Threshold Policies for Unit Commitment as a Markov Decision Process

Kabir Murjani - 23BEE064

Electrical Power System Analysis (3EE201CC24)*
Professor Santosh Vora, Professor Akhilesh Nimje

Abstract

Unit commitment (UC) is naturally a sequential decision problem, yet its obvious state of on/off status plus the exogenous price or load is not Markov: minimum up/down times and offline-dependent startup costs make feasible actions and costs depend on history. I prove that a capped time-in-status counter with values restores the Markov property and is minimal. For the 72 thermal units of RTS-GMLC the minimal state is $10^{48.6}$ times the status-only state. For a price-taking unit facing a stochastically monotone Markov price, I prove that with UT = DT = 1 and constant startup cost the optimal policy is a pair of time-varying price thresholds with hysteresis, via a recursion for the on–off value difference.

1 TL;DR

Unit commitment (UC) decides which generators are on in each period, subject to minimum up and down times and startup costs that grow with the time a unit has been cold. Deterministic UC is solved as a mixed-integer program; stochastic variants use scenarios or a Markov model of load or price. A Markov decision process (MDP) view promises *structural* results—threshold policies of the kind known for entry and exit under switching costs but first one must say what the state is. The obvious candidate, price and previous status (p_t, u_{t-1}), is insufficient: whether a unit may switch and what a start costs depend on how long it has been in its current status. DP treatments of generator operation with minimum up/down times therefore track the time

*Department of Electrical Engineering, Nirma University, Assignment

in status. I make that construction precise, prove it minimal, and ask what structure the optimal policy has on it.

2 Model

Price. Periods are $t = 0, \dots, T-1$. The price is $p_t = \mu_t + x_t$ with a deterministic profile μ_t and a Markov chain x_t on a grid $x_1 < \cdots < x_K$ with transition matrix P, obtained by discretising a mean-reverting AR(1) process with the RouInhorst. I write $p_{t,k} = \mu_t + x_k$ and $(Pf)(k) = \sum_l P_{kl} f(l)$.

Assumption 1. P is stochastically monotone: $\sum_{l \geq j} P_{kl}$ is nondecreasing in k for every j. Equivalently, Pf is nondecreasing whenever f is.

Rather than rely on a general result for the discretisations, the code asserts Assumption 1 for every matrix it builds (10,753 chains, no failure).

Unit. A unit has output limits $P^{\min} \leq P^{\max}$, marginal cost c, no-load cost F, minimum up and down times $\mathrm{UT}, \mathrm{DT} \geq 1$ and a nondecreasing, piecewise-constant startup cost $S(\tau)$ that depends on the number $\tau \geq 1$ of periods offline (hot/warm/cold). Let $T_{\mathrm{cold}} = \min\{\tau \geq 1 : S(\tau') = S(\tau)\ \forall \tau' \geq \tau\}$. The on-profit is

$$\begin{aligned} r(p) &= \max_{q \in [P^{\min}, P^{\max}]} (p-c)q - F \\ &= \begin{cases} (p-c)P^{\max} - F, & p \geq c, \\ (p-c)P^{\min} - F, & p < c, \end{cases} \end{aligned} \tag{1}$$

which is nondecreasing and convex; the off-profit is 0. I write $r_t(k) = r(p_{t,k})$.

Augmented state and timing. The decision state at the start of period t is $s_t = (k_t, u_{t-1}, \tau_t)$, where u_{t-1} is the status in period $t-1$ and τ_t counts the consecutive periods, ending at $t-1$, spent in that status, capped at UT if $u_{t-1} = 1$ and at $J := \max(\mathrm{DT}, T_{\mathrm{cold}})$ if $u_{t-1} = 0$. The action $a_t \in \{0, 1\}$ is the status in period t. The feasible set is $A(1, \tau) = \{1\}$ for $\tau < \mathrm{UT}$, $A(0, \tau) = \{0\}$ for $\tau < \mathrm{DT}$, and $\{0, 1\}$ otherwise. *Timing convention:* a unit started in period t ($u_{t-1} = 0$, $a_t = 1$) pays $S(\tau_t)$ and earns $r(p_t)$ in period t. The reward is $R_t(s, a) = a\, r_t(k) - a(1-u)S(\tau)$, and the counter moves deterministically: $\tau_{t+1} = \min(\tau_t + 1, \mathrm{cap}(u))$ if $a = u$ and $\tau_{t+1} = 1$ otherwise. With discount $\gamma \in (0, 1]$ and $V_T \equiv 0$,

$$\begin{aligned} V_t(k, u, \tau) = \max_{a \in A(u,\tau)} \Big\{ & R_t(k, u, \tau, a) \\ & + \gamma \big(P V_{t+1}(\cdot, a, \tau') \big)(k) \Big\}. \end{aligned} \tag{2}$$

Table 1: RTS-GMLC units used n_g is the number of counter values).

Unit	Type	$P^{\min}$ MW	$P^{\max}$ MW	c \$/MWh	F \$/h	UT h	DT h	T_{cold} h	S (hot/warm/cold) k\$	n_g
101_STEAM_3	Coal	30	76	16.4	349	8	4	12	7.1/10.3/11.2	20
115_STEAM_3	Coal	62	155	23.3	55	8	8	60	14.6/15.7/22.8	68
123_STEAM_3	Coal	140	350	21.7	546	24	48	96	20.6/21.4/36.7	120
107_CC_1	Gas CC	170	355	26.8	209	8	5	2	12.4/17.6/28.0	13
113_CT_1	Gas CT	22	55	28.9	487	3	3	1	1.8/4.4/5.7	6
115_STEAM_1	Oil ST	5	12	127.7	259	4	2	12	0.4/0.5/0.7	16
101_CT_1	Oil CT	8	20	101.0	278	1	1	1	0.05/0.05/0.05	2

Since the counter transition is deterministic, (2) is evaluated for all (k, τ) at once with index arrays; per period the cost is $O(K^2 n + Kn)$ with $n = \text{UT}+J$. For a decision state I write G_t for the advantage $Q_t(\text{on}) - Q_t(\text{off})$; ties are broken towards "on".

3 Non-Markovianity

In the *true* problem the history of statuses is given, including periods before $t = 0$, and the constraints are stated on it: a unit that switches on (off) must stay on (off) for UT (DT) periods, and a start after an offline run of length ℓ costs $S(\ell)$.

Proposition 1 (the status is not a sufficient state)**.** *Suppose* $\text{UT} \geq 2$*, or* $\text{DT} \geq 2$*, or* $S(\tau_1) \neq S(\tau_2)$ *for some* $\text{DT} \leq \tau_1 < \tau_2$*. Then there are histories with the same* (p_t, u_{t-1}) *but different feasible sets or different startup costs at* t*.*

Proof. If $\text{DT} \geq 2$, take histories ending in $(1, 0)$ and $(1, 0^{\text{DT}})$: starting at t is infeasible after the first and feasible after the second. If $\text{UT} \geq 2$, the histories $(0, 1)$ and $(0, 1^{\text{UT}})$ differ in whether a shut-down is alloId. In the third case the histories $(1, 0^{\tau_1})$ and $(1, 0^{\tau_2})$ both allow a start, at costs $S(\tau_1) \neq S(\tau_2)$. □

Consequently, even under a fixed policy, the status process need not be Markov. With $\text{UT} = 2$, $\text{DT} = 1$ and the policy "switch whenever alloId", the status sequence is $1, 1, 0, 1, 1, 0, \ldots$, so $\Pr(u_{t+1} = 0 \mid u_t = 1)$ is 0 or 1 depending on u_{t-1}.

Theorem 2 (the counter restores the Markov property)**.** *For every history, the feasible set and the reward at* t *are functions of* $s_t = (k_t, u_{t-1}, \tau_t)$ *and* a_t *alone, and* $\Pr(s_{t+1} \mid \textit{history}, a_t) = P_{k_t k_{t+1}} \mathbf{1}\{u_t = a_t,\ \tau_{t+1} = \tau'(u_{t-1}, \tau_t, a_t)\}$.

Hence (2) *is a finite MDP, and a deterministic Markov policy on s_t is optimal among all history-dependent policies.*

Proof. Let ℓ be the uncapped length of the current run, so $\tau_t = \min(\ell, \text{cap})$. A shut-down is alloId iff $\ell \geq \text{UT}$ iff $\tau_t = \text{UT}$. A start is alloId iff $\ell \geq \text{DT}$ iff $\tau_t \geq \text{DT}$, because $J \geq \text{DT}$. The startup cost is $S(\ell) = S(\min(\ell, J))$ because S is constant on $[T_{\text{cold}}, \infty)$ and $J \geq T_{\text{cold}}$. The run length evolves as $\ell \mapsto \ell + 1$ or 1, and capping commutes with this update: $\min(\ell + 1, m) = \min(\min(\ell, m) + 1, m)$. The price is exogenous and Markov. □

Minimality. View the counter as a deterministic Mealy machine $\mathcal{M}$ with input alphabet $\{0, 1\}$ (actions), a partial transition $\delta(q, a)$ that is defined iff a is feasible at q, and output $\lambda(q, a)$, the startup cost incurred. Two states q, q' are *equivalent* if every action word is feasible from both or from neither, and produces the same output sequence when feasible. A statistic of the history is *sufficient for the constraints* if histories with the same statistic have the same future feasible words and cost sequences. By the Myhill–Nerode theorem, the coarsest such statistic is the class of the history under this equivalence, and its range is the set of equivalence classes of reachable states.

Theorem 3 (minimality)**.** *Assume $S(\tau) > 0$ for $\tau \geq \text{DT}$. All $\text{UT} + J$ counter states are reachable and pairwise inequivalent. Hence any statistic sufficient for the constraints takes at least $\text{UT} + \max(\text{DT}, T_{\text{cold}})$ values, and the counter is minimal.*

Proof. The state on(i) is reached by the history $(0, 1^i)$ and off(j) by $(1, 0^j)$. I exhibit a distinguishing word for each pair. (a) on(i) vs. on(i'), $i < i'$: the word $1^{\text{UT}-i'}0$ is feasible from on(i'), but from on(i) the final shut-down comes after only $i + \text{UT} - i' < \text{UT}$ periods. (b) off(j) vs. off(j'), $j < j'$. If $j < \text{DT}$, the word $0^m 1$ with $m = \max(0, \text{DT} - j')$ is feasible from off(j') but not from off(j), since $j + m < \text{DT}$. If $\text{DT} \leq j$, then $J > \text{DT}$, so $J = T_{\text{cold}}$. The word $0^{T_{\text{cold}}-j'}1$ is feasible from both and costs $S(T_{\text{cold}})$ versus $S(T_{\text{cold}} - (j' - j)) \leq S(T_{\text{cold}} - 1) < S(T_{\text{cold}})$, by minimality of T_{cold} and monotonicity of S. (c) on(i) vs. off(j). If $i < \text{UT}$, the word 0 is feasible only from off(j). If $i = \text{UT}$ and $j < \text{DT}$, the word 1 is feasible only from on(UT). If $i = \text{UT}$ and $j \geq \text{DT}$, the word 1 costs 0 versus $S(j) > 0$. □

The positivity assumption is needed only in the last case. Without it (for example $S \equiv 0$, $\text{UT} = \text{DT} = 1$), Moore minimisation collapses the automaton to 1 class. The distinguishing words have length at most $\text{UT} + J$,

so the distinctions matter whenever that many periods remain. I check Theorem 3 by Moore partition refinement on 2,000 random units (Table 2).

Corollary 4 (value distinguishability)**.** *If also $P^{\min} > 0$, then for any two counter states $q \neq q'$ there is a deterministic price path of length $L \leq \mathrm{UT} + J$ (a one-state chain) such that $V_0(q) \neq V_0(q')$. Consequently, no value function defined on a statistic that merges q and q' is correct for every price input.*

Proof. Let w be a shortest distinguishing word; its prefixes are feasible from both states with equal costs. Set $p_t = c + M$ if $w_t = 1$ and $c - M$ otherwise, with M so large that $\min(MP^{\max} - F,\, MP^{\min} + F) > L \max_\tau S(\tau)$. Every deviation from w then loses more than any possible startup saving, so w is the unique optimal word wherever it is feasible. If w is feasible from q only, then $V_0(q') < V_0(q)$. If it is feasible from both, the values differ by the difference in the last startup cost. □

Proposition 5 (state-space size)**.** *A single unit with a K-state price chain has $K(\mathrm{UT} + J)$ states per period. A fleet of N units facing a shared net-load chain with $|\mathcal{D}|$ states has $|\mathcal{D}| \prod_{g=1}^{N} n_g$ states, with $n_g = \mathrm{UT}_g + \max(\mathrm{DT}_g, T_{\mathrm{cold},g})$, and 2^N joint actions. The joint counter is again minimal.*

Proof. The counting is immediate. For minimality, two joint states differ in some unit g. Apply the distinguishing word of Theorem 3 to unit g while every other unit keeps its status, which is always feasible and free, and the dispatch cost is the same under both joint states. So the joint outputs differ. □

Remark 6. Ramping limits $|q_t - q_{t-1}| \leq R$ would add the continuous previous output q_{t-1} to the state. I leave this extension aside.

4 Threshold structure

4.1 Base case: proved

Theorem 7 (thresholds with hysteresis)**.** *Let* $\mathrm{UT} = \mathrm{DT} = 1$*, $S(\tau) \equiv S \geq 0$, $\gamma \in (0, 1]$, and let Assumption 1 hold. The state is then (k, u). Let $W_t(k, a) = a\, r_t(k) + \gamma(PV_{t+1}(\cdot, a))(k)$ and $\Delta_t(k) = W_t(k, 1) - W_t(k, 0)$. Then:* (a) *$V_t(k, 1) - V_t(k, 0) = \mathrm{clamp}(\Delta_t(k), 0, S)$, hence $\Delta_t = r_t + \gamma P\, \mathrm{clamp}(\Delta_{t+1}, 0, S)$ with $\Delta_T \equiv 0$;* (b) *$\Delta_t(k)$ is nondecreasing in k for every t;* (c) *an off unit starts iff $p \geq p_t^{\uparrow} := \min\{p_{t,k} : \Delta_t(k) \geq S\}$, an on unit shuts down iff $p \leq p_t^{\downarrow} := \max\{p_{t,k} : \Delta_t(k) < 0\}$, and $p_t^{\downarrow} < p_t^{\uparrow}$ whenever both exist.*

Proof. (a) $V_t(k,1) = \max(W_1, W_0)$ and $V_t(k,0) = \max(W_1 - S, W_0)$. With $D = W_1 - W_0$, the difference is $\max(D,0) - \max(D-S,0)$, which equals 0, D or S according as $D \le 0$, $0 \le D \le S$ or $D \ge S$; that is, $\text{clamp}(D, 0, S)$. Substituting into the definition of Δ_t gives the recursion. (b) By backward induction. $\Delta_T \equiv 0$. If Δ_{t+1} is nondecreasing, then so is $\text{clamp}(\Delta_{t+1}, 0, S)$, and by Assumption 1 so is its image under P; r_t is nondecreasing because r is and x_k increases with k. (c) From on the unit stays iff $\Delta_t \ge 0$, and from off it starts iff $\Delta_t \ge S$; by (b) these are upper sets in k. Since $S \ge 0$, $\{\Delta_t \ge S\} \subseteq \{\Delta_t \ge 0\}$. □

Equivalently, $Q_t(k,u,a)$ has increasing differences in (k,a) and in (u,a), since $-Sa(1-u)$ is supermodular. So a^* is nondecreasing in (k,u), as Topkis's theorem predicts. For prices in the band $[p_t^{\downarrow}, p_t^{\uparrow})$ the unit keeps its current status.

Corollary 8 (break-even bracketing)**.** *Under Theorem 7, $r_t \le \Delta_t \le r_t + \gamma S$. A unit shuts down only if $r(p) < 0$ and starts only if $r(p) \ge (1-\gamma)S$. For $\gamma = 1$, $p_t^{\downarrow} < p_{\text{be}} \le p_t^{\uparrow}$, where $p_{\text{be}} = \min\{p : r(p) \ge 0\}$.*

Proof. Since $0 \le \text{clamp}(\cdot, 0, S) \le S$, part (a) gives $r_t \le \Delta_t \le r_t + \gamma S$. A shut-down requires $\Delta_t < 0$ and a start requires $\Delta_t \ge S$. □

Proposition 9 (thresholds rise towards the horizon)**.** *Under Theorem 7, if in addition $\mu_t \equiv \mu$, then $\Delta_t \ge \Delta_{t+1}$ pointwise for $t \le T-2$, so $p_t^{\uparrow}$ and $p_t^{\downarrow}$ are nondecreasing in t.*

Proof. With $r_t \equiv r$, I have $\Delta_t - \Delta_{t+1} = \gamma P[\text{clamp}(\Delta_{t+1}) - \text{clamp}(\Delta_{t+2})]$, and clamp and P preserve order. For the base step, $\text{clamp}(\Delta_{T-1}) \ge 0 = \text{clamp}(\Delta_T)$. Hence $\{\Delta_{t+1} \ge S\} \subseteq \{\Delta_t \ge S\}$ and $\{\Delta_t < 0\} \subseteq \{\Delta_{t+1} < 0\}$. □

4.2 Beyond the base case

With general $(\text{UT}, \text{DT}, S(\tau))$ the decision states are on(UT) and off(τ) with $\tau \ge \text{DT}$. The argument of Theorem 7 would need $V_t(\cdot, \text{on}, i) - V_t(\cdot, \text{off}, j)$ to be nondecreasing in price. This supermodularity fails in 2,237 of 3,000 random instances (Table 3), so the proof does not extend. Two positive results survive.

Theorem 10 (i.i.d. prices)**.** *If all rows of P are equal, then for every UT, DT, nondecreasing $S(\cdot)$ and γ, each start set $\{k : G_t(k,\tau) \ge 0\}$ ($\tau \ge \text{DT}$) is an upper set in price and each shut-down set is a loIr set.*

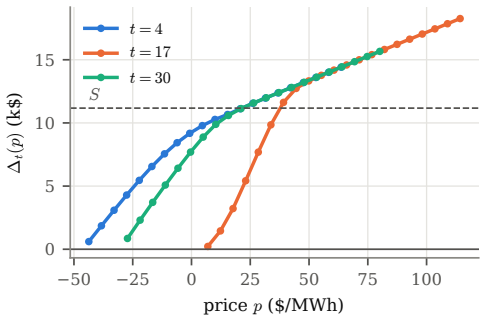


Figure 1: $\Delta_t(p)$ of Theorem 7 for the base-case variant of 101_STEAM_3 (UT = DT = 1, $S = S_{\text{cold}}$) at three periods: nondecreasing in p. An off unit starts where $\Delta_t \geq S$ (dashed line); an on unit shuts down where $\Delta_t < 0$, which does not occur at these periods.

Proof. The continuation term $\gamma(PV_{t+1}(\cdot, s'))(k)$ does not depend on k. So $G_t(k, \tau) = r_t(k) - S(\tau) + \text{const}$ and the stay-on advantage is $r_t(k) + \text{const}$, both nondecreasing in k. □

Proposition 11 (sawtooth in offline time)**.** *For any P (monotone or not),* UT, DT*, nondecreasing S and $\gamma \in [0, 1]$: (a) $V_t(k, \mathit{off}, m) \geq V_t(k, \mathit{off}, m')$ for* $\text{DT} \leq m \leq m' \leq J$*; (b) the start advantage satisfies $G_t(k, \tau+1) - G_t(k, \tau) \geq -[S(\tau+1) - S(\tau)]$. In particular, on each segment where S is constant, the start set at $\tau + 1$ contains the start set at τ (for every t, k), so $p_t^{\uparrow}(\tau)$ is nonincreasing in τ within the segment.*

Proof. (a) By backward induction; $V_T \equiv 0$. At t both states may start or wait. Starting yields $r_t - S(m) + \gamma PV_{t+1}(\text{on}, 1) \geq r_t - S(m') + \gamma PV_{t+1}(\text{on}, 1)$. Waiting yields $\gamma PV_{t+1}(\text{off}, m^+) \geq \gamma PV_{t+1}(\text{off}, m'^+)$ by the induction hypothesis, where $m^+ = \min(m + 1, J) \leq m'^+$. (b) $G_t(k, \tau) = r_t(k) - S(\tau) + \gamma P[V_{t+1}(\text{on}, 1) - V_{t+1}(\text{off}, \tau^+)](k)$, so the difference in τ equals $-[S(\tau+1) - S(\tau)]$ plus a term that is nonnegative by (a). □

Proposition 11 explains the dips in Figure **??**. Just before the hot-start window closes, the unit starts even at low prices ("start while it is still hot"), and its threshold then jumps at the cost step.

Proposition 12 (each relaxation can break the threshold property)**.** *Assumption 1 holds in all three examples. Take $c = 10$, $\gamma = 1$ and two price scenarios p^L and $p^H = p^L + \delta$ with $P = I$ in (a) and (b), which is stochastically monotone.* (a) UT $= 2$, DT $= 1$, $S \equiv 8$, $P^{\min} = 0$, $P^{\max} = 2$,

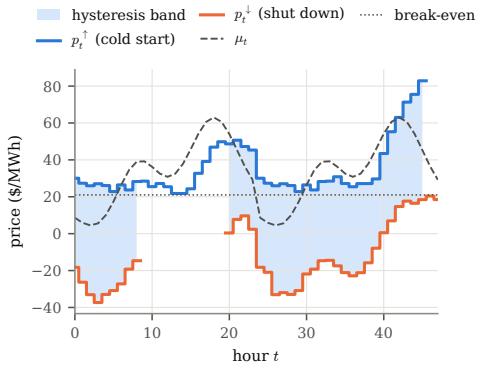


Figure 2: Start (cold) and shut-down thresholds of 101_STEAM_3 over the $T = 48$-hour horizon. Gaps: no price on the grid triggers the action.

$F = 9$, $p^L = (0, 35, 5)$, $\delta = 10$, *unit on with* $\tau = \mathrm{UT}$ *at* $t = 0$*: the unit* stays on *under* p^L *and* shuts down *under* p^H. (b) $\mathrm{UT} = 1$, $\mathrm{DT} = 2$, $S \equiv 9$, $P^{\min} = P^{\max} = 1$, $F = 3$, $p^L = (30, -15, 25, 15)$, $\delta = 5$, *unit off with* $\tau = 2$*: the unit* starts *under* p^L *and* waits *under* p^H. (c) $\mathrm{UT} = \mathrm{DT} = 1$, $S(\tau) = 0, 0, 5, 17$ *for* $\tau = 1, 2, 3, \geq 4$, $P^{\min} = 0$, $P^{\max} = 1$, $F = 7$, $\mu = (-15, 5, 10, 35, 10)$, $x \in \{0, 15\}$, $P = \left(\begin{smallmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{smallmatrix}\right)$, *unit off with* $\tau = 2$*:* $G_0 = 4.30$ *in the low state and* -1.30 *in the high state.*

Proof. (a) $r(p) = 2(p - 10) - 9$ for $p \geq 10$ and -9 otherwise, so $r(p^L) = (-9, 41, -9)$ and $r(p^H) = (-9, 61, 1)$. Staying on is worth $-9+41+\max(-9, 0) = 32$ under p^L. Shutting down and restarting at $t = 1$ is worth $-8+41-9 = 24$, because $\mathrm{UT} = 2$ forces the unit on at $t = 2$, and every other plan is worth at most 0. Under p^H the two plans are worth $-9+61+1 = 53$ and $-8+61+1 = 54$. Thus $G = 8 > 0$ at the low price and $G = -1 < 0$ at the high price. (b) $r = p - 13$, so $r(p^L) = (17, -28, 12, 2)$ and $r(p^H) = (22, -23, 17, 7)$. Under p^L the best start-now plan is to start and shut down at $t = 1$, worth 8; $\mathrm{DT} = 2$ then forbids a restart at $t = 2$, and a restart at $t = 3$ loses money. The best waiting plan starts at $t = 2$ and is worth $12 - 9 + 2 = 5$. Under p^H the best start-now plan runs throughout ($22 - 9 - 23 + 17 + 7 = 14$), while waiting is worth $17 - 9 + 7 = 15$. Thus $G = 3$ and $G = -1$. (c) Exact DP. Both scenarios have $r = -7$ at $t = 0$, and after waiting the optimal policy starts at $t = 1$ in both states, at cost $S(3) = 5$. Starting at $t = 0$ is free ($S(2) = 0$) but adds an hour at $r = -7$. In the low state the hour at $t = 1$ loses money as Ill, so the free start now is cheaper; in the high state that hour is profitable, and waiting avoids the loss at $t = 0$. □

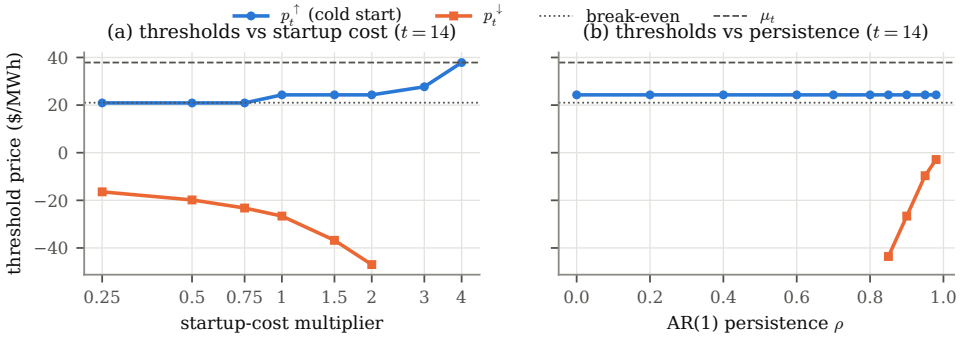


Figure 3: Thresholds of 101_STEAM_3 at $t = 14$ versus a multiplier on all startup costs (a) and versus ρ with the stationary s.d. fixed (b). Grid $K = 51$; missing points mean the action is never taken.

The hand-computed values in (a) and (b) are asserted against the DP in the pipeline, which also confirms them by exhaustive enumeration. The mechanisms are the ones one would expect. In (a), UT locks the unit on for a later hour whose profit changes sign with the price level. In (b), DT locks it out of a later window. In (c), a free start now competes with a paid start an hour later, which is preferable only when that hour is profitable. In each case the competing plans are on in different hours, so their values respond to a price increase at different rates; the kink of r at c in (1) (slope $P^{\min}$ below, $P^{\max}$ above) amplifies this in (a) and (c). An adversarial hill-climbing search (Section 5) finds violations for every relaxed class, including deterministic scenarios and RouInhorst chains with $\rho \leq 0.95$, $K \geq 11$. It finds none under the hypotheses of Theorems 7 and 10.

4.3 Three conjectures, tested

Conjecture 1 (as posed)**.** *With general* $(\mathrm{UT}, \mathrm{DT}, S(\tau))$*: (i)* $p_t^{\uparrow}(\tau)$ *is monotone in* τ*; (ii) thresholds rise towards the end of the horizon; (iii) thresholds decrease as the price persistence* ρ *increases.*

Resolution (Table 3). (i) is *false*: $p_t^{\uparrow}(\tau)$ falls within constant-cost segments (Proposition 11, Figure **??**) and rose or stayed level across cost steps in all but 234 of 102,552 cases. (ii) is *proved* in the base case with stationary prices (Proposition 9) and *false* in general: with stationary prices, 77 of 1,000 general instances have a start threshold, and 29 a shut-down threshold, that decreases in t somewhere. (iii) is *false even in the base case.* Varying $\rho \in \{0, \ldots, 0.95\}$ at fixed stationary variance, $p^{\uparrow}$ increased with ρ somewhere

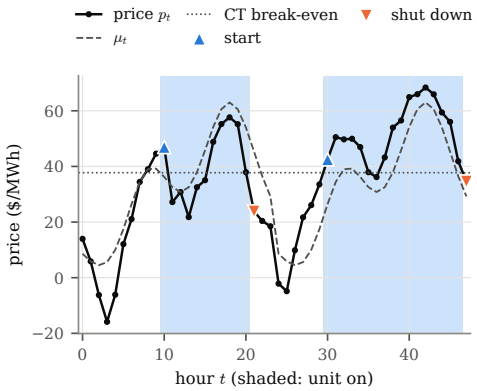


Figure 4: Optimal commitment of the gas CT 113_CT_1 (UT = 3, DT = 3, starting cold) on a simulated price path.

in 2.3% of the (instance, t) columns with the threshold above the mean price and in 40.6% of those below it; $p^{\downarrow}$ increased in 62.1% (general-case counts: `results/persistence.json`). This is consistent with persistence making low prices last.

5 Numerical results

Setup. Unit data come from the RTS-GMLC generator table, downloaded by the pipeline). I replace each unit's heat-rate curve by its chord betIen $P^{\min}$ and $P^{\max}$ and read startup costs as start heat × fuel price. I interpret the *Start Time Warm/Cold* columns as the offline durations at which the warm and cold costs apply, and round all durations up to whole hours (Table 1). The price process is *illustrative*: a two-day hourly profile μ_t betIen \$4 and \$63/MWh, and a RouInhorst AR(1) with $\rho = 0.9$, stationary s.d. \$12/MWh and $K = 21$, over $T = 48$. The fleet uses five RTS units (518 MW) and an illustrative net load (mean profile betIen 218 and 440 MW, $\rho = 0.9$, s.d. 6% of capacity, $|\mathcal{D}| = 9$, coarser than the price grid to keep the $N = 6$ timing run short), with unserved load and over-generation both priced at \$1,000/MWh. Expectations are exact; Monte Carlo (5,000 paths) serves only as a cross-check and to count overrides.

Policy structure (Figures **??**–4). On all five RTS units every decision state has a price threshold (Table 2). The coal unit's hysteresis band is wide (Figure 2): at night it keeps running far below break-even rather than pay $S(\tau)$ and wait out DT. Larger startup costs widen the band; as ρ grows, $p^{\downarrow}$

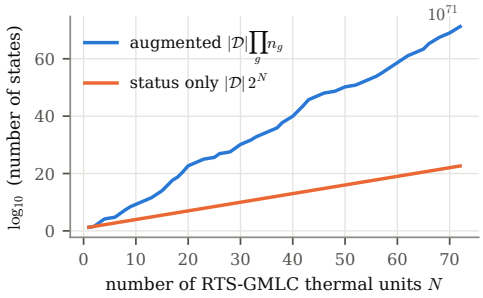


Figure 5: State-space size as the RTS-GMLC thermal units are added in file order.

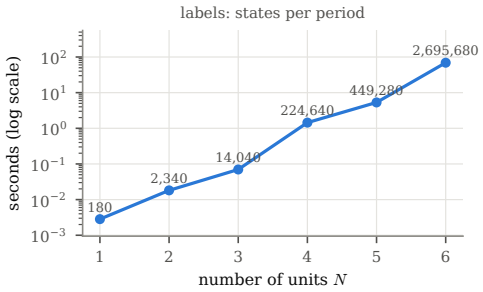


Figure 6: Exact fleet DP solve time; labels give the number of states per period.

rises while $p^{\uparrow}$ stays put (Figure 3).

State-space size (Table 5, Figures 5–6). Counters range from $n_g = 2$ to 120 (the 350 MW coal unit: UT = 24, DT = 48, $T_{\text{cold}} = 96$). For the 72 thermal units the minimal state has about $10^{71.2}$ elements, against $10^{22.6}$ for the insufficient status-only state. Exact DP is therefore limited to small fleets: 449,280 states and 2^5 actions per period ($N = 5$) solve in 5.4 s; $N = 6$ (2,695,680 states) takes 69.2 s (Table 6).

Cost of ignoring history (Table 4, Figure **??**). The naive model solves the MDP on (p, u) with UT = DT = 1 and one startup cost (hot, warm or cold). Its policy runs on the true system: illegal actions are overridden by keeping the status, the true $S(\tau)$ is charged, and the value is computed exactly by policy evaluation on the augmented chain. The mean-path baseline solves deterministic UC on $\mathbb{E}[p_t]$ by MIP (HiGHS via SciPy; Egret was not used; the fleet uses a deterministic DP) and evaluates the fixed schedule exactly. With hot costs the naive model loses up to 29.9% (fleet: 10.9%); with

cold costs at most 2.2% (fleet: 1.2%). For the flexible units the naive-hot policy over-cycles: the CC unit makes 1.6 expected starts per horizon instead of 0.6, the CT 1.9 instead of 1.3. The largest loss, for the 350 MW coal unit, is instead a feasibility effect. The naive model shuts the unit down expecting a cheap restart, and the 48-hour minimum down time then locks it out: it is off 38% of the hours (optimal: 1%), and in 18.6% of its hours the naive action is illegal. Elsewhere overrides are rare (Table 4). The mean-path schedule cannot react to deviations and is worst for the peaking CT (73.4%) and the fleet (23.2%).

Verification. Every check in Table 2 passes. Among the unclaimed properties (Table 3), the threshold property never failed on 3,000 random instances with smooth daily profiles (0 violations), although the adversarial search shows it can; starting at a current loss (1,393 instances) and shutting down at a current profit (35) show that Corollary 8 does not extend beyond the base case.

6 Conclusion

A unit's status is not a state; status plus a time-in-status counter capped at UT and $\max(\mathrm{DT}, T_{\mathrm{cold}})$ is, and no smaller statistic is. The two-threshold policy with hysteresis is a theorem only in a narrow case: minimum up times, minimum down times and offline-dependent startup costs each break single crossing, though thresholds survive under i.i.d. prices and start sets are nested in offline time. Thresholds Ire the rule in our smooth-profile instances; a general guarantee is open.

Table 2: Verification of the proved statements and of the implementation (exact DP; tolerance 10^{-9} relative to the value scale).

Check	Result	Tested	Pass	Fail
Price chains stochastically monotone (every chain built)	Assumption	10,753	10,753	0
DP optimum = exhaustive enumeration of all 2^{10} schedules (deterministic prices)	Thm. 2	200	200	0
Augmented-state DP = DP over full status histories (stochastic prices)	Thm. 2	30	30	0
DP optimum = HiGHS MIP optimum ($T = 24$, deterministic prices)	Thm. 2	100	100	0
Moore minimisation of the counter automaton leaves $\mathrm{UT} + \max(\mathrm{DT}, T_{\mathrm{cold}})$ classes	Thm. 3	2,000	2,000	0
Every pair of counter values has different optimal value on the constructed price path	Cor. 4	15,710	15,710	0
$\Delta_t = r + \gamma\,\mathbb{E}[\mathrm{clamp}(\Delta_{t+1}, 0, S)]$ (every t, p)	Thm. 7(a)	1,000	1,000	0
$\Delta_t(p)$ nondecreasing in p (every t)	Thm. 7(b)	1,000	1,000	0
Start/shut-down sets are upper/lower price sets	Thm. 7(c)	1,000	1,000	0
Hysteresis $p_t^{\downarrow} \le p_t^{\uparrow}$	Thm. 7(c)	1,000	1,000	0
Shut down only if $r(p) < 0$; start only if $r(p) \ge (1-\gamma)S$	Cor. 8	1,000	1,000	0
Stationary prices: $\Delta_t \ge \Delta_{t+1}$, thresholds nondecreasing in t	Prop. 9	500	500	0
i.i.d. prices: every decision state has a price threshold	Thm. 10	1,000	1,000	0
Start sets nested in τ within constant-S segments ($G_t(\tau{+}1) \ge G_t(\tau)$)	Prop. 11	3,000	3,000	0
Counterexamples (Prop. 12) reproduce: $G > 0$ at low price, $G < 0$ at high price	Prop. 12	3	3	0
Base-case RTS-GMLC units: every decision state has a price threshold	(numerical)	5	5	0
Exact policy evaluation agrees with Monte Carlo ($\lvert\cdot\rvert \le 4$ s.e.)	—	20	20	0
Mean-path HiGHS MIP = deterministic DP on the same path	—	5	5	0
Merit-order dispatch cost = LP optimum (HiGHS)	—	300	300	0
Fleet DP = enumeration of all joint schedules ($N = 2$, $T = 5$)	Thm. 2	5	5	0
Fleet: exact policy evaluation agrees with Monte Carlo (≤ 4 s.e.)	—	4	4	0

Table 3: Numerical findings on properties that are not claimed as theorems, and the adversarial search.

Property (not claimed as a theorem)	Tested	Violations	Unit
Threshold property (random smooth instances)	3,000	0	instances
Advantage G monotone in p	3,000	22	instances
$V(\text{on}, i) - V(\text{off}, j)$ monotone in p (supermodularity)	3,000	2,237	instances
Start only if $r(p) \geq (1-\gamma)S(\tau)$	3,000	1,393	instances
Shut down only if $r(p) \leq 0$	3,000	35	instances
B3(i): $p_t^{\uparrow}(\tau)$ nondecreasing across a cost step	102,552	234	pairs
B3(ii): stationary prices, start thresholds nondecreasing in t	1,000	77	instances
B3(ii): stationary prices, shut-down thresholds nondecreasing in t	1,000	29	instances
B3(iii) [B1]: $p^{\uparrow}$ nonincreasing in ρ (threshold above mean)	3,234	73	(inst, t)
B3(iii) [B1]: $p^{\uparrow}$ nonincreasing in ρ (threshold below mean)	3,680	1,494	(inst, t)
B3(iii) [B1]: $p^{\downarrow}$ nonincreasing in ρ	3,479	2,160	(inst, t)
Adversarial search for threshold violations (4 hill-climbing runs of 20,000 evaluations per class)			
Structural class	Evaluations	Runs with violation	
$\text{UT} = \text{DT} = 1$, constant S (Thm. 7)	80,000	0/4	
i.i.d. prices, any $\text{UT}, \text{DT}, S(\tau)$ (Thm. 10)	80,000	0/4	
$\text{UT} \geq 1$, $\text{DT} = 1$, constant S	8,495	4/4	
$\text{UT} = 1$, $\text{DT} \geq 1$, constant S	7,014	4/4	
$\text{UT} = \text{DT} = 1$, τ-dependent S	2,918	4/4	
$\text{UT} = \text{DT} = 1$, τ-dependent S, deterministic scenarios ($P = I$)	4,567	4/4	
any UT, DT, constant S	3,966	4/4	
any $\text{UT}, \text{DT}, S(\tau)$	4,789	4/4	
any $\text{UT}, \text{DT}, S(\tau)$; Rouwenhorst, $\rho \leq 0.95$, $K \geq 11$	4,001	4/4	

Table 4: Loss relative to the augmented-state optimum (percent of expected profit, or of expected cost for the fleet) and the largest rate of overridden illegal actions.

	Optimal	Naive (p, u) model, $S \equiv$			Mean-path	Max.
Instance	value ($)	hot	warm	cold	MIP	overrides
101_STEAM_3	52,651	0.2%	0.1%	0.1%	0.4%	0.01%
115_STEAM_3	95,586	0.0%	0.0%	0.1%	0.3%	0.04%
123_STEAM_3	217,719	29.9%	28.9%	2.2%	0.3%	18.62%
107_CC_1	165,661	4.5%	2.2%	0.0%	4.3%	0.55%
113_CT_1	7,439	16.8%	2.9%	0.0%	73.4%	0.04%
Fleet $N = 5$ (cost)	215,939	10.9%	1.7%	1.2%	23.2%	0.50%

Table 5: Counter sizes (hours; $n_g = \mathrm{UT} + \max(\mathrm{DT}, T_{\mathrm{cold}})$) for all RTS-GMLC thermal units.

Category	Units	UT	DT	T_{cold}	n_g
Coal	16	8–24	4–48	12–96	20–120
Gas CC	10	8	5	2	13
Gas CT	27	3	3	1	6
Oil CT	12	1	1	1	2
Oil ST	7	4	2	12	16

All 72 thermal units, $|\mathcal{D}| = 9$: augmented state $|\mathcal{D}| \prod_g n_g \approx 10^{71.2}$; status-only $|\mathcal{D}|\, 2^N \approx 10^{22.6}$.

Table 6: Exact DP wall-clock times on a laptop (NumPy); single units use $K = 21$, the fleet $|\mathcal{D}| = 9$. Times vary with machine load.

Problem	States per period	Actions	Solve time (s)
Single units (5)	126–2,520	2	$\leq$0.007
Fleet $N = 1$	180	2	0.00283
Fleet $N = 2$	2,340	4	0.0182
Fleet $N = 3$	14,040	8	0.0699
Fleet $N = 4$	224,640	16	1.44
Fleet $N = 5$	449,280	32	5.32
Fleet $N = 6$	2,695,680	64	69.2
Whole pipeline			266