Biometrika (2018), **xx**, x, *pp*. 1–6
Printed in Great Britain

Supplementary Material to "Wasserstein Covariance for Multiple Random Densities"

BY ALEXANDER PETERSEN

Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106, U.S.A.
petersen@pstat.ucsb.edu

AND HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, California 95616, U.S.A.
hgmueller@ucdavis.edu

SUMMARY

Section S.1 demonstrates a valid density estimator with the properties required by Assumption 1 in the paper. Section S.2 contains proofs of Theorems 1 and 2, while Section S.3 contains an additional theoretical result. Section S.4 provides additional visualizations for the connectivity and mortality analyses.

S.1. MODIFIED KERNEL DENSITY ESTIMATOR

A density estimator $\hat{f}$ that satisfies Assumption 1 is defined as follows. Assume one has an i.i.d. sample $W_1, \ldots, W_N \sim f$. Let K be a bounded pdf on $[-1, 1]$, symmetric about 0. We use an adapted version of the modified kernel density estimator (Petersen and Müller 2016)

$$\hat{f}(u) = \sum_{r=1}^{m} K\left(\frac{W_r - u}{h}\right) w(u,h) \Bigg/ \left\{ \sum_{r=1}^{m} \int_{D_f} K\left(\frac{W_r - v}{h}\right) w(v,h) \mathrm{d}v \right\}, \qquad (1)$$

for $u \in D_f = [a_f, b_f]$ and 0 elsewhere. The weight function

$$w(u,h) = \begin{cases} \left\{ \int_{(a_f - u)h^{-1}}^{1} K(v)\mathrm{d}v \right\}^{-1}, & a_f \le u \le a_f + h, \\ \left\{ \int_{-1}^{(b_f - u)h^{-1}} K(v)\mathrm{d}v \right\}^{-1}, & b_f - h \le u \le b_f, \\ 1, & \text{otherwise}, \end{cases}$$

is designed to remove the boundary bias of the standard kernel density estimator, which is a major issue due to the compact support of he densities, while simultaneously maintaining the integrated squared error rate and rendering an estimate that is nonnegative and integrates to one. This leads to the following proposition.

PROPOSITION 1. *With $\mathcal{D}$ as in Assumption 1, assume further that*

$$\sup_{f \in \mathcal{D}} \sup_{u \in D_f} \max\left\{ f(u), 1/f(u), |f'(u)| \right\} < M < \infty.$$

© 2016 Biometrika Trust

For generic $f \in \mathcal{D}$, given an i.i.d. sample $W_r \sim f$, $(r = 1, \ldots, N)$ and assuming D_f is known, then $\hat{f}$ given in (1) *satisfies Assumption 1 with $b_N = N^{-1/3}$, using a bandwidth $h = tN^{-1/3}$ for any $t > 0$.*

Proof. Proposition 1 of Petersen and Müller (2016) immediately implies that

$$\sup_{f \in \mathcal{D}} E\left\{d_2(\hat{f}, f)\right\} = O\left(N^{-1/3}\right)$$

under the given assumptions, where d_2 is the $L^2(I)$ metric and I is the interval in Assumption 1.

Choose $C > 0$ such that $I \subset [-C/2, C/2]$ and let F and $\hat{F}$ be the cdfs of f and $\hat{f}$. It is clear that $d_2(F, \hat{F}) \leq C d_2(f, \hat{f})$, and that $(F^{-1})'$ is uniformly bounded by M over $f \in \mathcal{D}$. By the mean value theorem, and using the change of variables $u = \hat{F}^{-1}(t)$,

$$\begin{aligned} d_W(f, \hat{f})^2 &= \int_0^1 \left\{F^{-1}(t) - \hat{F}^{-1}(t)\right\}^2 \mathrm{d}t \\ &\leq M^2 \left[\int_{D_f} \left\{F(u) - \hat{F}(u)\right\}^2 f(u) \mathrm{d}u \right. \\ &\qquad \left. + \int_{D_f} \left\{F(u) - \hat{F}(u)\right\}^2 \left\{\hat{f}(u) - f(u)\right\} \mathrm{d}u\right] \\ &\leq M^2 C(MC + 1) d_2(f, \hat{f})^2 \end{aligned}$$

by Cauchy-Schwarz, and the result follows. □

S.2. PROOFS OF THEOREMS 1 AND 2

Throughout the proofs, we use the notations $\langle f, g \rangle = \int_0^1 f(u)g(u)\mathrm{d}u$ for the standard inner product on $L^2[0,1]$, $\|f\| = (\langle f, f \rangle)^{1/2}$ for the corresponding L^2 norm and $\|\cdot\|_{[0,1]^2}$ for the standard $L^2([0,1]^2)$ norm. Let $X_i(t) = \{X_{i1}(t), \ldots, X_{ip}(t)\}^\top$ be the vector of quantile functions for subject i, and set $\nu(t) = E\{X_1(t)\}$, $\tilde{\nu}(t) = n^{-1}\sum_{i=1}^n X_i(t)$, and $X_i^c(t) = X_i(t) - \nu(t)$. Letting $U_i(s,t) = X_i^c(s)X_i^c(t)^\top$ and $T_n(s,t) = \{\tilde{\nu}(s) - \nu(s)\}\{\tilde{\nu}(t) - \nu(t)\}^\top$, set

$$\tilde{\mathcal{C}}_{jk}(s,t) = \frac{1}{n}\sum_{i=1}^n \{U_i(s,t)\}_{jk} - \{T_n(s,t)\}_{jk}, \quad 0 \leq s, t \leq 1.$$

Proof of Theorem 1.

Define

$$\tilde{\Sigma}_\oplus = \int_0^1 \left\{\frac{1}{n}\sum_{i=1}^n U_i(t,t) - T_n(t,t)\right\} \mathrm{d}t$$

and, for $0 \leq s, t \leq 1$, define the four dimensional array $D(s,t)$ with elements

$$\{D(s,t)\}_{jklm} = \mathrm{cov}\left[\{U_i(s,s)\}_{jk}, \{U_i(t,t)\}_{lm}\right].$$

Since $E\left[\{U_i(t,t)\}_{jk}\right] = \mathcal{C}_{jk}(t,t)$, if $R_i = \int_0^1 U_i(t,t)\mathrm{d}t$, then $E(R_i) = \Sigma_\oplus$. With $\tau_n = \int_0^1 T_n(t,t)\mathrm{d}t$,

$$\tilde{\Sigma}_\oplus - \Sigma_\oplus = \left(\frac{1}{n}\sum_{i=1}^n R_i\right) - \Sigma_\oplus - \tau_n. \tag{2}$$

Let $\mathcal{S}_{jklm} = \mathrm{cov}\{(R_i)_{jk}, (R_i)_{lm}\} = \int_0^1\int_0^1 \{D(s,t)\}_{jklm}\mathrm{d}s\,\mathrm{d}t$, so that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n R_i - \Sigma_\oplus\right) \to \mathfrak{C} \quad \text{weakly},$$

where $\mathfrak{C}$ is a zero-mean $p \times p$ Gaussian matrix with covariance structure $\mathcal{S}$. Next, with $\|\cdot\|_F$ denoting the Frobenius norm, it is easy to show that $\|\tau_n\|_F \le \sum_{j=1}^p \|\hat{\nu}_j - \nu_j\|^2$. Since each summand is $O_p(n^{-1})$, this proves that $\tau_n = o_p(n^{-1/2})$. Hence,

$$\sqrt{n}(\tilde{\Sigma}_\oplus - \Sigma_\oplus) \to \mathfrak{C} \quad \text{weakly}.$$

Next, due to Assumption 1, the functions X_{ij} and $\hat{X}_{ij}$ are uniformly bounded for all i and j, and this bound clearly carries over for the quantile averages $\tilde{\nu}_j = n^{-1}\sum_{i=1}^n X_{ij}$ and $\hat{\nu}_j = n^{-1}\sum_{i=1}^n \hat{X}_{ij}$. Hence, for some positive constant C,

$$\begin{aligned}
\left|\left(\hat{\Sigma}_\oplus\right)_{jk} - \left(\tilde{\Sigma}_\oplus\right)_{jk}\right| &= \left|\frac{1}{n}\sum_{i=1}^n \left(\langle \hat{X}_{ij} - \hat{\nu}_j, \hat{X}_{ik} - \hat{\nu}_k\rangle - \langle X_{ij} - \tilde{\nu}_j, X_{ik} - \tilde{\nu}_k\rangle\right)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^n \left(\langle \hat{X}_{ij} - \hat{\nu}_j, \hat{X}_{ik} - X_{ik}\rangle + \langle \hat{X}_{ij} - X_{ij}, X_{ik} - \tilde{\nu}_k\rangle\right)\right| \\
&\le \frac{1}{n}\left(\sum_{i=1}^n \left|\langle \hat{X}_{ij} - \hat{\nu}_j, \hat{X}_{ik} - X_{ik}\rangle\right| + \sum_{i=1}^n \left|\langle \hat{X}_{ij} - X_{ij}, X_{ik} - \tilde{\nu}_k\rangle\right|\right) \\
&\le \frac{C}{n}\left(\sum_{i=1}^n \|\hat{X}_{ik} - X_{ik}\| + \sum_{i=1}^n \|\hat{X}_{ij} - X_{ij}\|\right) \\
&= \frac{1}{n}\left\{\sum_{i=1}^n d_W(\hat{f}_{ik}, f_{ik}) + \sum_{i=1}^n d_W(\hat{f}_{ij}, f_{ij})\right\}.
\end{aligned}$$

The result then follows from Assumptions 1 and 2.

Proof of Theorem 2.

Adopting the same notations as in the proof of Theorem 1, let $R_{ijk} = (R_i)_{jk}$ and

$$\tilde{\Sigma}_\oplus(y_j, y_k) = n^{-1}\sum_{i=1}^n R_{ijk} - \tau_{njk},$$

where $\tau_{njk} = \int_0^1 T_{njk}(t,t)\mathrm{d}t$ can be shown to satisfy $p^{-2}\sum_{j,k=1}^p \tau_{njk}^2 = O_p(n^{-2})$. Using Assumptions 3 and 4, we immediately obtain

$$\left\|\Sigma_\oplus - \tilde{\Sigma}_\oplus\right\|^2_{[0,1]^2} = O\left[n^{-1} + \frac{1}{p^2}\sum_{j,k=1}^p \left\{\Sigma_\oplus(y_j,y_k) - \tilde{\Sigma}_\oplus(y_j,y_k)\right\}^2\right]$$
$$= O\left[n^{-1} + \frac{1}{p^2}\sum_{j,k=1}^p \left\{\Sigma_\oplus(y_j,y_k) - n^{-1}\sum_{i=1}^n R_{ijk}\right\}^2\right].$$

Observing $E(R_{ijk}) = \Sigma_\oplus(y_j,y_k)$ and $\mathrm{var}(R_{ijk}) = G(y_j,y_k)$, where

$$G(y,z) = \int_{[0,1]^2} \mathrm{cov}\left\{X_1^c(s;y)X_1^c(s;z), X_1^c(t;y)X_1^c(t;z)\right\}\mathrm{d}s\mathrm{d}t,$$

Assumption 5 implies

$$E\left[\frac{1}{p^2}\sum_{j,k=1}^p \left\{\Sigma_\oplus(y_j,y_k) - n^{-1}\sum_{i=1}^n R_{ijk}\right\}^2\right] = \frac{1}{np^2}\sum_{j,k=1}^p G(y_j,y_k) = O(n^{-1}),$$

so that $\left\|\Sigma_\oplus - \tilde{\Sigma}_\oplus\right\|^2_{[0,1]^2} = O_p(n^{-1})$, as claimed.

Finally,

$$\max_{j,k=1,\ldots,p}\left\{\tilde{\Sigma}_\oplus(y_j,y_k) - \hat{\Sigma}_\oplus(y_j,y_k)\right\} = O_p(b_N)$$

by Assumptions 1 and 2, so that

$$\left\|\tilde{\Sigma}_\oplus - \hat{\Sigma}_\oplus\right\|^2_{[0,1]^2} = O_p(b_N^2),$$

completing the proof.

S.3. THEOREM 3

THEOREM 3. *Under Assumptions 1 and 2, the covariance function estimates satisfy*

$$\int_0^1\int_0^1 \left\{\hat{\mathcal{C}}_{jk}(s,t) - \mathcal{C}_{jk}(s,t)\right\}^2 \mathrm{d}s\,\mathrm{d}t = O_p(n^{-1} + b_N^2).$$

Proof of Theorem 3.

Write $U_{ijk}(s,t) = \{U_i(s,t)\}_{jk}$. Since $E\{U_{ijk}(s,t)\} = \mathcal{C}_{jk}(s,t)$, Assumption 1 implies $E\left(\|X_{1j}\|^4\right) < \infty$, whence

$$E\left(\left\|\mathcal{C}_{jk} - \frac{1}{n}\sum_{i=1}^n U_{ijk}\right\|^2_{[0,1]^2}\right) = O(n^{-1}).$$

Furthemore, with $T_{njk}(s,t) = \{T_n(s,t)\}_{jk}$,

$$\left\|T_{njk}\right\|^2_{[0,1]^2} = \|\tilde{\nu}_j - \nu_j\|^2\|\tilde{\nu}_k - \nu_k\|^2 = O_p(n^{-2}),$$

so that $\|\mathcal{C}_{jk} - \tilde{\mathcal{C}}_{jk}\|^2_{[0,1]^2} = O_p(n^{-1})$. Lastly, under Assumption 1, there exists a constant C such that

$$\left\|\tilde{\mathcal{C}}_{jk} - \hat{\mathcal{C}}_{jk}\right\|^2_{[0,1]^2} \leq \frac{2C}{n}\left\{\sum_{i=1}^{n} d_W(f_{ij}, \hat{f}_{ij})^2 + \sum_{i=1}^{n} d_W(f_{ik}, \hat{f}_{ik})^2\right\} = O_p(b_N^2)$$

by applying Assumption 2, which proves the result.

S.4. ADDITIONAL FIGURES

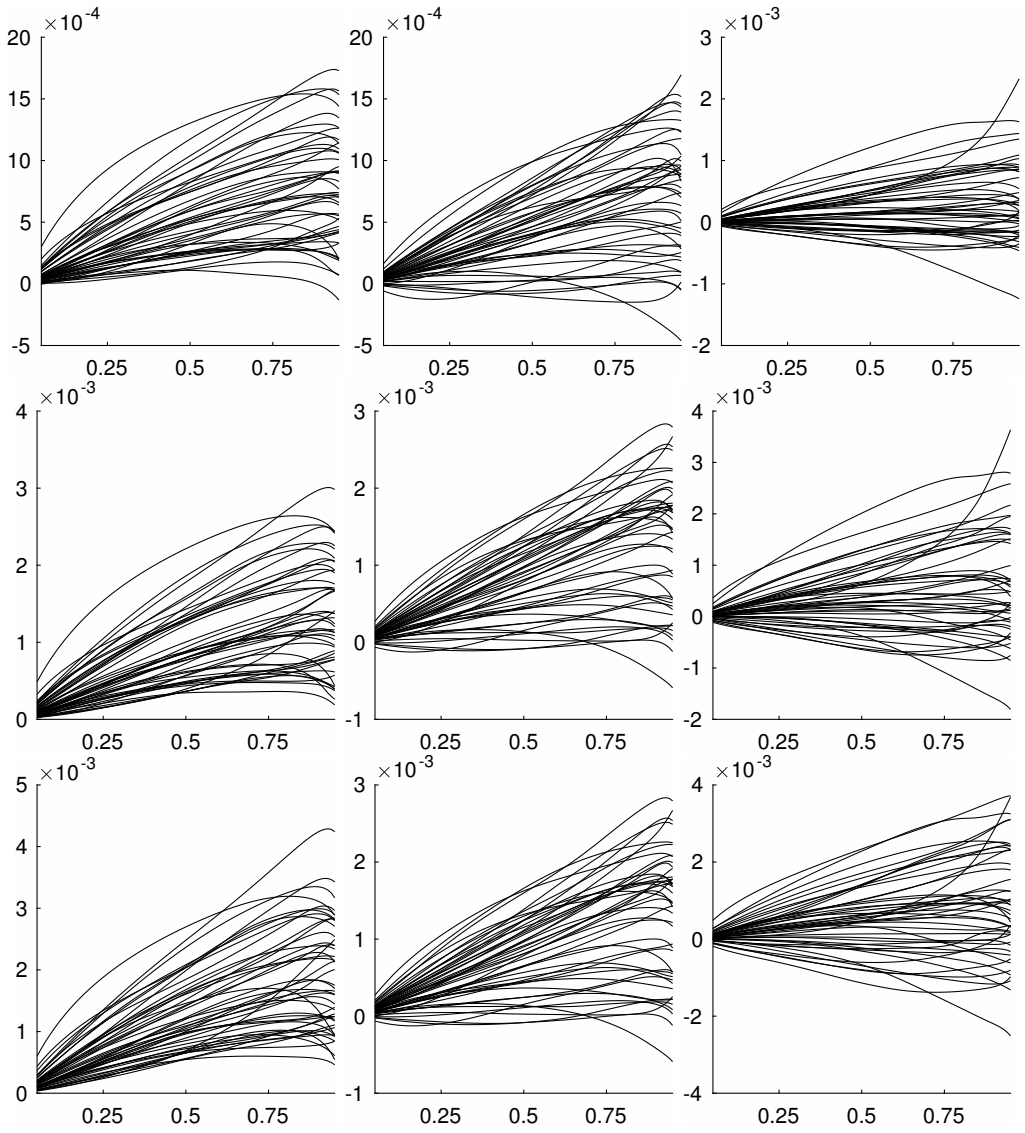


Fig. 5. Slices of covariance surface estimates $\hat{\mathcal{C}}_{jk}(s^*, t)$, $1 \leq j < k \leq 10$, for $s^* = 0.25, 0.5, 0.75$, corresponding to top, middle, and bottom rows, respectively. The left, middle, and right columns correspond to normal, mild cognitive impairment, and Alzheimer's subjects.

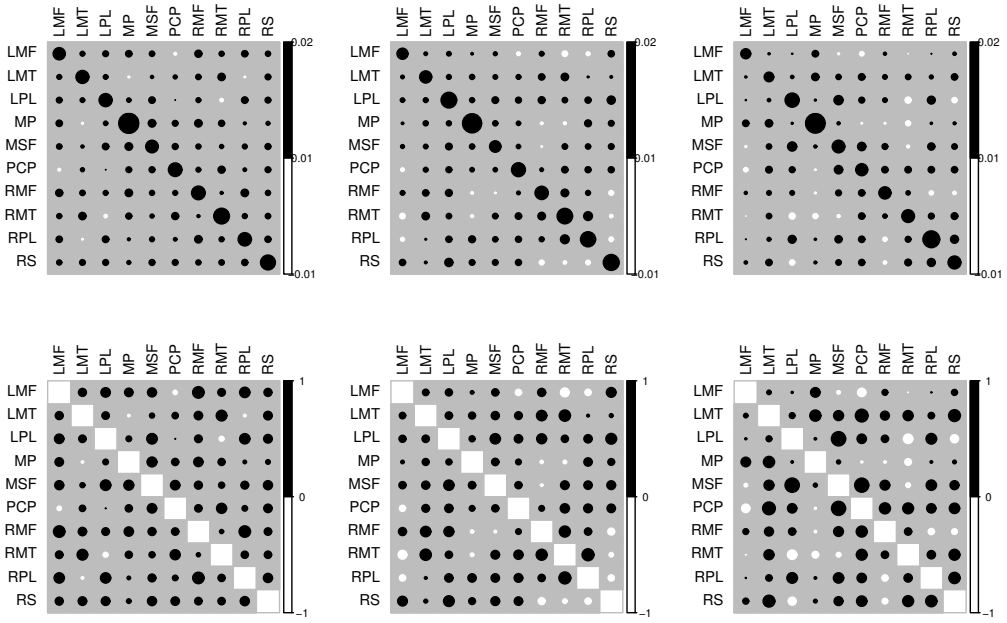


Fig. 6. Estimated covariance (top row) and correlation (bottom row) matrices of regional homogeneity scores for normal (left), mild cognitive impairment (middle) and Alzheimer's (right) subjects. Labels: LMF and RMF (left and right middle frontal), LPL and RPL (left and right parietal), LMT and RMT (left and right middle temporal), MSF (medial superior frontal), MP (medial prefrontal), PCP (posterior cingulate/precuneus) and RS (right supramarginal). Positive (negative) values are drawn in black (white) and larger circles correspond to larger absolute values.

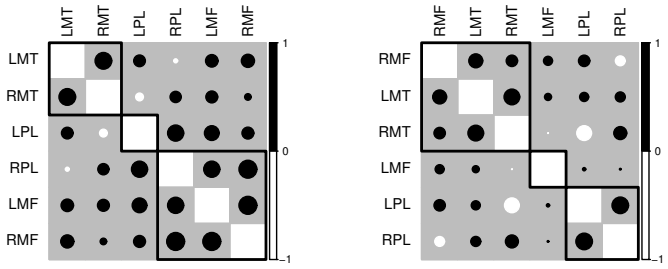


Fig. 7. Estimated regional homogeneity correlation submatrices corresponding to lateral hub pairs for normal (left) and Alzheimer's (right) subjects, after reordering of hubs by hierarchical clustering using Ward's criterion. Rectangles indicate the groupings when three clusters are used. Labels correspond to those in Figure 6 and further explanations can be found in the caption of Figure 6.

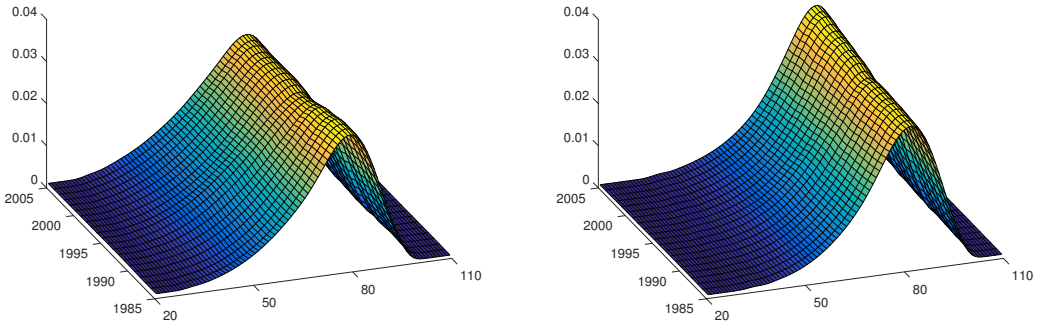


Fig. 8. Wasserstein mean density surfaces for mortality between 1980 and 2005, for Eastern European (left) and other (right) countries.