Supporting information for "A goodness-of-fit test for the functional linear model with functional response"

Eduardo García-Portugués[1,2], Javier Álvarez-Liébana[3],
Gonzalo Álvarez-Pérez[4,6], and Wenceslao González-Manteiga[5]

Abstract

This Supporting information contains an extra real data application.

Keywords: bootstrap, Cramér–von Mises statistic, functional data, goodness-of-fit, regularization.

The "Ontario dataset", constructed by the authors of Benatia et al. (2017), contains the hourly electricity consumption ($\mathcal{Y}$; measured in gigawatts) and smoothed temperature ($\mathcal{X}$; Celsius degrees) in the province of Ontario (Canada). More precisely, it features a set of $n = 368$ daily curves on 2010–2014, where only summer months are taken into account, while weekends and holidays are discarded (hence, the i-th datum is not necessarily consecutive in time to the $(i+1)$-th). The response is valued in $\mathbb{H}_2 = L^2([0, 24])$ and discretized in 25 equispaced grid points. Each temperature curve is valued in $\mathbb{H}_1 = L^2([-24, 48])$ and discretized in 73 equispaced grid points. The interval $[-24, 48]$ accounts for a 3-days window that is considered since the past and future temperatures of a given day may influence the demand of energy on that day. Thus, the response is also regressed on 24 past and future hours. The raw temperature records are smoothed by a local polynomial regression on a weighted average of the temperatures of 41 Ontarian cities, producing the smoothed temperature, finally shifted so its minimum is set to 0°.

[1]Department of Statistics, Carlos III University of Madrid (Spain).
[2]UC3M-Santander Big Data Institute, Carlos III University of Madrid (Spain).
[3]Department of Statistics and Operations Research and Mathematics Didactics, University of Oviedo (Spain).
[4]Department of Physics, University of Oviedo (Spain).
[5]Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela (Spain).
[6]Corresponding author. e-mail: gonzaloalvarez@uniovi.es.

We check whether there exists a linear relation in the Ontario dataset. This is inspired by the data application in Benatia et al. (2017), where a FLMFR featuring several seasonal dummies is considered. Therefore, testing the GoF of the "canonical" FLMFR allows to evaluate if a seasonal-free simplified model succeeds in describing the daily electricity consumption from the temperature alone. Based on the data-driven selection of $\tilde{p} = 7$ and $q = 4$, the PCvM test gave null p-value, rejecting emphatically the FLMFR. When testing for significance, the KMSZ, PSS, LZS, and PCvM tests clearly rejected with null p-values. Hence, a nontrivial and nonlinear functional relation between daily electricity consumption and the temperature is evidenced, and the seasonal-free version of Benatia et al. (2017)'s model is shown to be inadequate for modeling such relation.

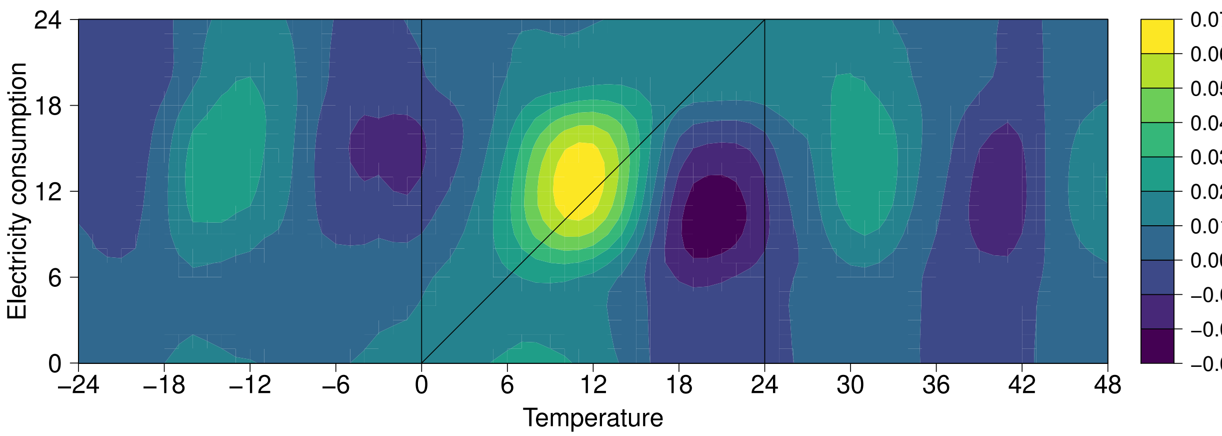

Figure 1: FPCR-L1S estimator $\hat{\beta}$ for the Ontario dataset. Note how $\hat{\beta}$ reflects the smoothness of the data, inherited by the FPC. The plot is coherent with Figure 11 in Benatia et al. (2017), yet ours is less centered at the diagonal, probably since no seasonal dummies were considered for fitting the FLMFR.

A referee and Associate Editor pointed out that the presence of temporal dependence in the data violates the iid assumption of our GoF test. Indeed, the data construction inherited from Benatia et al. (2017) employs 3-days overlapping windows that notably increases the serial dependency of the functional records. In order to investigate if this dependency was the leading rejection cause of the FLMFR, we have run our test retaining the 3-days windows but ensuring there are no overlaps in the observations. That is, we have considered only the curves for *day 1* (includes days 0, 1, 2), *day 4* (days 3, 4, 5), *day 7* (days 6, 7, 8), etc., properly handling weekends and holidays. The results are the same as in the original application: the FLMFR is emphatically rejected (null p-values) for the three possible subsettings of non-overlapping data and for different estimators. The no effect hypothesis is also rejected with null p-values. From this analysis, we are confident that the rejections with the original data are not primarily driven by temporal dependence (though still present in the non-overlapping data, e.g., by annual periodicity), and that a reduction in the complexity of the model in Benatia et al. (2017) through a seasonal-free version is not possible.

References

Benatia, D., Carrasco, M., and Florens, J. P. (2017). Functional linear regression with functional response. *J. Econometrics*, 201(2):269–291.