IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 197–212

DOI: 10.1214/12-IMSCOLL914

Efficient testing and estimation in two Lehmann alternatives to symmetry-at-zero models

W. J. Hall[1,*] **and Jon A. Wellner**[2,†]

University of Rochester and University of Washington

Abstract: We consider two variations on a Lehmann alternatives to symmetry-at-zero semiparametric model, with a real parameter θ quantifying skewness and a symmetric-at-0 distribution as a nuisance function. We show that a test of symmetry based on the signed log-rank statistic [A signed log-rank test of symmetry at zero (2011) University of Rochester] is asymptotically efficient in these models, derive its properties under local alternatives and present efficiency results relative to other signed-rank tests. We develop efficient estimation of the primary parameter in each model, using model-specific estimates of the nuisance function, and provide a method for choosing between the two models. All inference methods proposed are based solely on the signed ranks of the absolute values of the observations, the invariantly sufficient statistic. A simulation study is summarized and an example presented. Extensions to regression modeling are envisaged.

1. Introduction and summary

The simplest test of the null hypothesis of symmetry at 0 is the *sign test*. Additional nonparametric choices are provided by *signed rank tests*, again without other assumed structure, starting with the path-breaking signed rank test of Wilcoxon [20], and followed by a series of others—including *signed normal scores* and various robust and adaptive variations; see books such as [8, 10, 15], and [11] and references therein. Another is a signed rank test of [13] (see Section 3). Other tests of symmetry at 0 include those of Kolmogorov type, but not having the convenience of an asymptotically normal test statistic (and not considered here).

These tests are especially suitable for use with paired data differences—e.g., a treatment and placebo administered on the same subjects, or measures on both left and right sides of a body, or twin data—where a null hypothesis of symmetry at 0 is often appropriate. Other common applications are to testing for a median of zero in settings in which symmetry may be a reasonable additional assumption (but without this extra assumption, signed rank tests of median = 0 need not be consistent).

*Research supported by NHLBI grant 5R01 HL58751.
†Research supported in part by NSF grant DMS-0804587 and NI-AID grant 2R01 AI291968-04.
[1]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642-8630, e-mail: hall@bst.rochester.edu
[2]Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, e-mail: jaw@stat.washington.edu, url: http://www.stat.washington.edu/jaw/
AMS 2000 subject classifications: Primary 62G07, 62H12; secondary 62G05, 62G20
Keywords and phrases: Asymptotically uniformly most powerful test, signed log-rank test, proportional hazards, reversed hazards, semiparametric models, information bound

Although these tests are generally consistent against all stochastically ordered alternatives to symmetry at 0, evaluations have been restricted largely to shift alternatives whereas applications often show skewness. Asymptotic properties of signed rank tests in shift models are shape-dependent; for example, different tests are efficient against logistic (Wilcoxon) and normal (normal scores) shapes.

Only in semiparametric (or fully parametric) models is there an associated estimation problem, and essentially the only such models studied in the literature are shift models. For these, the center of symmetry is the real parameter of the model, and rank-based Hodges-Lehmann estimation methods [14] provide solutions. Again, efficiency varies with the shape of the distribution.

In a recent paper [9], a new signed rank test was introduced, aimed at skewed alternatives to symmetry at 0, and evaluated in two Lehmann alternatives models. Here, we show it to be asymptotically optimal in these semiparametric models, and we evaluate the asymptotic relative efficiency of other signed rank tests in these models. We go on to develop and evaluate efficient estimation of the associated real parameter, which has both skewness and hazard interpretations. Efficiency of tests and of estimation in these models are shape-free.

Note: There are tests for symmetry versus asymmetry without specifying a null center of symmetry, and also tests that assume a specified median and evaluate whether variation around that median is or is not symmetric; such tests are outside the context considered here.

2. Two Lehmann alternatives to symmetry-at-zero models

Consider a random sample $(X_1, \ldots, X_n)$ with an absolutely continuous distribution function (df) on $\mathbb{R}$. We confine attention to two related semiparametric *Lehmann alternatives to symmetry-at-0* models considered in [9], the first defined by the distribution function

$$F_{\theta,F}(x) = F(x)^{\theta} \tag{1}$$

and the second by the survival function

$$G_{\theta,F}(x) = G(x)^{1/\theta} \tag{2}$$

with θ belonging to an interval of positive reals including $\theta = 1$ and $F \equiv 1 - G \in \mathcal{F}_0$, the set of absolutely continuous d.f.'s symmetric at 0; write f for the density. Hence, $F(-x) = G(x)$ for all x. Our focus is on the parameter θ, with F a nuisance function. The models are related; specifically, the survival function for $Y \equiv -X$ when X follows Model (1) is given by (2) with θ replaced by $1/\theta$. Plots of densities for the models for several choices of F and θ appear in [9].

It is readily verified that the right tail area in each of Models (1) and (2) is increasing in θ, so θ quantifies the skewness within each family—to the left for $\theta < 1$ and to the right for $\theta > 1$. As a consequence, the medians—indeed, all quantiles—and the means (if existent) are increasing in θ.

Each of (1) and (2) is a *Lehmann alternatives family* [14], but here with a symmetric-at-0 baseline F. We will refer to these models as *LAS1* and *LAS2*, respectively. LAS2 is the popular *proportional hazards family* [5] of survival analysis while LAS1 is the *proportional reversed hazards family* ([6] and [7]), introduced in recent years in the reliability literature. But these latter two families are only defined on $\mathbb{R}^+$, where hazards concepts have more direct interpretation. Still, the reverse hazard at x is the density at x given $\{X \le x\}$, in contrast to the (forward) hazard, namely the density at x given $\{X \ge x\}$. Hence, θ is essentially the proportional-hazards parameter of the popular Cox models of survival analysis, but here with a restricted interpretation of 'hazard'. The distribution F is associated with 'no effect' and plays the role played by a baseline or control-group distribution in survival analysis settings.

An alternative parametrization of (1) and (2) replaces θ by e^ρ, common in proportional hazards modeling, and possibly more appropriate for estimation and confidence interval construction.

It is easily determined that the score for ρ in LAS2—not the efficient score but the "raw" or "ordinary" score in the semiparametric model—is $-1 + \exp(-\rho)\Lambda$ ($\Lambda = -\log G$) with information for ρ equal to 1. In LAS1, information for ρ is also equal to 1. But ρ and F are not orthogonal in these models, as is established in the Appendix. Specifically, we find in these one-sample models that the resulting information for ρ due to not knowing F is ≈ 0.8225, representing a 17.75% loss of information; see Section 5.

By contrast, in a two-sample PH model (without censoring)—with $X \sim 1 - G^{\exp(-\rho)}$, $Y \sim 1 - G^{\exp(\rho)}$, $X \perp Y$, $G = 1 - F$—it may be shown (omitted here) that the score for ρ in the model for (X, Y) is $\exp(-\rho)\Lambda(x) - \exp(\rho)\Lambda(y)$ with information for ρ equal to 2. Moreover, in this 2-sample model, ρ and F are orthogonal, so the score for ρ is the efficient score and there is no loss of information. Partial likelihood methods—not applicable in the models considered here—facilitate efficient inference.

The LAS models may be appropriate for modeling differences between paired measurements, say before and after treatment, with θ quantifying the treatment effect with null value unity. Exchangeability of the paired measurements under a null hypothesis of no treatment effect implies that $F \in \mathcal{F}_0$, while an effective treatment would often induce skewness. In the literature, focus has largely been on shift models $F(x - \theta)$, with symmetry preserved for all θ. Models (1) and (2) could easily accommodate regression, replacing ρ by $\beta' z$, say; see Section 8.

3. Efficient testing

We focus on the null hypothesis of symmetry. [9] introduced a linear signed-rank test, the *signed log-rank test SL*, with LAS models in mind. There $\theta = 1$ ($\rho = 0$) is consistent with the nonparametric null hypothesis that the df is in $\mathcal{F}_0$. Let $Z_i = \operatorname{sign} X_i$, $Y_i = |X_i|$, and let R_i^+ be the rank of Y_i among $(Y_1, \ldots, Y_n)$. The test SL is defined by the test statistic

$$SL_n = \sum_{i=1}^n Z_i\, a_n\big(R_i^+\big) = \sum_{i=1}^n a_n\big(Z_i R_i^+\big), \quad \text{where} \quad a_n(j) = \frac{1}{2}\log\left(\frac{n+1+j}{n+1-j}\right). \tag{3}$$

Its null distribution is symmetric around 0 with variance $\sigma_n^2 = \sum_{j=1}^n a_n(j)^2 \approx n\pi^2/12$. The test rejects in favor of stochastically larger alternatives ($\theta > 1$ in Models (1) and (2)) whenever SL_n, or its standardized form $SL_n^* = SL_n/\sigma_n$, is sufficiently large. (The exact standard error σ_n is strongly recommended.) Its null distribution is asymptotically normal, with Edgeworth corrections provided in [9].

Moreover, it was shown there to be a *locally most powerful signed-rank test* in Models (1), (2) and their union, and consistent against stochastically-ordered alternatives. Here we strengthen this by proving it to be *asymptotically uniformly most powerful*, as defined in [4], and give its power under local alternatives, along with ARE's relative to various other linear signed-rank tests. We follow the methodology there.

First, for comparison, the popular Wilcoxon signed rank test statistic W_n may be written as in (3) with $a_n(j) = j/(n+1)$, and likewise for the signed normal scores test statistic NS_n with $a_n(j) = \Phi^{-1}(\frac{1}{2}(n+1+j)/(n+1))$. (We choose the van der Waerden form of this test, first introduced by [19].) Koziol's signed rank test K_n [13] is equivalently based on $a_n(j) = \sqrt{2}\sin(\frac{\pi}{2}j/(n+1))$ and the sign test S_n on $a_n(j) = 1$ (here expressed as a signed rank test). Null distributions, variances and Edgeworth corrections are given by the same formulas (see [9] for the Edgeworth correction formulas); the asymptotic variance of W_n is $1/3$ while that for the others is unity. Optimality for W, NS and S is against F-specific shift alternatives, with F logistic for W, normal for NS and double-exponential (Laplace) for S. (The supports of W and S (in this form) being a limited number of values spaced 2 units apart, these two statistics should be moved one unit closer to zero as a continuity correction; the others have non-lattice, and much richer, supports.)

Returning to the SL test, and following [4], we consider local alternatives to the null hypothesis $\theta = 1$, namely $\theta_n = 1 + (h_\theta + o(1))/\sqrt{n}$ and $F_n(x) = \int_{-\infty}^x f_n(y)\,dy$ with $\sqrt{f}_n = \sqrt{f} + (h_f + \delta_n)/\sqrt{n}$ and $\|\delta_n\| = o(1)$; here, symmetry of f_n implies symmetry of the $\mathcal{L}_2$-function h_f. For one-sided local alternatives, $h_\theta > 0$.

We first derive the *effective (efficient) score*, depending on the nuisance $\mathcal{L}_2$-function $\sqrt{f}$ through F, or equivalently through the df $F^+ = 2F - 1$ on $\mathbb{R}^+$ for $|X|$.

Proposition 1. *The effective score for θ at $\theta = 1$ in the Lehmann alternatives Models (1) and (2) is*

$$s^*(x,F) \;=\; \frac{1}{2}\log\frac{F(x)}{G(x)} \;=\; \operatorname{sign}(x)\cdot\frac{1}{2}\log\frac{1+F^+(|x|)}{1-F^+(|x|)}; \tag{4}$$

the effective information, its variance, is $I^* = \pi^2/12 \approx 0.822467$.

Proof. We assume Model (1). As in the examples in [4], we first find the *directional score* (per observation) in direction (h_θ, h_f). Writing $g_{\theta,f}$ for the density, this score is (ignoring $o(1)$ terms)

$$s(x;h_\theta,h_f) = \sqrt{n}\,\log\left[\frac{g_{\theta_n,f_n}(x)}{g_{1,f}(x)}\right] = \big[1+\log F(x)\big]h_\theta + h'(x) \tag{5}$$

with $h'(x) = 2\,h_f(x)/\sqrt{f(x)} = h'(-x)$. The *directional information* (variance of s) is

$$\begin{aligned}
I(h_\theta,h_f) &= \int_{-\infty}^{\infty}\big\{\big[1+\log F(x)\big]h_\theta + h'(x)\big\}^2\,dF(x)\\
&= \int_0^\infty \big\{\big[(h'+h_\theta)+(\log F)h_\theta\big]^2 + \big[(h'+h_\theta)+(\log G)h_\theta\big]^2\big\}\,dF\\
&= \int_0^\infty \left\{2\left[(h'+h_\theta)+\frac{1}{2}(\log F+\log G)h_\theta\right]^2 + \frac{1}{2}[\log F-\log G]^2h_\theta^2\right\}dF.
\end{aligned}$$

The terms in square brackets in the last expression are orthogonal, and so minimization of I is achieved by choosing direction h_f for which $h' = -[1 + \frac{1}{2}\log(F\,G)]h_\theta$. Substituting in (5) and in I, we find the effective score as claimed (first form in (4)) and the effective information reduces to $I^* = \frac{1}{4}\int_{-\infty}^{\infty}\log^2(F/G)\,dF$. This integral clearly does not depend on F, so choose the standard logistic distribution function $F_L(x) = 1/(1+\exp(-x))$, with variance $\pi^2/3$, yielding $I^* = \frac{1}{4}\int x^2\,dF_L(x) = \pi^2/12$. The second form for s^* in (4) follows from the first. Verification in Model (2) is similar; or, the effective score for (2) is seen to be $-s^*(\cdot, G) = s^*(\cdot, F)$. $\square$

Let TL_n be the sum of the effective scores (4), which is $AN(0, nI^*)$ (asymptotically normal). To obtain an AE test, we need to replace $s^*(x, F)$ by an estimate, say $\hat{s}^*(x)$, for which the sum will equal $TL_n + o_p(\sqrt{n})$. To this end, write $\hat{s}^*(x) = \mathrm{sign}(x)\cdot\psi^+(F^+(|x|))$ where $\psi^+(u) = \frac{1}{2}\log[(1+u)/(1-u)] = \tanh^{-1}(u)$; note that ψ^+ integrates to 0 with an integrated square of $\pi^2/12$. In $\hat{s}^*(x_i)$, estimate $F^+(|x_i|)$ by $R_i^+/(n+1)$, resulting in $\sum \hat{s}^*(X_i) = SL_n$.

We now need to show the asymptotic equivalence of SL_n and TL_n. Using Lemma V.1.6.a of [8], we find from their Theorem V.1.7 and its proof the following (with o_{ms} indicating smaller order in mean-square):

Proposition 2. *$TL_n - SL_n = o_{ms}(\sqrt{n})$, and SL_n is $AN(0, nI^*)$ and $AN(0, \sigma_n^2)$.*

It follows, applying *LeCam's Third Lemma* to TL, that SL^* is asymptotically $N(\sqrt{I^*}\,h_\theta, 1)$ under local alternatives, to $\theta = 1$ and F—just as in [8] except now for Lehmann alternatives rather than the shift alternatives considered there and with F locally varying. We thus conclude:

Corollary. *Let Φ be the standard normal df, and define z_α by $\bar{\Phi}(z_\alpha) = \alpha$. The signed log-rank test, rejecting $\theta = 1$ in favor of $\theta > 1$ ($\rho > 0$) in Model (1) or (2) whenever $SL^* \geq z_\alpha$, is AUMP(α). Its local power under local alternatives with $\theta_n = 1 + (h_\theta + o(1))/\sqrt{n}$ is $\Phi(\sqrt{I^*}\,h_\theta - z_\alpha)$.*

The power at a particular θ_n near unity—that is, $\rho_n \equiv \log\theta_n$ near 0—may be estimated by $\Phi(\sigma_n\rho_n - z_\alpha)$. Similarly, a two-sided version of *SL*, rejecting for large $|SL_n|$, is *asymptotically uniformly most powerful unbiased* for (1), (2) and their union; see [4].

Following [19], the asymptotic relative efficiency (ARE) between two signed rank tests—based on scores a_n and a'_n, say—when one is asymptotically efficient is given by

$$\lim_{n\to\infty}\left\{\left[\sum_{j=1}^{n} a_n(j)\,a'_n(j)\right]^2 \Big/ \sum_{j=1}^{n} a_n(j)^2 \sum_{j=1}^{n} a'_n(j)^2\right\} \tag{6}$$

(or with sums replaced by integrals du with $u = j/n$). We find when a LAS model is correct that *ARE(NS,SL)* $\approx 99.2\%$, *ARE(W,SL)* $= 9/\pi^2 \approx 91.2\%$, *ARE(K,SL)* $\approx 84.5\%$ and *ARE(S,SL)* $= 12(\log 2/\pi)^2 \approx 58.4\%$. Each of these also has an interpretation when the other test considered is efficient. Hence, there is little to choose between SL and NS when either a LAS model or a normal shift model is correct, and—since it is well-known that $ARE(W, NS) = 3/\pi \approx 95.5\%$ under normal shift alternatives—with W not very far behind. Although these comparisons are asymptotic—and all limits in (6) were found to be slowly decreasing as n increased—simulation studies (Section 7) confirm that the ordering *SL* > *NS* > *W* in LAS models is maintained, even for small n.

To test other null hypotheses, say $\rho = \rho_o$ vs. $> \rho_o$, the appropriate effective scores are given by the ρ version of the scores in Proposition 3 in Section 5. To construct a test, replace F in the efficient score by a consistent estimate (see Section 4), sum over the x's, and treat the resulting statistic as $N(0, n\pi^2/12)$ under the null hypothesis and $AN(\sqrt{n}(\rho_n - \rho_o), n\pi^2/12)$ under local alternatives for which $|\rho_n - \rho_o|$ and $\|\sqrt{f_n} - \sqrt{f}\|$ are $O(1/\sqrt{n})$, or, estimate the variance of the score test statistic by the sum of squares of the scores or by the 'null variance', as noted in Section 6.

4. Estimating F

We assume LAS1 and start by acting as if θ is known. Write $\mathbb{F}_n$ for the empirical df of the data and $\mathbb{G}_n = 1 - \mathbb{F}_n$ for the empirical survival function; we use the form $\mathbb{F}_n(x) = \#\{i|x_i \le x\}/(n+1)$, $\in (0,1)$ at the observations. Then an obvious consistent estimate for F in Model (1) is $\tilde{F}_1 \equiv \mathbb{F}_n^{1/\theta}$. Another is obtained by noting that $\tilde{F}_1(-x)$ estimates $F(-x) = G(x)$. Hence, $\tilde{F}_2(x) \equiv 1 - \tilde{F}_1(-x) = \tilde{G}_1(-x)$ is also consistent for F, as is any weighted average $\tilde{F}$ of $\tilde{F}_1$ and $\tilde{F}_2$. To assure symmetry at 0, use weights adding to unity: $p(x) + p(-x) \equiv 1$. Then, for $x > 0$ and θ specified, F may be estimated by

$$\tilde{F}_n(x;\theta) = p(x)\,\mathbb{F}_n(x)^{1/\theta} + p(-x)\left[1 - \mathbb{F}_n(-x)^{1/\theta}\right]. \tag{7}$$

(If such a weighted $\tilde{F}_n$ is not everywhere monotone, it can be adjusted by starting at $\tilde{F}_n(0) = \frac{1}{2}$ and proceeding with estimation at successive positive x-values, forcing monotonicity by never allowing any decrease.) Note that p is any function from $\mathbb{R}$ to $(0,1)$ for which $p(-x) = 1 - p(x)$. The choice $p(x) \equiv \frac{1}{2}$ is a good choice in the neighborhood of $\theta = 1$ in that it minimizes the asymptotic variance of $\tilde{F}_n(x)$ there, but other choices may be more suitable for more distant θ-values since, when $\theta > 1$, there tend to be more positive observations than negative ones, and hence the first term in (7) would deserve higher weight than the second, and vice versa. A variational argument (see Appendix) shows that the choice of p which minimizes the pointwise variance of $\tilde{F}_n$ is

$$p_0(\cdot,\theta,F) = \frac{G^{2-\theta}(1-G^\theta) + F^{1-\theta}G - FG}{F^{2-\theta}(1-F^\theta) + G^{2-\theta}(1-G^\theta) + 2\,F^{1-\theta}G - 2\,FG} \quad \text{for } x > 0. \tag{8}$$

The asymptotic variance of $\tilde{F}_n$ in (7) is given in the Appendix; when $p \equiv \frac{1}{2}$, it is

$$(1/(4\theta^2))\{F^{2-\theta}(1-F^\theta) + G^{2-\theta}(1-G^\theta) - 2\,F^{1-\theta}G - 2\,FG\}$$

(and interchange F and G for $x < 0$). Note that when $\theta = 1$ this variance reduces to $\frac{1}{2}(F \vee G)|F - G|$ which is seen to be in agreement with [17]; see e.g. [18], page 746. The $\sqrt{n}$-consistency of (7) for F is sufficient for our needs.

For LAS2, the corresponding estimator of $G = 1 - F$ is given by

$$\tilde{G}_n(x;\theta) = p(-x)\mathbb{G}_n^\theta(x) + p(x)\left[1 - \mathbb{G}_n(-x)^\theta\right].$$

for any $p(x) + p(-x) = 1$ where $\mathbb{G}_n \equiv 1 - \mathbb{F}_n$. The same weighting (8) again minimizes the pointwise variance.

To estimate F when θ is unspecified or unknown, replace θ in (7) by a $\sqrt{n}$-consistent estimate; see Section 5.

5. Estimating θ

Efficient scores and *information bounds* for estimation of θ are not so easily derived as in the testing case of Section 3; derivations appear in the Appendix, leading to

Proposition 3. *The efficient (effective) score for θ, with subscript (m) for Model (m), is*

$$s^*_{(1)}(\cdot,\theta,F) = \frac{1}{2}(\Lambda/\theta - G\,\Lambda - F\,\overline{\Lambda})/F \;=\; \frac{1}{2}\left\{(\log(F/G) + \frac{\theta-1}{\theta}\frac{\log G}{F}\right\}, \tag{9}$$

$$s^*_{(2)}(\cdot,\theta,F) = -\frac{1}{2}(\theta\overline{\Lambda} - F\,\overline{\Lambda} - G\,\Lambda)/G \;=\; \frac{1}{2}\left\{\log(F/G) + (\theta-1)\frac{\log F}{G}\right\}, \tag{10}$$

where $\Lambda = -\log G$ is the cumulative hazard for F and $\overline{\Lambda} = -\log F$ is the cumulative reversed hazard. The respective information bounds for estimation of θ are $I^*_{(1)}(\theta) = \pi^2/(12\theta^2)$ and $I^*_{(2)}(\theta) = \pi^2\theta^2/12$. For estimation of $\rho = \log\theta$, multiply $s^*_{(1)}$ by θ and $s^*_{(2)}$ by $-1/\theta$, with information $= \pi^2/12 \approx 0.822467$ in each case.

Note that the first term in each of the score formulas, common to both, has the sign of the argument x while each second term has the sign of $(1-\theta)$; the two effects apparently balance in expectation since the scores have expectation zero. When $\theta = 1$, both formulas reduce to $\frac{1}{2}(\Lambda - \overline{\Lambda}) = \frac{1}{2}\log(F/G)$, in agreement with (4); the information is likewise in agreement.

For estimation, it may be best to convert to $\rho = \log\theta$ since it has an unrestricted range and the corresponding information (and asymptotic variance for an efficient estimate) is parameter-free. For comparison, the information for ρ when F is completely known is readily found to be unity, so there is a 17.75% loss in information when F has to be estimated.

Summing the efficient scores leads to the full-sample efficient score $S^*_{(m)}(\underline{\mathrm{x}},\theta,F)$ $(m = 1,2)$. To find an efficient estimate $\hat{\theta}$ of θ, we need to find θ and the resulting $\tilde{F}_n$ for which this total score is 0. Starting from an initial $\sqrt{n}$-consistent $\tilde{\theta}_0$, it is tempting to cycle through computation of $\tilde{F}_n$ from (7) and then find the root $\tilde{\theta}$ of the total score—simple, since 'total score $= 0$' yields a linear equation in θ—and repeat until adequate convergence; but convergence does not generally occur. Instead, we solve the score equation $S^*_n(\underline{\mathrm{x}},\theta,\tilde{F}_n(\theta)) = 0$ for θ computationally, starting from an initial $\tilde{\theta}_0$. The resulting solution is the efficient estimate $\hat{\theta}_n$, and $\hat{\rho}_n = \log\hat{\theta}_n$. Hence, $\sqrt{n}(\hat{\rho}_n - \rho)$ is $AN(0, 12/\pi^2)$ $(12/\pi^2 \approx 1.2159)$, but other variance estimates may be more suitable when sample sizes are moderate. Further details are given in the next section.

6. Calculating estimates; Fitting the model

For an initial estimate of θ in Model (1), the fact that $F \approx \mathbb{F}_n^{1/\theta}$ should be $\frac{1}{2}$ at $x = 0$ suggests the estimate $\tilde{\theta}_0 = \log[(n+1)/(n^- + \frac{1}{2})]/\log 2$ with n^- the number of negative observations. This estimate is $\sqrt{n}$-consistent, indeed $AN(\theta, [(2^\theta - 1)/\log^2 2]/n)$. For Model (2), estimate $1/\theta$ by the same formula but with n^- replaced by n^+, the number of positive observations. However, we will focus on LAS1, using the sign-change approach to deal with LAS2.

We use the following algorithm for fitting these models.

Step 1: Order the $|x|$'s as $0 < y_1 < \cdots < y_n$, carrying along the signs of the corresponding x's, so the reordered data are represented as $x_r = z_r y_r$ with r the rank of $|x_r|$. Then calculate the empirical (edf) at $\pm x_r$ for each r as $\mathbb{F}_n(w) = \#\{r|x_r \leq w\}/(n+1)$. These $2n$ numbers may be represented as functions of the signed ranks zr, but it is more convenient to retain the edf notation; everything that follows depends on the data only through these $2n$ numbers. Also, let $\mathbb{F}_n(0) = \frac{1}{2}[\mathbb{F}_n(-y_1) + \mathbb{F}_n(y_1)]$; we use this in $\tilde{\theta}_0$ instead of $(n^- + \frac{1}{2})/(n+1)$.

Step 2: Calculate $\tilde{\theta}_0$ and a preliminary estimate of F (at the $2n$ $\pm x_i$'s) as $\tilde{F}_n(\cdot;\tilde{\theta}_0)$ from (7) with $p \equiv \frac{1}{2}$.

Step 3: Update the estimate of F by first calculating $p(x)$ for each x from (8) using the current estimates of F and θ and then using (7) with $p = p_0$.

Step 4: Update the estimate of θ by iteratively solving the score equation $S^*_{(m)}(\underline{\mathrm{x}}, \theta, \tilde{F}_n(\cdot;\theta)) = 0$. Then iterate Steps 3 and 4 until adequate convergence, labeling the final estimates as $\hat{\theta}$ and $\hat{F}$; and $\hat{\rho} = \log \hat{\theta}$.

To fit LAS2, repeat the LAS1 algorithm after replacing all sample values by their negatives. The resulting $\hat{F}(-x)$ estimates $G(x)$ of (2) and the resulting $(\hat{\rho}, \hat{\theta})$ estimates $(-\rho, 1/\theta)$ of (2).

The asymptotic variance of $\hat{\rho}$ (in either model) is $12/(\pi^2 n)$; however, we recommend use of the null variance, namely, the reciprocal of the null information as in Section 3 (the sum of squares of the signed-rank scores $a_n(j)$), supported by numerical studies reported in Section 7; there, the asymptotic variance when $n = 100$ is found to be about 40% too small whereas the null variance is about 6% too large; neither may be reliable when the model is incorrect, however. Recall that the asymptotic information loss due to the necessity of having to estimate F was about 18%; apparently, this is an insufficient evaluation unless samples are very large. The null variance enables an asymptotic confidence interval for ρ and, by exponentiation, a confidence interval for θ, and, being free of data dependence, will be useful for sample size planning.

To evaluate the fit of each model and to choose between them, compute a measure of fit for each, using a norm such as *sup*, ℓ_1, ℓ_2, or a weighted version thereof, to compare the edf $\mathbb{F}_n$ with the fitted versions of (1) and/or (2), evaluated at the n observations, the negatives thereof and 0. Graphs of the fits along with $\mathbb{F}_n$, on the df or cumulative hazard scales, are recommended. To choose between the two models, choose the one with the better fit. (See the example in Section 8.)

A Fortran program for carrying out testing, estimation and model fitting is available from the first author.

7. A Monte Carlo study

We have evaluated these estimation and fitting methods in a Monte Carlo simulation study, summarized briefly here. For each of five values of ρ , we generated 10,000 samples of size $n = 100$ from LAS1, and similarly for five values of ρ and $n = 30$. In each case, we generated u_i from $U(0,1)$; then $x_i = 2\,u_i^{\exp(-\rho)} - 1$ is a simulated observation from (1) with parameter $\theta = \exp(\rho)$ and F uniform on $(-1,1)$. (Since the inference methods are invariant to $F \in \mathcal{F}_0$, we conveniently chose F to be uniform.) Moreover, $-x_i$ is a simulated value from LAS2 with parameter $\theta = \exp(-\rho)$; hence, simulation studies for LAS1 have LAS2 interpretations as well.

For each sample, we tested the null hypothesis $H_0 :\ \rho = 0$ by *SL*, *NS* and *W*, estimated ρ along with three estimates of its standard error (SE)—'estimated SE' is based on summing the squares of the fitted scores, 'null SE' is based on summing the squares of the null hypothesis scores, and 'asymptotic SE' $= \pi/\sqrt{12n}$—and checked whether the true ρ fell outside 95% confidence intervals based on respective

TABLE 1
Summary of Simulations of Model 1, part 1: $n = 30$, MCreps $= 10^4$

ρ	$\hat{\rho}$	θ	$\hat{\theta}$	rejections of $\rho = 0$ by			incorrect CI's, using $SE =$		
				SL	NS	W	est.	null	asym.
-0.6	-0.4387	0.549	0.645	74.75%	74.32%	71.87%	3.29%	5.98%	16.52%
-0.4	-0.2967	0.670	0.743	43.25%	42.88%	41.09%	2.87%	3.82%	11.56%
0.0	0.0437	1.000	1.045	4.90%	4.94%	4.84%	4.39%	3.26%	10.04%
0.4	0.4546	1.492	1.576	51.38%	51.20%	49.28%	11.58%	5.66%	13.82%
0.6	0.6921	1.822	1.998	87.05%	86.72%	85.28%	19.59%	9.17%	18.14%

SE estimates. We also calculated four ℓ_2 fits, comparing the estimated F with the true F, the fitted df with the true df (1), the edf with the true df (1), and the fitted df with the edf; the first three of these measure quality of fit with the true (known) model while the fourth compares the fit of the semiparametric LAS1 model versus a fully nonparametric model. Tables 1–4 provide a summary report of these simulation studies.

In Tables 2 and 4, 'bias' $= \hat{\rho} - \rho$, 'MC SE' is the standard deviation of the 10^4 estimates of ρ, all other SE's are squareroots of the averages of the 10^4 estimated variances, and the ℓ_2 values were averaged in squared form before extracting squareroots.

From Tables 1 and 3, it is seen that all three tests have valid significance levels (5%) and the powers at all non-zero ρ-values are in the order implied by the AREs in Section 3, namely $SL > NS > W$, but with W faring better than suggested by the asymptotics. For confidence interval construction, the most reliable SE is the null one, which is indeed quite reliable for $|\rho| \leq 0.4$ ($0.67 \leq \theta \leq 1.50$, say) or so with these sample sizes, but the estimated SE is a contender.

From Tables 2 and 4, we see that $\hat{\rho}$ is positively biased, especially as ρ moves away from 0. Again we see that the null SE is reasonably reliable (by comparing with the MC SE), and hence should be quite suitable for sample-size planning as well as confidence interval construction. The slowness of convergence to the asymptotic SE is notable, and was also reported in Section 3. Simulations with $n = 500$ (not shown) do show continued improvement. Still, this estimation problem has $n + 1$ parameters (F at the $|x|$'s and ρ) and only n observations. Notice that the fits of the estimated df to the true df are somewhat better than those of the edf to the true df, as they should be.

The upward bias of $\hat{\rho}$ remains a puzzle, awaiting further ongoing investigation.

8. An example

We illustrate with an example appearing in the textbook by [16] (taken from [12] and also appearing in [9]). The sample size is small ($n = 14$), but it still serves

TABLE 2
Summary of Simulations of Model 1, part 2: $n = 30$, MCreps $= 10^4$

ρ	bias	SE's $\times 100$				ℓ_2-fits $\times 100$			
	$\times 100$	MC	est.	null	asym.	est. F	est.df	edf	fit-edf
-0.6	16.13	16.75	31.27	21.84	16.56	4.69	6.64	7.05	3.77
-0.4	10.33	17.88	28.12	21.84	16.56	4.31	6.44	7.24	3.41
0.0	4.37	19.33	23.29	21.84	16.56	3.90	6.60	7.33	3.22
0.4	5.46	21.58	19.10	21.84	16.56	3.68	7.00	7.01	3.47
0.6	9.21	23.97	16.93	21.84	16.56	3.72	7.07	6.68	3.72

TABLE 3
Summary of Simulations of Model 1, part 1: $n = 100$, MCreps $= 10^4$

ρ	$\hat{\rho}$	θ	$\hat{\theta}$	rejections of $\rho = 0$ by			incorrect CI's, using $SE =$		
				SL	NS	W	est.	null	asym.
-0.4	-0.3518	0.670	0.703	92.41%	92.33%	90.38%	3.22%	5.66%	13.38%
-0.2	-0.1735	0.819	0.841	40.73%	40.33%	37.81%	3.16%	4.30%	11.14%
0.0	0.0162	1.000	1.016	4.75%	4.86%	4.89%	3.84%	3.88%	10.00%
0.2	0.2148	1.221	1.240	44.69%	44.54%	42.44%	5.48%	3.72%	9.90%
0.4	0.4230	1.492	1.527	96.87%	96.76%	95.74%	9.03%	4.56%	11.08%

usefully as an illustration, with Model (1) but not Model (2) apparently fitting well. The data represent reductions, over 25-week periods, in the forced vital capacity (FVC) in 14 patients with cystic fibrosis, measured while undergoing drug therapy and placebo, in turn. Differences, for drug therapy minus placebo therapy, form the sample of data, with values ranging from -178 to 680.

Some of the output from a Fortran program analyzing these data is summarized here. One-sided p-values for testing symmetry-at-0 by signed rank tests *SL* and W were 0.019, by *NS* and K were 0.018 and by the sign test S 0.031; 2-term Edgeworth corrections (see [9]) reduced each somewhat (between 0.001 and 0.003 units).

The θ parameter was estimated to be 2.19 ($\hat{\rho} = 0.782$—Table 5). Three versions of the standard error for $\hat{\rho}$ are shown in Table 5 along with the resulting confidence intervals for θ. Based on simulation studies (Section 7), the null SE is perhaps the most dependable. Here, however, the sample size is too small to consider these more than rough guides.

The empirical and fitted model distribution functions are graphed in Figure 1. Three measures of fit were computed, based on *sup*, ℓ_1 and ℓ_2 norms, each measuring the difference between the empirical and the fitted model at the $\pm x_i$'s and 0. We found $sup = 0.085$, $\ell_1 = 0.058$ and $\ell_2 = 0.054$. The estimated F is also graphed in Figure 1; it is unimodal and long-tailed and represents (according to LAS1) what the distribution would have looked like if the drug therapy were completely ineffective.

Since the simulation studies showed considerable bias when θ is not close to unity, this value of $\hat{\theta} = 2.19$ might better be interpreted as somewhat smaller, say 1.5. The formal interpretation of $\theta = 1.5$ is that the probability of a difference, drug therapy minus placebo, being near x given that it is at most x has been increased by 50% relative to what it would have been for an ineffectual treatment—for every x. For comparison, when fitting LAS2 to these data, we found $\hat{\theta} = 1.59$, $\hat{\rho} = 0.467$ (estimated SE $= 0.521$, with null and asymptotic SEs as before), so all three of the confidence intervals would include 0, in contradiction to each of the *SL*, *NS* and W tests. Moreover, each measure of fit was larger than for LAS1; in particular, $\ell_2 = 0.072$. This illustrates potential use of measures of fit for choosing between LAS1 and LAS2.

TABLE 4
Summary of Simulations of Model 1, part 2: $n = 100$, MCreps $= 10^4$

ρ	bias	SE's $\times 100$				ℓ_2-fits $\times 100$			
	$\times 100$	MC	est.	null	asym.	est. F	est.df	edf	fit-edf
-0.4	4.82	10.79	13.67	11.43	9.07	2.31	3.46	3.92	1.92
-0.2	2.65	10.77	12.63	11.43	9.07	2.21	3.49	4.00	1.84
0.0	1.62	10.65	11.72	11.43	9.07	2.12	3.55	4.02	1.84
0.2	1.48	10.62	10.88	11.43	9.07	2.05	3.62	3.97	1.89
0.4	2.30	10.94	10.01	11.43	9.07	1.99	3.70	3.86	1.99

TABLE 5
Estimates of θ and ρ and Confidence Intervals for θ in the Example

estimate	$\hat{\theta}$	$\hat{\rho}$	SE^*		method for SE	95% CI for θ
initial	2.10	0.742	0.333		estimated SE	1.38, 3.47
final	2.19	0.782	0.236		null SE	1.13, 4.23
					asymptotic SE	1.36, 3.51

*Estimated SE's; the null $SE = 0.336$, the asymptotic $SE = 0.242$.

9. Final comments: A regression extension

As an extension to Model (2), consider the semiparametric regression model with covariate (vector) z :

$$G_{\beta,F}(x|z) = G(x)^{\exp(-\beta z)}, \quad F = 1 - G \in \mathcal{F}_0, \tag{11}$$

a proportional hazards regression model with symmetric-at-0 baseline F. (An alternative model would extend Model (1).) If F were known, $\hat{\beta}$ would be the solution to $(1/n)\sum_i z_i \exp(-\beta z_i)\Lambda(x_i) = \bar{z}$, Λ the cumulative hazard corresponding to F. Otherwise, a consistent estimate of Λ in this model would be needed, along with associated information bounds and efficient scores for β.

A simpler version of (11) is a 2-sample extension of (1) or (2), with a common F but differing θ-values, e.g., as a model for 'after' minus 'before' treatment measures in males and females. If the common F requirement is abandoned, this could easily be analyzed by applying the methodology herein to each group, leading to group-specific estimates with independence between groups. But small groups and/or continuous covariates will require new methods to analyze the full regression model (11). Further experience with applications is first needed.

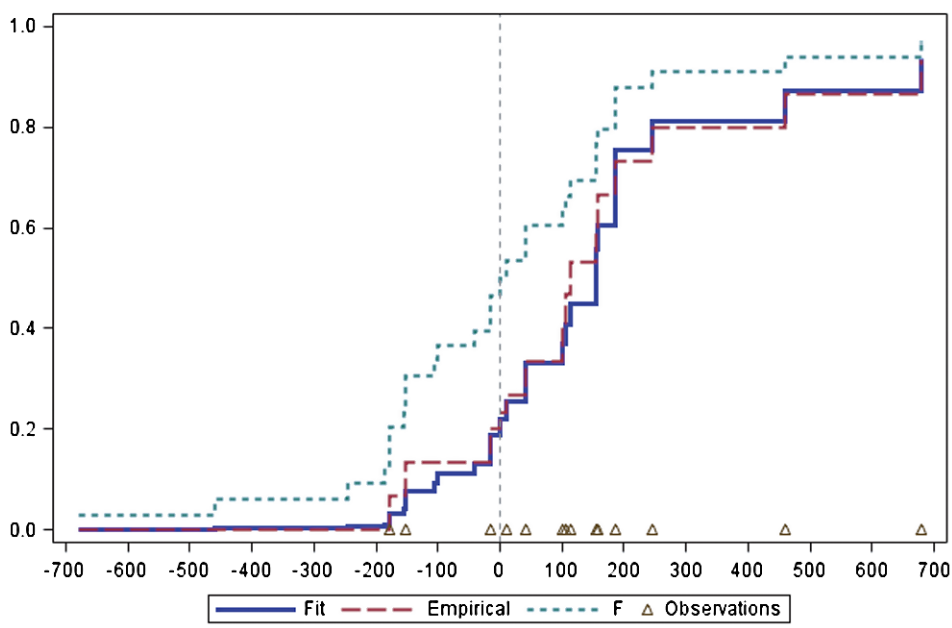

FIG 1. *Fitting of LAS1 to the empirical distribution and estimator of F in the example.*

In summary, the only earlier semiparametric model for paired-data differences is the shift model, requiring symmetry throughout. Resulting evaluations are dependent on the shape of the underlying symmetric F. The LAS models treated here have skewed alternatives, with the further advantage of evaluations free of dependence on F. The parameter has both a hazard and skewness interpretation.

Further investigations will include *(i)* seeking methods to reduce estimation bias, *(ii)* gaining experience with applications, *(iii)* evaluations when the LAS model is incorrect and *(iv)* development of the regression extension.

Appendix A: Information bounds and efficient scores for θ

The calculations in this section use the results and methods of [1] and [2, 3]. In particular we use the obvious "reverse" versions $\overline{R}$ and $\overline{L}$ of the R and L operators discussed in [2, 3], pages 420–424.

Suppose Model (1) holds: $F_\theta(x) = F(x)^\theta$ where F is continuous and symmetric about 0 and $\theta > 0$. We further assume that F has (symmetric) density f with respect to Lebesgue measure. Thus the density of the observations is given by $f(x;\theta,F) = \theta F(x)^{\theta-1} f(x) = \theta e^{-\theta\overline{\Lambda}(x)}\overline{\lambda}(x)$ where $\overline{\Lambda}(x) = \int_x^\infty F^{-1}\,dF = -\log F(x)$ and $\overline{\lambda} = f/F$; also, write $\Lambda = -\log(1-F)$. Thus the logarithm of the density is

$$\log f(x;\theta,F) = \log\theta + (\theta-1)\log F(x) + \log f(x) = \log\theta - \theta\overline{\Lambda}(x) + \log\overline{\lambda}(x).$$

Letting $\{f_\eta\}$ and $\{\overline{\lambda}_\gamma\}$ be parametric families through f and $\overline{\lambda}$, we find that the scores for θ and f (or $\overline{\lambda}$) are given by

$$\begin{aligned}(12)\qquad \dot{l}_1(x) &\equiv \dot{l}_\theta(x) = \frac{1}{\theta} + \log F(x) = \frac{1}{\theta} - \overline{\Lambda}(x),\\ \dot{l}_2 b(x) &\equiv \dot{l}_f b(x) = b(x) + (\theta-1)\frac{\int_{-\infty}^x b(y)\,dF(y)}{F(x)} = a(x) - \theta\int_x^\infty a(s)\,d\overline{\Lambda}(s),\end{aligned}$$

where

$$b(x) \equiv \frac{\partial}{\partial\eta}\log f_\eta(x)|_{\eta=0}, \quad a(x) \equiv \frac{\partial}{\partial\gamma}\log\overline{\lambda}_\gamma(x)|_{\gamma=0}.$$

By symmetry of all the densities f (and hence also f_η) under consideration,

$$\begin{aligned} b \in \mathbb{H}_0 &\equiv L_2^{0,even}(F)\\ &= \left\{b:\ \int b\,dF = 0,\ \int b^2\,dF < \infty,\ b(x) = b(-x)\ \text{ for all } x\in\mathbb{R}\right\}.\end{aligned}$$

We know that a and b are related by the $\overline{R}$ operator:

$$(13)\qquad a(x) = \overline{R}b(x) = b(x) - \frac{\int_{-\infty}^x b\,dF}{F(x)} = -E_F\{b(X) - b(x)|X \le x\}.$$

Since $b(x) = \overline{L}\circ\overline{R}b(x) \equiv \overline{R}b(x) - \int_x^\infty \overline{R}b\,d\overline{\Lambda} = b(x) - \int_{-\infty}^x dF/F(x) - \int_x^\infty \overline{R}b\,d\overline{\Lambda}$, it follows that $\int_{-\infty}^x b\,dF/F(x) = -\int_x^\infty \overline{R}b\,d\overline{\Lambda}$, and hence

$$(14)\qquad \begin{aligned}\dot{l}_2 b(x) &= \overline{R}b(x) + \theta\frac{\int_{-\infty}^x b\,dF}{F(x)} = \overline{R}b(x) - \theta\int_x^\infty \overline{R}b\,d\overline{\Lambda}\\ &= \overline{R}b(x) - \int_x^\infty \overline{R}b\,d(\theta\overline{\Lambda}) = \overline{L}_\theta\overline{R}b(x),\end{aligned}$$

where $\overline{L}_\theta$ is the (reverse) martingale or $\overline{L}-$operator corresponding to F_θ with reverse cumulative hazard $\theta\overline{\Lambda}$.

Our goal is to find $b^* \in \mathbb{H}_0$ satisfying $l_1^* \equiv \dot{l}_1 - \dot{l}_2 b^* \perp \dot{l}_2 b$ for all $b \in \mathbb{H}_0$, and to calculate the information for θ when F is symmetric but otherwise unknown given by

$$I(\theta) = E_\theta l_1^{*2}.$$

It follows from (12) that the information for θ when F is known is $I_0(\theta) = \theta^{-2}$, so that $I(\theta) \le I_0(\theta) = \theta^{-2}$ and $1/I(\theta) \ge 1/I_0(\theta) = \theta^2$. On the other hand, $1/I(\theta) \le (2^\theta - 1)/(\log 2)^2$, the asymptotic variance of the preliminary estimator of θ discussed in Section 5.

Thus we need to find b^* satisfying

$$\begin{aligned} 0 &= E\{(\dot{l}_1 - \dot{l}_2 b^*)\dot{l}_2 b\} \qquad \text{for all} \;\; b \in \mathbb{H}_0 \\ &= \langle \dot{l}_1 - \dot{l}_2 b^*, \dot{l}_2 b\rangle \;\;=\;\; \langle \dot{l}_2^T(\dot{l}_1 - \dot{l}_2 b^*), b\rangle_{\mathbb{H}_0} \end{aligned}$$

where now $\dot{l}_2^T = S \circ \dot{l}_{2,0}^T$ where $\dot{l}_{2,0}^T : L_2^0(F_\theta) \to L_2^0(F)$ is the adjoint of $\dot{l}_2$ in the unconstrained problem (with no symmetry imposed) and where $S : L_2^0(F) \to L_2^{0,even}(F) = \mathbb{H}_0$ is defined by

$$Sh(x) = \frac{1}{2}(h(x) + h(-x)).$$

Thus we want to calculate $b^* = (\dot{l}_2^T \dot{l}_2)^{-1}\dot{l}_2^T \dot{l}_1$. Now $\dot{l}_{2,0} = \overline{L}_\theta \circ \overline{R}$, so $\dot{l}_{2,0}^T = \overline{R}^T \circ \overline{L}_\theta^T = \overline{L} \circ \overline{R}_\theta$ using $\overline{R}^T = \overline{L}$ and $\overline{L}_\theta^T = \overline{R}_\theta$. Thus $\dot{l}_2^T = S \circ \overline{L} \circ \overline{R}_\theta$, and we find that

$$\begin{aligned} \dot{l}_2^T \dot{l}_1(x) &= S \circ \overline{L} \circ \overline{R}_\theta \dot{l}_1 = S \circ \overline{L} \circ \overline{R}_\theta\big(\theta^{-1}\overline{L}_\theta(1)\big) \;\;=\;\; S \circ \overline{L}\big(\theta^{-1}1\big) = \theta^{-1}S(1-\Lambda) \\ &= \theta^{-1}\left\{1 - \frac{\overline{\Lambda}(x) + \overline{\Lambda}(-x)}{2}\right\}. \end{aligned}$$

On the other hand, $\dot{l}_2^T \dot{l}_2 b = (\overline{L} \circ \overline{R}_\theta) \circ (\overline{L}_\theta \circ \overline{R}) b = \overline{L} \circ \overline{R} b = b - E_F b = b$ if $E_F b = 0$. Hence $(\dot{l}_2^T \dot{l}_2)^{-1} b = b$, and we conclude that

$$b^*(x) = \dot{l}_2^T \dot{l}_1(x) = \theta^{-1}\left(1 - \frac{\overline{\Lambda}(x) + \overline{\Lambda}(-x)}{2}\right). \tag{15}$$

Note that $\overline{\Lambda}(-x) = -\log F(-x) = -\log(1 - F(x)) = \Lambda(x)$, and hence if $X \sim F$, $\overline{\Lambda}(-X) = -\log(1 - F(X)) \stackrel{d}{=} -\log(1 - U) \stackrel{d}{=} -\log(U) \sim \text{Exponential}(1)$. Thus $E_F b^*(X) = 0$.

We now substitute expression (15) for b^* into the formula for l_1^* (after (15)), and then apply (15) and (13) to obtain

$$\begin{aligned} l_1^*(\cdot) &= \dot{l}_1 - \dot{l}_2 b^* = (1/\theta) - \overline{L}_\theta \overline{R} b^* = \overline{L}_\theta\left(\frac{1}{\theta} - \overline{R} b^*\right) \\ &= \overline{L}_\theta\left(\frac{1}{\theta} - b^* + \frac{\int_{-\infty}^{\cdot} b^* \, dF}{F}\right). \end{aligned} \tag{16}$$

We next calculate, from (15),

$$\theta \int_{-\infty}^{\cdot} b^* \, dF/F = 1 - \frac{1}{2F}\int_{-\infty}^{\cdot} \big(\overline{\Lambda}(y) + \overline{\Lambda}(-y)\big)\, dF(y) = \frac{1}{2F}(1-F)\Lambda - \frac{1}{2}\overline{\Lambda} \tag{17}$$

since $\Lambda(x) = \overline{\Lambda}(-x)$ and, by easy calculation,

$$\begin{aligned}
&\int_{-\infty}^{\cdot} \{\overline{\Lambda}(y) + \overline{\Lambda}(-y)\}\, dF(y) \\
&\quad = \int_{-\infty}^{\cdot} \{-\log F(y) - \log F(-y)\}\, dF(y) \\
&\quad = \int_{-\infty}^{\cdot} \{-\log F(y)\bigl(1 - F(y)\bigr)\}\, dF(y) \quad \text{since} \quad 1 - F(-y) = F(y) \\
&\quad = \int_0^v - \log u(1-u)\, du\big|_{v=0}^{F} \\
&\quad = (1-F)\log(1-F) - F\log F + 2F.
\end{aligned}$$

Thus, from (15), (16) and (17), we find that

$$l_1^*(x) = \frac{1}{2\theta}\overline{L}_\theta\Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr). \tag{18}$$

This yields, using $\overline{L}_\theta^T = \overline{R}_\theta$ and $\overline{R}_\theta \circ \overline{L}_\theta = I$,

$$\begin{aligned}
I(\theta) &= E_\theta l_1^{*2}(X) \\
&= \frac{1}{4\theta^2}\Bigl\langle \overline{L}_\theta\Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr), \overline{L}_\theta\Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr)\Bigr\rangle_{L_2^0(P_{\theta,F})} \\
&= \frac{1}{4\theta^2}\Bigl\langle \overline{R}_\theta\overline{L}_\theta\Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr), \Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr)\Bigr\rangle_{L_2(F)} \\
&= \frac{1}{4\theta^2}\Bigl\langle \Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr), \Bigl(\frac{1}{F(x)}\overline{\Lambda}(-x)\Bigr)\Bigr\rangle_{L_2(F)} \\
&= \frac{1}{4\theta^2}\int_0^1 \frac{(\log(1-u))^2}{u^2}\, du = \frac{\pi^2}{12\theta^2}.
\end{aligned}$$

To go further with the calculations, we proceed with calculation of $\overline{L}_\theta(\overline{\Lambda}(-x)/F(x))$. By definition,

$$\begin{aligned}
\overline{L}_\theta\Bigl(\frac{\overline{\Lambda}(-x)}{F(x)}\Bigr) &= \frac{1}{F(x)}\overline{\Lambda}(-x) - \theta\int_x^\infty \frac{1}{F(y)}\overline{\Lambda}(-y)\, d\overline{\Lambda}(y) \\
&= \frac{\overline{\Lambda}(-x)}{F(x)} - \theta\int_x^\infty \frac{-\log(1-F(y))}{F(y)^2}\, dF(y) \\
&= \frac{\overline{\Lambda}(-x)}{F(x)} - \theta\Bigl\{\overline{\Lambda}(x) + \frac{(1-F(x))}{F(x)}\overline{\Lambda}(-x)\Bigr\} \\
&= \frac{1}{F(x)}\bigl\{\Lambda(x) - \theta\bigl(F(x)\,\overline{\Lambda}(x) + G(x)\,\Lambda(x)\bigr)\bigr\}
\end{aligned}$$

since $\int_v^1 [-\log(1-u)]/u^2\, du = -[(1-v)/v]\log(1-v) - \log v$, $G = 1 - F$ and $\overline{\Lambda}(-x) = \Lambda(x)$. Hence, (18) becomes

$$l_1^*(x) = \frac{1}{2F(x)}\Bigl\{\frac{1}{\theta}\Lambda(x) - G(x)\Lambda(x) - F(x)\,\overline{\Lambda}(x)\Bigr\},$$

as claimed in Proposition 3. The efficient score for Model (2) is seen to be $-l_1^*(x)$ with θ replaced by $1/\theta$ and F and G interchanged, as in Proposition 3.

Appendix B: Ideal weights for estimation of F

We derive the formula for $p_0(x;\theta,F)$ in (8), the weights minimizing the variance of $\tilde{F}_n$ in (7). We only need it for $x>0$. To this end, for $x>0$,

$$\begin{aligned}
&\sqrt{n}\big(\tilde{F}_n(x)-F(x)\big)\\
&\quad= p(x)\sqrt{n}\big(\mathbb{F}_n^{1/\theta}(x)-F(x)\big)-p(-x)\sqrt{n}\big(\mathbb{F}_n(-x)^{1/\theta}-F(-x)\big)\\
&\quad= p(x)\frac{\mathbb{F}_n^{1/\theta}(x)-F(x)}{\mathbb{F}_n(x)-F^\theta(x)}\sqrt{n}\big(\mathbb{F}_n(x)-F^\theta(x)\big)\\
&\qquad - p(-x)\frac{\mathbb{F}_n(-x)^{1/\theta}-F(-x)}{\mathbb{F}_n(-x)-F(-x)^\theta}\sqrt{n}\big(\mathbb{F}_n(-x)-F(-x)^\theta\big)\\
&\quad\to_d p(x)\theta^{-1}F(x)^{1-\theta}\mathbb{U}\big(F(x)^\theta\big)-p(-x)\theta^{-1}F(-x)^{1-\theta}\mathbb{U}\big(F(-x)^\theta\big)\\
&\quad\equiv \mathbb{W}(x)-\mathbb{W}(-x)\equiv\mathbb{V}(x),
\end{aligned}$$

where $\mathbb{U}$ denotes a standard Brownian bridge process. Now

$$\begin{aligned}
\operatorname{Var}\big(\mathbb{V}(x)\big)&=\operatorname{Var}\big(\mathbb{W}(x)\big)+\operatorname{Var}\big(\mathbb{W}(-x)\big)-2\operatorname{Cov}\big(\mathbb{W}(x),\mathbb{W}(-x)\big)\\
&=\theta^{-2}\big(p^2A+q^2B-2pqC\big),
\end{aligned}$$

where

$$\begin{aligned}
A&\equiv\theta^2\operatorname{Var}\big(\mathbb{W}(x)\big)/p(x)^2\\
&=F(x)^{2(1-\theta)}F(x)^\theta\big(1-F(x)^\theta\big)=F(x)^{2-\theta}\big(1-F(x)^\theta\big),\\
B&\equiv\theta^2\operatorname{Var}\big(\mathbb{W}(-x)\big)/p(-x)^2\\
&=G(x)^{2-\theta}\big(1-G(x)^\theta\big)\quad\text{with}\quad G(x)=F(-x),\\
C&\equiv\theta^2\operatorname{Cov}\big(\mathbb{W}(x),\mathbb{W}(-x)\big)/\big(p(x)p(-x)\big)\\
&=F(x)^{1-\theta}G(x)^{1-\theta}(G(x)^\theta\big(1-F(x)^\theta\big)\\
&=F(x)^{1-\theta}G(x)-F(x)G(x)
\end{aligned}$$

and where $q(x)=p(-x)=1-p(x)$.

To minimize this variance w.r.t. choice of $p(x)$, we find

$$\frac{1}{2}\theta^2\frac{\partial}{\partial p}\operatorname{Var}(\mathbb{V})=pA-(1-p)B-(1-2p)C$$

which, when set to 0 and solved for $p=p_0$, yields $p_0=(B+C)/(A+B+2C)$, verifying (8).

References

[1] Begun, J. M., Hall, W. J., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

[2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD.

[3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer-Verlag, New York. Reprint of the 1993 original.

[4] Choi, S., Hall, W. J. and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861.

[5] Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

[6] Gupta, R. C., Gupta, P. L. and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Comm. Statist. Theory Methods* **27** 887–904.

[7] Gupta, R. C. and Gupta, R. D. (2007). Proportional reversed hazard rate model and its applications. *J. Statist. Plann. Inference* **137** 3525–3536.

[8] Hájek, J., Šidák, Z. and Sen, P. K. (1999). *Theory of Rank Tests*, 2nd ed. Probability and Mathematical Statistics, Academic Press Inc., San Diego, CA.

[9] Hall, W. J. (2011). A signed log-rank test of symmetry at zero. Tech. rep., University of Rochester, Department of Biostatistics.

[10] Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics.* John Wiley & Sons Inc., New York.

[11] Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods*, vol. 5 of *Kendall's Library of Statistics.* Edward Arnold, London.

[12] Knowles, M. R., Church, N. L. and Waltner, W. E. (1990). A pilot study of aerosolozed amiloride for the treatment of lung disease in cystic fibrosis. *N. Engl. J. Med.* **322** 1189–1194.

[13] Koziol, J. A. (1980). On a Cramér-von Mises-type statistic for testing symmetry. *J. Amer. Statist. Assoc.* **75** 161–167.

[14] Lehmann, E. L. (1953). The power of rank tests. *Ann. Math. Statistics* **24** 23–43.

[15] Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day Inc., San Francisco, Calif. With the special assistance of H. J. M. d'Abrera, Holden-Day Series in Probability and Statistics.

[16] Pagano, M. and Gauvreau, K. (1993). *Principles of Biostatistics.* Wadsworth, Belmont, CA.

[17] Schuster, E. F. (1975). Estimating the distribution function of a symmetric distribution. *Biometrika* **62** 631–635.

[18] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics.* John Wiley & Sons Inc., New York.

[19] van Eeden, C. (1963). The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *Ann. Math. Statist.* **34** 1442–1451.

[20] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics* **1** 80–83.