Technical Report 1: ARIMA Model for Forecasting ISEM DA Prices

February 13, 2025

1 Introduction

Why ARIMA for ISEM DA Price?

For PowerNI's forecast, the ISEM Day-Ahead (DA) Price is influenced by both short-term fluctuations and longer-term trends. An ARIMA model is a strong foundational choice because it addresses three key aspects of time series data:

1. **Autocorrelation:** The AR (Auto-Regressive) component captures the relationship between current and past values. The parameter p indicates how many past values are used.

2. **Stationarity:** To meet the requirement of constant mean and variance (stationarity), the model can difference the data as needed. The parameter d specifies how many differencing steps are applied.

3. **Error Shocks:** The MA (Moving Average) component models the influence of past error shocks, with the parameter q showing how many past error terms are included.

Intuitively, an ARIMA(p, d, q) model is structured as follows:

$$\underbrace{\nabla^d y_t}_{\text{Differenced Series}} = c + \underbrace{\sum_{i=1}^{p} \phi_i \, \nabla^d y_{t-i}}_{\text{AR terms}} + \underbrace{\varepsilon_t + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j}}_{\text{MA terms}},$$

where:

- $\nabla^d = (1 - L)^d$ is the differencing operator (with L as the lag operator),
- c is a constant,
- ϕ_i are the coefficients for the AR terms,
- θ_j are the coefficients for the MA terms, and
- ε_t represents white noise.

In simpler terms:

- **p** tells us how many past values (autoregressive terms) are used,
- **d** indicates the number of times we difference the data to remove trends, and
- **q** tells us how many past error terms (moving-average terms) are included.

For PowerNI, after training the ARIMA model on the ISEM DA Price data, the best fit was found to be ARIMA$(1, 0, 1)$:

$$p = 1, \quad d = 0, \quad q = 1.$$

This means no differencing was needed (i.e., the data was already stationary), and the model uses one AR term (ϕ_1) along with one MA term (θ_1). This outcome was validated through relevant statistical tests.

Ultimately, while the ARIMA model is a powerful tool for forecasting, it should be viewed as one component in a trader's toolkit to help PowerNI make informed decisions about energy trading and risk management.

2 Data and Code Overview

I used the `auto_arima` model (from the `pmdarima` Python library) to automatically fit ARIMA parameters. The entire codebase can be accessed at following GitHub repository [1]:

2.1 Directory Structure

The project structure of the GitHub Repo is:

```
powerni/
|-- README.md
|-- initial_analysis.ipynb
`-- forecasting/
    `-- ARIMA/
        |-- arima_model.ipynb
        `-- differencing.ipynb
```

We are interested in the notebooks within powerni/forecasting/ARIMA/.. :
differencing.ipynb:

- Handles data preprocessing for ARIMA (ensuring stationarity).
- Loads raw ISEM DA Price data.
- Runs Augmented Dickey-Fuller (ADF) tests for stationarity.
- Applies up to second-order differencing if needed.
- Saves original and differenced data; creates diagnostic plots.

arima_model.ipynb:

- Implements the ARIMA modeling using `auto_arima`.
- Splits data into 80% training, 20% testing.
- Fits the ARIMA model and generates predictions.
- Validates forecasts via:
  - Residual analysis and stationarity tests,
  - RMSE, MAE metrics,
  - Q-Q plots and histograms for residual normality.

2.2 Train–Test Split

We then split the data into 80% training and 20% testing:

```
train_size = int(len(df) * 0.8)
train = df.iloc[:train_size]
test = df.iloc[train_size:]
```

2.3 Fitting the Model

To automatically search for the best (p, d, q) combination, we employ a brute-force approach using `pmdarima`'s `auto_arima`, which is computationally feasible in this context:

```
import pmdarima as pm

model = pm.auto_arima(
    train['ISEM DA Price'],
    start_p=0, max_p=5,
    start_q=0, max_q=5,
    start_d=0, max_d=2,
    test='adf',
    seasonal=False,
    trace=True,
    stepwise=True
)

print("Best ARIMA order:", model.order)
```

Additionally, we manually verified that the original series was already stationary by running the Augmented Dickey-Fuller (ADF) test on the *ISEM DA Price* data in the **differencing.ipynb** notebook. The tests returned a sufficiently negative test statistic (e.g., -4.3580) and a p-value $\ll 0.05$, indicating we could safely set $d = 0$ (no differencing needed).
The final output from `auto_arima` indicates that ARIMA$(1, 0, 1)$ is the best fit:

- $\mathbf{p = 1}$: One past value (lag 1) influences the current value,
- $\mathbf{d = 0}$: No differencing is required,
- $\mathbf{q = 1}$: One lag of the error term is included.

2.4 Generating Predictions

Once the model is fit, we forecast the test set length:

```
predictions = model.predict(n_periods=len(test))
predictions = pd.Series(predictions, index=test.index)
```

3 Findings and Validation

3.1 Model Forecast vs. Actual

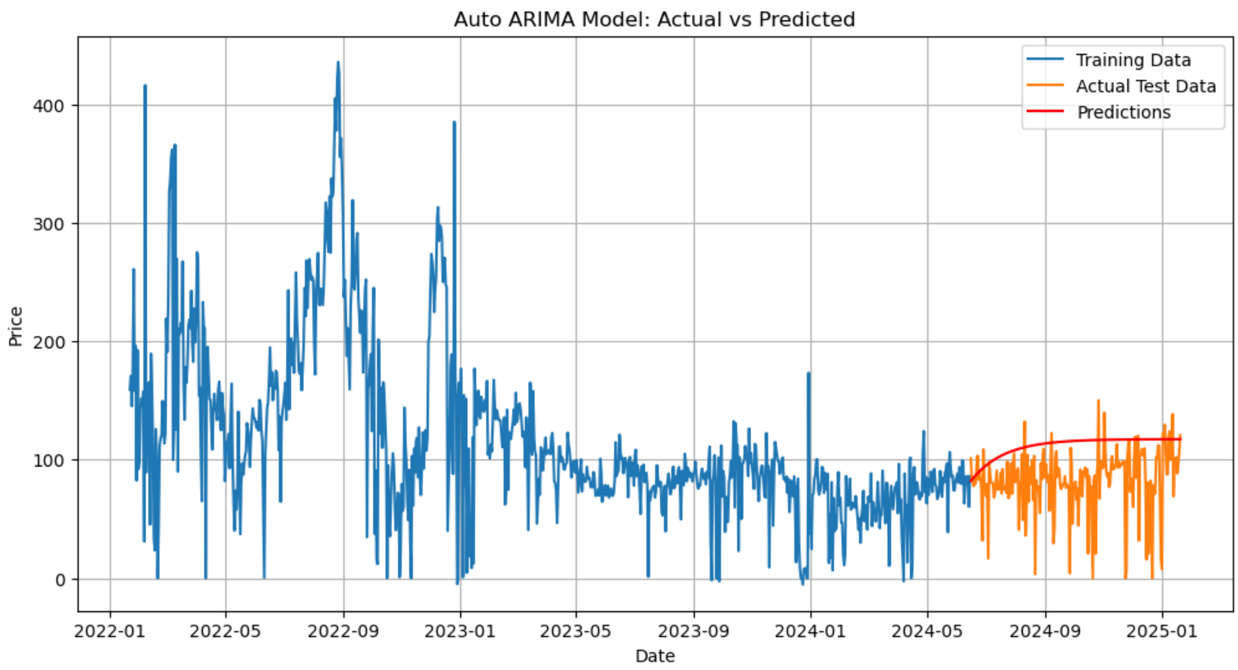

Figure 1: (Fig.1) ARIMA Model: Actual (blue/orange) vs. Predicted (red).

Explanation: Predicted forecast (red) follows the test set (orange) decently overall, indicating the model captures main trends. However does not capture the dips.

3.2 ACF and PACF of Residuals

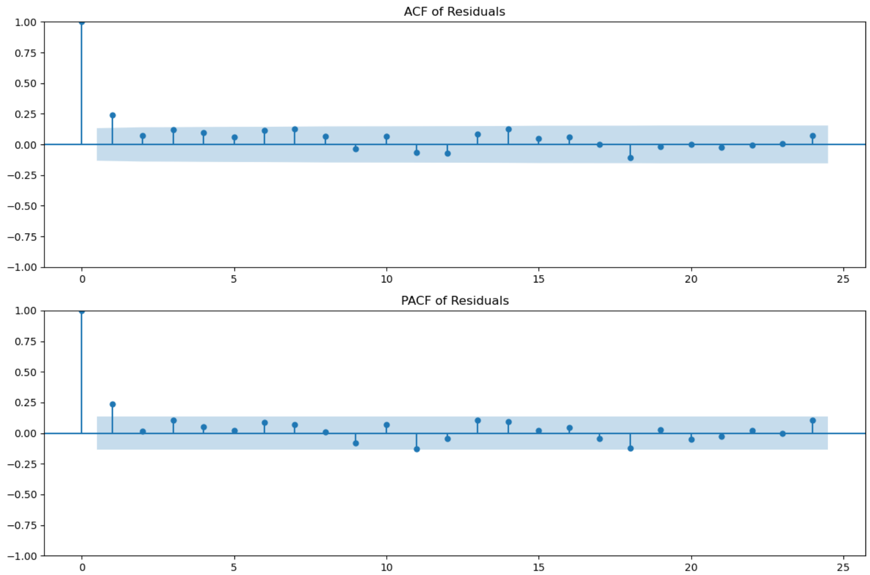

Figure 2: (Fig.2) ACF (left) and PACF (right) of the residuals.

Explanation: No strong autocorrelation peaks; i.e. suggests residuals behave like white noise. This is the ideal.

3.3 Q-Q Plot of Residuals

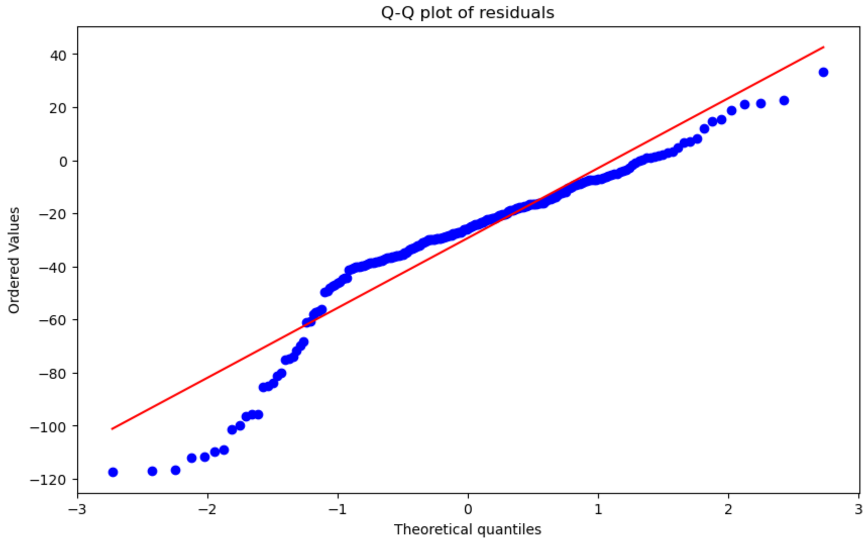

Figure 3: (Fig.3) Q-Q plot of the residuals.

Explanation: Residuals are near-normal, but slight deviation in tails is visible. This is not a major concern.

3.4 Error Metrics

Finally, and importantly we assess the accuracy of the model's metrics using the

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_t - \hat{y}_t)^2}, \quad \text{MAE} = \frac{1}{n}\sum|y_t - \hat{y}_t|,$$

The results obtained are:

- $RMSE \approx 40$
- $MAE \approx 31$

These values suggest moderate forecasting accuracy. In real terms, this means we are obtaining an average error of about 31 euros for the ISEM DA Price forecast.

4 Conclusion

After traning, the model ARIMA$(1, 0, 1)$ was stated to be the best fit to model forecasted prices. Residual analysis confirms no significant patterns remain that aren't explained by our model. While RMSE and MAE indicate moderate accuracy and visual inspection shows that the model doesnt explain the lower ISEM DA price fluctuations, the model can still be a valuable tool for trading decisions in a trader's toolkit. Feedback on additional validation and comparison with internal models would be appreciated.

References

[1] GitHub, *PowerNI Forecasting Models*, 2025. Available at: `https://github.com/jyoutir/powerni`