On optimal prediction of missing functional data with memory

Tommi Sottinen

University of Vaasa, Finland

This joint work with P. Ilmonen, N. Shafik, G. Van Bever, and L. Viitasaari

EcoSta 2026,
8-10 August 2026, Kyoto

Abstract

We consider the problem of reconstructing missing parts of functions based on their observed segments. It provides, for Gaussian processes and arbitrary bijective transformations thereof, theoretical expressions for the L^2-optimal reconstruction of the missing parts. These functions are obtained as solutions of explicit integral equations. In the discrete case, approximations of the solutions provide consistent expressions of all missing values of the processes. Rates of convergence of these approximations are provided. In the case of Gaussian processes with a parametric covariance structure, the estimation can be conducted separately for each function, and yields nonlinear solutions in presence of memory. Simulated examples show that the proposed reconstruction indeed fares better than the conventional interpolation methods in various situations. The methodology is also illustrated on S&P500 data.

Reference

The talk is based on the preprint

Van Bever, G., Ilmonen, P., Viitasaari, L., Shafik, N. and Sottinen, T. (2022) On optimal prediction of missing functional data with memory. arXiv preprint arXiv:2208.09925

which is under revision for Bernoulli.

Outline

Outline

Setting

In functional data analysis, the observed units are (i.i.d.) random curves $(Y^j_t)_{t \in I_j}$, $j = 1, \ldots, J$, where $I_j \subset [0, T] \subset \mathbb{R}$.

We assume that the observed sets I_j are

$$\Delta = [t^1_L, t^1_U] \cup \cdots \cup [t^n_L, t^n_L],$$

which may depend on j.

We want to estimate the conditional expectation

$$\hat{Y}^j_s = \mathbf{E}[Y^j_s | \mathcal{F}^Y_\Delta],$$

where Y^j_s is a missing value. Note that $\hat{Y}^j_s$ is optimal in the sense that it minimizes

$$\int_0^T \mathbf{E}[(Y^j_s - \hat{Y}^j_s)^2] \, \mathrm{d}s.$$

Outline

Gaussian Fredholm case, I

Suppose the underlying process X is Gaussian with

$$\sup_{t \in [0,T]} \mathbf{E}[X_t^2] < \infty.$$

Then it follows that there exists a Brownian motion W and a Fredholm kernel K such that

$$X_t = \int_0^T K(t,s)\,\mathrm{d}W_s$$

We assume that the first chaoses satisfy $H_T^X = H_T^W$.

Gaussian Fredholm case, II

Theorem

Let X be a Gaussian Fredholm process satisfying the assumptions above. Then

$$\hat{X}_t = \mathbf{E}[X_t|\mathcal{F}^X_\Delta] = \int_0^T f_t(s)\,\mathrm{d}W_s,$$

where f_t is the solution of the system of Fredholm integral equations

$$\int_0^T f_t(s)K(u_k,s)\,\mathrm{d}s = \int_0^T K(t,s)K(u_k,s)\,\mathrm{d}s,$$

for all k and $t_L^k \leq u_k \leq t_U^k$.

Gaussian Fredholm case, III

Theorem

Let X be a Gaussian Fredholm process satisfying the assumptions above. Then $X|\mathcal{F}_\Delta^X$ is Gaussian with random mean $\hat{X}$ and deterministic covariance

$$\rho(t, s|\Delta) = \int_0^T [K(t,u) - f_t(u)][K(s,u) - f_s(u)]\,\mathrm{d}u.$$

Outline

Gaussian example, I

Let X be Gaussian Volterra process:

$$X_t = \int_0^t K(t,s)\,\mathrm{d}W_s$$

with $H_t^X = H_t^W$. Let $\Delta = [0, t_U] \cup [t_L, T]$. Then

$$\hat{X}_t = \int_0^{t_U} K(t,s)\mathrm{d}W_s + c(t)\int_{t_U}^{t_L} K(t_L,s)\,\mathrm{d}W_s + \int_{t_L}^{T} g_t(s)\,\mathrm{d}W_s,$$

where

$$c(t) = \frac{\int_{t_U}^{t} K(t,s)K(t_L,s)\,\mathrm{d}s}{\int_{t_U}^{t_L} K(t_L,s)^2\mathrm{d}s}$$

Gaussian example, II

and g satisfies the Volterra equation of the first kind:

$$\int_{t_L}^{n} g_t(s) K(u,s)\,\mathrm{d}u = \int_{t_U}^{t_L} [K(t,s) - c(t)K(t_L,s)] K(u,s)\,\mathrm{d}u$$

Outline

Non-Gaussian case

Theorem

Let $Y_t = f(t, X_t)$ where X is a Gaussian Fredholm process satisfying the assumptions above and $f(t,\cdot)$ is bijective for all t. Then

$$\mathbf{E}[Y_t|\mathcal{F}_\Delta^Y] = \int_{-\infty}^{\infty} f\left(t, \hat{X}_t + z\sqrt{\rho(t,t|\Delta)}\right)\varphi(z)\mathrm{d}z,$$

where φ is the standard normal density.

Outline

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}_\Delta^Y]$, I

Let $Y^1, \dots, Y^J$ be i.i.d. observations on the sets $\Delta_1, \dots, \Delta_J$.

We assume that the transfer function $f(t,z)$ continuous and monotone in z.

We assume that the variance function $\mathbf{E}[X_t^2]$ is known.

Proposition

Let F_t and Q_t be the cdf and quantile function of Y_t. Then

$$f(t,z) = Q_t(\Phi_t(z)),$$

where Φ_t is the cdf of a centered Gaussian variable with variance $\mathbf{E}[X_t^2]$.

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}^Y_\Delta]$, II

This is how we estimate the predictors $\mathbf{E}[Y^j_t|\mathcal{F}^Y_{\Delta_j}]$ for $t \in \Delta_j$.

We assume we are given a discrete net π_N on points and the corresponding values Y_{t_j}, $t_j \in \Delta_j \cap \pi_N$. Let $K(t) = \{j; t \in \Delta_j \cap \pi_N\}$ and $J(t,s) = \{j; t, s \in \Delta_j \cap \pi_N\}$.

The estimation is then done by using the following steps

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}^Y_\Delta]$, III

$$\hat{F}_{t,K}(z) = \frac{1}{|K|+1} \sum_{k \in K} \mathbf{1}_{Y_t^k \leq z}.$$

$$\hat{f}_K(t,z) = \hat{Q}_{t,K}(\Phi_t(z)).$$

The proxy $X_{t,K}^{\mathrm{proxy}}$ for the value X_t is

$$X_{t,k}^{\mathrm{proxy}} = \Phi_t^{-1}(\hat{F}_{t,K}(Y_t)).$$

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}_\Delta^Y]$, IV

Let $t, s \in \pi_N$.

$$\hat{R}(t,s) = \frac{1}{|J|} \sum_{j \in J} X_{t,j}^{\text{proxy}} X_{s,j}^{\text{proxy}}.$$

Let $\mathbf{R}_{N,\Delta}$ be the matrix collecting the covariance estimators and let $\mathbf{R}_N = \mathbf{R}_{N,\pi_N}$. Let $\mathbf{b}_{N,\Delta}(t)$ collect the estimates $\hat{R}_J(t,s)$, $s \in \Delta \cap \pi_N$. For $j \in \pi_N \setminus \Delta_j$ approximate

$$\hat{X}_{t,N}^j = (a_j(t))' X_{\Delta_j,K}^{\text{proxy}},$$

where

$$a_j(t) = \mathbf{R}_{N,\Delta_j}^{-1} \mathbf{b}_{N,\Delta_j}(t).$$

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}_{\Delta}^{Y}]$, V

$$\hat{\rho}(t,t|\Delta\cap\pi_N)=\frac{1}{M}\sum_{m=1}^{M}\left(Z_{t,N}^{m}-\hat{Z}_{t,N}^{m}\right).$$

Here Z^m's are centered Gaussian vectors with covariance $\mathbf{R}_N$.

Finally we have to estimate the integral. This can be done by finite Riemann sums

$$I_{\varphi,L}(h)=\sum_{z_k\in[-L,L]}h(z_k)\varphi(z_k)(z_k-z_{k-1}),$$

for example.

CONSISTENT ESTIMATION OF $\mathbf{E}[Y|\mathcal{F}_\Delta^Y]$, VI

Let $t \in \pi_N \setminus \Delta_J$. For given L and M define

$$\hat{Y}^j_{t,L,M,K,N} = I_{\varphi,L}(\hat{h}_M(t,\cdot)),$$

where

$$\hat{h}_M(t,z) = \hat{f}_K(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}_M(t,t|\Delta_j \cap \pi_N)}).$$

Consistent estimation of $\mathbf{E}[Y|\mathcal{F}^Y_\Delta]$, VII

Theorem

Let $J^* = \min_{t,s\in\pi_N} |J(t,s)|$. *Assume* $\mathcal{F}^Y_{\Delta\cap\pi_n}$ *increases to* $\mathcal{F}^Y_\Delta$. *Then*

$$\lim_{L\to\infty}\lim_{N\to\infty}\lim_{J^*\to\infty}\lim_{M\to\infty}\hat{Y}^j_{t,L,M,K,N} = \mathbf{E}[Y_t|\mathcal{F}^Y_\Delta].$$

Thank you for listening!
Any questions?