Juan A. Cuesta-Albertos, Universidad de Cantabria
A Test for the Functional Linear Model.

Abstract: The history of this research started with the following question: Let $\mathbb{P} \neq \mathbb{Q}$ be two probabilities on a Hilbert space. How many one-dimensional marginal distributions can $\mathbb{P}$ and $\mathbb{Q}$ have in common? The answer (found in [2]) was that really few: it happens that if μ satisfies certain continuity condition, then μ-almost every marginals are different.

Here we extend this result to test the linearity of the regression a real r.v. Y on a Hilbert-valued random element $\mathbf{X}$. Being a bit more precise, let $\mathbb{H}$ be a Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$. We say that the regression of Y on $\mathbf{X}$ is linear if there exists $\beta \in \mathbb{H}$ such that

$$\mathbb{E}[Y|\mathbf{X}] = \langle \beta, \mathbf{X} \rangle, \text{ a.s.} \tag{1}$$

The problem is that this condition is not easy to be tested in practice because the regression space is too large. However, it is not too difficult to prove that (1) is equivalent to show that there exists $\beta \in \mathbb{H}$ such that

$$\mathbb{E}[Y - \langle \beta, \mathbf{X} \rangle | \langle \mathbf{X}, \mathbf{h} \rangle] = 0 \text{ a.s.},$$

for every $\mathbf{h} \in \mathbb{H}$. Now the regressor is real, what eases the problem, but this simplification is not completely satisfactory because it requires to compute to many conditional expectactions. This inconvenience can be fixed because it happens that we have been able to prove that if μ is defined on the Borel σ-algebra on $\mathbb{H}$, and satisfies certain continuity condition, then (1) holds iff

$$\mu\{\mathbf{h} \in \mathbb{H} : \mathbb{E}[Y - \langle \beta, \mathbf{X} \rangle | \langle \mathbf{X}, \mathbf{h} \rangle] = 0 \text{ a.s. } \} = 1.$$

In other words, if β weew known, the test of (1) would be equivalent to select randomly just one element $\mathbf{h} \in \mathbb{H}$ and test if $\mathbb{E}[Y - \langle \beta, \mathbf{X} \rangle | \langle \mathbf{X}, \mathbf{h} \rangle] = 0$ a.s. is satisfied. Thus, we would only need to compute one conditional expectation given a real r.v.

The only reamining problems is to find a good estimator of β. Here we use the estimator based on a Principal Components decomposition proposed in [1].

As stated from a theoretical point of view, only a randomly chosen conditional expectation is required. However, in practice, this implies some loss of power. To mitigate it, we propose to test several random projections and combine the obtained p-values using the False Discovery Rate to make the final decision.

This is joint work with Manuel Febrero-Bande and Eduardo García-Portugués.

References

[1] H. Cardot, A. Mas, and P. Sarda, *CLT in functional linear regression models*. Probab. Theory Relat. Fields **138**, 325–361 (2007).

[2] J.A. Cuesta-Albertos, R. Fraiman, and T. Ransford, *A sharp form of the Cramer-Wold theorem*. J. Theor. Probab. **20**, 201–209 (2007).