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

Tayler A. Blake* Yoonkyung Lee†

November 7, 2017

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

$$J_2(f) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \left(f^2_{x_1x_1} + f^2_{x_1x_2} + f^2_{x_2x_2}\right) dx_1 dx_2 \tag{1}$$

and in general,

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

$$\langle f, g\rangle = \sum_{\alpha_1+\alpha_2=m} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}}\right)\left(\frac{\partial^m g}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}}\right) dx_1 dx_2. \tag{2}$$

We suppose that $f \in \mathcal{X}$, the space of functions with partial derivatives of total order m belong to $\mathcal{L}_2(E^2)$. We endow $\mathcal{X}$ with seminorm $J^2_m(f)$; 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 $y_i = (y_{i1}, \ldots, y_{i,M_i})'$ is given by

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

where $v_{ijk} = \left(t_{ij} - t_{ik}, \frac{1}{2}(t_{ij} + t_{ik})\right) = (l_{ijk}, m_{ijk})$. We assume that $\phi^* \in \mathcal{X}$ and $e_{ij} \overset{\text{i.i.d.}}{\sim} N(0, 1)$. 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

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

*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 4, 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

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

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 S denote the $n \times d_0$ matrix with i-ν^{th} element $\eta_\nu(v_i)$, which we assume has full column rank; let Q 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 4 corresponds to the coefficient vectors c, d minimizing

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

where W is the matrix of autoregressive covariates constructed so that 4 and 8 are equivalent. Differentiating Q_λ with respect to c and d and setting equal to zero, we have that

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

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

So, the coefficients satisfy the normal equations

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

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

Let

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

then, the system defined by 10 and 11 may be written

$$\tilde{Y} = \tilde{S}\tilde{d} + \left(\tilde{Q} + \lambda\left(W'D^{-1}W\right)\right)\tilde{c} \tag{12}$$
$$\tilde{S}'\tilde{c} = 0 \tag{13}$$

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

$$\tilde{S} = FR = \begin{bmatrix} F_1 & F_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = F_1 R$$

where F is an orthogonal matrix; F_1 has dimension $n \times d_0$, and F_2 has dimension $n \times (n - d_0)$. Since $\tilde{S}'\tilde{c} = 0$, $\tilde{c}$ must belong to the subspace spanned by the columns of F_2, so

$$\tilde{c} = F_2\gamma$$

for some $\gamma \in \mathrm{R}^{n-d_0}$. Letting $M = W'D^{-1}W$ premultiplying 12 by F_2', it follows that

$$\tilde{c} = F_2\left[F_2'\left(\tilde{Q} + \lambda M\right)F_2\right]^{-1} F_2'\tilde{Y} \tag{14}$$

Using $\tilde{S} = F_1 R$, we can write

$$\tilde{d} = R^{-1}F_1'\left[\tilde{Y} - \left(\tilde{Q} + \lambda M\right)\tilde{c}\right] \tag{15}$$

1 Estimating the smoothing parameter

1.1 Cross Validation

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

$$\sum_{\substack{i,j \\ (i,j)\neq(k,l)}} \sigma_{ij}^{-2}\left(\tilde{y}_{ij} - \sum_{j'<j}\phi^*\left(v_{ijj'}\right)\tilde{y}_{ij'}\right)^2 + \lambda\tilde{J}_m^2\left(\phi^*\right), \tag{16}$$

where $\tilde{J}_m$ is the penalty term reparameterized according to the transformation defining $\tilde{c}$:

$$\tilde{J}_m^2\left(\phi^*\right) = \tilde{c}'Q\tilde{c}. \tag{17}$$

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

$$\sum_{i=1}^{N}\sum_{j=2}^{n_i}\tilde{\sigma}_{ij}^{-2}\left(\tilde{y}_{ij} - \hat{\tilde{y}}_{[ij]}\right)^2. \tag{18}$$

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

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

$$V_0\left(\lambda\right) = \sum_{k=1}^{n_y} \left(\tilde{\sigma}_k^{-1}\left(\tilde{y}_k - \hat{\tilde{y}}_k\right)\right)^2 / \left(1 - \tilde{a}_{kk}\left(\lambda\right)\right)^2 \tag{19}$$

where $\left\{\tilde{a}_{kk}\left(\lambda\right)\right\}$ are the diagonal elements of the smoothing matrix $\tilde{A}\left(\lambda\right)$ which satisfies

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

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

$$\bar{a}\left(\lambda\right) = n^{-1}\sum_{j=1}^{n}\tilde{a}_{jj}\left(\lambda\right) = n^{-1}\text{tr}\tilde{A}\left(\lambda\right).$$

The GCV function is defined

$$\begin{aligned} V\left(\lambda\right) &= \sum_{k=1}^{n} \left(\tilde{\sigma}_k^{-1}\left(\tilde{y}_k - \hat{\tilde{y}}_k\right)\right)^2 / \left(1 - \bar{a}\left(\lambda\right)\right)^2 \\ &= \frac{||\tilde{D}^{-1/2}\left(I - \tilde{A}\left(\lambda\right)\right)||^2}{\left[\text{tr}\left(I - \tilde{A}\left(\lambda\right)\right)\right]^2}, \end{aligned} \tag{20}$$

where $\tilde{D} =$ is the diagonal matrix with k^{th} diagonal element $\tilde{\sigma}_k^2$:

$$\begin{aligned} \tilde{D} = Cov\left(\tilde{e}\right) &= Cov\left(\tilde{Y} - \tilde{S}\tilde{d} - \tilde{Q}\tilde{c}\right) \\ &= Cov\left(W'D^{-1}e\right) \\ &= W'D^{-1}W \end{aligned} \tag{21}$$

From 12, $\tilde{A}\left(\lambda\right)\tilde{Y} = \tilde{Q}\tilde{c} + \tilde{S}\tilde{d}$, we can derive a simple expression for $I - \tilde{A}\left(\lambda\right)$:

$$\begin{aligned} \left(I - \tilde{A}\left(\lambda\right)\right)\tilde{Y} &= \lambda\left(W'D^{-1}W\right)\tilde{c} \\ &= \lambda M F_2\left[F_2'\left(\tilde{Q} + \lambda M\right)F_2\right]^{-1} F_2'\tilde{Y}, \end{aligned} \tag{22}$$

so that

$$I - \tilde{A}\left(\lambda\right) = \lambda M F_2\left[F_2'\left(\tilde{Q} + \lambda M\right)F_2\right]^{-1} F_2'.$$

1.2 Unbiased Risk Estimate

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

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

1.3 Generalized Maximum Likelihood

See pg. 68 of SS Anova Models.

$$M\left(\lambda\right)=n_y^{-1}||\left(I-A\left(\lambda\right)\right)D^{-1/2}Y||^2+2\mathrm{tr}A\left(\lambda\right)$$

2 Computation

The minimization of XXXXX lies within a space $\mathcal{H}\subseteq\{\phi^*:J\left(\phi^*\right)<\infty\}$ in which $J\left(\phi^*\right)$ is a square (semi) norm, or a subspace therein. The evaluation functional $\left[\boldsymbol{v}\right]\phi^*$, which appears in the first term in XXXXX, is assumed to be continuous in $\mathcal{H}$. A space in which the evaluation functional is continuous is called a reproducing kernel Hilbert space (RKHS) endowed with reproducing kernel (RK) $Q\left(\cdot,\cdot\right)$, a non-negative definite function satisfying

$$\langle Q\left(\boldsymbol{v},\cdot\right),phi^*\left(\cdot\right)\rangle$$

$\forall\phi^*\in\mathcal{H}$, where $\langle\cdot,\cdot\rangle$ is an inner product in $\mathcal{H}$. The norm and RK determine each other uniquely.

Let $\mathcal{N}_J=\{\phi^*:\ J\left(\phi^*\right)=0\}$ denote the null space of J, and consider the tensor sum decomposition

$$\mathcal{H}=\mathcal{N}_J\oplus\mathcal{H}_J.$$

The space $\mathcal{H}_J$ is a RKHS having $J\left(\phi^*\right)$ as the squared norm. The minimizer of XXXX has form

$$\phi^*\left(\boldsymbol{v}\right)=\sum_{\nu=1}^{d_0}d_\nu\eta\left(\boldsymbol{v}\right)+\sum_{i=1}^{n}c_iQ\left(\boldsymbol{v}_i,\boldsymbol{v}\right), \tag{23}$$

where $\{\eta_\nu\}$ is a basis for $\mathcal{N}_J$, and Q_J is the RK in $\mathcal{H}_J$.

For $\boldsymbol{v}\in\mathcal{X}$ where $\mathcal{X}$ is a product domain, ANOVA decompositions can be characterized by

$$\mathcal{H}=\bigoplus_{\beta=0}^{g}\mathcal{H}_\beta$$

and

$$J\left(\phi^*\right)=\sum_{\beta=0}^{g}\theta_\beta^{-1}J_\beta\left(\phi_\beta^*\right),$$

where $\phi^*_\beta \in \mathcal{H}_\beta$, J_β is the square norm in $\mathcal{H}_\beta$, and $0 < \theta_\beta < \infty$. This gives

$$\begin{aligned}
\mathcal{H}_0 &= \mathcal{N}_J \\
\mathcal{H}_J &= \bigoplus_{\beta=1}^{g} \mathcal{H}_\beta, \text{ and} \\
Q &= \sum_{\beta=1}^{g} \theta_\beta Q_\beta,
\end{aligned}$$

where Q_β is the RK in $\mathcal{H}_\beta$. The $\{\theta_\beta\}$ are additional smoothing parameters, which may or may not appear explicitly in notation to follow. The penalized likelihood is given by

$$\ell_\lambda(c, d) = \Big[Y - W(Sd + Qc)\Big]' D^{-1} \Big[Y - W(Sd + Qc)\Big] + \lambda c' Q c. \tag{24}$$

Letting $\tilde{Y} = D^{-1/2} Y$, $\tilde{S} = D^{-1/2} W S$, and $\tilde{Q} = D^{-1/2} W Q$, this may be written

$$\ell_\lambda(c, d) = \Big[\tilde{Y} - \tilde{S}d - \tilde{Q}c\Big]' \Big[\tilde{Y} - \tilde{S}d - \tilde{Q}c\Big] + \lambda c' Q c. \tag{25}$$

Taking partial derivatives with respect to d and c and setting equal to zero yields normal equations

$$\begin{aligned}
\tilde{S}'\tilde{S}d + \tilde{S}'\tilde{Q}c &= \tilde{S}'\tilde{Y} \\
\tilde{Q}'\tilde{S}d + \tilde{Q}'\tilde{Q}c + \lambda Q c &= \tilde{Q}'\tilde{Y},
\end{aligned} \tag{26}$$

which is equivalent to solving

$$\begin{bmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{Q} \\ \tilde{Q}'\tilde{S} & \tilde{Q}'\tilde{Q} + \lambda Q \end{bmatrix} \begin{bmatrix} d \\ c \end{bmatrix} = \begin{bmatrix} \tilde{S}'\tilde{Y} \\ \tilde{Q}'\tilde{Y} \end{bmatrix} \tag{27}$$

Fixing smoothing parameters λ and θ_β (hidden in Q and $\tilde{Q}$ if present), assuming that $\tilde{Q}$ is full

column rank, 27 can be solved by the Cholesky decomposition of the $(n + d_0) \times (n + d_0)$ matrix followed by forward and backward substitution. See **?**. Singularity of $\tilde{Q}$ demands special consideration. Write the Cholesky decomposition

$$\begin{bmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{Q} \\ \tilde{Q}'\tilde{S} & \tilde{Q}'\tilde{Q} + \lambda Q \end{bmatrix} = \begin{bmatrix} C_1' & 0 \\ C_2' & C_3' \end{bmatrix} \begin{bmatrix} C_1 & C_2 \\ 0 & C_3 \end{bmatrix} \tag{28}$$

where $\tilde{S}'\tilde{S} = C_1'C_1$, $C_2 = C_1^{-T}\tilde{S}'\tilde{Q}$, and $C_3'C_3 = \lambda Q + \tilde{Q}'\left(I - \tilde{S}\left(\tilde{S}'\tilde{S}\right)^{-1}\tilde{S}'\right)\tilde{Q}$. Using an exchange of indices known as pivoting, one may write

$$C_3 = \begin{bmatrix} H_1 & H_2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} H \\ 0 \end{bmatrix},$$

where H_1 is nonsingular. Define

$$\tilde{C}_3 = \begin{bmatrix} H_1 & H_2 \\ 0 & \delta I \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} C_1 & C_2 \\ 0 & \tilde{C}_3 \end{bmatrix}; \tag{29}$$

then

$$\tilde{C}^{-1} = \begin{bmatrix} C_1^{-1} & -C_1^{-1}C_2\tilde{C}_3^{-1} \\ 0 & \tilde{C}_3^{-1} \end{bmatrix}. \tag{30}$$

Premultiplying **??** by $\tilde{C}^{-T}$, straightforward algebra gives

$$\begin{bmatrix} I & 0 \\ 0 & \tilde{C}_3^{-T}C_3^TC_3\tilde{C}_3^{-1} \end{bmatrix}\begin{bmatrix} \tilde{d} \\ \tilde{c} \end{bmatrix} = \begin{bmatrix} C_1^{-T}\tilde{S}'\tilde{Y} \\ \tilde{C}_3^{-T}\tilde{Q}'\left(I - \tilde{S}\left(\tilde{S}'\tilde{S}\right)^{-1}\tilde{S}'\right)\tilde{Y} \end{bmatrix} \tag{31}$$

where $\left(\tilde{d}' \ \ \tilde{c}'\right)' = \tilde{C}'\left(d \ \ c\right)'$. Partition $\tilde{C}_3 = \begin{bmatrix} K & L \end{bmatrix}$; then $HK = I$ and $HL = 0$. So

$$\begin{aligned} \tilde{C}_3^{-T}C_3^TC_3\tilde{C}_3^{-1} &= \begin{bmatrix} K' \\ L' \end{bmatrix} C_3'C_3 \begin{bmatrix} K & L \end{bmatrix} \\ &= \begin{bmatrix} K' \\ L' \end{bmatrix} H'H \begin{bmatrix} K & L \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

If $L'C_3^TC_3L = 0$, then $L'\tilde{Q}'\left(I - \tilde{S}\left(\tilde{S}'\tilde{S}\right)^{-1}\tilde{S}'\right)\tilde{Q}L = 0$, so $L'\tilde{Q}'\left(I - \tilde{S}\left(\tilde{S}'\tilde{S}\right)^{-1}\tilde{S}'\right)\tilde{Y} = 0$. Thus, the linear system has form

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \tilde{d} \\ \tilde{c}_1 \\ \tilde{c}_2 \end{bmatrix} = \begin{bmatrix} * \\ * \\ 0 \end{bmatrix}, \tag{32}$$

which can be solved, but with c_2 arbitrary. One may perform the Cholesky decomposition of 27 with pivoting, replace the trailing 0 with δI for appropriate value of δ, and proceed as if $\tilde{Q}$ were of full rank.

It follows that

$$\widehat{\tilde{Y}} = \tilde{S}d + \tilde{Q}c = \begin{bmatrix} \tilde{S} & \tilde{Q} \end{bmatrix}\tilde{C}^{-1}\tilde{C}^{-T}\begin{bmatrix} \tilde{S}' \\ \tilde{Q}' \end{bmatrix}\tilde{Y} = \tilde{A}\left(\lambda\right)\tilde{Y}. \tag{33}$$

2.1 Minimization of GCV and GML scores with multiple smoothing parameters

The expression in permits the straightforward evaluation of the GCV score

$$V(\lambda, \boldsymbol{\theta}) = \frac{(1/n_y) \left\| \left(I - \tilde{A}(\lambda, \boldsymbol{\theta}) \right) \tilde{Y} \right\|^2}{\left[(1/n_y) \left(I - \tilde{A}(\lambda, \boldsymbol{\theta}) \right) \right]^2} \tag{34}$$

and the GML score

$$M(\lambda, \boldsymbol{\theta}) = \frac{(1/n_y) \tilde{Y}' \left(I - \tilde{A}(\lambda, \boldsymbol{\theta}) \right) \tilde{Y}}{\left[\det^+ \left(I - \tilde{A}(\lambda, \boldsymbol{\theta}) \right) \right]^{1/n_y}}. \tag{35}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_g)'$ denotes the vector of smoothing parameters associated with each RK. To minimize the functions $V(\lambda, \boldsymbol{\theta})$ and $M(\lambda, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and λ, we iterate as follows:

1. Fix $\boldsymbol{\theta}$; minimize $V(\lambda|\boldsymbol{\theta})$ or $M(\lambda|\boldsymbol{\theta})$ with respect to λ.
2. Update $\boldsymbol{\theta}$ using the current estimate of λ.

Executing step 1 follows immediately from the expression for the smoothing matrix. Step 2 requires evaluating the gradient and the Hessian of $V(\boldsymbol{\theta}|\lambda)$ or $M(\boldsymbol{\theta}|\lambda)$ with respect to $\boldsymbol{\kappa} = \log(\boldsymbol{\theta})$. Optimizing with respect to $\boldsymbol{\kappa}$ rather than on the original scale is motivated by two driving factors: first, $\boldsymbol{\kappa}$ is invariant to scale transformations. With examination of V and M and 33, it is immediate that the $\theta_\beta \tilde{Q}_\beta$ are what matter in determining the minimum. Multiplying the $\tilde{Q}_\beta$ by any positive constant leaves the θ_β subject to rescaling, though the problem itself is unchanged by scale transformations. The derivatives of $V(\cdot)$ and $M(\cdot)$ with respect to $\boldsymbol{\kappa}$ are invariant to such transformations, while the derivatives with respect to $\boldsymbol{\theta}$ are not. In addition, optimizing with respect to $\boldsymbol{\kappa}$ converts a constrained optimization ($\theta_\beta \geq 0$) problem to an unconstrained one.