Sequential Predictive Conformal Inference: Adaptive Prediction Intervals for Electricity Price Forecasting

Antonio Panico

Energy Finance Christmas Workshop, Wrocław, Poland
December 12, 2025

Department of Engineering for Industrial Systems and Technologies
University of Parma

Antonio Panico
PhD Candidate
University of Parma, Italy

Luigi Grossi
Professor of Economic Statistics
University of Parma

Outline

Introduction to Probabilistic Forecasting

  Literature Review

Methodology

Case Study EPEX & Results

Conclusion

Introduction to Probabilistic Forecasting

Probabilistic Forecasting in Modern Power Markets

UNIVERSITÀ DI PARMA

The Paradigm Shift

The integration of intermittent Renewable Energy Sources (RES) has transformed electricity prices (P_t) from deterministic trajectories into highly volatile stochastic processes. Point forecasts $\mathbb{E}[P_t]$ are no longer sufficient for optimal decision-making.

Core Objectives:

- **Quantify Uncertainty:** Generate Prediction Intervals (PIs) $\hat{C}_t^\alpha$ or full Predictive Densities $\hat{f}(P_t)$.
- **Risk Management:** Essential for derivative pricing, Value-at-Risk (VaR) calculations, and strategic bidding under non-Gaussian noise.

Literature Review: Probabilistic EPF Methods

UNIVERSITÀ DI PARMA

Category	Methodology	Key Reference
Benchmarks	• **Historical Simulation (HS):** Sample quantiles of past errors. • **Distributional:** Gaussian/Student-t/Johnson (S_U) fits. • **Bootstrapping:** Resampling residuals.	[Nowotarski and Weron, 2018]
QRA Family	• **QRA:** Quantile reg. on pool of point forecasts. • **Factor QRA (FQRA):** PCA on point forecasts first. • **Hybrid QRA:** Pre-filtering + Post-processing. • **Smoothing QRA (SQRA):** Kernel-based smoothing.	[Liu et al., 2017] [Maciejowska and Nowotarski, 2016] [Nowotarski and Weron, 2015] [Uniejewski, 2025]
Conformal	• **Inductive CP:** Split-conformal calibration. • **Normalized CP (NCP):** Adapts width to volatility. • **Adaptive CQR:** On-line conformalized NN ensembles.	[Vovk et al., 2005] [Kath and Ziel, 2021] [Romano et al., 2019] [Brusaferri et al., 2025]

Our Contribution: Agnostic Post-Processing

UNIVERSITÀ DI PARMA

The Practical Reality

Real-world decision-makers aggregate **heterogeneous sources** (TSOs, proprietary feeds, black-box ML).

The Limitation: These sources typically provide only **deterministic point forecasts** ($\hat{y}_t$), lacking rigorous uncertainty quantification.

Our Contribution: Agnostic Post-Processing

UNIVERSITÀ DI PARMA

The Practical Reality

Real-world decision-makers aggregate **heterogeneous sources** (TSOs, proprietary feeds, black-box ML).

The Limitation: These sources typically provide only **deterministic point forecasts** ($\hat{y}_t$), lacking rigorous uncertainty quantification.

Our Approach: The SPCI Wrapper

We treat any point forecast as a signal and calibrate its residuals dynamically.

- **Input:** *Any* stream of point forecasts (model-agnostic).
- **Mechanism:** Sequential Predictive Conformal Inference (**SPCI**) adapts to non-stationarity in the error distribution.
- **Output:** Reliable, adaptive prediction intervals for operational use.

Methodology

Classical Approach: Inductive Conformal Prediction (ICP)

UNIVERSITÀ DI PARMA

Definition (ICP): Let $\mathcal{D}_{cal} = \{(X_i, Y_i)\}_{i=1}^{n}$ be a calibration set. Define the non-conformity score as the absolute residual $R_i = |Y_i - \hat{f}(X_i)|$ [Fontana et al., 2023].

The prediction interval $\hat{C}_{1-\alpha}(X_{n+1})$ is constructed as:

$$\hat{C}(X_{n+1}) = \left[\hat{f}(X_{n+1}) \pm \hat{Q}_{1-\alpha}(\{R_i\}_{i\in\mathcal{D}_{cal}})\right] \quad (1)$$

The Limitation (Exchangeability Violation):

- ICP guarantees validity only if data is **exchangeable**: $P(z_1, \ldots, z_n) = P(z_{\pi(1)}, \ldots, z_{\pi(n)})$ [Fontana et al., 2023].
- In time series, residuals ϵ_t exhibit **serial correlation** (e.g., volatility clustering).
- *Result:* A static quantile $\hat{Q}$ fails to adapt to distribution shifts, leading to coverage violations [Barber et al., 2023, Tibshirani et al., 2019].

The General Solution: SPCI Framework

UNIVERSITÀ DI PARMA

The SPCI Hypothesis: Since residuals in time series are not i.i.d., the conditional distribution of the *next* residual ϵ_t is predictable given the filtration of past errors $\mathcal{F}_{t-1}$ [Xu and Xie, 2023].

Algorithm (Sequential Predictive Conformal Inference):

1. **Base Prediction:** Train point predictor $\hat{f}$ to obtain residuals ϵ_t.
2. **Dynamic Quantile Regression:** Instead of a static histogram, train a regressor $\mathcal{Q}$ to forecast the residual quantile:

$$\hat{q}_t^{(\tau)} = \mathcal{Q}\left(\tau \mid \{\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-w}\}\right) \tag{2}$$

3. **Interval Construction:**

$$\hat{C}_t(X_t) = \left[\hat{f}(X_t) + \hat{q}_t^{(\beta^*)}, \quad \hat{f}(X_t) + \hat{q}_t^{(1-\alpha+\beta^*)}\right] \tag{3}$$

DMQ Model Specification

Let τ_{j^*} be a **reference quantile** (e.g., the median). The model defines the entire distribution relative to this anchor using positive distance processes $\eta_{j,t}$ [Catania and Luati, 2023].

The Quantile Process:

$$q_t^{\tau_j} = \begin{cases} q_t^{\tau_{j+1}} - \eta_{j,t} & \text{if } \tau_j < \tau_{j^*} \quad \text{(Lower Tail)} \\ q_t^{\tau_{j^*}} & \text{if } \tau_j = \tau_{j^*} \quad \text{(Reference)} \\ q_t^{\tau_{j-1}} + \eta_{j,t} & \text{if } \tau_j > \tau_{j^*} \quad \text{(Upper Tail)} \end{cases} \tag{4}$$

Advantage 1 (Strict Positivity): The distance is defined as $\eta_{j,t} = \exp(\xi_{j,t})$. Since $\exp(\cdot) > 0$, the quantiles **cannot cross** by construction, solving the primary defect of classical quantile regression.

Evolution of the Parameters

UNIVERSITÀ DI PARMA

The dynamics of the reference level and the distances are governed by **Generalized Autoregressive Score (GAS)** updates.

Reference Quantile Dynamics:

$$q_t^{\tau_{j^*}} = \bar{q}^{\tau_{j^*}}(1-\beta) + \beta q_{t-1}^{\tau_{j^*}} + \alpha u_{t-1}^{\tau_{j^*}} \tag{5}$$

Distance Dynamics:

$$\xi_{j,t} = \bar{\xi}_j(1-\phi) + \phi \xi_{j,t-1} + \gamma u_{t-1}^{\tau_j} \tag{6}$$

- $\theta = (\alpha, \beta, \phi, \gamma)$ are static parameters estimated via the Hogg function.
- u_t is the **forcing variable** (the score), which drives the adaptation of the interval width based on the gradient of the loss.

The Score: Driving Adaptation

UNIVERSITÀ DI PARMA

The "Hit" Variable: The core signal is the deviation of the realization from the expectation:

$$z_{i,t} = \mathbb{I}(y_t \leq q_t^{\tau_i}) - \tau_i \tag{7}$$

The Forcing Variables (The Gradient): Derived from the gradient of the aggregate check loss:

$$u_t^{\tau_j} \propto \frac{\partial}{\partial \xi_{j,t}} \sum_{k=1}^{J} \rho_{\tau_k}(y_t - q_t^{\tau_k}) \quad \text{(Updates Distance/Scale)} \tag{8}$$

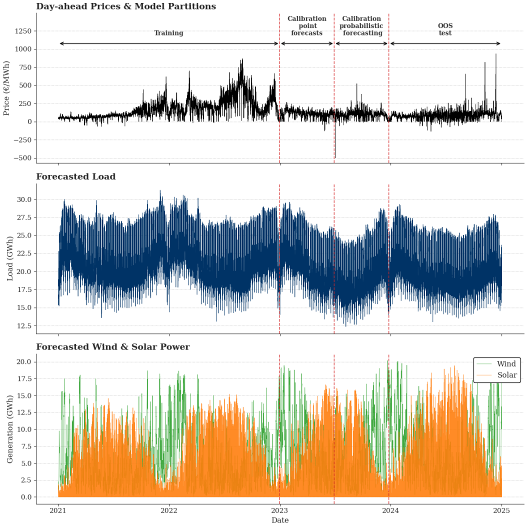
$$u_t^{\tau_{j^*}} \propto \frac{\partial}{\partial q_t^{\tau_{j^*}}} \sum_{k=1}^{J} \rho_{\tau_k}(y_t - q_t^{\tau_k}) \quad \text{(Updates Reference/Location)} \tag{9}$$

Advantage 2 (Adaptivity): DMQ uses these gradients to react instantly. A volatility spike increases $u_t^{\tau_j}$ (expanding widths η), while a structural break in price levels activates $u_t^{\tau_{j^*}}$, shifting the entire distribution.

Case Study EPEX & Results

Experimental Setup: The German Market

UNIVERSITÀ DI PARMA



Experimental Setup: The ARX Model

UNIVERSITÀ DI PARMA

The point forecasts are generated using an **ARX (AutoRegressive with Exogenous inputs)** structure. For a given time step t and transformation $f(\cdot)$:

$$p_{d,h} = \beta_{h,0} + \underbrace{\sum_{i=1}^{7} \beta_{h,i}\, p_{d-i,h}}_{\text{Sum: Previous Days}} + \underbrace{\sum_{j=1}^{G} \gamma_{h,j}\, p_{d-1,h-j}}_{\text{Sum: Previous day hours}} + \underbrace{\sum_{k=2}^{7} \delta_{h,k} D_k}_{\text{Sum: Week Days}} + \underbrace{\beta_{h,L} L_{d,h} + \beta_{h,W} W_{d,h} + \beta_{h,S} S_{d,h}}_{\text{Exogenous: Load, Wind, Solar}} + \varepsilon_{d,h} \tag{10}$$

We employ 5 variance stabilizing transformations—the **Logistic**, **Robust Box-Cox** ($\lambda = 0.5$), **Inverse Hyperbolic Sine** (asinh), **Mirror Log** (mlog, $c = 1/3$), and **N-PIT**—as defined in [Uniejewski et al., 2018] to yield 5 different point forecasts estimated via ridge regression.

Experimental Setup: Ensemble Strategy

UNIVERSITÀ DI PARMA

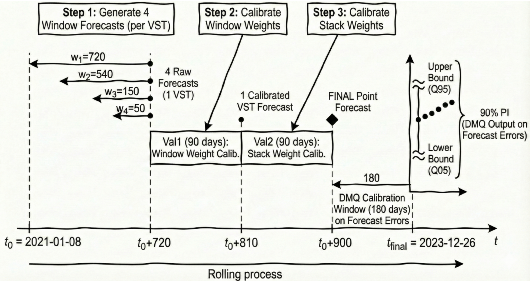
1. Window Calibration (Ridge on $\mathcal{V}_1$)
For each VST model (m), we aggregate forecasts from 4 window lengths (l) [Wang et al., 2023, Hubicka et al., 2019]. Weights w_l are learned via Ridge Regression (minimizing MSE) on recent history $\mathcal{V}_1$:

$$\hat{y}_{t,m}^{calib} = \sum_{l \in L} w_l \cdot \hat{y}_{t,m,l} \quad \text{s.t.} \quad \sum w_l = 1,\ w_l \geq 0$$

---

2. VST Stacking (Median Reg. on $\mathcal{V}_2$)
We combine the 5 calibrated VST models (m) using Quantile Regression ($\tau = 0.5$) on a separate window $\mathcal{V}_2$. This minimizes Absolute Error (MAE), making the forecast robust to outliers:

$$\hat{Y}_t = \sum_{m \in M} \beta_m \cdot \hat{y}_{t,m}^{calib} \quad \left(\beta^* = \underset{\beta}{\text{argmin}} \sum |e_i| \right)$$



Result: *A robust point forecast anchoring our probabilistic methods.*

Performance Evaluation: Benchmarking

UNIVERSITÀ DI PARMA

The Naive Benchmark Definition

Let $P_{d,h}$ be the price on day d at hour h. The naive forecast $\hat{P}^{\text{naive}}_{d,h}$ is defined as:

$$\hat{P}^{\text{naive}}_{d,h} = \begin{cases} P_{d-1,h} & \text{if } d \in \{\text{Tue, Wed, Thu, Fri}\} \quad \text{(Persistence)} \\ P_{d-7,h} & \text{if } d \in \{\text{Sat, Sun, Mon}\} \quad \text{(Weekly Lag)} \end{cases} \tag{11}$$

Metrics Comparison

The table below compares the 5 VST-ARX models and the Stacked ensemble against the Naive benchmark over the test period (2021–2024).

Method	**MAE** (€/MWh)	**RMSE** (€/MWh)	**sMAPE** (%)
Naive Benchmark	30.43	47.50	56.51
VST: Logistic	10.70	17.81	26.42
VST: Robust Box-Cox	8.52	15.56	23.21
VST: Arcsinh	**7.28**	**11.90**	**21.23**
VST: Mirror Log (mlog)	13.99	76.71	26.68
VST: N-PIT	10.41	17.50	26.34
Stacked Ensemble	*8.01*	*14.92*	*21.86*

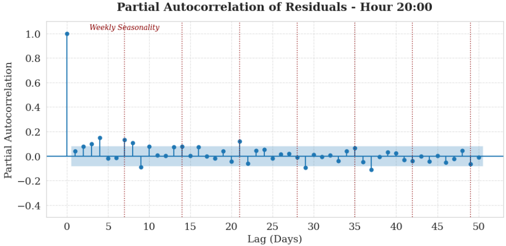
Residual Analysis: Autocorrelation Structure

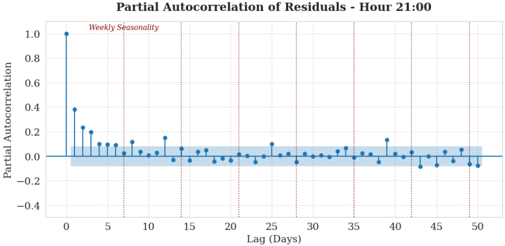
UNIVERSITÀ DI PARMA

We examine the **Partial Autocorrelation Function (PACF)** of the model residuals to assess the quality of the ARX filtration. Ideally, residuals should be white noise (no significant lags).

Key Observations:

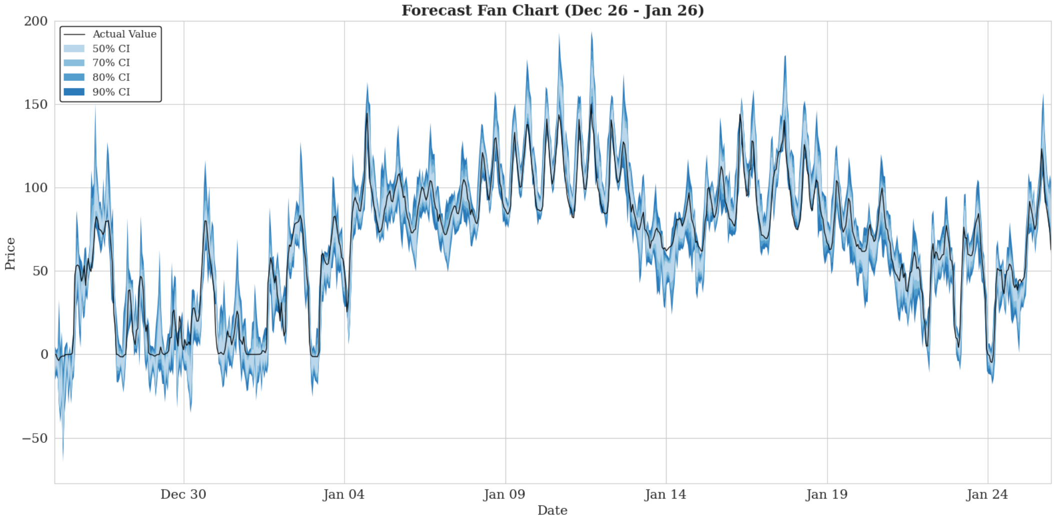
- **Weekly Seasonality:** Significant spikes at lags $k = 7, 14, 21$ (marked in red) persist in the residuals.
- **Implication:** The standard ARX model with weekday dummies accounts for the *average* weekly pattern but fails to capture the full dynamic weekly cyclicity of prices.





Forecast Analysis: First Month Performance

UNIVERSITÀ DI PARMA



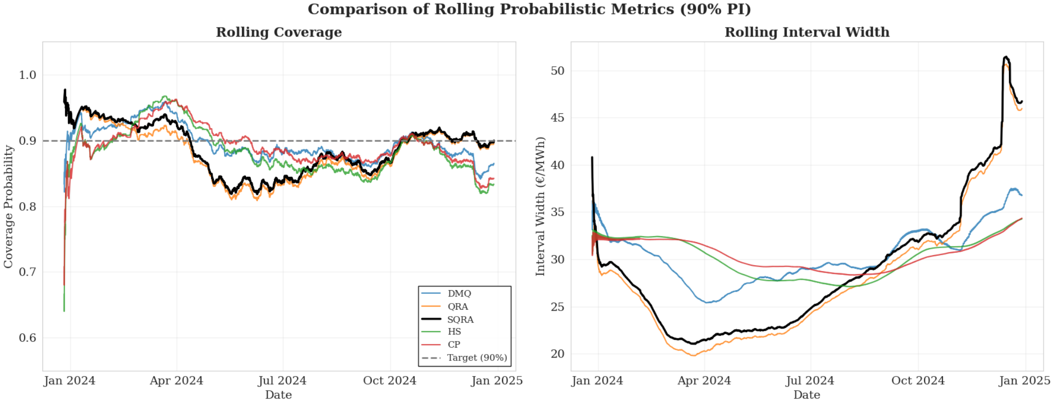
Global Forecast Performance Metrics

UNIVERSITÀ DI PARMA

Method	Interval	PICP	Pinball	Average Width	Winkler	p_{uc} (Kupiec)
DMQ	50%	0.510	3.283	10.97	26.26	0.072
	70%	**0.700**	2.592	17.25	34.55	0.932
	80%	0.792	2.102	22.09	42.04	0.075
	90%	**0.894**	1.458	30.62	58.33	0.053
QRA	50%	0.480	3.101	9.24	24.81	0.000
	70%	0.668	2.468	15.28	32.90	0.000
	80%	0.768	2.007	20.28	40.15	0.000
	90%	0.879	1.373	28.99	54.94	0.000
SQRA	50%	**0.501**	**3.099**	9.68	**24.79**	0.915
	70%	0.686	**2.465**	15.90	**32.87**	0.006
	80%	0.781	**2.001**	20.88	**40.01**	0.000
	90%	0.886	**1.366**	29.82	**54.63**	0.000
HS	50%	0.494	3.368	10.51	26.94	0.288
	70%	0.687	2.680	16.80	35.73	0.006
	80%	0.781	2.188	21.64	43.75	0.000
	90%	0.882	1.526	30.31	61.03	0.000
CP	50%	0.508	3.388	11.07	27.10	0.143
	70%	0.705	2.693	17.62	35.91	0.332
	80%	**0.796**	2.192	22.49	43.83	0.399
	90%	0.891	1.519	30.71	60.75	0.006

Visual Results: Adaptivity (Day View)

UNIVERSITÀ DI PARMA



Conclusion

Conclusion

UNIVERSITÀ DI PARMA

Methodological Contribution

- **Shift:** Transitioned from **classical Conformal Prediction** approach to **Sequential Predictive Conformal Inference (SPCI)**.
- **Implementation:** Operationalized SPCI via a rolling DMQ model on stacked residuals.

Key Empirical Results (German Market)

- **Performance:** Outperforms state-of-the-art benchmarks. Notably, it is the **only model to pass the Kupiec test** across all confidence levels (null hypothesis not rejected).

Thanks for your attention!

Open discussion welcome

Feedback appreciated

References

Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani, R. J. (2023).
Conformal prediction beyond exchangeability.
The Annals of Statistics, 51(2):816–845.

Brusaferri, A., Ballarino, A., Grossi, L., and Laurini, F. (2025).
On-line conformalized neural networks ensembles for probabilistic forecasting of day-ahead electricity prices.
Applied Energy, 398:126412.

Catania, L. and Luati, A. (2023).
Semiparametric modeling of multiple quantiles.
Journal of Econometrics, 237(2):105365.

- Fontana, M., Zeni, G., and Vantini, S. (2023).
Conformal prediction: a unified review of theory and new challenges.
Bernoulli, 29(1):1–23.
- Hubicka, K., Marcjasz, G., and Weron, R. (2019).
A note on averaging day-ahead electricity price forecasts across calibration windows.
IEEE Transactions on Sustainable Energy, 10(1):321–323.
- Kath, C. and Ziel, F. (2021).
Conformal prediction interval estimation and applications to day-ahead and intraday power markets.
International Journal of Forecasting, 37(2):777–799.

- Liu, B., Nowotarski, J., Hong, T., and Weron, R. (2017).
  Probabilistic load forecasting via quantile regression averaging on sister forecasts.
  IEEE Transactions on Smart Grid, 8(2):730–737.
- Maciejowska, K. and Nowotarski, J. (2016).
  A hybrid model for gefcom2014 probabilistic electricity price forecasting.
  International Journal of Forecasting, 32(3):1051–1056.
- Nowotarski, J. and Weron, R. (2015).
  Computing electricity spot price prediction intervals using quantile regression and forecast averaging.
  Computational Statistics, 30(3):791–803.

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

Romano, Y., Patterson, E., and Candès, E. (2019).
Conformalized quantile regression.
In *Advances in Neural Information Processing Systems*, volume 32, pages 3538–3548.

Tibshirani, R. J., Barber, R. F., Candès, E. J., and Ramdas, A. (2019).
***Conformal prediction under covariate shift*.**
Curran Associates Inc., Red Hook, NY, USA.

References v

Uniejewski, B. (2025).
Smoothing quantile regression averaging: A new approach to probabilistic forecasting of electricity prices.
Journal of Commodity Markets, 39:100501.

Uniejewski, B., Weron, R., and Ziel, F. (2018).
Variance stabilizing transformations for electricity spot price forecasting.
IEEE Transactions on Power Systems, 33(2):2219–2229.

Vovk, V., Gammerman, A., and Shafer, G. (2005).
***Algorithmic Learning in a Random World*.**
Springer, New York.

Wang, X., Hyndman, R. J., Li, F., and Kang, Y. (2023).
Forecast combinations: An over 50-year review.
International Journal of Forecasting, 39(4):1518–1547.

Xu, C. and Xie, Y. (2023).
Conformal prediction for time series.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(10):11575–11587.