Web-based supporting materials for "Empirical likelihood based inference for functional means with application to wearable device data"

by Hsin-wen Chang* Ian W. McKeague†

July 7, 2022

Contents

*Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
†Department of Biostatistics, Columbia University, New York, NY 10032.

1 Proof of Theorem 1

For this proof and all other proofs in the supplement, o_p, o, O_p and O refer to *uniform* bounds over $a \in [\alpha_1, \alpha_2]$.

1.1 The First Part

In this part we show $\sqrt{n}\{f_n(\hat{\mu}) - f_n(\mu)\}(a) \xrightarrow{d} U(a)$ in $\ell^\infty([\alpha_1, \alpha_2] \setminus I_\delta)$ as $n \to \infty$, where I_δ can be constructed as $\bigcup_{\ell=1}^{J}(c_\ell - \delta/J, c_\ell)$, c_ℓ for $\ell = 1, \ldots, J < \infty$ are the (finitely many) discontinuities of $\mu(\cdot)$, and δ can be any small positive number that is less than the minimal distance between any two c_ℓ's and between any c_ℓ and α_1 or α_2, divided by 2. Since most parts of the proof still work when $[\alpha_1, \alpha_2] \setminus I_\delta$ is replaced by $[\alpha_1, \alpha_2]$, we will use $[\alpha_1, \alpha_2]$ as default in this subsection and use $[\alpha_1, \alpha_2] \setminus I_\delta$ only when needed (especially in the last paragraph of this subsection). It suffices to check the conditions of a changing-class Donsker Theorem (Kosorok, 2008b,a). To that end, we make use of an "almost" uniform continuity property of a right-continuous function $G(a)$ that has bounded variation on $[\alpha_1, \alpha_2]$: for all $\varepsilon > 0$, there exists a $\delta' > 0$ such that for all $a, b \in [\alpha_1, \alpha_2]$, $a < b$ and $(a, b]$ does not include a discontinuity of magnitude $\geq \varepsilon/2$, $b - a < \delta'$ implies $|G(b) - G(a)| < \varepsilon$ (this can be shown along the lines of the proof of the Heine–Cantor Theorem). This implies that several quantities appearing later in the proof, namely, $E\{V(a)\}$, $E\{D(a)\}$ and the sample paths of $T(a)$, are almost uniformly continuous, where recall that V is defined in Remark 4 after Theorem 1 as the total variation of T over $[\alpha_1, a]$, $D(a) = V(a) - T(a)$, $f_{n,a}(g) = f_n(g)(a)$, $F(g) = |g(\alpha_1)| +$ total variation of g over $[\alpha_1, \alpha_2]$ for $g \in \mathcal{B} \cap \mathcal{C}$, $\mathcal{B} = \{g : [\alpha_1, \alpha_2] \mapsto [0, \tau], g$ is of bounded variation$\}$, and $\mathcal{C} = \{g : [\alpha_1, \alpha_2] \mapsto [0, \tau], g$ is right-continuous$\}$.

The classes of functions in question are $\mathcal{F}_n = \{f_{n,a} : \mathcal{B} \cap \mathcal{C} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2] \setminus I_\delta\}$. The classes $\mathcal{F}_n$ satisfy the almost measuarable Suslin condition, by the separability of $\mathcal{F}_n$ shown as follows. By definition of $\mathbf{G}_n$, we have $\sup_{a \in [\alpha_1, \alpha_2]} \inf_{b \in \mathbf{G}_n} |f_{n,b}(T) - f_{n,a}(T)| = 0$ almost surely. This implies the desired separability $P\{\sup_{a \in [\alpha_1, \alpha_2]} \inf_{b \in \mathbf{G}_n} |f_{n,b}(T) - f_{n,a}(T)| > 0\} = 0$. The same function F (as defined above) will be used as an envelope function for every $\mathcal{F}_n$. Since $EV^2(\alpha_2)$ and $ET^2(\alpha_1)$ are finite, $EF^2(T) < \infty$. This and the dominated convergence theorem imply that $EF^2(T)I\{F(T) > \eta\sqrt{n}\} \to 0$ as $n \to \infty$, for each $\eta > 0$. The bounded uniform entropy integral condition is shown as follows. Each $g \in \mathcal{B} \cap \mathcal{C}$ satisfies $g = v(g) - d(g)$, where $v(g)(a)$ is the total variation of g over $[\alpha_1, a]$, $d(g) = v(g) - g$, and both $v(g)(a)$ and $d(g)(a)$ are non-decreasing functions of $a \in [\alpha_1, \alpha_2]$ (Carothers, 2000, Theorem 13.5). By linearity of $f_{n,a}$, we have that $f_{n,a}(g) = f_{n,a} \circ v(g) - f_{n,a} \circ d(g)$. Hence for any finitely discrete probability measure Q with $\|F\|_{Q,2} > 0$ (see, e.g., Kosorok, 2008b, page 18) and any $f_{n,a}, f_{n,b} \in \mathcal{F}_n$, $\|f_{n,a} - f_{n,b}\|_{Q,2} \leq \|f_{n,a} \circ v(g) - f_{n,b} \circ v(g)\|_{Q,2} + \|f_{n,a} \circ d(g) - f_{n,b} \circ d(g)\|_{Q,2}$ by the

triangle inequality, where $\|f\|_{Q,2} \equiv \{Qf^2\}^{1/2}$, which implies

$$N(\varepsilon \|F\|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq N(\varepsilon \|F\|_{Q,2}/2, \mathcal{V}_n, L_2(Q)) N(\varepsilon \|F\|_{Q,2}/2, \mathcal{D}_n, L_2(Q))$$

for any $\varepsilon \in (0,1)$. Here $N(\varepsilon, \mathcal{G}, L_2(Q))$ is the minimal number of balls of radius ε with respect to the $L_2(Q)$-norm $\|f\|_{Q,2}$ needed to cover the set $\mathcal{G}$, $\mathcal{V}_n = \{f_{n,a} \circ v : \mathcal{B} \cap \mathcal{C} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$, and $\mathcal{D}_n = \{f_{n,a} \circ d : \mathcal{B} \cap \mathcal{C} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$. By Lemma 9.18 of Kosorok (2008b), for $\mathcal{G}_n = \mathcal{V}_n$ or $\mathcal{D}_n$, $N(\varepsilon \|F\|_{Q,2}/2, \mathcal{G}_n, L_2(Q)) \leq N_{[]}(\varepsilon \|F\|_{Q,2}/2, \mathcal{G}_n, L_2(Q))$, where $N_{[]}(\varepsilon, \mathcal{G}, L_2(Q))$ is the minimal number of ε-brackets with respect to the $L_2(Q)$-norm needed to cover the set $\mathcal{G}$. Further, we can show that $N_{[]}(\varepsilon \|F\|_{Q,2}/2, \mathcal{G}_n, L_2(Q)) \leq 8/\varepsilon^2$ for all n, for $\mathcal{G}_n = \mathcal{V}_n$ or $\mathcal{D}_n$, by extending a proof (van der Vaart and Wellner, 1996, Example 2.11.16) involving right-continuous monotone stochastic processes. These results imply $N(\varepsilon \|F\|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq 64/\varepsilon^4$. We then obtain the bounded uniform entropy integral condition

$$\limsup_{n\to\infty} \sup_Q \int_0^1 \sqrt{\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}_n, L_2(Q))} d\varepsilon \leq \int_0^1 \sqrt{\log(64/\varepsilon^4)} d\varepsilon < \infty.$$

The covariance of the limiting process at the index pair (a,b) is $\lim_{n\to\infty} E\{f_{n,a}(T) - Ef_{n,a}(T)\}\{f_{n,b}(T) - Ef_{n,b}(T)\} = ET(a)T(b) - ET(a)ET(b)$, where the equality follows by the almost uniform continuity of the sample path of $T(a)$, Slutsky's lemma and the dominated convergence theorem. The same reasoning implies $\lim_{n\to\infty} \rho_n(a,b)$ exsits for all $a, b \in [\alpha_1, \alpha_2]$ and equals $\rho(a,b) = [E\{T(a) - T(b)\}^2]^{1/2}$, where $\rho_n(a,b) = [E\{f_{n,a}(T) - f_{n,b}(T)\}^2]^{1/2}$.

The last condition for the changing classes Donsker Theorem is that for all deterministic sequences $\{a_n\}, \{b_n\} \subset [\alpha_1, \alpha_2] \setminus I_\delta$, $\rho_n(a_n, b_n) \to 0$ if $\rho(a_n, b_n) \to 0$, which we show as follows. Given $\varepsilon > 0$, we want to show there exists $N \in \mathbb{N}$ such that for all $n > N$, $\rho_n(a_n, b_n) < \varepsilon$. Since $\rho_n(a_n, b_n) \leq \sqrt{\tau E\,|f_{n,a_n}(T) - f_{n,b_n}(T)|}$, it suffices to show $E\,|f_{n,a_n}(T) - f_{n,b_n}(T)| < \varepsilon^2/\tau$. By the almost uniform continuity of $E\{V(a)\}$ and $E\{D(a)\}$, there exists a $\delta_E > 0$ such that for all $a, b \in [\alpha_1, \alpha_2]$, $a < b$ and $(a,b]$ does not include a discontinuity of magnitude $\geq \varepsilon^2/(12\tau)$, $b - a < \delta_E$ implies $E\{V(b) - V(a)\}, E\{D(b) - D(a)\} < \varepsilon^2/(6\tau)$. Denote the mesh of $\mathbf{G}_n$ by u_n. By $\rho(a_n, b_n) \to 0$ and $u_n = o(1)$ as $n \to \infty$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $\rho(a_n, b_n) < \varepsilon^2/(3\tau)$ and $u_n < \min\{\delta/J, \delta_E\}$. Then for each $n > N$, note that $(a_n, a_n + u_n]$ and $(b_n, b_n + u_n]$ do not contain any discontinuity of $\mu(a)$ (and hence any discontinuity of $E\{V(a)\}$ and $E\{D(a)\}$) because $u_n < \delta/J$ and $a_n, b_n \in [\alpha_1, \alpha_2] \setminus I_\delta$.

This leads to

$$
\begin{aligned}
&E\left|f_{n,a_n}(T)-f_{n,b_n}(T)\right| \\
&\le E\left|f_{n,a_n}(T)-T(a_n)\right|+E\left|T(a_n)-T(b_n)\right|+E\left|f_{n,b_n}(T)-T(b_n)\right| \\
&\le E\left|f_{n,a_n}(V)-V(a_n)\right|+E\left|f_{n,a_n}(D)-D(a_n)\right|+\rho(a_n,b_n) \\
&+E\left|f_{n,b_n}(V)-V(b_n)\right|+E\left|f_{n,b_n}(D)-D(b_n)\right| \\
&\le E\left\{V(a_n+u_n)-V(a_n)\right\}+E\left\{D(a_n+u_n)-D(a_n)\right\}+\varepsilon^2/(3\tau) \\
&+E\left\{V(b_n+u_n)-V(b_n)\right\}+E\left\{D(b_n+u_n)-D(b_n)\right\} \\
&<4\times\varepsilon^2/(6\tau)+\varepsilon^2/(3\tau)=\varepsilon^2/\tau,
\end{aligned}
$$

where the first inequality is due to the triangle inequality, the second inequality is due to $T(a)=V(a)-D(a)$ and the Cauchy-Schwarz inequality, the third inequality is due to monotonicity of $E\{V(a)\}$ and $E\{D(a)\}$, and the last inequality is due to $u_n<\delta_E$ and the almost uniform continuity of $E\{V(a)\}$ and $E\{D(a)\}$. Since all the conditions for the changing classes Donsker Theorem are satisfied, we have $\sqrt{n}\left\{f_{n,a}(\hat{\mu})-f_{n,a}(\mu)\right\}$ converges weakly in $\ell^\infty([\alpha_1,\alpha_2]\setminus I_\delta)$, as $n\to\infty$, to a tight, mean zero Gaussian process $U(a)$ (as defined in Section 2.2) with covariance function at the index pair (a,b) being $\mathrm{Cov}\{T(a),T(b)\}$.

1.2 The Second Part

It suffices to show $\sqrt{n}\{f_n(\mu)(a)-\mu(a)\}\to 0$ uniformly over $a\in[\alpha_1,\alpha_2]\setminus I_\delta$ as $n\to\infty$. Since the sample paths of $T(a)$ has bounded variation, it can be shown that $\mu(a)$ is also of bounded variation. By the Lebesque decomposition of a function of bounded variation, and the condition that $\mu(a)$ is right continuious with finitely many jumps, we can write $\mu(a)=h_1(a)+h_2(a)$, where $h_1(a)$ is a continuous function of bounded variation and $h_2(a)=\sum_{b\le a}\{\mu(b)-\mu(b-)\}$ is a piecewise-constant, right continuous function of bounded variation. The desired result follows if we can show $\sqrt{n}\{f_n(h_1)(a)-h_1(a)\}\to 0$ and $\sqrt{n}\{f_n(h_2)(a)-h_2(a)\}\to 0$ uniformly over $a\in[\alpha_1,\alpha_2]\setminus I_\delta$ as $n\to\infty$.

We first deal with $h_1(a)$. By the condition of bounded right-hand β-Dini derivatives, we can show that $D^+(h_1,\beta)(a)$ and $D_+(h_1,\beta)(a)$ are also bounded over $a\in[\alpha_1,\alpha_2]$. Then by a similar reasoning as in the results 34.4 and 34.5 of McShane (1944), we can show that for all $a,b\in[\alpha_1,\alpha_2]$,

$$
|h_1(a)-h_1(b)|\le\kappa\,|a-b|^\beta\,, \tag{S.1}
$$

where $\kappa<\infty$ is a bound of $|D^+(h_1,\beta)(a)|$ and $|D_+(h_1,\beta)(a)|$ over $a\in[\alpha_1,\alpha_2]$. Given $\eta>0$, by the assumption placed on the mesh u_n of $\mathbf{G}_n$, there exists an $N_1=N_1(\eta)\in\mathbb{N}$

such that for all $n > N_1$, $\sqrt{n}u_n^\beta \le \eta/\kappa$. Then

$$\left|\sqrt{n}\left\{f_n(h_1)(a) - h_1(a)\right\}\right| \le \sup_{h \le u_n}\left\{\frac{|h_1(a+h) - h_1(a)|}{h^\beta} \times \sqrt{n}h^\beta\right\} \le \kappa \times \sqrt{n}u_n^\beta \le \eta.$$

This means $\sup_{a\in[\alpha_1,\alpha_2]} \sqrt{n}\{f_n(h_1)(a) - h_1(a)\} \to 0$ as $n \to \infty$.

As for $h_2(a)$, there exists an $N_2 = N(\delta) \in \mathbb{N}$ such that for all $n > N_2$, $u_n < \delta/J$, leading to

$$\sup_{a\in[\alpha_1,\alpha_2]\setminus I_\delta} \sqrt{n}\{f_n(h_2)(a) - h_2(a)\} = 0.$$

This means $\sqrt{n}\{f_n(h_2)(a) - h_2(a)\} \to 0$ uniformly over $a \in [\alpha_1, \alpha_2] \setminus I_\delta$ as $n \to \infty$.

2 Details regarding remarks after Theorem 1

2.1 Constructing a simultaneous confidence band for $\mu(\cdot)$ for essentially all $a \in [\alpha_1, \alpha_2]$

Let $r_n(\mu)(a)$ be some random function of $\mu(\cdot)$ in $\ell^\infty[\alpha_1, \alpha_2]$. In this subsection, we explain how we use the limiting process $r_L(a)$ of $r_n(\mu)(a)$ in $\ell^\infty\{[\alpha_1, \alpha_2] \setminus I_\delta\}$ for each sufficiently small $\delta > 0$ to construct a simultaneous confidence band for $\mu(\cdot)$ for essentially all $a \in [\alpha_1, \alpha_2]$. Here we use "essentially all" a to mean that it is for all a except for a set I_δ having arbitrarily small Lebesgue measure δ, and both δ and I_δ are defined in Supplement Section 1.1. This result can then be applied to $r_n(\mu)(a) = |\sqrt{n}\{f_n(\hat{\mu}) - \mu\}(a)|$ in Theorem 1, or other confidence bands/test statistics in the manuscript that involve the same construction (i.e. the same $r_n(\mu)(a)$ and critical value is used) for each sufficiently small $\delta > 0$. Furthermore, the same reasoning holds when the limiting process needs to be estimated by bootstrap.

For each sufficiently small $\delta > 0$, an asymptotic $100(1-\alpha)\%$ simultaneous confidence band for $\mu(\cdot)$ over $[\alpha_1, \alpha_2] \setminus I_\delta$ is

$$\{(a, \tilde{\mu}(a)) : r_n(\tilde{\mu})(a) \le c_{\alpha,\delta}, a \in [\alpha_1, \alpha_2] \setminus I_\delta\}, \tag{S.2}$$

where $c_{\alpha,\delta}$ denotes the upper α-quantile of $\sup_{a\in[\alpha_1,\alpha_2]\setminus I_\delta} r_L(a)$. Let c_α be the upper α-quantile of $\sup_{a\in[\alpha_1,\alpha_2]} r_L(a)$. Since $c_{\alpha,\delta} \le c_\alpha$, replacing $c_{\alpha,\delta}$ by c_α in (S.2) still gives an asymptotic $100(1-\alpha)\%$ simultaneous confidence band for $\mu(\cdot)$ over $[\alpha_1, \alpha_2] \setminus I_\delta$ for every δ:

$$\{(a, \tilde{\mu}(a)) : r_n(\tilde{\mu})(a) \le c_\alpha, a \in [\alpha_1, \alpha_2] \setminus I_\delta\}.$$

Essentially, this band can be constructed over the entire interval $[\alpha_1, \alpha_2]$, because it is valid for every sufficiently small $\delta > 0$ and because c_α does not depend on δ. Therefore, we term it as a simultaneous confidence band for $\mu(\cdot)$ for essentially all $a \in [\alpha_1, \alpha_2]$.

2.2 Boundedness of the right-hand β-Dini derivatives

Recall that in Theorem 1, we need the condition that $D^+(\mu,\beta)(a)$ and $D_+(\mu,\beta)(a)$ are bounded over $a \in [\alpha_1,\alpha_2]$ for some $\beta > 0$. Here we provide some examples of functions satisfying this condition. Taking $\mu(a) = a^\gamma$, $a \in [0,1]$, for some $0 < \gamma < 1$, note that $\mu(a)$ has uniformly bounded right-hand β-Dini derivatives over $a \in [0,1]$ when $\beta = \gamma$. However, the right-hand β-Dini derivatives are unbounded at $a = 0$ for any $\beta > \gamma$. Another example in which we have uniformly bounded right-hand β-Dini derivatives for any $\beta < 1/2$ (although not for $\beta = 1/2$) is the sample path of a Brownian motion; but note that our $T(a)$ cannot be a Brownian motion due to the bounded variation assumption.

Note that if $D^+(\mu,\beta)(a)$ is uniformly bounded by some positive constant κ for some $\beta > 1$, then under the condition of the first part of Theorem 1, $\mu(a)$ is a piecewise constant function of a. The proof is as follows. Since

$$\left|\limsup_{h\to 0+} \frac{\mu(a+h)-\mu(a)}{h}\right| = \left|\limsup_{h\to 0+} \frac{\mu(a+h)-\mu(a)}{h^\beta} h^{\beta-1}\right| \le \kappa \lim_{h\to 0+} h^{\beta-1} = 0,$$

and similarly for $|\liminf_{h\to 0+}[\{\mu(a+h)-\mu(a)\}/h]|$, the right-hand derivative of $\mu(a)$ exists and equals zero. Under the assumption (from the first part of the theorem) that $T(a)$ is right-continuous in a (which leads to the right-continuity of $\mu(a)$), this means that $\mu(a)$ is a piecewise constant function of a.

3 Bootstrap in Section 2.2

3.1 Proof of Corollary 1

To show bootstrap consistency of $f_{n,a}(U_n^*)$, we write it as $\sum_{i=1}^n (W_{ni}-1) f_{n,a}(T_i)/\sqrt{n}$ $= \sum_{i=1}^n (W_{ni}-1)\{f_{n,a}(T_i) - f_{n,a}(\mu)\}/\sqrt{n}$, where recall from Section 2.2 that W_{ni} is the number of times that $f_n(T_i)(a)$ is redrawn from $\{f_n(T_1)(a),\ldots,f_n(T_n)(a), a \in [\alpha_1,\alpha_2]\}$. We first introduce a partly Poissonized version of $f_{n,a}(U_n^*)$. The reason for Poissonization is to remove the dependence among the multinomial distributed $W_{n1},\ldots,W_{nn}$, so that a changing classes bootstrap central limit theorem can be utilized. Secondly, we show the conditional asymptotic equivalence of $f_{n,a}(U_n^*)$ and the partly Poissonized version of $f_{n,a}(U_n^*)$. Finally we show the bootstrap consistency of the partly Poissonized version of $f_{n,a}(U_n^*)$. Then by a routine extension of the conditional Slutsky's lemma (Cheng, 2015, Appendix A.2, (i)) to the case of random elements of a metric space, we have the bootstrap consistency of $f_{n,a}(U_n^*)$. The aforementioned bootstrap consistency results are provided conditional on $T_1, T_2, \ldots$. The extension to conditioning on the discretized data is postponed until the last part of the proof.

The partly Poissonized version is obtained by taking a Poisson number N_n (instead

of n) of bootstrap draws from the original sample, where N_n has mean n and is independent of the original sample. Then $W_{N_n,i}$ is the resulting number of times that $T_i(a)$ is redrawn from the original sample, $i = 1, \ldots, n$. This construction leads to i.i.d. $W_{N_n,1}, \ldots, W_{N_n,n}$, each having Poisson distribution with mean 1. The partly Poissonized version of $f_{n,a}(U_n^*)$ is defined as $f_{n,a}(U_{n,N_n}^*)$, where $U_{n,N_n}^*(a) = \sum_{i=1}^n (W_{N_n,i} - 1)\{T_i(a) - \mu(a)\}/\sqrt{n}$.

To show the conditional asymptotic equivalence of $f_{n,a}(U_n^*)$ and $f_{n,a}(U_{n,N_n}^*)$, we use similar reasoning as in the proof establishing such equivalence for classes of functions not changing with n (van der Vaart and Wellner, 1996, page 347–348) and the following additional result. We show in the following that conditional on the observed processes $T_1, T_2, \ldots$ a.s., $\sum_{i \in I_n^j} \{f_{n,a}(T_i) - f_{n,a}(\mu)\}/\#I_n^j = o_p(1)$ for any j, where I_n^j is the set of indices $i \in \{1, \ldots, n\}$ such that $|W_{N_n,i} - W_{ni}| \geq j$. Since $\sup_{a \in [\alpha_1,\alpha_2]} \left| \sum_{i \in I_n^j} \{f_{n,a}(T_i) - f_{n,a}(\mu)\}/\#I_n^j \right| \leq \sup_{a \in [\alpha_1,\alpha_2]} \left| \sum_{i \in I_n^j} \{T_i(a) - \mu(a)\}/\#I_n^j \right|$ by definition of $f_{n,a}$, it amounts to showing the right hand side is $o_p(1)$ conditional on the observed processes a.s. This is true (van der Vaart and Wellner, 1996, Lemma 3.6.16) because the class of evaluation functions $\mathcal{F}_\mathcal{C} = \{q_a : \mathcal{B} \cap \mathcal{C} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$ is strong P-Glivenko–Cantelli, where $q_a(g) = g(a)$ for $g \in \mathcal{B}$, and P is the distribution of T. The proof utilizes the fact that the class of functions $\mathcal{F} = \{q_a : \mathcal{B} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$ is P-Donsker (see Supplement Section 11), Donsker preservation for a restriction in the sample space (van der Vaart and Wellner, 1996, Theorem 2.10.6 and page 200), and the fact that a Donsker class is also strong Glivenko–Cantelli (Kosorok, 2008b, Lemma 8.17).

Now we show the bootsrap consistency of $f_{n,a}(U_{n,N_n}^*)$, which we decompose as

$$\frac{N_n}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{W_{N_n,i}}{N_n/n} - 1 \right) \{f_{n,a}(T_i) - f_{n,a}(\mu)\} + \sqrt{n} \left(\frac{N_n}{n} - 1 \right) \{f_{n,a}(\hat{\mu}) - f_{n,a}(\mu)\}. \tag{S.3}$$

In the first term of (S.3), the quantity after N_n/n satisfies the conditions for a changing classes bootstrap central limit theorem (Kosorok, 2008b, Theorem 11.23) as follows. All the conditions except one has been checked in Supplement Section 1.1. The one additional condition is that the $W_{N_n,i}$ are positive, i.i.d. random variables independent of $T_1, T_2, \ldots$ with finite variance, which holds by the Poissonization construction. This, $N_n/n = 1 + o(1)$ a.s. (by the strong law of large numbers) and another routine extension of the conditional Slutsky lemma (Cheng, 2015, Appendix A.2, (ii)) to the case of random elements of a metric space, imply the bootstrap consistency of the first term of (S.3). In the second term, the quantity in the brackets is $o(1)$ a.s., by the fourth paragraph of Supplement Section 4 and the fact that $\sup_{a \in [\alpha_1,\alpha_2]} |f_{n,a}(\hat{\mu} - \mu)| \leq \sup_{a \in [\alpha_1,\alpha_2]} |\hat{\mu}(a) - \mu(a)|$ per definition of $f_{n,a}$. This, the conditional weak convergence of $\sqrt{n}(N_n/n - 1)$ as $n \to \infty$ (due to the central limit theorem), and aforementioned conditional Slutsky's lemma imply the second term of

(S.3) converges weakly to zero as $n \to \infty$, conditional on $T_1, T_2, \ldots$, in probability. This conditional weak convergence to zero implies conditional convergence in probability to zero as $n \to \infty$, based on a similar proof as in the unconditional version. Combining the results from the two terms using a routine extension of the conditional Slutsky's lemma (Cheng, 2015, Appendix A.2, (i)) to the case of random elements of a metric space, we have the bootstrap consistency of $f_{n,a}(U^*_{n,N_n})$. Then by the same extension of the conditional Slutsky's lemma and the conditional asymptotic equivalence of $f_{n,a}(U^*_n)$ and $f_{n,a}(U^*_{n,N_n})$, we have the bootstrap consistency of $f_{n,a}(U^*_n)$.

We now show the desired bootstrap consistency of $f_{n,a}(U^*_n)$ given the discretized data $\{f_j(T_i), i = 1, \ldots, j, j = 1, 2, \ldots\}$ in probability. Let the σ-fields generated by the fully-observed and discretely-observed data be denoted $\mathcal{A} = \sigma(T_1, T_2, \ldots)$ and $\mathcal{A}_0 = \sigma(f_j(T_i), i = 1, \ldots, j, j = 1, 2, \ldots)$, respectively. Since it can be shown that f_j is measurable for each $j = 1, 2, \ldots$, we have $\mathcal{A}_0 \subset \mathcal{A}$. Let U' be a copy of U that is independent of $T_1, T_2, \ldots$, and let BL_1 be the set of all real-valued functions on $\ell^\infty\{[\alpha_1, \alpha_2] \setminus I_\delta\}$ with Lipschitz norm bounded by 1 and $\sup_{g \in \ell^\infty\{[\alpha_1,\alpha_2]\setminus I_\delta\}} |h(g)| \leq 1$, where I_δ is defined in Supplement Section 1.1. Then

$$\begin{aligned}
&\sup_{h \in BL_1} \left| E\left\{h\left(f_{n,a}\left(U^*_n\right)\right) | \mathcal{A}_0\right\} - Eh\left(U\right)\right| \\
=&\sup_{h \in BL_1} \left| E\left\{h\left(f_{n,a}\left(U^*_n\right)\right) | \mathcal{A}_0\right\} - E\left\{h\left(U'\right) | \mathcal{A}_0\right\}\right| \\
\leq& E\left[\sup_{h \in BL_1} \left| E\left\{h\left(f_{n,a}\left(U^*_n\right)\right) | \mathcal{A}\right\} - Eh\left(U\right)\right| | \mathcal{A}_0\right] \xrightarrow{P} 0,
\end{aligned}$$

where we use the tower property of conditional expectation (see, e.g., Shao, 2003, Proposition 1.10 (v)) and $E\left\{h\left(U'\right) | \mathcal{A}_0\right\} = Eh\left(U'\right) = Eh\left(U\right) = E\left\{h\left(U'\right) | \mathcal{A}\right\}$; the convergence is due to the conditional version of the dominated convergence theorem (see, e.g., Shao, 2003, Proposition 1.10 (x)), $\sup_{h \in BL_1} \left| E\left\{h\left(f_{n,a}\left(U^*_n\right)\right) | \mathcal{A}\right\} - Eh\left(U\right)\right| \xrightarrow{P} 0$ (the bootstrap consistency of $f_{n,a}(U^*_n)$), and arguing along subsquences.

3.2 Details in implementing bootstrap and computing $c^*_{NS,\alpha}$

A bootstrap sample is obtained by drawing n curves independently with replacement from the data $\{f_n(T_1)(a), \ldots, f_n(T_n)(a), a \in [\alpha_1, \alpha_2]\}$. Based on this bootstrap sample, compute a value of $\sup_{a \in [\alpha_1, \alpha_2]} |f_n(U^*_n(a))|$, where recall from Section 2.2 that $U^*_n(a) = \sqrt{n}\{\hat{\mu}^*(a) - \hat{\mu}(a)\}$, $\hat{\mu}(a) = \sum_{i=1}^n T_i(a)/n$ is the full trajectory of the mean of the original sample, $\hat{\mu}^*(a) = \sum_{i=1}^n W_{ni} T_i(a)/n$ is the full trajectory of the mean of the bootstrap sample, and W_{ni} is the number of times that $f_n(T_i)(a)$ is redrawn from $\{f_n(T_1)(a), \ldots, f_n(T_n)(a), a \in [\alpha_1, \alpha_2]\}$.

Repeat the procedure in the previous paragraph B times to obtain B bootstrapped values for $\sup_{a \in [\alpha_1, \alpha_2]} |f_n(U^*_n(a))|$; our simulation study (see Section 3) uses $B = 1000$.

Let $c^*_{NS,\alpha}$ denote the upper α-quantile of these B bootstrapped values of $\sup_{a\in[\alpha_1,\alpha_2]} |f_n(U^*_n(a))|$. Then the asymptotic $100(1-\alpha)\%$ NS simultaneous confidence band for $\mu(\cdot)$ is computed as $f_n(\hat{\mu})(a) \pm n^{-1/2} c^*_{NS,\alpha}$ for essentially all $a \in [\alpha_1, \alpha_2]$.

4 Proof of Theorem 2

For simplicity of exposition, we first study the large sample behavior of the fully observed trajectories of the EL statistic process $-2\log \mathcal{R}(\mu)(\cdot)$. Then we discretize this large sample result and study the asymptotics of $-2\log f_n(\mathcal{R}(\mu))(\cdot)$ accordingly. But note that, as in Section 2.3, the quantities we use in the process of this proof are available to us only in their discretized forms. Since most parts of the proof still work when $[\alpha_1, \alpha_2] \setminus I_\delta$ is replaced by $[\alpha_1, \alpha_2]$, we will use $[\alpha_1, \alpha_2]$ as default in this subsection and use $[\alpha_1, \alpha_2] \setminus I_\delta$ only when needed (especially in the last paragraph of this subsection).

We need to show weak convergence of $-2\log \mathcal{R}(\mu)(\cdot)$ as a random element of $\ell^\infty[\alpha_1, \alpha_2]$ when $n \to \infty$. The denominator of (2) is obviously n^{-n}. The constrained optimum of the numerator is found by Lagrange's method to be $\prod_{i=1}^n \tilde{p}_i(a)$, where $\tilde{p}_i(a) = [n\{1 + \tilde{\lambda}(a)\tilde{Y}_i(a)\}]^{-1}$, $\tilde{Y}_i(a) = T_i(a) - \tilde{\mu}(a)$, and $\tilde{\lambda}(a)$ satisfies the estimating equation $\sum_{i=1}^n \tilde{p}_i(a)\tilde{Y}_i(a) = 0$. It follows that $-2\log \mathcal{R}(\tilde{\mu})(a) = 2\sum_{i=1}^n \log\{1 + \tilde{\lambda}(a)\tilde{Y}_i(a)\}$. In the sequel, we use $-2\log \mathcal{R}(\mu)(a)$, Y_i and λ to denote the corresponding quantities at the true value $\mu(a)$.

The first step is show that $\lambda(a) = O_p(1/\sqrt{n})$. This step uses the inequality

$$|\lambda(a)|\overline{\sigma}(a) \le \zeta(a)\overline{T}(a)\left\{1 + |\lambda(a)|Z(a)\right\}, \tag{S.4}$$

where $\zeta(a)$ is such that $\lambda(a) = \zeta(a)|\lambda(a)|$ and $|\zeta(a)| = 1$, $\overline{T}(a) = \sum_{i=1}^n Y_i(a)/n = \hat{\mu}(a) - \mu(a)$, $\overline{\sigma}^2(a) = \sum_{i=1}^n Y_i^2(a)/n$, and $Z(a) = \max_{i=1,\ldots,n} |Y_i(a)|$. Here (S.4) is shown using the estimating equation

$$n^{-1}\sum_{i=1}^n \frac{Y_i(a)}{1 + \lambda(a)Y_i(a)} = 0 \tag{S.5}$$

and a similar argument to Section 11.2 of Owen (2001), which leads to

$$\zeta(a)\overline{T}(a) = \frac{|\lambda(a)|}{n}\sum_{i=1}^n \frac{Y_i^2(a)}{1 + \lambda(a)Y_i(a)},$$

and

$$|\lambda(a)|\overline{\sigma}^2(a) \le \left\{\frac{|\lambda(a)|}{n}\sum_{i=1}^n \frac{Y_i^2(a)}{1 + \lambda(a)Y_i(a)}\right\}\left\{1 + |\lambda(a)|Z(a)\right\}.$$

Then (S.4) and the large sample properties of $\overline{T}(a)$, $\overline{\sigma}^2(a)$ and $Z(a)$ are used to establish the asymptotic order of $\lambda(a)$ as follows. By Theorem S.1 (see Supplement Section 11),

we have $\overline{T}(a) = O_p(1/\sqrt{n})$. Furthermore,

$$\overline{\sigma}^2(a) = \sigma^2(a) + o(1) \tag{S.6}$$

a.s. This follows by Glivenko–Cantelli preservation (Kosorok, 2008b, Lemma 9.28), and the fact that $\mathcal{F}$ is P-Donsker (see Supplement Section 11) and strong P-Glivenko–Cantelli (Kosorok, 2008b, Lemma 8.17). In (S.6), note the assumption that $\inf_{a\in[\alpha_1,\alpha_2]} \sigma^2(a) > 0$ means $\overline{\sigma}^2(a) \geq \inf_{a\in[\alpha_1,\alpha_2]} \sigma^2(a) + o(1)$ is bounded away from 0 as $n \to \infty$ a.s.

As for $Z(a)$, we begin with showing $\sup_{a\in[\alpha_1,\alpha_2]} Z(a) \leq \max_{i=1,\ldots,n}\{V_i(\alpha_2)+|T_i(\alpha_1)|\}+\sup_{a\in[\alpha_1,\alpha_2]}|\mu(a)|$ by the decomposition of $T_i(a)$ and monotonicity. Utilizing the finite second moment conditions for $V(\alpha_2)$ and $T(\alpha_1)$, we obtain $\max_{i=1,\ldots,n} T_{\alpha_1,i}$, $\max_{i=1,\ldots,n} V_{\alpha_2,i} = o(\sqrt{n})$ a.s. (Owen, 2001, Lemma 11.2), and $\sup_{a\in[\alpha_1,\alpha_2]}|\mu(a)| \leq EV(\alpha_2) + E|T(\alpha_1)| < \infty$ by Jensen's inequality. The three strings of (in)equalities above imply $\sup_{a\in[\alpha_1,\alpha_2]} Z(a) = o(\sqrt{n})$ a.s. These large sample properties of $\overline{T}(a)$, $\overline{\sigma}^2(a)$ and $Z(a)$, together with (S.4), lead to $\lambda(a) = O_p(n^{-1/2})$.

Based on the asymptotic order and the fact that $\lambda(a)Y_i(a) = O_p(n^{-1/2})o(\sqrt{n}) = o_p(1)$, we apply Taylor's theorem and get

$$-2\log\mathcal{R}(\mu)(a) = 2n\lambda(a)\overline{T}(a) - n\lambda^2(a)\overline{\sigma}^2(a) + o_p(1). \tag{S.7}$$

We also expand (S.5) around 0 and get

$$0 = \frac{1}{n}\sum_{i=1}^{n} Y_i(a)\left\{1 - \lambda(a)Y_i(a) + O(\lambda^2(a)Y_i^2(a))\right\} = \overline{T}(a) - \lambda(a)\sigma^2(a) + o_p(n^{-1/2}),$$

which implies

$$\lambda(a) = \sigma^{-2}(a)\overline{T}(a) + o_p(n^{-1/2}) \tag{S.8}$$

and

$$\lambda(a)\overline{T}(a) = \lambda^2(a)\sigma^2(a) + o_p(n^{-1}). \tag{S.9}$$

Substituting (S.9) into (S.7) gives $-2\log\mathcal{R}(\mu)(a) = n\lambda(a)\overline{T}(a) + o_p(1)$. This and (S.8) imply

$$-2\log\mathcal{R}(\mu)(a) = \hat{\Psi}^2(a) + o_p(1), \tag{S.10}$$

where $\hat{\Psi}(a)$ is given in Remark 2 of Theorem 2.

As mentioned in the beginning of this Supplement Section, to study the asymptotics of $-2\log f_n(\mathcal{R}(\mu))(\cdot)$, we consider discretizing (S.10). Since $\sup_{a\in[\alpha_1,\alpha_2]}|-2\log f_n(\mathcal{R}(\mu)-\hat{\Psi}^2)(a)| \leq \sup_{a\in[\alpha_1,\alpha_2]}|-2\log\mathcal{R}(\mu)(a) - \hat{\Psi}^2(a)|$ by definition of f_n, we have

$$-2\log f_n(\mathcal{R}(\mu))(a) = f_n(\hat{\Psi})^2(a) + o_p(1). \tag{S.11}$$

The uniform convergence of the first term on the right hand side of (S.11) can only be established in $\ell^\infty([\alpha_1, \alpha_2] \setminus I_\delta)$, where I_δ can be constructed as $\bigcup_{\ell=1}^{Q}(d_\ell - \delta/Q, d_\ell)$, d_ℓ for $\ell = 1, \ldots, Q < \infty$ are the (finitely many) discontinuities of $\mu(\cdot)$ and $\sigma^2(\cdot)$, and δ can be any small positive number that is less than the minimal distance between any two d_ℓ's and between any d_ℓ and α_1 or α_2, divided by 2. The uniform convergence of the denominator $f_n(\sigma^2)(a)$ on the right hand side of (S.11) to $\sigma^2(a)$ is obtained as follows. First, $\sigma^2(a)$ can be shown to be of bounded variation and right continuous by the right-continuity of $T(a)$ and the dominated convergence theorem. This implies $\sigma^2(a)$ satisfies the almost uniform continuity property described in Supplement Section 1.1. Then for all $\varepsilon > 0$, there exists a $\delta_\sigma > 0$ such that for all $a, b \in [\alpha_1, \alpha_2]$, $a < b$ and $(a, b]$ does not include a discontinuity of magnitude $\geq \varepsilon/2$, $b - a < \delta_\sigma$ implies $\left|\sigma^2(b) - \sigma^2(a)\right| < \varepsilon$. Then there exists an $N = N(\varepsilon, \delta)$ such that for all $n > N$, $u_n < \min\{\delta/Q, \delta_\sigma\}$, where recall from Supplement Section 1.1 that u_n is the mesh of $\mathbf{G}_n$. Then for all $a \in [\alpha_1, \alpha_2] \setminus I_\delta$, $(a, a + u_n]$ does not include any discontinuities. Then the almost uniform continuity of $\sigma^2(a)$ and $u_n < \delta_\sigma$ imply

$$\sup_{a \in [\alpha_1, \alpha_2] \setminus I_\delta} \left| f_n(\sigma^2)(a) - \sigma^2(a) \right| < \varepsilon.$$

By this uniform convergence of $f_n(\sigma^2)(a)$, the first part of Theorem 1, Slutsky's Lemma and the continuous mapping theorem, we have the desired result.

5 Handling data that violate the nonzero variance condition

In this section we explain situations when violation of the nonzero variance condition can occur, and utilize existing approaches in the literature to deal with them.

5.1 Adapting Nair's two-step approach to the simultaneous confidence bands for the functional means

In constructing the EL and Wald-type EP bands, there can be situations when the endpoints α_1 and α_2 do not respect the $\inf_{a \in [\alpha_1, \alpha_2]} \sigma^2(a) > 0$ condition. One such example occurs in our application to occupation time, where the marginal variance of $L(a)$ can shrink to zero as a approaches α_2 (see Supplement Section 5.3 below for more details). Existing bands in the literature (Degras, 2011; Cao et al., 2012; Choi and Reimherr, 2018) have a similar problem because they use standardized estimators in forming the simultaneous confidence bands. In this case, one can still construct the bands, but with some modifications we describe as follows. Note in our simulation study in Section 3 we show the bands before and after modifications both work well, for the $\inf_{a \in [\alpha_1, \alpha_2]} \sigma^2(a) > 0$ and $\inf_{a \in [\alpha_1, \alpha_2]} \sigma^2(a) = 0$ scenarios, respectively.

For application to occupation time, to modify the bands to respect the nonzero variance condition, we adapt the following two-step approach that has been proposed in Nair (1984) for the equal precision confidence bands of a survival function, where zero variance occurs when the survival function is 0 or 1. First, construct the prescribed band up to a certain point $\hat{r} = \hat{r}(z)$ in terms of $z \in [0, 1]$, where $\hat{r} = \sup\{a : f_n(\hat{\mu})(a)/\tau > z\}$; we use $z = 0.05$ in our simulation study and data analysis, as it works well in our experience. Second, use a principled approach as follows to extend the confidence band beyond $\hat{r}$: for the upper boundary of the confidence bands beyond the right endpoint, we use the same upper bound (of the confidence bands) at $\hat{r}$, according to the monotonicity of the mean occupation time. As for the lower boundary of the confidence bands beyond $\hat{r}$, we use the lower bound of the mean occupation time 0 for the EL band and $l_{\text{cb}} I\{l_{\text{cb}} < 0\}$ for other bands, where l_{cb} is the lower bound of the specific band at $\hat{r}$. This $l_{\text{cb}} I\{l_{\text{cb}} < 0\}$ is devised to preserve monotonictiy (as best as possible) of the lower boundary when $l_{\text{cb}} < 0$.

5.2 Adapting Uno et al.'s selection approach to the functional ANOVA tests

The nonzero variance condition is also needed in certain functional ANOVA tests for which pointwise test statistics are standardized or of F-type. These include our EL test and existing Wald-type ANOVA tests such as the maximally-selected F-statistic of Zhang et al. (2019) and the integrated F-statistic GPF (Zhang and Liang, 2014). For any of the aformentioned tests, to deal with data that violate the nonzero variance condition, we can adapt the Uno et al. (2015) approach to automatically choose the test calibration that gives the smallest p-value among $\{B(z), z \in \mathbf{Z}\}$ as follows, where $B(z)$ is the test constructed up to $\hat{r}_k(z) = \min_{j=1,\ldots,k} \sup\{a : f_n(\hat{\mu}_j)(a)/\tau > z\}$ and $\mathbf{Z}$ is a grid of points in $[0, 1]$ generalizing the choice of z mentioned in the previous subsection. First, we use the bootstrap method mentioned in our manuscript to estimate the null distribution of $\{B(z), z \in \mathbf{Z}\}$. Let D denote the resulting 1000 bootstrapped values for $\{B(z), z \in \mathbf{Z}\}$. Then the p-value $p(z) = \mathrm{P}\{B(z) > b(z)\}$ can be estimated from D, where $b(z)$ is the value of $B(z)$ based on the observed data. Let $p_s = \min_{z \in \mathbf{Z}} p(z)$. A small p_s should lead to a rejection of H_0. Since this p-value is obtained after selection, we can no longer compare it with the original level of significance; instead, a post-selection approach is needed. The key is to approximate the null distribution of $P_s = \min\{P(z) : z \in \mathbf{Z}\}$, where $P(z) = S_{B(z)}(B(z))$ and $S_{B(z)}(b)$ is the survival function of $B(z)$. This can be done by using D to compute one minus the empirical cumulative distribution function (cdf) of $B(z)$ at each $B(z)$, leading to an estimate for $P(z)$ and hence P_s. This allows us to estimate the post-selection p-value $\mathrm{P}(P_s < p_s)$ of the test accordingly. Note that this procedure does not require double bootstrap nor a huge $\mathbf{Z}$—we find a very simple grid $\mathbf{Z} = \{0, 0.01, 0.05\}$ suffices in our simulations and data analysis.

5.3 Variances of occupation time at extreme values of a

The marginal variance of $L(a)$ near $a = 0$ is bounded away from 0 because different subjects spent different time in 0 activity. However, the marginal variance of $L(a)$ can shrink to zero as a approaches α_2, for example when $L(a) \to 0$ a.s. as $a \to \alpha_2$. This may be due to the fact that α_2 (e.g., if we set it to the maximal activity level in the device) is not achievable by the study subjects. By the dominated convergence theorem, we then have $\mathrm{Var}\{L(a)\} \to 0$ as $a \to \alpha_2$.

6 Bootstrap in Section 2.3

6.1 Proof of Corollary 2

It suffices to show that $\sup_{a\in[\alpha_1,\alpha_2]\setminus I_\delta} \left|f_n(\hat{S} - \sigma)(a)\right| \to 0$ a.s. This, the uniform convergence of $f_n(\sigma^2)(a)$ to $\sigma^2(a)$ in the last paragraph of Supplement Section 4, the bootstrap consistency of $f_n(U_n^*)(a)$ in Supplement Section 3.1, and a routine extension of the proof for the conditional Slutsky's lemma in Cheng (2015) to the case of random elements of a metric space, imply that $[f_n(U_n^*), f_n(\hat{S})]^T$ is bootstrap consistent for $[U, \sigma]^T$ in $\{\ell^\infty([\alpha_1, \alpha_2] \setminus I_\delta)\}^2$. The desired result follows by the continuous mapping theorem for the bootstrap (see, e.g., Kosorok, 2008b, Theorem 10.8).

The first sentence in the previous paragraph is true due to the strong consistency of $\hat{S}$ for σ in $\ell^\infty([\alpha_1, \alpha_2])$ and the fact that $\sup_{a\in[\alpha_1,\alpha_2]} \left|f_n(\hat{S} - \sigma)(a)\right| \le \sup_{a\in[\alpha_1,\alpha_2]} \left|\hat{S}(a) - \sigma(a)\right|$. The strong consistency of $\hat{S}$ for σ is in turn due to Glivenko–Cantelli preservation (Kosorok, 2008b, Lemma 9.28) and the fact that $\mathcal{F}$ is P-Donsker (see Supplement Section 11).

6.2 Details in implementing bootstrap and computing $c^*_{EL,\alpha}$ and $c^*_{EP,\alpha}$

A bootstrap sample is obtained by drawing n curves independently with replacement from the data $\{f_n(T_1)(a), \ldots, f_n(T_n)(a), a \in [\alpha_1, \alpha_2]\}$. Based on this bootstrap sample, compute a value of M_n^* and a value of $\sup_{a\in[\alpha_1,\alpha_2]} \left|f_n(\hat{\Psi}^*)(a)\right|$, where recall from Sections 2.2 and 2.3 that $M_n^* = \sup_{a\in[\alpha_1,\alpha_2]} f_n(\hat{\Psi}^*)^2(a)$, $\hat{\Psi}^*(a) = U_n^*(a)/\hat{S}(a)$, $U_n^*(a) = \sqrt{n}\{\hat{\mu}^*(a) - \hat{\mu}(a)\}$, $\hat{\mu}(a) = \sum_{i=1}^n T_i(a)/n$, $\hat{\mu}^*(a) = \sum_{i=1}^n W_{ni} T_i(a)/n$, W_{ni} is the number of times that $f_n(T_i)(a)$ is redrawn from $\{f_n(T_1)(a), \ldots, f_n(T_n)(a), a \in [\alpha_1, \alpha_2]\}$, and $\hat{S}(a) = [\sum_{i=1}^n \{T_i(a) - \hat{\mu}(a)\}^2/n]^{1/2}$ is the sample version of $\sigma(a)$.

Repeat the procedure in the previous paragraph B times to obtain B bootstrapped values for M_n^* and B bootstrapped values for $\sup_{a\in[\alpha_1,\alpha_2]} \left|f_n(\hat{\Psi}^*)(a)\right|$; our simulation study (see Section 3) uses $B = 1000$. Let $c^*_{EL,\alpha}$ and $c^*_{EP,\alpha}$ denote the upper α-quantile of these B bootstrapped values of M_n^* and $\sup_{a\in[\alpha_1,\alpha_2]} \left|f_n(\hat{\Psi}^*)(a)\right|$, respectively. Then the desired asymptotic $100(1-\alpha)\%$ EL simultaneous confidence band for $\mu(\cdot)$ for essentially all $a \in [\alpha_1, \alpha_2]$ is computed as $\{(a, \tilde{\mu}(a)) : -2 \log f_n(\mathcal{R}(\tilde{\mu}))(a) \le c^*_{EL,\alpha}, \tilde{\mu} \in \mathcal{D}_n\}$,

where $\mathcal{D}_n$ is defined back in Section 2.3 as the class of functions of the form in (1). The asymptotic $100(1-\alpha)\%$ EP simultaneous confidence band for $\mu(\cdot)$ is computed as $f_n(\hat{\mu})(a) \pm n^{-1/2} c^*_{EP,\alpha} f_n(\hat{S})(a)$ for essentially all $a \in [\alpha_1, \alpha_2]$.

Regarding performance, the use of the aforementioned bootstrap method is computationally more efficient than an alternative bootstrap procedure based on calculating the local EL ratio in each of the bootstrap sample (see, e.g., Owen, 2001, Section 3.3 for the bootstrap-calibrated EL for a multivariate mean), since the alternative calculation requires performing optimization for each $a \in \mathbf{G}_n$. As for the theoretical performance, both bootstrap procedures are asymptotically first-order equivalent. A more detailed comparison could be done based on their higher-order properties (see, e.g., Hall, 1992), but this is outside the scope of the current paper.

7 The EL confidence band

7.1 Construction and asymptotic consistency

For each $\delta > 0$ defined in Supplement Section 1.1, to obtain an EL-based counterpart of the Wald-type NS band (see Section 2.2)

$$\left\{(a, \tilde{\mu}(a)) : \sqrt{n}\,|\{f_n(\hat{\mu}) - \tilde{\mu}\}(a)| \le c^*_{NS,\alpha}, a \in [\alpha_1, \alpha_2] \setminus I_\delta\right\},$$

the complete trajectory of $\tilde{\mu}(a)$ (instead of its discretized version) is needed in the local EL statistic. Since $-2\log f_n(\mathcal{R}(\tilde{\mu})(\cdot)$ depends on $f_n(\tilde{\mu})(a)$ rather than $\tilde{\mu}(a)$, it is replaced by $\tilde{\mu}(a)$, leading to

$$-2\log \check{\mathcal{R}}(\tilde{\mu})(a) = 2\sum_{i=1}^{n} \log[1 + f_n(\tilde{\lambda})(a)\{f_n(T_i)(a) - \tilde{\mu}(a)\}].$$

Then the EL-based counterpart of the Wald-type NS band is

$$\left\{(a, \tilde{\mu}(a)) : -2\log \check{\mathcal{R}}(\tilde{\mu})(a) \le c^*_{EL,\alpha}, a \in [\alpha_1, \alpha_2] \setminus I'_\delta\right\}, \tag{S.12}$$

where I'_δ is the I_δ defined at the last paragraph of Supplement Section 4. Note the band in (S.12) is equivalent to the EL simultaneous confidence band in Section 2.3 by definition.

We now show that the band in (S.12) has asymptotic confidence level $100(1-\alpha)\%$. For simplicity of exposition, the sup notation below represents supremum over $a \in [\alpha_1, \alpha_2] \setminus I'_\delta$; let $M_{n,\delta} = \sup\{-2\log f_n(\mathcal{R}(\mu))(a)\}$ and $M^*_{n,\delta} = \sup f_n(\hat{\Psi}^*)^2(a)$. By the bootstrap consistency result of Corollary 2, a similar reasoning as in Lemma 23.3 of van der Vaart (2000) (by replacing $(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, $(\hat{\theta}^*_n - \hat{\theta}_n)/\hat{\sigma}^*_n$, T and $\hat{\xi}_{n,\alpha}$ there with $M_{n,\delta}$, $M^*_{n,\delta}$, $\sup\{U^2(a)/\sigma^2(a)\}$ and $c^*_{EL,\alpha}$), and the fact that the limiting random vari-

able $\sup\{U^2(a)/\sigma^2(a)\}$ has a continuous limiting distribution, it suffices to show that $\sup -2\log\check{\mathcal{R}}(\mu)(a)$ has the same limiting distribution as $M_{n,\delta}$. This follows from

$$-2\log\check{\mathcal{R}}(\mu)(a)\xrightarrow{d}U^2(a)/\sigma^2(a) \tag{S.13}$$

in $\ell\{[\alpha_1,\alpha_2]\setminus I'_\delta\}$ as $n\to\infty$ and the continuous mapping theorem. To show (S.13), first rearrange the terms to see that $-2\log\check{\mathcal{R}}(\mu)(a) = 2\sum_{i=1}^n \log[1+f_n(\lambda Y_i)(a)+f_n(\lambda)(a)\{f_n(\mu)(a)-\mu(a)\}]$. We consider discretizing the large sample results in Supplement Section 4: we have $f_n(\lambda Y_i)(a)=o_p(1)$ due to $\sup_{a\in[\alpha_1,\alpha_2]}|f_n(\lambda Y_i)(a)|\le \sup_{a\in[\alpha_1,\alpha_2]}|\lambda Y_i(a)|$ by the definition of f_n. A similar reasoning will be used in the following discretization of the large sample results. Discretizing the result $\lambda(a)=O_p(1/\sqrt{n})$ in Supplement Section 4, we have $f_n(\lambda)(a)=O_p(1/\sqrt{n})$. This and the result in Supplement Section 1.2 (based on the conditions of the second part of Theorem 1) gives $n\sup[f_n(\lambda)(a)\{f_n(\mu)(a)-\mu(a)\}]\xrightarrow{P}0$. Applying Taylor's theorem, we get

$$\sup\left\{-2\log\check{\mathcal{R}}(\mu)(a)-2nf_n(\lambda\overline{T})(a)+nf_n(\lambda^2\overline{\sigma}^2)(a)\right\}\xrightarrow{P}0. \tag{S.14}$$

Discretizing (S.9) using a similar reasoning for obtaining $f_n(\lambda Y_i)(a)=o_p(1)$ above, we have $f_n(\lambda\overline{T}-\lambda^2\sigma^2)(a)=o_p(n^{-1})$, so that $f_n(\lambda\overline{T})(a)=f_n(\lambda^2\sigma^2)(a)+o_p(n^{-1})$ by linearity of f_n. Substituting this into (S.14) gives $\sup\{-2\log\check{\mathcal{R}}(\mu)(a)-nf_n(\lambda\overline{T})(a)\}\xrightarrow{P}0$. This and the discretized (S.8) imply

$$\sup\left\{-2\log\check{\mathcal{R}}(\mu)(a)-f_n(\hat{\Psi})^2(a)\right\}\xrightarrow{P}0.$$

Then (S.13) follows by the same reasoning as the last paragraph of Supplement Section 4.

Remark. From the above proof, we can see that the conditions of the second part of Theorem 1 only come into play when we deal with the asymptotics relevant to the complete trajectory of $\mu(a)$. This complete trajectory is needed in constructing the confidence bands, but not in constructing the asymptotics of the discretized EL statistics in Sections 2.3 and 2.4. Therefore, besides the nonzero variance conditions, Theorems 2 and 3 only involve the conditions of the first part of Theorem 1 or their k-sample version.

7.2 Monotonicity of the EL band in a

In this section, we show if the observed processes are monotone in a, as is the case for $L(a)$, then the lower and upper boundaries of the EL band will respect this monotonicity. We first illustrate this monotonicity in Figure S.1 below based on a simulation model for $L(a)$ in Section 3.1, and then provide a proof later. In Figure S.1, we can see $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a_j)$ as a function of $\tilde{\mu}(a_j)$ is U-shaped for each of the given a_j, $j=1,2$.

Futhermore, for $a_1 \leq a_2$, the function $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a_1)$ as a function of $\tilde{\mu}(a_2)$ is to the right of $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a_2)$ as a function of $\tilde{\mu}(a_2)$. Hence, a horizontal line through $c^*_{EL,\alpha}$ will intersect the sample path of $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a_2)$ earlier than that of $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a_1)$. Therefore, the resulting confidence region for $\mu(a_2)$ is to the left of that of $\mu(a_1)$.

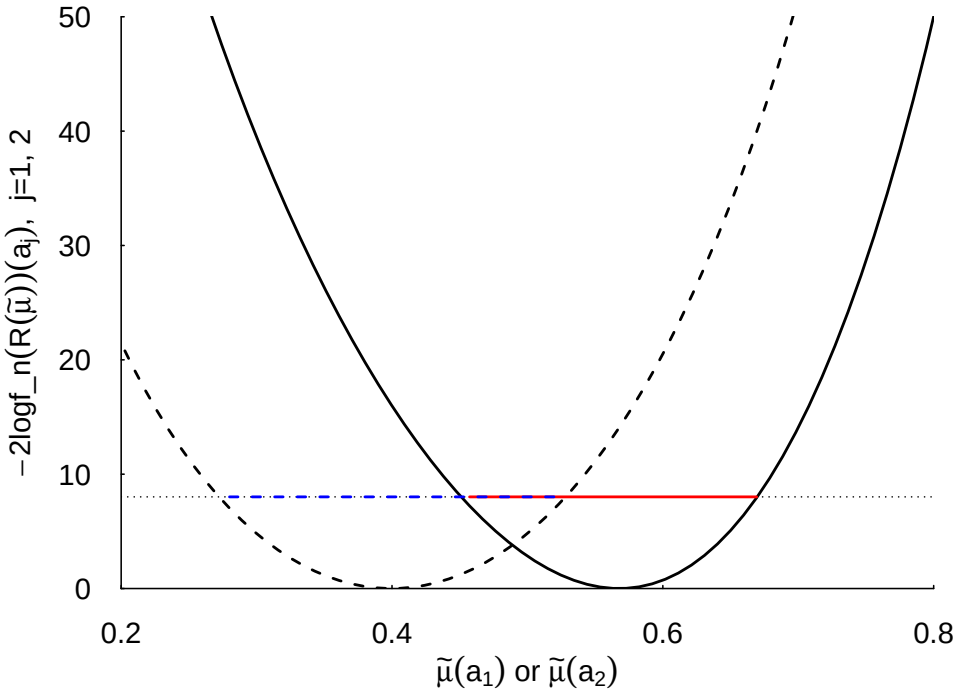

Figure S.1: Simulated $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a)$ for some $a = a_1$ (solid) or a_2 (dashed) such that $a_1 \leq a_2$, under the simulation model used to assess the performance of the confidence bands, $\nu = 2$ and $n = 50$. The dotted horizontal line is at $c^*_{EL,\alpha}$, the solid red line segment is the confidence region for $\mu(a_1)$, and the dashed blue line segment is the confidence region for $\mu(a_2)$.

Without loss of generality, suppose the observed process $T(a)$ is non-increasing in a. We first prove for each a, $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a)$ is U-shaped as a function of $\tilde{\mu}(a)$. Second, we show for $a_1 \leq a_2$ and b such that $\tilde{\mu}(a) = b$ over $a \in [\alpha_1, \alpha_2]$,

(i) if $b \leq \hat{\mu}(a_2)$, then $-2\log f_n(R(\tilde{\mu}))(a_1) \geq -2\log f_n(R(\tilde{\mu}))(a_2)$

(ii) if $b \geq \hat{\mu}(a_1)$, then $-2\log f_n(R(\tilde{\mu}))(a_1) \leq -2\log f_n(R(\tilde{\mu}))(a_2)$

(iii) if $\hat{\mu}(a_2) \leq b \leq \hat{\mu}(a_1)$, then the magnitude of $-2\log f_n(R(\tilde{\mu}))(a_1)$ versus $-2\log f_n($ $R(\tilde{\mu}))(a_2)$ will be determined by how close b is to $\hat{\mu}(a_2)$ and $\hat{\mu}(a_1)$: if b is closer

to $\hat{\mu}(a_2)$, then $-2\log f_n(R(\tilde{\mu}))(a_1) \geq -2\log f_n(R(\tilde{\mu}))(a_2)$, and vice versa if b is closer to $\hat{\mu}(a_1)$.

This can easily be understood graphically from Figure S.1. Finally, by the argument in the previous paragraph, we have that the lower bound of the line segment $\{(a_1, \tilde{\mu}(a_1)) : -2\log f_n(\mathcal{R}(\tilde{\mu}))(a_1) \leq c^*_{EL,\alpha}\}$ is no less than that of the line segment $\{(a_2, \tilde{\mu}(a_2)) : -2\log f_n(\mathcal{R}(\tilde{\mu}))(a_2) \leq c^*_{EL,\alpha}\}$, and the same holds for the upper bounds.

Given $a \in [\alpha_1, \alpha_2]$, to show the sample path of $-2\log f_n(\mathcal{R}(\tilde{\mu}))(a)$ is U-shaped in $\tilde{\mu}(a)$, it suffices to show $-2\log \mathcal{R}(\tilde{\mu})(b_a)$ is U-shaped in $\tilde{\mu}(b_a)$. Thus, we restrict attention to $a \in \mathbf{G}_n$ in studying the shape of $-2\log \mathcal{R}(\tilde{\mu})(a)$ as follows. First note that $-2\log R(\tilde{\mu})(a) = 2\sum_{i=1}^n \log[1 + \tilde{\lambda}(a)\{T_i(a) - \tilde{\mu}(a)\}]$ is a function of $n+1$ of the $n+2$ variables $\tilde{\lambda}(a)$, $T_i(a)$ and $\tilde{\mu}(a)$, because one of the variables will be determined by the estimating equation $\sum_{i=1}^n \tilde{p}_i(a)\{T_i(a) - \tilde{\mu}(a)\} = 0$, where recall from Supplement Section 4 that $\tilde{p}_i(a) = [n\{1 + \tilde{\lambda}(a)(T_i - \tilde{\mu})(a)\}]^{-1}$. Since the data $T_i(a)$ changes with a and we have fixed a here, $-2\log R(\tilde{\mu})(a)$ can be viewed as a function of $\tilde{\mu}(a)$ only. The derivative of $-2\log R(\tilde{\mu})(a)$ with respect to $\tilde{\mu}(a)$ can be shown to be $-2n\tilde{\lambda}(a)$. Next, to study the sign of the derivative, we study how the sign of $\tilde{\lambda}(a)$ is affected as $\tilde{\mu}(a)$ changes. The relationship between $\tilde{\lambda}(a)$ and $\tilde{\mu}(a)$ can be seen from the estimating equation $\sum_{i=1}^n \tilde{p}_i(a)\{T_i(a) - \tilde{\mu}(a)\} = 0$. Rewriting the equation as $\sum_{i:T_i(a)\neq\tilde{\mu}(a)} 1/\{1/(T_i - \tilde{\mu})(a) + \tilde{\lambda}(a)\}/n = 0$ and investigating the partial derivatives of its left-hand side, we see that the left-hand side $\equiv \mathrm{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ is non-increasing in $\tilde{\lambda}(a)$ and $\tilde{\mu}(a)$, respectively. When $\tilde{\mu}(a) = \hat{\mu}(a)$, we have $\tilde{\lambda}(a) = 0$ and $-2\log R(\tilde{\mu})(a) = 0$. As $\tilde{\mu}(a)$ decreases from $\hat{\mu}(a)$, $\mathrm{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ increases, and thus $\tilde{\lambda}(a)$ needs to increase in order to make $\mathrm{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ decrease back to 0 (to satisfy the estimating equation). This means for $\tilde{\mu}(a) \leq \hat{\mu}(a)$, $\tilde{\lambda}(a) \geq 0$, leading to a non-positive derivative of $-2\log R(\tilde{\mu})(a)$ with respect to $\tilde{\mu}(a)$. By a similar reasoning, we can show for $\tilde{\mu}(a) \geq \hat{\mu}(a)$, $\tilde{\lambda}(a) \leq 0$, leading to a non-negative derivative of $-2\log R(\tilde{\mu})(a)$ with respect to $\tilde{\mu}(a)$. Thus, $-2\log \mathcal{R}(\tilde{\mu})(a)$ is U-shaped in $\tilde{\mu}(a)$.

Secondly, to show (i)–(iii) in the second paragraph of this Section, by a similar reasoning as in the previous paragraph, it suffices to restrict attention to $a \in \mathbf{G}_n$ and study (i)–(iii) in terms of $-2\log \mathcal{R}(\tilde{\mu})(a)$. Since $b = \tilde{\mu}(a_1) = \tilde{\mu}(a_2)$, the change from a_1 to a_2 results in changes in $T_i(a)$, and $\tilde{\lambda}(a_j)$ $(j = 1, 2)$ will be determined by the estimating equation as mentioned in the previous paragraph, it amounts to study the shape of $-2\log R(\tilde{\mu})(a)$ as a function of $T_i(a)$, $i = 1, \ldots, n$, while holding $\tilde{\mu}(a)$ fixed at b. First we derive the (partial) derivative of $-2\log R(\tilde{\mu})(a)$with respect to $T_i(a)$ as

$$2n\tilde{\lambda}(a)\tilde{p}_i(a),$$

$i = 1, \ldots, n$. To study the sign of this derivative, we study how the sign of $\tilde{\lambda}(a)$ is affected as each $T_i(a)$ changes. We use the fact that $\mathrm{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ is non-

increasing in $\tilde{\lambda}(a)$ (by the previous paragraph) and non-decreasing in each $T_i(a)$ because $\partial \text{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)/\partial T_i(a) \geq 0$. As each $T_i(a)$ increases, $\text{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ increases, and thus $\tilde{\lambda}(a)$ needs to increase in order to make $\text{LHS}(\tilde{\mu}, \tilde{\lambda}, T_1, \ldots, T_n)(a)$ decrease back to 0 (to satisfy the estimating equation). Since $-2 \log R(\tilde{\mu})(a)$ would be 0 at some point $(T_1(a), \ldots, T_n(a)) = (t_1, \ldots, t_n)$ such that their sample mean equals $\tilde{\mu}(a)$, as each $T_i(a)$ increases from this point, $\tilde{\lambda}(a)$ increases and is non-negative. Thus, by the partial derivative of $-2 \log R(\tilde{\mu})(a)$ displayed above, we have that $-2 \log R(\tilde{\mu})(a)$ increases. Likewise, as each $T_i(a)$ decreases from this point, $\tilde{\lambda}(a)$ decreases and is non-positive, so $-2 \log R(\tilde{\mu})(a)$ increases. These imply that $-2 \log R(\tilde{\mu})(a)$ as a function of each $T_i(a)$ is U-shaped as well, with the minimum occurring at $(t_1, \ldots, t_n)$.

Thus, to study (i)–(iii) in terms of $-2 \log \mathcal{R}(\tilde{\mu})(a)$ for $a \in \mathbf{G}_n$, we first determine the point $(t_1, \ldots, t_n)$ where $-2 \log R(\tilde{\mu})(a)$ as a function of $(T_1(a), \ldots, T_n(a))$ hits 0, and then see which side $(T_1(a_1), \ldots, T_n(a_1))$ and $(T_1(a_2), \ldots, T_n(a_2))$ fall onto the U-shaped $-2 \log R(\tilde{\mu})(a)$ curve. Under (i) when $b \leq \hat{\mu}(a_2)$, suppose that $-2 \log R(\tilde{\mu})(a)$ as a function of $(T_1(a), \ldots, T_n(a))$ would be 0 at the set of coordinates $(t_1, \ldots, t_n)$ such that their sample mean equals b. This set can be chosen to be pointwise no greater than $(T_1(a_2), \ldots, T_n(a_2))$ by $b \leq \hat{\mu}(a_2)$, so that $(T_1(a_2), \ldots, T_n(a_2))$ is on the right-hand arm of this U-shaped $-2 \log R(\tilde{\mu})(a)$ curve. By $\hat{\mu}(a_2) \leq \hat{\mu}(a_1)$ and monotonicity of $T_i(a)$ we know $(T_1(a_1), \ldots, T_n(a_1))$ is further away from $(t_1, \ldots, t_n)$ than $(T_1(a_2), \ldots, T_n(a_2))$. In this case, $-2 \log R(\tilde{\mu})(a_1) \geq -2 \log R(\tilde{\mu})(a_2)$. Similarly, we can show under (ii) when $b \geq \hat{\mu}(a_1)$, $-2 \log R(\tilde{\mu})(a_1) \leq -2 \log R(\tilde{\mu})(a_2)$. Furthermore, under (iii) when $\hat{\mu}(a_2) \leq b \leq \hat{\mu}(a_1)$, the point $(t_1, \ldots, t_n)$ where $-2 \log R(\tilde{\mu})(a) = 0$ can be chosen to be pointwise no less than $(T_1(a_2), \ldots, T_n(a_2))$ and no greater than $(T_1(a_1), \ldots, T_n(a_1))$. This means $(T_1(a_2), \ldots, T_n(a_2))$ and $(T_1(a_1), \ldots, T_n(a_1))$ will be on the left-arm and right-arm of the U-shaped $-2 \log R(\tilde{\mu})(a)$ curve, respectively. As a result, the magnitude of $-2 \log R(\tilde{\mu})(a_1)$ versus $-2 \log R(\tilde{\mu})(a_2)$ will be determined by how close $(T_1(a_2), \ldots, T_n(\ a_2))$ and $(T_1(a_1), \ldots, T_n(a_1))$ are to the point $(t_1, \ldots, t_n)$ where $-2 \log R(\tilde{\mu})(a) = 0$, or equivalently how close b is to $\hat{\mu}(a_2)$ and $\hat{\mu}(a_1)$. When the point $(t_1, \ldots, t_n)$ is closer to $(T_1(a_2), \ldots, T_n(a_2))$ (or equivalent when b is closer to $\hat{\mu}(a_2)$), $-2 \log R(\tilde{\mu})(a_1) \geq -2 \log R(\tilde{\mu})(a_2)$, and otherwise $-2 \log R(\tilde{\mu})(a_1) \leq -2 \log R(\tilde{\mu})(a_2)$. This establishes the desired second claim.

8 Proof of Theorem 3

For simplicity of exposition, we first study the large sample behavior of the fully observed trajectories of the EL statistic process $-2 \log \mathcal{R}_k(\cdot)$. Then we discretize this large sample result and study the asymptotics of $-2 \log f_n(\mathcal{R}_k)(\cdot)$ accordingly. But note that, as in Sections 2.3 and 2.4, the quantities we use in the process of this proof are available to us only in their discretized forms.

Clearly the denominator of (3) is $\prod_{j=1}^{k}\prod_{i=1}^{n_j}(1/n_j)$. The constrained optimum in the numerator of (3) can be found by the Lagrange method and is given by $\prod_{j=1}^{k}\prod_{i=1}^{n_j}\hat{p}_{ij}(a)$, where $\hat{p}_{ij}(a) = n^{-1}\hat{\gamma}_j^{-1}[1 + \{\hat{\lambda}_{j-1}(a) - \hat{\lambda}_j(a)\}g_{ij}(\hat{\mu}_{j,EL})(a)]^{-1}$, $\hat{\gamma}_j = n_j/n$ (as defined in Section 2.4), $g_{ij}(\tilde{\mu}_j)(a) = \{T_{ij}(a) - \tilde{\mu}_j(a)\}/\hat{\gamma}_j$ for any given $\tilde{\mu}(a)$ $(i = 1, \ldots, n_j)$, $\hat{\lambda}_0(a) \equiv \hat{\lambda}_k(a) \equiv 0$, $\hat{\mu}_{j,EL}(a)$ and $\hat{\lambda}_j(a)$ satisfy the system of estimating equations $\sum_{i=1}^{n_j}\hat{p}_{ij}(a)g_{ij}(\hat{\mu}_{j,EL})$ $(a) = 0$ and $\hat{\mu}_{j+1,EL}(a) - \hat{\mu}_{j,EL}(a) = 0$ for $j = 1, \ldots, k$, and $\hat{\mu}_{k+1,EL}(a) \equiv \hat{\mu}_{k,EL}(a)$. This results in

$$-2\log\mathcal{R}_k(a) = 2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\log\left[1 + \left\{\hat{\lambda}_{j-1}(a) - \hat{\lambda}_j(a)\right\}g_{ij}(\hat{\mu}_{j,EL})(a)\right]. \tag{S.15}$$

Note that $g_{ij}(\hat{\mu}_{j,EL})(a) = o(\sqrt{n})$ a.s., because $|g_{ij}(\hat{\mu}_{j,EL})(a)| \le \max_{1\le i\le n_j}|g_{ij}(\hat{\mu}_{j,EL})(a)|$, which is shown to be $o(\sqrt{n})$ a.s. in Supplement Section 9.

Let $\widehat{\Delta}_j(a) = \hat{\lambda}_{j-1}(a) - \hat{\lambda}_j(a)$ for $j = 1, \ldots, k$. By the fact that $\widehat{\Delta}_j(a) = O_p(n^{-1/2})$ (see Supplement Section 9) and $g_{ij}(\hat{\mu}_{j,EL})(a) = o(\sqrt{n})$ a.s., we apply Taylor's theorem to (S.15) and get

$$\begin{aligned}
-2\log\mathcal{R}_k(a) &= 2\sum_{j=1}^{k}\left\{\widehat{\Delta}_j(a)\sum_{i=1}^{n_j}g_{ij}(\hat{\mu}_{j,EL})(a) - \frac{\widehat{\Delta}_j^2(a)}{2}\sum_{i=1}^{n_j}g_{ij}^2(\hat{\mu}_{j,EL})(a)\right\} + o_p(1)\\
&= 2\sum_{j=1}^{k}\left[\widehat{\Delta}_j(a)\sum_{i=1}^{n_j}g_{ij}(\mu_j)(a) - n\widehat{\Delta}_j(a)\left\{\hat{\mu}_{j,EL}(a) - \mu_j(a)\right\} - \frac{n\widehat{\Delta}_j^2(a)}{2}\theta_j(a)\right] + o_p(1),
\end{aligned} \tag{S.16}$$

where $\theta_j(a) = \sigma_j^2(a)/\gamma_j$, and the last equality follows by (S.27), (S.28) and the fact that $\hat{\mu}_{j,EL}(a) - \mu_j(a) = O_p(n^{-1/2})$ (see Supplement Section 9). By expanding $\hat{\gamma}_j\sum_{i=1}^{n_j}\hat{p}_{ij}(a)$ $g_{ij}(\hat{\mu}_{j,EL})(a) = 0$ around $(\hat{\mu}_{j,EL}(a), \widehat{\Delta}_j(a)) = (\mu_j(a), 0)$, we get

$$0 = \frac{1}{n}\sum_{i=1}^{n_j}g_{ij}(\mu_j)(a) - \frac{\widehat{\Delta}_j(a)}{n}\sum_{i=1}^{n_j}g_{ij}^2(\mu_j)(a) - \{\hat{\mu}_{j,EL}(a) - \mu_j(a)\} + o_p(n^{-1/2}), \tag{S.17}$$

$j = 1, \ldots, k$. Multiplying both sides by $n\widehat{\Delta}_j(a)$, we have $\widehat{\Delta}_j(a)\sum_{i=1}^{n_j}g_{ij}(\mu_j)(a) - n\widehat{\Delta}_j(a)\{\hat{\mu}_{j,EL}(a) - \mu_j(a)\} = \widehat{\Delta}_j^2(a)\sum_{i=1}^{n_j}g_{ij}^2(\mu_j)(a) + o_p(1) = n\widehat{\Delta}_j^2(a)\theta_j(a) + o_p(1)$. This implies that (S.16) can be written as

$$\sum_{j=1}^{k}\widehat{\Delta}_j^2(a)n\theta_j(a) + o_p(1). \tag{S.18}$$

On the other hand, by $\hat{\mu}_{j,EL}(a) = \hat{\mu}_{j+1,EL}(a)$ and $\mu_j(a) = \mu_{j+1}(a)$ for $j = 1, \ldots, k-1$, (S.17) leads to $\widehat{\Delta}_j(a)\sum_{i=1}^{n_j}g_{ij}^2(\mu_j)(a)/n - \widehat{\Delta}_{j+1}\sum_{i=1}^{n_{j+1}}g_{i,j+1}^2(\mu_{j+1})(a)/n = \sum_{i=1}^{n_j}g_{ij}$ $(\mu_j)(a)/n - \sum_{i=1}^{n_{j+1}}g_{i,j+1}(\mu_{j+1})(a)/n + o_p(n^{-1/2})$. This can be expressed in matrix form

as

$$-\Theta(a)\hat{\boldsymbol{\lambda}}(a) = \hat{\boldsymbol{\psi}}(a) + o_p(n^{-1/2}), \tag{S.19}$$

where $\hat{\boldsymbol{\psi}}(a) \equiv [\bar{g}_1 - \bar{g}_2, \ldots, \bar{g}_{k-1} - \bar{g}_k]^T$, $\hat{\boldsymbol{\lambda}}(a) \equiv [\hat{\lambda}_1(a), \ldots, \hat{\lambda}_{k-1}(a)]^T$, $\Theta(a)$ is the $(k-1) \times (k-1)$ non-singular tridiagonal matrix

$$\begin{bmatrix} \hat{\theta}_1 + \hat{\theta}_2 & -\hat{\theta}_2 & 0 & \cdots & & & 0 \\ -\hat{\theta}_2 & \hat{\theta}_2 + \hat{\theta}_3 & -\hat{\theta}_3 & 0 & \cdots & & 0 \\ 0 & -\hat{\theta}_3 & \hat{\theta}_3 + \hat{\theta}_4 & -\hat{\theta}_4 & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \vdots & & & & & & 0 \\ 0 & \cdots & & & -\hat{\theta}_{k-2} & \hat{\theta}_{k-2} + \hat{\theta}_{k-1} & -\hat{\theta}_{k-1} \\ 0 & \cdots & & & 0 & -\hat{\theta}_{k-1} & \hat{\theta}_{k-1} + \hat{\theta}_k \end{bmatrix}, \tag{S.20}$$

$\bar{g}_j = \bar{g}_j(\mu_j)(a)$, $\hat{\theta}_j = \hat{\theta}_j(\mu_j)(a)$, and $\bar{g}_j(\tilde{\mu}_j)(a) = n^{-1}\sum_{i=1}^{n_j} g_{ij}(\tilde{\mu}_j)(a)$ and $\hat{\theta}_j(\tilde{\mu}_j)(a) = n^{-1}\sum_{i=1}^{n_j} g_{ij}^2(\tilde{\mu}_j)(a)$ for any given $\tilde{\mu}_j(a)$. (Note in the definition of $\hat{\boldsymbol{\psi}}(a)$ and $\Theta(a)$ above we surppress the dependence on $\mu_j(a)$ and a.) In general a recursive algorithm is used for inverting tridiagonal matrices (Usmani, 1994), but we can take advantage of the more specific structure of $\Theta(a)$ and obtain an explicit expression in a similar fashion as in the supplement of Chang and McKeague (2019). The latter paper analyzes a matrix having the same structure as $\Theta(a)$, except with different $\hat{\theta}_j$ (derived in the setting of a nonparametric k-sample testing problem with right-censored survival data). Multiplying both sides of $\hat{\boldsymbol{\lambda}}(a) = -\Theta^{-1}(a)\{\hat{\boldsymbol{\psi}}(a) + o_p(n^{-1/2})\}$ by $-\Lambda$, where

$$\Lambda \equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & & & 0 \\ -1 & 1 & 0 & 0 & \cdots & & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & & \\ \vdots & & & & & & 0 \\ 0 & \cdots & & & 0 & -1 & 1 \\ 0 & \cdots & & & 0 & 0 & -1 \end{bmatrix}_{k \times (k-1)},$$

we obtain

$$\widehat{\boldsymbol{\Delta}}(a) = \Lambda\Theta^{-1}(a)\left\{\Lambda^T \boldsymbol{A}(a) + o_p(n^{-1/2})\right\}, \tag{S.21}$$

where $\widehat{\boldsymbol{\Delta}}(a) = -\Lambda\hat{\boldsymbol{\lambda}}(a)$ and $\boldsymbol{A}(a) = [\bar{g}_1(\mu_1(a)), \ldots, \bar{g}_k(\mu_k(a))]^T$. By inserting the ex-

plicit expression for $\Theta^{-1}(a)$ along the lines of Supplement Section S.1.2 of Chang and McKeague (2019), we obtain

$$[\Lambda\Theta^{-1}(a)\Lambda^T]_{ij}(a) = \begin{cases} \dfrac{1}{\hat{\theta}_i(\mu)(a)\hat{\phi}(a)} \displaystyle\sum_{l\neq i} \prod_{g\in E_l} \hat{\theta}_g(\mu)(a), & i = j, \\ -\dfrac{1}{\hat{\theta}_i(\mu)(a)\hat{\phi}(a)} \displaystyle\prod_{g\in E_j} \hat{\theta}_g(\mu)(a), & i \neq j, \end{cases} \tag{S.22}$$

where $E_j = \{1,\ldots,k\} \setminus \{j\}$ and $\hat{\phi}(a) = \sum_{l=1}^k \prod_{g\in E_l} \hat{\theta}_g(\mu)(a)$. Inserting (S.22) into (S.21), algebraic manipulation of the j-th element in the leading term of (S.18) then leads to

$$\widehat{\Delta}_j^2(a) n\theta_j(a) = \left\{ \hat{\Psi}_j(a) - \frac{1}{\phi(a)} \sum_{l=1}^k \prod_{g\in E_l} \theta_g(a) \hat{\Psi}_l(a) \frac{\sqrt{\theta_l(a)}}{\sqrt{\theta_j(a)}} \right\}^2 + o_p(1) \tag{S.23}$$

for $j = 1,\ldots,k$, where $\phi(a) = \sum_{l=1}^k \prod_{g\in E_l} \theta_g(a)$. The leading term above can be expressed in terms of the weights $w_j(a)$ and the $\hat{\Psi}_j(a)$ processes as $w_j(a)\{\hat{\Psi}_j(a)/\sqrt{w_j(a)} - \check{\Psi}(a)\}^2$. This, (S.16) and (S.18) imply $-2\log\mathcal{R}_k(a)$ is asymptotically equivalent to the weighted sum of squares between blocks $\widehat{\mathrm{SSB}}(a)$.

We now discretize the above result to study the asymptotics of $-2\log f_n(\mathcal{R}_k)(\cdot)$, as mentioned in the beginning of this Supplement Section. Since $\sup_{a\in[\alpha_1,\alpha_2]} |f_n(-2\log\mathcal{R}_k - \widehat{\mathrm{SSB}})(a)| \leq \sup_{a\in[\alpha_1,\alpha_2]} |-2\log\mathcal{R}_k(a) - \widehat{\mathrm{SSB}}(a)|$ by definition of f_n, we have

$$-2\log f_n(\mathcal{R}_k)(a) = f_n(\widehat{\mathrm{SSB}})(a) + o_p(1).$$

On the other hand, the uniform convergence of $\hat{\Psi}_j(a)$ to $\Psi_j(a)$ as $n\to\infty$ can be obtained by the uniform convergence of $\bar{g}_j(\mu_j)(a)$ obtained in Supplement Section 9. Application of the continuous mapping theorem then leads to $\widehat{\mathrm{SSB}}(a) \xrightarrow{d} \mathrm{SSB}(a)$ in $\ell^\infty([\alpha_1,\alpha_2])$ as $n\to\infty$.

9 Asymptotic Orders of $\widehat{\Delta}_j(a)$

If $\widehat{\Delta}_j(a) = 0$ for all $j = 1,\ldots,k$, then they are trivially $O_p(n^{-1/2})$, and $\hat{\mu}_{j,EL}(a) - \mu_j(a) = \bar{g}_j(\mu_j)(a) = O_p(n^{-1/2})$, where the second equality holds by adding the subscript j to the $\sqrt{n}$-consistency of $\hat{\mu}(a)$ given in Supplement Section 11. On the complementary event, let $\eta_j(a)$ be such that $\widehat{\Delta}_j(a) = \eta_j(a)\max_{1\leq l\leq k}\left|\widehat{\Delta}_l(a)\right|$, $j = 1,\ldots,k$. Denote $\widehat{\Delta}_j(a) g_{ij}(\hat{\mu}_{j,EL})(a)$ by $\zeta_{ij}(a)$. Substituting $1/\{1+\zeta_{ij}(a)\} = 1 - \zeta_{ij}(a)/\{1+\zeta_{ij}(a)\}$ into

$\eta_j(a)\hat{\gamma}_j \sum_{i=1}^{n_j} \hat{p}_{ij}(a) g_{ij}(\hat{\mu}_{j,EL})(a) = 0$, we get

$$0 = \frac{\eta_j(a)}{n} \sum_{i=1}^{n_j} g_{ij}(\hat{\mu}_{j,EL})(a) - \frac{\eta_j(a)\widehat{\Delta}_j(a)}{n} \sum_{i=1}^{n_j} \frac{g_{ij}^2(\hat{\mu}_{j,EL})(a)}{1+\zeta_{ij}(a)} \tag{S.24}$$

$$= \frac{\eta_j(a)}{n} \sum_{i=1}^{n_j} g_{ij}(\hat{\mu}_{j,EL})(a) - \frac{\eta_j^2(a) \max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right|}{n} \sum_{i=1}^{n_j} \frac{g_{ij}^2(\hat{\mu}_{j,EL})(a)}{1+\zeta_{ij}(a)}. \tag{S.25}$$

Note that $1+\zeta_{ij}(a) > 0$ by $\hat{p}_{ij}(a) > 0$ for all i, j. Thus we can obtain the following (in)equalities:

$$\frac{\eta_j^2(a) \max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right|}{n} \sum_{i=1}^{n_j} g_{ij}^2(\hat{\mu}_{j,EL})(a) \le \frac{\eta_j^2(a) \max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right|}{n} \sum_{i=1}^{n_j} \frac{g_{ij}^2(\hat{\mu}_{j,EL})(a)}{1+\zeta_{ij}(a)} \times$$

$$\left\{1 + \max_{1\le i\le n_j} \zeta_{ij}(a)\right\} \le \frac{\eta_j(a)}{n} \sum_{i=1}^{n_j} g_{ij}(\hat{\mu}_{j,EL})(a) \left\{1 + \max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right| \max_{1\le i\le n_j} |g_{ij}(\hat{\mu}_{j,EL})(a)|\right\},$$

where the last equality follows from (S.25). This can be written as

$$\max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right| \eta_j^2(a)\hat{\theta}_j(\hat{\mu}_{j,EL})(a) \le \eta_j(a)\bar{g}_j(\hat{\mu}_{j,EL})(a) \left\{1 + \max_{1\le l\le k} \left|\widehat{\Delta}_l(a)\right| Z_j(\hat{\mu}_{j,EL})(a)\right\}, \tag{S.26}$$

where $Z_j(\hat{\mu}_{j,EL})(a) = \max_{1\le i\le n_j} |g_{ij}(\hat{\mu}_{j,EL})(a)|$. By adding the subscript j in the proof that leads to (S.6) and the $\sqrt{n}$-consistency of $\hat{\mu}(a)$ given in Supplement Section 11, we can show the uniform convergence of $\hat{\theta}_j(\mu_j)(a)$ and $\bar{g}_j(\mu_j)(a)$ as $n \to \infty$, leading to

$$\hat{\theta}_j(\hat{\mu}_{j,EL})(a) = \theta_j(a) + \left\{\bar{g}_j(\mu_j)(a) + \mu_j(a) - \hat{\mu}_{j,EL}(a)\right\}^2 / \hat{\gamma}_j + o_p(1) \tag{S.27}$$

and

$$\bar{g}_j(\hat{\mu}_{j,EL})(a) = \mu_j(a) - \hat{\mu}_{j,EL}(a) + O_p(n^{-1/2}). \tag{S.28}$$

In (S.27), the fact that $\left\{\bar{g}_j(\mu_j)(a) + \mu_j(a) - \hat{\mu}_{j,EL}(a)\right\}^2 \ge 0$ means $\hat{\theta}_j(\hat{\mu}_{j,EL})(a) \ge \inf_{a\in[\alpha_1,\alpha_2]} \theta_j(a) + o_p(1)$. As for $Z_j(\hat{\mu}_{j,EL})(a)$, first note that $\sup_{a\in[\alpha_1,\alpha_2]} Z_j(\hat{\mu}_{j,EL})(a) \le \sup_{a\in[\alpha_1,\alpha_2]} \max_{1\le i\le n_j} |g_{ij}(\mu_j)(a)| + \sup_{a\in[\alpha_1,\alpha_2]} |\mu_j(a)| / \hat{\gamma}_j + \sup_{a\in[\alpha_1,\alpha_2]} |\hat{\mu}_{j,EL}(a)| / \hat{\gamma}_j$ by triangle inquality. Second, $\sup_{a\in[\alpha_1,\alpha_2]} |\hat{\mu}_{j,EL}(a)| \le \sup_{a\in[\alpha_1,\alpha_2]} \max_{1\le i\le n_j} |T_{ij}(a)| \le \max_{1\le i\le n_j} \{V_{ij}(\alpha_2) + |T_{ij}(\alpha_1)|\}$, where $V_{ij}(a)$ is the total variation of $T_{ij}(\cdot)$ over $[\alpha_1, a]$. Using a similar reasoning as in the proof that leads to $Z(a) = o(\sqrt{n})$ a.s. in Supplement Section 4, we have that $\max_{1\le i\le n_j} |g_{ij}(\mu_j)(a)| = o(\sqrt{n})$ a.s., $\sup_{a\in[\alpha_1,\alpha_2]} |\mu_j(a)| < \infty$, and $|\hat{\mu}_{j,EL}(a)| = o(\sqrt{n})$ a.s. Thus, $Z_j(\hat{\mu}_{j,EL})(a) = o(\sqrt{n})$ a.s. From these results and

(S.26), we have

$$\frac{\max_{1\le l\le k}\left|\widehat{\Delta}_l(a)\right|\eta_j^2(a)\hat{\theta}_j(\hat{\mu}_{j,EL})(a)}{1+\max_{1\le l\le k}\left|\widehat{\Delta}_l(a)\right|\max_{1\le l\le k}Z_l(\hat{\mu}_{l,EL})(a)} \le \eta_j(a)\left\{\mu_j(a)-\hat{\mu}_{j,EL}(a)\right\}+O_p(n^{-1/2}), \tag{S.29}$$

$j=1,\ldots,k$. By adding up (S.29) for $j=1,\ldots,k$ and the fact that $\sum_{j=1}^k\widehat{\Delta}_j(a)=0$ and $\mu_j(a)-\hat{\mu}_{j,EL}(a)=\mu_{j+1}(a)-\hat{\mu}_{j+1,EL}(a)$ for $j=1,\ldots,k-1$, we have

$$\max_{1\le l\le k}\left|\widehat{\Delta}_l(a)\right|\sum_{j=1}^k\frac{\eta_j^2(a)\hat{\theta}_j(\hat{\mu}_{j,EL})(a)}{1+\max_{1\le l\le k}\left|\widehat{\Delta}_l(a)\right|\max_{1\le l\le k}Z_l(\hat{\mu}_{l,EL})(a)} \le O_p(n^{-1/2}). \tag{S.30}$$

This, $Z_j(\hat{\mu}_{j,EL})(a)=o(\sqrt{n})$ a.s., $\sum_{j=1}^k\eta_j^2(a)>0$, $\hat{\theta}_j(\hat{\mu}_{j,EL})(a)\ge\inf_{a\in[\alpha_1,\alpha_2]}\theta_j(a)+o_p(1)$ (shown above), and the assumption that $\inf_{a\in[\alpha_1,\alpha_2]}\sigma_j^2(a)>0$ imply $\max_{1\le l\le k}\left|\widehat{\Delta}_l(a)\right|=O_p(n^{-1/2})$. These imply that the l.h.s. of (S.29) is $O_p(n^{-1/2})$ for all j. By the fact that $\sum_{j=1}^k\widehat{\Delta}_j(a)=0$, for a fixed $\jmath$ there exists $\jmath'\neq\jmath$ such that $\widehat{\Delta}_\jmath(a)$ and $\widehat{\Delta}_{\jmath'}(a)$ have different signs. Without loss of generality suppose $\widehat{\Delta}_\jmath(a)>0,\widehat{\Delta}_{\jmath'}(a)<0$. Then (S.29) for the $\jmath$-th and $\jmath'$-th samples imply that both $\hat{\mu}_{\jmath,EL}(a)-\mu_\jmath(a)$ and $\mu_\jmath(a)-\hat{\mu}_{\jmath,EL}(a)$ are bounded above by $O_p(n^{-1/2})$ terms, using the fact that $\mu_j(a)-\hat{\mu}_{j,EL}(a)=\mu_{j+1}(a)-\hat{\mu}_{j+1,EL}(a)$ for $j=1,\ldots,k-1$. And thus $\hat{\mu}_{\jmath,EL}(a)-\mu_\jmath(a)=O_p(n^{-1/2})$.

10 Bootstrap in Section 2.4

A bootstrap sample is obtained by drawing n_j curves independently with replacement from the j-th sample $\{f_n(T_{1j})(a),\ldots,f_n(T_{n_jj})(a),a\in[\alpha_1,\alpha_2]\}$, for $j=1,\ldots,k$. Based on this bootstrap sample, compute a value of K_n^*, where recall from Sections 2.3 and 2.4 that $K_n^*=\sup_{a\in[\alpha_1,\alpha_2]}f_n(\widehat{\mathrm{SSB}}^*)(a)$,

$$\widehat{\mathrm{SSB}}^*(a)=\sum_{j=1}^k\hat{w}_j(a)\left\{\frac{\hat{\Psi}_j^*(a)}{\sqrt{\hat{w}_j(a)}}-\check{\Psi}^*(a)\right\}^2,$$

$\hat{w}_j(a)\propto\hat{\gamma}_j/\hat{S}_j^2(a)$ are normalized to sum to 1 across the groups, $\hat{\Psi}_j^*(a)=U_{nj}^*(a)/\hat{S}_j(a)$, $U_{nj}^*(a)=\sqrt{n_j}\{\hat{\mu}_j^*(a)-\hat{\mu}_j(a)\}$, $\hat{\mu}_j(a)=\sum_{i=1}^{n_j}T_{ij}(a)/n_j$, $\hat{\mu}_j^*(a)=\sum_{i=1}^{n_j}W_{nij}T_{ij}(a)/n_j$, W_{nij} is the number of times that $f_n(T_{ij})(a)$ is redrawn from $\{f_n(T_{1j})(a),\ldots,f_n(T_{n_jj})(a),a\in[\alpha_1,\alpha_2]\}$, $\hat{S}_j(a)=[\sum_{i=1}^{n_j}\{T_{ij}(a)-\hat{\mu}_j(a)\}^2/n_j]^{1/2}$ is the sample version of $\sigma_j(a)$, and $\check{\Psi}^*(a)=\sum_{j=1}^k\sqrt{\hat{w}_j(a)}\hat{\Psi}_j^*(a)$.

Repeat the procedure in the previous paragraph B times to obtain B bootstrapped values for K_n^*; our simulation study (see Section 3) uses $B=1000$. Let $c_{K,\alpha}^*$ denote the upper α-quantile of these B bootstrapped values of K_n^*. To calibrate the test, we compare $c_{K,\alpha}^*$ with our test statistic K_n based on the original data. We reject H_0

if $K_n > c^*_{K,\alpha}$. As for calibrating the Wald-type test $K_{n,Wald}$, due to the asymptotic equivalence of $K_{n,Wald}$ and K_n, again we reject H_0 if $K_{n,Wald} > c^*_{K,\alpha}$.

Similar to the point we made in Supplement Section 6.2, the use of the aforementioned bootstrap method is computationally more efficient than an alternative bootstrap procedure based on calculating the local EL ratio in each of the bootstrap sample. As for the theoretical performance, again both bootstrap procedures are asymptotically first-order equivalent.

11 Results based on Fully Observed Trajectories

Here we provide the results throughout the paper based on fully observed trajectories, for comparison purposes.

Theorem S.1. *Suppose the sample path of $T(a)$ is of bounded variation, and $ET^2(a)$ is bounded over $a \in [\alpha_1, \alpha_2]$. Then*

$$\sqrt{n}\,\{\hat{\mu}(a) - \mu(a)\} \xrightarrow{d} U(a)$$

in $\ell^\infty([\alpha_1, \alpha_2])$ as $n \to \infty$.

Remark. Comparing the conditions of Theorem 1 with those in Theorem S.1, we see that additional assumptions of right-continuity, finiteness in the number of discontinuities, and boundedness of $D^+(\mu, \beta)(a)$ and $D^-(\mu, \beta)(a)$ are needed to take the domain discretization into account. These assumptions are key for the discretized observations to approximate the full trajectories. Note that the right-continuity assumption can be replaced by left-continuity, with b_a in the definition of $f_n(g)(a)$ changing to the closest point on $\mathbf{G}_n$ to the left of a, and limits changing to left-hand ones instead in defining the β-Dini derivatives.

Proof. It suffices to show the class of evaluation functions $\mathcal{F}$ is P-Donsker, where $\mathcal{F}$ is defined in Supplement Section 3.1. We start with decomposing this class. Each $g \in \mathcal{B}$ satisfies $g = v(g) - d(g)$, where recall that $\mathcal{B}$, $v(g)$ and $d(g)$ are defined in Supplement Section 1.1. Further, for $e(g) = v(g)$ or $d(g)$, let $r(e(g)) = s(e(g))$ and $\ell(e(g)) = e(g) - s(e(g))$, where $s(e(g))(a)$ is the sum of left discontinuities of $e(g)$ up to $a \in [\alpha_1, \alpha_2]$. It can be shown that $r(e(g))(a)$ and $\ell(e(g))(a)$ is non-decreasing right- and left-continiuous in a, respectively. Let the classes $\mathcal{V}_r = \{r \circ v : \mathcal{B} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$, $\mathcal{V}_\ell = \{\ell \circ v : \mathcal{B} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$, $\mathcal{D}_r = \{r \circ d : \mathcal{B} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$, and $\mathcal{D}_\ell = \{\ell \circ d : \mathcal{B} \mapsto \mathbb{R}, a \in [\alpha_1, \alpha_2]\}$. Then we can see that $\mathcal{F} = \mathcal{V}_r + \mathcal{V}_\ell - \mathcal{D}_r - \mathcal{D}_\ell$.

Now we show that $\|P\|_{\mathcal{G}} < \infty$ for $\mathcal{G} = \mathcal{V}_r, \mathcal{V}_\ell, \mathcal{D}_r$ or $\mathcal{D}_\ell$ to enable an application of Donsker preservation in the next paragraph (see, e.g., van der Vaart and Wellner,

1996, Theorem 2.10.6), where $\|P\|_{\mathcal{G}} = \sup\{|Pg| : g \in \mathcal{G}\}$. $\|P\|_{\mathcal{V}_r}, \|P\|_{\mathcal{D}_r} < \infty$ because $E\{r(V)(a)\}$ and $E\{r(D)(a)\}$ are bounded by being cadlag on $[\alpha_1, \alpha_2]$. Next we have $\|P\|_{\mathcal{V}_\ell} < \infty$ because $0 \leq \ell(V)(a) \leq V(a)$ a.s. and $\sup_{a\in[\alpha_1,\alpha_2]} EV(a) \leq EV(\alpha_2) < \infty$. Then by triangle inequality we have $\|P\|_{D_\ell} < \infty$.

We can show that $\mathcal{V}_r$ and $\mathcal{D}_r$ are P-Donsker, by the proof for establishing bracketing entropy for right-continuous monotone stochastic processes (see, e.g., van der Vaart and Wellner, 1996, Example 2.11.16). Modifying the proof by including the right endpoint instead of the left in forming the brackets, it can be shown that $\mathcal{V}_\ell$ and $\mathcal{D}_\ell$ are P-Donsker. Therefore, by Donsker preservation (see, e.g., van der Vaart and Wellner, 1996, Theorem 2.10.6) we have the desired result.

□

Corollary S.1. *Under the conditions of Theorem S.1, U_n^* converges weakly to U as $n \to \infty$, given $T_1, T_2, \ldots$, in probability.*

Proof. To show bootstrap consistency of $U_n^*(a)$, we first write it as $\sum_{i=1}^n (W_{ni}-1)\{T_i(a)-\mu(a)\}/\sqrt{n}$. The bootstrap consistency of $U_n^*(a)$ (in $\ell^\infty([\alpha_1, \alpha_2])$) holds due to the bootstrap central limit theorem (see, e.g., Kosorok, 2008b, Theorem 2.6) and the fact that $\mathcal{F}$ is P-Donsker.

□

Theorem S.2. *Suppose the conditions of Theorem S.1 hold. In addition, suppose $\inf_{a\in[\alpha_1,\alpha_2]} \sigma^2(a) > 0$. Then we have $-2\log \mathcal{R}(\mu)(a) \xrightarrow{d} U^2(a)/\sigma^2(a)$ in $\ell[\alpha_1, \alpha_2]$ as $n \to \infty$, where the process $U(a)$ is defined in Section 2.2.*

Remark. This result is a stochastic process version of the univariate Wilks type theorem of Owen (2001), page 16.

Proof. We can use the same proof in the second paragraph through (S.10) in Supplement Section 4. Then by the $\sqrt{n}$-consistency of $\hat{\mu}(a)$ given in Theorem S.1 and the continuous mappting theorem, we have the desired result.

□

Corollary S.2. *Under the conditions of Theorem S.2, $\hat{\Psi}^{*2}(a)$ converges weakly to $U^2(a)/\sigma^2(a)$ in $\ell^\infty([\alpha_1, \alpha_2])$ as $n \to \infty$, given $T_1, T_2, \ldots$, in probability.*

Proof. To prove Corollary S.2, we use the bootstrap consistency of U_n^* shown in Corollary S.1 and the strong consistency of $\hat{S}$ shown in Supplement Section 6.1. Then by a routine extension of the proof for the conditional Slutsky's lemma in Cheng (2015) to the case of random elements of a metric space, we have $[U_n^*, \hat{S}]^T$ is bootstrap consistent

for $[U, \sigma]^T$ in $\{\ell^\infty([\alpha_1, \alpha_2])\}^2$. The desired result follows by the continuous mapping theorem for the bootstrap (see, e.g., Kosorok, 2008b, Theorem 10.8).

□

Theorem S.3. *Suppose the conditions of Theorem S.2 hold for each group indexed by $j = 1, \ldots, k$. Then, under H_0, the fully-observed version $K_{n,\text{full}}$ of K_n satisfies*

$$K_{n,\text{full}} \xrightarrow{d} \sup_{a \in [\alpha_1, \alpha_2]} \text{SSB}(a)$$

as $n \to \infty$.

Proof. We can use the same proof in the second paragraph through the second to the last paragraph in Supplement Section 8. Then by the uniform convergence of $\bar{g}_j(\mu_j)(a)$ obtained in Supplement Section 9 and the continuous mapping theorem, we have the desired result.

□

12 Derivation and illustration of $E\{L(a)\}$ in Section 3.1

It suffices to compute $L(a)$ based on $X(t)$ having $\pi(t) = t$. For clarity, we first consider $X(t)$ without domain discretization. This leads to $L(a) = \max\left\{1 - \max\left(\varepsilon c_a^4/c_1, U\right), 0\right\}$, where c_a is the closest point in $\{0\} \cup \mathbb{N}$ to the right of a and accounts for the floor function in the definition of $X(t)$, and $\varepsilon \sim \text{Log-normal}(0, \nu_L^2) + 10^{-6}$ here is defined in Section 3.1 and different from ε used in the previous sections in this Supplement. Denote the cdf and density of $\varepsilon c_a^4/c_1$ as Q_L and q_L, resepctively. Letting $c_1 = 10^{10}$, $c_2 = 10^{-6}$ and $A = \max\left(\varepsilon c_a^4/c_1, U\right)$, the cdf and density of A can be computed as

$$P\left(\varepsilon c_a^4/c_1 \le x\right) P\left(U \le x\right) = Q_L(x)\left(xI\left\{0 < x < 1\right\} + I\left\{x \ge 1\right\}\right)$$

and

$$q_L(x)\left(xI\left\{0 < x < 1\right\} + I\left\{x \ge 1\right\}\right) + Q_L(x) I\left\{0 < x < 1\right\},$$

respectively, where $Q_L(x) = \Phi\left[\left\{\ln h(x,a) - \log\left(c_a^4/c_1\right)\right\}/\nu_L\right] I\{h(x,a) > 0\}$, $q_L(x) = \exp\left[-\left\{\ln h(x,a) - \log\left(c_a^4/c_1\right)\right\}^2/(2\nu_L^2)\right]/\{h(x,a)\nu_L\sqrt{2\pi}\} I\{h(x,a) > 0\}$, and $h(x,a) = x - c_2 c_a^4/c_1$. Thus, we can write $E\{L(a)\}$ as $Q_L(1) - \int_0^1 A^2 q_L(A) dA - \int_0^1 A Q_L(A)\, dA$. Since

$$\int_0^1 A Q_L(A) dA = \int_0^1 A \int_0^A q_L(y) dy dA = \frac{1}{2} Q_L(1) - \int_0^1 \frac{y^2}{2} q_L(y) dy,$$

where the last equality follows from Fubini's theorem, we have

$$E\{L(a)\} = \frac{1}{2}Q_L(\tau) - \frac{1}{2}\int_0^1 A^2 q_L(A)dA.$$

Computing $E\{L(a)\}$ amounts to deriving a more explicit expression for the last term in the above display:

$$\begin{aligned}
&\int_0^1 A^2 q_L(A)dA = \int_{c_2c_a^4/c_1}^1 x^2 \frac{1}{h(x,a)\nu_L\sqrt{2\pi}} \exp\left\{-\left(\ln h(x,a) - \mu_a\right)^2/(2\nu_L^2)\right\} dx \\
&= \exp\left(2\mu_a + 2\nu_L^2\right)\Phi\left\{\frac{\log\left(1 - c_2c_a^4/c_1\right) - \mu_a}{\nu_L} - 2\nu_L\right\} + 2c_2c_a^4/c_1 \exp\left(\mu_a + \nu_L^2/2\right) \\
&\times\Phi\left\{\frac{\log\left(1 - c_2c_a^4/c_1\right) - \mu_a}{\nu_L} - \nu_L\right\} + \left(c_2c_a^4/c_1\right)^2 \Phi\left\{\frac{\log\left(1 - c_2c_a^4/c_1\right) - \mu_a}{\nu_L}\right\},
\end{aligned}$$

where $\mu_a = \log(c_a^4/c_1)$ and the last equality follows by a change of variables and completing the square.

Taking the more complicated scenario of domain discretization of $X(t)$ into account, recall from Section 3.1 that $\mathbf{H}_n$ is a finite grid of points $t \in [0,1)$ at which $X(t)$ is observed; by convention, we add ∞ as an extra point. Then we have $L(a) = \max\left\{1 - \max\left(c_a^H, U^H\right), 0\right\}$, where c_a^H and U^H is the smallest point in $\mathbf{H}_n$ that is no less than $\varepsilon c_a^4/c_1$ and U, respectively. Note that c_a^H and U^H are discrete random variables taking values in $\mathbf{H}_n$. Here we illustrate how to compute their distributions numerically. We use the fact that for each $t \in \mathbf{H}_n$, $c_a^H \geq t$ if and only if $\varepsilon c_a^4/c_1 > t_-$, where t_- is the largest element in $\mathbf{H}_n$ that is less than t and 0_- is defined as 0; by a similar reasoning, we have $U^H \geq t$ if and only if $U > t_-$. Therefore, the quantities $P\left(c_a^H \geq t\right) = P\left(\varepsilon c_a^4/c_1 > t_-\right) = 1 - Q_L(t_-)$ and $P\left(U^H \geq t\right) = P\left(U > t_-\right) = 1 - t_-$ can be computed for each $t \in \mathbf{H}_n$. Similarly, $P\left(c_a^H > t\right) = P\left(\varepsilon c_a^4/c_1 \geq t_+\right)$ and $P\left(U^H > t\right) = P\left(U \geq t_+\right)$ can be computed for each $t \in \mathbf{H}_n$, where t_+ is the smallest element in $\mathbf{H}_n$ that is greater than t and ∞_+ is defined as ∞. Subtracting from 1 the four quantities computed above, we get $P\left(c_a^H < t\right)$, $P\left(c_a^H \leq t\right)$, $P\left(U^H < t\right)$, and $P\left(U^H \leq t\right)$. Letting $A_H = \max\left(c_a^H, U^H\right)$, we get $P\left(A_H \leq t\right) = P\left(c_a^H \leq t\right) P\left(U^H \leq t\right)$ and similarly for the $< t$ version. This leads to $P\left(A_H = t\right) = P\left(A_H \leq t\right) - P\left(A_H < t\right)$, so that $E\{L(a)\}$ is readily obtained by

$$\sum_{t \in \mathbf{H}_n} \max\{1 - t, 0\} P\left(A_H = t\right).$$

The following Figure S.2 shows (part of) a simulated sample path of X along with $E\{L(a)\}$ for $\nu_L = 2$.

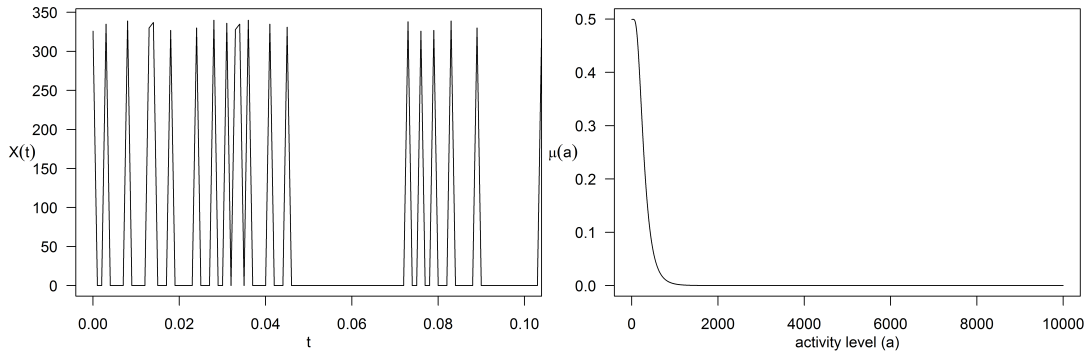

Figure S.2: A simulated sample path of $X(\cdot)$ (left) up to $t = 0.1$, along with the mean occupation time $E\{L(a)\}$ (right) under the second simulation example in Section 3.1 used to assess the performance of the confidence bands, $\nu_L = 2$.

13 Additional numerical details and results

13.1 Rationale for covariance in the first paragraph of Section 3.1

The non-smooth covariance function

$$\mathrm{Cov}\{J(a), J(b)\} = (0.6 + \nu_T) I\{a < 0.25, a = b\} + 0.6 I\{a \geq 0.25, a = b\} + 1.5 I\{a, b < 0.25, a \neq b\} + 0.5 I\{a \text{ or } b \geq 0.25, a \neq b\}$$

is the sum of an exchangeable part (with variances 0.6 and covariances 0.5) and a part having a jump at $a = 0.25$: $\nu_T I\{a < 0.25, a = b\} + I\{a, b < 0.25, a \neq b\}$ for some $\nu_T > 1$.

13.2 Simulating $U(a)$ directly using estimate of $\mathrm{Cov}\{T(a), T(b)\}$

In calibrating our EL procedures, instead of bootstrap, one can simulate an estimated verison of $U(a)$ based on repeated generations of Gaussian process over the grid $\mathbf{G}_n$ with the covariance function of the process $\mathrm{Cov}\{T(a), T(b)\}$ estimated by the sample covariance. Here we compare the performance of the EL simultaneous confidence band using this method with the one using bootstrap (the results of which have been shown in Section 3.1), based on the first simulation example with $\nu_T = 10$ in Section 3.1. For one dataset with $n = 100$ with 1000 bootstrap samples and 1000 replications of the Gaussian process, it takes about 45.78 and 47.52 seconds to construct the two EL bands, respectively, on a server with Intel Xeon CPU E5-2620 v4 @ 2.10 GHz 2.10 GHz and 384 GB RAM. The $U(a)$ simulation approach, therefore, has a slightly longer computation time than the bootstrap approach.

The empirical coverage rates, average widths and range-violation of the EL bands

Table S.1: Simulation study for 95% simultaneous confidence bands: empirical coverage (percentage), average width (rounded to four decimal places, in parenthesis) and range-violation rate (percentage, in square brackets); 1000 Monte Carlo replications, 1000 bootstrap samples and 1000 replications of Gaussian process, jump parameter $\nu_T = 10$, $n = 100, 200$.

	$\nu_T = 10$	
Calibration methods	$n = 100$	$n = 200$
Bootstrap	94.5 (0.4735) [0]	94.8 (0.3535) [0]
$U(a)$ simulation	94.5 (0.4746) [0]	94.9 (0.3532) [0]

are given in Table S.1. The empirical coverage rates of the two calibration methods are similar, both close to the nominal 95% level. They are both range-repecting. Compared with bootstrap, the band based on simulating the limiting $U(a)$ is wider in $n = 100$ and slightly shorter in $n = 200$.

13.3 Measurement error

We first consider a measurement error version of our general framework, and then specialize to the occupation time setting. The following measurement error model is just to illustrate that as long as the mean of the observed (measurement-error-contaminated) process remains the same as the mean of the measurement-error-free process, our methods continue to work. We utilize a multiplicative measurement error model to account for non-negative sample paths that appeared in our application to occupation time, in contrast to additive measurement errors typically employed in the functional data analysis literature (Wang et al., 2016). The observed process $\{T^o(a), a \in [r_1, r_2]\}$ is given by $T^o = f_n(T)f_n(\varepsilon)$, where $\varepsilon = \{\varepsilon(a)\}$ is a right-continuous error process of bounded variation, $E\{\varepsilon(a)|T\} = 1$ and $E\varepsilon^2(a) < \infty$ for all $a \in [\alpha_1, \alpha_2]$. When $\varepsilon \equiv 1$ a.s., the model reduces to the case without measurement error. More constraints can be added to ε according to the desired features for T^o, as shown in the next paragraph. The sample mean and its bootstrap based on observation of T^o has the same form of limit process as without measurement error (Theorem 1 and Corollary 1), except that the covariance function now becomes $\mathrm{Cov}\{(T\varepsilon)(a), (T\varepsilon)(b)\}$. The same result will continue to hold for other (perhaps more sophisticated) measurement error models, as long as the mean of the observed process $E(T^o)$ is the same as the target functional mean $E(T)$. For the above reasons, the results in Section 2 of the main text continue to hold for their measurement-error-contaminated counterparts.

Now we consider a simulation example in the occupation time setting which explicitly

specifies measurement error. We generated the measurement-error-free occupation time $L(a)$ the same way as the second simulation example in Section 3.1. The observed process $\{L^o(a), a \in [r_1, r_2]\}$ is given by $L^o = f_n(L)f_n(\varepsilon)$, where $\varepsilon = \{\varepsilon(a)\}$ is a right-continuous error process of bounded variation, $E\{\varepsilon(a)|L\} = 1$ a.s. and $E\varepsilon^2(a) < \infty$ for all $a \in [r_1, r_2]$. Note that $\sup_{a\in[\alpha_1,\alpha_2]} L(a) = L(0) < \tau$ a.s. because $L(0) = \tau$ represents the impossible case that a subject always has activity. When $\varepsilon = 1$ a.s., the model reduces to the case without measurement error. As mentioned in the previous paragraph, more constraints are needed according to the desired features for L^o: to make sure $L^o \in [0, \tau]$, for $a \in \mathbf{G}_n$, we need $0 \le \varepsilon(a) \le \tau/L(a)$ when $0 < L(a) < \tau$, and to ensure that L^o is non-increasing, for $a, b \in \mathbf{G}_n$, we need $\varepsilon(b)/\varepsilon(a) \le L(a)/L(b)$ for $a < b$. To satisfy these conditions, conditionally on L, for each $a \in \mathbf{G}_n$ we generate independent $\varepsilon(a)$ from Uniform$(C_1/Y_1, Y_2/C_2)$, where

$$\begin{aligned} Y_1 = &\min_{(a,b):a,b\in\mathbf{G}_n, a<b, L(a)>L(b), L(b)\neq 0} \sqrt{L(a)/L(b)} \\ &+\delta_L I[\{(a,b) : a, b \in \mathbf{G}_n, a < b, L(a) > L(b), L(b) \neq 0\} = \emptyset], \end{aligned}$$

$\delta_L > 0$, $Y_2 = \min(Y_1, Y_3)$, $Y_3 = \min_{a\in\mathbf{G}_n}\{\tau/L(a)\}$, $C_1 = (1-c)Y_1 + cC_3$ for some $c \in [0, 1]$, $C_2 = (Y_1Y_2)/(2Y_1 - C_1)$, $C_3 = \max\{2Y_1 - Y_1Y_2, 1\}$, and the minimum in the definition of Y_1 is set to 1 when there are no $a, b \in \mathbf{G}_n$ such that $a < b, L(a) > L(b), L(b) \neq 0$. The magnitude of measurement error is controlled by c and δ_L, with larger values indicating a larger measurement error; $c, \delta_L = 0$ corresponds to no measurement error. In the first dataset simulated from $\nu_L = 1.5$, $\delta_L = 7$ and $c = 1$ results in $\sup_{a\in[\alpha_1,\alpha_2]} \text{Var}\{\varepsilon(a)|L\} = 0.019$, which is 20% of the maximal variance of $L(a)$, whereas $\delta_L = 7$ and $c = 0.5$ results in $\sup_{a\in[\alpha_1,\alpha_2]} \text{Var}\{\varepsilon(a)|L\} = 0.005$, which is 5% of that maximal variance.

Empirical coverage rates of $E\{L(a)\}$, average widths, range-violation and monotonicity preservation of the various bands are given in Table S.2. We see similar results as those produced by the measurement-error-free counterparts in Table 1 (also provided on the right-side of Table S.2). This suggests our method outperforms the competing methods under varying degrees of measurement error.

13.4 Controlling the functional means and variances used in Section 3.2

In this subsection, we utilize the subscript j in the notation to denote the j-th group. Let Σ_j be the beta random variable Beta(ι_j, ω_j) independent of $L_j(a)$, $j = 1, \ldots, k$. Then the functional mean and variance for the j-th group is

$$E\{L_j(a)\Sigma_j\} = E[\Sigma_j E\{L_j(a)|\Sigma_j\}] = E\{L_j(a)\}\frac{\iota_j/\omega_j}{1+\iota_j/\omega_j} \tag{S.31}$$

Table S.2: Simulation study assessing the effect of measurement error contamination on the 95% simultaneous confidence bands: empirical coverage (percentage), average width (in parenthesis), range-violation rate (percentage, in square brackets) and average number of confidence band boundaries that satisfy monotonicity (rounded to two decimal places, in curly brackets); 1000 Monte Carlo replications, 1000 bootstrap samples, $n = 100$, log-normal scale parameter $\nu_L = 1.5, 2$, measurement error parameters $\delta_L = 7$, $c = 0.5$ and 1.

	Error-contaminated				Error-free	
tests	$\nu_L = 1.5$		$\nu_L = 2$		$\nu_L = 1.5$	$\nu_L = 2$
	$c = 0.5$	$c = 1$	$c = 0.5$	$c = 1$		
EL	94.1 (0.10) [0] {2}	94.0 (0.10) [0] {2}	94.3 (0.10) [0] {2}	94.4 (0.10) [0] {2}	93.9 (0.10) [0] {2}	94.3 (0.10) [0] {2}
EP	91.4 (0.09) [0] {2}	91.2 (0.09) [0] {2}	92.1 (0.09) [0] {2}	92.2 (0.09) [0] {2}	91.7 (0.09) [0] {2}	92.2 (0.09) [0] {2}
NS	93.8 (0.14) [100] {2}	93.6 (0.14) [100] {2}	93.7 (0.14) [100] {2}	93.7 (0.14) [100] {2}	93.9 (0.14) [100] {2}	93.6 (0.14) [100] {2}
MFD	90.7 (0.09) [0.4] {0.19}	90.7 (0.09) [0.4] {0.19}	90.9 (0.09) [2] {1.25}	90.9 (0.09) [2] {1.25}	90.6 (0.09) [0.4] {0.19}	90.8 (0.09) [2.2] {1.24}
MFDbs	98.5 (0.09) [0] {2}	98.6 (0.09) [0] {2}	99.0 (0.09) [0.1] {2}	99.0 (0.09) [0.1] {2}	98.6 (0.09) [0] {2}	99.0 (0.09) [0.1] {2}
Cao1	89.0 (0.09) [0] {2}	89.0 (0.09) [0] {2}	88.5 (0.09) [0] {2}	88.4 (0.09) [0] {2}	89.1 (0.09) [0] {2}	88.4 (0.09) [0] {2}
Cao2	90.1 (0.09) [0] {2}	89.8 (0.09) [0] {2}	89.3 (0.09) [0] {2}	89.4 (0.09) [0.1] {2}	89.7 (0.09) [0] {2}	89.0 (0.09) [0] {2}
Geo	75.0 (0.16) [100] {0}	74.9 (0.16) [100] {0}	74.2 (0.16) [100] {0}	74.0 (0.16) [100] {0}	74.9 (0.16) [100] {0}	74.4 (0.16) [100] {0}

and

$$\operatorname{Var}\{L_j(a)\Sigma_j\} = \operatorname{Var}\{L_j(a)\}\left\{\frac{\iota_j}{\omega_j^2\left(\iota_j/\omega_j+1\right)^2\left(\iota_j/\omega_j+1+1/\omega_j\right)}+\left(\frac{\iota_j/\omega_j}{1+\iota_j/\omega_j}\right)^2\right\}$$
$$+\left[E\{L_j(a)\}\right]^2\frac{\iota_j}{\omega_j^2\left(\iota_j/\omega_j+1\right)^2\left(\iota_j/\omega_j+1+1/\omega_j\right)}, \tag{S.32}$$

respectively, where the equality in (S.32) is due to the law of total variance. From these equations, we can see that both the functional means and variances contain moments of $L_j(a)$ that depend on parameters of the underlying Ornstein–Uhlenbeck processes, where the dependence has no closed form expressions. However, we can control $E\{L_j(a)\Sigma_j\}$ and $\operatorname{Var}\{L_j(a)\Sigma_j\}$ through different parameters, using ι_j/ω_j (see (S.31)) and (ι_j, ω_j) (see (S.32)), respectively. For example, to keep $E\{L_j(a)\Sigma_j\}$ the same across the groups, we need to ensure the parameters of the underlying Ornstein–Uhlenbeck processes and ι_j/ω_j are the same for all $j = 1, \ldots, k$. While doing that, by (S.32), we can still change either ι_j or ω_j so that $\operatorname{Var}\{L_j(a)\Sigma_j\}$ are different for different j.

13.5 Raw activity curves for accelerometery data

A graphical comparison on the basis of sample means for the (unregistered) raw activity data is provided in Figure S.3. For clarity, different days of data are displayed in separate panels, for four consecutive days. A comparison of the sample means (groups 1:3) suggests higher mean raw activity in the younger group of veterans, which is consistent with the results given in Table 3. However, the data are extremely noisy, so functional smoothing and curve-registration techniques would be needed to provide an effective bias-variance tradeoff and to validate a functional ANOVA testing approach in terms of a direct analysis of the raw activity data. In contrast, the corresponding mean occupation time curves in Figure S.4 are smooth apart from barely visible jump discontinuities, which is consistent with what we observe for individual occupation time curves in Section 2.1. Further, these curves are automatically aligned on the grid of activity levels. These features of occupation time motivate our non-smoothing approach without curve alignment, as explained in the Introduction and Section 2.1.

14 Missing Sensor Readings

In computing the occupation time $a \mapsto L(a)$, if X is not observed for part of the study period $[0, \tau]$ (i.e., there are missing sensor readings), it is necessary to rescale the corresponding occupation time to compensate. Here we propose a simple imputation method (essentially mean imputation) of doing this.

In the functional data analysis literature, there is very little discussion of how to handle missing data. In the context of sensor data, Song et al. (2019) pointed out

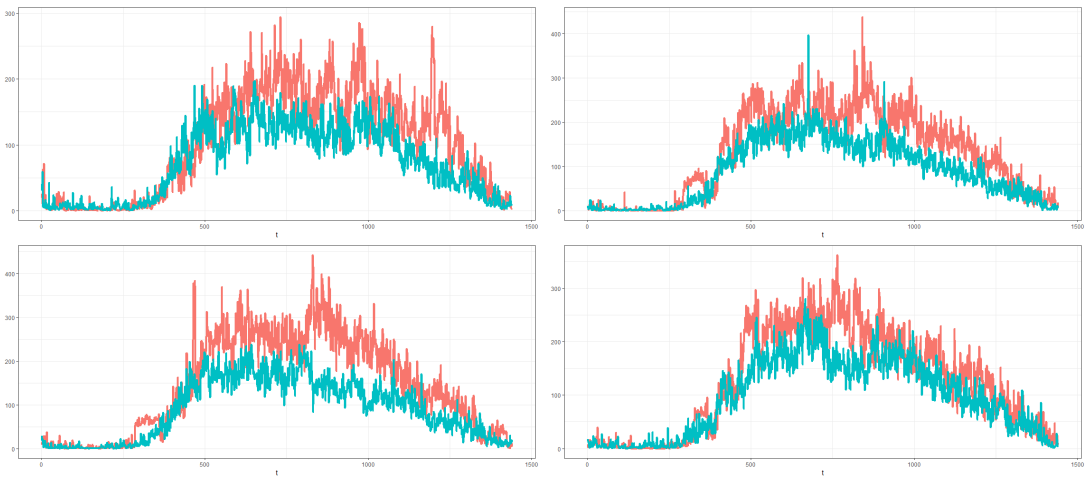

Figure S.3: Comparison of veterans aged 75-and-older (group 1, acqua) and veterans aged 65–74 (group 3, coral) in terms of sample means of the raw activity curves over four consecutive days. Time unit = 1 minute on each horizontal axis, as in Figure 1.

the potential for bias caused by the presence of associations between the device wear pattern (causing missing sensor readings) and the underlying physical activity process. To address this problem, they use a semiparametric mean regression model for panel count data (defined as cumulative minutes of moderate to vigorous physical activity), in which the observation scheme and the event process are allowed to be dependent through covariates and an unobserved (multiplicative) frailty term. However, their approach still faces the curve registration problem mentioned earlier, because the baseline mean function in their model (Wang et al., 2013) is the same for all subjects and the frailty term is non-time-dependent. Yang and Wang (2021) treated device nonwear as "window censoring" and utilized techniques from recurrent event data to handle the missing data problem. However, their theory only dealt with one time point and did not consider the effect of discretization as we do.

In analyzing the NHANES data, we used the following simple rescaling of the observed occupation time to compensate for the missing sensor readings in a few subjects. Let the observed occupation time above level a be denoted O_a, and let Υ denote the total time during $[0, \tau]$ that $X(t)$ is *not* observed. The idea is to impute $L(a)$ by the rescaled version $O_a\tau/(\tau - \Upsilon)$ of O_a. This rescales the observed occupation time by taking into account that it only refers to the proportion $1 - \Upsilon/\tau$ of the full follow-up interval $[0, \tau]$. More sophisticated imputation approaches might be considered (e.g., estimating the mean and covariance function of X, and using a Gaussian process simulation to fill in the missing data), but in our numerical experiments we have found that the results using the suggested imputation method are remarkably insensitive to mild levels of missingness.

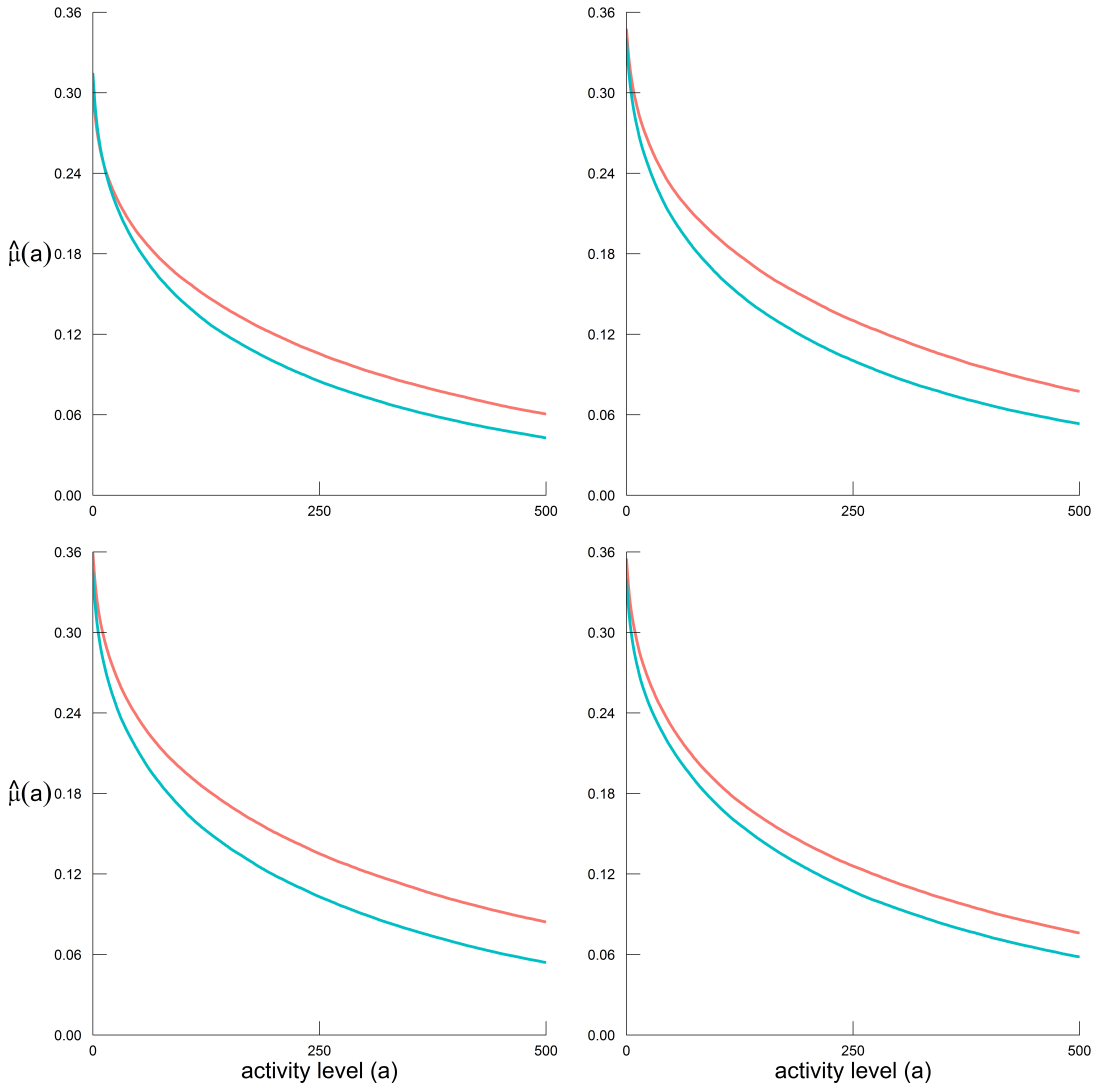

Figure S.4: Comparison of veterans aged 75-and-older (group 1, acqua) and veterans aged 65–74 (group 3, coral) in terms of sample mean occupation time (as percentage of one day) up to $a = 499$ over four consecutive days.

15 References

Cao, G., Yang, L. and Todem, D. (2012) Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, **24**, 359–377.

Carothers, N. L. (2000) *Real Analysis*. Cambridge Series on Statistical and Probabilistic Mathematics. New York: Cambridge University Press.

Chang, H.-w. and McKeague, I. W. (2019) Nonparametric testing for multiple survival

functions with non-inferiority margins. *Annals of Statistics*, **47**, 205–232.

Cheng, G. (2015) Moment consistency of the exchangeably weighted bootstrap for semiparametric m-estimation. *Scandinavian Journal of Statistics*, **42**, 665–684.

Choi, H. and Reimherr, M. (2018) A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**, 239–260.

Degras, D. A. (2011) Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, **21**, 1735–1765.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Kosorok, M. R. (2008a) Errata for *Introduction to Empirical Processes and Semiparametric Inference* (Kosorok, 2008, Springer). `http://www.bios.unc.edu/ kosorok/errata.pdf`.

— (2008b) *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

McShane, E. J. (1944) *Integration*. Princeton University Press.

Nair, V. N. (1984) Confidence bands for survival functions with censored data: a comparative study. *Technometrics*, **26**, 265–275.

Owen, A. B. (2001) *Empirical Likelihood*. Boca Raton: Chapman & Hall/CRC.

Shao, J. (2003) *Mathematical Statistics*. Springer-Verlag New York Inc, 2nd edn.

Song, J., Swartz, M. D., Gabriel, K. P. and Basen-Engquist, K. (2019) A semiparametric model for wearable sensor-based physical activity monitoring data with informative device wear. *Biostatistics*, **20**, 287–298.

Uno, H., Tian, L., Claggett, B. and Wei, L. J. (2015) A versatile test for equality of two survival functions based on weighted differences of Kaplan–Meier curves. *Statistics in Medicine*, **34**, 3680–3695.

Usmani, R. A. (1994) Inversion of a tridiagonal Jacobi matrix. *Linear Algebra and its Applications*, **212–213**, 413–414.

van der Vaart, A. W. (2000) *Asymptotic Statistics*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016) Functional data analysis. *Annual Review of Statistics and Its Application*, **3**, 257–295.

Wang, X., Ma, S. and Yan, J. (2013) Augmented estimating equations for semiparametric panel count regression with informative observation times and censoring time. *Statistica Sinica*, **23**, 359–381.

Yang, Y. and Wang, M.-C. (2021) Analyzing wearable device data using marked point processes. *Biometrics*, **77**, 54–66.

Zhang, J.-T., Cheng, M.-Y., Wu, H.-T. and Zhou, B. (2019) A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis*, **132**, 3–17. Special Issue on Biostatistics.

Zhang, J.-T. and Liang, X. (2014) One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, **41**, 51–71.