Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Seasonality-Aware Dual-Memory Pricing for Electric Vehicle Charging Stations under Non-Stationary Demand

Xinzi Wang

2008wangxinzi@163.com

Shenyang University

Article

Keywords: electric vehicle charging, dynamic pricing, non-stationary demand, online learning, seasonal estimation, dual-memory prediction, dynamic pseudo-regret

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11437863/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Seasonality-Aware Dual-Memory Pricing for Electric Vehicle Charging Stations under Non-Stationary Demand

Xinzi Wang [1,*]

1 College of Liberal Arts and Law, Shenyang University, Shenyang 110044, China
*email: 2008wangxinzi@163.com

Abstract

Electric-vehicle (EV) charging prices are selected from partial price-response feedback while demand combines recurrent calendar patterns with structural changes. The central learning challenge is to retain reusable seasonal information while adapting to recent market conditions. We develop and evaluate seasonality-aware dual-memory pricing (SA-DMP), which combines long- and short-horizon seasonal demand forecasts through context-specific weights updated from previously observed prediction losses. In six controlled 100-day demand environments with 20 paired seeds per scenario, SA-DMP achieves lower mean dynamic pseudo-regret than an architecture-matched fixed-weight dual-memory policy in all six environments. Its macro-average oracle-normalized reduction is 0.441 percentage points (95% seed-cluster bootstrap interval: 0.315–0.564 percentage points). A separate exploratory, post-development calibration study using disjoint test seeds yields a 0.463-percentage-point macro-average reduction relative to a validation-selected fixed blend. An eight-connector operational simulation complements the learning comparison by characterizing multi-hour occupancy, abandonment, and time-of-use procurement in the evaluation of contribution margin. At the same time, absolute contribution margin remains below a validation-selected time-of-use tariff in the tested operational environments, delineating the deployment boundary of the adaptive gate. The findings identify the incremental value of loss-adaptive seasonal memory weighting for non-stationary EV charging-price learning under bandit feedback.

Keywords: electric vehicle charging; dynamic pricing; non-stationary demand; online learning; seasonal estimation; dual-memory prediction; dynamic pseudo-regret

1. Introduction

Electric-vehicle charging operators must set retail prices while balancing uncertain demand, electricity procurement costs, customer price response, and limited charging capacity. Demand varies systematically by hour and weekday, yet it also changes with customer composition and local operating conditions. The central pricing challenge is to learn from partial feedback without relying excessively on observations that no longer represent the current demand environment.

A charging price is an online action: after a price is posted, the operator observes the corresponding session demand but not the demand that would have occurred at other prices. This is a bandit-feedback problem, related to the classical stochastic upper confidence bound (UCB) analysis [1,2], the adversarial EXP3 family [3], and non-stationary bandit methods based on windowing, discounting and restart schedules [4,5]. The pricing problem is an instance of dynamic pricing with demand learning [6]. This bandit literature motivates a price-learning design that combines recurring seasonal structure with rapid adaptation under selected-price feedback.

There is an important distinction between seasonal variation and structural non-stationarity. Morning and evening peaks may recur predictably, so forgetting all prior history after a temporary demand fluctuation can unnecessarily discard information. Conversely, a method that preserves months of observations may respond slowly when a

competing station enters the market or the local population of charging users changes. A dual-memory design is therefore appealing: a longer predictor captures recurring patterns, while a shorter predictor places greater emphasis on the recent market state.

Existing non-stationary bandit pricing methods resolve this tension with a fixed forgetting rule: stationary UCB-type policies pool all observations and adapt slowly after structural change, whereas sliding-window, discounted, and restart-based policies impose a single prespecified forgetting timescale whatever the demand pattern [4,5]. Under selected-price feedback, no existing charging-pricing method specifies how a learner should allocate weight between reusable seasonal memory and recent market evidence as demand conditions change. This open design question motivates the present study.

This study isolates the contribution of adaptive memory combination while matching seasonal adjustments, historical windows, and price-selection rules. The evaluation includes fixed-weight ablations, independently tuned conventional online-learning policies, and fixed and time-of-use pricing controls. A complementary operational experiment evaluates electricity contribution margin under multi-hour connector occupancy, providing a distinct assessment of tariff performance beyond the hourly learning benchmark.

The contributions are threefold. First, we formulate a reproducible EV charging-pricing benchmark with selected-price feedback and a separately computed dynamic pseudo-regret comparator. Second, we evaluate an online loss-based combination of seasonal long- and short-memory demand forecasts against fixed-weight methods sharing the same architecture. Third, we characterize the gate's performance across recurrent patterns, structural changes, and customer-response shifts, and separately quantify contribution margin under connector-constrained operation. Together, these comparisons identify the gate's learning contribution while distinguishing it from operational tariff selection.

2. Related Work and Research Positioning

2.1. Sequential Decision-Making under Changing Rewards

Multi-armed bandits describe sequential decisions with incomplete feedback and the exploration–exploitation trade-off [1,2]. Sliding-window and switching UCB policies address stale observations when rewards change [4], while variation-aware models motivate dynamic regret as a performance criterion [5], and tuned sliding-window policies attain optimal dynamic regret under such variation budgets without prior knowledge of their magnitude [7]. Adversarial bandit methods such as EXP3 offer a different approach to randomized exploration [3]. These methods form complementary reference policies for assessing adaptation under the common seasonal and non-stationary simulation protocols.

2.2. Seasonality and Adaptive Forecast Combination

EV charging demand contains both repeated calendar structure and changing market conditions. Field studies based on observed charging records document systematic hourly, weekday, and seasonal structure in charging demand and station load [8,9]. Calendar-based adjustments can improve prediction when a price–time context has enough historical observations; however, long-memory estimates adapt slowly after a structural shift. SA-DMP combines forecasts with two time horizons and updates their mixture according to previous prediction losses. The method is related to prediction-with-experts and adaptive aggregation [10], including the classical problem of tracking a shifting best expert as the identity of the better predictor changes over time [11]. The focus here is the controlled evaluation of that combination within a price-dependent, partial-feedback charging model.

2.3. Charging Pricing and Data Availability

EV charging pricing has been investigated under different tariff formats and assumptions about user flexibility. Limmer [12] reviews dynamic-pricing designs and distinguishes flexibility in charging timing, energy requirements, duration, and location. These forms of flexibility are not interchangeable when defining a pricing benchmark.

Learning-based pricing is established in EV charging. Soltani et al. [13] track price-dependent charging behavior through online learning and use the learned response in pricing decisions. Qureshi et al. [14] formulate non-stationary pricing at a commercial charging station as a multi-armed bandit problem. Narayan et al. [15] study reinforcement-learning-based charging prices under unknown customer behavior, while Zhong et al. [16] couple demand prediction and service-fee optimization in an iterative deep-reinforcement-learning approach. Building on these approaches [13–16], this study targets a specific design decision: how a seasonal pricing learner should allocate weight between historical and recent demand evidence using only selected-price feedback.

Operational models can involve decisions beyond retail price selection. Wang et al. [17] investigate joint real-time charging pricing and scheduling, Liu et al. [18] couple time-of-use pricing with coordinated charging schedules, and Mrkos et al. [19] represent cross-period charging capacity in reservation-based pricing. We focus the decision space on retail price selection and evaluate its consequences in a separate simulator with multi-hour connector occupancy and abandonment.

Observed behavior may also respond differently across customer groups and fee designs. Luo and Wang [20] analyze charging behavior and user composition following changes to public-station charging and post-charging parking fees. Evidence of heterogeneous charging behavior motivates the two-segment price-response sensitivity analysis. The numerical parameterization is specified in Sections 5 and 7.

2.4. Mechanistic Contribution: Loss-Adaptive Seasonal Memory Blending

The method builds on the established families of non-stationary UCB [4,5], calendar-dependent forecasting, and online loss-weighted prediction [10]. Our contribution is the integration and matched evaluation of a context-specific, loss-adaptive short/long-memory gate for price-dependent, partially observed EV charging demand. The architecture-matched ablation identifies the incremental effect of updating the mixture weight online while holding the seasonal estimator, histories, and price-selection rule constant.

In application terms, charging-price learning under non-stationary demand [14], real-time price-response tracking [13], and forecasting–pricing feedback [16] provide relevant precedents. The specific question tested here is whether a past-prediction-loss gate adds value to an architecture-matched seasonal combination of short- and long-memory forecasts under selected-price feedback. We use within-simulator, matched-information comparisons to identify the contribution of online memory weighting.

2.5. Evaluation Objectives and Controlled Comparisons

H1 (algorithmic gate) compares SA-DMP with a pre-frozen, architecture-matched fixed-weight dual-memory predictor while matching the seasonal estimator, observation histories, arm-selection rule, and exploration scale. The primary comparator uses the fixed mixture weight frozen during method development. A complementary post-development calibration study selects a fixed mixture weight on separate validation seeds and evaluates it on disjoint test seeds; its exploratory design is specified in Section 5.4.

H2 (non-stationarity) evaluates how the paired gate effect responds to changes in arrival dispersion, customer composition, and regime-switch timing. H3 addresses the deployment-relevance question of whether an improvement in simplified dynamic pseudo-regret translates into greater realized electricity contribution margin under multi-hour occupancy and time-of-use procurement. We evaluate this question separately using

architecture-matched fixed-memory pricing and validation-selected tariff controls; the two evaluation layers have different outcomes and capacity-state representations.

Together, these comparisons identify the incremental algorithmic effect of the gate, characterize its response to demand shifts, and distinguish learning performance from connector-constrained operating margin. Scenario-wise paired effects, uncertainty intervals, and the complete policy comparisons are reported to support these evaluation objectives.

3. Problem Formulation

3.1. Decision Process and Information Structure

Consider a single EV charging station making an hourly decision for t = 1, …, T. The station selects a retail price p_t from a discrete set $P = \{p_1, \ldots, p_k\}$. At the beginning of the hour it observes a calendar context c_t and its electricity procurement cost w_t. The context belongs to a three-class off-peak, shoulder, or peak partition in the reference simulator. The policy does not observe the contemporaneous demand-response function or the profit associated with prices it did not choose.

Let $A_t \sim \text{Poisson}(\lambda_t)$ denote potential arrivals. A hypothetical driver accepts a price p with probability

$$s_t(p) = \left[1 + \exp\left(\frac{p - v_t}{\tau}\right)\right]^{-1}, \tag{1}$$

where v_t is a time-varying location parameter for synthetic willingness to pay and $\tau > 0$ is its scale. Under independent Poisson thinning, the accepted number of customers obeys

$$D_t(p) \sim \text{Poisson}\big(\lambda_t s_t(p)\big). \tag{2}$$

Both λ_t and v_t may vary, making the distribution of accepted demand non-stationary. The simulator samples common potential customer arrivals and willingness-to-pay values for all evaluated policies, but an individual online policy only sees its selected-price outcome.

3.2. Served Demand and Contribution Margin

The reference model caps completed charging sessions within each hour at M, and assigns a fixed energy amount E kWh per completed session. The served count, electricity volume, and realized contribution margin are

$$N_t(p) = \min\{D_t(p), M\}, \qquad Q_t(p) = EN_t(p), \tag{3}$$

$$r_t(p) = (p - w_t)Q_t(p). \tag{4}$$

The expectation used by the independent offline evaluator is

$$\mu_t(p) = (p - w_t)E\,\mathbb{E}[\min\{D_t(p), M\}]. \tag{5}$$

We report electricity contribution margin, defined as charging revenue minus modeled energy procurement cost, rather than accounting net profit. The reference benchmark uses an hourly service cap; Section 7 models cross-hour connector occupancy explicitly.

3.3. Dynamic Pseudo-Regret

For an online policy selecting p_t, the simulation reports expected dynamic pseudo-regret

$$\bar{R}_T^{\text{dyn}} = \sum_{t=1}^{T}\left[\max_{p\in\mathcal{P}}\mu_t(p) - \mu_t(p_t)\right]. \tag{6}$$

The offline oracle supplies the period-specific expected-margin reference for evaluating the learner and uses the true simulation mechanism to evaluate every candidate price. Oracle information remains inaccessible to the online policy. Realized contribution margin $\Sigma_t\, r_t(p_t)$ is reported as a separate outcome. The oracle calculation can be reproduced from the capacity-capped Poisson expectation used by the simulator.

4. Proposed Seasonality-Aware Dual-Memory Pricing Method

4.1. Price-Specific Histories and Pooled Seasonal Adjustment

The policy observes the normalized served fraction $y_t = N_t(p_t)/M \in [0,1]$. Historical records are partitioned by price arm k and calendar context c. To capture weekly variation within these contexts, a two-dimensional harmonic feature is defined as

$$\boldsymbol{\phi}(t) = [\sin(2\pi t/168),\ \cos(2\pi t/168)]^{\mathrm{T}}. \tag{7}$$

A ridge-regularized harmonic coefficient β_t is fitted to previous normalized outcomes after subtracting a separate within-arm historical mean. This yields a pooled seasonal correction without mechanically equating the demand observed at different posted prices. Harmonic fitting is enabled only when the amount of available data and a minimum explanatory-gain condition are satisfied. The harmonic component is a predictive calendar adjustment, not a causal price-response estimator.

For arm k, let $H_{k,c}^{(H)}(t)$ contain its earlier observations in the current context over the most recent H hours. The seasonally adjusted demand estimate is schematically

$$\hat{q}_{k,t}^{(H)} = \mathrm{clip}_{[0,1]}\left[\bar{y}_{k,c}^{(H)} + \hat{\boldsymbol{\beta}}_t^{\mathrm{T}}\left(\boldsymbol{\phi}(t) - \bar{\boldsymbol{\phi}}_{k,c}^{(H)}\right)\right]. \tag{8}$$

The implementation uses up to $H^L = 672$ hours of history for the long-memory estimator and $H^s = 168$ hours for the short-memory estimator. Both predictions are computed solely from information available before choosing the current price. The archived frozen settings are $\alpha = 0.005$, $\eta = 8$, $b = 2$, $\delta = 0.98$, and $\varrho\max = 0.85$, with minimum long- and short-memory counts of 5 and 3. Harmonic fitting requires at least 70 eligible observations, ridge penalty 15, explained-gain threshold 0.035, and coefficient-norm threshold 0.05. All hyperparameters are fixed to the archived settings before evaluation and are provided for replication.

4.2. Online Blending Based on Previous Prediction Errors

Let $\hat{q}_{k,t}^{L}$ and $\hat{q}_{k,t}^{s}$ denote the long- and short-memory predictions. For calendar context c, the learner stores a recursively discounted measure of the short model's historical squared-error advantage:

$$G_{c,t+1} = \delta G_{c,t} + \left(y_t - \hat{q}_{k_t,t}^{L}\right)^2 - \left(y_t - \hat{q}_{k_t,t}^{S}\right)^2. \tag{9}$$

The loss difference is added only when both relevant predictors meet their minimum price-specific observation counts; otherwise the statistic decays without receiving an invalid comparison. Positive $G_{c,t}$ indicates that the short predictor has, on recent selected-price observations, accumulated less squared error. After each selected-context update, the implementation clips the advantage statistic to [−4, 4] and the logistic argument to [−40, 40] for numerical stability; Equations (9)–(10) show the underlying unclipped recursions. The short-memory mixing weight is

$$\rho_{c,t} = \rho_{\max}\left[1 + \exp\left(-\eta G_{c,t} + b\right)\right]^{-1}, \tag{10}$$

and the combined demand estimate becomes

$$\hat{q}_{k,t} = \left(1 - \rho_{c,t}\right)\hat{q}_{k,t}^{L} + \rho_{c,t}\hat{q}_{k,t}^{S}. \tag{11}$$

If there is insufficient history for either component on a candidate arm, the implementation temporarily falls back to the long-memory prediction. The gate is updated after the

selected-price demand is observed, so its current value cannot access the yet-unrealized y_t.

4.3. Optimistic Price Selection

An optimistic served-demand score with a count-based exploration bonus is constructed as

$$U_{k,t} = \min\left\{1, \hat{q}_{k,t} + \sqrt{\frac{\alpha \log(t+2)}{2\max\left(1, n_{k,t}^{L}\right)}}\right\}, \tag{12}$$

where $n_{k,t}^{L}$ is the count of relevant long-memory observations. Unobserved arm/context combinations are initially assigned the maximal normalized score of one. The price decision is

$$p_t \in \operatorname{argmax}_{p_k \in \mathcal{P}} (p_k - w_t) EM\, U_{k,t}. \tag{13}$$

Equation (12) uses a count-based exploration bonus within the empirical pricing policy. The bonus follows the standard UCB form rather than a calibrated finite-sample confidence interval, and performance is assessed empirically through paired dynamic pseudo-regret in Section 6 rather than through a formal regret bound. Figure 1 summarizes the information flow and the timing of the selected-price feedback update.



Figure 1. SA-DMP online decision sequence. The blending state for period t is computed from outcomes observed before t; the selected-price outcome updates the state used in later periods.

4.4. Implementation and Primary Matched Ablation

Algorithm 1 (SA-DMP). Initialize arm- and context-specific histories and the online gating statistics. At each hour: (i) read context and known wholesale cost; (ii) for each arm–context, fit the ridge-regularized seasonal correction only when at least 70 eligible observations are available and the explained-gain (0.035) and coefficient-norm (0.05) thresholds are met, and otherwise disable that correction; (iii) predict each price's long- and short-window normalized served demand; (iv) blend those forecasts using the previously updated context gate, falling back to the long-memory prediction on arms whose observation counts miss the minima (5 long, 3 short); (v) select the arm maximizing Eq. (13); (vi) observe only its realized served demand, record contribution margin, update predictor histories, and add the squared-error difference to the context's discounted advantage statistic only when both predictors met their observation minima on the selected arm (otherwise the statistic decays without an update); and (vii) proceed to the next hour.

The primary matched comparator, Fixed-ϱ = 0.8, uses $\varrho_{c,t} = 0.8$ when short- and long-window observation-count requirements are satisfied and otherwise matches SA-DMP. All other model components remain unchanged. This ablation isolates the incremental effect of learning ϱ online rather than the value of introducing a second window. The

reference implementation maintains bounded historical buffers for both forecasting horizons; its runtime depends on the maximum window and context observations.

5. Numerical Experimental Design

5.1. Simulator and Scenario Definitions

The reference benchmark uses the following prespecified synthetic demand and tariff parameters. The candidate prices are 0.9, 1.1, 1.3, 1.5, 1.7, 1.9, and 2.1 CNY/kWh. Each synthetic charging session delivers E = 20 kWh and the hourly throughput cap is M = 12 sessions. The three time-of-use contexts use procurement costs of 0.45, 0.63, and 0.82 CNY/kWh, respectively, with baseline hypothetical arrival intensities of 5, 9, and 16 vehicles per hour. The baseline logistic willingness-to-pay scale is $\tau = 0.24$ CNY/kWh.

Parameter magnitudes were chosen to reproduce publicly documented operating conditions rather than to favor the proposed policy. The candidate price range of 0.9–2.1 CNY/kWh brackets retail tariffs commonly reported for public charging in dynamic-pricing surveys [12], and the two-segment willingness-to-pay specification reflects the heterogeneous user responses to charging and post-charging parking fee changes documented in field studies [20]. Within this range, a uniform 0.2 CNY/kWh step yields seven arms and trades price resolution against learning speed: a finer grid would leave some arms too sparse to accumulate the required within-arm observations within the 100-day horizon, whereas a coarser grid would compress the oracle margin differences the benchmark is designed to measure. The same seven prices support the price-response diagnostics in Section 7.4. A session energy of 20 kWh, an hourly throughput cap of 12 sessions, and the constant 22-kW active charging rate used in Section 7 correspond to typical public Level-2 charging configurations. These choices define a deliberately ordinary operating point, so that the observed policy differences arise from demand non-stationarity rather than from an extreme parameterization.

The six demand processes in Table 1 combine routine calendar variation with stationary, abrupt, gradual, periodic, mixed, and temporary-shock changes. Paired random seeds ensure each competing policy encounters the same latent arrivals and willingness-to-pay realizations, while receiving feedback only for its selected prices.

Table 1. Demand processes used in the numerical evaluation; full parameter definitions are included in the accompanying code. CNY denotes Chinese yuan.

Scenario	Demand change mechanism	Principal evaluation question
Stationary	No structural parameter change; routine day-part classes remain	Does the learner forget unnecessarily?
Abrupt	At halfway, arrival intensity is multiplied by 0.72 and willingness center falls by 0.45 CNY/kWh	How quickly does it react to a breakpoint?
Gradual	Arrival intensity and willingness center decline continuously	Can it track ongoing drift?
Periodic	Additional weekly sinusoidal fluctuations	Can it exploit repeated cycles?
Mixed	Abrupt reduction combined with weekly variation and mild trend	Does adaptation remain useful under simultaneous effects?
Seasonal shocks	Weekly cycles with two brief negative impulses that subsequently disappear	Does it avoid treating temporary shocks as permanent shifts?

5.2. Competing Policies and Parameter Calibration

The comparison includes standard UCB, fixed sliding-window UCB (SW-UCB), discounted UCB (D-UCB), block-restarted UCB (R-UCB), a context-specific EXP3 reference (C-EXP3), the seasonality-only policy (CP-season-only), Fixed-ϱ = 0.8, and SA-DMP. R-UCB and C-EXP3 are implemented as the reference variants specified in the reproducibility archive, with their tuning grids and observation rules documented.

The five conventional bandit families are tuned over pre-specified finite grids using 60 simulated days and eight validation seeds, with the equal-scenario average normalized pseudo-regret as the criterion. SA-DMP, fixed-weight dual memory, and CP-season-only use frozen parameters from the preceding development stage. The independent benchmark then evaluates eight policies over 100 days and 20 new seeds in each of six scenarios. Every policy is exposed to the same underlying stochastic realizations within each scenario–seed pair, with no access to unselected-price feedback. The selected conventional settings are UCB α = 0.005; SW-UCB α = 0.005 and window 672 h; D-UCB α = 0.005 and discount 0.999; R-UCB α = 0.005 and restart block 1,008 h; and C-EXP3 γ = 0.25. The full candidate grids, complete eight-policy run-level outcomes, frozen scenario configurations, and source hashes are supplied in the reproducibility archive.

5.3. Evaluation Metrics and Inference

Primary performance is cumulative dynamic pseudo-regret in CNY, computed from expected counterfactual margins. A normalized value divides pseudo-regret by total oracle expected margin for each run. Realized contribution margin and policy wall-clock time are additional descriptive outputs. The principal incremental-effect estimand is the difference between Fixed-ϱ = 0.8 and SA-DMP, matched within scenario and seed. Positive values indicate SA-DMP obtains lower pseudo-regret.

We compute paired percentile bootstrap confidence intervals by resampling complete random-seed trajectories within each experiment. The primary benchmark uses 6,000 draws per scenario–comparator pair and 12,000 seed-cluster draws for six-scenario macro averages. The distribution-shift analysis uses 5,000 draws per scenario and 10,000 macro seed-cluster draws; the fixed-weight study uses 10,000 scenario draws and 15,000 macro draws. The connector-constrained margin analysis uses 7,000 paired draws per comparison. Scenario-level two-sided paired sign-flip tests enumerate all sign assignments (2^20 for the primary benchmark and 2^12 for the distribution-shift experiments), with Holm correction within each six- or three-scenario family. The fixed-weight macro test enumerates 2^10 assignments of seed-level equal-scenario mean differences. Sign-flip inference assumes exchangeability of paired signs under the null. The intervals quantify seed-level uncertainty under the stated simulation mechanisms; scenario-wise intervals are not simultaneous or multiplicity-adjusted confidence bands. The archive-statistics reproduction script supplies the exact tests and reproduces the archived bootstrap algorithms without rerunning simulated policy trajectories.

Independent numerical checks confirm the analytic expected-service evaluator using a separately implemented Poisson tail-sum formula and Monte Carlo samples across 63 demand–context–price combinations (Document S1). Reproducibility checks and execution provenance are summarized in Supplementary Document S2 (also included within File S1); the detailed oracle audit is reported in Document S1.

Generative AI use. SciMaster (accessed on 13 and 16 October 2026; version not recorded) assisted with supplementary verification code, checking archived summary statistics, manuscript editing, and revision of the presentation figures. It was not used to generate simulated demand data or policy outcomes. The author reviewed and verified all assisted code, figures, and text before submission.

5.4. Fixed-Weight Calibration Sensitivity Study

After examining the primary 100-day benchmark, we performed a separate, explicitly exploratory sensitivity study to assess sensitivity to fixed-weight calibration. Fixed short-memory mixture weights of 0, 0.25, 0.50, 0.65, 0.75, 0.80, 0.85, and 1.00 were evaluated using 35-day simulations, five validation seeds, and all six previously specified scenario families. The selection criterion was the equal-scenario average oracle-normalized dynamic pseudo-regret. The selected weight was frozen before evaluation on ten disjoint seeds with 65-day horizons. The comparison included SA-DMP, the validation-selected fixed weight, and the archived fixed weight 0.8. All forecasts and price decisions remained online, and oracle rewards remained inaccessible to the policies.

This post-development sensitivity analysis is reported as exploratory robustness evidence because the scenario families had been used during method development. Validation/test seed separation assesses fixed-weight calibration on new stochastic realizations of those specified families.

6. Results

6.1. Primary Architecture-Matched Comparison

SA-DMP achieves lower mean cumulative dynamic pseudo-regret than Fixed-$\varrho = 0.8$ in each of the six primary demand environments (Table 2). The table foregrounds the architecture-matched comparison through paired CNY reductions and oracle-normalized gains. Figure 2 shows the paired reductions and scenario-wise uncertainty intervals. Complete eight-policy means, including the conventional-policy and seasonal-only comparisons, are retained in Supplementary Document S2 and the run-level archive. Among the eight evaluated policies, SA-DMP attains the lowest mean pseudo-regret in three of the six scenarios, with discounted UCB, restarted UCB, and the seasonal-only policy leading in one scenario each.

Table 2. Primary architecture-matched comparison: mean cumulative dynamic pseudo-regret and paired improvements over 20 seeds per scenario, each spanning 100 days. Δ = Fixed-$\varrho = 0.8$ minus SA-DMP; positive values favor SA-DMP. Gains in pp are absolute reductions in oracle-normalized pseudo-regret. Each displayed mean and paired difference is rounded independently from unrounded run-level data; subtracting displayed means can differ by 1 CNY.

Scenario	SA-DMP (CNY)	Fixed-$\varrho = 0.8$ (CNY)	Paired reduction Δ (CNY)	Normalized gain (pp)
Stationary	6,493	7,569	1,076	0.445
Abrupt	9,346	10,098	752	0.456
Gradual	7,977	8,404	428	0.276
Periodic	17,136	18,156	1,020	0.409
Mixed	15,573	16,076	503	0.300
Seasonal shocks	18,390	20,211	1,822	0.760

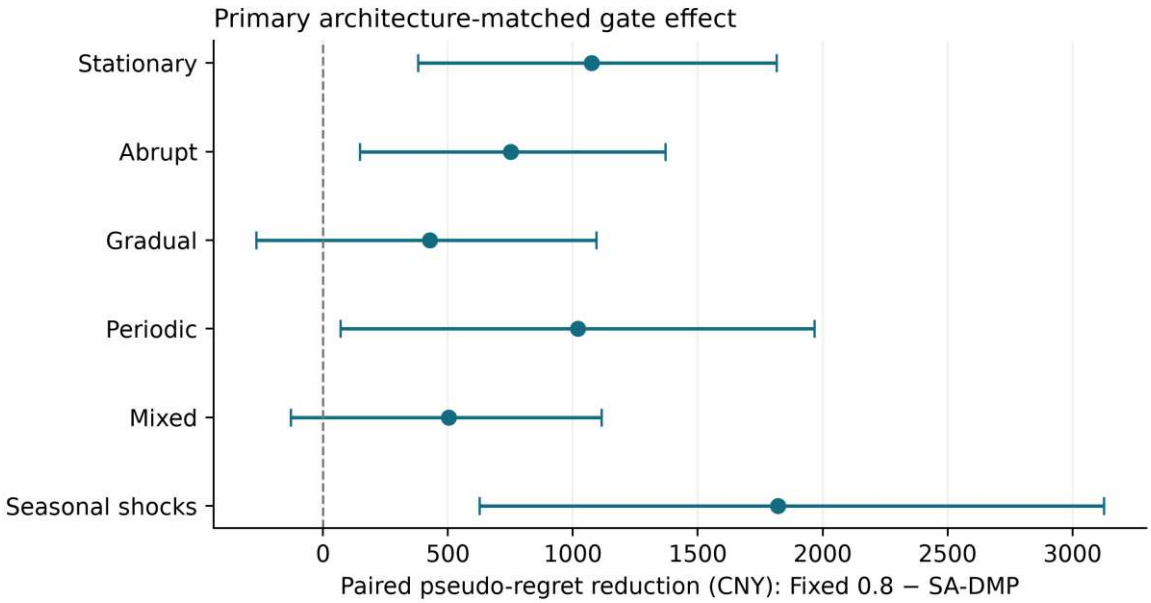


Figure 2. Primary matched comparison: paired reduction in cumulative dynamic pseudo-regret (Fixed-ϱ = 0.8 minus SA-DMP) across six demand environments. Positive values favor SA-DMP. Whiskers are scenario-wise 95% percentile bootstrap intervals based on complete paired seed trajectories; they are not simultaneous or Holm-adjusted confidence bands.

6.2. Incremental Effect of Online Blending

Across the six scenarios, SA-DMP decreases the macro-average oracle-normalized dynamic pseudo-regret by 0.441 percentage points compared with fixed-weight dual memory (95% seed-cluster bootstrap interval: 0.315–0.564 percentage points). This is an absolute reduction in normalized pseudo-regret, computed from the six-scenario, 20-seed matched benchmark. Table 3 reports the scenario-specific paired magnitudes and corrected hypothesis tests.

The paired mean improvement is positive in all six scenarios, ranging from CNY 428 under gradual drift to CNY 1,822 under temporary seasonal shocks. Holm-adjusted paired sign-flip tests identify statistically significant differences in the stationary and seasonal-shock environments (Table 3). The paired estimates and intervals in Figure 2 characterize how the magnitude of the gate effect varies by demand pattern.

Table 3. Paired sign-flip tests for Fixed-ϱ = 0.8 versus SA-DMP, with Holm correction across six scenarios. Here, pp denotes percentage points. Displayed paired reductions are calculated before whole-CNY rounding; uncertainty intervals are shown in Figure 2.

Scenario	Paired reduction (CNY)	Normalized gain (pp)	Two-sided p	Holm p
Stationary	1,076	0.445	0.0066	0.0397
Abrupt	752	0.456	0.0249	0.0997
Gradual	428	0.276	0.2425	0.2734
Periodic	1,020	0.409	0.0560	0.1680
Mixed	503	0.300	0.1367	0.2734
Seasonal shocks	1,822	0.760	0.0100	0.0499

6.3. Calibration-Robust Comparison with a Validation-Selected Fixed Blend

The fixed-weight calibration grid selects ϱ = 0.85 with the lowest six-scenario mean normalized pseudo-regret on the validation seeds (8.981%), compared with 9.228% for ϱ = 0.80. On the disjoint 65-day, ten-seed test, SA-DMP achieves lower mean pseudo-regret than the validation-selected ϱ = 0.85 in each scenario. Its macro-average normalized reduction is 0.463 percentage points (95% seed-cluster bootstrap interval: 0.277–0.653; exploratory two-sided paired sign-flip p = 0.0039). Relative to ϱ = 0.80, the macro-average reduction is 0.380 percentage points (95% interval: 0.136–0.645), with five of six scenario mean differences favoring SA-DMP. Table 4 retains both fixed alternatives; Figure 3 shows the paired normalized differences against the validation-selected weight.

Table 4. Fixed-weight calibration sensitivity experiment: mean cumulative dynamic pseudo-regret (CNY), 65-day horizon and ten paired test seeds per scenario. The exploratory, post-development study design is specified in Section 5.4.

Scenario	SA-DMP	Fixed-ϱ = 0.85	Fixed-ϱ = 0.80
Stationary	5,137	5,883	5,739
Abrupt	8,136	8,644	8,743
Gradual	6,972	7,117	7,205
Periodic	11,964	12,748	12,755
Mixed	11,183	11,680	11,144
Seasonal shocks	11,619	12,809	12,647

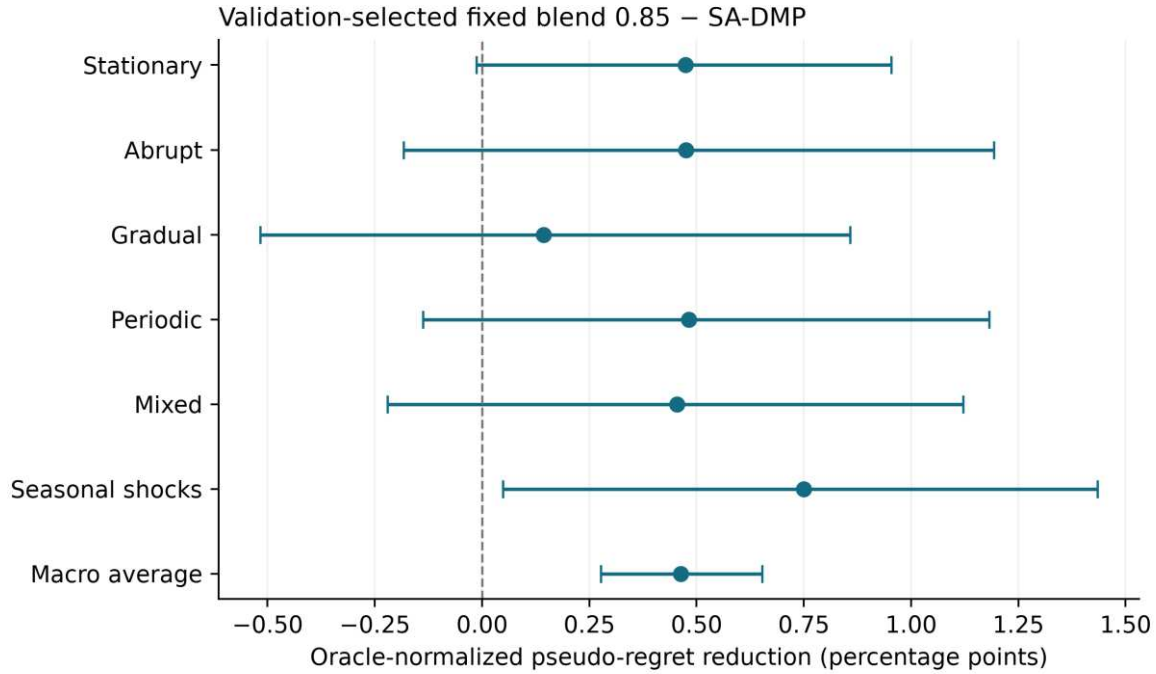


Figure 3. Calibration sensitivity: paired reduction in oracle-normalized dynamic pseudo-regret (validation-selected Fixed-ϱ = 0.85 minus SA-DMP). Positive differences favor SA-DMP. Whiskers show scenario-wise 95% percentile bootstrap intervals; the macro-average interval uses equal-scenario seed-cluster resampling. Scenario-wise intervals are not simultaneous confidence bands.

The aggregate paired reduction against the validation-selected fixed weight is positive, with the largest scenario-level normalized contrast under seasonal shocks. The macro-average interval describes the equal-scenario effect, not statistical significance in every scenario: five of the six scenario-wise intervals include zero. This calibration comparison complements the demand-distribution sensitivity analysis in Section 6.4.

6.4. Sensitivity to Arrival Dispersion and Customer Composition

The distribution-shift experiments examine the gate effect under Gamma–Poisson (negative-binomial) arrivals and a two-segment willingness-to-pay mixture. Three 40-day experiments use 12 paired seeds each (Table 5). The paired mean reduction in pseudo-regret is CNY 1,094 under the weekend mixture, while the changing-customer-mix and irregular-regime cases yield differences of CNY −143 and −918, respectively. The equal-scenario normalized difference is −0.091 percentage points (95% seed-cluster bootstrap interval: −0.716 to +0.607), and none of the scenario-level sign-flip tests reaches 0.05 after Holm correction. The direction and uncertainty of these contrasts characterize sensitivity to the structure of demand variation.

Table 5. Sensitivity of the paired gate effect to distribution shifts (40-day horizon, 12 paired seeds per scenario). Paired gain = Fixed-ϱ = 0.8 minus SA-DMP; negative values favor the fixed blend. Holm-adjusted p-values are shown.

Distribution shift	SA-DMP (CNY)	Fixed-ϱ = 0.8 (CNY)	Paired gain (CNY)	Holm p
Weekend mixture	6,452	7,547	+1,094	0.1963
Changing customer mix	8,313	8,170	−143	0.5806
Irregular regimes	8,695	7,777	−918	0.1963

7. Engineering-Constrained Numerical Evaluation

The reference online-learning benchmark in Sections 5–6 uses an independent hourly throughput constraint. The operational experiment below instead tracks connector occupancy across hours and demand-dependent admission decisions. Because these dynamics create path-dependent capacity states, we evaluate the resulting electricity contribution margins directly rather than applying the reference benchmark's one-period pseudo-regret comparator.

7.1. Connector-Constrained Charging-Station Simulation

The operational simulator represents an eight-connector, 22-kW AC charging station. Retail prices are selected hourly from 0.9–2.1 CNY/kWh, while electricity procurement costs vary across time-of-use contexts. Drivers arrive according to a Gamma–Poisson process and belong to two willingness-to-pay groups. Requested energy and connection duration vary across sessions; connectors remain occupied across hours, and arrivals abandon service if their waiting tolerance is exceeded. The data-generating parameters and simulation code are supplied in File S1. In this layer, the learner receives clipped normalized admitted energy, $\min\{1, Q_{admitted}/(8 \times 35\ \text{kWh})\}$, rather than the reference model's fixed-energy served fraction. The 35-kWh factor is a learner normalization constant. Prespecified illustrative procurement rates are CNY 0.86/kWh during 17:00–22:00, CNY 0.48/kWh during 23:00–07:00, and otherwise CNY 0.65/kWh on weekdays or CNY 0.61/kWh on weekends.

The principal outcome is electricity contribution margin booked at admission, computed from accepted charging energy and the retail–procurement price spread. This energy-margin measure is distinct from accounting net profit. Section 7.3 checks an alternative allocation of procurement costs across the actual charging hours.

A separate validation sample (five independent seeds, 90 days each, combined-change scenario) was used to select a fixed tariff from seven candidate prices and a three-context tariff from 27 candidate tuples based on 1.3, 1.5 and 1.7 CNY/kWh. The frozen validation winners were a flat price of 1.5 CNY/kWh and a night/normal/peak schedule of 1.3/1.5/1.5 CNY/kWh. Both were assessed on 12 unseen seeds for each of three 90-day scenario families (combined changes, smooth decline and stationary demand); seven policy labels were evaluated per scenario–seed pair (252 runs). The validated flat control is intentionally identical to the prespecified fixed-1.5 control. Table 6 distinguishes the matched gate contrast from absolute tariff margins. Figure 4 compares SA-DMP with Fixed blend 0.85; complete policy comparisons are retained in Supplementary Document S2 and File S1.

Table 6. Mean 90-day electricity contribution margin (CNY) under connector-constrained operation, with 12 paired test seeds per scenario. The final row reports SA-DMP minus Fixed blend 0.85, calculated from unrounded seed-level margins. Means and differences are independently rounded to whole CNY, so subtracting displayed means can differ by 1 CNY.

Policy	Combined (CNY)	Smooth (CNY)	Stationary (CNY)
SA-DMP	99,635	92,851	107,222
Fixed blend 0.85	98,782	92,277	105,888
Validated flat 1.5	106,354	100,148	113,826
Validated TOU	108,142	101,940	115,142
SA-DMP − Fixed blend 0.85	+853	+574	+1,333

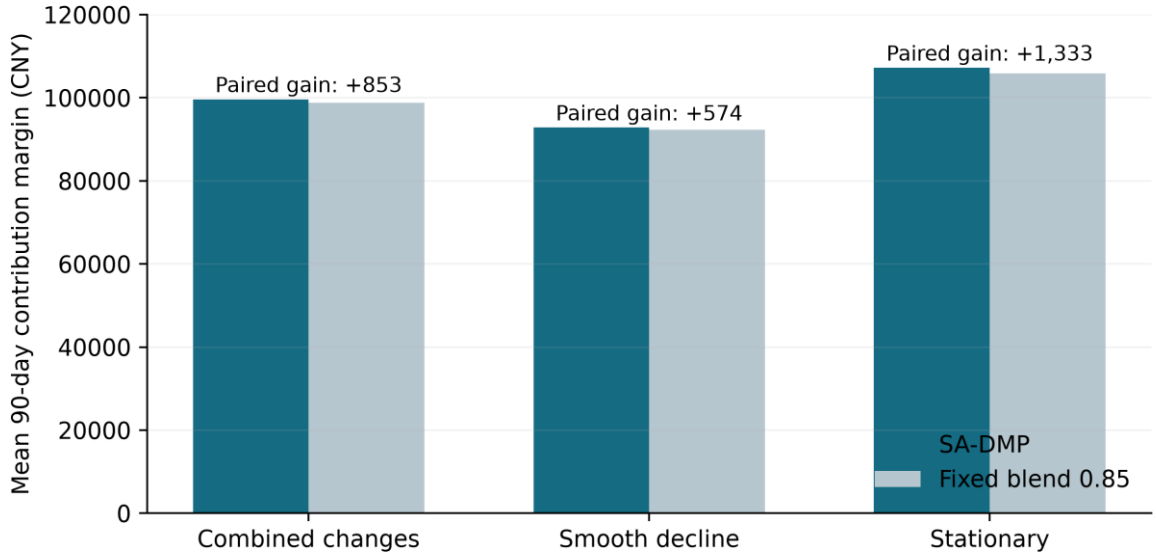


Figure 4. Architecture-matched operational comparison: mean 90-day electricity contribution margin for SA-DMP and Fixed blend 0.85 across three demand scenarios. Annotations report paired mean gains calculated before rounding. Uncertainty intervals are shown in Figure 5; validation-selected flat and TOU margins remain in Table 6.

7.2. Gate Increment and Absolute Tariff-Margin Comparisons

The connector-constrained evaluation assesses the adaptive gate under multi-hour charging occupancy. SA-DMP yields higher mean 90-day electricity contribution margin

than the matched fixed blend in the combined-change, smooth-decline, and stationary scenarios by CNY 853, CNY 574, and CNY 1,333, respectively (Table 6). The corresponding scenario-wise 95% paired bootstrap intervals are CNY 88–1,834, CNY −831–1,893, and CNY 442–2,415 (Figure 5). The smooth-decline interval includes zero; these are unadjusted scenario-wise intervals. Exact two-sided paired sign-flip tests over all 2^12 sign assignments give p = 0.068, 0.435, and 0.016 for the combined-change, smooth-decline, and stationary contrasts; after Holm correction across the three scenarios, only the stationary increment remains significant (adjusted p = 0.047).

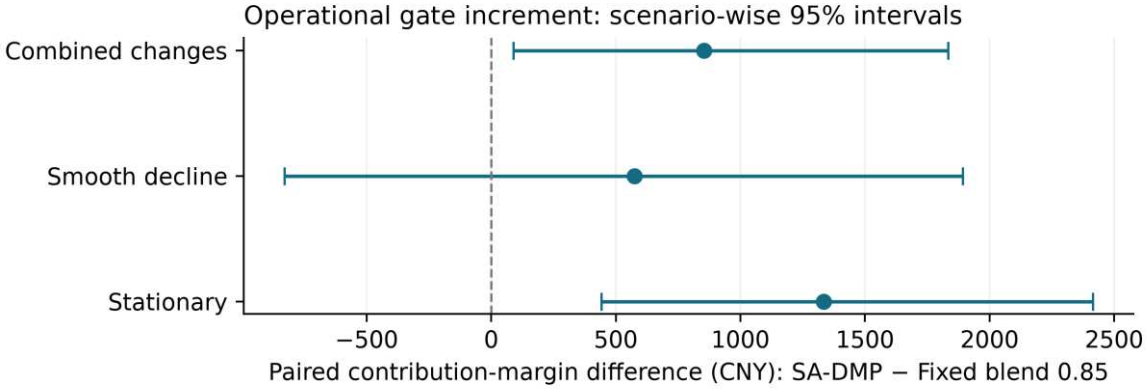


Figure 5. Matched-seed operational gate effect: SA-DMP minus Fixed blend 0.85 in 90-day electricity contribution margin across the three operational scenarios. Positive values favor SA-DMP. Whiskers show scenario-wise 95% paired percentile bootstrap intervals, without multiplicity adjustment. Complete TOU and flat-tariff contrasts are preserved in Supplementary Document S2 and File S1.

Absolute tariff-margin performance is assessed separately. In the combined-change environment, SA-DMP earns CNY 99,635, while the validation-selected TOU tariff earns CNY 108,142. The paired SA-DMP-minus-TOU difference is CNY −8,507 (95% paired bootstrap interval: −9,603 to −7,506). The exact paired sign-flip test attains the smallest two-sided p-value achievable with 12 seeds (p = 0.0005), as it also does in the smooth-decline and stationary scenarios. The selected TOU tariff also has higher mean absolute margin in the smooth-decline and stationary scenarios (Table 6). These contrasts address tariff selection under the specified customer-response and occupancy model, rather than the isolated effect of memory-weight adaptation. Two structural asymmetries qualify this comparison: the adaptive policy's booked margin includes the cost of price exploration, which the fixed and time-of-use controls do not incur, and the time-of-use schedule was itself selected on validation seeds drawn from the same scenario families. The margin gap therefore partly reflects exploration overhead and tariff calibration rather than inferior execution of the learned policy.

The financial decomposition describes an energy-throughput versus unit-margin trade-off. Relative to the validation-selected 1.5 CNY/kWh flat tariff, the booked-margin gap is CNY −6,719, comprising a CNY −7,024 volume term and a CNY +305 unit-margin term, alongside 8,553 kWh of lower admitted energy. This exact per-seed accounting decomposition is descriptive, not causal: posted prices jointly affect acceptance, congestion, occupancy, and admitted customer composition.

7.3. Charging-Hour Procurement-Cost Check

The charging-hour cost-allocation check confirms that the main operational ranking is stable to the procurement booking convention. The calculation assumes a constant 22-kW active charging rate, without power tapering or conversion losses. Detailed policy

margins and cost-allocation assumptions are retained in Supplementary Document S2 and File S1.

7.4. Price-Response Sensitivity Analysis

Price-response sensitivity characterizes the model-implied demand-volume mechanism behind the operational margin comparison. The two-segment willingness-to-pay specification is evaluated at the seven candidate retail prices; the complete acceptance and elasticity grid is provided in Supplementary Document S2 and File S1.

For a willingness-to-pay distribution Logistic(μ, σ), price acceptance is $s(p; \mu, \sigma) = 1/[1 + \exp((p - \mu)/\sigma)]$. With price-sensitive customer proportion π, the mixture is $a(p; \pi) = \pi\, s(p; 1.32, 0.18) + (1 - \pi)s(p; 1.84, 0.21)$, and local point elasticity is $\varepsilon(p; \pi) = p\, a'(p; \pi)/a(p; \pi)$. At p = 1.5 CNY/kWh, the model-implied elasticity is −2.529 for $\pi = 0.54$ and −3.177 for $\pi = 0.68$, indicating that the assumed customer composition changes the local demand response.

The capacity-free per-arrival margin diagnostic $(p - c)a(p; \pi)$, for procurement cost c, isolates customer price response. The operational simulator evaluates the joint consequences of response, connector occupancy, and abandonment. These elasticity diagnostics are model-implied; they support interpretation of the simulated throughput–margin trade-off.

8. Discussion

The matched experiments identify a consistent mean reduction in dynamic pseudo-regret from loss-adaptive seasonal memory weighting across the six primary demand environments. The connector-constrained experiment evaluates a different outcome—electricity contribution margin—under path-dependent station capacity. Separating these outcomes allows the gate's incremental learning effect and the operational tariff comparison to be assessed on their own terms.

The operational comparison relates to joint pricing and charging scheduling [17,18] and reservation-based cross-period connector allocation [19]. Those formulations incorporate decision variables and information beyond the present retail-pricing simulator. Here, the relative financial performance of fixed, time-of-use, and adaptive tariffs is evaluated for the specified customer-response and connector-occupancy model. The matched fixed-blend contrast isolates memory-weight adaptation, while validation-selected tariff controls assess absolute margin.

The six-scenario matched ablation identifies the incremental effect of the loss-adaptive gate, with a 0.441-percentage-point macro-average reduction in oracle-normalized pseudo-regret. The disjoint-seed calibration study yields a 0.463-percentage-point macro-average reduction against the validation-selected fixed blend. These aggregate effects describe improvements under the specified simulation mechanisms; the demand-shift analysis in Section 6.4 identifies settings in which the paired contrast changes direction.

The scenario pattern of the gate effect is itself informative. The largest gain arises under temporary seasonal shocks (0.760 pp), in which the learner must alternate rapidly between trusting recent observations during a shock and trusting long history after the shock disappears—precisely the setting in which online weight adjustment has room to act. The smallest gain arises under gradual drift (0.276 pp), in which long- and short-memory forecasts diverge slowly and a fixed 0.8 blend remains close to the best achievable weighting. The same mechanism rationalizes the reversed contrasts reported in Section 6.4: when regime changes arrive at irregular times, the discounted advantage statistic (discount factor 0.98, an effective memory of roughly 50 hours) can mistake transient error differences for a structural advantage of the short predictor, so the gate overreacts where a fixed weight would have remained passive. The recorded gate trajectories illustrate both cases: under irregularly timed regime changes the mixing weight swings widely and

mistimes the switches, whereas under the persistent weekend structure it settles near zero and correctly suppresses the short-memory predictor (Supplementary Fig. S2.1). The gate is therefore most valuable when the divergence between the two memories is large, temporary, and self-correcting, and it can be counterproductive when advantage signals are dominated by irregular timing rather than by persistent differences in predictive reliability.

In operational terms, the paired reductions of CNY 428–1,822 per station per 100 days correspond to approximately CNY 1,600–6,600 per station per year, obtained through a change in the learning rule alone, with no modification to hardware, tariff structure, or information requirements. Scaled to a network of several hundred stations, the aggregate magnitude would be material to operating budgets under the simulated conditions. These figures derive from the specified synthetic demand mechanisms and serve as a scale reference rather than a field estimate.

The distinguishing feature is a context-dependent online gate evaluated with price-specific seasonal histories and matched algorithmic controls under selected-price feedback. The long predictor retains recurring calendar information, while the short predictor emphasizes recent observations; past forecast losses determine their relative influence. The hourly-throughput and connector-constrained evaluation layers make explicit how the modeled capacity state changes the interpretation of pricing outcomes.

The approach also differs methodologically from earlier learning-based charging-price studies. Qureshi et al. [14] apply a non-stationary multi-armed bandit directly to price selection, while Narayan et al. [15] and Zhong et al. [16] train reinforcement-learning agents whose state and reward representations are specified in advance. Those methods adapt the pricing policy as a whole but do not separate reusable seasonal evidence from recent market evidence within the demand model. SA-DMP instead holds the seasonal estimator and the price-selection rule fixed and adapts only the memory-weighting gate, which makes the incremental effect of memory adaptation identifiable through the architecture-matched ablation. The corresponding trade-off is a narrower adaptation target: representation-level changes, such as the arrival of a customer segment with a different response function, lie outside the gate's adaptation scope and are better handled by policy-level learning formulations.

The price-response diagnostics in Section 7.4 support interpretation of the simulated throughput–margin trade-off. Heterogeneous responses reported for charging and parking fee changes [20] motivate evaluating customer composition, while the present diagnostics retain the energy-only response function specified for this simulation.

The paired analyses support an aggregate within-simulator gate effect, while the stress tests characterize demand-response settings in which the direction of the contrast changes. Validation on independently specified demand regimes and observed price-response data represents a natural next step for strengthening external inference.

In terms of the evaluation objectives in Section 2.5, H1 is supported: paired mean pseudo-regret favors SA-DMP in all six primary environments, and the macro-average interval excludes zero. H2 is supported with a direction: the gate effect varies systematically with the structure of demand variation and reverses under irregularly timed regime changes. H3 receives a qualified negative answer: the gate increment remains positive under connector-constrained operation, but absolute contribution margin stays below a validation-selected time-of-use tariff in each tested environment, so the adaptive gate in its present form complements rather than replaces calibrated tariff schedules.

The evaluation is also bounded in horizon and parameter coverage. The primary horizon of 100 days (90 days in the operational layer) spans weekly seasonality and short-horizon structural changes but not a full annual cycle, so annual-scale seasonal effects and multi-year drift remain untested. In addition, the gate's steepness and bias were frozen at

development-stage values and, unlike the mixture weight, were not subjected to a separate calibration sensitivity study. These bounds follow from the controlled-benchmark design rather than from observed failures of the method.

A natural extension is to integrate station-level connector occupancy and empirically calibrated price response into the online decision state. Such a learner could combine forecast-error adaptation with admission opportunities and price-exploration costs. Operationally this requires per-session transaction records—arrival times, connection durations, delivered energy, and posted and accepted prices—or a designed price-response experiment; methodologically it amounts to a contextual pricing learner whose state includes connector occupancy and queue length.

8.1. Computational Implementation

In the archived reference benchmark, SA-DMP and the architecture-matched fixed blend have similar measured simulation times: 1.587 and 1.585 s per 100-day run, respectively. Both share the seasonal dual-memory computation; the measured difference is 0.002 s at the displayed precision. Complete timings and their environment limitations are documented in Supplementary Document S2. These stored simulation times are descriptive, not measurements of deployment latency.

The exploratory calibration design and inferential conventions are specified in Sections 5.3–5.4. Execution provenance and the scope of sampled source-to-archive replication are recorded in the reproducibility materials.

8.2. Implications for Tariff Design

Across the primary matched benchmark, loss-adaptive blending lowers mean pseudo-regret in all six demand environments. The validation-selected fixed-weight comparison further assesses sensitivity to calibration. In the operational layer, matched fixed-memory and calibrated tariff controls provide distinct information about incremental gate value and absolute electricity contribution margin.

The next step is a connector-state-aware pricing learner that jointly uses forecast-error adaptation and operational admission information. Suitable observed charging transactions or designed price-response experiments would enable empirical response calibration for that extension.

9. Conclusions

We developed and evaluated SA-DMP, a seasonality-aware charging-price learner that updates the balance between long- and short-horizon demand forecasts from previously observed prediction losses. Across six controlled demand environments, SA-DMP achieves lower mean dynamic pseudo-regret than architecture-matched fixed-weight dual memory, with a 0.441-percentage-point macro-average improvement in oracle-normalized performance (95% seed-cluster bootstrap interval: 0.315–0.564). A complementary exploratory, post-development calibration study yields a 0.463-percentage-point macro-average gain against a validation-selected fixed blend on disjoint test seeds. In the eight-connector operational simulation, the adaptive gate also improves mean contribution margin over the matched fixed blend in all three tested demand environments. The operational evaluation distinguishes this gate increment from absolute tariff-margin performance under customer acceptance and multi-hour occupancy. At the same time, a validation-selected time-of-use tariff achieves higher absolute contribution margin in each tested environment (CNY 108,142 versus 99,635 for SA-DMP in the combined-change case), delineating the deployment boundary of the adaptive gate. These results establish the measurable incremental value of loss-adaptive memory combination under selected-price feedback, delineate the conditions under which this advantage holds, and motivate connector-state-aware extensions for operational tariff design.

Supplementary information: The following supporting information is supplied with this article: Document S1: Independent verification of the capacity-capped Poisson oracle; Document S2: Complete policy comparisons, operational diagnostics, and reproducibility checks (also included within File S1); File S1: SA-DMP reproducibility archive (source code, frozen configurations, run-level datasets, calibration records, tariff and cost audits, analysis scripts, and the independent capacity-capped Poisson verification script, 63-cell dataset, summary, and diagnostic figure).

Author Contributions: Conceptualization, X.W.; methodology, X.W.; software, X.W.; validation, X.W.; formal analysis, X.W.; investigation, X.W.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W.; visualization, X.W. The author has read and agreed to the published version of the manuscript.

Funding: This work was supported by a commissioned technical consulting project, "Construction of an Enterprise Compliance Management and Legal Risk Prevention and Control System," entrusted by Shenyang Xinjiacheng Automobile Rescue Co., Ltd. and undertaken by Shenyang University. No contract or grant number was assigned for this project.

Data availability: The data presented in this study were generated computationally; no field-station transaction or meter records were used. The run-level datasets (960-run primary benchmark, 252-run connector-constrained benchmark, 144-run distribution-shift benchmark, fixed-weight validation/test records, and the 63-cell oracle audit), frozen configurations, source code, and analysis scripts are provided in File S1. The original execution environment was not recorded; the scope and environment of the sampled reproduction are documented in Supplementary Document S2 and SUBMISSION_REPRODUCTION_NOTES.md within File S1. Revised presentation figures are calculated from archived outcomes, not new simulation runs.

Acknowledgments: This work was supported by a commissioned technical consulting project from Shenyang Xinjiacheng Automobile Rescue Co., Ltd., which motivated the author's investigation of algorithmic pricing methods in the automotive-service sector. During the preparation and revision of this manuscript, the author used SciMaster (accessed on 13 and 16 October 2026; version not recorded) for supplementary verification code, checking archived statistics, manuscript editing, and figure presentation. The author reviewed and edited all assisted outputs and takes full responsibility for the content of this publication.

Competing interests: The author declares no competing interests.

References

1. Auer, P., Cesa-Bianchi, N. & Fischer, P. Finite-time analysis of the multiarmed bandit problem. Mach. Learn. 47, 235–256 (2002). https://doi.org/10.1023/A:1013689704352
2. Bubeck, S. & Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Found. Trends Mach. Learn. 5, 1–122 (2012). https://doi.org/10.1561/2200000024
3. Auer, P., Cesa-Bianchi, N., Freund, Y. & Schapire, R. E. The nonstochastic multiarmed bandit problem. SIAM J. Comput. 32, 48–77 (2002). https://doi.org/10.1137/S0097539701398375
4. Garivier, A. & Moulines, E. On upper-confidence bound policies for switching bandit problems. In Algorithmic Learning Theory (ALT 2011), Lecture Notes in Computer Science Vol. 6925 (eds Kivinen, J. et al.) 174–188 (Springer, 2011). https://doi.org/10.1007/978-3-642-24412-4_16
5. Besbes, O., Gur, Y. & Zeevi, A. Stochastic multi-armed-bandit problem with non-stationary rewards. In Advances in Neural Information Processing Systems 27 (eds Ghahramani, Z. et al.) 199–207 (Curran Associates, 2014).
6. den Boer, A. V. Dynamic pricing and learning: historical origins, current research, and new directions. Surv. Oper. Res. Manag. Sci. 20, 1–18 (2015). https://doi.org/10.1016/j.sorms.2015.03.001
7. Cheung, W. C., Simchi-Levi, D. & Zhu, R. Learning to optimize under non-stationarity. In Proc. 22nd Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2019), Proceedings of Machine Learning Research Vol. 89 (eds Chaudhuri, K. & Sugiyama, M.) 1079–1087 (PMLR, 2019).

8. Soufiani, E., Almehdawe, E. & Tang, Y. Modelling the stochastic charging behaviour of electric vehicles: a validated framework with user charging data. Transportmetrica A Transp. Sci. (2025). https://doi.org/10.1080/23249935.2025.2599392
9. Guo, Z., Bian, H., Zhou, C., Ren, Q. & Gao, Y. An electric vehicle charging load prediction model for different functional areas based on multithreaded acceleration. J. Energy Storage 73, 108921 (2023). https://doi.org/10.1016/j.est.2023.108921
10. Cesa-Bianchi, N. & Lugosi, G. Prediction, Learning, and Games (Cambridge University Press, 2006). https://doi.org/10.1017/CBO9780511546921
11. Herbster, M. & Warmuth, M. K. Tracking the best expert. Mach. Learn. 32, 151–178 (1998). https://doi.org/10.1023/A:1007424614876
12. Limmer, S. Dynamic pricing for electric vehicle charging—a literature review. Energies 12, 3574 (2019). https://doi.org/10.3390/en12183574
13. Soltani, N. Y., Kim, S.-J. & Giannakis, G. B. Real-time load elasticity tracking and pricing for electric vehicle charging. IEEE Trans. Smart Grid 6, 1303–1313 (2015). https://doi.org/10.1109/TSG.2014.2363837
14. Qureshi, U. et al. Dynamic pricing for electric vehicle charging at a commercial charging station in presence of uncertainty: a multi-armed bandit reinforcement learning approach. In Proc. Int. Conf. on Data Science and Applications (ICDSA 2022), Lecture Notes in Networks and Systems Vol. 552 (eds Saraswat, M. et al.) 625–635 (Springer, 2023). https://doi.org/10.1007/978-981-19-6634-7_44
15. Narayan, A., Krishna, A., Misra, P., Vasan, A. & Sarangan, V. A dynamic pricing system for electric vehicle charging management using reinforcement learning. IEEE Intell. Transp. Syst. Mag. 14, 122–134 (2022). https://doi.org/10.1109/MITS.2022.3198019
16. Zhong, X., Zhou, Y., Yang, Z., Yang, W. & Lu, Z. Dynamic pricing for electric vehicle charging: an iterative two-step approach based on deep reinforcement learning. Res. Transp. Bus. Manag. 64, 101535 (2026). https://doi.org/10.1016/j.rtbm.2025.101535
17. Wang, S., Bi, S. & Zhang, Y.-J. A. Reinforcement learning for real-time pricing and scheduling control in EV charging stations. IEEE Trans. Ind. Inform. 17, 849–859 (2021). https://doi.org/10.1109/TII.2019.2950809
18. Liu, Y. et al. An aggregator-based dynamic pricing mechanism and optimal scheduling scheme for the electric vehicle charging. Front. Energy Res. 10, 1037253 (2023). https://doi.org/10.3389/fenrg.2022.1037253
19. Mrkos, J., Komenda, A., Fiedler, D. & Vokřínek, J. Online dynamic pricing for electric vehicle charging stations with reservations. IEEE Trans. Intell. Transp. Syst. 26, 13882–13897 (2025). https://doi.org/10.1109/TITS.2025.3565906
20. Luo, S. & Wang, Y. The impacts of rate surge on electric vehicle charging behaviors: evidence from California. J. Econ. Behav. Organ. 236, 107048 (2025). https://doi.org/10.1016/j.jebo.2025.107048

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles.zip