ell_n(\beta)\big) \\
&= \sum_{j=1}^{n} n\pi_{nj} \int_0^\tau \Big(n^{-1/2}a^\top(Z_j(t) - \bar{\mathbb{Z}}_n(t)) \\
&\qquad\qquad - 1/2n^{-1}a^\top \mathbb{V}_n(t)a - r_n(t, n^{-1/2}a)\Big)\,dN_j(t) \\
&:= a^\top \mathbb{U}_n - 1/2a^\top \mathbb{J}_n^* a - r_n(a),
\end{aligned}$$

where

$$\mathbb{U}_n = n^{1/2} \sum_{j=1}^n \pi_{nj} \int_0^\tau (Z_j(t) - \bar{\mathbb{Z}}_n(t))\, dN_j(t),$$

$$\mathbb{J}_n^* = \sum_{j=1}^n \pi_{nj} \int_0^\tau \mathbb{V}_n(t)\, dN_j(t), \quad r_n(a) = \sum_{j=1}^n n\pi_{nj} \int_0^\tau r_n(t, n^{-1/2}a)\, dN_j(t).$$

One observes that $\ell_n^*(a)$ is maximized at $\hat{a} = \sqrt{n}(\hat{\beta}_n - \beta)$. Since $W_u = E(u(R)u(R)^\top)$ is positive definite and $S_n = n^{-1}\sum_{j=1}^n u(R_j)u(R_j)^\top \xrightarrow{p} W_u$, it follows that $\lambda_{\min}(S_n) = \lambda_n \geq \lambda_0$ for some constant $\lambda_0 > 0$. Noticing (K) implies $u_n^*\|\zeta_n\| = o_p(1)$, hence the first inequality in (4.46) implies $n\pi_{nj} \leq \frac{1}{\lambda_0 - u_n^*\|\zeta_n\|} \leq c_1$ for some constant $c_1 > 0$. Since $Z_j(t), j = 1, 2, \ldots$ are uniformly bounded processes (bounded by c_2), it follows, in view of (4.48), that

$$|r_n(a)| \leq c_1 \sum_{j=1}^n \int_0^\tau 4/3|a|^3(2c_2)^3 n^{-3/2}\, dN_j(t) = O(n^{-1/2}).$$

Therefore $r_n(a) = o_p(1)$ for every finite a. Recall that it is shown in the Introduction that ℓ_n hence ℓ_n^* is convex. Using the convex argument (e.g. Hjort and Pollard (1993)), the desired result now follows from

$$\mathbb{U}_n \implies \mathscr{N}(0, \Sigma_1), \quad \text{and} \tag{4.49}$$

$$\mathbb{J}_n^* \xrightarrow{p} J = \int_0^\tau J(t) h_0(t)\, dt, \tag{4.50}$$

which are shown below. Here $\Sigma_1 = J - C(\mathscr{I}(Z,Y))W_u^{-1}C(\mathscr{I}(Z,Y))^\top$ and $J(t) = s_0(t)V(t) = s_2(t) - s_1(t)^{\otimes 2}/s_0(t)$. Note first that by Proposition 4.1 and (4.3) we have

$$\begin{aligned}\sum_{j=1}^n \pi_{nj} \int_0^\tau Z_j(t) dN_j(t) &= \frac{1}{n}\sum_{j=1}^n \int_0^\tau Z_j(t)\, dN_j(t) \\ &\quad - \frac{1}{n}\sum_{j=1}^n \int_0^\tau Z_j(t)\, dN_j(t) u_j^\top \zeta_n + r_n,\end{aligned} \tag{4.51}$$

where r_n is the remainder term whose square is bounded by the right hand side of (4.3). Hence by (4.8),

$$\|r_n\|^2 \leq o_p(n^{-1}) \|\frac{1}{n}\sum_{j=1}^n \int_0^\tau Z_j\, dN_j\|^2 = o_p(n^{-1}). \tag{4.52}$$

Denote the i-the component of Z_j by $Z_{j,i}$ for $i = 1, \ldots, k$. By Lemma 4.1, for any $\eta > 0, \epsilon > 0$ we have

$$P\Big(\Big\|\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dM_j(t)\Big\|^2 \geq \epsilon\Big) \leq \sum_{i=1}^{k} P\Big(\Big(\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_{j,i}(t)\,dM_j(t)\Big)^2 \geq \epsilon/k\Big)$$
$$\leq \frac{k^2\eta}{\epsilon} + \sum_{i=1}^{k} P\Big(\frac{1}{n^2}\sum_{j=1}^{n}\int_0^\tau Z_{j,i}^2(t)\,d\langle M_j, M_j\rangle(t) \geq \eta\Big).$$

By taking $\eta = \epsilon^2$ and in view of the uniform boundedness of $Z_j(t)$, the last sum tends to zero as n goes to infinity, hence letting $\epsilon \to 0$ gives

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dM_j(t) = o_p(1).$$

Analogously,

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dM_j(t) \otimes u_j = o_p(1),$$

where $\otimes$ is the Kronecker product. Thus using the decomposition $dN_j = dM_j + dA_j$ and the law of large number we derive

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dN_j(t) \otimes u_j = \frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dA_j(t) \otimes u_j + o_p(1)$$
$$= E\big(\int_0^\tau Z(t)Y(t)\exp(\beta^\top Z(t))h_0(t)\,dt \otimes u(R)\big) + o_p(1).$$

This, (4.51), (4.52), the two inequalities in (4.2), (A1) and (4.8) yield

$$\begin{aligned}\sum_{j=1}^{n}\pi_{nj}\int_0^\tau Z_j(t)dN_j(t) &= \frac{1}{n}\sum_{j=1}^{n}\int_0^\tau Z_j(t)\,dN_j(t)\\ &- E\big(\int_0^\tau Z(t)Y(t)\exp(\beta^\top Z(t))h_0(t)\,dtu(R)^\top\big)W_u^{-1}\bar{u}_n + o_p(n^{-1/2}).\end{aligned} \tag{4.53}$$

Analogously, it follows from Proposition 4.1 and Theorem 4.1 that

$$\mathbb{S}_n(t,\beta) = S_n(t,\beta) + O_p(n^{1/2}), \quad \text{and} \tag{4.54}$$

$$\bar{\mathbb{Z}}_n(t) = \bar{Z}_n(t) + O_p(n^{-1/2}), \tag{4.55}$$

uniformly in $t \in [0, \tau]$. Thus, recalling (2.8) and observing that $R_{nj} = \int_0^\tau \bar{\mathbb{Z}}_n(t)\, dN_j(t)$ is square-integrable, we apply (4.3) to get

$$\begin{aligned} \sum_{j=1}^n \pi_{nj} \int_0^\tau \bar{\mathbb{Z}}_n(t)\, dN_j(t) = \int_0^\tau \bar{Z}_n(t) \frac{d\bar{N}_n(t)}{n} \\ - \frac{1}{n} \sum_{j=1}^n \int_0^\tau \bar{Z}_n(t)\, dN_j(t) u_j^\top \zeta_n + r_n, \end{aligned} \tag{4.56}$$

where the remainder satisfies

$$r_n = \|\zeta_n\|^4 u_n^{(4)} \|\big(n^{-1} \int_0^\tau \bar{Z}_n(t)\, d\bar{N}_n(t)\big)^{\otimes 2}\| = o_p(n^{-1/2}).$$

It is easily seen that

$$\bar{Z}_n(t) = e(t) + o_p(1). \tag{4.57}$$

By the second inequality of (4.46), (4.49)-(4.53) and (4.56)-(4.57), we achieve

$$\mathbb{U}_n = n^{-1/2} \sum_{j=1}^n \int_0^\tau (Z_j(t) - \bar{Z}_n(t))\, dN_j(t) - C(\mathscr{I}(Y, Z)) W_u^{-1} n^{1/2} \bar{u}_n + o_p(1).$$

It is a standard result that the first term on the right side of the above display converges in distribution of $\mathscr{N}(0, J)$, see e.g. Theorem 8.2.1. of Fleming and Harrington [4], while the second term converges in distribution $\mathscr{N}(0, C(\mathscr{I}(Z, Y)) W_u^{-1} C(\mathscr{I}(Z, Y))^\top)$. Since the second term is (asymptotically) the projection of the first term onto the closed linear subspace spanned by $u_1, \ldots, u_m$, it follows $\mathbb{U}_n$ satisfies (4.49). We are now left to prove (4.50). Analogous to (4.54), one obtains

$$\sum_{j=1}^n \mathbb{p}_{nj} Z_j(t)^{\otimes 2} = \sum_{j=1}^n p_{nj} Z_j(t)^{\otimes 2} + O_p(n^{-1/2}), \quad \text{and}$$

$$\mathbb{V}_n(t) = \sum_{j=1}^n \mathbb{p}_{nj} Z_j(t)^{\otimes 2} - \bar{\mathbb{Z}}_n(t)^{\otimes 2} = V_n(t) + o_p(1).$$

Accordingly,

$$\mathbb{J}_n^* = \sum_{j=1}^n \pi_{nj} \int_0^\tau V_n(t)\, dN_j(t) + o_p(1) = n^{-1} \int_0^\tau V_n(t) d\bar{N}_n(t) + o_p(1) = J + o_p(1),$$

where the last equality uses (i) of [6]. This proves (4.50) and completes the proof. □

References.

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

[2] Bravo, F. (2010). Efficient M-estimators with auxiliary information. *J. Statist. Plan. Infer.* **140** 11: 3326–3342.

[3] Tang, C. Y. and Leng, C. (2012). An empirical likelihood approach to quantile regression with auxiliary information. *Statist. & Probabil. Lett.* **82**, 29–36.

[4] Fleming, T.R. and Harrington, D.P. (2005). *Counting processes and survival analysis.* John Wiley & Sons, Inc.

[5] Hellerstein, J. and Imbens, G. W. (1999). Imposing moment restrictions from auxiliary data by weighting. *Review of Economics and Statistics* **81**: 1–14.

[6] Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. *Statistical Research Report.*

[7] Lu, X. and Tsiatis, A.A. (2008). Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika* **95**, 679–694.

[8] Müller, U.R. and Keilegom, V.I. (2012). Efficient quantile regression with auxiliary information. *Electron. J. Statist.* **6**, 1200–1219.

[9] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

[10] Owen, A. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, London.

[11] Peng, H. (2014). Efficient estimation in models with side information: an easy empirical likelihood approach. The *Manuscript* can be found at the URL: `http://www.math.iupui.edu/~hpeng/preprint.html.`

[12] Peng, H. and Schick, A. (2013). Maximum empirical likelihood estimation and related topics. The *Manuscript* can be found at the URL: `http://www.math.binghamton.edu/anton/preprint.html.`

[13] Peng, H. and Schick, A. (2012). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**, 954–981

[14] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 30–325.

[15] Vaart, van der A.W. (2000). *Asymptotic Statistics.* Cambridge University Press.

[16] Yuan, A., He, W., Wang, B. and Qin, G. (2012). U-statistics with side information. *J. Multiv. Analy.* **111**, 20 – 38.

[17] Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Infer.* **44**, 77 – 94.

[18] Zhang, B. (1997). Quantile processes in the presence of auxiliary information. *Ann. Inst. Statist. Math.* **49**, 35 – 55.

Indiana University Purdue University
Indianapolis
Department of Mathematical Sciences
Indianapolis, IN 46202-3216
USA
E-mail: hpeng@math.iupui.edu