



# On principal components regression with Hilbertian predictors

Ben Jones<sup>1</sup> · Andreas Artemiou<sup>1</sup>

Received: 6 April 2018 / Revised: 23 September 2018 / Published online: 6 December 2018  
© The Institute of Statistical Mathematics, Tokyo 2018

## Abstract

We demonstrate that, in a regression setting with a Hilbertian predictor, a response variable is more likely to be more highly correlated with the leading principal components of the predictor than with trailing ones. This is despite the extraction procedure being unsupervised. Our results are established under the conditional independence model, which includes linear regression and single-index models as special cases, with some assumptions on the regression vector. These results are a generalisation of earlier work which showed that this phenomenon holds for predictors which are real random vectors. A simulation study is used to quantify the phenomenon.

**Keywords** Conditional independence · Hilbertian random variables · Principal components regression · Elliptical distributions · Cauchy distribution

---

✉ Ben Jones  
JonesBL7@cardiff.ac.uk  
Andreas Artemiou  
ArtemiouA@cardiff.ac.uk

<sup>1</sup> School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4AG, UK