SCIENCE ■ PASSION ■ TECHNOLOGY

Institute of Statistics ■ Kopernikusgasse 24/III, 8010 Graz, Austria ■ *stat.tugraz.at*

Estimating the conditional distribution in functional regression problems

Siegfried Hörmann [1], Thomas Kuenzer [1], Gregory Rice [2]

[1] Institute of Statistics, Graz University of Technology, Graz, Austria.
[2] Department of Statistics and Actuarial Science, University of Waterloo, Canada.

Abstract. We consider the problem of consistently estimating the conditional distribution $P(Y \in A|X)$ of a functional data object $Y = (Y(t) : t \in [0, 1])$ given covariates X in a general space, assuming that Y and X are related by a functional linear regression model. Two natural estimation methods for this problem are proposed, based on either bootstrapping the estimated model residuals, or fitting functional parametric models to the model residuals and estimating $P(Y \in A|X)$ via simulation. We show that under general consistency conditions on the regression operator estimator, which hold for certain functional principal component based estimators, consistent estimation of the conditional distribution can be achieved, both when Y is an element of a separable Hilbert space, and when Y is an element of the Banach space of continuous functions on the unit interval. The latter results imply that sets A that specify path properties of Y that are of interest in applications can be considered, such as the maximum of the curve. Our methods have numerous applications in the context of constructing prediction sets, quantile regression and VaR estimation. Compared to direct modelling these curve properties using scalar-on-function regression, modelling the whole response distribution and extracting the curve properties in a second step allows us to harness the full information contained in the functional data to fit the regression model and achieve better results. We study the proposed methods in several simulation experiments and real data analysis of electricity price curves and show that they outperform both the non-parametric kernel estimator and functional binary regression.

References

D. Bosq. *Linear processes in function spaces: theory and applications.* Lecture Notes in Statistics. Springer, New York, 2000.

H. Dette, K. Kokot, and A. Aue. Functional data analysis in the Banach space of continuous functions. *The Annals of Statistics*, 48(2), 2020.

S. Hörmann and P. Kokoszka. Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884, 2010.

S. Hörmann, T. Kuenzer, and G. Rice. Estimating the conditional distribution in functional regression problems. *arXiv preprint arXiv:2105.01412*, 2021.

A regression problem

We observe a strictly stationary process $(Y_k, X_k)_{k\in\mathbb{Z}} \subset H_2 \times H_1$ that following a general functional linear regression model

$$Y_k = \varrho(X_k) + \varepsilon_k. \tag{1}$$

Here Y_k is a curve on $[0, 1]$, ϱ is a linear operator and X_k is an element of a Hilbert space. This setup includes linear function-on-function regression and functional autoregressive models.

We want to estimate the conditional distribution of Y given X, i.e. $P(Y \in A|X)$, for sets of interest $A \subset H_2$ that might stem from transformations $T(Y)$ such as the maximum of the curve.

Consistent estimation is possible only for continuity sets of the distribution, i.e. sets A with $P(Y \in \partial A) = 0$. This property depends strongly on the choice of the space H_2 and its norm. For many problems, a stronger metric than the common L^2-metric is needed. For example, prediction bands $A = \{y\colon \lambda(t : a(t) < y(t) < b(t)) = 1\}$ cannot be used in $L^2[0, 1]$ since $\partial A = A$. For such path properties, the space of continuous functions on $[0, 1]$ equipped with the supremum norm is more appropriate.

We propose simple procedures to estimate $P(Y \in A|X)$. Replacing ϱ with a consistent estimator $\hat\varrho_n$ based on the sample $(Y_k, X_k)_{1\le k\le n}$, we estimate $P(Y \in A|X)$ by one of two bootstrap methods:

1. **i.i.d. bootstrap:** we resample the estimated residuals $\hat\varepsilon_{k,n} = Y_k - \hat\varrho_n(X_k)$ and use the empirical distribution of $\hat\varrho_n(X) + \hat\varepsilon_{k,n}$, obtaining $\hat P_n^{\mathsf{B}}(Y \in A|X)$.
2. **Gaussian bootstrap:** assuming Gaussianity of the model errors ε_k, we model Y conditioned on X as a Gaussian process with mean $\hat\varrho_n(X)$, and covariance estimated from the residual sequence $\hat\varepsilon_{k,n}$, obtaining $\hat P_n^{\mathsf{G}}(Y \in A|X)$.

We use the truncated principal components based estimator $\hat\varrho_n$. Calculating the empirical covariance and cross-covariance by

$$\widehat C_{XX} = \frac{1}{n}\sum_{k=1}^{n} X_k \otimes X_k, \quad \text{and} \quad \widehat C_{YX} = \frac{1}{n}\sum_{k=1}^{n} Y_k \otimes X_k,$$

we define the estimator of ϱ as

$$\hat\varrho_n(x) := \sum_{i=1}^{T_n} \frac{1}{\hat\lambda_i} \widehat C_{YX}\, \hat v_i \otimes \hat v_i(x), \tag{2}$$

where $\hat\lambda_i$ and $\hat v_i$ are the non-increasing eigenvalues and the eigenfunctions of $\widehat C_{XX}$.

Example: Level sets

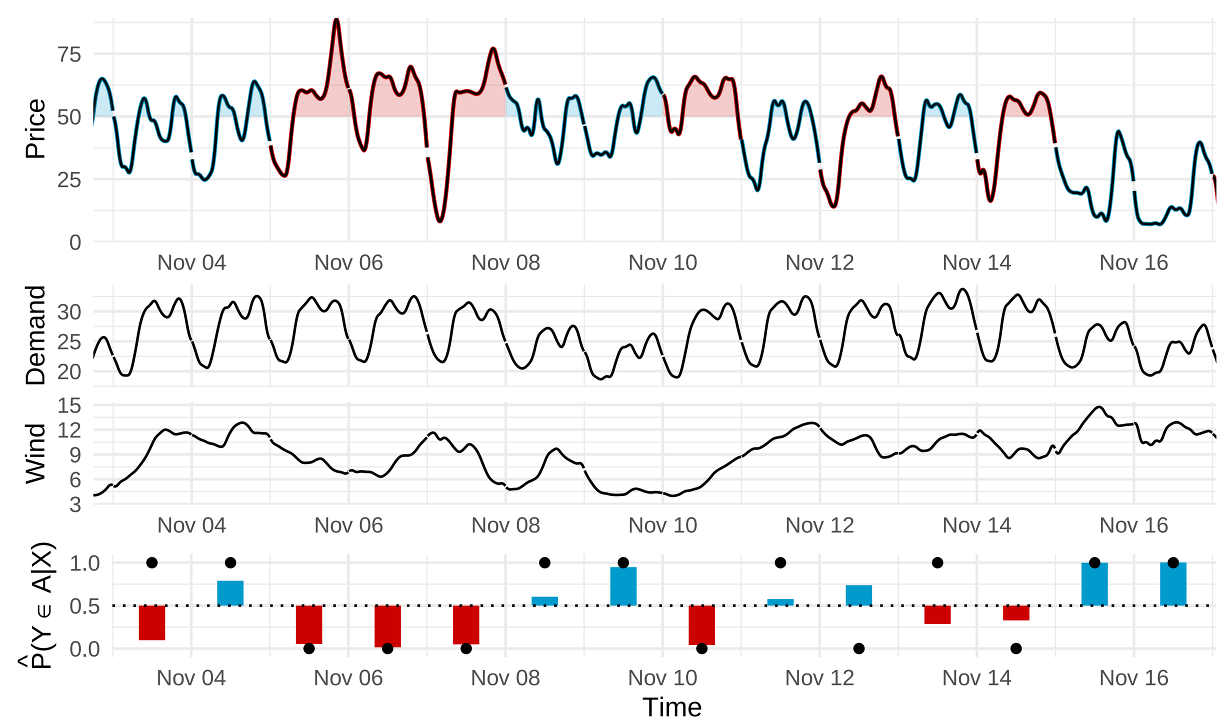
Let

$$A_{\alpha,z} := \{y \in H_2\colon \lambda(t\colon y(t) > \alpha) \le z\}$$

for some $z \in [0, 1]$ and $\alpha \in \mathbb{R}$. The *level set* $A_{\alpha,z}$ contains curves that stay a limited amount of time z above a threshold α.

As a practical example, we look at electricity data and model the price using an FARX(7) model that includes several covariates. We predict the probability that the curve of the next day falls in a level set $A_{\alpha,z}$. Forecasting whether price or demand curves will spend prolonged periods of time above certain levels is useful in anticipating volatility in continuous intraday electricity markets and planning for peak loads. This falls within the scope of the general problem we consider.

Our approach outperforms competing methods such as the non-parametric Nadaraya–Watson estimator and the functional binary regression model, both in the real data scenario and in simulation studies. Other applications of our method include prediction bands and quantile predictions. Its results are competitive to other, more specialized methods and more complex bootstrap schemes. For rare events A, the Gaussian bootstrap delivers an additional improvement in performance compared to the i.i.d. bootstrap.



Top: *electricity price with the covariates demand and wind energy production. Price curves are colored blue or red according to whether or not they lie in the level set $A = \{y \in H_2 : \lambda(t : y(t) > 50) \le 0.5\}$.* ***Bottom:*** *estimated conditional probability $\hat P^B(Y_k \in A|X_k)$ with the decision threshold ½.*

Consistency

For asymptotic consistency, we select the truncation parameter T_n as

$$T_n = \max\{j \ge 1\colon \hat\lambda_j \ge m_n^{-1}\}, \quad \text{with} \quad m_n \to \infty.$$

The sequence m_n decides at which point to truncate the estimator $\hat\varrho_n$.

Assumption 1.

(i) The process $(X_k)_{k\in\mathbb{Z}}$ has mean zero, and is L^4-m-approximable in the separable Hilbert space H_1.

(ii) The noise $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a mean zero, i.i.d. sequence in the space $H_2 = C[0, 1]$, satisfies $E\|\varepsilon_k\|_\infty^4 < \infty$ and is independent from $(X_j)_{j\le k}$ for all $k \in \mathbb{Z}$.

(iii) For some $0 < \alpha \le 1$, the noise ε_k a.s. satisfies the Hölder condition

$$|\varepsilon_k(t) - \varepsilon_k(s)| < M_k\,|t - s|^\alpha$$

where M_k is a r.v. independent from X_k, with $EM_k^2 < \infty$.

(iv) $\varrho\colon H_1 \to H_2$ is a bounded linear operator that satisfies

$$\forall\, x \in H_1\colon \quad |\varrho x(t) - \varrho x(s)| \le M_\varrho \|x\|\,|t - s|^\alpha,$$

where M_ϱ is a finite constant.

The Hölder condition is fulfilled by many stochastic processes such as the (fractional) Brownian motion. While the assumption that H_1 be a Hilbert space is typically no restriction, care needs to be taken in the case of an FAR model. If ϱ is a kernel operator or X_k is sufficiently smooth, this presents no problem, as we can choose suitable Hilbert spaces to embed the time series.

Theorem 1. *Suppose that Assumption 1 holds and we define $\hat\varrho_n$ as in (2) with $m_n = o(n^{\alpha/2})$. If $P(Y \in \partial A) = 0$, then $\hat P_n^{\mathsf{B}}(Y \in A|X) \xrightarrow{P} P(Y \in A|X)$ as $n \to \infty$, i.e. the i.i.d. residual bootstrap consistently estimates the conditional distribution of the response.*

If additionally $(\varepsilon_k)_{k\ge1}$ are Gaussian random variables, then $\hat P_n^{\mathsf{G}}(Y \in A|X) \xrightarrow{P} P(Y \in A|X)$ as $n \to \infty$, i.e. the Gaussian bootstrap is also consistent.