Prediction in regression models with continuous observations

Andrey Pepelyshev

joint with Holger Dette and Anatoly Zhigljavsky

13th MODA
Southampton, UK
July 11, 2023

The model

Consider the common linear regression model

$$y(t) = \theta_1 f_1(t) + \ldots + \theta_m f_m(t) + \varepsilon(t)\ , \ \ t \in \mathcal{T} \subset \mathbb{R}^d$$

- functions $f_1(t), \ldots, f_m(t)$ are linearly independent and continuous,
- a random error field $\varepsilon(t)$ has the zero mean and the covariance kernel $K(t,s) = \mathrm{E}[\varepsilon(t)\varepsilon(s)]$,
- parameters $\theta_1, \ldots, \theta_m$ are unknown and have to be estimated.

Suppose that we observe one realization of a random field.

Prediction with discrete observation

The best linear unbiased predictor (BLUP) of $y(t_0)$ is

$$\hat{y}(t_0) = f^\top(t_0)\hat{\theta}_{\mathrm{BLUE}} + K_{t_0}^\top \Sigma^{-1}(Y - X\hat{\theta}_{\mathrm{BLUE}}),$$

where $\Sigma = \big(K(t_i, t_j)\big)_{i,j=1}^N$, $K_{t_0} = \big(K(t_0, t_1), \ldots, K(t_0, t_N)\big)^\top$, $X = (f(t_1), \ldots, f(t_N))^\top$, $Y = (y(t_1), \ldots, y(t_N))^\top \in \mathbb{R}^N$ and

$$\hat{\theta}_{\mathrm{BLUE}} = (X^\top\Sigma^{-1}X)^{-1}X^\top\Sigma^{-1}Y.$$

The BLUP satisfies the unbiased condition $\mathbb{E}[\hat{y}(t_0)] = \mathbb{E}[y(t_0)]$ and minimizes the mean squared error $\mathrm{MSE}(\tilde{y}(t_0)) = \mathbb{E}\left(y(t_0) - \tilde{y}(t_0)\right)^2$ in the class of all linear unbiased predictors $\tilde{y}(t_0)$; its mean squared error is

$$\mathrm{MSE}(\hat{y}(t_0)) = K(t_0, t_0) - \begin{bmatrix} f(t_0) \\ K_{t_0} \end{bmatrix}^\top \begin{bmatrix} 0 & X^\top \\ X & \Sigma \end{bmatrix}^{-1} \begin{bmatrix} f(t_0) \\ K_{t_0} \end{bmatrix}.$$

Gradient-enhanced kriging is possible.

BLUE for discrete observation

A general estimator

$$\hat{\theta} = G^T Y, \quad Y = (y(t_1), \dots, y(t_N))^T,$$

where G is a $N \times m$-matrix.
The condition of unbiasedness $\mathbb{E}[\hat{\theta}] = \theta$ means

$$G^T X = \mathbf{1}_{m \times m},$$

where $X = (f(t_1), \dots, f(t_N))^T$ is a $N \times m$-matrix.

$$\mathrm{Var}(\hat{\theta}) = G^T \Sigma G \to \min_{unbiased \;\; G}$$

By Gauss-Markov theorem, the best linear unbiased estimator (BLUE) is

$$G_*^T = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1},$$

where $\Sigma = (K(t_i, t_j))_{i,j=1}^N$.

BLUE for discrete observation with derivatives

A general estimator

$$\hat{\theta} = G_0^T Y_0 + G_1^T Y_1, \quad Y_0 = (y(t_1), \ldots, y(t_N))^T,$$

$$Y_1 = (y'(t_1), \ldots, y'(t_N))^T,$$

where G_0 and G_1 are $N \times m$-matrices.
The covariance matrix is

$$\mathrm{Var}(\hat{\theta}) = G^T \Sigma G,$$

where $G = \begin{pmatrix} G_0 \\ G_1 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{00} & \Sigma_{10}^\top \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix}$ is a block matrix,
$\Sigma_{00} = \big(K(t_i, t_j)\big)_{i,j=1}^N$, $\Sigma_{10} = \big(\frac{\partial}{\partial t_i} K(t_i, t_j)\big)_{i,j=1}^N$,
$\Sigma_{11} = \big(\frac{\partial^2}{\partial t_i \partial t_j} K(t_i, t_j)\big)_{i,j=1}^N$.

Continuous observation without derivatives

A general estimator

$$\hat{\theta}_G = \int_{\mathcal{T}} G_0(dt) y(t),$$

where $G_0(dt)$ is a signed vector-measure.
The condition of unbiasedness $\mathbb{E}[\hat{\theta}] = \theta$ means

$$\int_{\mathcal{T}} f(t) G_0^T(dt) = \mathbf{1}_{m \times m}.$$

The covariance matrix of any unbiased estimator $\hat{\theta}_G$ is

$$\mathrm{Var}(\hat{\theta}_G) = \int_{\mathcal{T}} \int_{\mathcal{T}} K(t,s) G_0(dt) G_0^T(ds).$$

The continuous BLUE minimizes this matrix and may not exist.

Continuous observation with derivatives

A general estimator

$$\hat{\theta}_G = \int_{\mathcal{T}} G_0(dt)y(t) + \ldots + \int_{\mathcal{T}} G_q(dt)y^{(q)}(t),$$

where $G_0(dt), \ldots, G_q(dt)$ are signed vector-measures.
The condition of unbiasedness $\mathbb{E}[\hat{\theta}] = \theta$ means

$$\int_{\mathcal{T}} f(t)G_0^T(dt) + \ldots + \int_{\mathcal{T}} f^{(q)}(t)G_q^T(dt) = \mathbf{1}_{m\times m}.$$

The covariance matrix of any unbiased estimator $\hat{\theta}_G$ is

$$\mathrm{Var}(\hat{\theta}_G) = \sum_{i=0}^{q}\sum_{j=0}^{q}\int_{\mathcal{T}}\int_{\mathcal{T}} \frac{\partial^{i+j}K(t,s)}{\partial t^i \partial s^j} G_i(dt)G_j^T(ds).$$

The continuous BLUE minimizes this matrix and may not exist.

Representation of continuous BLUE

Let $\zeta_0, \dots, \zeta_q$ be some signed vector-measures defined on $\mathcal{T}$ such that the $m \times m$ matrix

$$C = \sum_{i=0}^{q} \int_{\mathcal{T}} \zeta_i(dt) \left(f^{(i)}(t) \right)^T$$

is non-degenerate. Define $G = (G_0, \dots, G_q)$, where G_i are the signed vector-measures and $G_i(dt) = C^{-1}\zeta_i(dt)$ for $i = 0, \dots, q$.

Then the estimator $\hat{\theta}_G$ is unbiased.

Solution for continuous BLUE

Let $K(\cdot,\cdot) \in C^q([A,B]\times[A,B])$ for some $q \geq 0$. Suppose that the process $\{y(t)|t \in [A,B]\}$ along with its q derivatives can be observed at all $t \in \mathcal{T} \subseteq [A,B]$. Assume also that all components of $f(\cdot)$ are q times differentiable. Let $\zeta_0, \ldots, \zeta_q$ be signed vector-measures defined on $\mathcal{T}$ such that the matrix C is non-degenerate. Define $G = (G_0, \ldots, G_q)$, $G_i(dt) = C^{-1}\zeta_i(dt)$ for $i = 0, \ldots, q$.

The estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is the BLUE if and only if

$$\sum_{i=0}^{q} \int_{\mathcal{T}} K^{(i)}(t,s)\zeta_i(dt) = f(s) \quad \forall s \in \mathcal{T}.$$

In this case, the covariance matrix of $\hat{\theta}_G$ is $\mathrm{Var}(\hat{\theta}_G) = C^{-1}$.

Properties

To construct the BLUE, we have to solve

$$\sum_{i=0}^{q} \int_{\mathcal{T}} K^{(i)}(t,s)\zeta_i(dt) = f(s) \quad \forall s \in \mathcal{T}.$$

- For $q = 0$ the BLUE measure is a solution of Fredholm integral equation of the first kind.
- We can solve the equation individually for each component of the vector of regression functions $f(t)$.
- The BLUE measure may not exist.
- For $q > 0$ the BLUE measure is not unique [due to possibility of integration by parts].

BLUP without derivatives

Assume that (1) the best linear unbiased estimator (BLUE) $\hat{\theta}_{\mathrm{BLUE}} = \int_{\mathcal{T}} y(t)G(dt)$ exists,
(2) there exists a signed measure $\zeta_{t_0}(dt)$ which satisfies

$$\int_{\mathcal{T}} K(t,s)\zeta_{t_0}(dt) = K(t_0,s), \;\; \forall s \in \mathcal{T}.$$

Then the BLUP measure Q_* exists and is given by

$$Q_*(dt) = \zeta_{t_0}(dt) + c^\top G(dt),$$

where $c = f(t_0) - \int_{\mathcal{T}} f(t)\zeta_{t_0}(dt)$. The MSE of the BLUP $\hat{y}_{Q_*}(t_0) = \int_{\mathcal{T}} y(t)Q_*(dt)$ is given by

$$\mathrm{MSE}(\hat{y}_{Q_*}(t_0)) = K(t_0,t_0) + c^\top Df(t_0) - \int_{\mathcal{T}} K(t,t_0)Q_*(dt)\,,$$

where $D = \int_{\mathcal{T}}\int_{\mathcal{T}} K(t,s)G(dt)G^\top(ds)$ is the covariance matrix of $\hat{\theta}_{\mathrm{BLUE}} = \int_{\mathcal{T}} y(t)G(dt)$.

BLUP with derivatives 1

Consider the problem of prediction of $y^{(p)}(t_0)$, the p-th derivative of y at a point $t_0 \notin \mathrm{T}_p$, where $0 \le p \le q$.
A general linear predictor of the p-th derivative $y^{(p)}(t_0)$ can be defined as

$$\hat{y}_{p,Q}(t_0) = \int \mathbf{Y}^\top(t)\mathbf{Q}(dt) = \sum_{i=0}^{q} \int_{\mathrm{T}_i} y^{(i)}(t) Q_i(dt).$$

The estimator $\hat{y}_{p,Q}(t_0)$ is unbiased if $\mathbb{E}[\hat{y}_{p,Q}(t_0)] = \mathbb{E}[y^{(p)}(t_0)]$, which is equivalent to

$$\int \mathbf{F}(t)\mathbf{Q}(dt) = f^{(p)}(t_0),$$

where $\mathbf{F}(t) = \big(f(t), f^{(1)}(t), \dots, f^{(q)}(t)\big)$ is a $m \times (q+1)$-matrix.

BLUP with derivatives 2

Assume that

(1) The best linear unbiased estimator (BLUE) $\hat{\theta}_{\mathrm{BLUE}} = \int \mathbf{G}(dt)\mathbf{Y}(t)$ exists, where $\mathbf{G}(dt)$ is some signed $m \times (q+1)$-matrix measure (that is, the j-th column of $\mathbf{G}(dt)$ is a signed vector measure defined on $\mathcal{T}_j$);

(2) There exists a signed vector-measure $\zeta_{p,t_0}(dt)$ (of size $q+1$) which satisfies the equation

$$\int \mathbf{K}^\top(t,s)\zeta_{p,t_0}(dt) = \frac{\partial^p K(s,t_0)}{\partial t_0^p}, \;\; \forall s \in \mathrm{T}_i \, ,$$

where $\mathbf{K}(t,s) = \left(\frac{\partial^j K(t,s)}{\partial s^j}\right)_{j=0}^{q}$ is a $(q+1)$-dimensional vector.

BLUP with derivatives 3

Under above assumptions,
the BLUP measure $\mathbf{Q}_*$ exists and is given by

$$\mathbf{Q}_*(dt) = \zeta_{p,t_0}(dt) + \mathbf{G}^\top(dt)c_p,$$

where

$$c_p = f^{(p)}(t_0) - \int \mathbf{F}(t)\zeta_{p,t_0}(dt).$$

The MSE of the BLUP $\hat{y}_{p,Q_*}(t_0)$ is given by

$$\mathrm{MSE}(\hat{y}_{p,Q_*}(t_0)) = \left.\frac{\partial^{2p}K(t,s)}{\partial t^p \partial s^p}\right|_{\substack{t=t_0\\ s=t_0}} + c_p^\top D f^{(p)}(t_0) - \int \mathbf{K}^\top(t,t_0)\mathbf{Q}_*(dt),$$

where

$$D = \int\int \mathbf{G}(dt)\mathbb{K}(t,s)\mathbf{G}^\top(ds)$$

is the covariance matrix of $\hat{\theta}_{\mathrm{BLUE}} = \int \mathbf{G}(dt)\mathbf{Y}(t)$.

Examples of continuous BLUP 1

Consider the model

$$y(t) = \theta + \varepsilon(t), \;\; t = (t_1, t_2) \in \mathcal{T} = [0, 1]^2,$$

$$\mathrm{K}(t, t') = \mathbb{E}[\varepsilon(t)\varepsilon(t')] = \exp\left\{-\lambda\left[|t_1 - t_1'| + |t_2 - t_2'|\right]\right\},$$

where $\lambda > 0$ and $t = (t_1, t_2), t' = (t_1', t_2') \in [0, 1]^2$.
The BLUP at the point $T = (T_1, T_2)$ is $\int_{\mathcal{T}} y(t)\mathrm{Q}_*(dt)$, where

$$\mathrm{Q}_*(dt) = \zeta_T(dt) + c\,\mathrm{G}(dt) \;\; \text{with} \;\; c = 1 - \int_{\mathcal{T}} \zeta_T(dt),$$

where $\mathrm{G}(dt)$ is the BLUE measure and does not depend on $T = (T_1, T_2)$.

Examples of continuous BLUP 2

Define

$$G(dt_i) = \frac{1}{2+\lambda}\left[\delta_0(dt_i) + \delta_1(dt_i) + \lambda dt_i\right], \quad t_i \in [0,1].$$

Define

$$\zeta_{T_i}(dt_i) = \begin{cases} e^{-\lambda|T_i|}\delta_0(dt_i), & \text{if} \quad T_i \le 0, \\ \delta_{T_i}(dt_i), & \text{if} \quad 0 \le T_i \le 1, \\ e^{-\lambda(T_i-1)}\delta_1(dt_i), & \text{if} \quad T_i \ge 1. \end{cases}$$

Then we have $\mathrm{Q}_*(dt) = \zeta_T(dt) + c\,\mathrm{G}(dt)$ with

$$\zeta_T(dt) = \zeta_{T_1}(dt_1)\zeta_{T_2}(dt_2), \quad \mathrm{G}(dt) = G(dt_1)G(dt_2).$$

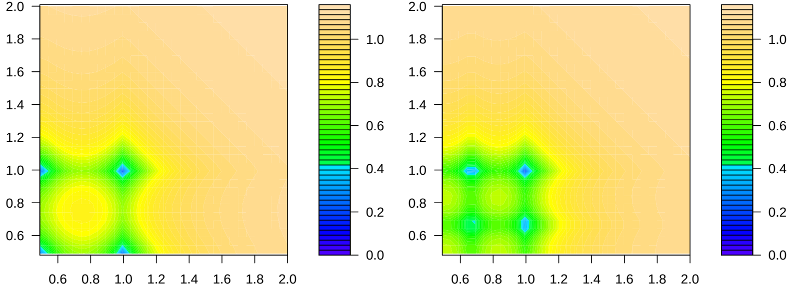
Examples of continuous BLUP 3

For $T_1 \leq 0$ we obtain

$$Q_*(dt) = \begin{cases} e^{-\lambda|T_1|}\delta_0(dt_1)\delta_{T_2}(dt_2) + \left(1 - e^{-\lambda|T_1|}\right) G(dt), & \text{if } 0 \leq T_2 \leq 1, \\ e^{-\lambda|T_1|-\lambda|T_2|}\delta_0(dt_1)\delta_0(dt_2) + \left(1 - e^{-\lambda|T_1|-\lambda|T_2|}\right) G(dt), & \\ \qquad \text{if } \; T_2 \leq 0, & \\ e^{-\lambda|T_1|-\lambda(T_2-1)}\delta_0(dt_1)\delta_1(dt_2) + \left(1 - e^{-\lambda|T_1|-\lambda|T_2-1|}\right) G(dt), & \\ \qquad \text{if } \; T_2 \geq 1. & \end{cases}$$

Similar formulas can be obtained for $0 < T_1 < 1$ and $T_1 \geq 1$.

Examples of continuous BLUP 4



The square root of the MSE of the BLUP for the $N \times N$-point equidistant design at points $(i/(N-1), j/(N-1))$ with $N = 3$ (left) and $N = 4$ (right) and the exponential kernel with $\lambda = 2$.

Examples of continuous BLUP for Matern 3/2

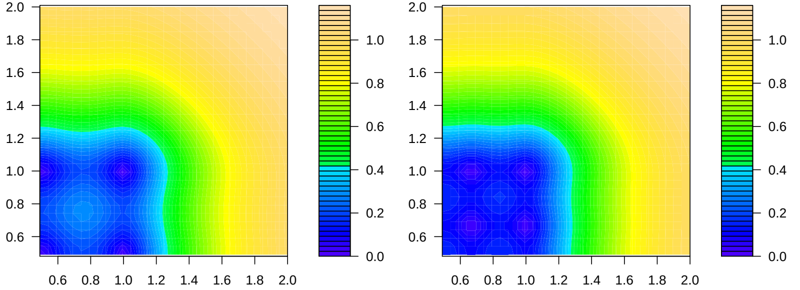
Consider the model $y(t) = \theta + \varepsilon(t)$, $t \in [A, B]$ with Matérn 3/2 covariance kernel $K(t,s) = (1 + \lambda|t - s|)e^{-\lambda|t-s|}$.
Let $t_0 > B$ and we want to predict $\hat{y}_{0,Q_*}(t_0)$. Define
$z_{t_0,B} = (1 + \lambda(t_0 - B))e^{-\lambda(t_0-B)}$, $\quad z_{t_0,1,B} = (t_0 - B)e^{-\lambda(t_0-B)}$,
$C = 1 + \lambda(B - A)/4$, $c_0 = 1 - z_{t_0,B}$). The BLUP measure is

$$\begin{aligned}\mathbf{Q}_*(dt) \;=\; & 0.5c_0\delta_A(dt)/C + (0.5c_0/C + z_{t_0,B})\delta_B(dt) + 0.25c_0\lambda dt/C \\ & -0.25c_0/(C\lambda)\delta_{A,y'}(dt) + (z_{t_0,1,B} + 0.25c_0/(C\lambda))\delta_{B,y'}(dt)\end{aligned}$$

The corresponding BLUP is

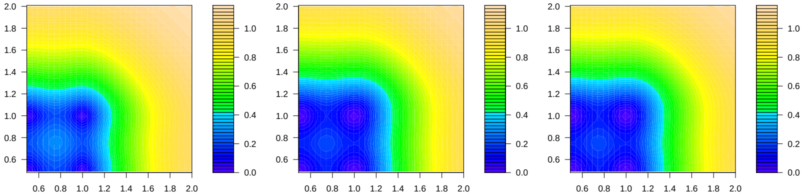
$$\begin{aligned}\hat{y}_{0,Q_*}(t_0) \;=\; & 0.5c_0 y(A)/C + (0.5c_0/C + z_{t_0,B})y(B) + 0.25c_0\lambda\int_A^B y(t)dt/C \\ & -0.25c_0/(C\lambda)y'(A) + (z_{t_0,1,B} + 0.25c_0/(C\lambda))y'(B).\end{aligned}$$

Examples of continuous BLUP for Matern 3/2



Square root of the MSE of the BLUP for the design $\xi_{N^2,0,0,0}$ with $N = 3$ (left) and $N = 4$ (right), and the Matérn 3/2 product-kernel with $\lambda = 2$.

Examples of continuous BLUP for Matern 3/2



Square root of the MSE of the BLUP for the design $\xi_{N^2,0,0,0}$ (left), $\xi_{N^2,4N-4,4N-4,4N-4}$ (center) and ξ_{N^2,N^2,N^2,N^2} (right) with $N = 3$ and the Matérn 3/2 product-kernel with $\lambda = 2$.

The MSE is the same for $\xi_{N^2,4N-4,4N-4,4N-4}$ and ξ_{N^2,N^2,N^2,N^2}.

References

- Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2023). Prediction in regression models with continuous observations. Stat papers. accepted.
- Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2019). The BLUE in continuous-time regression models with correlated errors. Annals of Statistics, 47, 1928-1959.
- Parzen, E. (1961). An approach to time series analysis. The Annals of Mathematical Statistics, 32(4):951–989.
- Näther, W. (1985). Effective observation of random fields. Teubner Verlagsgesellschaft, Leipzig.