TOPICS IN ECONOMETRICS AND STATISTICS

Summer term, 2021

Tim Mensinger
University of Bonn

tmensinger@uni-bonn.de

1. Introduction

2. Heuristics

3. Identification and Estimation

4. Generalizations and Extensions

Introduction

INTRODUCTION

MOTIVATION

Paper

Liebl et al. (2020)

Superconsistent estimation of points of impact in non-parametric regression with functional predictors

Dominik Poß, Dominik Liebl and Alois Kneip,

University of Bonn, Germany

Hedwig Eisenbarth,

Victoria University of Wellington, New Zealand

Tor D. Wager

Dartmouth College, Hanover, USA

and Lisa Feldman Barrett

Northeastern University, Boston, Massachusetts General Hospital and Harvard Medical School, Boston, and Massachusetts General Hospital, Charlestown, USA

[Received May 2019. Final revision May 2020]

Summary. Predicting scalar outcomes by using functional predictors is a classical problem in functional data analysis. In many applications, however, only specific locations or time points of the functional predictors have an influence on the outcome. Such 'points of impact' are typically unknown and must be estimated in addition to estimating the usual model components. We show that our points-of-impact estimator enjoys a superconsistent rate of convergence and does not require knowledge or pre-estimates of the unknown model components. This remarkable result facilitates the subsequent estimation of the remaining model components as shown in the theoretical part, where we consider the case of non-parametric models and the practically relevant case of generalized linear models. The finite sample properties of our estimators are assessed by means of a simulation study. Our methodology is motivated by data from a psychological experiment in which the participants were asked to rate their emotional state continuously while watching an affective video eliciting a varying intensity of emotional reactions.

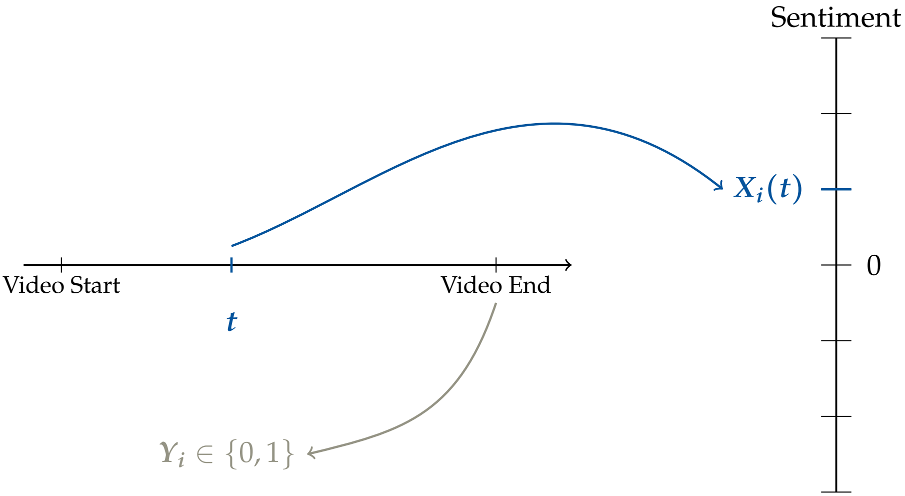
Keywords: Emotional stimuli; Functional data analysis; Non-parametric regression; On-line video rating; Quasi-maximum-likelihood; Variable selection

The Setting

Liebl et al. (2020)

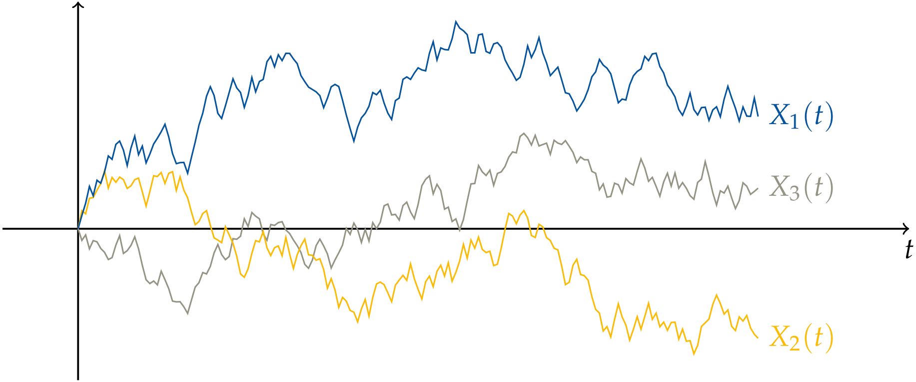
Figure 1: Snapshot of YouTube video

The Setting

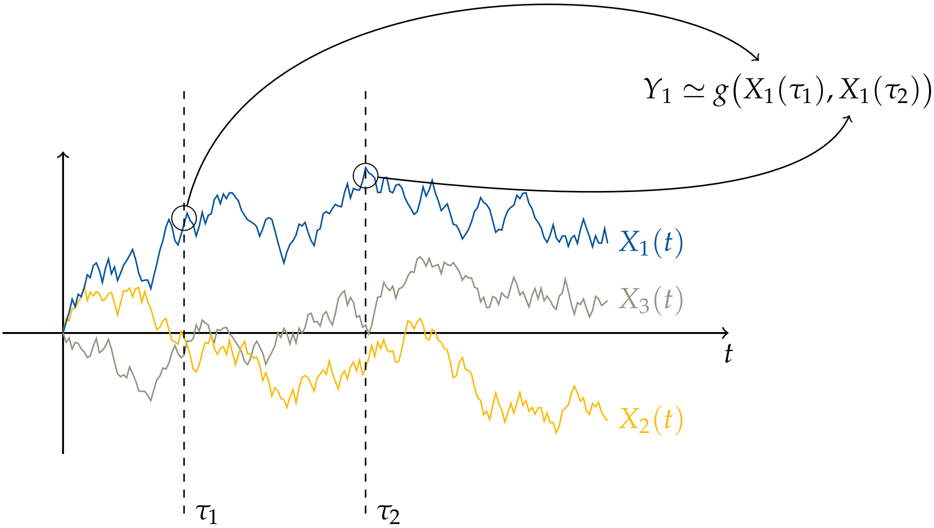


The Data

Regressors



The Problem



Introduction

The Model

The Model

$$\boxed{Y_i = g\left(X_i(\tau_1), \ldots, X_i(\tau_S)\right) + \epsilon_i}$$

- $X_i = \{X_i(t) : t \in [0,1]\}$ square integrable process
- $\mathbb{E}\left[\epsilon_i \mid X_i(t)\right] = 0$ exogeneity
- $g : \mathbb{R}^S \to \mathbb{R}$ *unknown* link function
- $\tau_1, \ldots, \tau_S \in [0,1]$ *unknown* points of impact

Approach

1. stage	Estimate S and $\tau_1, \ldots, \tau_S$
2. stage	Estimate link function g

Focus here: 1. stage

Heuristics

Plan

- Simple linear model $Y_i = \beta_0 + \sum_{s=1}^{S} \beta_s X_i(\tau_s) + \epsilon_i$
- Gaussian Process Regressors $X_i \sim \mathcal{GP}(0, \sigma)$

Task: Think of how to identify/estimate τ_s for different σ

Kernels: White, Brownian Motion, Matern, RBF

First insight: Investigate object $\mathbb{E}\left[Y_i X_i(t)\right]$

Building Intuition

What is the connection between the kernel and a sample trajectory?

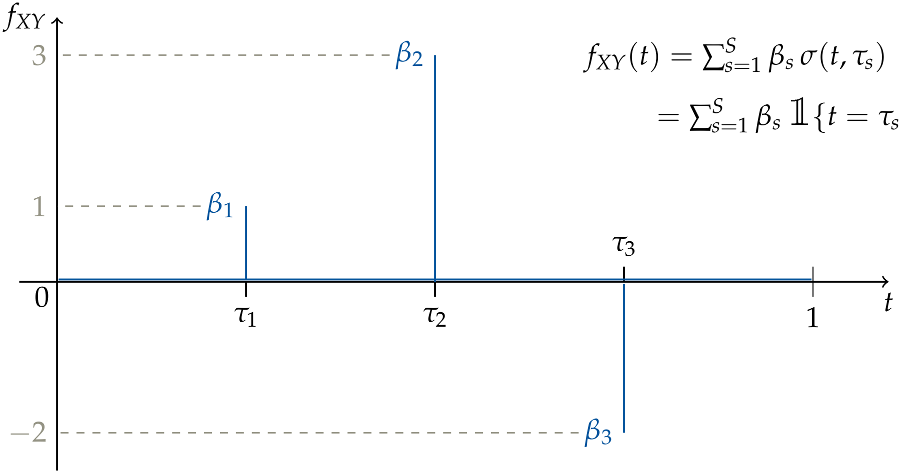
$\Longrightarrow$ Visualization

General Representation

$$
\begin{aligned}
f_{XY}(t) &\stackrel{def}{=} \mathbb{E}\left[Y_i X_i(t)\right] \\
&= \mathbb{E}\left[\left(\beta_0 + \sum_{s=1}^{S} \beta_s X_i(\tau_s) + \epsilon_i\right) X_i(t)\right] \\
&= \sum_{s=1}^{S} \beta_s \, \mathbb{E}\left[X_i(\tau_s) X_i(t)\right] \\
&= \sum_{s=1}^{S} \beta_s \, \sigma(t, \tau_s)
\end{aligned}
$$

White Kernel

$$\sigma(s,t) = \mathbb{1}\{s = t\}$$



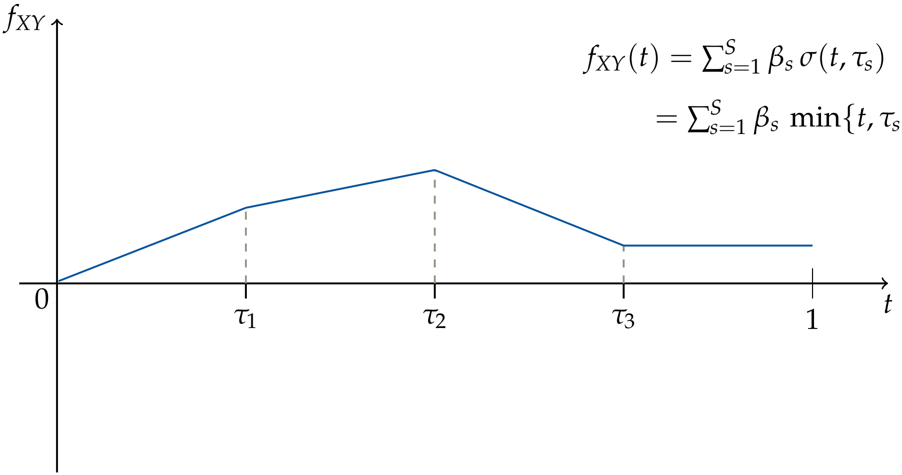
$$f_{XY}(t) = \sum_{s=1}^{S} \beta_s \, \sigma(t, \tau_s)$$
$$= \sum_{s=1}^{S} \beta_s \, \mathbb{1}\{t = \tau_s\}$$

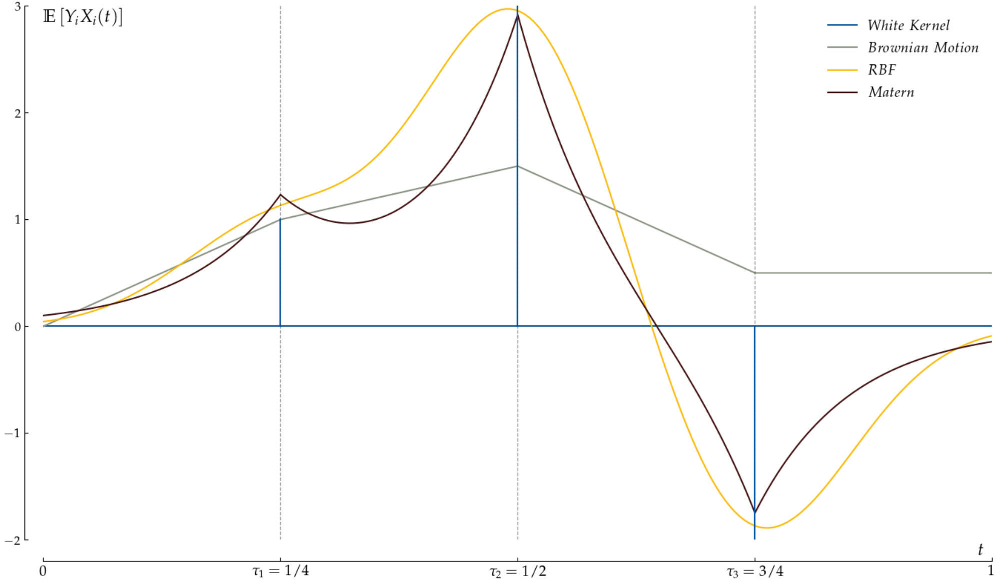
Brownian Motion

$$\sigma(s,t) = \min\{s,t\}$$

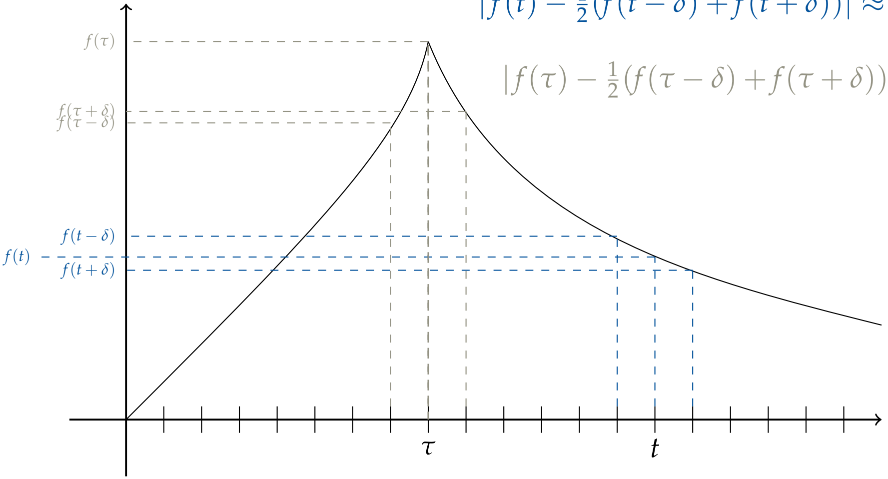
$$f_{XY}(t) = \textstyle\sum_{s=1}^{S} \beta_s \, \sigma(t, \tau_s)$$

$$= \textstyle\sum_{s=1}^{S} \beta_s \, \min\{t, \tau_s\}$$




$\mathbb{E}[Y_i X_i(t)]$
White Kernel
Brownian Motion
RBF
Matern
3
2
1
0
−1
−2
t
0
$\tau_1 = 1/4$
$\tau_2 = 1/2$
$\tau_3 = 3/4$
1

How do we find τ?

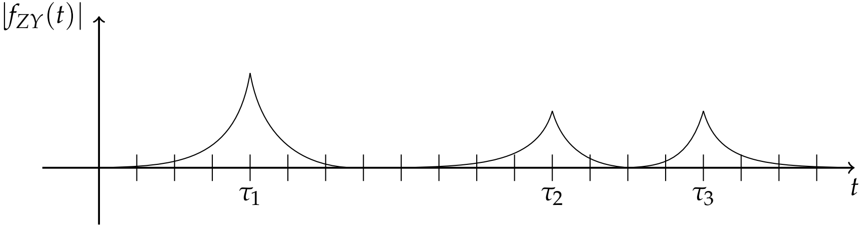


$$|f(t) - \tfrac{1}{2}(f(t-\delta) + f(t+\delta))| \approx 0$$

$$|f(\tau) - \tfrac{1}{2}(f(\tau-\delta) + f(\tau+\delta))| \gg 0$$

How do we find τ?

$$f_{ZY}(t) \stackrel{def}{=} f_{XY}(t) - \frac{1}{2}\left(f_{XY}(t-\delta) + f_{XY}(t+\delta)\right)$$



Estimation?

- $Z_{\delta,i}(t) := X_i(t) - (X_i(t+\delta) + X_i(t-\delta))\,/2$
- $\implies \mathbb{E}\left[Z_{\delta,i}(t)Y_i\right] = f_{XY}(t) - \frac{1}{2}\left(f_{XY}(t+\delta) + f_{XY}(t-\delta)\right) = f_{ZY}(t)$
- Estimate by: $\hat{f}_{ZY}(t) = n^{-1}\sum_{i=1}^{n} Z_{\delta,i}(t)Y_i$
- But what about δ? $\implies$ Visualization

Identification and Estimation

IDENTIFICATION AND ESTIMATION

IDENTIFICATION

Theorem 1

Let X_i be a Gaussian process and $g : \mathbb{R}^S \to \mathbb{R}$ be an arbitrary function with continuous partial derivatives almost everywhere such that

$$0 < \left| \mathbb{E}\left[\frac{\partial}{\partial x_s} g(X_i(\tau_1), \ldots, X_i(\tau_S)) \right] \right| < \infty .$$

Define

$$\vartheta_s = \mathbb{E}\left[\frac{\partial}{\partial x_s} g(X_i(\tau_1), \ldots, X_i(\tau_S)) \right]$$

then,

$$f_{XY}(t) = \mathbb{E}\left[X_i(t) Y_i \right] = \sum_{s=1}^{S} \vartheta_s \sigma(t, \tau_s)$$

Assumption 1

Given the kernel σ, there exists open $\Omega \supset [0,1]^3$ and twice continuously differentiable function $\omega : \Omega \to \mathbb{R}$ as well as some $\kappa \in (0,2)$ such that $\forall s,t \in [0,1]$

$$\sigma(s,t) = \omega(s,t,|s-t|^{\kappa}) \,.$$

Moreover, $0 < \inf\{c(t) : t \in [0,1]\}$, where

$$c(t) = -\frac{\partial}{\partial z}\omega(t,t,z)|_{z=0} \,.$$

Can show that, under Assumption 1 and Theorem 1, as $\delta \to 0$

$$f_{ZY}(t) = \mathbb{E}\left[Z_{\delta,i}(t)Y_i\right] = \begin{cases} \mathcal{O}(\delta^\kappa) & \text{, if } t = \tau_s \text{ for some } \tau_s \in \{\tau_1, \ldots, \tau_S\} \\ \mathcal{O}(\delta^2) & \text{, else} \end{cases}$$

$$\frac{1}{n}\sum_{i=1}^{n} Z_{\delta,i}(t)Y_i - \mathbb{E}\left[Z_{\delta,i}(t)Y_i\right] = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\delta^\kappa / n}\right)$$

As seen in visualization, need sensible choice of δ

- δ too small (e.g. $\delta^\kappa \sim n^{-1}$) $\implies$ estimation noise dominates
- δ too big $\implies$ cannot distinguish between neighboring points

ESTIMATION

Estimation

- X_i observed over p equidistant points t_j in $[0, 1]$
- Possibly $p \gg n$
- Estimate $\{\tau_s\}$ using local maxima of $|\hat{f}_{ZY}(t)| = |n^{-1} \sum_{i=1}^n Z_{\delta,i}(t) Y_i|$
- Algorithm 1: Given $\delta > 0$ determine $\hat{\tau}_1, \dots, \hat{\tau}_{M_\delta}$ with $M_\delta \in \mathbb{N}$

Finally,

$$\hat{S} = \min \left\{ \ell \in \mathbb{N}_0 : \left| \frac{\sum_{i=1}^n Z_{\delta,i}(\hat{\tau}_{\ell+1}) Y_i}{\{\sum_{i=1}^n Z_{\delta,i}(\hat{\tau}_{\ell+1})^2\}^{1/2}} \right| < \lambda \right\},$$

and we select only the first $\hat{\tau}_s$, where $s = 1, \dots, \hat{S}$.

Algorithm 1

compute $\hat{f}_{XY}(t_j) = \sum_i X_i(t_j) Y_i / n$, for all $j = 1, \ldots, p$

choose $\delta > 0$ s.t. $\exists\, k_\delta \in \mathbb{N}$ with $1 \leq k_\delta < (p-1)/2$ and $\delta = k_\delta / (p-1)$

define $\mathcal{J}_\delta = \{k_\delta + 1, \ldots, p - k_\delta\}$ and set $\ell = 1$

compute $\hat{f}_{ZY}(t_j) = \hat{f}_{XY}(t_j) - (\hat{f}_{XY}(t_j + \delta) + \hat{f}_{XY}(t_j - \delta))/2$, for all $j \in \mathcal{J}_\delta$

while $|\mathcal{J}_\delta| \neq \varnothing$ **do**

  estimate $\hat{\tau}_\ell = \operatorname{argmax} \left\{ |\hat{f}_{ZY}(t_j)| : \text{for } t_j \text{ with } j \in \mathcal{J}_\delta \right\}$

  update $\mathcal{J}_\delta \leftarrow \mathcal{J}_\delta \setminus [\hat{\tau}_\ell - \sqrt{\delta}, \hat{\tau}_\ell + \sqrt{\delta}]$

  update $\ell \leftarrow \ell + 1$

return $\{\hat{\tau}_\ell\}$

Asymptotics

Can we say something about the convergence rate of $\hat{\tau}_s$?

Asymptotics

Can we say something about the convergence rate of $\hat{\tau}_s$?

Assumption 1 + 2 and Theorem 1 + 2 $\implies$ superconsistent rates

Assumption 2

(a) $X_1, \ldots, X_n \overset{iid}{\sim} X$, where X is a Gaussian process

(b) $\exists\, 0 < \sigma_{|y|} < \infty$ s.t. $\forall m \geq 1 : \mathbb{E}\left[|Y_i|^{2m}\right] \leq 2^{m-1} m! (\sigma_{|y|})^{2m}$

Condition (b) is fulfilled, for example,if Y_i is bounded, as in the logistic regression case, or if (a) holds and the errors ϵ_i are sub-Gaussian and g has bounded partial derivatives.

Theorem 2

(Under Assumption 1, 2 and Theorem 1)

If $\delta = \delta_n \to 0$ as $n \to \infty$ with $n\delta^\kappa / |\log \delta| \to \infty$ and $\delta^\kappa / n^{-\kappa+1} \to 0$, then

(i) $\max_{\ell=1,\dots,\hat{S}} \min_{s=1,\dots,S} |\hat{\tau}_\ell - \tau_s| = \mathcal{O}_{\mathbb{P}}(n^{-1/\kappa})$

(ii) $\exists\, D \in (0, \infty)$ such that when algorithm 1 is applied with threshold

$\lambda = \lambda_n = A\sqrt{\sigma^2_{|y|}/n \log(1/\delta)}$, with $A > D$, and $\delta^2 = \mathcal{O}(n^{-1})$, then,

$\mathbb{P}\left[\hat{S} = S\right] \to 1$ as $n \to \infty$.

Generalizations and Extensions

Generalizations

- Non-Gaussian processes

  $\rightarrow$ Build on framework of elliptical processes (heavy-tailed, logistic, t)

Generalizations

- Non-Gaussian processes

  $\rightarrow$ Build on framework of elliptical processes (heavy-tailed, logistic, t)

- Generalizing covariance assumption

  $\rightarrow$ Do not need rough process with non-diffb. kernel at diagonal, rather need the kernel to be fewer times differentiable at the diagonal than off the diagonal

Extensions

- Non-IID data (time-series)

Extensions

- Non-IID data (time-series)
- Higher-dim. index set, e.g. $[0,1]^2$ instead of $[0,1]$

Extensions

- Non-IID data (time-series)
- Higher-dim. index set, e.g. $[0,1]^2$ instead of $[0,1]$
- Sparsely sampled data

References

Kneip, A., Poß, D., & Sarda, P. (2016). Functional linear regression with points of impact. *The Annals of Statistics*.

Poß, D., Liebl, D., Kneip, A., Eisenbarth, H., Wager, T. D., & Barrett, L. F. (2020). Superconsistent estimation of points of impact in non-parametric regression with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

st –. (2021). Gaussian process visualization.

YouTube-video. (2009). Youtu.be/9f6upujifay.

Contact

Slides and codes are hosted on GitHub.
For further questions contact me via email.

github.com/timmens/topics-metrics-2021
tmensinger(at)uni-bonn.de