Supplementary Material: Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach.

October 12, 2015

SIMULATION STUDY

In this supplement we conduct a simulation study to evaluate the performance of the proposed method for functional principal component analysis (FPCA) of generalized quantiles. We run the simulation for independent as well as for autocorrelated functional observations to demonstrate robustness to temporal dependence. We follow the simulation setup of Guo et al. (2015) and Tran et al. (2014), who both suggest alternative approaches for modeling independent functional tail event curves.

The data $Y_{j,k}, j = 1, \ldots, T, k = 1, \ldots, N$ is simulated from the model

$$Y_{j,k} = \mu(t_j) + \alpha_{1,k} f_1(t_j) + \alpha_{2,k} f_2(t_j) + \varepsilon_{j,k} \tag{1}$$

where t_j are equidistant sampling points in $[0, 1]$ with $t_j = j/T$, $\mu(t) = 1 + t + exp\{-(t-0.6)^2/0.05\}$ is the mean function, $f_1(t) = \sqrt{2}\sin(2\pi t)$ and $f_2(t) = \sqrt{2}\cos(2\pi t)$ are the principal component functions and $\alpha_{1,k}$ and $\alpha_{2,k}$ are principal component scores. The principal component scores are generated either (1) independently from a $\mathrm{N}(0, 36)$ and $\mathrm{N}(0, 9)$ distribution, respectively or (2) from a VAR(1) process with

$$\Phi_1 = \begin{pmatrix} -0.5 & -0.2 \\ 0.2 & 0.5 \end{pmatrix}.$$

The error $\varepsilon_{j,k}$ is generated from three different distributions as specified in Table (1), where the first one is a light-tailed distribution, the second one is heavy-tailed and the third one exhibits heteroscedasticity. The simulation is run 200 times with two different setups of sample sizes: $T = 100$ data points per curve and $N = 20$ curves and $T = 150$ data points per curve and $N = 50$ curves. We evaluate the estimates of the $\tau = 0.05$ and $\tau = 0.95$ expectile curves based on the mean squared error (MSE). The MSE of the k-th τ-expectile curve is computed as

$$MSE_k = \frac{1}{T}\sum_{j=1}^{T}\{l_{\tau,k}(t_j) - \widehat{l}_{\tau,k}(t_j)\}^2, \tag{2}$$

where $l_{\tau,k}(t)$ denotes the theoretical expectile and $\widehat{l}_{\tau,k}(t)$ denotes the estimated expectile. Summary statistics of the mean squared errors and the average run time in seconds of the simulations are given in Table (1). The magnitude of the average MSE does not differ substantially between the independent and the autocorrelated case. This confirms that the quality of the proposed methodology is not sensitive to temporal dependence between functional observations. The methodology performs worst for the fat tailed distribution, but well handles heteroscedasticity. As a benchmark, we apply the methods proposed by Guo et al. (2015) and Tran et al. (2014) to the described simlation setup with independent functional observations. Guo et al. (2015) propose an estimation algorithm that jointly estimates a collection of generalized quantile curves. Tran et al. (2014) develop an analogue of PCA for generalized quantiles and propose three different estimation algorithms called BottomUp (BUP), TopDown (TD) and PrincipalExpectile (PEC) algorithm. For a detailed description of the algorithms we refer to their work. The simulation results of the four benchmark methods are given in Table (2). It can be seen that in terms of average MSE our methodology outperforms the benchmark methods for almost all specifications. Only for the fat-tailed distribution in combination with the large sample size ($T = 150, N = 50$) the method proposed by Guo et al. (2015) performs slightly better.

		$T = 100, N = 20$		$T = 150, N = 50$	
		(1)	(2)	(1)	(2)
		$\tau = 0.05$			
$\varepsilon \sim \mathrm{N}(0, 0.5)$	Mean	0.0433	0.0407	0.0259	0.0234
	SD	0.0285	0.0281	0.0177	0.0168
	AT	3.3900	3.7200	9.8000	10.3200
$\varepsilon \sim t(5)$	Mean	0.2447	0.2242	0.1480	0.1401
	SD	0.2508	0.2407	0.1644	0.1571
	AT	3.7100	4.6600	11.6200	12.0700
$\varepsilon \sim \mathrm{N}(0, \mu(t)0.5)$	Mean	0.0521	0.0518	0.0499	0.0501
	SD	0.0379	0.0354	0.0385	0.0393
	AT	3.2600	3.7700	9.8700	12.2000
		$\tau = 0.95$			
$\varepsilon \sim \mathrm{N}(0, 0.5)$	Mean	0.0448	0.0400	0.0254	0.0233
	SD	0.0295	0.0271	0.0174	0.0170
	AT	3.4400	3.7200	10.1300	11.0100
$\varepsilon \sim t(5)$	Mean	0.2444	0.2290	0.1465	0.1428
	SD	0.2565	0.2396	0.1590	0.1729
	AT	3.7500	4.2300	10.3700	12.2500
$\varepsilon \sim \mathrm{N}(0, \mu(t)0.5)$	Mean	0.0564	0.0518	0.0416	0.0500
	SD	0.0381	0.0340	0.0286	0.0389
	AT	3.5100	3.8200	10.2100	11.8000

Table 1: Mean and standard deviation (SD) of MSE and average run time in seconds (AT) based on 200 simulation runs for independent PC scores (1) and autocorrelated PC scores (2).

		$T = 100, N = 20$				$T = 150, N = 50$			
		Guo	BUP	TD	PEC	Guo	BUP	TD	PEC
					$\tau = 0.05$				
$\varepsilon \sim \mathrm{N}(0, 0.5)$	Mean	0.5800	0.3921	0.1566	0.1313	0.0749	0.2290	0.0704	0.0729
	SD	2.2901	0.2371	0.0121	0.0156	0.4637	0.1887	0.0039	0.0047
$\varepsilon \sim t(5)$	Mean	0.4513	1.1577	0.7842	0.3735	0.0993	0.5780	0.4379	0.1818
	SD	1.4959	0.5303	0.1575	0.0664	0.3444	0.2064	0.0911	0.0196
$\varepsilon \sim \mathrm{N}(0, \mu(t)0.5)$	Mean	0.5308	0.6576	0.2968	0.2292	0.0923	0.3241	0.1336	0.1204
	SD	1.7210	0.4211	0.0241	0.0337	0.3376	0.2894	0.0075	0.0094
					$\tau = 0.95$				
$\varepsilon \sim \mathrm{N}(0, 0.5)$	Mean	0.4753	0.3833	0.1567	0.1343	0.0760	0.2316	0.0707	0.0725
	SD	1.8573	0.2143	0.0129	0.0197	0.4298	0.1956	0.0043	0.0045
$\varepsilon \sim t(5)$	Mean	0.5104	1.1637	0.7643	0.3785	0.0988	0.5998	0.4408	0.1819
	SD	1.6699	0.5796	0.1135	0.0653	0.3294	0.4100	0.0835	0.0195
$\varepsilon \sim \mathrm{N}(0, \mu(t)0.5)$	Mean	0.5260	0.7514	0.3713	0.2806	0.0879	0.3247	0.1341	0.1625
	SD	1.6950	1.1349	1.0524	0.0378	0.3207	0.2617	0.0087	0.0103

Table 2: Mean and standard deviation (SD) of MSE based on 200 simulation runs for independent PC scores with four different benchmark methods.

References

Guo, M., Zhou, L., Härdle, W. K., and Huang, J. Z. (2015), "Functional Data Analysis of Generalized Regression Quantiles," *Statistics and Computing*, 25, 189–202.

Tran, N. M., Osipenko, M., and Härdle, W. K. (2014), "Principal Component Analysis in an Asymmetric Norm," *Humboldt University Berlin, CRC 649 Discussion Paper.*