When do Electricity Price Forecasts Create Operational Value? A Regime-Aware Evaluation Case Study

Paulo Batista[1,2(✉)], Eduardo Monteiro[2], Margarida Vieira[2], and Manuel Ferreira[3]

[1] Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
pauloricardolb@ua.pt
[2] INEGI, Campus da FEUP, R. Dr. Roberto Frias 400, 4200-465 Porto, Portugal
[3] Microprocessador, Rua de Cidres, 1444 Matosinhos, Portugal

Abstract. Whether electricity price forecasts generate operational value in industrial settings—beyond improving predictive accuracy—remains an open question, particularly under extended 168-h horizons where forecast skill degrades, and tariff-facing decisions must be made against a plausible fixed-tariff baseline.

This work evaluated a leakage-safe hybrid model of OMIE gross hourly prices in a backtest of 15 audited Portuguese weeks, balanced across low-, average-, and high-volatility regimes (5/5/5), and compared a fixed BAU tariff (0.0829 €/kWh) with a forecast-informed dynamic hourly tariff, under fixed production conditions.

The electricity price forecast is assessed with MAE, RMSE, bias, empirical 90% interval coverage, and Diebold-Mariano tests against naïve lag-24 and lag-168 persistence. Then, the forecast's cost ranking quality was assessed, along with realized proxy-cost gains and downside risk.

The results show the electricity price model with mean MAEs of 24.1, 26.4, and 43.0 €/MWh across low-, average-, and high-volatility weeks, and empirical coverage of 66.4%, 58.5%, and 39.8%.

The model beat lag-24 persistence in all regimes ($p \leq 0.001$) but beat lag-168 persistence only in average-volatility weeks ($p < 0.001$). However, the broad tariff ordering remained stable (Spearman ρ 0.67–0.69, pairwise accuracy 0.747–0.749). Mean realized gains relative to BAU were €559.3, €458.0, and €311.2, with positive outcomes in 60%, 80%, and 80% of weeks, and underperformance probabilities of 40%, 20%, and 20%.

The results indicate a conditional decision value driven more by tariff-order preservation than by point accuracy alone. This study contributes an applied, regime-aware backtesting framework that evaluates forecast value under bounded industrial flexibility—a setting underrepresented in the existing forecast-value literature, which predominantly addresses trading and market-operator contexts.

Keywords: Electricity price forecasting · Dynamic electricity tariffs · Electricity-based decision support system · Demand response

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16768, pp. 669–683, 2027.
https://doi.org/10.1007/978-3-032-30488-9_41

1 Introduction

Industrial electricity purchasing is a non-trivial cost component in many industrial settings, and its managerial importance increases when tariff structures vary over time rather than remaining flat. Under time-varying pricing, the cost consequences of consumption depend not only on total electricity use but also on when that use is billed, so hourly price variability can materially affect industrial cost exposure [1]. In practice, this means that firms facing differentiated hourly tariffs confront a cost environment in which the temporal price structure becomes part of the decision problem rather than a background accounting detail. Such exposure is especially relevant when electricity-intensive processes operate over multiple hours and when procurement conditions can shift materially within a planning horizon. This does not imply that all industrial settings can freely shift demand or that dynamic tariffs necessarily yield savings in every operational setting. It does imply, however, that tariff design and price timing matter for cost management wherever electricity procurement is embedded in production decisions. That relevance helps explain the appeal of electricity-price forecasting for operational use.

Forecasts are often evaluated using point-error criteria because these metrics summarize predictive performance in a compact, comparable way. Yet the literature has noted that predictive skill and downstream decision value are distinct and should not be treated as interchangeable [2]. A model can outperform another one in aggregate error terms while still failing to support better operational decisions, and the reverse can also occur when the information most relevant to a decision is only partially captured by conventional accuracy summaries.

In adjacent energy applications, value-oriented forecasting has likewise been framed as a decision-support problem rather than a pure prediction problem, because a forecast may be operationally useful or useless depending on how its errors interact with the actions it informs [3].

Moreover, in industrial electricity management, the scientific issue is therefore narrower than whether a model reduces forecast error on average. The more relevant question is whether the forecast preserves enough economically useful information to support better tariff-facing decisions than a plausible business-as-usual baseline. This paper addresses that gap through a deliberately narrower design. It evaluates the decision value of a 168h electricity-price forecasting model in an industrial setting by comparing a fixed-tariff Business-As-Usual (BAU) baseline with a forecast-informed dynamic hourly tariff - Policy_Variant_1.

It is important to note that the question must also be evaluated under realistic operational constraints: industrial demand response is not unconstrained, and firms often face technical, organizational, and economic barriers that limit how far power consumption can be adapted to volatile electricity conditions [4]. Even when some flexibility exists, it is typically bounded rather than unlimited, and the benefits of demand response participation depend on the feasible operational adjustment margin rather than on abstract assumptions of full responsiveness [1]. This motivates our real-case study design, selecting a production process in a real-world industrial setting to reproduce a real-world anchored flexibility case.

This work was part of a broader research–industry partnership on the development of an *Intelligent system for dynamic adjustment of electric power consumption*,

led by Microprocessador - Sistemas Digitais, S.A., a tool which orchestrates different electricity-related models, such as a forecast gross hourly electricity price model, forecast hourly electricity price demand models for specified industrial production process, plus local PV solar plant production forecast to self-consumption. This tool will feed a broader industrial management cockpit (including production process planning and optimization) and will be integrated as an additional layer into production plan optimization routines.

2 Related Work

A substantial share of the electricity-price forecasting literature is organized around predictive-error performance, commonly reported using aggregate-level measures such as RMSE and MAE [5, 6]. Those metrics remain useful for judging statistical fit. Still, they do not automatically establish whether a forecast improves downstream operational decisions, because economically consequential errors may not coincide with the forecasts that minimize aggregate point error [5].

From a strict forecast point of view, the literature points out that simple baselines in electricity price forecasting are hard to beat, and decomposition/hybridization can improve accuracy, despite the hybrid/ensemble methods often being among the strongest options, despite most of the studies being focused on day-ahead settings [6–8], but no class wins universally.

Related work on forecast evaluation, therefore, matters not because alternative metrics can substitute directly for decision tests, but because they can reveal performance dimensions that simple aggregate errors may hide [9, 10]. In a decision-focused industrial study, this distinction is central: the relevant question is not only whether prices are predicted closely in level terms, but whether the forecast preserves information useful for a tariff choice relative to a fixed baseline.

Beyond standard forecast-error evaluation, a bounded strand of literature studies the value of forecasts through their economic consequences. The forecast-value literature [11] treats both conceptual and empirical aspects, as forecasts are useful for decision-making rather than as a simple extension of predictive accuracy. In that perspective, a forecast can be statistically better yet operationally unhelpful for a particular decision rule, or statistically imperfect yet still economically useful, depending on how forecast errors interact with the action being taken. That perspective is directly relevant here, but its alignment with the fixed-plan industrial tariff response remains limited, since supporting evidence often comes from neighboring markets or trading settings rather than from the specific operational comparison considered in this paper. Accordingly, the existence of forecast-value literature helps justify decision-oriented evaluation. Still, it does not by itself settle how forecast quality should be assessed against a fixed-tariff industrial baseline.

A further constraint in the adjacent literature concerns assumptions about demand-side responsiveness. In industrial environments, tariff responses are conditioned by technical, organizational, economic, and regulatory barriers, so operational relevance should be framed in terms of bounded flexibility rather than unrestricted responsiveness or optimization [4]. Survey-level work [12] reinforces this caution by showing that industrial

demand response is substantively relevant but uneven in adoption and implementation, making broad assumptions about maneuverability unsafe in the present context. This caution is especially relevant in settings where production continuity, process coupling, and organizational constraints limit how far hourly price signals can be translated into operational adjustment. For that reason, industrial demand-response literature is most useful here when it disciplines realism in the response setting, not when it is treated as if it already validated schedule redesign or a broader decision-support architecture.

Taken together, these literatures motivate a deliberately narrow positioning for the present study: a focus on the empirical gap of whether forecast quality translates into value relative to a fixed-tariff BAU baseline, under the imperfection of a 168h horizon forecast.

The relevant comparison is therefore not forecasting performance in general, but the bounded question of whether a forecast-informed tariff choice can outperform a fixed baseline under a constrained industrial action space. This places this work within forecast-value research, with a clear, applied, and real-world application.

3 Study Scope and Method

3.1 Decision Problem and Policy Scope

This study evaluates an electricity-purchase decision problem using a backtesting framework anchored in shop-floor management, comparing two policy configurations under a common production plan (held fixed for this study): a fixed-tariff business-as-usual baseline (BAU) and a forecast-informed dynamic hourly tariff policy (Policy_Variant_1). The forecast horizon (168h) was selected as a (7-day) scenario, which can be particularly interesting from the point of view of the general industrial production plan management, even though it will pose important technical challenges for any forecast modelling exercise.

The particular case study is based on an opportunistic selection of a real-world operational production process, which is particularly useful for feeding the exercise with demand data informed by real-world use-case conditions. From that specific production process, a fixed production plan was selected for a "typical" (average consumption) week in 2023. This restriction is intended to isolate the contribution of the electricity-price forecast to tariff valuation and policy selection, so that any later performance differences are not confounded with schedule resequencing, load shifting, or broader production re-optimization, which, obviously, will open the door for future research.

The electricity demand associated with the selected production plan is obtained not from real electricity consumption records but from an electricity demand model (outside the scope of this study) to better simulate a real-world use context, where all these models live – the DSS tool mentioned before. This implies that the study's results should be analyzed with caution, as the values will depend on the precision of the demand forecast model. Thus, the results should be considered primarily for comparison rather than for absolute values (in particular in cost analysis). Moreover, it is important to note that the production plan has nearly 20% free time overall, which describes a bounded production flexibility scenario.

Within BAU, the decision signal is the fixed tariff. Within Policy_Variant_1, the decision signal is the point forecast of gross market hourly electricity prices over the 168-h horizon, which informs exposure to the dynamic tariff schedule. The focal comparison is thus restricted to how forecast information changes tariff-facing decisions under fixed operating conditions.

3.2 Electricity-Price Forecasting Model (EPF Model)

The study uses an electricity-price forecaster that generates hourly OMIE gross-price predictions over a 168 h horizon for a specified forecast origin. Model training and inference are restricted to information available up to origin - 1h, and the same pre-origin boundary is preserved in both the backtest exercise (this one) and production mode, incorporated into the DSS models. The design was intended to maintain a leakage-safe multi-step forecasting workflow and turn a possible backtest into a simulation exercise using historical data to estimate past performance, risk, and viability before risking actual capital.

The electric price forecast literature has noted that decomposing prices into structured and stochastic components can be a simple but straightforward strategy to improve forecasts, as sophisticated models tend to struggle to beat the simplest naïve approach [13]. Combining a naive-type forecast with a parameter-rich correction model is an attractive framework due to its simplicity, variance-control architecture by design, and performance [14–16]. Moreover, this framework allows for coupling a versatile set of models to learn from the residuals. Thus, classical or machine learning models, such as LightGBM can be adopted to learn only the deviation from a strong seasonal naive anchor, not the whole price level. This is usually a good idea in electricity markets, where the level is partly seasonal and partly driven by hard-to-predict shocks, and the residual models can be focused on this hard forecast problem.

In this study, a leakage-safe 168-h electricity-price forecaster hybrid model is implemented, with the hourly OMIE gross price (target) decomposed into a repeated lag-24 persistence baseline and a bucket-specific residual term. Residuals are learned with Python *MLForecast* and *lightgbm.LGBMRegressor* on hourly data, short-memory autoregressive residual lags (1, 2, 3, 4, 5, 6, 12), 24 h rolling mean and standard deviation transforms on the first three lags, and one interesting property of the machine learning learner for this case, a set of exogenous features.

Separate residual models are fitted for forecast lead buckets 1–24 h, 25–48 h, and 49–168 h, and their predictions are shrunk by bucket-specific damping coefficients (0.20, 0.20, 0.15) before being added back to the lag-24 baseline. These weights were imposed as a conservative regularization rule, motivated by forecast-combination evidence that simple/shrunk weights can outperform unstable estimated weights [17, 18] and as a first, simple, exploratory specification that can establish a future model development baseline. A drawback is that this specification is structurally designed for smoothness and persistence, which limits the model's capacity to respond to weekend transitions, holidays, renewable ramp events, outages, fuel shocks, scarcity episodes, and negative price regimes.

The exogenous information is included in the residual model learner as calendar, astronomical, renewable-proxy, demand-proxy, and interaction features. Note that the

last three types of features are engineered variables built by combining spatial information (e.g., the geographic locations of power plants, weighted by capacity, and the most relevant demand points, weighted by population) and engineering-based features from weather forecast parameters. The feature engineering procedures try to proxy market fundamentals, such as renewable production availability or consumption patterns. Future-feature construction reuses a persisted transformer bundle generated during training, ensuring the same transformation logic is applied at inference time. These inputs should nevertheless be interpreted as pure operational proxies rather than as a full structural representation of electricity-market fundamentals, but are chosen to proxy those elements in an environment of lack / asymmetric information.

3.3 Evaluation Design and Metrics

The evaluation design uses a 168h forecast horizon, for which a scenario-based protocol follows:

- built around 15 historical weekly forecast origins (forecast issues origin on Monday, 0:00:00)
- a fixed-tariff BAU policy baseline of 0.0829 €/kWh, based on Eurostat [19] electricity prices components for non-household consumers, in 2024, for Portugal, and considering the consumption band ID (2 000 MWh to 19 999 MWh)
- volatility-stratified week audited cases (low – average – high) are chosen from 2025 real market gross hourly OMIE (Portugal) electricity prices data (published by OMIE).

The analysis operates on a balanced 5/5/5 audited weeks with spacing enforced within each bucket (30 days).

The evaluation framework is structured to keep forecast validity, economic relevance, and policy value (including downside risks) separately. Each component of the evaluation design addresses different aspects of decision usefulness and should not be collapsed into a single performance score [20, 21].

Thus, it follows:

Forecast validity is assessed against leakage-safe one-shot persistence comparators at lags 24 and 168, complemented by point-error metrics such as MAE, RMSE, Bias and Diebold-Mariano significance tests [22]. Additional *empirical interval coverage,* defined as the share of observed prices that fall within the model's forecast interval, will also be used. This design remains methodologically consistent with current electricity-price forecasting practice [7, 10] while adapting to this study's objectives.

Economic relevance assumes it depends not only on the level of accuracy but also on the preservation of tariff ordering; this analysis is implemented with *ranking diagnostics metrics* through Spearman's ρ, Kendall's τ, plus pairwise ordering accuracy. Additional *precision at k lags* and *load-weighted cheap-hour capture*, alongside regime summaries and statistical comparisons, were performed to provide a broader view.

Finally, a generalized *policy-value* evaluation was developed based on proxy-based cost comparisons between BAU and Policy_Variant_1, with a downside-risk layer to represent the probability of underperformance, such as the VaR_{95} and $CVaR_{95}$ metrics. Additional semideviation and mean excess cost are calculated to go beyond average

policy-value summaries alone, which do not fully characterize adverse risk exposure [23].

It is important to highlight that the evaluation outputs must be interpreted conservatively, in particular with respect to the cost layer: realized cost remains a proxy rather than field-deployment evidence. Moreover, evidence is limited to 15 audited scenario weeks and should be read as sensitivity-analysis evidence rather than independent-sample inference.

4 Results

4.1 Forecast Validity

Forecast validity was evaluated using weekly point-error metrics, leakage-safe one-shot persistence benchmarks, lead-wise error profiles, and empirical interval coverage. Performance was materially regime-dependent. Low-volatility weeks showed the lowest mean error, whereas high-volatility weeks showed the highest error, the strongest negative bias, and the weakest empirical interval coverage.

Against leakage-safe one-shot 24 h persistence, the forecast model achieved lower squared loss in low-, average-, and high-volatility weeks. The comparison with the leakage-safe one-shot 168 h persistence was more selective: superiority was observed only in average-volatility weeks, whereas low- and high-volatility weeks did not favor the model.

Horizon-wise diagnostics also indicated non-uniform validity across the 168h run, with MAE lowest near lead hour 2 (15 €/MWh) and peaking near hour 57 (56 €/MWh).

Empirical interval coverage remained below the nominal 90% level in all three regimes. These findings are limited to the 15 selected audited weeks (5 per volatility class) and should therefore be read as sensitivity-analysis evidence rather than independent-sample inference (Tables 1, 2 and Fig. 1).

Table 1. Forecast Accuracy by Volatility Regime

Regime	n weeks	MAE (€/MWh)	RMSE (€/MWh)	Bias (€/MWh)	Coverage (%)
Low	5	24.1	29.5	-10.1	66.4
Average	5	26.4	33.2	-10.1	58.5
High	5	43.0	53.9	-38.5	39.8

Regime-level averages of weekly point-accuracy metrics. Coverage is empirical 90% prediction-interval coverage; nominal target is 90%. Negative bias indicates systematic underprediction of actual prices.

Sign convention: DM statistic is computed on (model squared loss − benchmark squared loss). Negative DM → model has lower loss (model preferred). Positive DM → benchmark has lower loss.

Table 2. Diebold–Mariano Tests Against Leakage-Safe Persistence Benchmarks

Regime	ΔLoss lag-24h	DM stat lag-24h	p lag-24h	Favored lag-24h
Low	−40.7	−4.4	0.000	Model
Average	−65.8	−4.6	0.000	Model
High	−50.3	−3.4	0.001	Model
Regime	ΔLoss lag-168h	DM stat lag-168h	p lag-168h	Favored lag-168h
Low	60.6	0.6	0.563	Benchmark
Average	−398.4	−4.0	0.000	Model
High	118.3	0.6	0.540	Benchmar

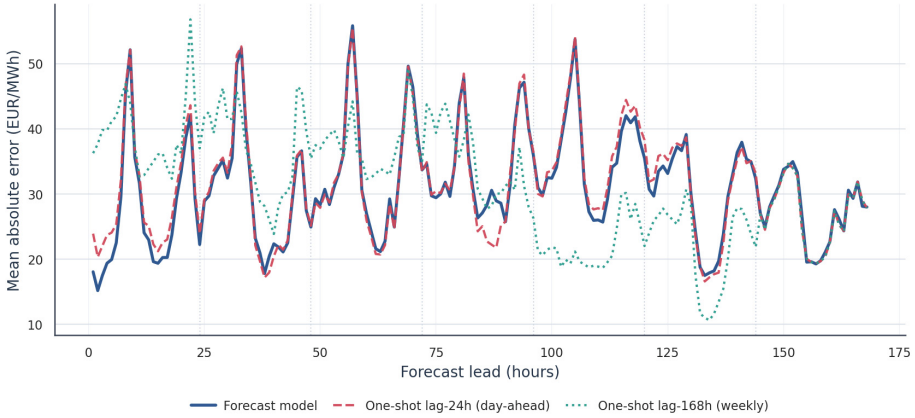

Fig. 1. Lead-Wise Forecast Error vs Benchmarks.

4.2 Ranking Quality

Ranking quality was evaluated through broad rank agreement, top-k cheap-hour precision, and load-weighted cheap-hour capture for the current 168 h run [11]. Broad weekly ordering skill was materially positive and highly similar across volatility regimes, with Spearman ρ around 0.67–0.69, Kendall τ around 0.49–0.50, and pairwise ordering accuracy around 0.747–0.749. By contrast, narrow cheap-hour retrieval was weak at $k = 6$: mean precision@6 was 0.000 in low-volatility weeks, 0.067 in average-volatility weeks, and 0.100 in high-volatility weeks. Retrieval improved as the selected cheap-hour window widened to $k = 12$ and $k = 24$, with the strongest broad-window performance in average-volatility weeks (precision@24 = 0.483; capture@24 = 0.434). These results indicate that the model preserves coarse merit order better than exact hour placement: rank agreement remains positive, but narrow-set retrieval is unstable, and exact placement remains noisy [24] (Fig. 2).

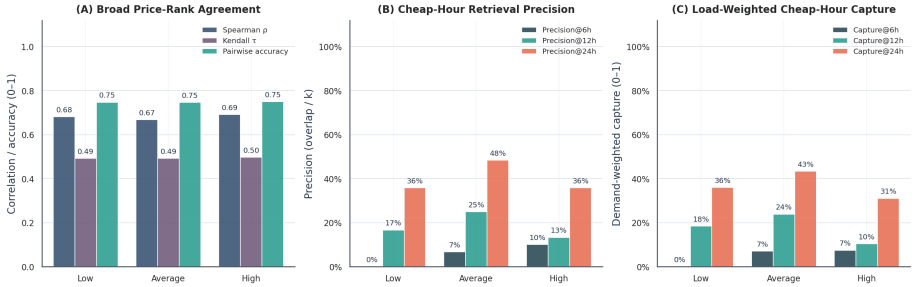

Fig. 2. Composite Tariff-Ordering Quality

4.3 Policy Value

In forecast-cost terms, Policy_Variant_1 yields positive mean gains relative to BAU in all three volatility regimes, with the largest expected gain in high-volatility weeks (1 266 €), followed by low-volatility (753 €) and average-volatility weeks (652 €).

Realized proxy-cost outcomes also remain positive on average in all regimes. Still, they are materially smaller and differently ordered: low-volatility weeks show the highest mean realized gain (559 €), followed by average-volatility (458 €) and high-volatility weeks (311 €).

Decision value is not uniform across scenarios. Positive realized gains occur in 80.0% of average-volatility scenarios, 80.0% of high-volatility scenarios, and 60.0% of low-volatility scenarios, while negative realized outcomes are still observed across all regimes. This dispersion is especially visible in the low-volatility subset, where realized gains range from −1 226 € to 1 662 € in the current selected set (Table 3 and Figs. 3, 4).

Table 3. Realized Policy Value vs BAU by Volatility Regime

Volatility Regime	n scenarios	Mean gain (€)	Median gain (€)	Std dev (€)	Positive gain (%)	Worst case (€)
Low	5	559.3	874.6	1 250.3	60%	−1 226.3
Average	5	458.0	346.4	478.1	80%	−89.7
High	5	311.2	215.5	493.9	80%	−315.3
All regimes	*15*	*442.8*	*346.4*	*769.9*	*73%*	*−1 226.3*

Realized gain = Policy_Variant_1 cost minus BAU cost per scenario week; negative values indicate savings. "Positive gain (%)" is the share of scenarios where Policy_Variant_1 outperformed BAU. "Worst case" is the deepest single-scenario loss. Production plan is fixed; gains arise from tariff-response timing only.

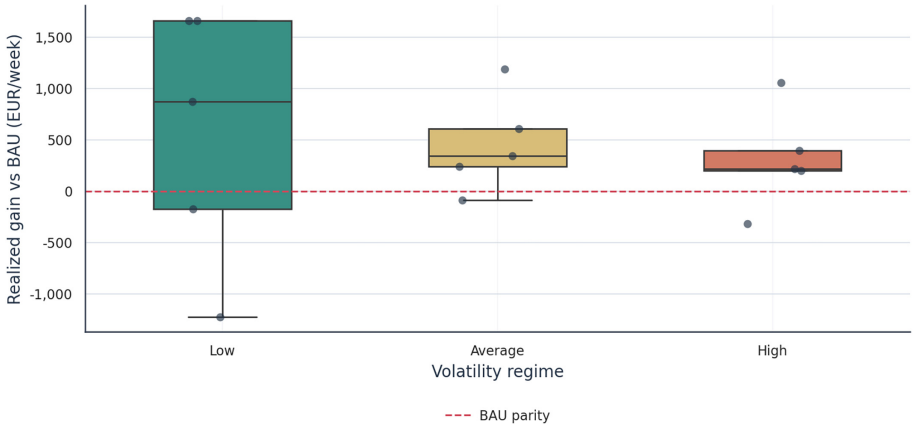

Fig. 3. Realized Gain vs BAU by Volatility Regime

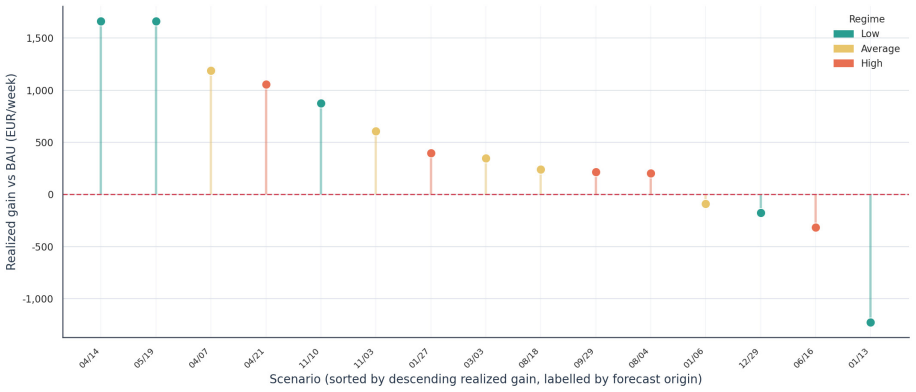

Fig. 4. Scenario-Level Realized Gain vs BAU (Sorted)

Scenarios sorted by descending realized gain across all regimes. Points below the dashed red line represent weeks where Policy_Variant_1 exceeded BAU cost (4 of 15 scenarios).

4.4 Risk and Robustness

Risk and robustness were evaluated using underperformance probability and bootstrap excess-cost tail summaries under the fixed-plan design. The results show that downside risk was non-zero in all three volatility regimes, and the mean probability that Policy_Variant_1 underperformed BAU was highest in low-volatility weeks (0,391), compared with 0,202 in high-volatility and 0,146 in average-volatility scenarios.

Under the pipeline excess-cost sign convention, regime-mean VaR95, CVaR95, and mean excess cost remained negative across all regimes, indicating that the average

selected-scenario tail summaries still favored Policy_Variant_1, even though adverse weeks were not eliminated.

Robustness was nevertheless materially regime-dependent: low-volatility weeks were the most uneven subset, combining the lowest positive realized-gain share (0.60), the deepest realized worst case (−1,226 €), and the widest dispersion in both underperformance probability (std = 0.54) and VaR95 (std = 791 €).

The results show that favorable average gains do not settle robustness in this study: positive regime means can coexist with meaningful reversal risk, so robustness must be assessed with probability, tail, and dispersion evidence together rather than with regime means alone (Table 4 and Fig. 5).

Table 4. Downside Risk by Volatility Regime

Volatility Regime	Underperf. Prob	VaR95 (EUR)	CVaR95 (EUR)	Mean excess cost (EUR)
Low	40%	−471.7	−448.3	−562.1
Average	20%	−367.7	−343.5	−458.5
High	20%	−210.0	−180.8	−311.9

Underperf. prob." is the bootstrap probability that Policy_Variant_1 realized a higher cost than BAU. VaR95 and CVaR95 are the 95th percentile and conditional tail excess cost estimates under the pipeline sign convention; negative values indicate that even the tail favored Policy_Variant_1. Mean excess cost is the regime-mean of the per-scenario excess cost. Production plan is fixed.

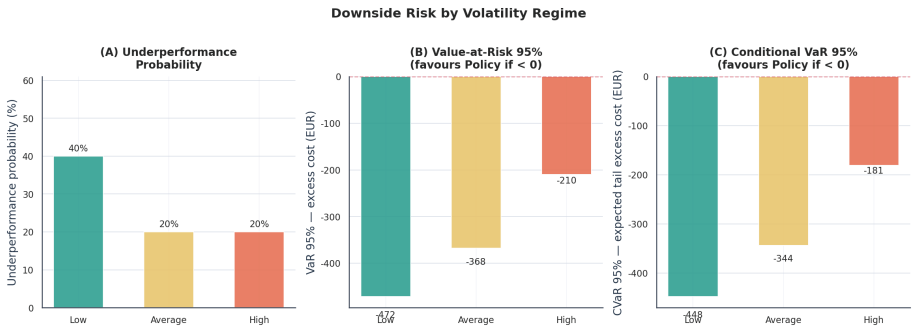

Fig. 5. Downside Risk Composite by Volatility Regime

5 Discussion

The results support that operational usefulness in this setting cannot be inferred solely from point-accuracy summaries. Within the audited fixed-plan comparison, forecast-validity performance was mixed and regime-dependent, including selective superiority over leakage-safe one-shot 168 h persistence and below-nominal interval coverage across regimes.

By contrast, broad tariff-ordering skill remained materially positive even when exact cheap-hour retrieval was weak, especially at $k = 6$.

Read together, those patterns suggest that the decision value observed here is more consistent with preservation of economically relevant tariff ordering than with aggregate error statistics alone. The point is not that ranking information universally dominates point accuracy across all forecasting problems: in this study setting, a forecast can be operationally informative even when conventional error summaries do not, by themselves, provide a sufficient account of usefulness. That interpretation aligns with the literature, which argues that decision-linked forecast evaluation can reveal distinctions that aggregate accuracy measures may obscure, although this argument remains application-specific rather than universal [3, 25].

The value interpretation must therefore remain conditional. In the audited setting, the mean realized proxy gain relative to BAU was positive in all three volatility regimes. Still, negative realized outcomes occurred in each regime, and the probability of underperformance remained non-zero throughout the audited set. This is why average improvement cannot be read as robustness. The evidence instead indicates that volatility shapes both opportunity and downside exposure, while the relationship is not uniform enough to justify a simple "more volatility, more value" claim.

The highest downside frequency in the audited set occurs in low-volatility weeks, reinforcing the analytical distinction between mean gains and tail risk. More broadly, this supports a cautious reading in which forecast-informed tariff response can create value without eliminating the possibility of unfavorable realized outcomes. That reading is consistent with warnings that modeled or target-state performance should not be collapsed into unconditional operational success claims [26].

Several design limits bound these interpretations. First, the evaluation is a proxy-cost sensitivity analysis rather than deployment evidence. Second, it covers 15 selected audited scenario weeks rather than an independent sample designed for broad statistical generalization. Third, the policy scope is narrow: the study compares only BAU and Policy_Variant_1 under an unchanged production plan. Fourth, the results should not be extended to other sites, operating logics, or industrial sectors without additional evidence. Those are internal design facts, not minor caveats.

More generally, the literature on energy-system modelling and electricity-price forecasting cautions against reading model- or data-specific findings as direct evidence of real-world behavior or as broad external validity [8, 27].

Taken together, the discussion supports a narrow practical implication. For this audited fixed-plan application, forecast-informed dynamic tariff response appears conditionally useful when the forecast preserves tariff ordering well enough to improve decisions relative to BAU.

6 Conclusions

In this study, the decision value of the 168h electricity-price forecast is explained more by the preservation of economically relevant tariff ordering than by point error alone under the evaluated fixed-plan BAU versus Policy_Variant_1 setting.

Mean realized proxy gains are positive across the selected audited scenario set, but they are non-uniform across regimes and remain exposed to adverse cases and non-zero downside risk; accordingly, the evidence supports only cautious, context-bound use claims for fixed-plan forecast-informed tariff response.

Holding the production plan fixed is analytically useful in this evaluation design because it isolates tariff valuation under an unchanged schedule. Still, it does not validate joint optimization or full operational adaptation.

The evidence is therefore bound to a realized proxy-cost sensitivity analysis over 15 selected audited weeks and to BAU versus Policy_Variant_1 only, so broader operational validation under richer policies and independent samples is still needed.

Acknowledgments. The authors will thank the industrial partner for the provided dataset (undisclosed due to privacy concerns) on production process features (including overall recorded electricity consumption and the associated production plan.

This work has been supported by the European Union under the Next Generation EU, through a grant of the Portuguese Republic's Recovery and Resilience Plan (RRP) Partnership Agreement, within the scope of the project PRODUTECH R3 – "Agenda Mobilizadora da Fileira das Tecnologias de Produção para a Reindustrialização".

The information provided here reflects only the author's view, which does not necessarily reflect those of the European Union or of the European Commission. Neither the European Union nor the granting authority can be held responsible for them. The European Commission is not responsible for any use that may be made of the information herewith included.

Disclosure of Interests. Manuel Ferreira is an employee of Microprocessador – Sistemas Digitais S.A., which commercializes decision-support system solutions to manage electricity consumption and supported the study with real data on production process consumption. However, the company had no role in the study data analysis, manuscript preparation, or decision to publish.

References

1. Siddiquee, S., Agyeman, K., Bruton, K., Howard, B., O'Sullivan, D.: A data-driven framework for quantifying demand response participation benefit of industrial consumers. IEEE Trans. Ind. Appl. **60**, 2577–2587 (2024). https://doi.org/10.1109/tia.2023.3334218
2. Turner, S., Bennett, J., Robertson, D., Galelli, S.: Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. Hydrol. Earth Syst. Sci. **21**, 4841–4859 (2017). https://doi.org/10.5194/hess-21-4841-2017
3. Carrière, T., Kariniotakis, G.: An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading. IEEE Trans. Smart Grid **10**, 6933–6944 (2019). https://doi.org/10.1109/tsg.2019.2914379
4. Leinauer, C., Schott, P., Fridgen, G., Keller, R., Ollig, P., Weibelzahl, M.: Obstacles to demand response: why industrial companies do not adapt their power consumption to volatile power generation. Energy Policy (2022). https://doi.org/10.1016/j.enpol.2022.112876
5. Maciejowska, K., Lipiecki, A., Uniejewski, B.: Statistical and economic evaluation of forecasts in electricity markets: beyond RMSE and MAE. Energy Convers. Manag. **356**, 121408 (2026). https://doi.org/10.1016/j.enconman.2026.121408
6. Weron, R.: Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int. J. Forecast. **30**(4), 1030–1081 (2014). https://doi.org/10.1016/j.ijforecast.2014.08.008

7. Lago, J., Marcjasz, G., Schutter, B., Weron, R.: Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark, arXiv, vol. abs/2008.08004 (2020). https://doi.org/10.1016/j.apenergy.2021.116983
8. O'Connor, C., Bahloul, M., Prestwich, S., Visentin, A.: A review of electricity price forecasting models in the day-ahead, intra-day, and balancing markets. Energies (2025). https://doi.org/10.3390/en18123097
9. Strøm, E., Gundersen, O.E.: Performance metrics for multi-step forecasting measuring win-loss, seasonal variance and forecast stability: an empirical study. Appl. Intell. **54**, 10490–10515 (2024). https://doi.org/10.1007/s10489-024-05715-4
10. Nowotarski, J., Weron, R.: Recent advances in electricity price forecasting: a review of probabilistic forecasting. Renew. Sustain. Energy Rev. **81**, 1548–1568 (2018). https://doi.org/10.1016/j.rser.2017.05.234
11. Kath, C., Ziel, F.: The value of forecasts: quantifying the economic gains of accurate quarter-hourly electricity price forecasts. Energy Econ. (2018). https://doi.org/10.1016/j.eneco.2018.10.005
12. Shoreh, M., Siano, P., Shafie-khah, M., Loia, V., Catalão, J.: A survey of industrial applications of demand response. Electr. Power Syst. Res. **141**, 31–49 (2016). https://doi.org/10.1016/j.epsr.2016.07.008
13. Xiong, X., Qing, G.: A hybrid day-ahead electricity price forecasting framework based on time series. Energy (2022). https://doi.org/10.1016/j.energy.2022.126099
14. Uniejewski, B., Weron, R.: Efficient forecasting of electricity spot prices with expert and LASSO models. Energies (2018). https://doi.org/10.3390/en11082039
15. Marcjasz, G., Uniejewski, B., Weron, R.: Beating the Naïve—combining LASSO with Naïve intraday electricity price forecasts. Energies (2020). https://doi.org/10.3390/en13071667
16. Billé, A.G., Gianfreda, A., Del Grosso, F., Ravazzolo, F.: Forecasting electricity prices with expert, linear, and nonlinear models. Int. J. Forecast. **39**(2), 570–586 (2023). https://doi.org/10.1016/j.ijforecast.2022.01.003
17. Clemen, R.T.: Combining forecasts: a review and annotated bibliography. Int. J. Forecast. **5**(4), 559–583 (1989). https://doi.org/10.1016/0169-2070(89)90012-5
18. Wang, X., Hyndman, R.J., Li, F., Kang, Y.: Forecast combinations: an over 50-year review. Int. J. Forecast. **39**(4), 1518–1547 (2023). https://doi.org/10.1016/j.ijforecast.2022.11.005
19. Eurostat: Electricity prices components for non-household consumers - annual data (from 2007 onwards), European Commission (2025). https://doi.org/10.2908/nrg_pc_205_c
20. Koutsandreas, D., Spiliotis, E., Petropoulos, F., Assimakopoulos, V.: On the selection of forecasting accuracy measures. J. Oper. Res. Soc. **73**, 937–954 (2021). https://doi.org/10.1080/01605682.2021.1892464
21. Giudici, P., Raffinetti, E.: RGA: a unified measure of predictive accuracy. Adv. Data Anal. Classif. **19**, 67–93 (2024). https://doi.org/10.1007/s11634-023-00574-2
22. Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. J. Bus. Econ. Stat. **20**(1), 134–144 (2002). https://doi.org/10.1198/073500102753410444
23. Xiao, D., Chen, H., Cai, W., Wei, C., Zhao, Z.: Integrated risk measurement and control for stochastic energy trading of a wind storage system in electricity markets. Prot. Control Mod. Power Syst. **8**, 1–11 (2023). https://doi.org/10.1186/s41601-023-00329-3
24. Ziel, F., Weron, R.: Day-ahead electricity price forecasting with high-dimensional structures: univariate vs. multivariate modeling frameworks. Energy Econ. **70**, 396–420 (2018). https://doi.org/10.1016/j.eneco.2017.12.016

25. Vogler, A., Ziel, F.: Event-based evaluation of electricity price ensemble forecasts. Forecasting (2021). https://doi.org/10.3390/forecast4010004
26. Torralba-Díaz, L., et al.: Identification of the efficiency gap by coupling a fundamental electricity market model and an agent-based simulation model. Energies **13**, 3920 (2020). https://doi.org/10.3390/en13153920
27. Trutnevyte, E.: Does cost optimization approximate the real-world energy transition? Energy **106**, 182–193 (2016). https://doi.org/10.1016/j.energy.2016.03.038