Assignment3PEF_AY23_24
Group 14

Idda, Raffaeli, Riondato, Stillo

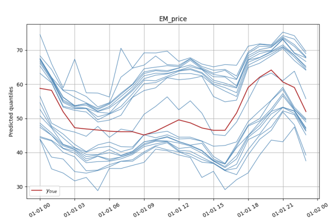
1 Implement Johnson's SU distributional DNN

Our aim was to implement the Johnson's SU model, which represents a specific type of deep neural network (DNN) model utilized for data prediction. This type of model can be useful when predicting data that are not well described by the normal distribution and require more flexible modeling of heavy tails or skewness. For this reason its implementation is very similar to that of normal but in addition skewness and tailweight are considered.

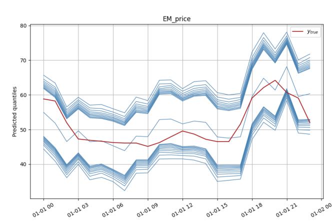
Note: To complete this point it was enough to slightly modify the code by adding few taylor-made lines for the Johnson'SU distribution.

2 Compare QR-DNN, Normal-DNN, JSU-DNN

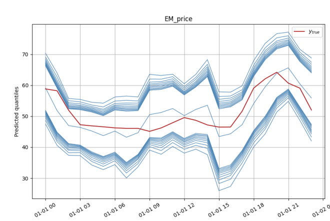
We were asked to compare deep neural network (DNN) models based on different distributions (QR, Normal, JSU) in terms of Pinball score. The goal is to obtain a lower Pinball Score, which indicates greater accuracy of the model in predicting data over a range of quantiles. We initially choose to concentrate on hour by hour one-day-ahead prediction (from 1-Jan-2017 to 2-Jan-2017) using the three distinct models. For each model, we trained it on a consistent time window spanning approximately two years and performed calibration by executing it 10 times. Below, we present graphs illustrating the predicted price hour by hour. In these graphs, the point prediction is highlighted in red, while various quantiles (ranging from 0% to 100% with intervals of 5%) are depicted in blue:



(a) QR EM_price

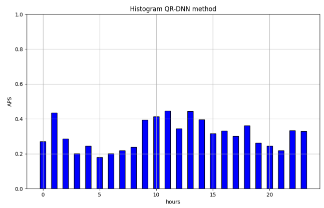


(b) Normal EM_price

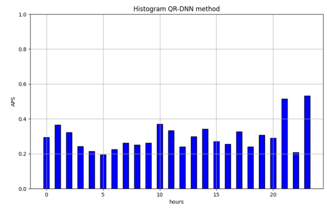


(c) JSU EM_price

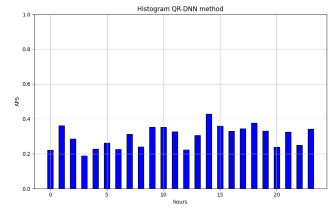
We chose to evaluate the performance of each model by calculating pinball scores for every quantile and prediction, yielding a (24,21) matrix. To provide a clearer interpretation of the results, we computed the Average Pinball Scores (APS) across the different quantiles. These APS values were then depicted in a histogram, showcasing 24 APS values, one for each hour of the predicted day. Finally, we computed the average of the APS values to obtain a comprehensive indicator of the models' quality.



(a) QR histogram



(b) Normal histogram



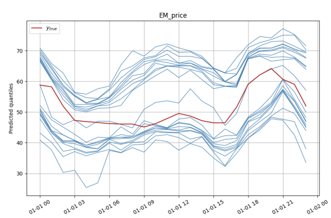
(c) JSU histogram

Low values are also preferable for APS because it indicates higher model accuracy.

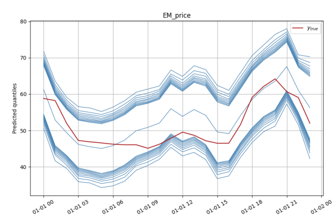
NOTE:
When we examined the Pinball Score graphs, we found it challenging to discern any noticeable improvement as the scores fluctuated significantly across the three approaches, simulation by simulation. However, upon closer inspection of the predicted prices value graphs, it became evident that the Quantile Regression (QR) method yielded notably poor estimations. Conversely, both the Normal and JSU approaches demonstrated substantial enhancements in predicted values. In particular, the JSU approach emerged as the most effective, evidenced by more accurate quantiles and median values compared to the other methods.

2.1 Random hyperparameter search

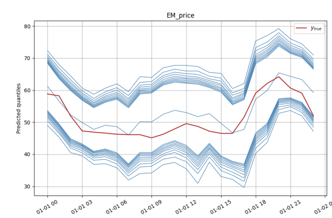
We performed the same procedure but changing the optuna model from grid_search to random:



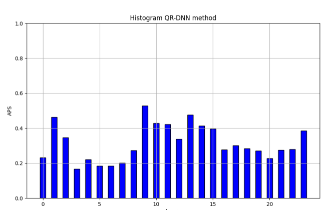
(a) QR EM_price



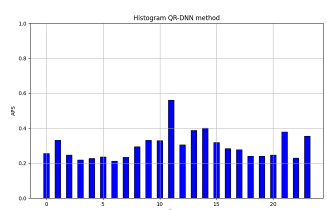
(b) Normal EM_price



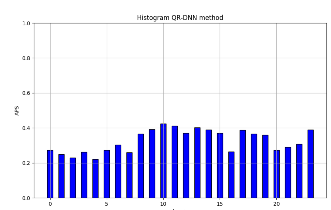
(c) JSU EM_price



(a) QR histogram



(b) Normal histogram

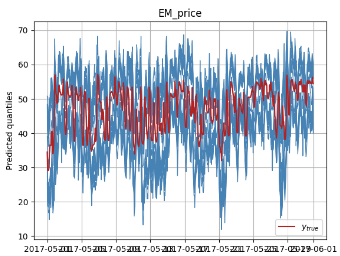


(c) JSU histogram

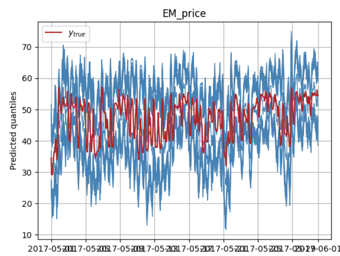
It is noticeable that switching to random parameters research does not reveal significant differences in the distribution and values of pinball scores. However, observing the distributions of prices and their quantiles, a slight improvement is noticed, particularly in the case of QR.

2.2 Test recalibration (1 time) on May 2017

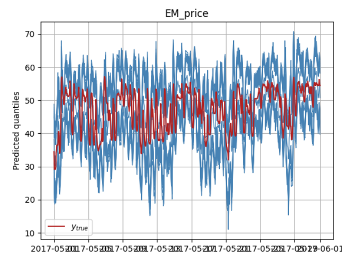
By changing the number of trials to 1 and increasing the prediction period considering all of May, we obtained the following results:



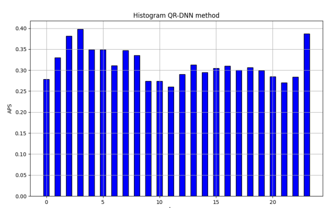
(a) QR EM_price



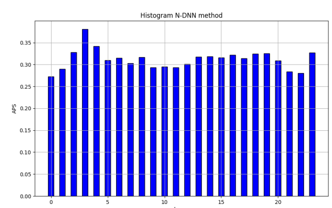
(b) Normal EM_price



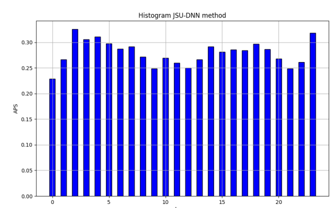
(c) JSU EM_price



(a) QR histogram



(b) Normal histogram



(c) JSU histogram

3 Performance Evaluation with Winkler score

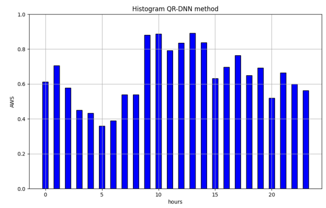
The Winkler's score assesses probabilistic forecasts formed as Prediction Intervals at discrete coverage levels $1 - \alpha$. This score is a proper scoring rule used to evaluate the accuracy of these intervals, and it is defined as

$$Winkler_n = \begin{cases} \delta_n, & \text{if } y_n \in [\hat{L}_n, \hat{U}_n] \\ \delta_n + \frac{2}{1-\alpha}(\hat{L}_n - y_n), & \text{if } y_n < \hat{L}_n \\ \delta_n + \frac{2}{1-\alpha}(y_n - \hat{U}_n), & \text{if } y_n > \hat{U}_n \end{cases}$$

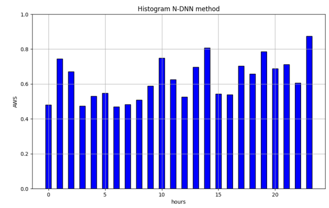
where $\delta_n = \hat{U}_n - \hat{L}_n$ is the width of the α-PI, with $\hat{L}_n, \hat{U}_n$ lower and upper bounds respectively.

The Winkler score rewards narrow prediction intervals, indicating higher confidence in the forecast when the observed value falls within the predicted interval. On the other hand, it penalizes the occurrence of test observations outside the predicted interval, with the penalty being proportional to the distance of the observation from the nearest bound of the interval.

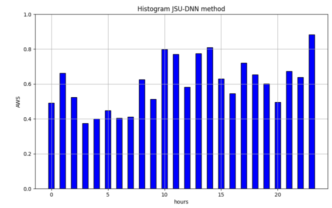
We decided to calculate AWS because it provides a summary measure of the overall goodness-of-fit of the model, taking into account all data points and all quantile levels.
Overall, higher scores indicate more accurate prediction intervals, while lower scores suggest that the intervals are less reliable.



(a) QR Winkler



(b) Normal Winkler

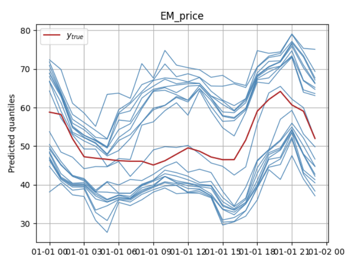


(c) JSU Winkler

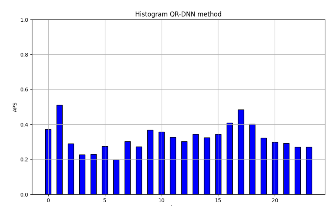
4 Preprocess the data by the arcsinh(.)

Applying the arcsinh function to the data can help ensure that the data are more evenly distributed and are less affected by extreme values, which can improve the performance of models.

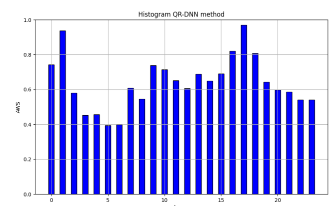
We applied the arcsinh function $\left(\sinh^{-1}\left(\frac{p-\xi}{\lambda}\right)\right)$, to the TARG__EM_price dataset column. Then, before the computation of the pinball score and Winkler's score, we retransformed the data by applying sinh to test_predictions.



(a) QR EM_price



(b) QR histogram



(c) QR Winkler

NOTE:
We attempted to implement pre-processing on market data considering standard rescaling factors ($\xi = 0$ and $\lambda = 1$) and also experimented with the mean and standard deviation of the market data as ξ and λ respectively. However, in both cases, we were not able to detect any kind of improvement in the results. In conclusion, it's possible that we misunderstood the given assignment or the implementation modatilities of the arcsinh pre-processing method.