Statistics and Its Interface Volume 15 (2022) 157–158

Discussion on "Estimation of Hilbertian varying coefficient models"*

Ming-Yen Cheng

The suggested model covers a broad range of settings in functional data analysis and the proposed smooth backfitting method is advantageous. Here I discuss its relationship with several alternative approaches.

Keywords and phrases: Functional data, Random objects, Semiparametric regression, Structure identification.

The smooth backfitting (SBF) technique, originated by [11], has been developed for various structured semiparametric models for scalar responses [9, 10, 19] and functional/longitudinal responses [12, 13, 21]. Other L_2-norm based models for funcational/longitudinal responses include [4], [8], [22], etc. For density-valued data, [15] introduced a transformation approach, and [5] considered an additive model that describes a density response in terms of level sets at different levels.

[7] introduced additive regression with Hilbertian responses and Euclidean predictors and utilizes SBF in the estimation; equipped with an inner product, the Hilbert space is very general. To cope with discrete predictors and interactions between different predictors, this discussion paper introduces Hilbertian varying coefficient models, which include the simplest varying coefficient models (2.1), the additive varying coefficient models (2.3) and the additive models by [7] as special cases. The most appealing property of the SBF technique in the considered setting is that it operates on the Hilbertain responses as a whole with the Bochner integrals being the same as Lebesgue integrals, and so the computation is simple and efficient. One interesting problem is identification of the nonlinear and linear parts of the continuous predictors and the interaction terms. Another is about choice of the metric which plays an important role in the methodology and theory, see the literature mentioned below.

[16] considered Fréchet regression of random objects on Euclidean predictors, given some metric. There are also a variety of methods that specifically target the case where Euclidean predictors are paired with responses that reside on a finite-dimensional Riemannian manifold [6, 17, 20]. The kernel and spline type methods have been proposed for the case where both predictors and responses are elements of finite-dimensional Riemannian manifolds [1, 18].

A recent approach to including random distributions as predictors in complex regression models is to transform the densities of these distributions to unconstrained functions in the Hilbert space equipped with L_2 norm, e.g., by the log quantile density transformation [15] and then to employ functional regression models where the transformed functions serve as predictors and the responses are either also the transformed functions or scalars [3, 14], whence established methods for functional regression become applicable. However, these methods do not cover spaces of probability measures under the Wasserstein metric, where the tangent spaces are subspaces of infinite-dimensional Hilbert spaces. Wasserstein regression works better for regression modeling when comparing it to the "unadapted" transformation approaches [2].

*Supported by the Research Grants Council GRF grant HKBU12304120 and the Hong Kong Baptist University grant RC-ICRS-17-18.

Received 29 March 2021

REFERENCES

[1] Banerjee, M., Chakraborty, R., Ofori, E., Okun, M. S., Viallancourt, D. E. and Vemuri, B. C. (2016). A nonlinear regression technique for manifold valued data with applications to medical image analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4424–4432.

[2] Chen, Y., Lin, Z. and Müller, H.-G. (2020). Wasserstein regression. *arXiv preprint arXiv:2006.09660*.

[3] Chen, Z., Bao, Y., Li, H. and Spencer Jr, B. F. (2019). LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data. *Mechanical Systems and Signal Processing* **121** 655–674.

[4] Chiou, J.-M., Müller, H.-G. and Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 405–423. MR1983755

[5] Delicado, P. and Vieu, P. (2017). Choosing the most relevant level sets for depicting a sample of densities. *Computational Statistics* **32** 1083–1113. MR3703579

[6] Hinkle, J., Muralidharan, P., Fletcher, P. T. and Joshi, S. (2012). Polynomial regression on Riemannian manifolds. In *European Conference on Computer Vision* 1–14. Springer.

[7] Jeon, J. M. and Park, B. U. (2020). Additive regression with Hilbertian responses. *Annals of Statistics* **48** 2671–2697. MR4152117

[8] Jiang, C.-R. and Wang, J.-L. (2011). Functional single index models for longitudinal data. *Annals of Statistics* **39** 362–388. MR2797850

[9] Lee, Y. K., Mammen, E. and Park, B. U. (2012). Flexible generalized varying coefficient regression models. *Annals of Statistics* **40** 1906–1933. MR3015048

[10] LINTON, O., SPERLICH, S. and VAN KEILEGOM, I. (2008). Estimation of a semiparametric transformation model. *Annals of Statistics* **36** 686–718. MR2396812
[11] MAMMEN, E., LINTON, O. and NIELSEN, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* **27** 1443–1490. MR1742496
[12] MÜLLER, H.-G. and YAO, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103** 1534–1544. MR2504202
[13] PARK, B. U., CHEN, C.-J., TAO, W. and MÜLLER, H.-G. (2018). Singular additive models for function to function regression. *Statistica Sinica* **28** 2497–2520. MR3839871
[14] PETERSEN, A., CHEN, C.-J. and MÜLLER, H.-G. (2019). Quantifying and visualizing intraregional connectivity in resting-state functional magnetic resonance imaging with correlation densities. *Brain connectivity* **9** 37–47.
[15] PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics* **44** 183–218. MR3449766
[16] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics* **47** 691–719. MR3909947
[17] SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. and ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 192–199. Springer.
[18] STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM Journal on Imaging Sciences* **3** 527–563. MR2736019
[19] YU, K., PARK, B. U. and MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *Annals of Statistics* **36** 228–260. MR2387970
[20] YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74** 697–719. MR2965956
[21] ZHANG, X., PARK, B. U. and WANG, J.-L. (2013). Time-varying additive models for longitudinal data. *Journal of the American Statistical Association* **108** 983–998. MR3174678
[22] ZHU, H., LI, R. and KONG, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of statistics* **40** 2634. MR3097615

Ming-Yen Cheng
Department of Mathematics
Hong Kong Baptist University
Kowloon Tong
Hong Kong
E-mail address: chengmingyen@hkbu.edu.hk