Bivariate Thin-plate Splines Models for Nonparametric Covariance Estimation with Longitudinal Data

Tayler A. Blake* Yoonkyung Lee†

January 9, 2018

The theoretical foundations of the thin-plate spline was laid in the seminal work of **?**. For a bivariate function $f\left(x_1, x_1\right)$, the usual thin-plate spline functional ($d = m = 2$) is given by

$$J_2\left(f\right) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(f_{x_1x_1}^2 + f_{x_1x_2}^2 + f_{x_2x_2}^2\right)dx_1dx_2 \tag{1}$$

and in general,

For $d = 2$, define the inner product of functions f and g as follows:

$$\langle f, g\rangle = \sum_{\alpha_1+\alpha_2=m}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial^m f}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}}\right)\left(\frac{\partial^m g}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}}\right)dx_1dx_2. \tag{2}$$

We suppose that $f \in \mathcal{X}$, the space of functions with partial derivatives of total order m belong to $\mathcal{L}_2\left(E^2\right)$. We endow $\mathcal{X}$ with seminorm $J_m^2\left(f\right)$; for such $\mathcal{X}$ to be a reproducing kernel Hilbert space, i.e. for the evaluation functionals to be bounded in $\mathcal{X}$, if it necessary and sufficient that $2m > d$. For $d = 2$, we require $m > 1$.

The data model for a random vector $y_i = \left(y_{i1}, \ldots, y_{i,M_i}\right)'$ is given by

$$y_{ij} = \sum_{k<j}\phi^*\left(v_{ijk}\right)y_{ik} + \sigma\left(v_{ijk}\right)e_{ij} \tag{3}$$

where $v_{ijk} = \left(t_{ij} - t_{ik}, \frac{1}{2}\left(t_{ij} + t_{ik}\right)\right) = \left(l_{ijk}, m_{ijk}\right)$. We assume that $\phi^* \in \mathcal{X}$ and $e_{ij} \overset{\text{i.i.d.}}{\sim} N\left(0,1\right)$. If we have a random sample of observed vectors $y_1, \ldots, y_N$ available for estimating ϕ^*, then we take ϕ^* to be the minimizer of

$$Q_\lambda\left(\phi^*\right) = \sum_{i=1}^{N}\sum_{j=2}^{n_i}\sigma_{ij}^{-2}\left(y_{ij} - \sum_{k<j}\phi^*\left(v_{ijk}\right)y_{ik}\right)^2 + \lambda J_m^2\left(\phi^*\right) \tag{4}$$

*The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

†The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

where $\sigma_{ij}^2 = \sigma^2(t_{ij})$. The null space of the penalty functional $J_m^2(\phi^*)$, denoted $\mathcal{H}_0$, corresponds to the $d_0 = \binom{2+m-1}{2}$-dimensional space spanned by the polynomials in two variables of total degree $< m$. For example, for $d = m = 2$, we have that $d_0 = 3$, and the null space of J_2^2 is spanned by η_1, η_2, and η_3 where

$$\eta_1(v) = 1, \quad \eta_2(v) = l, \quad \eta_2(v) = m.$$

In general, we let $\eta_1, \ldots, \eta_{d_0}$ denote the d_0 monomials of total degree less than m.

? showed that if the $\{v_{ijk}\}$ are such that the least squares regression of $\{y_{ijk}\}$ on $\eta_1, \ldots, \eta_{d_0}$ is unique, then there exists a unique minimizer of **??**, ϕ_λ^*, which has the form

$$\phi_\lambda^*(v) = \sum_{\nu=0}^{d_0} d_\nu \eta_\nu(v) + \sum_{v_i \in \mathcal{V}} c_i E_m(v, v_i) \tag{5}$$

where $\mathcal{V}$ denotes the set of unique within-subject pairs of observed $\{v_{ijk}\}$. E_m is a Green's function of the m-iterated Laplacian. Let

$$E_m(\tau) = \begin{cases} \theta_{m,d}|\tau|^{2m-d}\log|\tau| & 2m-d \text{ even} \\ \theta_{m,d}|\tau|^{2m-d} & 2m-d \text{ odd} \end{cases} \tag{6}$$

$$\theta_{md} = \begin{cases} \frac{(-1)^{\frac{d}{2}+1+m}}{2^{2m-1}\pi^{\frac{d}{2}}(m-1)!\left(m-\frac{d}{2}\right)!} & 2m-d \text{ even} \\ \frac{\Gamma\left(\frac{d}{2}-m\right)}{2^{2m}\pi^{\frac{d}{2}}(m-1)!} & 2m-d \text{ odd} \end{cases} \tag{7}$$

Defining $|v - v_i| = \left[(l-l_i)^2 + (m-m_i)^2\right]^{1/2}$, then we can write

$$E_m(v, \tilde{v}) = E_m(|v - \tilde{v}|)$$

Formally, we have that

$$\Delta^m E_m(\cdot, v_i) = \delta_{v_i},$$

so

$$\Delta^m \phi_\lambda^*(v) = 0 \text{ for } v \neq v_i, \ \ i = 1, \ldots, n$$

where $n = |\mathcal{V}|$.

The kernel E_m is not positive definite, but rather *conditionally positive definite*....

Stack the N observed response vectors $y_1, \ldots, y_N$ less their first element y_{i1} into a single vector Y of dimension $n_y = \left(\sum_i M_i\right) - N$. Let B denote the $n \times d_0$ matrix with i-ν^{th} element $\eta_\nu(v_i)$, which we assume has full column rank; let K denote the $n \times n$ kernel matrix with i-j^{th} element

$E_m(v_i, v_j)$, and let D denote the $n_y \times n_y$ diagonal matrix of innovation variances σ^2_{ijk}. The ϕ^* minimizing **??** corresponds to the coefficient vectors c, d minimizing

$$\begin{aligned} Q_\lambda(c,d) &= -\ell(Y|c,d) + \lambda J^2_m(\phi^*) \\ &= (Y - W(Bd + Kc))' D^{-1} (Y - W(Bd + Kc)) + \lambda c' K c \end{aligned} \tag{8}$$

where W is the matrix of autoregressive covariates constructed so that **??** and **??** are equivalent.

Differentiating Q_λ with respect to c and d and setting equal to zero, we have that

$$\begin{aligned} \frac{\partial Q_\lambda}{\partial c} &= KW'D^{-1}[W(Bd + Kc) - Y] + \lambda K c = 0 \\ &\iff W'D^{-1}W\left[Bd + Kc\right] + \lambda c = W'D^{-1}Y \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\partial Q_\lambda}{\partial d} &= B'W'D^{-1}[W(Bd + Kc) - Y] = 0 \\ &\iff -\lambda B' c = 0 \end{aligned}$$

So, the coefficients satisfy the normal equations

$$Y = W\left[Bd + \left(K + \lambda\left(W'D^{-1}W\right)^{-1}\right)c\right] \tag{10}$$

$$B'c = 0 \tag{11}$$

Let

$$\begin{aligned} \tilde{K} &= \left(W'D^{-1}W\right) K \left(W'D^{-1}W\right) \\ \tilde{c} &= \left(W'D^{-1}W\right)^{-1} c \\ \tilde{B} &= \left(W'D^{-1}W\right) B \\ \tilde{d} &= d \\ \tilde{Y} &= W'D^{-1}Y \end{aligned}$$

then, the system defined by **??** and **??** may be written

$$\tilde{Y} = \tilde{B}\tilde{d} + \left(\tilde{K} + \lambda\left(W'D^{-1}W\right)\right)\tilde{c} \tag{12}$$

$$\tilde{B}'\tilde{c} = 0 \tag{13}$$

Using the QR decomposition of $\tilde{B}$, we may write

$$\tilde{B} = \tilde{Q}\tilde{R} = \begin{bmatrix}\tilde{Q}_1 & \tilde{Q}_2\end{bmatrix}\begin{bmatrix}\tilde{R} \\ 0\end{bmatrix} = \tilde{Q}_1\tilde{R}$$

where $\tilde{Q}$ is an orthogonal matrix; $\tilde{Q}_1$ has dimension $n \times d_0$, and $\tilde{Q}_2$ has dimension $n \times (n - d_0)$. Since $\tilde{B}'\tilde{c} = 0$, $\tilde{c}$ must belong to the subspace spanned by the columns of $\tilde{Q}_2$, so

$$\tilde{c} = \tilde{Q}_2\gamma$$

for some $\gamma \in \mathrm{R}^{n-d_0}$. Premultiplying **??** by $\tilde{Q}_2'$, it follows that

$$\tilde{c} = \tilde{Q}_2\left[\tilde{Q}_2'\left(\tilde{K} + \lambda\left(W'D^{-1}W\right)\right)\tilde{Q}_2\right]^{-1}\tilde{Q}_2'\tilde{Y} \tag{14}$$

Using $\tilde{B} = \tilde{Q}_1\tilde{R}$, we can write

$$\tilde{d} = \tilde{R}^{-1}\tilde{Q}_1'\left[\tilde{Y} - \left(\tilde{K} + \lambda\left(W'D^{-1}W\right)\right)\tilde{c}\right] \tag{15}$$

1 Estimating the smoothing parameter

1.1 Cross Validation

Let $\phi^{*[kl]}$ be the minimizer of

$$\sum_{\substack{i,j \\ (i,j)\neq(k,l)}} \sigma_{ij}^{-2}\left(y_{ij} - \sum_{j'<j}\phi^*\left(v_{ijj'}\right)y_{ij'}\right)^2 + \lambda J_m^2\left(\phi^*\right). \tag{16}$$

The *ordinary cross validation function* $V_0\left(\lambda\right)$ is given by

$$\sum_{i=1}^{N}\sum_{j=2}^{n_i}\sigma_{ij}^{-2}\left(y_{ij} - \hat{y}_{ij}^{[ij]}\right)^2 \tag{17}$$

where $\hat{y}_{ij}^{[ij]} = \sum_{k<j}\phi^{[ij]*}\left(v_{ijk}\right)y_{ik}$. The value of λ minimizing $V_0\left(\lambda\right)$ is the OCV estimate.

Indexing the y_{ij} using a single integer $k = 1, \ldots, n_y$, when the innovation variances are known, it can be shown that $V_0(\lambda)$ can be written

$$V_0(\lambda) = \sum_{k=1}^{n_y} \left(\sigma_k^{-1}(y_k - \hat{y}_k)\right)^2 / \left(1 - a_{kk}(\lambda)\right)^2 \tag{18}$$

where $\{a_{kk}(\lambda)\}$ are the diagonal elements of the smoothing matrix $A(\lambda)$ which satisfies

$$\hat{Y} = A(\lambda) Y.$$

The *generalized cross validation function* $V(\lambda)$ is obtained by replacing a_{kk} by

$$\bar{a}(\lambda) = n_y^{-1} \sum_{j=1}^{n_y} a_{jj}(\lambda) = n_y^{-1} \mathrm{tr} A(\lambda).$$

The GCV function is defined as

$$\begin{aligned} V(\lambda) &= \sum_{k=1}^{n_y} \left(\sigma_k^{-1}(y_k - \hat{y}_k)\right)^2 / \left(1 - \bar{a}(\lambda)\right)^2 \\ &= \frac{||D^{-1/2}(I - A(\lambda))||^2}{\left[\mathrm{tr}(I - A(\lambda))\right]^2} \end{aligned} \tag{19}$$

Since $\tilde{A}(\lambda)\tilde{Y} = \tilde{K}\tilde{c} + \tilde{B}\tilde{d}$, from **??**, we can derive a simple expression for $I - \tilde{A}(\lambda)$:

$$\begin{aligned} \left(I - \tilde{A}(\lambda)\right) Y &= \lambda \left(W' D^{-1} W\right) \tilde{c} \\ &= \lambda \left(W' D^{-1} W\right) \tilde{Q}_2 \left[\tilde{Q}_2' \left(\tilde{K} + \lambda \left(W' D^{-1} W\right)\right) \tilde{Q}_2\right]^{-1} \tilde{Q}_2' \tilde{Y}, \end{aligned} \tag{20}$$

so

$$I - \tilde{A}(\lambda) = \lambda \left(W' D^{-1} W\right) \tilde{Q}_2 \left[\tilde{Q}_2' \left(\tilde{K} + \lambda \left(W' D^{-1} W\right)\right) \tilde{Q}_2\right]^{-1} \tilde{Q}_2'.$$

Then the GCV criterion can be written

$$V(\lambda) = \frac{n_y^{-1} \tilde{Y}' \tilde{Q}_2' \left[\tilde{Q}_2' \left(\tilde{K} + \lambda M\right) \tilde{Q}_2\right]^{-1} \tilde{Q}_2' M^2 \tilde{Q}_2 \left[\tilde{Q}_2' \left(\tilde{K} + \lambda M\right) \tilde{Q}_2\right]^{-1} \tilde{Q}_2' \tilde{Y}}{\left[n_y^{-1} \mathrm{tr} M \tilde{Q}_2 \left[\tilde{Q}_2' \left(\tilde{K} + \lambda M\right) \tilde{Q}_2\right]^{-1} \tilde{Q}_2'\right]^2} \tag{21}$$

where $M = W' D^{-1} W$.

1.2 Unbiased Risk Estimate

$$M(\lambda) = \frac{\left(D^{-1/2}Y\right)'(I - A(\lambda))\left(D^{-1/2}Y\right)}{\left[\det^{+}(I - A(\lambda))\right]^{1/(n-d_0)}}$$

where $\det^{+}(\cdot)$ denotes the product of the non-zero eigenvalues.

1.3 Generalized Maximum Likelihood

$$U(\lambda) = n_y^{-1}||(I - A(\lambda))D^{-1/2}Y||^2 + 2\text{tr}A(\lambda)$$