WORKing papers in Management Science

WORMS/23/01

Trading on short-term path forecasts of intraday electricity prices. Part II – Distributional Deep Neural Networks

Grzegorz Marcjasz[1]
Tomasz Serafin[1]
Rafał Weron[1]

[1] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Poland

WORMS is a joint initiative of the Management Science departments of the Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

Trading on short-term path forecasts of intraday electricity prices. Part II – Distributional Deep Neural Networks

Grzegorz Marcjasz[a,1], Tomasz Serafin[a], Rafał Weron[a]

[a]*Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland*

Abstract

We propose a novel electricity price forecasting model tailored to intraday markets with continuous trading. It is based on distributional deep neural networks with Johnson SU distributed outputs. To demonstrate its usefulness, we introduce a realistic trading strategy for the economic evaluation of ensemble forecasts. Our approach takes into account forecast errors in wind generation for four German TSOs and uses the intraday market to resolve imbalances remaining after day-ahead bidding. We argue that the economic evaluation is crucial and provide evidence that the better performing methods in terms of statistical error metrics do not necessarily lead to higher trading profits.

Keywords: Intraday electricity market, Probabilistic forecast, Path forecast, Prediction bands, Trading strategy, Neural networks

1. Introduction

The European power trading landscape is undergoing significant changes as the generation from renewable energy sources (RES), such as wind and solar, continues to grow, accompanied by ongoing market integration and active demand-side management Grossi and Nan (2019); Maciejowska (2020). They make it more difficult to balance the supply and demand sides in the power system, mainly due to high uncertainty regarding the RES generation during the day-ahead (DA) auction. Therefore, we observe the shift towards shorter time horizons in electricity trading. The day-ahead market, which traditionally played a crucial role in electricity trading in Europe, is now slowly losing the market share to the intraday (ID) trading. Between years 2021 and 2022, the volume traded on European intraday markets (operated by EPEX) increased by 9%, while the day-ahead – decreased by 5% (EPEX, 2023).

This gradual change of focus is not yet visible in the electricity price forecasting (EPF) literature. The search of Scopus-indexed[1] articles reveals that only around 3% of EPF articles consider the topic of intraday electricity price forecasting.

Among the existing literature, researchers focus on few distinct topics. Kiesel and Paraschiv (2017) investigate the bidding behavior of German intraday electricity market participants and link the RES generation forecast errors to the electricity price changes. Narajewski and Ziel (2020a) and Marcjasz et al. (2020) focus on forecasting the ID3 index – the most commonly used proxy for the German intraday price (see Section 2). Janke and Steinke (2019) conduct a forecasting study with the focus on the quantiles of the price distribution for the last three hours of trading before the delivery. Linear regression models and an ensemble of neural networks are compared to several naive benchmarks. Narajewski and Ziel (2020b) and Serafin et al. (2022) propose ensemble forecasting methods for the continuous intraday markets, which in case of the latter paper, are used as a basis for a trading strategy which serves as a tool for the economic evaluation of electricity price forecasts. This particular direction is recently gaining attention of researchers and as Hong et al. (2020) and Maciejowska et al. (2023) argue, it is an important aspect of the model evaluation that at the same time is commonly overlooked in the literature.

In this paper, we address the aforementioned existing literature gap and extend the trading strategy proposed by Serafin et al. (2022) with a more realistic (from a perspective of a wind power plant owner) set of assumptions. More precisely, we consider wind generation forecast errors and use the intraday market to cover the imbalance left after the day-ahead bidding. Additionally we argue that the economic assessment of the forecast is the key factor in choosing the optimal approach from the perspective of the decision maker. Moreover, we propose a novel ensemble prediction model based on the well-performing machine learning approach of Marcjasz et al. (2023) and show that – albeit the results of the statistical evaluation are not unanimous – it is the best among tested methods in all trading simulations we performed.

The rest of this paper is structured as follows. In Section 2 we describe datasets used in this study. In Section 3 we provide the description of forecasting models while in Section 4 we introduce the "building blocks" that the models use – from point forecasting methods, through probabilistic and path trajectories to prediction band generation algorithm. In Section 5

*Corresponding author

Email address: grzegorz.marcjasz@pwr.edu.pl (Grzegorz Marcjasz)

[1] the EPF articles were queried using `TITLE-ABS-KEY(''electricity price*'' AND (''forecast*'' OR ''predict*'' ))` query, while the intraday EPF ones — `TITLE-ABS-KEY(''electricity price*'' AND (''forecast*'' OR ''predict*'' ) AND (''intraday'' OR ''intra-day''))`

we introduce trading strategies that are used for the economic evaluation of forecasts. Section 6 demonstrates the results of statistical and economic evaluation and provides a discussion on the applicability of both approaches. Lastly, Section 7 concludes the findings of this paper.

2. Data

2.1. Market description

Unlike the auction-based markets, the German intraday continuous trading does not have a single price for the product (i.e., for the delivery of a set amount of electricity over a given period). Instead, the price depends on the moment of entering the market – and as a result, we are presented with the price trajectory. The trajectory starts at 16:00 on the day preceding the delivery and ends 5 minutes before the delivery begins. The exchange lists three price indices: IDfull, ID1 and ID3, that are computed as a volume-weighted average price of transactions in the whole trading period, last hour before the delivery and last 3 hours before the delivery, respectively. While the indices are informative – they provide an approximation of the price via a single value – they do not present the whole information, especially regarding the trading opportunities.

We use the same dataset as Serafin et al. (2022) to facilitate the comparison. The dataset comprises the transaction data (price, volume and timestamp) for the hourly contracts on the German intraday electricity market covering period from 15.06.2017 to 29.09.2019. The first 364 days serve as an initial calibration window for the point forecasts, followed by three 91-day calibration periods: for the probabilistic forecast estimation based on the point ones (for the LASSO-based method), for the path forecasts and finally for the simultaneous coverage probability (see Section 5.3). This leaves a 200-day out-of-sample test period for the path forecasts. The data split is visualized in Figure 3 in Serafin et al. (2022).

The data contains the raw info for each of the executed transactions (timestamp, volume and price). To make the data better suited for modeing, we use an aggregated view of the market data. From the raw transaction data, we extract volume-weighted average prices (VWAPs) of the 15-minute timeframes that constitute the ID3 index, for a total of 12 subperiods. However, the first subperiod only considers 10 minutes of the data (in the modeling framework, we use information that is available 3 hours before the delivery, we allow 5 minutes for gathering the data and running the models) and the last two subperiods are ignored (as the last 30 minutes of trading is limited – only trades within the control zones are allowed). We therefore use 10 subperiods for evaluation of the strategy.

Having the VWAPs for the 10 subperiods $t_1, \dots, t_{10}$, we can use them as an approximation of the price trajectory – and as Serafin et al. (2022) state – it also is more realistic for selling larger volumes than the prices of single transactions.

2.2. Exogenous data

Aside from the market data, we also have exogenous series that are used in the model. Firstly, we have German wind power generation data – hourly values describing the forecasted (day-ahead forecast) the generation and the series of actual (observed) values. We assume that the actual data is available with only a small delay (such data is publicly available, e.g. on the ENTSO-E platform). Secondly, we have similar data regarding the forecasted and actual load for Germany. For the depiction of the exogenous data (the day-ahead forecasts) we refer the reader to Figure 4 in Serafin et al. (2022).

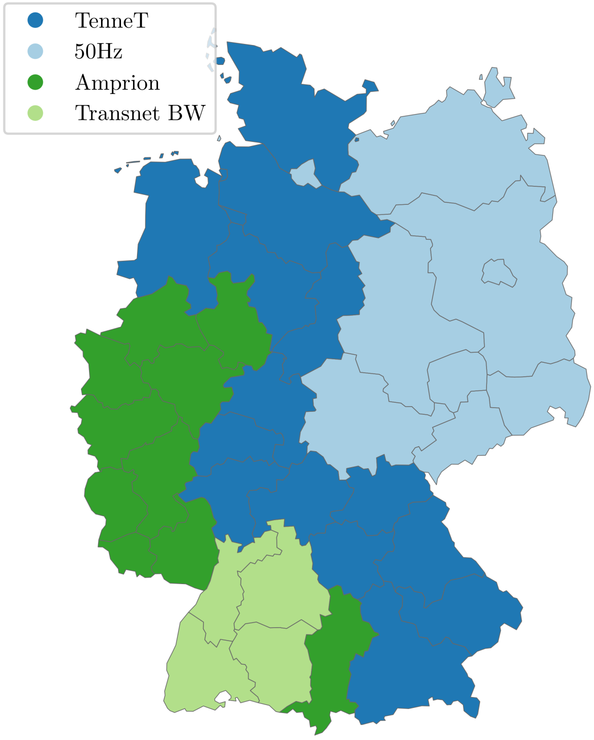

Figure 1: Map of Germany with the approximate geographic division to four zones covered by the service of respective TSOs.

2.3. Wind data for realistic strategy

The wind data described in the previous paragraph correspond to the nation-wide values. However, to better approximate what is the imbalance after the day-ahead bid and the update of the wind generation forecasts, we need to take the location of a power plant (as the wind gusts are not uniform over the whole country). We use the wind generation forecasts from the four German transmission system operators (TSOs): Amprion, 50Hertz, TenneT and Transnet-BW. For each zone, we have a set of two forecasts of the zonal wind generation, the day-ahead one (which is a basis for computing the volume sold on the day-ahead market) and the one closer to delivery (assumed to be equal to the generation), the difference of which needs to be purchased or sold on the intraday market (see Section 5.2.2).

3. Models

This section describes the three models we use for the generation of prediction bands – each comprises of various “building

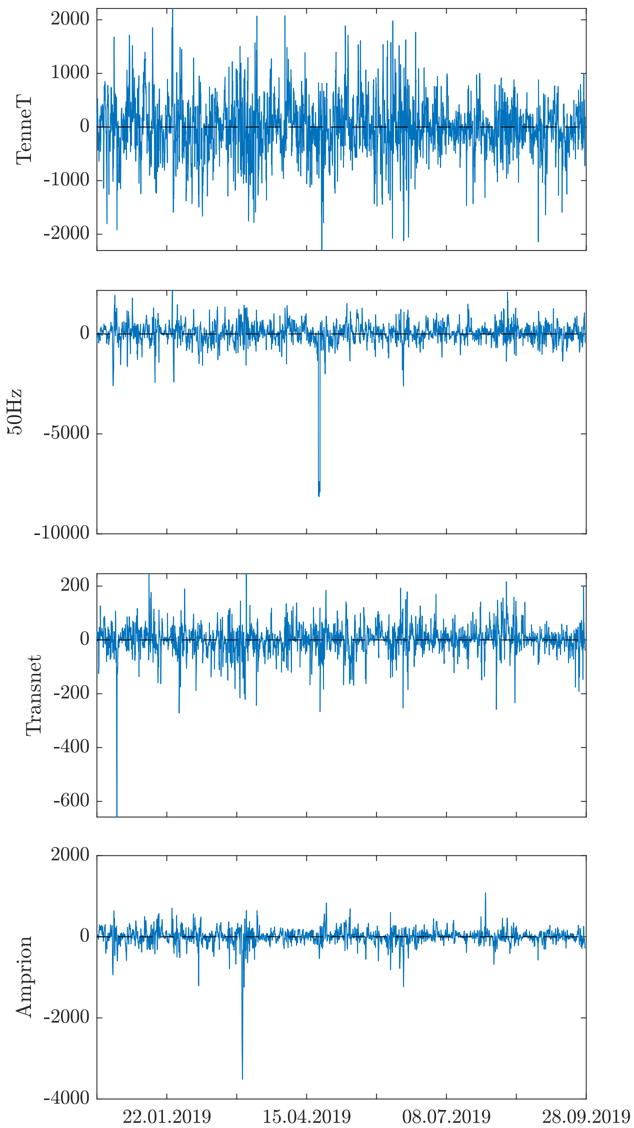

Figure 2: Hourly zonal imbalance plots for four German delivery zones.

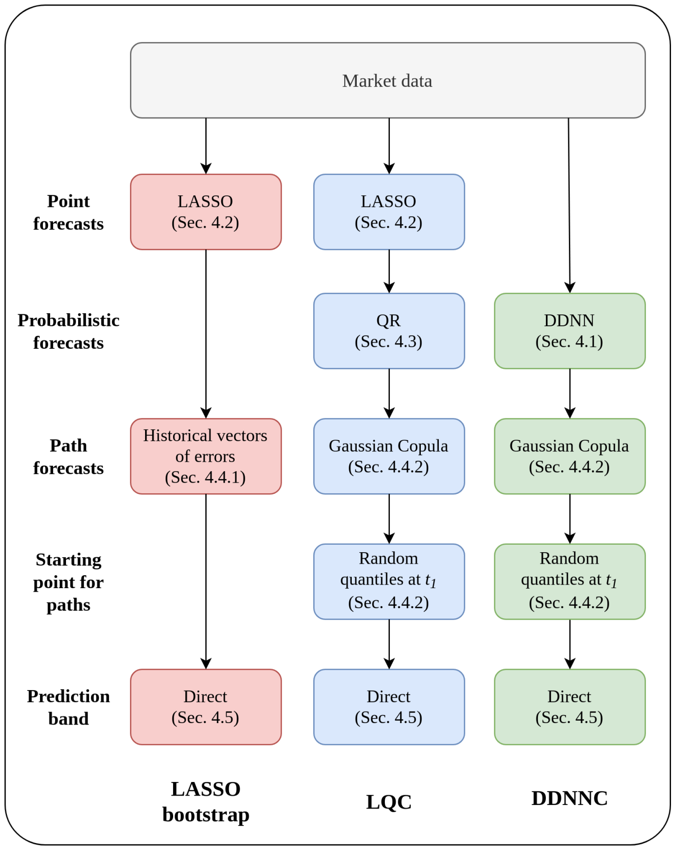

Figure 3: Flowchart presenting the "building blocks" of the forecasting approaches introduced in Section 3, based on the computational techniques described in Section 4.

blocks" (see Figure 3), however all of them use the same approach to obtain the prediction bands from the path forecasts – the Direct method described in Section 4.5. The respective steps are introduced in Section 4.

3.1. The DDNNC approach

The novel approach we introduce combines distributional neural networks and Gaussian copula-modeled temporal dependencies. As described in Section 4.1, the neural network outputs the probability density function directly – there is no intermediate point forecast created in the process. The steps for creating the path forecasts from the probabilistic one, choosing the starting point for the paths and prediction band computation are identical to the LQC method of Serafin et al. (2022). More specifically, first trajectory forecasts are computed, with temporal dependencies between the sub-periods modeled using a Gaussian copula. Next, vectors of innovations are affixed to randomly drawn values from the probabilistic forecast for t_1 sub-period and eventually, the Direct approach is used for computing the prediction bands.

3.2. The LQC approach

The so-called **LQC** approach proposed in Serafin et al. (2022) comprises three main parts: LASSO point forecasting model, quantile regression (to obtain probabilistic forecasts) and – like the DDNNC approach – copula-modeled structure of temporal dependencies. The QR is used to compute 99 percentile forecasts based on the point predictions. The 99 percentiles are linearly interpolated to obtain more granular quantiles, there is also extrapolation to the minimum and maximum prices for the extreme values.

3.3. LASSO bootstrap approach

Lastly, we use the better of the two point forecast-based methods described in Serafin et al. (2022): **LASSO bootstrap**. The approach uses LASSO point predictions as the base for obtaining price paths - it additionally samples vectors of historical point forecast errors to "correct" for the observed temporal dependency. This particular method proved to be an extremely well performing benchmark despite its simplicity.

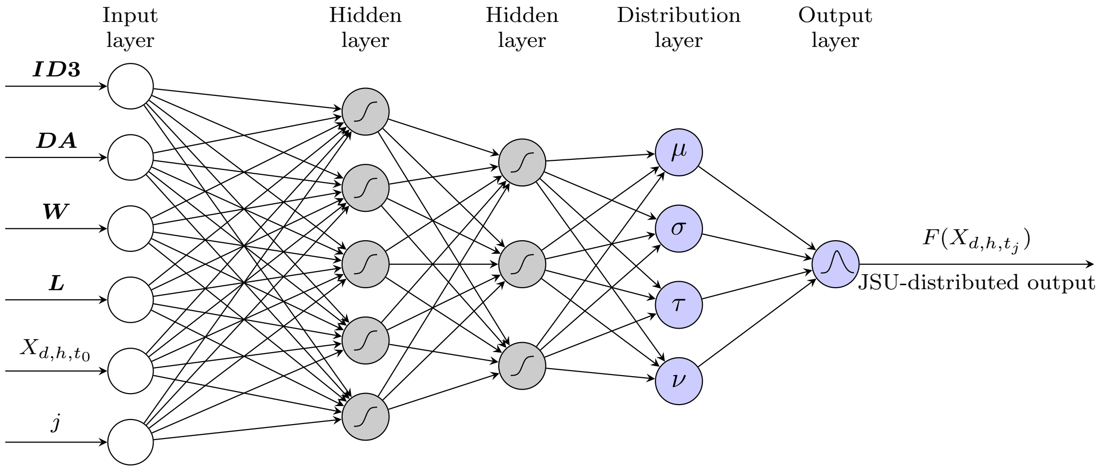

Figure 4: Visualization of the deep distributional neural network (DDNN) structure with the inputs similar to the LASSO defined in Eq. 1, and two hidden layers. The size and activation functions were chosen automatically in the hyper-parameter optimization study. For simplicity inputs in bold represent vectors of respective variables, i.e. $\boldsymbol{ID3}$ and $\boldsymbol{DA}$ represent lagged DA and ID3 values whereas $\boldsymbol{W}$ and $\boldsymbol{L}$ correspond to forecasted and actual values of load and wind generation used in in Eq. 1. Note that the neural network outputs the distribution of price for day d, hour h and j-th time period t_j, $j = 1 : 10$.

4. Methodology

4.1. DDNN

Distributional deep neural networks are feed-forward networks that – compared to their point counterparts – differ only slightly in the structure (see Fig. 4) and are trained to minimize the log-likelihood instead of error metrics such as mean absolute error (Marcjasz et al., 2023). The network used in this study is a deep structure with outputs that create (via a four-part parameter layer corresponding to the four distribution parameters) an output in form of a Johnson SU distribution (Johnson, 1949). The model for predicting the distributions of VWAP for a given j-th subperiod ($j = 1, ..., 12$) for delivery at day d, hour h consists of the following 102 inputs:

1. 21 past ID3 index values (newest avaiable at the time of forecasting)
2. 25 day-ahead prices, ranging from day $d-1$, hour h to day d, hour h
3. the day-ahead forecast of 25 hourly values of wind generation and day-ahead load forecast (day $d-1$, hour h to day d, hour h)
4. the actual wind power production and observed load for the last observed hour (4 hours preceding the delivery) and hour h of day $d-1$
5. last VWA price of the 15-minute interval (period spanning from 3h15m to 3h before the delivery)
6. a multi-valued indicator variable corresponding to the modeled subperiod j

Note, that we do not use dummies corresponding to the weekday or hour of the day in the model in a fashion similar to the day-ahead models, as this information is strongly correlated with the day-ahead prices and load forecasts. The model, however uses a dummy to mark the training samples coming from j-th subperiod. The original formulation of the model proposed by Serafin et al. (2022) did not have this information, as 12 separate models were trained, one for each future horizon (see Section 4.2). Based on a limited numerical study, for this particular deep neural network model it is beneficial to use only one model that has a vastly larger set of training samples. Moreover, unlike for the LASSO model (that follows the original formulation of Serafin et al. (2022)), the input data is not preprocessed in any way (aside from the batch normalization applied in the neural network). The non-linear model structure is expected to fit well to the non-linear patterns in the data (Hill et al., 1994; Jędrzejewski et al., 2022).

The calibration window used in the neural network training was 364 days ($24 \cdot 12 \cdot 364$ samples), 20% of which were randomly left out as an unseen data for the validation. Whenever the forecast error on the validation set did not improve in the last 50 iterations over the whole dataset, the training is assumed to be finished (and the weights from iteration that yields the lowest validation error are restored). The process is called early stopping and is a common practice in the literature (Lago et al., 2018; Yao et al., 2007).

Due to the randomness of the neural network training process, the results of consecutive runs (training processes – the trained network is deterministic) vary – for a more robust performance, it is a standard practice to train multiple neural networks (in this case, using the same data) and treat their combined ouptut as the final outcome of the model (here, horizontal (qAve) averaging was used, following Marcjasz et al. (2023)). The results reported in this paper correspond to the ensemble of 5 identical (w.r.t. the structure and input data) neural networks trained separetely, using the same hyper-parameter set (see paragraph below). Also, Fig. 7 presents the impact and variance of the profit from the trading strategy depending on the size of the ensemble.

The aforementioned hyper-parameter set was chosen in an additional optimization study, in which only the initial calibration data (the first 364 days of the dataset) was used to avoid an *ex-post* optimization. The last 13 weeks (91 days) of that were allocated as a hyper-parameter validation dataset. The process of hyper-parameter optimization, in simple terms, was an iterative training of neural networks using different hyper-parameter values and evaluating them on the hyper-parameter evaluation dataset. In more detail, for each hyper-parameter set, there were 7 neural networks trained – each was evaluated on only 13 of the 91 validation days. This was done to make the hyper-parameter optimization more robust, as it limits the impact of the randomness on the optimization process. The hyper-parameter sets were chosen using Tree-structured Parzen Estimator implemented in the optuna package for Python – so the consecutive trials were based on the history of values the tested so far. Finally, the network was trained five times independently (using a random starting points for the weights in the model) with the chosen hyper-parameter set – the predictons made using these five trained networks constitute an ensemble. During the hyper-parameter calibration, the following hyper-parameters were determined:

- activation function for both hidden layers, (independently) chosen from sigmoid, relu, elu, tanh, softplus and softmax
- the number of neurons in both hidden layers (independently) – an integer from 16 to 1024
- the initial learning rate for the ADAM algorithm (float ranging from 10^{-7} to 10^{-1})
- dropout application (yes/no) and, if yes, the dropout rate – float from 0.0 to 1.0

4.2. Generating LASSO point forecasts

The LASSO-estimated point forecasts were generated using the same model as in Serafin et al. (2022), with the baseline model for the VWA of hour h on day d is (for the j-th subperiod before the delivery) given by:

$$
\begin{aligned}
X_{d,h,t_j} = {} & \beta_0 + \underbrace{\sum_{i=4}^{24} \beta_{i-3} \mathrm{ID3}_{d,h-i} + \sum_{i=0}^{24} \beta_{22+i} \mathrm{DA}_{d,h-i}}_{\text{past ID3 and past/forward-looking DA prices}} \\
& + \underbrace{\sum_{i=0}^{24} \beta_{47+i} \widehat{W}_{d,h-i} + \beta_{72} W_{d,h-4} + \beta_{73} W_{d,h-24}}_{\text{wind generation forecasts and past values}} \\
& + \underbrace{\sum_{i=0}^{24} \beta_{74+i} \widehat{L}_{d,h-i} + \beta_{99} L_{d,h-4} + \beta_{100} L_{d,h-24}}_{\text{load forecasts and past values}} \\
& + \underbrace{\beta_{101} X_{d,h,t_0}}_{\text{last VWA price}} + \varepsilon_{d,h,t_j},
\end{aligned} \tag{1}
$$

where $\mathrm{ID3}_{d,h}$ denotes the value of the ID3 price index for day d and hour h, $\mathrm{DA}_{d,h}$ is the day-ahead price for day d and hour h, $\widehat{W}_{d,h}$ and $W_{d,h}$ are, respectively, the day-ahead predicted and actual wind generation for day d and hour h, $\widehat{L}_{d,h}$ and $L_{d,h}$ are the day-ahead predicted and actual system-wide load for day d and hour h, respectively, and X_{d,h,t_0} is the last known VWA price, i.e., the VWA price of transactions between 3 hours and 15 minutes and 3 hours before the delivery. The last regressor is widely used in the literature on forecasting the ID3 index prices (Marcjasz et al., 2020; Narajewski and Ziel, 2020a). Note, that for the sake of simplicity the notation $\mathrm{ID3}_{d,h-i}$ refers to the ID3 index value i hours before the day d and hour h even though the $h - i$ might be negative.

Note, that the inputs are identical to the ones used in the DDNN methdod (Section 4.1), with the omission of the variable indicating the modeled subperiod j. Instead, 12 separate models are constructed, one for each day d, hour h and subperiod j – although the inputs are exactly the same for each j, the modeled dependencies can be different, as LASSO method automatically limits the impact of less relevant input values, effectively creating 12 (possibly) different models (constructed as subsets of the baseline model) for each day and hour. As in the original paper, the calibration window had length of 364 days.

However, unlike in the DDNN model, the input data series undergo a variance stabilizing transformation, following Serafin et al. (2022) description. Each input series is independently normalized by subtracting the in-sample median and dividing by the in-sample median absolute deviation adjusted by the 75-th percentile of the standard normal distribution. Finally, the *area hyperbolic sine* is applied as the so-called variance stabilizing transformation (Uniejewski et al., 2018). This allows the data to be better suited for the linear model (and normalize the variances of all input series, which is beneficial for the LASSO method).

The model is estimated using the LASSO operator (Tibshirani, 1996), that implicitly (via the regularization of the model's coefficients) selects only the relevant inputs (note, that this results in a set of 24 hourly models that possibly use a different information set). The regularization parameter is in this study chosen automatically from a set of 50 values (that are automatically computed) through a cross validation procedure with 3 folds. The method is implemented in *scikit-learn* library for Python (Pedregosa et al., 2011).

4.3. Computing quantile forecasts using LASSO point estimates

Having the point LASSO forecasts as described in the previous Section, we use quantile regression with 91-day calibration window to generate an approximation of the probabilistic forecast. For each of the percentiles, we estimate its based on the previous forecast values and the actual values. Since we do it separately for each of the 10 sub-periods, we might observe a so-called quantile crossing (i.e., non-monotonic approximation of the percentiles), we prevent it by sorting the percentile estimates, as suggested by Maciejowska and Nowotarski (2016), Serafin et al. (2019) and Serafin et al. (2022).

4.4. Generating path forecasts

The study uses two different schemes of obtaining the path forecast (which are later used to construct the prediction bands in Section 4.5; note however, that in general the path forecast is not required for the prediction band to be generated (see e.g., AQL method in Serafin et al. (2022)). One of the approaches is based on the point forecasts (Section 4.4.1) and one uses the probabilistic forecast (Section 4.4.2). Both methods introduce time-dependency in the generated scenarios based on the historical forecasts and realized actual price paths.

4.4.1. *Historical point forecast errors*

The first method utilizes the point forecasts from the LASSO model as the base for generated scenarios. The time-dependency between prices in consecutive time points is introduced by adding a vector of past errors of the LASSO point model to the forecast for day d in the following way:

$$\widetilde{X}_{d,h,t_j} = \hat{X}_{d,h,t_j} + \varepsilon_{d^*,h,t_j},$$

where $\hat{X}_{d,h,t_j}$ is the point forecast obtained using LASSO, $\varepsilon_{d^*,h,t_j} = \hat{X}_{d^*,h,t_j} - X_{d^*,h,t_j}$ and d^* is a randomly selected day from the past 182 days.

4.4.2. *Gaussian Copula*

The second approach generates price scenarios based on the forecasted quantiles (from the probabilistic model – LASSO with QRA or DDNN) while the time-dependency is modeled with the Gaussian copula, similarly to Serafin et al. (2022). Using 91-day rolling calibration window, we estimate Σ – the temporal correlation matrix of transformed quantile coverage errors of probabilistic forecasts. Later, we simulate the trajectories by selecting quantiles in consecutive periods that are inter-correlated based on the estimated Σ. For the more detailed description see Pinson et al. (2009), Gneiting et al. (2007) and Janke and Steinke (2020).

4.5. *Determining prediction bands from path forecasts*

Following Serafin et al. (2022), we construct the prediction bands from the pool of simulated trajectories in order to later use them for the trading strategies (see Section 5). Prediction bands, unlike a set of prediction intervals, take into consideration the temporal dependency of the price forecast evolution in consecutive time points. Each prediction band (upper or lower) is characterized by the *simultaneous coverage probability* (SCP), which is the probability that the whole price trajectory lies below (for upper) or above (for lower) the band. Note that in the strategies we use for the economic evaluation of path forecasts, we make a decision of either selling or purchasing the electricity and therefore, at the time of the decision, only upper (for selling) or lower (for buying) prediction band is taken into consideration.

More formally, the SCP for the upper prediction band B^U_{d,h,t_j} can be written as:

$$\mathbb{P}\left(X_{d,h,t_j} \leq B^U_{d,h,t_j}, \forall_j\right) = \text{SCP},$$

while for the lower B^L_{d,h,t_j}:

$$\mathbb{P}\left(B^L_{d,h,t_j} \leq X_{d,h,t_j}, \forall_j\right) = \text{SCP}.$$

The algorithm we employ for the construction of the prediction bands is similar to the one proposed by Staszewska (2007). Since the simultaneous coverage property requires the price paths to respect the prediction band in each time point, the procedure comes down to rejecting the forecasted trajectories containing extreme points (maximum values for upper and minimum values for lower prediction band) from the whole simulated sample until SCP % of trajectories remain. Then, the prediction band is created by selecting the maximum (or minimum) values of the remaining paths at each consecutive time point. For the reference see both panels of Figure 5 – light-gray dotted lines represent rejected trajectories, dark-gray solid lines the remaining trajectories, while the solid red line depicts the derived prediction band.

4.6. *Evaluation of path forecasts*

The path forecasts in this paper are evaluated twofold: first based on the statistical measures and later in context of the economic measures. The statistical evaluation is the standard literature approach for the ranking the accuracy of forecasting methodologies (Hyndman and Koehler, 2006; Maciejowska and Nowotarski, 2016; Makridakis et al., 2018). However, the statistical evaluation might not always be straightforward. Lago et al. (2021) note that the relative accuracy of different models might change when we consider various error metrics and suggest to report multiple well-defined error measures, suitable for the type of data (e.g., in case of electricity prices that can have close to 0 or even negative values, percentage errors lead to incorrect conclusions). Therefore, in this paper we use three well-known scoring metrics suitable for the evaluation of path forecasts: Energy Score, Variogram Score and Dawid-Sebastiani Score (Scheuerer and Hamill, 2015). As Scheuerer and Hamill points out, each of these have its shortcomings in sensitivity to certain types of the forecast biases (see Sections 4.6.1–4.6.3).

Moreover, in practice a manager has to make one decision – and multiple sources (error measures) might point to different suggested actions (Kolassa, 2020). Moreover, the optimal choice should be determined by the expectations of the decision-maker (for example maximization of the profits or reduction of the risk). However, the statistical evaluation does not provide the necessary information since there is no clear relation between the error measures and the expected outcome of the decision (such as profit or VaR maximization), making it unclear if the accuracy of better methods (with regards to the statistical error metrics) corresponds in practice to improved financial results.

Hence, there is a need for a more universal evaluation, ideally one that addresses the aforementioned issues, for example a market simulation that uses the forecasts as an automatic decision support system (Janczura and Wójcik, 2022; Kath and Ziel, 2018; Maciejowska et al., 2019; Serafin et al., 2022; Uniejewski, 2023). In this paper, we propose a market simulation approach based on a simple trading strategy to determine whether the best forecast in terms of the statistical measures would be also a top performer in the context of economic evaluation from the perspective of the power producer.

4.6.1. Energy Score

The *energy score* is defined by Gneiting and Raftery (2007):

$$\mathrm{ES}_{d,h} = \frac{1}{M}\sum_{i=1}^{M}\left\|\widetilde{X}^i_{d,h} - X_{d,h}\right\|_2 - \frac{1}{M(M-1)}\sum_{i=1}^{M-1}\sum_{l=i+1}^{M}\left\|\widetilde{X}^i_{d,h} - \widetilde{X}^l_{d,h}\right\|_2, \quad (2)$$

where $\widetilde{X}^i_{d,h} = \left(\widetilde{X}^i_{d,h,t_1}, \ldots, \widetilde{X}^i_{d,h,t_{10}}\right)$ is the i-th path forecast for day d and hour h, $X_{d,h}$ is the corresponding actual VWA price path and M is the number of generated paths, see Section 4.4 for details. The energy score is a strictly proper scoring rule and a useful tool for evaluating forecasts, including ensemble forecasts, as it generalizes the continuous ranked probability score (CRPS; Hersbach (2000)). However, it has been observed that the energy score may lack sensitivity to misspecifications in the correlations between different components (Pinson and Girard, 2012; Pinson and Tastu, 2013).

4.6.2. Dawid-Sebastiani score

The *Dawid-Sebastiani score* – a multivariate scoring rule based on the mean vector and covariance matrix of the predictive distribution (Dawid and Sebastiani (1999)) – is defined by:

$$\mathrm{DSS}_{d,h} = \ln\left(\det\left(\mathbf{S}_{d,h}\right)\right) + \mathbf{K}^T\mathbf{S}^{-1}_{d,h}\mathbf{K} \quad (3)$$

where $\mathbf{K}_{d,h} = (K_{d,h,t_1}, \ldots, K_{d,h,t_{10}})$ is a vector of 10 differences, each taking the form of:

$$K_{d,h,t_j} = X_{d,h,t_j} - \frac{1}{M}\sum_{i=1}^{M}\widetilde{X}^i_{d,h,t_j}$$

and $\mathbf{S}_{d,h}$ is the covariance matrix estimated from the simulated scenarios. This scoring rule corresponds to the logarithmic score for multivariate Gaussian predictive distributions and remains a proper scoring rule for a broader class of probability distributions. However, Scheuerer and Hamill (2015) argue that the score calculation is very sensitive to the small sample size, hence it is not always a good choice for ensemble forecast evaluation (see e.g., Table 2 in Feldmann et al. (2015)). Note, that in case of the forecasting exercise considered in this paper, the ensemble size is large enough for the score to be applicable.

4.6.3. Variogram score

Lastly, we use the variogram score which has been proposed as an alternative proper scoring rule by Scheuerer and Hamill (2015). The *variogram score* of order p (VS-p) is defined by:

$$\mathrm{VS}_{d,h} = \sum_{i=1}^{10}\sum_{j=1}^{10} w_{i,j}\left(\left|X_{d,h,t_i} - X_{d,h,t_j}\right|^p - \frac{1}{M}\sum_{l=1}^{M}\left|\widetilde{X}^l_{d,h,t_i} - \widetilde{X}^l_{d,h,t_j}\right|^p\right)^2, \quad (4)$$

where $w_{i,j} = \frac{1}{100}$. This scoring rule has been shown to be more discriminative in context of misspecifications in the correlations structure of ensemble forecasts than two metrics described earlier. However, the types of biases and misspecifications are unknown in the forecasting task, and different values of the p parameter yield a scoring rule that is sensitive to different types of errors (for details see Scheuerer and Hamill (2015)). Therefore, the optimal value is not known in advance – we use two recommended values ($p \in \{0.5, 1\}$) here, and come to a completely different conclusions between the two. Not knowing the source of the errors, we are unable to discern a better model using the variogram score.

5. Trading strategies

In order to evaluate the path forecasts in terms of the economic results, we use and extend the prediction band-based trading strategy proposed by Serafin et al. (2022). The original approach assumes the position of energy producer that owns intermittent renewable energy sources or manages multiple such sources own by different entities (similarly to Li and Park (2018) or Kath et al. (2020)). It simulates a surplus of 1MWh of electricity sold in the intraday market each hour of the day. Our first extension is assuming that the decision maker, instead of excess generation, faces a deficit of 1MWh of electricity which has to be covered on the short-term market for each hour. This strategy provides a different view on the challenges posed by the renewable generation sources – and the combination of both sides, which is the second extension we propose, allow for a realistic evaluation of the daily operations of RES producer.

The third trading strategy mimics the actual uncertainty of the wind power generation (forecasted day before the delivery) and better relates to the challenges of the day-to-day operations faced by a RES producing company. We use the data from 4 German TSOs (see 1) that contain two wind generation forecasts: day ahead $\widehat{W}^{\mathrm{DA}}_{d,h}$ and intraday $\widehat{W}^{\mathrm{ID}}_{d,h}$ (see Section 2.3). We assume that the energy producer have an installed capacity of roughly $\omega = 1\%$ (for the Transnet-BW zone) or $\omega = 0.1\%$ (for the remaining three zones) of the total wind power capacity in the respective zone. Based on the forecasts, the decision-maker submits offers to sell $\omega\widehat{W}^{\mathrm{DA}}_{d,h}$ MWh of electricity on the day-ahead market and then has to balance his/her position on the intraday market based on the updated value of the generation forecast $\omega\widehat{W}^{\mathrm{ID}}_{d,h}$.

For each hour we compute:

$$\Delta_{d,h} = \omega\widehat{W}^{\mathrm{ID}}_{d,h} - \omega\widehat{W}^{\mathrm{DA}}_{d,h}. \quad (5)$$

$\left|\Delta_{d,h}\right|$ represents the volume that needs to be sold ($\Delta_{d,h} > 0$) or purchased ($\Delta_{d,h} < 0$) during the intraday market continuous trading.

In all cases, we assume that the impact of our trades on the prices on the intraday market is negligible and ignore the transaction costs. The problem can be then summarized as finding the optimal time to enter the market for each individual hourly delivery period.

5.1. Naive strategies

Following Serafin et al. (2022), use three naive strategies that are not based on any generated forecasts. In the first strategy, $\mathrm{Naive}_{\mathrm{first}}$, the market participant always enters the market during the first period t_0. The second strategy, $\mathrm{Naive}_{\mathrm{last}}$, involves taking the required position in the trading period closest to the delivery, t_{10}. The last ($\text{Naive}_{\text{avg}}$) strategy assumes that the total traded volume is split into 10 evenly-sized transactions throughout all periods $t_1 \ldots, t_{10}$.

5.2. Prediction band-based strategies

Having the prediction band generated based on the path forecasts, we use it as a time-varying price level of the recommended limit order (placed every 15 minutes). More precisely, the points from the prediction band (upper or lower depending on the trade direction) correspond to the prices of the limit orders (buy or sell) placed on the market in the consecutive 15-minute subperiods until one of the orders is filled. If none of the limit orders gets filled, we assume that the electricity is sold at the last VWA price, as in the $\text{Naive}_{\text{last}}$ strategy.

5.2.1. Fixed-volume sell/buy strategies

Serafin et al. (2022) introduces a novel strategy that uses prediction bands generated from forecasted price paths to support the decision-making of the company managing the renewable energy sources. It was assumed that the producer had to sell the excess of the electricity generation over the day-ahead bid, with a fixed volume order of 1MWh placed on the market each hour of each day. However, the stochastic RES generation might force the decision-maker to purchase the electricity instead. In this study, we address that and provide the results of not only always selling the electricity on the intraday market, but also always buying the same amount (since the optimal points of entry are different for both sides of the trade, the problems are similar, but with a different solution; see Fig. 5).

5.2.2. Realistic market simulation

This study, aside from considering separate perspectives of both the buyer and the seller, proposes a new, realistic market simulation, in which we assume the position of the decision-maker in a wind power plant. Each day, the manager offers 100% of the forecasted electricity generation for each hour of the next day (based on the day-ahead generation forecast). However closer to the delivery, a new, more precise forecast is available – and there will be a surplus or a shortage of electricity generated versus the day-ahead offer. Like in the fixed-volume strategies (see Section 5.2.1), the decision-maker needs to therefore balance it on the intraday market and the problem becomes an optimization of the time to enter the market. The main advantage of such an approach is getting rid of the unrealistic assumption that the balancing volume and direction are constant across all hours – here, we implicitly consider the correlation between the change of the wind generation forecast (day-ahead versus closer to the delivery) and the volume and direction of the balancing transaction. See Section 5 and Eqn. (5) for more details.

5.2.3. Ex-ante selection of the simultaneous coverage probability

As described in 4.5, we can derive a prediction band (upper or lower – depending on the trading direction) from a collection of path forecasts. However, we need to first specify SCP (simultaneous coverage probability), and its optimal value will vary – both in time and depending on whether we buy or sell. Following the methodology of Serafin et al. (2022), we leave out a 91-day long rolling calibration window to fit the optimal (most profitable) SCP. The selection is done independently for both the upper and the lower bands, based on the subset of hours for which the respective band type was used for trading (upper for selling and lower for buying). In order to confirm the validity of our approach, we compared the results of the automatic choice of SCP with the ex-post selected values for one of the German zones in Figure 6 – as can be seen, the automatic approach (red surface) yields profit very close to the optimal *ex-post* choices. Therefore, the results section of this paper will concentrate on the auto-SCP methods only.

5.3. Crystal-ball strategies

It is worth noting, that in the context of the proposed strategies there is a maximum and a minimum possible profit that can be extracted from trading activities. Therefore, we, following Serafin et al. (2022), introduce two additional reference strategies: $\text{Ref}_{\max}$ and $\text{Ref}_{\min}$, which always enter the market in the subperiods guaranteeing the best and the worst execution prices, respectively. These strategies can be treated as a baseline for the economic evaluation of other methods and additionally they provide a reference point. Given that the realistic market simulation will have different volumes traded for different zones, we can't compare the raw profits between them. Hence, we define the *fraction of realized trading potential* (FRTP) – a metric allowing for the explicit and qualitative comparison of the results, computed as follows.

$$\text{FRTP}_{\text{method}} = 100\% \cdot \frac{\text{Profit}_{\text{method}} - \text{Ref}_{\min}}{\text{Ref}_{\max} - \text{Ref}_{\min}}, \tag{6}$$

where $\text{Profit}_{\text{method}}$ corresponds to the sum of hourly profits of the trading strategy using the model's forecasts on the 200-day test period.

6. Results

As demonstrated in the literature (Kolassa (2020)), the selection of the "best" model (based on statistical evaluation) heavily relies on the choice of evaluation measure. Consequently, in the subsequent section, we will present the outcomes of both statistical and economic evaluation of the generated path forecasts. This approach aims to provide a comprehensive assessment of the forecasted data, taking into account not only statistical accuracy but also its relation to the economic performance.

6.1. Statistical measures

We evaluate the quality of forecasted price paths using three statistical measures especially suitable for this purpose: Energy Score (Gneiting and Raftery (2007), see Section 4.6.1), variogram score (Scheuerer and Hamill (2015), see Section 4.6.3) and Dawid-Sebastiani score (Dawid and Sebastiani (1999), see Section 4.6.2)). Results, calculated on the last 200 days

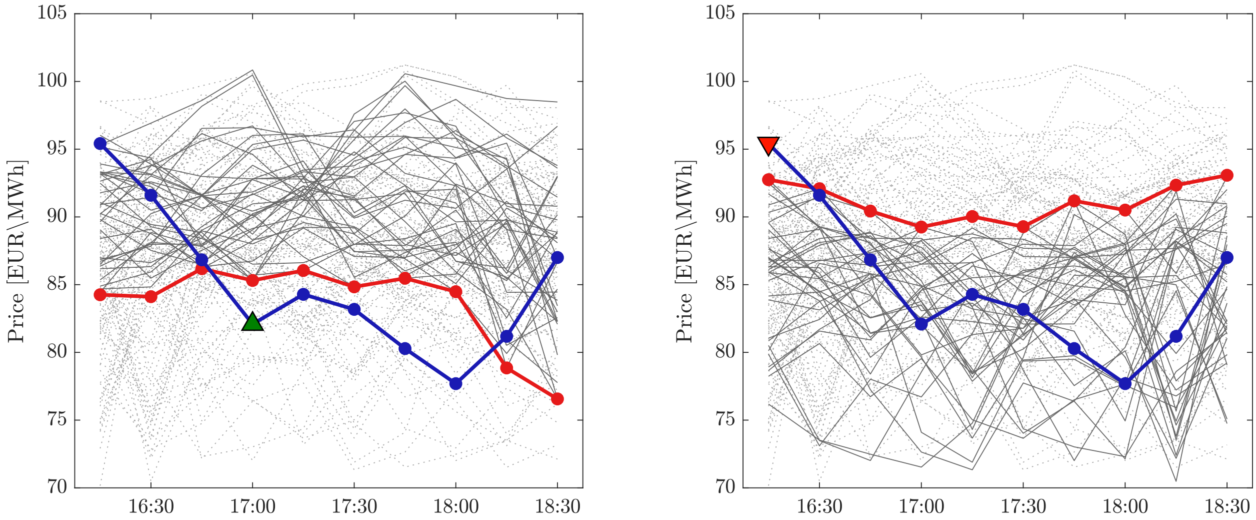

Figure 5: Exemplary trading situations based on lower (*left* panel) and upper (*right* panel) prediction bands with SCP 40%, derived from the same set of simulated trajectories. Green and red triangles mark the moments and prices of filled buy and sell orders, respectively.

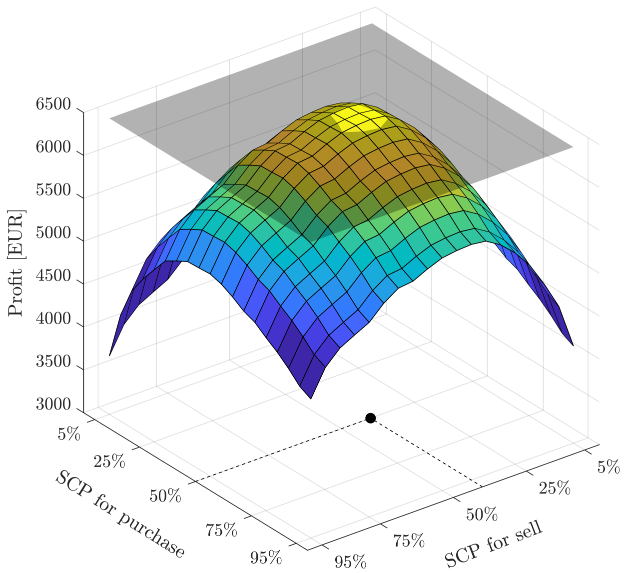

Figure 6: Mesh plot of profits from the realistic trading strategy (Section 5.2.2) for the TenneT zone, based on the trajectories from the DDNNC model. Parallel plane represents the profit from the automated selection of SCP.

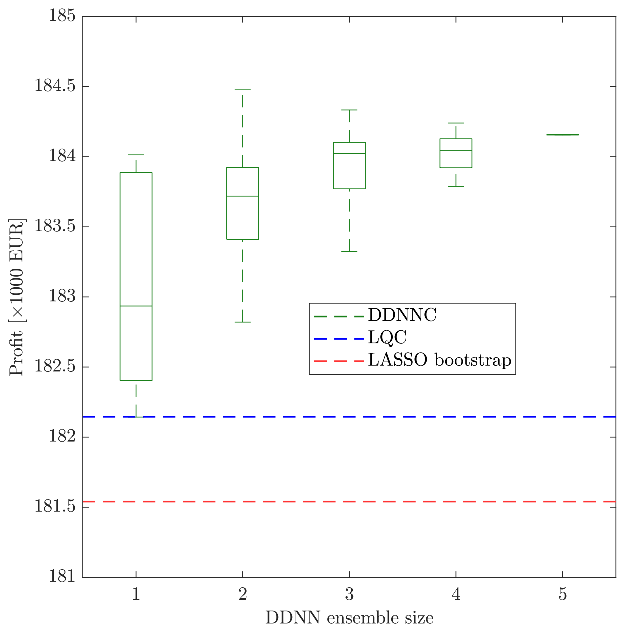

Figure 7: Profits from the fixed-volume selling strategy for different DDNN ensemble sizes. Boxplots were created based on all possible combination of a given number of forecasts from the pool of 5. For reference, solid lines correspond to LQC and LASSO bootstrap profits.

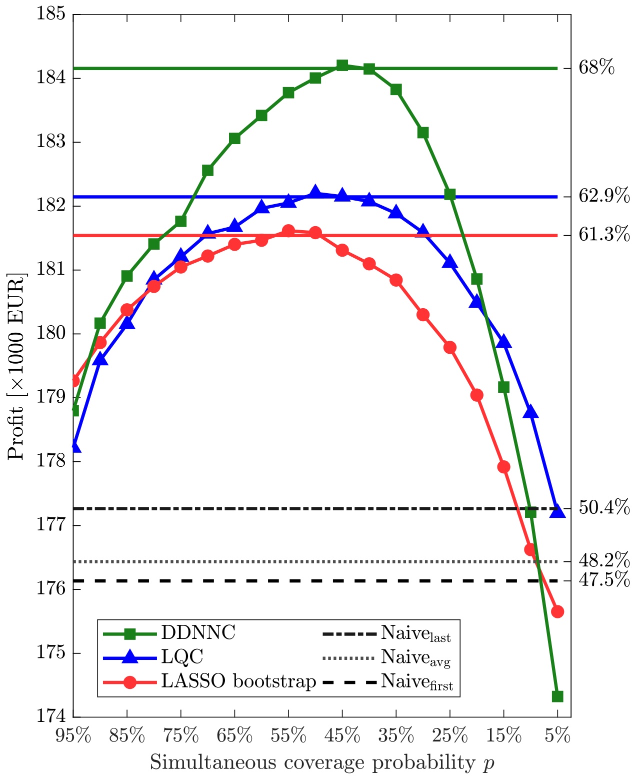

Figure 8: Profits from the fixed-volume selling strategy. On the left y-axis there is a nominal profit, whereas on the right – FRTP (see Section 5.3).

of the out-of-sample period, are presented in Table 1 and divided into peak (hours 8 to 19) and off-peak (remaining twelve hours of the day) periods. It is evident that various evaluation measures identify distinct top-performing models. Interestingly, in over half of the test cases the simplest considered approach (LASSO-bootstrap) exhibits the best performance – for both peak and off-peak test periods. The DDNNC approach outperforms its LASSO-QR-based counterpart with regards to almost every considered measure. The good performance of LASSO-bootstrap most probably stems from a completely different construction than two other models – it uses actual historical price evolution directly instead of estimating the dependency structure. Overall, it is difficult to ultimately pick the best model based on the presented results, the only clear conclusion is that DDNNC outperforms LQC. These results confirm that statistical evaluation of forecasts might not provide universal conclusions. In the next section, we present the results of the economic evaluation of path forecasts using trading strategies from Section 5.

6.2. *Trading profits*

Firstly, we will discuss the profits of the fixed-volume strategy (to provide a comparison to the original results published by Serafin et al. (2022)). In Figure 8 we can see the profits of a fixed-volume strategy that exclusively involves selling electricity on the market. Note, that this Figure is comparable with Fig. 10 of Serafin et al. (2022) – the Naive and LASSO-based methods are identical. The newly-proposed DDNNC model outperforms all remaining approaches, with the profit of the auto-SCP strategy higher than LQC by ca. 2000 EUR. This amount translates to approximately 5 percentage point improvement over the LQC model in context of the maximum profit achievable from forecasting for this particular strategy. For the sake of clarity, we omit the corresponding plot for the strategy that involves buying electricity – in this particular case, the DDNNC also performs better than other approaches with the profit higher than LQC by ca. 1500 EUR. Interestingly, the LASSO bootstrap performs better than LQC in this case (by almost 1000 EUR). Note, that although the averaging of multiple DDNN probabilistic forecasts is crucial in achieving such results, Figure 7 shows that even in the worst-case scenario (i.e., using only a single realization with the lowest out-of-sample profit) in the fixed-volume sell strategy, DDNN performs comparably to the LQC approach.

Secondly, we will discuss the results of the novel realistic strategy. In Table 2, we present the minimum and maximum possible profits (Ref$_{min}$ and Ref$_{max}$, respectively; see Section 5.3) alongside the FRTP defined in Eqn. (6), which described the percentage of the maximum possible gain achieved by the respective model. Note, that all Naive strategies, in general achieve FRTP of ca. 50% – further emphasizing the viability of the benchmarking approach (the Naive methods can be compared to a coin toss). Here, the DDNNC outperforms other methods in every case, and LASSO-based approaches trade places for the second result. Moreover, the outperformance is significant – DDNNC is better than the second best approach by 1.6 to 4.4 percentage points, with FRTP ranging from 65.4% to 67.3%.

7. Conclusions

In this paper we addressed an existing literature gap regarding evaluation of the electricity price path forecasts for the German intraday market. Firstly, we have used multiple scoring rules for the statistical evaluation of simulated price trajectories. Secondly, we proposed an extension of the simple trading strategy of Serafin et al. (2022), as a more realistic way for the economic assessment of ensemble forecasts. We make an important argument, that statistical and economical evaluation might lead to contrary conclusions regarding the best-performing-model selection. Moreover, we argue that from the practical perspective of the decision maker, the latter approach provides a clear outlook on the performance of the proposed models.

Additionally, we propose a novel path forecasting methodology that uses deep distributional neural networks of Marcjasz et al. (2023) as a replacement for the point and probabilistic forecasting steps in the LQC approach. This machine learning approach performs the best among the considered methods, significantly improving upon the results of the LQC method with regards to almost every metric.

Our results show that for the German intraday market, three statistical evaluation metrics: energy score, Dawid-Sebastiani score and Variogram Score were unable to discern the best model unanimously – depending on the metric, LASSO bootstrap approach traded the first place with DDNNC. On the other

Table 1: The results of the statistical evaluation of trajectory forecasts. The best results in each category (metric and daytime) are marked with bold.

	ES		DSS		VS-0.5		VS-1	
	peak	off-peak	peak	off-peak	peak	off-peak	peak	off-peak
DDNNC	8.46	6.67	**32.93**	30.55	0.53	0.43	**25.47**	**15.35**
LQC	9.33	7.02	33.47	29.81	0.58	0.45	37.05	16.42
LASSO-bootstrap	**6.21**	**5.17**	42.20	**29.71**	**0.50**	**0.42**	37.72	28.04

Table 2: Profits of automated strategies for different zones. The bold results are the best in each row. The results from each model represents the FRTP (see Section 5.3)

	Ref_{min} [EUR]	Ref_{max} [EUR]	**DDNNC** [%]	**LQC** [%]	**LASSOb** [%]	$\textbf{Naive}_{\text{first}}$ [%]	$\textbf{Naive}_{\text{last}}$ [%]	$\textbf{Naive}_{\text{avg}}$ [%]
TenneT	-2110	10685	**66.9**	62.3	63.8	49.9	49.7	50.2
. 50Hz	-7824	4906	**67.3**	63.3	61.6	48.1	51.4	49.8
Transnet	-9834	1952	**65.4**	62.2	63.8	49.3	50.1	49.6
Amprion	-3593	200	**65.6**	61.2	60.5	53.6	45.4	50.7

hand, a market simulation (in both the simpler and the more realistic form) always favored DDNNC in our testing – this results holds both for the whole Germany and also the four zones that we used for the evaluation. This implicates that the DDNNC model outperforms the other approaches in the context of easily quantifiable (and universal) economic measures.

Moreover, we further justify the attractiveness of the economic evaluation framework of Serafin et al. (2022) and its applicability to more realistic trading simulations. It provides a well-defined measure of the potential economic impact of forecast quality improvement (as we do know the minimum and maximum possible profits) and all naive methods are comparable to a coin toss – they achieve ca. 50% of the FRTP.

Acknowledgments

This work was partially supported by the Ministry of Science and Higher Education (MNiSW, Poland) through Diamond Grants No. 0009/DIA/2020/49 (to T.S.) and no. 0219/DIA/2019/48 (to G.M.) and the National Science Center (NCN, Poland) through grant No. 2018/30/A/HS4/00444 (to R.W.).

Author contributions

T.S., R.W. – Conceptualization; G.M., T.S. – Investigation; G.M., T.S. – Software; R.W. – Validation; T.S. – Writing, original draft; R.W. – Writing, review & editing.

References

Dawid, A.P., Sebastiani, P., 1999. Coherent dispersion criteria for optimal experimental design. Annals of Statistics , 65–81.

ENTSO-E, 2022. Entso-e transparency platform. `https://transparency.entsoe.eu`. Accessed: 2023-02-27.

EPEX, 2023. Annual report 2022. `https://www.eex.com/fileadmin/Global/News/Group/News/20230124_EEX_Group_Annual_Volume_Report.pdf`. Date accessed: 03.08.2023.

Feldmann, K., Scheuerer, M., Thorarinsdottir, T.L., 2015. Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous gaussian regression. Monthly Weather Review 143, 955–971.

Gneiting, T., Balabdaoui, F., Raftery, A., 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society B 69, 243–268.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Grossi, L., Nan, F., 2019. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. Technological Forecasting and Social Change 141, 305–318.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15, 559–570.

Hill, T., Marquez, L., O'Connor, M., Remus, W., 1994. Artificial neural network models for forecasting and decision making. International Journal of Forecasting 10, 5–15.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. IEEE Open Access Journal of Power and Energy 7, 376–388.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International journal of forecasting 22, 679–688.

Janczura, J., Wójcik, E., 2022. Dynamic short-term risk management strategies for the choice of electricity market based on probabilistic forecasts of profit and risk measures. the german and the polish market case study. Energy Economics 110, 106015.

Janke, T., Steinke, F., 2019. Forecasting the price distribution of continuous intraday electricity trading. Energies 12, 4262.

Janke, T., Steinke, F., 2020. Probabilistic multivariate electricity price forecasting using implicit generative ensemble post-processing, in: Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems – PMAPS 2020, p. 9183687.

Jędrzejewski, A., Lago, J., Marcjasz, G., Weron, R., 2022. Electricity price forecasting: The dawn of machine learning. IEEE Power and Energy Magazine 20, 24–31.

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36, 149–176.

Kath, C., Nitka, W., Serafin, T., Weron, T., Zaleski, P., Weron, R., 2020. Balancing generation from renewable energy sources: Profitability of an energy trader. Energies 13, 205.

Kath, C., Ziel, F., 2018. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. Energy Economics 76, 411–423.

Kiesel, R., Paraschiv, F., 2017. Econometric analysis of 15-minute intraday electricity prices. Energy Economics 64, 77–90.

Kolassa, S., 2020. Why the "best" point forecast depends on the error or accuracy measure. International Journal of Forecasting 36, 208–211.

Lago, J., De Ridder, F., De Schutter, B., 2018. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy 221, 386–405.

Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy 293, 116983.

Li, S., Park, C.S., 2018. Wind power bidding strategy in the short-term elec-

tricity market. Energy Economics 75, 336–344.
Maciejowska, K., 2020. Assessing the impact of renewable energy sources on the electricity price level and variability – a quantile regression approach. Energy Economics 85, 104532.
Maciejowska, K., Nitka, W., Weron, T., 2019. Day-ahead vs. intraday – forecasting the price spread to maximize economic benefits. Energies 12, 631.
Maciejowska, K., Nowotarski, J., 2016. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. International Journal of Forecasting 32, 1051–1056.
Maciejowska, K., Uniejewski, B., Weron, R., 2023. Forecasting electricity prices. URL: https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-667, doi:10.1093/acrefore/9780190625979.013.667.
Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2018. The m4 competition: Results, findings, conclusion and way forward. International Journal of Forecasting 34, 802–808.
Marcjasz, G., Narajewski, M., Weron, R., Ziel, F., 2023. Distributional neural networks for electricity price forecasting. Energy Economics , 106843URL: https://www.sciencedirect.com/science/article/pii/S0140988323003419, doi:https://doi.org/10.1016/j.eneco.2023.106843.
Marcjasz, G., Uniejewski, B., Weron, R., 2020. Beating the naïve – combining LASSO with naïve intraday electricity price forecasts. Energies 13, 1667.
Narajewski, M., Ziel, F., 2020a. Econometric modelling and forecasting of intraday electricity prices. Journal of Commodity Markets 19, 100107.
Narajewski, M., Ziel, F., 2020b. Ensemble forecasting for intraday electricity prices: Simulating trajectories. Applied Energy 279, 115801.
Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.
Pinson, P., Girard, R., 2012. Evaluating the quality of scenarios of short-term wind power generation. Applied Energy 96, 12–20.
Pinson, P., Madsen, H., Nielsen, H.A., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12, 51–62.
Pinson, P., Tastu, J., 2013. Discrimination ability of the energy score. DTU Informatics.
Scheuerer, M., Hamill, T.M., 2015. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Monthly Weather Review 143, 1321–1334.
Serafin, T., Marcjasz, G., Weron, R., 2022. Trading on short-term path forecasts of intraday electricity prices. Energy Economics 112, 106125.
Serafin, T., Uniejewski, B., Weron, R., 2019. Averaging predictive distributions across calibration windows for day-ahead electricity price forecasting. Energies 12, 256.
Staszewska, A., 2007. Representing uncertainty about response paths: The use of heuristic optimisation methods. Computational Statistics & Data Analysis 52, 121–132.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B 58, 267–288.
Uniejewski, B., 2023. Smoothing quantile regression averaging: A new approach to probabilistic forecasting of electricity prices arXiv:2302.00411.
Uniejewski, B., Weron, R., Ziel, F., 2018. Variance stabilizing transformations for electricity spot price forecasting. IEEE Transactions on Power Systems 33, 2219–2229.
Yao, Y., Rosasco, L., Caponnetto, A., 2007. On early stopping in gradient descent learning. Constructive Approximation 26, 289–315.