"On the Evaluation of Multivariate Event Probability Predictions in Electricity Price Forecasting"

Probabilistic forecasts are of ever increasing importance in the electricity price forecasting literature. They seek to predict the full conditional distribution of electricity prices and thus any statistical evaluation procedure aims at assessing the statistical validity of such distribution forecasts. Yet, as optimal statistical properties do not necessarily translate into optimal properties of a prediction in the sense of the problem to which it serves as an input, other evaluation techniques linked to the addressed problem have been considered in the literature. From such a practical perspective it may often be sufficient to specify a statistical event of interest, predict the associated probability and evaluate the quality of the prediction. As relevant events often depend on the entire price path in electricity price forecasting, they constitute multivariate events.

We propose a framework to evaluate predictions of multivariate event probabilities in electricity price forecasting, addressing the issue of evaluating problem-specific predictions with a statistically sound approach. The considered framework is illustrated using several events in the context of German day-ahead electricity prices, which have been chosen to represent typical considerations arising in the daily operation and optimization of generation assets and portfolios. Specifically, we forecast the probability of a given number of consecutive hours exhibiting negative prices, the probability of a pumped storage power plant being profitable during the day as well as the probability of the base-peak spread surpassing a predefined threshold. To forecast the associated probabilities expert models for electricity prices from the literature are estimated using a rolling window approach and simulated using a multivariate normal distribution, a multivariate t-distribution and a bootstrap approach. Across the ensemble of simulated paths of future electricity prices, we calculate the relative frequency of the considered event and report it as the predicted probability.

The proposed evaluation framework allows for both the assessment of calibration of individual forecast models and also facilitates comparisons across different models. To assess the calibration of individual predictions the Seillier-Moiseiwitsch and Dawid calibration test for probability forecasts is considered. The Quadratic Probability Score in combination with the Diebold Mariano test is used to establish statistically significant differences in predictive ability between the considered model. To gain further insights about the sources of predictive deficiency the Murphy Decomposition of the Quadratic Probability Score is also examined.

We use well-established expert models from the electricity price forecasting literature to calculate probability forecasts and present an appropriate evaluation framework. We conclude that the proposed models accurately predict probabilities of multivariate events related to the complete path of electricity prices and can be evaluated using the proposed framework. It thus offers a statistically sound and easily applicable compromise between a statistical evaluation of the full conditional predictive distribution and other methodologies based on the operational problem in consideration.

- References –

Bradley, P. S., Bennett, K. P., & Demiriz, A. (2000). Constrained k-means clustering. Microsoft Research, Redmond, 1-8.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business \& Economic Statistics, 13(6), 253-263.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281-291.

Lahiri, K., & Wang, J. G. (2013). Evaluating probability forecasts for GDP declines using alternative methodologies. International Journal of Forecasting, 29(1), 175-190.

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and Forecasting, 8(2), 281-293.

Seillier-Moiseiwitsch, F., & Dawid, A. P. (1993). On testing the validity of sequential probability forecasts. Journal of the American Statistical Association, 88(421), 355-359.

Steffen, B., & Weber, C. (2016). Optimal operation of pumped-hydro storage plants with continuous time-varying power prices. European Journal of Operational Research, 252(1), 308-321.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. Organizational Behavior and Human Performance, 30,132-156.