Germany Unplugged: Forecasting Grid Congestion to Reduce Welfare Losses

Team 30

April 12, 2025

Abstract

Due to the complex dynamics of the electricity market, grid congestion is hard to prevent and therefore produces large economic costs and welfare losses. In this paper, we propose two methodologies to forecast network congestion and rationalize the role of the transmission system operator through the lenses of a toy model. After investigating the principal components that drive redispatches, we run a time series analysis using an INAR(1) model which accounts for the count nature of the data and provides an interpretable structure for forecasting. As an alternative specification, we consider a fully flexible Recurrent Neural Network (RNN) model that allows to account for the unbalancedness of the congestion frequency attaining high forecasting accuracy (0.94). RNN outperforms INAR(1), as it better captures the true negative congestion events in the test data and therefore potentially reduces the costs related to readjusting energy production. Finally, we run a counterfactual analysis in which we artificially introduce a 75% cap on the price of electricity and study the effects on congestion. We find that introducing this policy would significantly increase the frequency of redispatches.

1 Introduction

The cost of grid congestion management in the European Union reached 4.2 billion euros in 2023[1]. Given that renewable energy, whose production is highly location-specific and volatile, becomes more and more integrated in the European energy supply, this figure is likely to grow in the next few years (Staffell and Pfenninger 2018; Schmidt and Staffell 2023). As an example, the increasing integration of photovoltaic (PV) and wind power into the German electricity grid, imposes new challenges in congestion management or redispatch. We argue that accurately forecasting congestion

[1]This figure comes from the European Union Agency for the Cooperation of Energy Regulators website

reduces the hidden cost of redispatches, leading to a more efficient allocation of economic resources and improving European welfare (Titz, Pütz, and Witthaut 2024; Zhou, Tesfatsion, and Liu 2011).

This paper focuses on forecasting congestion in the German electricity grid operated by Tennet DE, the largest Transmission System Operator (TSO) in Germany as of 2024[2].

We start our analysis by plotting the series of cumulative dispatches over the period January 2020-July 2023 and employing a Principal Component Analysis (PCA) to reduce the dataset's dimensionality into four easily interpretable components. We find that the main principal components are related to: (1) prices of oil, futures on carbon emissions and other commodities; (2) - (3) production of electricity from renewable sources in the area operated by Tennet DE and neighboring areas; (4) one day-ahead forecasts of energy production from different sources (Section 2.2). However, PCA is unsuited for forecasting and policy evaluation due to its purely decompositional nature.

To guide our empirical analysis, we rationalize the problem of the TSO as the one of a social planner that seeks at maximizing the efficiency of the energy supply network while reducing negative externalities due to grid congestion. The model is illustrated in Section 3. In the model, producers depart from standard perfect competition as they are allowed to have different cost functions. In our framework, the TSO effectively observes and integrates in a welfare function the costs of electricity transportation and redispatches. The TSO aims at minimizing the welfare costs related to redispatches and misalignment between energy needs and production. Eventually, the main goal of the planner is reducing prediction errors both for negative and positive events.

To forecast grid congestion, we employ two state-of-the-art techniques (see Sections 4 and 5): (1) a INAR(1) model with Poisson-distributed innovations, a time-series count model that can account for seasonality in the data; (2) a Recurrent Neural Network (RNN) model whose loss function accounts for the unbalancedness of the data (redispatches are more likely than no-redispatches in a given hour), and allows for full flexibility in the relation between covariates and the probability of congestion. We repeat the estimation for any congestion direction, for only *up* redispatches and only *down* redispatches. We find high accuracy for both models (0.93% and 0.92% for INAR and RNN respectively, when evaluated on general congestion). However, INAR performs much worse than RNN in specificity (0.78% against 0.94% respectively, when evaluated on the whole sample), a metric that our theoretical model hints to be relevant. Furthermore, we benchmark these two specifications against a naive forecast model (Giannone, Reichlin, and Sala 2004). As expected, we find that both models vastly outperform the benchmark with the RNN actually scoring 0.90 in an index from 0 to 1 that measures improvement on the naive forecast.

Guided by our model, in which prices do not fully incorporate information, we propose a coun-

[2]The figure comes from Tennet DE own's website

terfactual analysis in Section 6. We analyze a price cap at the 75^{th} percentile on gas and electricity prices with our INAR(1) model. We find that this policy increases the number of necessary upward redispatches. A price cap reduces the informative nature of the prices and thus leads consumers and producers to underestimate the aggregate future demand. When, the day after, the realization of consumed electricity is higher than the forecast the TSO is obliged to integrate additional energy into the grid, using an upward redispatches. Given the infra-marginal design of the electricity market, the necessity for more electricity may be satisfied by firms that are using fossil fuels for production, since these commodities have higher marginal costs than renewable sources. Thus, given our model in Section 3 and counterfactual estimation in Section 6 we argue that price caps reduce prices' information content and increase fossil fuels usage.

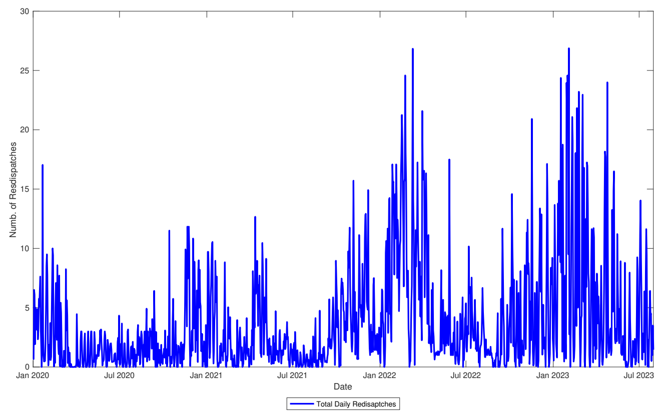
The paper proceeds as follows: Section 2 describes the data and explains our descriptive analysis; Section 3 deals with our theoretical model and develops economic intuitions on the problem at hand; Section 4 presents the econometric theory at the back of our estimation techniques; Section 5 deals with the results from the INAR(1) and the RNN specifications, benchmarking them against a naive forecast approach; Section 6 describes our counterfactual estimation, hypothesizing a 75% price cap on energy prices; finally, Section 7 concludes.

2 Data

2.1 Main Variables of Interest

The data available for the forecasting task are 62,784 hourly observations from 1st of January 2020 at midnight to the 31 July 2023 at 11 pm. Each hour is recorded twice since both directions of the redispatch correction action (i.e. up and down) are accounted for. The dataset contains 69 variables, which accounts for three main groups relevant in the energy congestion forecasting: regional weather data in Germany, such as wind speed and air temperature; electricity data, including energy consumption and prices; and market data, as price of oil and carbon emissions. Our variable of interest is instead the number of concurrent redispatches, a count variable which shows the amount of energy transferals in each given hour. An optimally planned energy production and consumption would lead to zero redispatches in either directions. As advised, we recode the number of energy transferals as a binary variable, and a clear unbalancedness in the dataset is displayed, with energy redispatches accounting for almost 75% of the dataset. To still consider its count nature, we decide to model it in a INAR(1), checking if any difference in model's performance is present. Before the analysis, we divide the dataset into training set (43,778 observations, from 1st January 2020 to 30th June 2022), validation set for the tuning of the RNN model parameters (8,830 observations, from 1 of July 2022 to 31st of December 2022) and test set to evaluate both models' performances (10,176

Figure 1: Time Series - Redispacthes (full sample)



Notes: *the figure displays the daily time-series of the number of redispatches. Observations are grouped at a 24-hour level and represent the sum of up and down redispatches.*

observations, from 1st of January 2023 to 31st July 2023).

2.2 Descriptive Statistics

Figure 1 displays the daily number of redispatches in the whole period (January 2020 to July 2023). Notice the high seasonality implied by data: in the days after January we consistently find a spike in the number of redispatches. Moreover, 2023 displays a steady increase in the variance of the redispacthes and on their level: we reckon this may be due to the inflationary phenomena that hit the commodity prices in the end of 2022. These phenomena may have limited the ability of the grid operator to accurately forecast electricity demand and supply. Figure 1 points at the fact we need to account for seasonality when forecasting hour-by-hour redispatches: we will do that with an INAR(1) model and by the means of a rolling window in our RNN approach (see Section 4).

Our INAR(1) process, specified in section 4, requires the data to be stationary. We check the stationarity of the displacements time series through an Augmented Dickey-Fuller test (ADF) (Elliott, Rothenberg, and Stock 1992) and rejected the null hypothesis of a unit root in the data with a null p-value. Moreover, we find evidence of cointegrating relations between the commodity and electricity prices pointing towards the fact that these markets are highly connected and influence each other in the determination of prices.

We employ a principal component analysis to reduce the data dimensionality on the correlation matrix of the variables in the training period (January 1, 2020 to June 30, 2022). Although different rules have been proposed in the literature to optimally select the number of components to isolate (Jolliffe 2002), we employ three hard set and well-established conditions: (i.) the eigenvalue λ_i for Principal Component i must be greater or equal than 1; (ii.) the cumulative variance explained by the last chosen component j must be at least 70%; (iii.) there must be a sharp kink in the value of the eigenvalues:

$$\lambda_j - \lambda_{j+1} < \lambda_i - \lambda_{i+1}, \quad \forall i \quad \text{where} \quad i < j$$

This criterion is consistent with the standard selective criteria in factor model selection (Ahn and Horenstein 2013) and is met by the first four component. We do not report the loads due to the high number of variables. However, consider that the 1^{st} component can be easily interpretable as the price factor of the data, with high loads on oil, carbon, electricity and natural gas prices. The 2^{nd} and 3^{rd} components summarize information on weather conditions, with both having high weights on sunshine duration, wind speed and air temperature. The difference lies in the fact that the former weights these conditions in the area under the TSO of interest while the latter has similar weights on neighboring German states. The 4^{th} component is that representing the forecasts, a day-ahead, of the production of electricity both total and from renewable sources.

Table 1 presents the Pearson's correlation between our outcome variable, the number of concurrent dispatches on the grid, and our components. In light of our model (Section 3) we expect to see high correlations of redispatches with prices and production of renewable energy, as detailed in Equation 3.3. This is indeed what we observed in the PCA-reduced data: correlations with energy prices are relatively high, since they are unable to absorb externalities, and renewable energy production, since they have a major effect on how electricity is priced (see Equation 3.2). Table 1 points exaclty at this. Moreover, the correlation of displacement with one-day-ahead forecasts of production can reasonably be considered zero: our forecast models aim at filling this gap by optimizing grid management.

Table 1: Principal Components across Different Samples

	Full Sample	Sample - Downs	Sample - Ups
PC-1 (Prices)	0.31	0.35	0.26
PC-2 (Grid Renewable Production)	0.13	0.15	0.10
PC-3 (Bordering Renewable Production)	0.43	0.41	0.47
PC-4 (One Day Ahead Forecasts)	-0.02	-0.03	-0.02

Note: Pearson's correlation between number of displacements by hour and first nine principal components. PCA ran on the training set (01 Jan 2020 to 30 Jun 2022). Due to the PCA loadings we interpret PC-1 as commodity and energy prices; PC-2 and PC-3 as production of renewable energy from weather related events; PC-4 as one day-ahead forecasts at the disposal of the TSO. The procedure was repeated for the whole sample (whether a redispatch happened), only for *downward* and only for *upward* redispatches.

3 Model

We model a non competitive energy market in which each producer maximizes its own profit, but does not internalize grid operating costs such as moving energy and operating displacements. Order the producers by ascending marginal cost, the price is uniquely defined on the market as

$$p \text{ such that } \pi_j = 0, \text{ where } j \Rightarrow D(p) = S(p) \tag{3.1}$$

Equation 3.1 implies that the market price is determined by the marginal producer: producers face different cost functions, reflecting whether for example they produce from renewable sources, and the price is determined by the producer that satisfies the marginal demand. Each producer maximises its own profit as in Equation 3.2.

$$\pi_i = pq_i - c(q_i) \Rightarrow p = \frac{\partial c(q_i)}{\partial q_i} \geq 0 \ \forall i \geq j \tag{3.2}$$

Equation 3.2 states that all infra-marginal producers receive positive profits from their operation. However, neither the producers nor the consumers are able to identify two kinds of externalities. In this set up, the grid operator works as a social planner maximizing Equation 3.3.

$$\max \mathbb{U} = D(p) - [v(p, q_F) + \lambda w(d, \tilde{q}_t) + \delta z(\alpha x_t^+ + (1-\alpha)x_t^-)] \tag{3.3}$$

where $D(p)$ is the benefit consumers derive from electricity; $v(p, q_F)$ is the cost of producing electricity as internalized by the firms (notice we let the quantity depend on q_F, the total production of energy from renewable sources which have very low marginal costs); the two externalities:

- $w(d, \tilde{q}_t)$ represents the *societal cost* the grid must sustain to transfer energy, both as connections and power lost due to traveled distances. This is a function of the geographical distribution of producers and consumers d and the load-gap (either in excess or in shortage) in the grid $\tilde{q}$.
- $z(\alpha x_t^+ + (1-\alpha)x_t^-)$ which represents the cost society bears when the grid controller must operate redispatches in order to restore power balance between demand and supply. Dispatches are allowed to be *upward* or *downward* and entail different costs in these two instances. Moreover, the weight α associated to the upward redispatch entails the asymmetric importance the TSO can assign to the directions of adjustments. These weights can be calibrated based on the social planner's preferences.

The dimensionality of this problem can be reduced if we focus sequentially on different parts of the objective function. Indeed, the v function can be considered as exogenous given the (quasi) inelastic energy demand and the inability of producers to switch from a source of production to another within the time span of the model. The two externalities behave differently as the quantities on which they depend move in the very short term and possibly within the hour. The w function could be minimized by correctly predicting the energy needs of the economy and, hence, reducing the misalignment between the actual load and the optimal one (i.e., $\tilde{q}_t$ close to 0). Moreover, this burden could be reduced by reducing the distance between production and consumption sites in the region/country. Of course the optimal placing of these sites could be investigated with an analysis of the weather patterns and by extending the network according to the specificity of the regions.

However, we want to deduct some simple insights into the effect of redispatches on the overall societal cost. So, the focus of the all problem could be re-conducted to the minimization of some loss function related to $z(\cdot)$. If we assume that $\frac{\partial z(\cdot)}{\partial x_t^+} > 0$ and the same goes with downward redispatches x_t^- then minimizing the argument is enough to minimize the whole function.

$$\mathcal{L} = \alpha x_t^+ + (1-\alpha)x_t^- \quad \text{with } \alpha \in [0,1] \tag{3.4}$$

If we manage to correctly predict upward and downward dispatches then we should be able to anticipate them and reduce their amount in the following hours reducing, in turn, the organizational cost associated to any of these adjustments.

However, we want to have accurate predictions in both cases, i.e. when dispatches happen and when the network is working as planned That is the case because our model suggests that the total welfare cost depends also on the efficiency of the network and hence correctly anticipating the network congestion is crucial to improve efficiency of the operations in the energy distribution while minimizing externalities.

4 Empirical Methodology

4.1 Empirical Specifications

4.1.1 Naive Approach - Baseline Benchmark

To ease the comparison between methods and to get a sense of their actual performance, we first consider the naive forecasting approach (Gürses-Tran, Flamme, and Monti 2020). This model's results will work as a baseline for more sophisticated models. We thus forecast our series as:

$$\hat{y}_{t+h|t} = y_{t+h-24} \tag{4.1}$$

that is we assume that the best predictor oh hour t is the same hour of the day 24-hours before.

4.1.2 INAR(1) with Poisson-distributed Innovations

We model the number of displacements as a first-order integer valued autoregressive process (INAR(1)) process with Poisson-distributed innovations, as defined in Equation 4.2.

$$Y_t = \varphi Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim \text{Poisson}(\lambda) \tag{4.2}$$

This implies that

$$\varepsilon_t = \frac{\lambda^k * e^{-\lambda}}{k!} \tag{4.3}$$

with k being the count of redispatches happened within the hour.

In other words, displacement at time t are autoregressive in an AR(1) process at the innovations follow a Poisson process for which we assume that its mean $\lambda = \beta X'$, where X is our set of controls[3]. The β coefficients capture the effect of the controls we include in the ε_t residual part on the outcome variable. We estimate the model by conditional maximum likelihood, since it has been shown in Monte-Carlo simulation to score better than CLS (Conditional Least Square) and YW (Yule–Walker approach) estimation methods both in terms of empirical bias and mean-square error (Bourguignon et al. 2016).

In practice, we employ Conditional Maximum Likelihood to first estimate β. The first step of the estimation process allows us to retrieve the β and hence the parameter λ of the Poisson residuals ε_t. After that, again using CML algorithm we estimate the autoregressive parameter φ which we then use to compute the one-step-ahead forecast.

$$\hat{Y}_t = \hat{\varphi} Y_{t-1} + \mu_\varepsilon \tag{4.4}$$

or eventually the k-step-ahead forecast

$$\hat{Y}_t = \hat{\varphi}^k Y_{t-1} + \frac{1-\varphi^k}{1-\varphi}\mu_\varepsilon \tag{4.5}$$

where

$$\mu_\varepsilon = \mathbb{E}_t[\varepsilon_t] \tag{4.6}$$

With this model we expect to explain some of the autocorrelation of the outcome variable with

[3]Including commodity prices, energy prices and weather conditions (at the national level)

its first lag and model the residual part with some of the covariates we have in the dataset. We employ the covariates that are more correlated with the PCA components and indeed we confirm that prices and wind conditions are the most relevant regressors to explain the variance of the error term structure. This simple process (Al-Osh and Alzaid 1987) has been shown to be able to account for seasonality trends in the data (Bourguignon et al. 2016), thus being a good *a priori* fit for our problem at hand.

Since we are interested in whether a displacement occurs rather than the number of displacements we then discretize the data by applying

$$\begin{cases} \tilde{y}_t = 0 & \text{if} \quad \hat{y}_t = 0 \\ \tilde{y}_t = 1 & \text{if} \quad \hat{y}_t \geq 0 \end{cases}$$

The INAR(1) model with Poisson innovations guarantees an interpretable benchmark: we lose in accuracy and forecasting power but gain in interpretability. Since it allows for the estimation of the β mapping, we chose to employ this model in Section 6, where we study a partial equilibrium with a price cap on electricity and focus on its effects on displacements.

However, given the computational constraints and the complexity of the class of INAR models, we acknowledge the simplifications we had to apply to this specification and, hence, we do not rely only on this approach to accurately forecast the one-day ahead redispatches. We could explore more features of this class of models by accurately selecting the optimal number of lags to be included and put some more structure on the modeling side of the residuals.

4.1.3 Recurrent Neural Network

Given the ultimate goal of avoiding energy displacement and their unnecessary costs for the grid operator, we turned to the state-of-art model in time-series forecasting, i.e. recursive neural network (RNN) (Masini, Medeiros, and Mendes 2021).

As Multilayer Neural Networks (MNN), it takes an input vector of p variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ and it computes an output Y through the activation of each layer $\ell = 1, \ldots, L$, which has K_ℓ activations $A_k^{(\ell)}$, where $k = 1, \ldots K_\ell$ are the units for each layer.

Nonetheless, as pointed out in James et al. (2021), RNN takes into account the time dependency of the data: as opposed to MNN an ordered sequence $\mathbf{X} = X_1, X_2, \ldots, X_L$, where each X_ℓ is a p-dimensional vector, and p represents the cross-sectional dimension and ℓ defined at time t,

$$\mathbf{X}_\ell^{\mathbf{T}} = (X_{\ell 1}^T, X_{\ell 2}^T, \ldots X_{\ell p}^T)$$

Consistent with the previous notation, the number of layers is as many as the elements of the sequence. In our scenario, $p = 76$, as the number of covariates, and $L = 24$, since we preferred a rolling-window approach, compared to an expanding-window one, to account for the weekly seasonality in the data. Moreover, to address the seasonal trends of redispatches, we followed Gürses-Tran, Flamme, and Monti 2020 and one-hot-encoded the hours of the day, as well as months and weekdays.

The RNN hidden layers consist of K units $A_\ell^T = (A_{\ell 1}^T, A_{\ell 2}^T, \dots, A_{\ell K}^T)$, and they take as input both the vector X_ℓ and the activation vector $A_{\ell-1}$ from the previous step in the sequence, such that the model learns recursively,

$$A_{\ell k}^T = g\left(w_{k0} + \sum_{j=1}^{p} w_{kj} X_{\ell j}^T + \sum_{s=1}^{K} u_{ks} A_{\ell-1,s}^T\right)$$

where, in our model, $g(\cdot)$ is an Hyperbolic Tangent Activation Function and $k = 1, \dots, K$ is a parameter to be tuned through the validation set. The output is then computed as:

$$O_\ell = \sigma\left(\beta_0 + \sum_{k=1}^{K} \beta_k A_{\ell k}^T\right)$$

where we selected $\sigma(\cdot)$ to be a linear function, because the sigmoid was not compatible with the chosen weighted loss function, which was used in the validation task to account for the unbalancedness of the data. In particular we consider a simple weight class w defined as:

$$w = \frac{\#\ \text{total}}{\#\text{positives}} - 1$$

Due to computational burden and extended training time, we selected only two parameters to be tuned: number of units in the hidden layers K and the learning rate lr, i.e. how much the model weights w and u are adjusted in each update. The validation was carried out through Grid Search assuming reasonable values for those parameters, i.e. $K = [16, 32, 64]$ and $lr = [10^{-2}, 10^{-3}, 10^{-4}]$, and assessing which set of parameters was minimizing the balance-corrected binary cross-entropy loss function:

$$\mathcal{L}(\hat{y}_t, y_t; w) = -\left[w_1 y_t \log(\hat{y}_t) + w_0 (1 - y_t) \log(1 - \hat{y}_t)\right]$$

and the corresponding minimization problem:

$$\mathcal{L}(\hat{y}_t, y_t; w) = -\left[w_1 \cdot y_t \cdot \log(\hat{y}_t) + w_0 \cdot (1 - y_t) \cdot \log(1 - \hat{y}_t)\right] \tag{4.7}$$

where w are the balancing weights, $\hat{y}(.)$ is the predicted probability for a dispatch to happen at hour t, ρ is the set of relevant parameters for the RNN, and Θ is the set of adminssible values for the collection (K, lr, ρ). We use the validation set to select optimally K and lr. The chosen parameters for our final RNN model are $K = 32$ and $lr = 10^{-2}$.

5 Results

To visualize the performance of our forecast models, we consider two simple graphs. First, the test set daily prediction error that computes the difference between the true share of hours when at least one dispatch happened in each day, and the predicted share considering the test set. Second, the 24-hours forecast of the average day congestion, considering days in the test set. Both these graphs are repeted considering the forecast of downward and upward dispatch separately. In Section 5.1, we present results for the INAR(1) model and in Section 5.2 we present results for the RNN model. In Section 5.3, we compare the performance of the two models with the naive baseline.

5.1 INAR(1) with Poisson Innovations

In Figure 2, we plot the daily prediction error of the INAR(1) model. In Panel A, we consider downward redispatch, and in Panel B, upward redispatch. One key finding we can see in this graph is that for both outcomes the model tends to predict higher share of redispatches as compared to the true data. This is most likely due to the fact that the model does not take into account in any way the low frequency of zero events in the data and therefore tends to over-forecast in days with relatvely many zeros. This explanation is consistent with the fact that the error is most of the time zero or very close to zero and exhibits negative spikes in some particular dates.

In Figure 3, we show the average full-day hour-by-hour forecast in the test set. In Panel A, we consider downward redispatch, and in Panel B, upward redispatch. The outcome variable is the count of dispatches events, not the probability for a dispatch to occur. The dashed black line is the true data, the solid blue line is the forcast, and the confidence bounds are computed as in (Bourguignon et al. 2016) at the 95% level. The forecast is fairly accurate, with slight under estimation during night hours and slight overestimation during light hours. The performance is very similar considering upward and downward redispatches. The mis-specification error that arises from the low frequency of zeros is most likely smoothed away by the averaging procedure.

5.2 Recurrent Neural Network

In Figure 4, we plot the daily prediction error of the RNN model. Notably, the prediction error is more uniformely distributed between above and below the zero line. This is expected given that we are explicitely accounting for the imbalanced nature of the data with the weighting scheme w, and the balance-corrected binary cross-entropy loss function. Nevertheless, there are some error spikes in particular dates while staying on average close to zero. The performance does not vary much considering downward or upward dispatches.

In Figure 6, we plot the hourly prediction error of the RNN model. We show the average full-day hour-by-hour forecast in the test set. The outcome variable is the probability of dispatches event to happen in each hour during the average full-day in the test set. The dashed black line is the true data, the solid blue line is the forecast, and the confidence bounds are computed as the forecast plus or minus 1.96 times the estimated variance of the true data (Gürses-Tran, Flamme, and Monti 2020). We can see that the forecast is fairly accurate, but there is quite a difference between downward and upward dispatches. In particular, night hours are overestimated by approximately 0.10 in the downward dispatch.

5.3 Performance Comparison

To evaluate and compare the performance of our two forecast models, we consider various standard performance metrics evaluated on the test set: accuracy, sensitivity, specificity, and F1 score. Accuracy is defined as the share of correct predictions, sensitivity is the true positive rate, specificity the true negative rate, and the F1 score is the harmonic mean of precision and sensitivity. The AUC measures the area under the Receiver Operating Characteristic (ROC) curve and it reflects the model ability to distinguish among the two classes, which, especially in unbalanced dataset, could lead to a trade-off. In Figure 6, we plot the ROC curve. In the following tables, for each direction, we report each metric across the naive forecast, RNN, and the INAR(1).

As a first model assessment, we can notice from Table 2 that, unsurprisingly, the naive benchmark model is quite accurate in forecasting true positive, while it lacks in the true negative prediction, independently from the considered direction. The reason is to be found in the minority of energy dispatch observations, constituting only one fourth of the dataset, yet it could also signals a rather weak hourly seasonality. Accounting for that, the critical metric to comment on in the model comparison is the specificity, as the other metrics are considerably elevate even for the naive benchmark.

We then compute the relative improvement measure over naive benchmark specificity for both our models. Reassuringly, both of their performance overcome the baseline. However, while the INAR(1) improves it only of 60%, the RNN manages in increasing it of almost 90%, giving a primary insight into the potential preferred model.

As for the other measures, the accuracy is extremely elevate in both models, which is was rather

Table 2: Naive Model Classification Performance Across Directions

Metric	**Upward**	**Downward**	**Any Direction**
Accuracy	0.73	0.73	0.73
Specificity	0.43	0.43	0.42
Sensitivity	0.82	0.82	0.83
F1 Score	0.63	0.62	0.62
RNN Improvement	0.89	0.89	0.90
INAR Improvement	0.58	0.60	0.62

Note: Accuracy, specificity, sensitivity, and F1 Score assess baseline naive model performance. RNN and INAR improvement values represent relative improvements over the naive baseline specificity in each directional category.

unexpected for the INAR(1). A plausible explanation is that it exploits the autocorrelation between redispatches, which is has shown to be extremely high even for a single lag. Hence, even modeling the congestion dynamics only through a Markov chain leads to an accurate forecast. However, its main flaw is the poor specificity compared to the RNN, which is due to its deficiency in capturing correctly underlying correlation dynamics. The RNN instead, through the balance-corrected binary cross-entropy loss function, manages to model correctly also the absence of redispatches. This is confirmed also from the RNN's AUC value, which essentially shows no trade-off between true positive rate and the misclassifications of negative ones.

Given the high level of accuracy and sensitivity we have in the two models, we would rather suggest the implementation of RNN as it would help predicting negative events which are still crucial in the light of our model to reduce the efficiency cost.

6 Structural Counterfactual Analysis

Our economic analysis in Section 3 assumes that prices, not incorporating the welfare loss due to externalities in grid management, have an effect on the number of redispatches operated by the grid manager, our social planner.

In this section, we run counterfactual scenarios using our INAR(1) model with Poisson innovations. We chose not to repeat the exercise with the Recurrent Neural Network framework since it is hardly interpretable. Specifically, we study a 75% price cap on electricity prices. We expect that reducing the volatility of prices by censoring them, that is introducing a price cap, should weaken

Table 3: RNN and INAR Model Performance Across Directions

	RNN			**INAR**		
	Upward	**Downward**	**Any Dir.**	**Upward**	**Downward**	**Any Dir.**
Accuracy	0.94	0.94	0.92	0.93	0.93	0.93
Specificity	0.94	0.94	0.94	0.74	0.76	0.78
Sensitivity	0.92	0.87	0.91	0.91	0.98	0.98
F1 Score	0.91	0.90	0.89	0.95	0.95	0.96
AUC	0.96	0.96	0.97	–	–	–

Note: Metrics reflect classification performance of RNN and INAR models by direction. AUC values are not available for the INAR model in this evaluation.

their signaling power.

Indeed, the results presented in Figure 7 show that the number of predicted upward redispatches is generally higher within the 24h window. The red solid line is the baseline forcast with real data, while the blue line is the estimated counterfactual. We believe this is linked to the production decisions that are mainly based on price levels and marginal costs. Given the reduced signal given by prices, our model - and potentially the real producers - are less likely to predict correctly energy needs and will under-predict it. This implies that the TSO will eventually need to push the production upwards and include also marginal producers into the energy mix.

7 Conclusion

This paper analyzes congestion in the German electricity grid. We start with a Principal Component Analysis, identifying commodity prices and weather conditions as the main drivers of redispatches.

We then develop an economic model entailing both imperfect competition on the energy market and externalities on the producers' side. This model guides the economic interpretation of the econometric results.

We thus forecast the total number of redispatches on the grid, the downward and the upward ones. To do so, we employ a INAR(1) model with Poisson-distributed innovations and a Recurrent Neural Network with reweighting to account for the data unbalancedness. We find very high accuracy for both models but the INAR(1) lags behind in specificity. We test these models against a naive benchmark and prove that they improve on it.

Finally, we show that a price-cap would increase ridespatches, lowering efficiency.

References

Ahn, Seung C and Alex R Horenstein (2013). "Eigenvalue ratio test for the number of factors". In: *Econometrica* 81.3, pp. 1203–1227.

Bourguignon, Marcelo et al. (2016). "A Poisson INAR(1) process with a seasonal structure". In: *Journal of Statistical Computation and Simulation* 86.2, pp. 373–387. DOI: 10.1080/00949655.2015.1015127. eprint: https://doi.org/10.1080/00949655.2015.1015127. URL: https://doi.org/10.1080/00949655.2015.1015127.

Elliott, Graham, Thomas J Rothenberg, and James H Stock (1992). *Efficient tests for an autoregressive unit root.*

Giannone, Domenico, Lucrezia Reichlin, and Luca Sala (2004). "Monetary policy in real time". In: *NBER macroeconomics annual* 19, pp. 161–200.

Gürses-Tran, Gonca, Hendrik Flamme, and Antonello Monti (2020). "Probabilistic load forecasting for day-ahead congestion mitigation". In: *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, pp. 1–6.

James, Gareth et al. (2021). "Deep Learning". In: *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer US, pp. 403–460. ISBN: 978-1-0716-1418-1. DOI: 10.1007/978-1-0716-1418-1_10. URL: https://doi.org/10.1007/978-1-0716-1418-1_10.

Jolliffe, Ian T (2002). "Choosing a subset of principal components or variables". In: *Principal component analysis*, pp. 111–149.

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes (2021). *Machine Learning Advances for Time Series Forecasting*. arXiv: 2012.12802 [econ.EM]. URL: https://arxiv.org/abs/2012.12802.

Al-Osh, Mohamed A and Aus A Alzaid (1987). "First-order integer-valued autoregressive (INAR (1)) process". In: *Journal of Time Series Analysis* 8.3, pp. 261–275.

Schmidt, Oliver and Iain Staffell (2023). *Monetizing energy storage: A toolkit to assess future cost and value*. Oxford University Press.

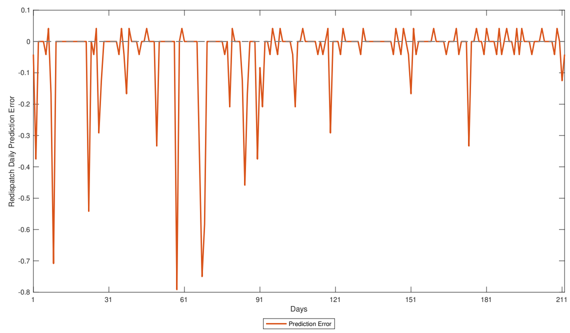
Staffell, Iain and Stefan Pfenninger (2018). "The increasing impact of weather on electricity supply and demand". In: *Energy* 145, pp. 65–78.

Titz, Maurizio, Sebastian Pütz, and Dirk Witthaut (2024). "Identifying drivers and mitigators for congestion and redispatch in the German electric power system with explainable AI". In: *Applied Energy* 356, p. 122351.

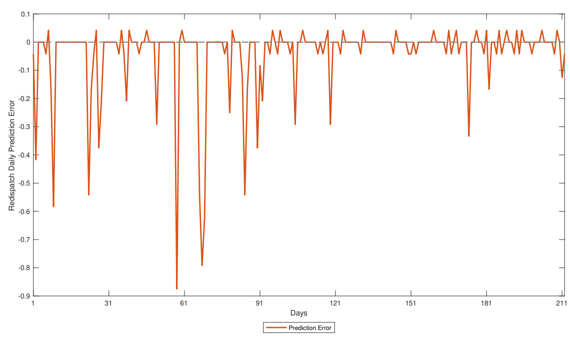
Zhou, Qun, Leigh Tesfatsion, and Chen-Ching Liu (2011). "Short-term congestion forecasting in wholesale power markets". In: *IEEE Transactions on Power Systems* 26.4, pp. 2185–2196.

Figures

Figure 2: INAR(1) - Daily Prediction Error



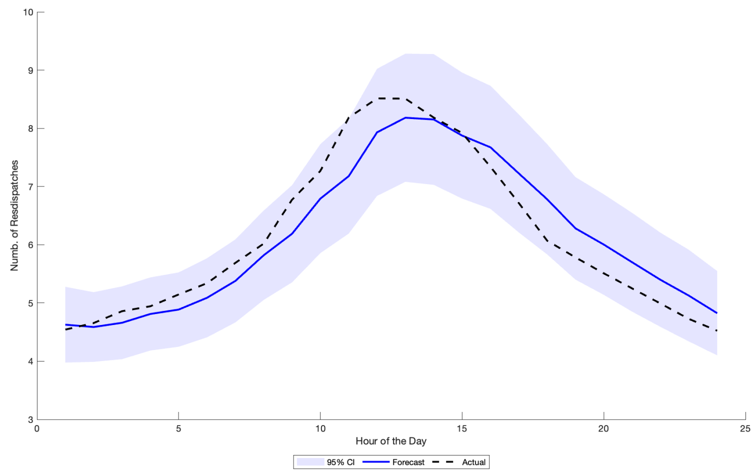
A.Downward Daily Prediction Error



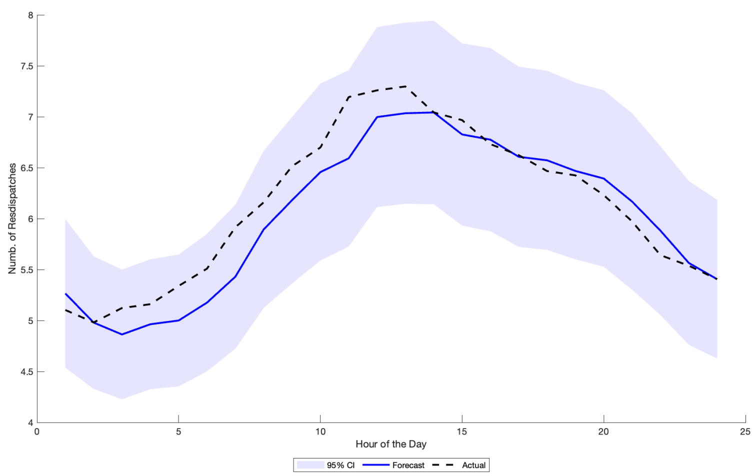
A.Upward Daily Prediction Error

Notes: The figure displays the daily prediction error of the INAR model, that is the difference between the predicted and the actual share of dispatches by day. Panel (A) displays the estimates for downward redispatches; Panel (B) displays them for upward dispatches.

Figure 3: INAR(1) - Hourly Prediction Errors: (a) downward and (b) upward redispatch prediction errors over time.



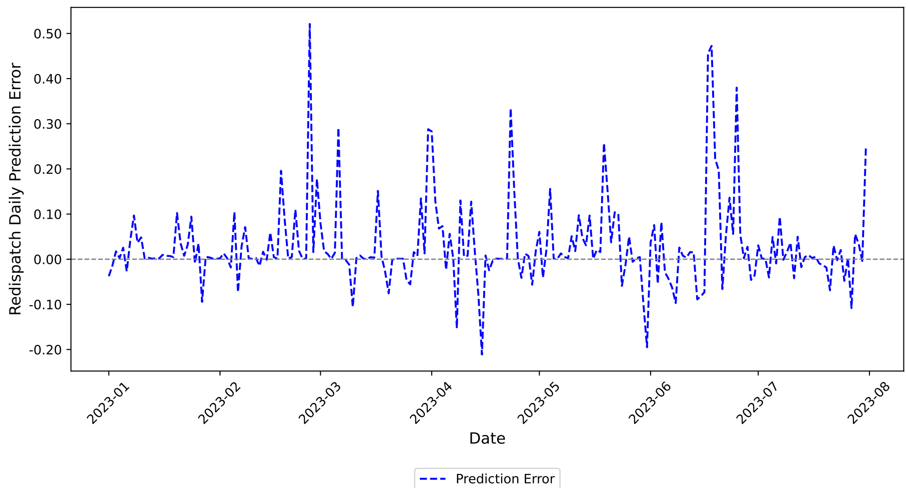
A.Downward Prediction Error



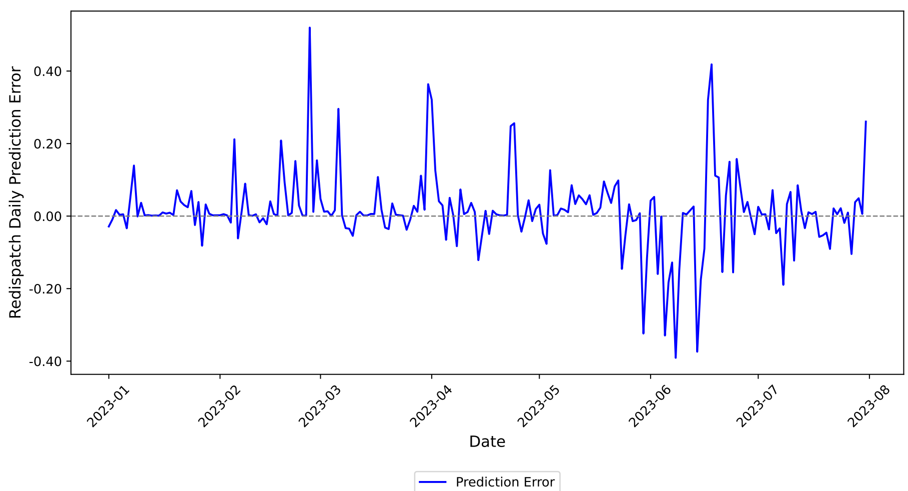
B.Upward Prediction Error

Notes: the figure displays the hourly INAR prediction against true data, that is the predicted and true probability for a dispatch to happen in each hour during the average day of the training set. Panel (A) displays the estimates for downward redispatches; Panel (B) displays them for upward dispatches.

Figure 4: RNN - Daily Prediction Errors: (a) downward and (b) upward redispatch prediction errors over time.



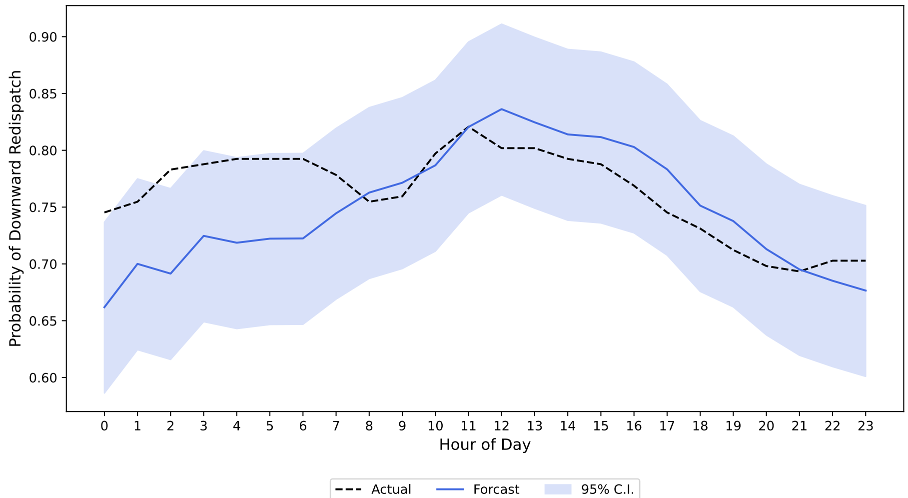
A.Downward Prediction Error



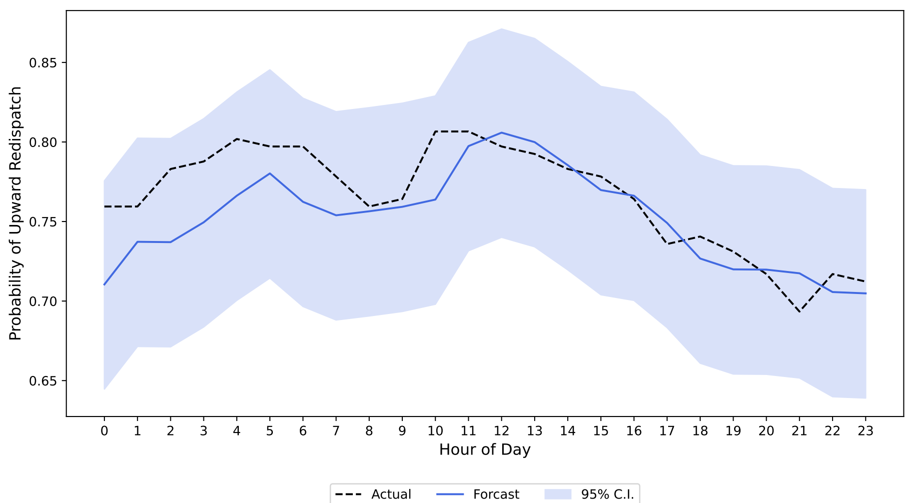
B.Upward Prediction Error

Notes: the figure displays the daily prediction error of the RNN model, that is the difference between the predicted and the actual share of dispatches by day. Panel (A) displays the estimates for downward redispatches; Panel (B) displays them for upward ones.

Figure 5: RNN - Hourly Prediction Errors: (a) downward and (b) upward redispatch prediction errors over time.



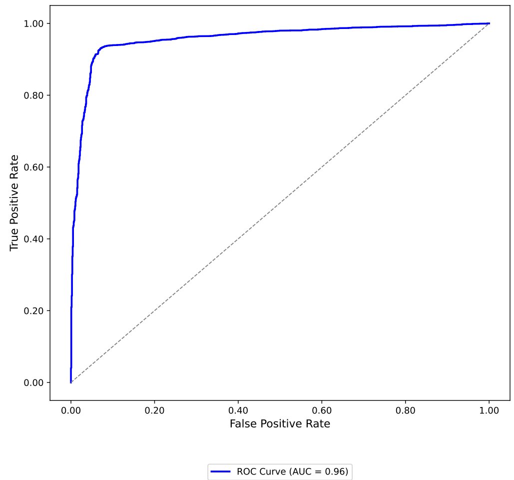
A.Downward Prediction Error



B.Upward Prediction Error

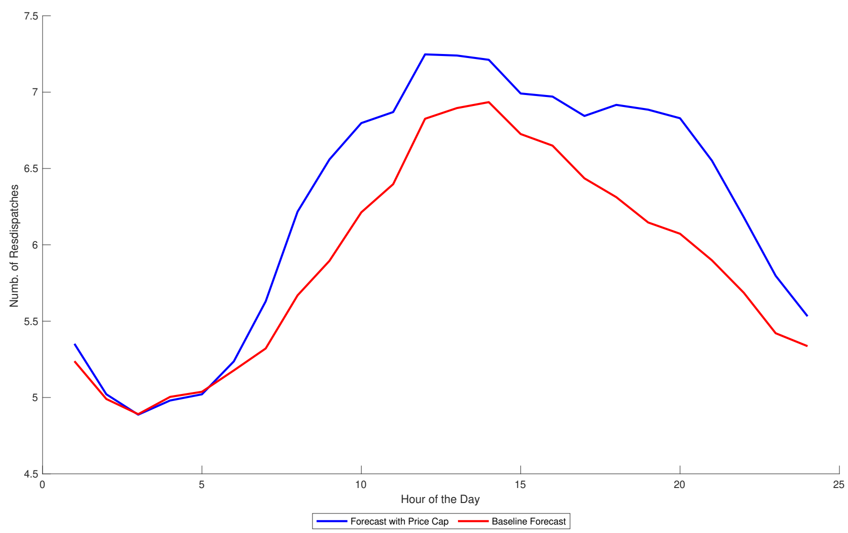
Notes: the figure displays the hourly RNN prediction against true data, that is the predicted and true probability for a dispatch to happen in each hour during the average day of the training set. Panel (A) displays the estimates for downward redispatches; Panel (B) displays them for upward ones.

Figure 6: RNN - ROC curve



Notes: the figure displays ROC curve. Sentivity is on the y-axis while (1 - accuracy) is displayed on the x-axis. The largest the area under the curve the better the forecast. The figure was created on the test set using the forecasted probabilites.

Figure 7: Counterfactual analysis given the introduction of a price cap



Notes: the figure displays our counterfactaul scenario with a 75% price cap. The model used was the INAR(1) with Poisson innovations. The red line represents the basic forecast while the blue line represents the forecast after the introduction of a price cap.