**Vol. 36, No. 2 (2026) | DOI: 10.37190/ord/217347**

OPEN ACCESS

**Operations Research and Decisions**

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

# Arbitrage in short-term electricity markets: Economic value of multidimensional ensemble forecasts

Katarzyna Maciejowska*[1] Weronika Nitka[1]

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

* *Corresponding author, email address: katarzyna.maciejowska@pwr.edu.pl*

**Abstract**

Generated quantity and market prices are major risk sources for electricity market participants, especially renewable producers with intermittent generation. Before delivery, wind and solar power plants must decide their day-ahead bids, which, through deviations from actual generation, determine the direction and volume of intraday trade. We identify an arbitrage opportunity between these short-term markets that increases expected profits and reduces trading risk. To exploit it, we propose a novel *multiple-split* approach for estimating multidimensional probabilistic forecasts of market fundamentals. This semi-parametric method produces ensembles preserving true error correlations and supports forecasting of linear and nonlinear functions such as price spreads, residual load, or trading income. Using German electricity market data, we demonstrate superior predictive and economic performance over standard benchmarks.

**Keywords:** *electricity market, trading strategy, probabilistic forecasting, ensemble forecasting, multivariate forecasting*

## 1. Introduction

The growing share of intermittent energy sources in the energy mix, together with the introduction of new short-term electricity markets, for example the intraday (ID) market, has increased the complexity of trading decisions faced by generation utilities. Because of the sequential structure of electricity markets, bids placed at one stage affect the range of options available for subsequent trading. Consequently, for renewable energy sources (RES) any unanticipated change in wind or solar generation – often driven by volatile weather conditions – requires adjustments in the ID market. On one hand, these dynamics make financial outcomes more difficult to predict, complicating the comparison of trading strategies and increasing the risk associated with decision-making. On the other hand, new market mechanisms create opportunities for portfolio optimization [11]. In short-term markets, utilities can earn additional profits by strategically allocating their production between the DA and ID markets.

Received 3 November 2025, accepted 25 January 2026, published online 25 January 2026
ISSN 2391-6060 (Online)/© 2026 Authors

The costs of publishing this issue have been co-financed by the Department of Operations Research and Business Intelligence at the Faculty of Management, Wrocław University of Science and Technology, Wrocław, Poland

The existing literature analyzes arbitrage primarily from the perspective of submitting a fixed-volume bid, typically 1 MWh [8, 20]. In such cases, financial gains depend mainly on the accuracy of price forecasts. However, this approach neglects the uncertainty surrounding generation levels, which is particularly relevant for RES. Unlike conventional power plants, wind and solar generators cannot precisely predict their future output. As a result, forecast errors in generation require real-time market participation to balance positions, introducing additional risk. Rational decision-making for DA offers therefore requires considering not only expected price spreads but also potential generation forecast errors.

Only a limited number of studies examine market selection and portfolio construction from the perspective of RES utilities [see 11, 19]. Their findings indicate that balancing trading activities across sequential markets helps reduce risk and stabilize income. Nevertheless, two important research gaps remain. First, the literature rarely considers generation reduction or full curtailment, leaving utilities exposed to financial losses during periods of negative electricity prices [29]. As such price events become increasingly frequent, this issue gains particular relevance for small and medium-sized RES producers. Second, the impact of a decision maker's risk appetite on portfolio construction has been addressed only marginally, typically through a limited set of trading schemes [19].

In this study, we aim to develop and evaluate a trading framework that addresses the above challenges. We consider a medium-sized RES utility that decides on the amount of electricity to offer in the day-ahead market on the day preceding delivery. The utility is assumed not to speculate but may withhold a fraction of its predicted generation to be sold later in the intraday market, closer to the time of physical delivery. The decision is based on the predicted distribution of income and reflects multiple sources of uncertainty arising from imperfect knowledge of both prices and generation levels. Depending on its risk attitude, the utility may choose to participate in the market or to curtail generation. In the latter case, production may be halted or redirected – for example, to charge a battery for later use – which remains beyond the scope of this study.

To meet these requirements, the proposed trading strategy relies on multidimensional probabilistic forecasts of electricity prices and renewable generation. Unlike point forecasts, probabilistic predictions approximate the entire conditional distribution, thereby allowing for assessment of both the expected value and the associated uncertainty of the forecasted variable. Although more computationally demanding, probabilistic forecasting has become increasingly popular in the electricity market literature [for a comprehensive review, see 24].

Three general approaches dominate the construction of probabilistic forecasts. The first is post-processing of point predictions, typically using methods such as quantile regression averaging [23, 33]. The second relies on the analysis of forecast errors, as in historical simulations or conformal prediction frameworks [9]. The third involves direct modeling of the data distribution, for example through GARCH-type models [3, 8] or distributional neural networks [21]. These approaches characterize the distribution of future observations using different representations. As noted in [7], the distinction between forecast types is, to some extent, artificial. For example, quantiles can be derived from an ensemble, and ensembles can be generated via sufficiently fine approximations of an underlying distribution.

Whether in the context of point or probabilistic forecasting, most existing techniques are designed to predict a single random variable. However, in many practical applications, the dependencies between variables are equally important and should be carefully modeled [7, 26]. For instance, a wind farm operator can explore the correlation between wind generation and electricity price forecast errors to improve trading performance [15]. An unexpected increase in renewable generation typically leads to a decrease in both DA and ID prices, and this relationship should be incorporated into the offer curve. [32] demonstrate how a multidimensional forecast of non-shiftable load, renewable generation, and electricity prices can enhance the management of a virtual power plant (VPP). Similarly, joint forecasting of electricity prices can support the design of trading strategies that optimize offers across multiple market segments, as shown in [12] and [19].

In this study, we propose a novel *multiple split* (MS) approach to estimate ensemble forecasts of electricity prices and market fundamentals that approximate the multidimensional distribution of considered variables. The *multiple split* approach is inspired by resampling-based probabilistic forecasting methods, notably [16] (*multiple split conformal* predictions) and [2] (*jackknife+*). Similarly to [5], we use univariate models to predict the expected values of individual variables and a multidimensional distribution of forecast errors to calculate probabilistic forecasts. In our approach, the sample is randomly divided into disjoint sets [1]. The subsets are then used to estimate the parameters, calculate the out-of-sample point prediction, and the forecast errors. The results are finally aggregated in the form of an ensemble of predictions. Since the split is repeated multiple times, the method captures both the uncertainty of the parameter estimates and the stochastic nature of the data.

The *multiple split* described in this work enhances the existing literature in various directions:

(i) It extends the analysis from a univariate to a multivariate framework. Due to a simultaneous division of the sample used for the prediction of all variables, the resulting forecast errors maintain the true correlation structure.

(ii) Unlike in previous articles, such as [9], random splitting is repeated multiple times to decrease the variability of the outcomes and reduce the dependence of the results on a particular division of the data. However, contrary to jackknife+ approach of [2], we do not consider all possible divisions of the original sample and hence reduce the computation time.

(iii) The results obtained by the splits are stored as an ensemble rather than the distributions (parametric or a set of quantiles) as in [16] or [9]. Therefore, the approach does not require the aggregation of the results obtained from individual splits with probabilistic prediction combination methods such as Bonferroni averaging [16]. Moreover, the ensemble forecasts can be easily used to generate a prediction of any linear or non-linear function of the original data. In such a case, a new ensemble is constructed by applying the function to the set of multidimensional predictions. This can be utilized to easily compute non-trivial risk measures, e.g. expected shortfall.

The proposed method is evaluated using data from the German electricity market. Forecast accuracy is measured through prediction interval coverage and the continuous ranked probability score (CRPS), and compared with established benchmarks such as Quantile Regression Machine and conformal predictions. The MS method generates prediction intervals with empirical coverage close to the nominal level and

[1]Several names have been assigned to approaches that are based on the division of the sample into disjoint subsets: split method, leave-$k$-out (LKO), $d$ delete jackknife, or cross-validation (CV) [2]. Although similar, the approaches differ in terms of applications and specifications.

**Table 1.** Evaluation window

| | Period A | Period B |
|---|---|---|
| Time | 2019/07/01 – 2021/06/30 | 2021/07/01 – 2023/06/30 |

frequently passes the Kupiec test for interval validity. It performs particularly well for functions of fundamental variables such as the price spread (a difference between day-ahead and intraday prices) and residual load (load minus RES).

Beyond forecasting performance, the study demonstrates the practical value of multidimensional probabilistic forecasts for trading strategy optimization with an example of a small RES utility. Arbitrage between short-term markets increases average profits while reducing financial risk, as measured by the 5% Value-at-Risk (VaR) and 10% Expected Shortfall (ES). Furthermore, generation curtailment enhances financial outcomes and allows strategies to be flexibly adjusted to varying levels of risk aversion.

This article is structured as follows. Section 2 presents the main characteristics of the dataset explored in the study. Next, the forecasting methods are presented in Section 3. Section 4 describes the trading strategy of a RES utility. Finally, the results of the empirical study are presented in Section 5. Section 6 provides conclusions.

## 2. Data

In this article, we analyze a dataset describing the EPEX SPOT market in Germany (`http://www.epexspot.com`, `https://transparency.entsoe.eu`). The data cover four and a half calendar years, from January 1, 2018 to June 30, 2023, with an hourly resolution. The dataset is divided into two main parts: an initial training window (from January 1, 2018 to June 30, 2019) and an evaluation window (from July 1, 2019 to June 30, 2023). The evaluation period encompasses highly diverse market conditions. The first half is characterized by relatively low price levels and low volatility, while the second half reflects the effects of the war in Ukraine and the subsequent gas crisis, marked by substantially higher and more volatile electricity prices. To account for these differences, the evaluation window is further divided into two sub-periods: Period A July 1, 2019 – June 30, 2021 and Period B July 1, 2021 – June 30, 2023 (see Table 1).

The dataset employed in this study comprises day-ahead market prices ($DA$) and intraday price indices ($ID$)[2], along with both actual and forecasted values of key fundamental variables, namely system Load ($L$) and generation from Renewable Energy Sources (RES, $R$), calculated as the combined output of wind and solar generation. In addition, the dataset includes commodity prices for coal (API2), natural gas (TTF), crude oil (Brent), and EU emission allowances (EUA), sourced from `https://investing.com`.

The time series of electricity prices and generation data are presented in Figure 1 and Figure 2, respectively. Figure 1 shows the average daily levels of DA, ID, and the corresponding Price Spread ($PS$), defined as the difference between the two market prices. Figure 2 presents the mean daily values of Load and RES generation, complemented by the Residual Load ($RL$), which approximates the share of

[2]The $ID$ index is based on a continuous intraday market and is computed as the volume-weighted average price of all traded offers for a selected delivery period.

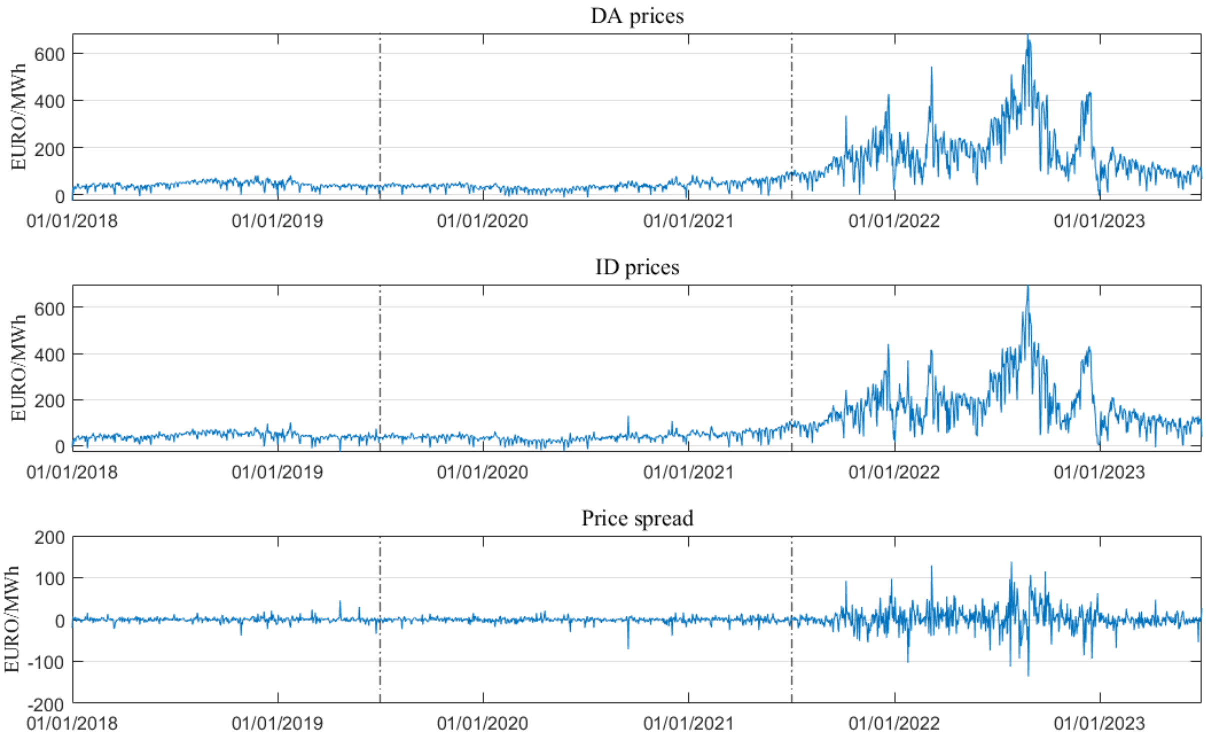


**Figure 1.** Average daily level of electricity prices and the price spread (EPEX, Germany). Vertical lines mark the boundaries between the training window and evaluation Period A and B.

demand met by conventional power plants: $RL = L - RES$.

All time series are pre-processed to account for time zone changes. The missing values (corresponding to the transition from winter to summer) were replaced by the arithmetic averages of the two nearest values. The doubled values (corresponding to the change from summer to winter) are replaced by their arithmetic mean.

# 3. Forecasting methods

In this paper, electricity prices and variables that describe the generation structure in consecutive hours are interpreted as separate time series (products). Their point forecasts are calculated using univariate models. While the overall model specification remains constant throughout the day, the parameter estimates are obtained independently for each hour. To replicate realistic trading conditions, it is assumed that all computations are conducted at 11:00 a.m., thereby restricting the information set to data available at that time.

Probabilistic forecasts are derived from point predictions obtained using autoregressive models with exogenous variables (ARX), a framework widely employed in the literature. The use of ARX models can be motivated by their very low computational complexity and high interpretability, coupled with well-established performance. The proposed methodology can be readily extended to incorporate alternative point forecasting approaches, including those based on conventional time series techniques as well as modern machine learning methods.

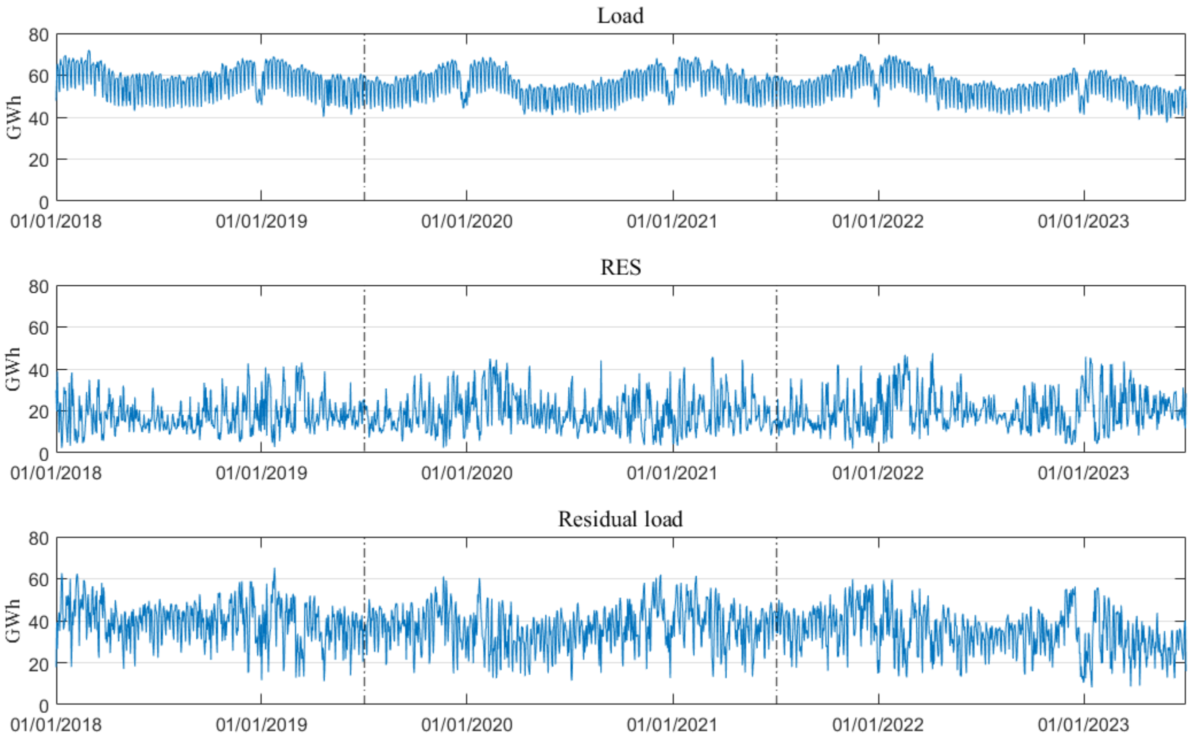


**Figure 2.** Average daily level of Load, RES generation and Residual Load (Germany). Vertical lines mark the boundaries between the training window and evaluation Period A and B.

## 3.1. Variance stabilizing transformation

Electricity prices exhibit both positive and negative spikes, typically driven by unpredictable weather conditions or transmission failures [22]. These extreme events can distort electricity price forecasts, as outliers tend to pull model coefficients toward values that fit the spikes, thereby reducing forecast accuracy under normal market conditions. Variance-stabilizing transformations (VSTs) are commonly applied to mitigate this effect by compressing the tails of the data distribution, which in turn allows forecasting models to achieve higher predictive accuracy [34].

In this study, the dependent variables are first standardized by subtracting the sample mean and dividing by the sample standard deviation. Subsequently, the DA and ID prices together with the price spread are transformed using the area hyperbolic sine (asinh) transformation. After generating forecasts, the inverse transformation and de-standardization are applied to recover the final predicted values.

## 3.2. ARX models

ARX type of models is widely used in the electricity price forecasting (EPF) literature [see 3, 24]. In these models, the expected value of an endogenous variable is expressed as a linear function of its past values (the autoregressive component) and a set of explanatory variables. When the model structure is predetermined and not subject to statistical selection or validation, it is often referred to as an expert

**Table 2.** Model specification

| Dep. variable | Lags | Explanatory variables: Hourly ($X_{t,h}$) | Explanatory variables: Daily ($X_t$) |
|---|---|---|---|
| $R_{t,h}$ | 1 | $FR_{t,h-1}$, $FR_{t,h}$, $FR_{t,h+1}$ | $FR_t^{ave}$, $FR_t^{min}$, $FR_t^{max}$ |
| $L_{t,h}$ | 1, 2, 7 | $FL_{t,h}$, $FR_{t,h}$ | $FL_t^{ave}$, $FL_t^{min}$, $FL_t^{max}$ |
| $RL_{t,h}$ | 1, 2, 7 | $FRL_{t,h}$ | $FRL_t^{ave}$, $FRL_t^{min}$, $FRL_t^{max}$ |
| $DA_{t,h}$ | 1, 2, 7 | $FL_{t,h}$, $FR_{t,h}$ | $DA_{t-1}^{ave}$, $DA_{t-1}^{min}$, $DA_{t-1}^{max}$, $DA_{t-1,24}$, $G_{t-2}$ |
| $ID_{t,h}$ | 1, 2, 7 | $FL_{t,h}$, $FR_{t,h}$ | $DA_{t-1}^{ave}$, $DA_{t-1}^{min}$, $DA_{t-1}^{max}$, $DA_{t-1,24}$, $G_{t-2}$ |
| $PS_{t,h}$ | 1, 2, 7 | $FL_{t,h}$, $FR_{t,h}$ | $DA_{t-1}^{ave}$, $DA_{t-1}^{min}$, $DA_{t-1}^{max}$, $DA_{t-1,24}$, $G_{t-2}$ |

Remarks: The model for $R_{t,h}$ includes a constant term, while the remaining models incorporate seven weekday dummy variables. Variables prefixed with $F$ represent TSO forecasts published day-ahead (with $FRL_{t,h} = FL_{t,h} - FR_{t,h}$). The abbreviations $ave$, $min$ and $max$ refer to the mean, minimum, and maximum values of a given variable on a selected day.

model. In this study, the following form of the model is adopted:

$$Y_{t,h} = D_t\alpha_h + \sum_{p\in\Theta} Y_{t-p,h}\rho_{p,h} + X_{t,h}\beta_h + X_t\gamma_h + \varepsilon_{t,h}, \tag{1}$$

where $Y_{t,h}$ denotes an endogenous variable observed on day $t$ at hour $h$. The model consists of four main components: an autoregressive part with lags $p \in \Theta$ that captures short-term dependencies and weekly seasonality, vectors of hourly, $X_{t,h}$, and daily, $X_t$, exogenous variables, and a vector of deterministic components, $D_t$. The vector of model parameters, $\theta_h = [\alpha_h', \rho_{1,h}, ..., \beta_h', \gamma_h']'$, is indexed by $h$ as it is estimated independently for each hour. Finally, $\varepsilon_{t,h}$ represents the random noise.

The lag structure, $\Theta$, and the selection of explanatory variables are presented in Table 2 and depend on the chosen exogenous variable. The model specifications – particularly those describing electricity prices, load, and RES – are consistent with approaches previously adopted in the literature [see 14, 20, 29].

Due to the scheduling of market activities, forecasts are assumed to be generated in the morning at 11:00, when information on generation levels and structures for hours later than 10:00 is not yet available. Therefore, for hours $h \geq 11$, the variables $W_{t-1,h}$, $R_{t-1,h}$, and $L_{t-1,h}$ are replaced by their corresponding TSO forecasts, denoted as $FW_{t-1,h}$, $FR_{t-1,h}$, and $FL_{t-1,h}$, respectively. Similarly, the intraday price level $ID_{t-1,h}$ is substituted with its day-ahead counterpart $DA_{t-1,h}$. Furthermore, let $G_t$ denotes a $(1\times4)$ vector of closing commodity prices (coal, gas, oil, and EUA) available in the evening of day $t$. To prevent data leakage, this vector is included in the models with a lag of $t-2$.

In this study, two alternative approaches to forecasting residual load and price spread are employed: the direct and the indirect methods. In the direct approach, these variables are modeled using ARX specifications, as defined in equation (1) and detailed in Table 2. In the indirect approach, forecasts of $PS_{t,h}$ and $RL_{t,h}$ are derived as linear functions of the predicted values of their individual components – electricity prices, load, and renewable energy generation. The direct approach is applied in the Quantile Regression Machine and conformal prediction frameworks, which rely on univariate models, whereas the indirect approach is adopted in the multiple split method, which leverages the multidimensional dependence structure of forecasts.

### 3.3. Quantile regression machine

The quantile regression machine (QRM) is a method for constructing probabilistic forecasts from point predictions. It was introduced by [28] and inspired by the quantile regression averaging (QRA) approach proposed by [18]. This simple yet accurate technique employs quantile regression (QR) to model and predict the quantiles of the conditional distribution of a selected variable. The model uses as inputs a constant term and the best available point forecast. In [28], the point forecast is computed as the average of predictions obtained from a single model estimated using calibration windows of different lengths. In the present study, since only one point prediction is available, it is used directly as the explanatory variable in the quantile regression.

Let us denote by $Q_\tau(Y_{t,h})$ the $\tau$ quantile of $Y_{t,h}$. Then in QRM it is assumed that

$$Q_\tau(Y_{t,h})^{QRM} = \alpha_{\tau,h} + \hat{Y}_{t,h}\beta_{\tau,h}, \tag{2}$$

where $\hat{Y}_{t,h}$ is a point forecast of $Y_{t,h}$ and $\alpha_{\tau,h}$, $\beta_{\tau,h}$ are parameters that depend on $\tau$ and change with the analyzed quantiles. Since the method uses predictions $\hat{Y}_{t,h}$, it requires additional data to calibrate the model parameters.

The coefficients of (2) can be estimated by minimizing the sum of *pinball scores* [$PS$, see 10]. A $PinS_{t,h}$ for a given day $t$ and hour $h$ is defined as:

$$PinS_{t,h}(\tau) = \begin{cases} (1-\tau)(Q_\tau(Y_{t,h}) - Y_{t,h}) & \text{for } Y_{t,h} < Q_\tau(Y_{t,h}), \\ \tau(Y_{t,h} - Q_\tau(Y_{t,h})) & \text{for } Y_{t,h} \geq Q_\tau(Y_{t,h}). \end{cases} \tag{3}$$

The estimation process can be carried out for 99 percentiles: $\tau = 0.01, ..., 0.99$ and therefore QRM can be used to approximate the entire distribution of $Y_{t,h}$. A prediction interval $PI$ with a nominal coverage $1-\alpha$ can be estimated as

$$PI_{1-\alpha}^{QRM} = [\ \hat{Q}_{\alpha/2}^{QRM}(Y_{T+1,h}),\quad \hat{Q}_{(1-\alpha/2)}^{QRM}(Y_{T+1,h})\ ], \tag{4}$$

where $\hat{Q}_\tau^{QRM}(Y_{T+1,h}) = \hat{\alpha}_{\tau,h} + \hat{Y}_{T+1,h}\hat{\beta}_{\tau,h}$ is a predicted quantile of $Y_{T+1,h}$.

### 3.4. Conformal predictions

Conformal prediction has recently gained attention in electricity price forecasting as a robust framework for quantifying predictive uncertainty. By calibrating prediction intervals using past forecast errors, it adapts naturally to the highly volatile and non-stationary nature of electricity markets. Recent studies [e.g. 17, 25] have demonstrated that conformal and adaptive conformal methods can yield more reliable and dynamically responsive uncertainty estimates compared to classical techniques such as QR.

The algorithm consists of three steps. Firstly, similar to QRM, the point forecasts $\hat{Y}_{t,h}$ are computed using ARX models and a moving window approach. Next, the predictions are compared with the actual realizations of $Y_{t,h}$, and forecast errors are calculated: $e_{t,h} = Y_{t,h} - \hat{Y}_{t,h}$. They are finally used to predict selected quantiles of a distribution and to construct prediction intervals.

In a simple form of the conformal prediction method, $PI$ with nominal coverage $1-\alpha$ is estimated as

$$PI_{1-\alpha}^{Conf} = [\ \hat{Y}_{T+1,h} - Q_{1-\alpha}(|e_{t,h}|), \hat{Y}_{T+1,h} + Q_{1-\alpha}(|e_{t,h}|)\ ], \tag{5}$$

where $Q_\tau(|e_{t,h}|)$ is the $\tau$ empirical quantile of the pool of absolute values of forecast errors $e_{t,h}$.

Using PIs, quantiles of a distribution are computed as

$$Q_\tau^{Conf}(Y_{T+1,h}) = \begin{cases} \hat{Y}_{T+1,h} - Q_{1-2\tau}(|e_{t,h}|) & \text{for } \tau < 0.5, \\ \hat{Y}_{T+1,h} + Q_{2\tau-1}(|e_{t,h}|) & \text{otherwise.} \end{cases} \tag{6}$$

The corresponding distribution is, therefore, symmetric around the point prediction $\hat{Y}_{T+1,h}$. A similar approach has been adopted by [17].

## 3.5. Multiple split method

In this research, we propose a *multiple split* forecasting method. It is an extension of an approach known in the literature under the name *split conformal prediction* or *inductive conformal inference* described by [2, 9, 16]. The main idea of this forecasting scheme is to use a random split of the data to construct probabilistic predictions. First, the training data is divided into two disjoint windows: estimation and calibration. The first (estimation) subset is used to estimate the model parameters, which are then applied to calculate point predictions of the observations both within the calibration window and out-of-sample (i.e. target point prediction). The forecast errors are then estimated by computing the difference between the actual observations and their predictions in the calibration subset. Finally, the errors are used to approximate the distribution of the dependent variable.

The MS approach extends previous methodologies in several important ways. First, the random data split is performed multiple times to enhance forecast accuracy and reduce outcome variability. In the remainder of the article, the number of splits is denoted by $N$, and the method is referred to as MS($N$). Unlike leave-$k$-out or delete-$d$ jackknife procedures [see 30], the MS method does not consider all possible sample partitions, as doing so would substantially increase computational complexity while offering limited gains in predictive performance. Second, forecast errors from the training windows are used directly to construct an ensemble of forecasts, rather than to estimate distributional quantiles. Consequently, the final ensemble is obtained by aggregating results from individual splits, without the intermediate step of averaging quantiles or prediction intervals, as required in [16]. Moreover, when the objective is to produce a probabilistic forecast of a function of the analyzed variables, this can be achieved simply by applying the function to the elements of the ensemble. Finally, the MS method is applied to forecast a multidimensional random variable. When the forecast errors are computed for several variables simultaneously, they preserve the natural correlation structure, eliminating the need for explicit modeling of their interdependencies. Hence, resampling-based techniques are particularly useful in this context, as they avoid parametric modeling of the joint distribution while allowing to obtain a multidimensional forecast.

In the *multiple split* method, all variables are forecasted jointly and collected in a $K$-dimensional vector $Y_{t,h} \in \mathbb{R}^K$. Suppose the data-generating process can be represented by a linear model:

$$Y_{t,h} = Z_{t,h}A + e_{t,h}, \tag{7}$$

where $Z_{t,h}$ denotes a vector of exogenous variables, $A$ is a matrix of coefficients (with the $k$-th column corresponding to the $k$-th element of $Y_{t,h}$), and $e_{t,h} \in \mathbb{R}^K$ is a vector of residuals. The parameters in $A$ can be estimated either jointly or equation-by-equation, depending on the model specification and the researcher's preference. In this study, each equation is specified separately, as shown in Table 2; therefore, the latter approach is adopted.

Assume that a sample $S$ is observed for periods $t = 1, \dots, T$, and the goal is to produce a forecast for period $T + 1$. Each iteration of the proposed algorithm consists of $N$ independent random splits of the available sample. The procedure for a single split ($i = 1, \dots, N$) is illustrated schematically in Fig. 3. In the diagram, the estimation subset and associated steps are marked in green. Gray shading represents calibration periods. Striped boxes show the final forecasts. The placement of boxes indicates the order of operations (from left to right and from top to bottom), while the arrows show the dependence on the results from previous steps.

The $i$th split consists of the following steps:

1. The sample is randomly divided into estimation ($S^{(i)}_{estim}$, green color Fig. 3) and calibration ($S^{(i)}_{calib}$, gray color Fig. 3) subsets such that $S = S^{(i)}_{estim} \cup S^{(i)}_{calib}$ and $S^{(i)}_{estim} \cap S^{(i)}_{calib} = \emptyset$.
2. The data in the estimation window, $S^{(i)}_{estim}$, is used to estimate the parameters of models used for forecasting, $\hat{A}^{(i)}$, which are next employed to calculate predictions for periods $t \in \{S^{(i)}_{calib}, T+1\}$: $\hat{Y}^{(i)}_{t,h} = Z_{t,h}\hat{A}^{(i)}$
3. For each observation in the calibration window, $t \in S^{(i)}_{calib}$, the vector of forecast errors is calculated as $\hat{e}^{(i)}_{t,h} = Y_{t,h} - \hat{Y}^{(i)}_{t,h}$
4. The ensemble of predictions is constructed as

$$\Psi_i = \{y \in R^K : y = \hat{Y}^{(i)}_{T+1,h} + \hat{e}^{(i)}_{t,h}, t \in S^{(i)}_{calib}\}$$

Steps 1 – 4 are repeated $N$ times, and new sets of forecasts are added to the pool

$$\Psi = \bigcup_{i=1}^{N} \Psi_i. \tag{8}$$

Finally, the ensemble of predictions is used to construct prediction intervals

$$PI^{MS}_{1-\alpha} = [\ Q_{\alpha/2}(\Psi), \quad Q_{(1-\alpha/2)}(\Psi)\ ], \tag{9}$$

where $1 - \alpha$ is the nominal coverage level of PI.

## 3.6. Forecast evaluation

Probabilistic forecasts discussed in previous sections represent two popular types of prediction: quantile forecasts and ensemble forecasts. Although it is possible to estimate quantiles from a set of predictions, it is more difficult to generate a diversified ensemble from a set of quantiles.

First, using QRM, conformal predictions, or MS method, we approximate the distribution of the variables of interest with 99 quantiles and construct prediction intervals of nominal coverage 90%, 95%

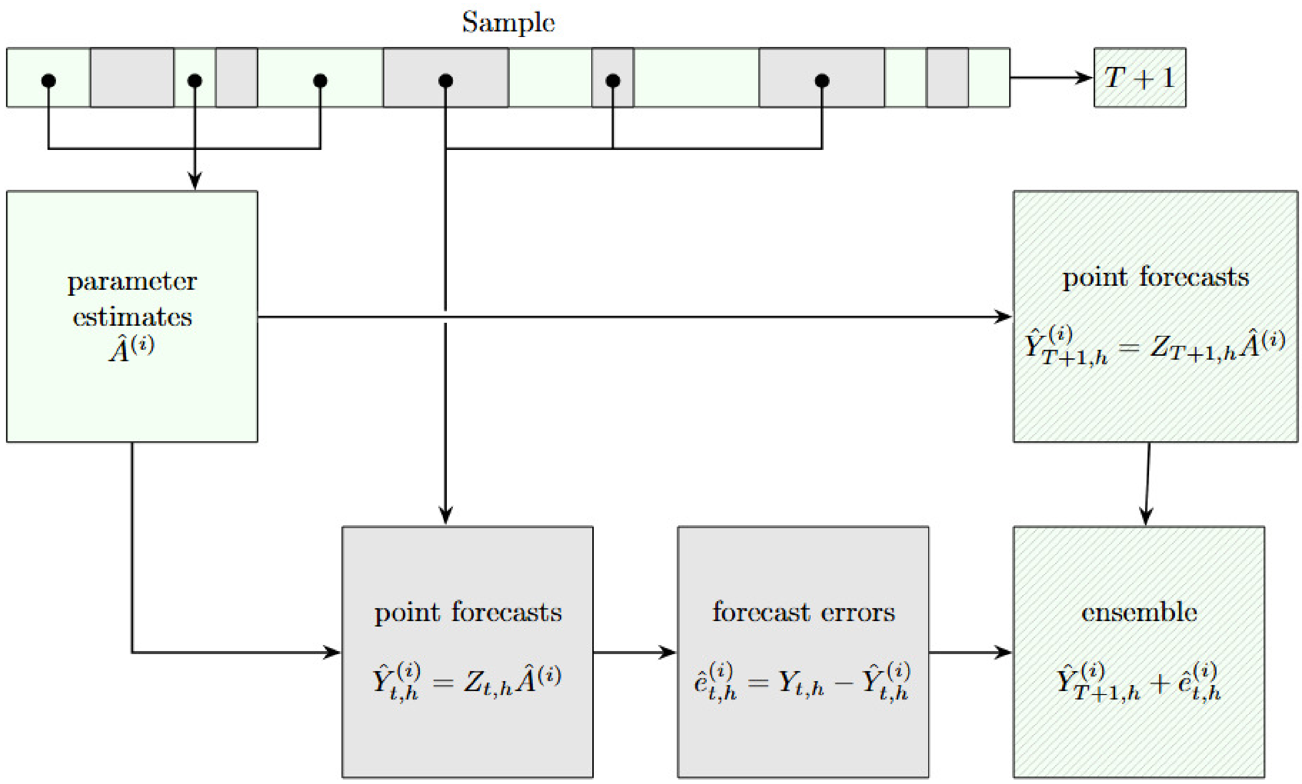


**Figure 3.** Schematic illustration of the algorithm.

and 98%. The precision of PIs is evaluated with the *PI coverage probability* (PICP). The measure is based on the average number of observations in the testing window that fall into the PI. For an hour $h$, it is calculated as

$$\text{PICP}_h = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}(Y_{t,h} \in PI_{t,h}), \tag{10}$$

where $\mathbb{I}(Y_{t,h} \in PI_{t,h})$ is an indicator function that takes the value 1 when the variable $Y_{t,h}$ falls within the prediction interval $PI_{t,h}$ and zero otherwise. Finally, the values $PICP$s for individual hours are averaged

$$\text{PICP} = \frac{1}{24}\sum_{h=1}^{24}\text{PICP}_h. \tag{11}$$

To evaluate whether $PICP_h$ is sufficiently close to the nominal coverage, we perform a Kupiec test [13] for each hour of the day separately. The null hypothesis states that the empirical coverage equals the nominal level, whereas under the alternative, the $PICP_h$ differs significantly from $1-\alpha$. In this article, we report the percentages of hours for which we were unable to reject the null at the significance level 5%. Therefore, the closer the measure is to one, the more successful the prediction method is in providing PIs of a predefined probability level.

Finally, to assess the quality of the estimated quantiles, we use the CRPS measure described by [6]. It evaluates both the calibration and the sharpness of the distribution. For a particular day, $t$, and hour, $h$, the CRPS can be approximated as the arithmetic mean of pinball scores (3) over 99 percentiles

$$\text{CRPS}_{t,h} = \frac{1}{99} \sum_{\tau=0.01}^{0.99} PinS_{t,h}(\tau). \tag{12}$$

In order to capture the ability of models to predict tails of the distribution, an additional modification of the CRPS is presented

$$\text{CRPS}_{t,h}^{tails} = \frac{1}{10} \left( \sum_{\tau=0.01}^{0.05} PinS_{t,h}(\tau) + \sum_{\tau=0.95}^{0.99} PinS_{t,h}(\tau) \right), \tag{13}$$

in which pinball is averaged only over the extreme quantiles. The final measures, CRPS and CRPS-tail, are computed as the mean of $\text{CRPS}_{t,h}$ and $\text{CRPS}_{t,h}^{tails}$ over the whole sample, respectively.

The lower the value of CRPS, the better the approximation of the distribution. It should be noted that the sharpness of the distribution has a significant impact on the measure. However, sharpness is a criterion that should be analyzed only when the calibration is correct. Furthermore, since the pinball score is a loss function used to estimate the QRM, the CRPS may favor the results that arise from this forecasting method. Thus, we believe that CRPS should be analyzed together with other measures, such as PICP.

# 4. Support of the decision process of a wind farm

In this research, we show how the joint prediction of different variables describing electricity market can be used to support the decision-making process of an electricity producer. We analyze the trading decisions of a company that owns wind and solar farms spread throughout Germany. The amount of energy produced, $r_{t,h}$, is assumed to be small enough not to influence market prices. Each day, utility generates electricity, which is sold in the power energy exchange. The company has access to the day-ahead and intraday markets and does not speculate; i.e., it aims to sell the entire generated energy in either of these markets. Although purchasing energy on the intraday market may be necessary in the case of an overestimated generation forecast, this operation is never done intentionally.

Each day, the actions of the company are as follows:

1. On a day before delivery, $t-1$:
   - Forecast the generation level, $\hat{r}_{t,h}$, and electricity prices.
   - Choose a fraction $q \in [0, 1]$ of $\hat{r}_{t,h}$ to offer in the DA market.
   - Submit an unlimited bid to the DA market for the quantity $q\hat{r}_{t,h}$ or curtail.
2. On the delivery day, $t$, (if the generation is not curtailed):
   - Self-balance the positions to match the actual generation $r_{t,h}$.
   - If the initial DA offer ($q\hat{r}_{t,h}$) is less than the actual generation $r_{t,h}$, sell the excess in ID market. Otherwise, purchase the missing amount in ID market.

In this paper, it is assumed that the company faces some operational and maintenance costs $C_{O\&M}$, which influence the level of marginal profits. For simplicity, these costs are kept constant and expressed in EUR per 1 MWh of production. Other types of costs are interpreted as fixed costs and, therefore, do not influence the decision-making process. Then, the profit of the company can be calculated as follows:

$$\pi_{t,h}(q) = \underbrace{q\hat{r}_{t,h}DA_{t,h}}_{\text{DA market}} + \underbrace{(r_{t,h} - q\hat{r}_{t,h})ID_{t,h}}_{\text{ID market}} - \underbrace{r_{t,h}C_{O\&M}}_{\text{O\&M costs}} . \tag{14}$$

It should be noted here that the level and distribution of income depend on three main sources of uncertainty: associated with the unknown level of generation, $r_{t,h}$, and market prices: $DA_{t,h}$, $ID_{t,h}$. Since profit is a nonlinear function of these variables, its distribution is nontrivial. Resampling and simulation methods, such as the MS method, are of great help in estimating its probabilistic forecasts.

## 4.1. Trading strategies

### 4.1.1. Naïve strategies

In this research, two naïve strategies are considered that are based on predefined values of $q_{t,h}$. The first one, called **DA strategy**, assumes that the entire predicted generation, $\hat{r}_{t,h}$, is sold in the DA market, hence sets $q_{t,h} = 1$. Since the strategy responds only to the fluctuations of generation but does not adjust to the market situation, its profitability depends primarily on the accuracy of RES generation forecasts. Alternatively, the company may decide to trade only in the ID market (**ID strategy**). Then, it assigns $q_{t,h} = 0$, which eliminates the need to generate day-ahead predictions of $r_{t,h}$, as the entire production is sold in the continuous market.

DA strategy is the most intuitive and widely adopted approach and is therefore used as the benchmark in the subsequent analysis.

### 4.1.2. Data driven strategies

This article presents three data-driven approaches that utilize probabilistic forecasts of profits to select $q_{t,h}$. To optimize the choice, a distribution of $\pi_{t,h}(q)$ is constructed – by applying (14) to a set of joint forecasts of electricity prices and generation – for different values of $q$ from a grid in the interval $[0, 1]$. Next, three distinct optimization criteria are applied to determine the proportion of predicted generation offered in DA market.

The first and most natural approach to choose $q_{t,h}$ is to maximize the expected value of the profits, henceforth called the **expected profit strategy**. It is based on a point forecast of income, calculated here as the median of the ensemble of $\pi_{t,h}(q)$. The value of $q$ is chosen as the one for which the median is the highest. This strategy is expected to bring high income at the cost of increased risk.

The next **expected shortfall strategy** is particularly relevant for traders who are more risk-averse. It aims to maximize the expected shortfall (ES), which represents the average value of profits in the lower tail of its distribution. Here, ES is calculated as the mean of the elements in the ensemble that fall below the 10% quantile. The ES is a coherent risk measure [see 1] and therefore is perceived as a better indicator of tail risk than, for example Value-at-Risk (VaR).

The final strategy, called **Sharpe ratio strategy**, aims to find a balance between maximizing profits and minimizing risk. To do this, the Sharpe ratio (expected profit divided by its standard deviation, [31]) is calculated for different values of $q$. The optimal $q$ is chosen so that the ratio is maximized.

The three data-driven strategies can be summarized as follows:

1. Expected profit strategy ($E\pi$):

$$q_{t,h} = \underset{q\in[0,1]}{\operatorname{argmax}}\, Q_{50\%}(\pi_{t,h}(q)),$$

   where $Q_\tau(\pi(q))$ is a $\tau$ quantile of $\pi(q)$.

2. Expected shortfall strategy ($ES$):

$$q_{t,h} = \underset{q\in[0,1]}{\operatorname{argmax}}\, \mathrm{E}(\pi_{t,h}(q)|\pi_{t,h}(q) < Q_{10\%}(\pi_{t,h}(q))),$$

3. Sharpe ratio strategy ($SR$):

$$q_{t,h} = \underset{q\in[0,1]}{\operatorname{argmax}}\, \mathrm{SR}(q) = \underset{q\in[0,1]}{\operatorname{argmax}}\, \frac{\bar{\pi}_{t,h}(q)}{\sigma_{t,h}(q)},$$

   where $\bar{\pi}_{t,h}(q)$ and $\sigma_{t,h}(q)$ denote the average value and the standard deviation of the predicted profits in the ensemble of $\pi_{t,h}(q)$, respectively.

It should be noted that the presented approaches are applicable solely to ensemble predictions, such as MS, which allow for a straightforward construction of probabilistic forecasts of any function of a vector of random variables. In the case of other types of distribution approximations, for example by quantiles as in QRM, the application of the methodology would require building a separate model to directly predict the distribution of $\pi(q)$ for each value of q, making the process cumbersome and time consuming. Moreover, there is no direct link between distribution percentiles and measures such as ES or SR used here for selection of $q_{t,h}$.

#### 4.1.3. Curtailment

In recent years, one could observe a frequent occurrence of negative prices in electricity markets (see Fig. 1). The existence of this phenomenon implies that even an optimal bidding strategy can result in financial losses. To avoid such a situation, the company may decide to curtail production. It means that it either stops the generation (for example the turbines are turned off) or stores electricity. The second solution becomes an attractive alternative as more and more investments are made in the development of energy storage systems.

In our research, the production curtailment implies that no electricity is sold in any of the markets, which simulates temporal shut down of the production without storing the unsold generation. To account for the risk aversion of company owners, the curtailment decision is based on probabilistic forecast of profits. We assume that the generation is stopped, when a selected quantile of the profit distribution falls below zero: $Q_\tau(\pi_{t,h}(q_{t,h})) < 0$ and that the company is engaged in trade when $Q_\tau(\pi_{t,h}(q_{t,h})) \geq 0$.

The selection of $\tau$ reflects the risk tolerance of traders. A risk-neutral traders would likely base their decision on the center of the distribution (for example, the median), whereas a risk-averse would place bids based on pessimistic scenarios (low quantiles). Traders who are reluctant to halt production may consider high quantiles that exceed 50%. Finally, as the aversion to curtailment increases, the selected quantile may converge to one, which will represent the strategy that assumes daily trade (i.e. without stopping generation).

#### 4.1.4. Oracle

The results of various trading strategies are compared with those of an Oracle, which describes a trading strategy based on perfect forecasts. Knowledge about future prices and generation is used in two ways. First, to select the optimal market: either DA or ID. Hence, $q_{t,h}$ takes on either value zero or one. Next, it is decided whether to curtail production. If the future profits are negative, the company will cease generation and await a price increase.

### 4.2. Evaluation of trading strategies

The performance of the presented trading strategies is evaluated according to the level of average profits, the trading risk, and the frequency of generation curtailment. Hence, we provide the company with a broad perspective that encompasses the potential preferences and aversions of the traders.

When the income level is considered, two quantities are reported: the average total profit and the average profit per trade. The main difference between these measures results from the approach to days when production is curtailed. The average total profits include the entire testing period and is calculated as the mean of individual hourly profits:

$$\bar{\pi} = \frac{1}{T}\frac{1}{24}\sum_{t=1}^{T}\sum_{h=1}^{24}\pi_{t,h}(q_{t,h}), \tag{15}$$

where $q_{t,h}$ is the optimal value of the parameter selected using the chosen strategy. On the contrary, the profit per trade, $\tilde{\pi}$, is calculated only for periods when the trade occurs and disregards observations when the generation is curtailed. These two measures, $\bar{\pi}$ and $\tilde{\pi}$, are different only for strategies with a stopping rule.

To better understand the significance of the curtailment, the frequency of trade is computed. It shows how often the stopping rule is not violated. We expect that frequent curtailment of generation may result in a rise in the average profit per trade and a fall in the total profit.

Finally, the trading risk of strategies without curtailment is evaluated using the VaR and ES measures. $VaR_{5\%}$ is calculated as the 5% quantile of the average profit $\pi_{t,h}$, and $ES_{10\%}$ represents the average value of profit below its 10% quantile. The measures are not applied directly to strategies with stoppling rule, because they bring zero profits in days of curtailment. This consequently affects how profits are distributed and complicates the interpreting of the results.

## 5. Results

The forecasting experiment is based on a rolling window scheme, in which each hour of the two-year validation period is predicted individually. We use a calibration window of 365 days to estimate the parameters of ARX models and construct probabilistic forecasts using the MS method. Additional 182 days are utilized to apply QRM and conformal prediction schemes.

## 5.1. Statistical accuracy of probabilistic forecasts

To evaluate forecast accuracy, we first consider four fundamental variables that define electricity markets: DA price, ID price, total load, and RES generation. Next, the ability to predict a linear combination of these variables is evaluated: the price spread (DA price minus ID price) and the residual load (total load minus RES generation).

### 5.1.1. Electricity prices and price spread

First, we consider the probabilistic forecasts of electricity prices and the price spread (DA price minus ID price). The results are presented in Table 3, which reports empirical coverage probabilities for three prediction interval (PI) levels – 90%, 95%, and 98% – along with the average frequencies of the Kupiec test indicating correct coverage and the values of the CRPS and CRPS-tail measures. The outcomes are analyzed separately for two out-of-sample subperiods: Period A and Period B.

The results show that all forecasting methods produce PIs that are too narrow, resulting in PICPs below their nominal levels. However, the ensemble forecasts – particularly MS(20) – appear to be better calibrated to the data. For instance, the PICP of the QRM method falls well below the nominal coverage, whereas for MS(20) the PICP is much closer to $1-\alpha$. This observation is confirmed by the Kupiec test. In Period A, the QRM method correctly constructs PIs in fewer than 7% of cases, while for MS(20) the PICPs are not statistically different from the nominal level in over 79% of cases for DA, 87% for ID, and 97% for PS. The Conformal Prediction method provides PIs that are more accurate than those from QRM, although they still fall short of the performance achieved by the MS(20) ensemble approach.

In Period B, the relative performance of the analyzed methods remains unchanged, but the overall quality of predictions declines significantly. Neither QRM nor Conformal Prediction is able to provide PIs with correct coverage. In the case of MS(20), the results show that only in 10–20% of cases does the Kupiec test indicate that PI coverage is sufficiently close to the nominal level. For PS, the performance of Conformal Prediction and MS(20) improves, reaching 8.3% and 44.4%, respectively.

Additionally, the forecasting methods are compared using the CRPS and CRPS-tail measures. Since the CRPS metric is derived from the pinball score, which is minimized when estimating quantile regression, it is not surprising that the QRM method outperforms the other approaches in this regard. However, when the tails of the distribution are considered, the results indicate that MS(20) provides the lowest CRPS-tail values for most of the analyzed variables.

It is worth noting that MS(20) performs particularly well when predicting the price spread. The Kupiec test confirms correct calibration of PIs in 97.5% and 44.4% of cases in Periods A and B, respectively. Hence, under both low and high price volatility, the method is able to produce PIs with PICPs close to the nominal level. Unlike QRM and Conformal Prediction, the MS approach captures the multidimensional distribution of prices and therefore helps control the risk associated with forecasting both markets simultaneously.

**Table 3.** Forecast accuracy measures for electricity prices and price spread

| | DA | | ID | | Price spread | |
|---|---|---|---|---|---|---|
| | Period A | Period B | Period A | Period B | Period A | Period B |
| PICP | QRM | | | | | |
| 90% | 85.07% | 85.00% | 85.31% | 84.91% | 86.89% | 85.16% |
| 95% | 91.19% | 91.17% | 91.47% | 91.03% | 92.53% | 90.94% |
| 98% | 95.31% | 95.14% | 95.78% | 95.13% | 96.44% | 95.09% |
| Kupiec | 0.00% | 2.78% | 6.94% | 1.39% | 26.39% | 0.00% |
| CRPS | **1.7815** | **10.5919** | **2.8236** | **14.0027** | **2.3443** | **10.3616** |
| CRPS-tails | 0.5587 | 3.1882 | 0.9213 | **3.9585** | 0.8620 | **3.2778** |
| PICP | Conformal predictions | | | | | |
| 90% | 87.21% | 85.76% | 87.09% | 86.42% | 88.22% | 86.26% |
| 95% | 93.50% | 91.12% | 93.68% | 91.42% | 94.07% | 92.39% |
| 98% | 97.09% | 96.32% | 97.25% | 96.15% | 97.58% | 96.24% |
| Kupiec | 54.17% | 1.39% | 59.72% | 4.17% | 91.67% | 8.33% |
| CRPS | 1.8833 | 11.2870 | 2.9398 | 14.7900 | 2.3631 | 10.4816 |
| CRPS-tails | 0.5898 | 3.3777 | 0.9724 | 4.2540 | 0.8529 | 3.2890 |
| PICP | MS(1) | | | | | |
| 90% | 85.54% | 83.19% | 86.54% | 83.82% | 88.51% | 84.97% |
| 95% | 92.03% | 90.08% | 92.62% | 90.31% | 93.87% | 91.13% |
| 98% | 96.40% | 95.24% | 96.72% | 95.37% | 97.26% | 95.68% |
| Kupiec | 8.33% | 0.00% | 23.61% | 1.39% | 76.39% | 11.11% |
| CRPS | 1.9365 | 11.4651 | 2.9993 | 15.3764 | 2.4758 | 11.6684 |
| CRPS-tails | 0.5890 | 3.3831 | 0.9229 | 4.6299 | 0.8230 | 3.8493 |
| PICP | MS(20) | | | | | |
| 90% | 88.28% | 85.71% | 88.75% | 86.42% | 90.48% | 87.47% |
| 95% | 94.09% | 91.91% | 94.07% | 92.32% | 95.13% | 92.65% |
| 98% | 97.42% | 96.08% | 97.45% | 96.15% | 97.89% | 96.66% |
| Kupiec | **79.17%** | **11.11%** | **87.50%** | **19.44%** | **97.22%** | **44.44%** |
| CRPS | 1.8666 | 11.1147 | 2.9059 | 14.8214 | 2.3930 | 11.1261 |
| CRPS-tails | **0.5481** | **3.1632** | **0.8741** | 4.3573 | **0.7885** | 3.6107 |

### 5.1.2. Load, RES and Residual load

The results of the forecasting experiment for variables describing electricity generation: Load, RES, and Residual Load, are presented in Table 4. Unlike in the case of electricity prices, the behavior of these variables does not change substantially over time. As shown in Fig. 2, the patterns of Load and RES remain stable throughout the entire out-of-sample period. This stability is reflected in the forecast accuracy. Although differences between Periods A and B exist, they are minor, particularly when the CRPS measures are considered.

Similar to the case of electricity prices, for the generation structure the QRM method produces prediction intervals that are too narrow, as confirmed by the Kupiec test. In contrast, the PICPs obtained from the Conformal Prediction and MS(20) methods are much closer to the nominal levels, and the Kupiec test fails to reject the null hypothesis in more than 50% of cases for all variables. In many cases, this percentage exceeds 90%, indicating very high forecasting accuracy.

When the two methods, conformal prediction and MS(20), are compared, the latter yields prediction intervals with coverage probabilities closer to the nominal levels (except Load in Period B). Moreover, both the CRPS and CRPS-tail metrics indicate that the MS(20) approach outperforms the alternative methods. Finally, MS(20) proves particularly effective in forecasting Residual Load, which resembles the case of the price spread. In both cases, leveraging the multidimensionality of predictions leads to enhanced forecast accuracy.

## 5.2. Economic value of forecasts

To evaluate the economic value of the ensemble MS method, probabilistic forecasts are used to support the decision-making process of a generation utility. For computational simplicity, the utility's production is assumed to be proportional to the total RES generation in Germany, with $r_{t,h} = 0.001R_{t,h}$. Hence, the same model specification as for $R_{t,h}$ is used to predict $r_{t,h}$. If additional information on the utility's specific generation becomes available, it can be easily incorporated into the forecasting framework.

The company's profits depend on operational and maintenance costs. For wind generation, which represents the dominant component of RES, O&M costs typically account for 25–35% of the levelized cost of energy [27], corresponding to approximately USD 10–30 per MWh [4]. [35] estimate these costs at around USD 11 per MWh for onshore installations, with a declining trend expected in the future. Considering that the exchange rate between USD and EUR has fluctuated between 0.9 and 1 USD/EUR, this study assumes an O&M cost of $C_{O\&M} = 10$ EUR/MWh.

First, we analyze the strategies without curtailment. The results are presented in Table 5, where the naïve strategies (DA and ID) with predefined levels of $q = 1$ or $q = 0$ are compared with data-driven strategies based on the MS(1) and MS(20) forecasting methods. The evaluation is based on the average income level ($\bar{\pi}$) and two risk measures, $VAR_{5\%}$ and $ES_{10\%}$, as described in the previous sections.

The results show that the average profit levels change significantly over time. They oscillate around 4 EUR in Period A and increase to 27–29 EUR in Period B. Although the mean of the profit distribution rises by more than 20 EUR, the associated $VAR_{5\%}$ and $ES_{10\%}$ increase only by 4–6 EUR.

When the naïve strategies are compared, the results indicate that the DA strategy is less profitable in Period A than its ID counterpart. A shift in trading activity from the DA to the ID market increases

**Table 4.** Forecast accuracy measures for load, RES and residual load

| | Load | | RES | | Residual load | |
|---|---|---|---|---|---|---|
| | Period A | Period B | Period A | Period B | Period A | Period B |
| PICP | QRM | | | | | |
| 90% | 87.10% | 88.97% | 88.73% | 88.02% | 87.35% | 87.09% |
| 95% | 92.32% | 93.56% | 93.88% | 93.01% | 92.88% | 92.56% |
| 98% | 96.00% | 96.44% | 96.81% | 96.45% | 96.20% | 96.03% |
| Kupiec | 9.72% | 56.94% | 73.61% | 30.56% | 22.22% | 11.11% |
| CRPS | 0.5349 | **0.4750** | 0.4956 | 0.5170 | 0.7607 | 0.7521 |
| CRPS-tails | 0.1423 | 0.1314 | **0.1319** | **0.1336** | 0.1963 | 0.1913 |
| PICP | Conformal predictions | | | | | |
| 90% | 88.27% | 90.20% | 89.68% | 88.69% | 89.23% | 89.26% |
| 95% | 93.51% | 95.27% | 94.53% | 93.76% | 94.39% | 94.84% |
| 98% | 96.61% | 97.61% | 97.47% | 97.22% | 97.38% | 97.35% |
| Kupiec | 58.33% | 95.83% | 97.22% | 80.56% | 95.83% | 98.61% |
| CRPS | 0.5300 | 0.4930 | 0.4927 | 0.5241 | 0.7468 | 0.7756 |
| CRPS-tails | 0.1460 | 0.1345 | 0.1340 | 0.1422 | 0.1965 | 0.1962 |
| PICP | MS(1) | | | | | |
| 90% | 86.37% | 90.03% | 89.75% | 88.14% | 87.83% | 88.37% |
| 95% | 92.29% | 94.87% | 94.52% | 93.55% | 93.35% | 93.71% |
| 98% | 96.14% | 97.76% | 97.53% | 96.93% | 96.87% | 96.97% |
| Kupiec | 11.11% | **100.0%** | 98.61% | 63.89% | 54.17% | 70.83% |
| CRPS | 0.5413 | 0.5045 | 0.5045 | 0.5279 | 0.7442 | 0.7156 |
| CRPS-tails | 0.1474 | 0.1342 | 0.1365 | 0.1434 | 0.1953 | 0.1876 |
| PICP | MS(20) | | | | | |
| 90% | 88.35% | 91.72% | 90.63% | 89.30% | 89.48% | 89.97% |
| 95% | 93.82% | 96.18% | 95.30% | 94.46% | 94.65% | 94.95% |
| 98% | 97.10% | 98.28% | 97.94% | 97.57% | 97.63% | 97.95% |
| Kupiec | **73.61%** | 84.72% | **100.0%** | **98.61%** | **97.22%** | **100.0%** |
| CRPS | **0.5260** | 0.4884 | **0.4942** | **0.5169** | **0.7250** | **0.6957** |
| CRPS-tails | **0.1391** | **0.1289** | 0.1341 | 0.1391 | **0.1881** | **0.1795** |

**Table 5.** Profits and risks, strategies without curtailment

| | | Period A | | | Period B | | |
|---|---|---|---|---|---|---|---|
| Model | Strategy | $\bar{\pi}$ | $ES_{10\%}$ | $VaR_{5\%}$ | $\bar{\pi}$ | $ES_{10\%}$ | $VaR_{5\%}$ |
| ARX | $DA$ | 4.244 | -3.425 | -1.346 | 27.872 | 1.369 | 3.148 |
| | $ID$ | 4.373 | -4.358 | -1.347 | 27.552 | -0.205 | 1.517 |
| MS(1) | $E\pi$ | 4.397 | -3.653 | -1.200 | 28.925 | 1.576 | 3.331 |
| | $ES$ | 4.335 | -3.339 | -1.246 | 28.601 | 1.787 | 3.402 |
| | $SR$ | 4.292 | -3.584 | -0.851 | 28.225 | 1.633 | 3.521 |
| MS(20) | $E\pi$ | **4.417** | -3.653 | -1.209 | **29.060** | 1.822 | 3.432 |
| | $ES$ | 4.333 | **-3.323** | -1.215 | 28.651 | **1.828** | 3.497 |
| | $SR$ | 4.288 | -3.586 | **-0.806** | 28.275 | 1.816 | **3.556** |

income by 3.04%. In contrast, the results for Period B are reversed, with the DA strategy generating profits that are 1.15% higher than those of the ID strategy. Moreover, the ID strategy is associated with a higher level of risk, as reflected by lower values of both $VAR_{5\%}$ and $ES_{10\%}$.

The data-driven strategies, particularly MS(20), not only improve the average profit relative to the DA strategy but also reduce risk exposure. The $E\pi$ portfolio construction method leads to profit increases of 4.07% and 4.26% in Periods A and B, respectively. Slightly smaller gains are observed for the $ES$ strategy, which produces 2.10% and 2.79% higher revenues in the corresponding periods. The $SR$ strategy yields profit improvements of approximately 1–1.5% over the DA benchmark. Although the data-driven strategies achieve higher profitability, they also reduce risk levels. As expected, the $ES$ strategy exhibits the highest expected shortfall among all approaches, while the $SR$ strategy yields the highest VaR. The differences between strategies are most pronounced in Period A, which is characterized by higher trading risk than Period B.

Next, we analyze the results for the strategies that incorporate a stopping rule. If trading and curtailment decisions are based on perfect forecasts, as in the Oracle strategy, the company achieves revenues exceeding those of the DA strategy by 17.63% and 10.51% in Periods A and B, respectively. The gains obtained from data-driven strategies based on imperfect forecasts are illustrated in Fig. 4, which presents the percentage change in average profits relative to the DA strategy. In the figure, the solid and dashed lines represent strategies with and without a stopping rule, respectively. The additional revenue depends on the trader's attitude toward risk, represented by the profit quantile $\tau$ used in the curtailment decision. Low values of $\tau$ correspond to risk-averse behavior, whereas $\tau > 0.5$ indicates a reluctance to reduce production. As $\tau$ approaches 1, the trader's behavior converges to that of a strategy without curtailment.

The outcomes for Period A and Period B differ qualitatively. In Period A, which is characterized by relatively low price levels and a high risk of losses, generation curtailment leads to a substantial increase in average profits relative to both the DA strategy and the strategies without a stopping rule. In contrast, during Period B, greater risk aversion results in lower profits compared with the strategies without curtailment. For very low values of $\tau$, revenues even fall below the DA strategy level. Consequently, the optimal choice of $\tau$ depends on market conditions and appears to oscillate around 30% for Period A and 70% for Period B.

The decline in average profits is associated with the opportunity cost of halting generation. The

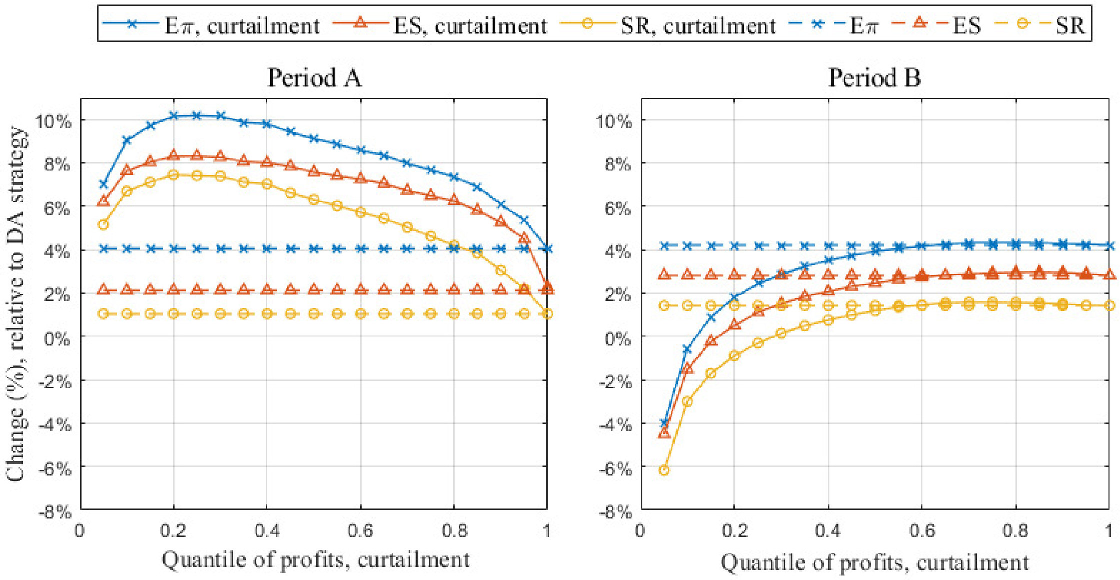


**Figure 4.** Average profits, percentage change relative to DA strategy.

average profit measure includes zero income from days when no trading occurs. Therefore, we also analyze the average profit per trade, calculated using only the days when energy is sold in either the DA or ID market. The results presented in Fig. 5, are expressed as a percentage change relative to the DA benchmark. They show that strategies incorporating a stopping rule are effective in increasing profit per trade, yielding income gains of up to 26% and 12% in Periods A and B, respectively. Moreover, the revenue per trade never falls below the DA strategy level. This outcome demonstrates that curtailment-based strategies are successful in mitigating the risk of losses.

Trading based on curtailment strategies involves temporary shutdowns of generation. The trading frequencies for the analyzed strategies, along with the Oracle benchmark, are presented in Fig. 6. Conditional on the profit quantile used as the threshold in the stopping rule, trading occurs in 84–100% of hours. The optimal trading frequency implied by the Oracle strategy is lower in Period A than in Period B, which aligns with the previous findings indicating that the former period is associated with a higher risk of losses.

In Period A, the highest average profits are achieved for $\tau$ values around 0.3, corresponding to a trading frequency of approximately 94–95%. Thus, curtailing generation during the 5–6% most risky hours increases profits by up to 10% relative to the DA benchmark and by 6% compared with the counterpart without curtailment. In Period B, the best performance is obtained for $\tau$ values close to 0.7, corresponding to a trading frequency of about 99%. Hence, the utility can slightly reduce production without incurring any financial losses. These results indicate that curtailment strategies based on profit quantiles represent a rational and profitable market behavior.

### 5.2.1. Portfolio analysis

Finally, we take a closer look at the trading portfolios selected by the various strategies. The data-driven approaches discussed above are characterized by different levels of profit and associated risk. As with other commodities, contracts offering higher income are typically accompanied by greater uncertainty.

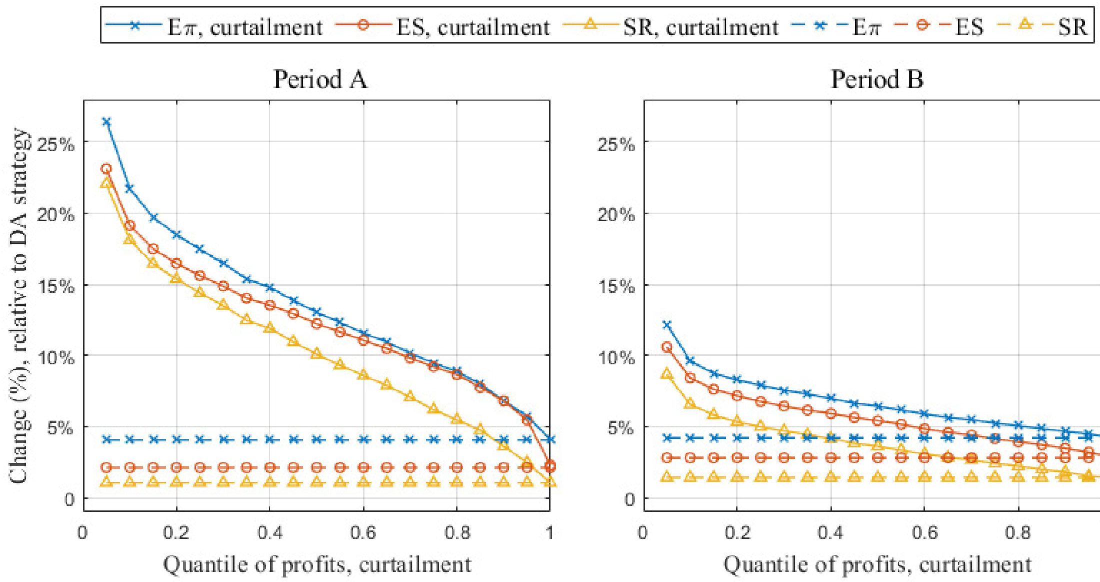


**Figure 5.** Average profits per trade, percentage change relative to DA strategy.

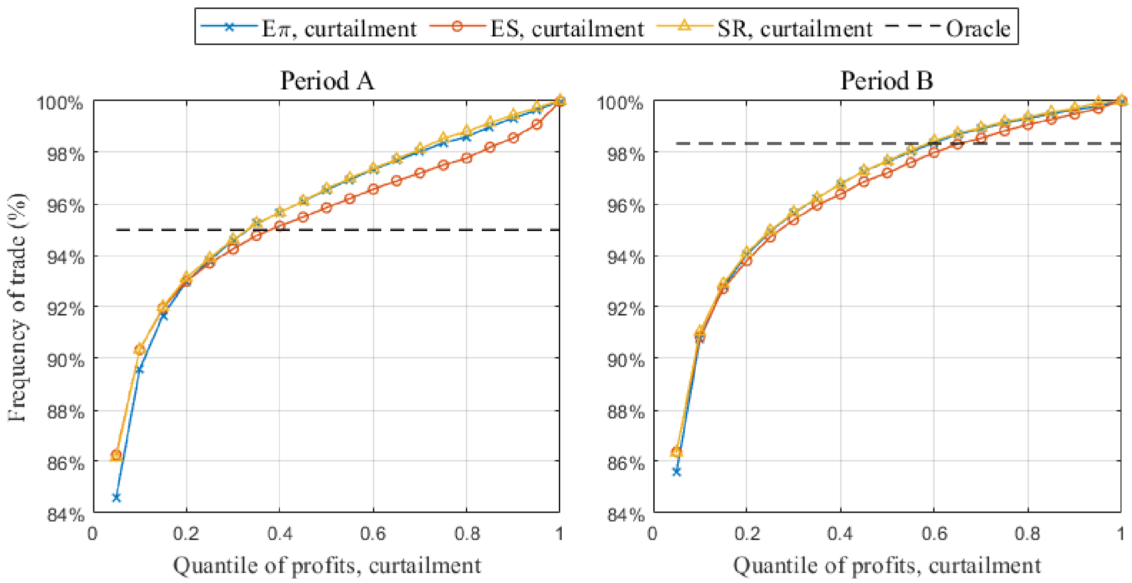


**Figure 6.** Frequency of trade, strategies with curtailment

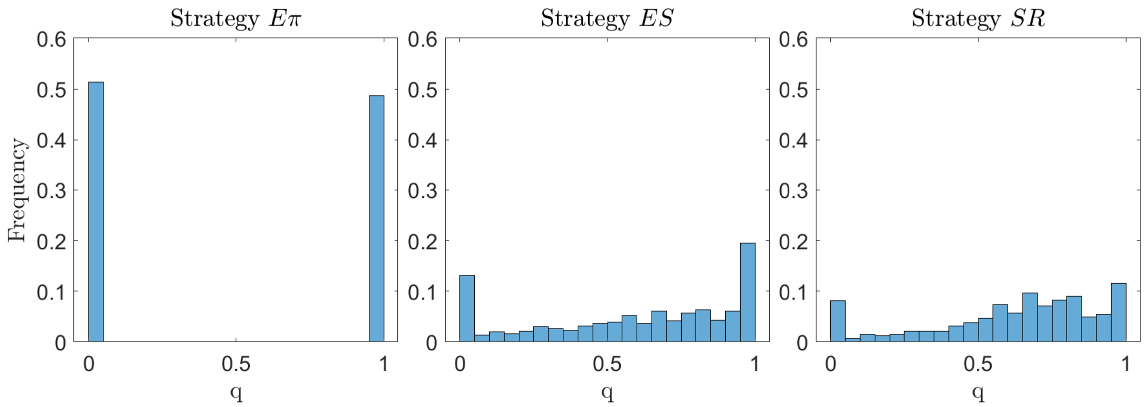


**Figure 7.** Histogram of $q_{t,h}$ selected by data-driven strategies: $E\pi$, $ES$ and $SR$.

Therefore, the choice of strategy depends on the utility owners' risk appetite. To make an informed decision, it is also useful to understand the source of the observed differences.

Figure 7 presents histograms of the decision variable $q_{t,h}$ for strategies without generation curtailment. The results reveal notable differences in the recommended trading patterns. The strategy that maximizes expected profit tends to select only extreme values of $q_{t,h}$, allocating either the entire forecasted generation to the DA market or leaving all production to the ID market. Moreover, it chooses the ID market in 51.3% of cases and the DA market in 48.7

In contrast, the $SR$ strategy produces the most diversified portfolios and frequently recommends selling electricity in both markets. It selects $q_{t,h} = 0$ or $q_{t,h} = 1$ in only 7.6% and 8.9% of cases, respectively. Most of the time, more than half of the predicted generation is offered in the DA market (with $q_{t,h} \geq 0.5$ occurring in 73.7% of cases), and in 30.9% of periods, the strategy recommends values of $q_{t,h} \geq 0.8$.

As observed earlier, the $ES$ strategy lies between the $E\pi$ and $SR$ approaches. It selects intermediate values of $q_{t,h}$ in more than 70% of cases and relatively often recommends allocating the entire generation to the DA market, approximately 16.2% of the time. Overall, in approximately 65% of cases, at least half of the produced electricity is sold in the DA market.

In conclusion, strategies that explicitly account for market risk, such as $ES$ and $SR$, tend to recommend trading patterns resembling those of active market participants, who typically favor the DA market while limiting exposure to the real-time ID market.

# 6. Conclusions and discussion

This study develops and evaluates a new multidimensional ensemble forecasting approach – **the multiple split (MS)** method – and demonstrates its practical relevance for short-term electricity trading. The method extends resampling-based inference to a multivariate framework, enabling the generation of joint probabilistic forecasts for electricity prices and market fundamentals. Results obtained for the German market show that the proposed approach yields well-calibrated prediction intervals and outperforms benchmark methods such as the Quantile Regression Machine and conformal prediction in terms of empirical coverage. Moreover, for variables exhibiting more stable and predictable behavior, such as load and renewable generation, the MS method produces sharper predictive distributions compared with the benchmark approaches.

The forecasts are subsequently applied to a decision-making problem faced by a small renewable energy utility operating in the day-ahead (DA) and intraday (ID) markets. Specifically, the utility decides on the share of predicted generation, $q$, to offer in the DA market. To balance its position, the remaining portion of production is either sold or purchased in the ID market. The company operates under uncertainty, knowing neither the next day's generation nor future electricity prices.

Two naïve strategies are considered: the DA strategy, in which the entire predicted generation is offered in the DA market, and the ID strategy, which allocates all production to the ID market. These are compared with data-driven approaches that utilize probabilistic forecasts of profits. The optimal share $q$ is selected by maximizing the expected profit, expected shortfall, or Sharpe ratio. Additionally, a stopping rule is introduced to allow generation curtailment, ensuring that the utility sells electricity only when a chosen profit quantile is positive.

The results can be summarized as follows:

- The DA and ID strategies are outperformed by data-driven approaches in terms of both profitability and risk control.
- Implementing data-driven strategies without generation curtailment increases profits by 1–4%, relative to the DA benchmark, while simultaneously reducing risk.
- Strategies incorporating generation curtailment based on profit quantiles are particularly effective, increasing revenues by up to 10% and reducing exposure to losses.
- The choice among data-driven strategies depends on the utility's risk preferences. The strategy that maximizes expected profit yields the highest income but exhibits higher risk (lower ES and VaR values), whereas strategies based on ES or SR provide a more balanced trade-off between profit and risk.

The findings underscore the **economic value of ensemble forecasts** in energy markets. By providing a coherent representation of uncertainty and dependence across prices and generation, the MS predictions are easily transformed into probabilistic forecasts of profits, enabling market participants to make better-informed trading and curtailment decisions. In particular, the approach allows for the estimation and comparison of expected shortfall across alternative trading strategies – an assessment that is not directly attainable using conventional density forecasting methods.

The results encourage further research on properties and applications that go beyond the presented analysis. First, it would be interesting to develop a rule to choose the number and proportion of splits. Here, we show the results of only two scenarios MS(1) and MS(20) in which these quantities were arbitrarily selected. A more comprehensive analysis in this area is recommended. Next, since the size of the prediction intervals does not change much over time, one could consider applying locally weighted methods to condition the length of the PI on the fluctuating market situation. This can increase the sharpness of the distribution and improve the statistical performance of the method. Finally, MS can be combined with point forecasting methods other than ARX, for example machine learning. Due to a simple construction and flexibility in selecting the number of splits, it can be used together with methods that are computationally burdensome.

# Funding

This work was supported by the Ministry of Education and Science (MEiN, Poland) through Diamond Grant No. 0027/DIA/2020/49 (to W.N.) and the National Science Center (NCN, Poland) through SONATA BIS grant No. 2019/34/E/HS4/00060 (to K.M).


# References


[1] Acerbi, C., and Tasche, D. On the coherence of expected shortfall. *Journal of Banking & Finance 26*, 7 (2002), 1487–1503.

[2] Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. Predictive inference with the jackknife+. *The Annals of Statistics 49*, 1 (2021), 486–507.

[3] Billé, A. G., Gianfreda, A., Del Grosso, F., and Ravazzolo, F. Forecasting electricity prices with expert, linear, and nonlinear models. *International Journal of Forecasting 39*, 2 (2023), 570–586.

[4] Costa, Á. M., Orosa, J. A., Vergara, D., and Fernández-Arias, P. New tendencies in wind energy operation and maintenance. *Applied Sciences 11*, 4 (2021), 1386.

[5] Durante, F., Gianfreda, A., Ravazzolo, F., and Rossini, L. A multivariate dependence analysis for electricity prices, demand and renewable energy sources. *Information Sciences 590* (2022), 74–89.

[6] Gneiting, T., Balabdaoui, F., and Raftery, A. E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 69*, 2 (2007), 243–268.

[7] Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST 17*, 2 (2008), 211–235.

[8] Janczura, J., and Wójcik, E. 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* (2022), 106015.

[9] Kath, C., and Ziel, F. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting 37*, 2 (2021), 777–799.

[10] Koenker, R., and Hallock, K. F. Quantile regression. *Journal of Economic Perspectives 15*, 4 (2001), 143–156.

[11] Kraft, E., Russo, M., Keles, D., and Bertsch, V. Stochastic optimization of trading strategies in sequential electricity markets *European Journal of Operational Research 308*, 1 (2023), 400–421.

[12] Kumbartzky, N., Schacht, M., Schulz, K., and Werners, B. Optimal operation of a CHP plant participating in the German electricity balancing and day-ahead spot market. *European Journal of Operational Research 261*, 1 (2017), 390–404.

[13] Kupiec, P. H. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives 3*, 2 (1995), 73–84.

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

[15] Lee, D., Shin, H., and Baldick, R. Bivariate probabilistic wind power and real-time price forecasting and their applications to wind power bidding strategy development. *IEEE Transactions on Power Systems 33*, 6 (2018), 6087–6097.

[16] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. Distribution-free predictive inference for regression. *Journal of the American Statistical Association 113*, 523 (2018), 1094–1111.

[17] Lipiecki, A., Uniejewski, B., and Weron, R. Postprocessing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression. *Energy Economics 139* (2024), 107934.

[18] Liu, B., Nowotarski, J., Hong, T., and Weron, R. Probabilistic load forecasting via Quantile Regression Averaging on sister forecasts. *IEEE Transactions on Smart Grid 8*, 2 (2017), 730–737.

[19] Maciejowska, K. Portfolio management of a small RES utility with a structural vector autoregressive model of electricity markets in Germany. *Operations Research and Decisions 32*, 4 (2022), 75–90.

[20] Maciejowska, K., Nitka, W., and Weron, T. Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Economics 99* (2021), 105273.

[21] Marcjasz, G., Narajewski, M., Weron, R., and Ziel, F. Distributional neural networks for electricity price forecasting. *Energy Economics 125* (2023), 106843.

[22] Muğaloğlu, E., Kiliç, E., Keskin, H., and Buğra Selçuklu, S. Dynamics of electricity price volatility and its impacts on energy investments. *Renewable Energy 256*, Part F (2026), 124383.

[23] Nowotarski, J., and Weron, R. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics 30*, 3 (2015), 791–803.

[24] Nowotarski, J., and Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews 81*, Part 1 (2018), 1548–1568.

[25] O'Connor, C., Bahloul, M., Rossi, R., Prestwich, S., and Visentin, A. Conformal prediction for electricity price forecasting in the day-ahead and real-time balancing market. *Energy and AI 21* (2025), 100571.

[26] Pinson, P. Wind energy: Forecasting challenges for its operational management. *Statistical Science 28*, 4 (2013), 564–585.

[27] Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., and Jiang, Z. Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews 144* (2021), 110886.

[28] Serafin, T., Uniejewski, B., and Weron, R. Averaging predictive distributions across calibration windows for day-ahead electricity price forecasting. *Energies 12*, 13 (2019), 2561.

[29] Serafin, T., and Weron, R. Loss functions in regression models: Impact on profits and risk in day-ahead electricity trading. *Energy Economics 148* (2025), 108596.

[30] Shao, J., and Wu, C. F. J. A general theory for jackknife variance estimation. *The Annals of Statistics 17*, 3 (1989), 1176–1197.

[31] Sharpe, W. F. The Sharpe ratio. *The Journal of Portfolio Management 21*, 1 (1994), 49–58.

[32] Toubeau, J.-F., Nguyen, T.-H., Khaloie, H., Wang, Y., and Vallée, F. Forecast-driven stochastic scheduling of a virtual power plant in energy and reserve markets. *IEEE Systems Journal 16*, 4 (2022), 5212–5223.

[33] Uniejewski, B., and Weron, R. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics 95* (2021), 105121.

[34] Uniejewski, B., Weron, R., and Ziel, F. Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems 33* (2018), 2219–2229.

[35] Wiser, R., Bolinger, M., and Lantz, E. Assessing wind power operating costs in the United States: Results from a survey of wind industry experts. *Renewable Energy Focus 30* (2019), 46–57.