Convergence to multiple price equilibria in algorithmic collusion

Krzysztof Siegieda[a]

Abstract. Simulation studies of algorithmic collusion often summarise results through average prices and profits. Such summaries conceal the variety and distribution of distinct prices, cycles, and collusive regimes. This limitation is salient given that game theory predicts multiple equilibria, implying a distribution rather than a single outcome. This paper examines the heterogeneity hidden by aggregate metrics in a sequential pricing duopoly with Q-learning agents. Using an established simulation framework for studying actions of pricing algorithms, the simulations cover three memory configurations: symmetric one-period, one-period versus none, and one-period versus two-period memory. K-means cluster analysis of steady-state prices reveals a spectrum: monopoly focal price equilibria, collusive focal price equilibria at non-monopoly grid points and incomplete Edgeworth cycles with supra-competitive averages. These regimes are consistent with the multiplicity of the Markov-perfect equilibria in the game theory literature. In a symmetric baseline, the monopoly outcome is a minority. Memory asymmetry eliminates or reduces cycling and increases the share of fixed-price outcomes, stabilising rather than breaking collusion. Aggregate numbers collapse qualitatively distinct results into a single figure; cluster analysis uncovers the hidden heterogeneity behind those averages and shows that algorithmic coordination spans a spectrum of regimes rather than a single outcome. For the competition policy, instruments that distinguish between pricing regimes are more informative than price-level benchmarks alone.

Keywords: algorithmic collusion, cluster analysis, reinforcement learning, Q-learning, pricing algorithms

JEL: K21, L13, L41, C63, C73

---

[a] Student at Collegium of Economic Analysis, SGH Warsaw School of Economics, ul. Madalińskiego 6/8, 02-513 Warszawa, Poland, e-mail: ks142569@student.sgh.waw.pl, ORCID: https://orcid.org/0009-0009-9839-9948.

1. Introduction

The widespread adoption of algorithmic pricing has fundamentally changed the dynamics of competition in many markets. From petrol retailing to e-commerce, firms increasingly delegate pricing decisions to self-learning algorithms that adjust prices in real time on the basis of the observed market conditions. A growing body of research warns that such algorithms, even without explicit communication, may autonomously learn to coordinate on supra-competitive prices (Calvano, Calzolari, Denicolo, Pastorello (2020); Ezrachi & Stucke, 2016). This possibility – the ‘algorithmic collusion’ – challenges the foundations of the competition policy, which traditionally relies on evidence of explicit agreements between human decision-makers (Harrington, 2018).

A key insight from the recent literature is that simple reinforcement learning algorithms, such as tabular Q-learning, are capable of converging to collusive outcomes in both simultaneous and sequential pricing games (Calvano et al., 2020; Klein, 2021). Klein (2021), in particular, demonstrates that in a sequential duopoly with a limited set of discrete prices, Q-learning agents consistently coordinate on outcomes that yield average per-period profits well above the competitive benchmark. Siegieda and Zawisza (2026) extend this framework to the asymmetric memory configurations and show that memory asymmetry stabilises the collusion rather than breaks it. However, like much of the existing literature, these studies report outcomes through aggregate metrics – mean prices and mean profits – that compress information and may hide the full distribution of pricing behaviours across simulation runs.

This paper argues that to fully understand the nature and stability of algorithmic collusion, one must look beyond averages and examine the distribution of the outcomes directly. An average price of, for example, 0.42 could reflect a mixture of fundamentally different behaviours: some runs converge to a stable monopoly price, some others to alternative focal prices, and the rest to volatile price cycles whose mean happens to lie in a similar range. From a competition-policy perspective, these distinctions matter. A market trapped in a stable collusive equilibrium at the monopoly price poses a different threat to consumer welfare than a market characterised by Edgeworth-like price cycles with a supra-competitive average. Yet standard aggregate metrics fail to capture this diversity.

I propose cluster analysis as a systematic method for identifying distinct pricing regimes – groups of simulation runs that share similar dynamic patterns – and apply it to the canonical sequential pricing framework studied by Klein (2021). The approach extracts a small number of features characterising the steady-state pricing behaviour, and then applies K-means clustering to uncover the natural groupings present in the data. Rather than reducing the simulation output to a handful of aggregate statistics, this method preserves the multi-dimensional structure of the results and allows a more comprehensive economic interpretation. Three memory configurations, i.e. symmetric one-period baseline, one-period versus no memory and one-period versus two-period memory are analysed in the study. The analysis starts from the symmetric baseline, where any heterogeneity uncovered arises endogenously from the stochastic learning dynamics rather than from built-in asymmetries. The clustering methodology has a general character and can be extended to finer price grids and alternative learning algorithms in future research.

2. Literature review

2.1. Foundations of algorithmic collusion

The question whether pricing algorithms can learn to collude without explicit communication has attracted intense scholarly attention since the seminal legal analyses of Ezrachi and Stucke (2016) and Mehra (2016). These early contributions indicated the possibility that autonomous algorithms might achieve outcomes economically equivalent to human cartels yet fall outside the existing antitrust laws because they lack an explicit agreement. Subsequent economic literature provided robust simulation evidence that this concern is justified. Calvano et al. (2020) show that independent Q-learning agents in a simultaneous-move duopoly can learn to sustain supra-competitive prices through reward-punishment strategies, with outcomes satisfying the Nash criterion. Klein (2021) extends this analysis to a sequential pricing framework based on Maskin and Tirole (1988), demonstrating that Q-learning agents with one-period memory consistently achieve supra-competitive profits. Klein, however, does not systematically report the distribution of final prices across runs, nor does he examine whether qualitatively different pricing behaviours underlie the aggregate results. Other studies reinforce these findings in related settings (Abada & Lambin, 2020; Waltman & Kaymak, 2008). A common feature across this strand of literature is the reliance on aggregate performance metrics, which, while informative about central tendencies, provide no information about the distribution of outcomes or the potential existence of distinct behavioural clusters.

2.2. Algorithmic heterogeneity and collusion

A related strand of research examines whether algorithmic collusion persists when the competing algorithms are not identical. Most existing work considers heterogeneity across algorithm classes. Den Boer et al. (2022) show that

Q-learning agents can be exploited by simpler bandit algorithms, leading to more competitive outcomes. Abada et al. (2024) and Bichler et al. (2024) find that alternative algorithms often outperform Q-learners and push prices downward, suggesting that mixing fundamentally different algorithm types might undermine collusion. Much less is known about heterogeneity within the same algorithm family – for example, when all firms use Q-learning but differ in memory depth. Abada et al. (2025) explicitly identify this as a gap. Siegieda and Zawisza (2026) address it by systematically varying memory depth in a sequential pricing duopoly, showing that memory asymmetry stabilises rather than breaks collusion. This paper builds directly on their work by applying cluster analysis to decompose aggregate results into their constituent pricing regimes and by tracing how regime composition changes across symmetric and asymmetric memory configurations.

2.3. Pricing dynamics and the analysis of regimes

Beyond the collusion-versus-competition dichotomy, several studies examine the finer structure of pricing dynamics. Maskin and Tirole (1988) characterise two types of Markov-perfect equilibria in sequential pricing: focal price equilibria and Edgeworth price cycle equilibria. Empirical research on retail petrol markets shows patterns consistent with Edgeworth cycles, with considerable variation in cycle length and amplitude (Byrne & de Roos, 2019; Eckert, 2013), suggesting that substantial heterogeneity can exist even within a broadly cyclic regime. In the algorithmic collusion literature, Klein (2021) notes that finer price grids produce Edgeworth-like cycles but does not systematically categorise their forms or their frequency in relation to fixed-price outcomes. Likewise Calvano et al. (2020), who treat convergence to different collusive price levels as incidental. This paper addresses this gap by applying cluster analysis to the steady-state pricing data from many simulation runs, identifying the natural groupings that emerge from the learning dynamics and checking whether they correspond to the canonical equilibrium types or to previously undocumented patterns.

3. Model and methodology

3.1. Sequential pricing duopoly

The environment follows the sequential pricing framework of Maskin and Tirole (1988), as adapted by Klein (2021). Two companies produce a homogeneous good at zero marginal cost and compete in an infinite-horizon, discrete-time setting. In each period $t = 1, 2, \ldots$, only one of them adjusts its price, while the other firm's price remains fixed. The firms alternate deterministically, so that Firm 1 moves in odd periods and Firm 2 in even periods. The price set by the company inactive in a given period is carried over from the previous period.

Both firms choose prices from a common discrete grid: $P = \{0, \frac{1}{k}, \ldots, 1\}$, with $k = 6$ price intervals, yielding the set $\{0, 0.167, 0.333, 0.5, 0.667, 0.833, 1\}$. This is the same grid used by Klein in his baseline specification, and it allows the direct comparability of results.

Demand is allocated according to an undercutting rule. When firm i sets price p_i and firm j sets price p_j, firm i captures the entire market and faces a linear demand schedule if $p_i < p_j$, splits the market equally if $p_i = p_j$, and sells nothing if $p_i > p_j$. Formally:

$$D_i\left(p_i, p_j\right) = \begin{cases} 1 - p_i \text{ if } p_i < p_j \\ \frac{1}{2}(1 - p_i) \text{ if } p_i = p_j. \\ 0 \text{ if } p_i > p_j \end{cases} \tag{1}$$

Given zero marginal costs, firm i's stage profit is simply:

$$\pi_i(p_i, p_j) \ = \ p_i \ \cdot \ D_i(p_i, p_j). \tag{2}$$

The monopoly price on the grid is $p^m \ = \ 0.5$, with the associated monopoly profit $\pi^m \ = \ 0.125$. Following Klein, I take as the competitive benchmark the average per-period profit of the most competitive Markov-perfect equilibrium (MPE) that corresponds to an Edgeworth price cycle, as described by Maskin and Tirole (1988), which yields approximately 0.061 for $k \ = \ 6$.

The alternating-move structure implies that, under the Markov assumption that strategies condition only on payoff-relevant variables, the relevant state for the firm about to move is simply the last price set by its competitor. This one-period memory structure is the natural starting point for the analysis and the focus of this paper.

3.2. Q-learning algorithm

Both companies are modelled as independent Q-learning agents (Watkins & Dayan, 1992). Q-learning is a model-free reinforcement learning algorithm that seeks to estimate the long-run value of taking a particular action in a particular state, without requiring knowledge of the transition dynamics or the competitor's strategy.

Each firm i maintains a Q-matrix $Q_i(s, a)$ of the size $|S| \ \times \ |P|$, where s in S is the state (the competitor's last price) and a in P is the chosen price. In the baseline symmetric configuration, both firms use one-period memory, so $|S| = |P| = 7$.

After each period in which firm i acts, it updates the Q-value corresponding to the state-action pair (s, a) that was just obtained, according to:

$$Q_i(s,a) \leftarrow (1 - \alpha)Q_i(s,a) + \alpha\,[\pi + \delta\,\pi' + \delta^2 \max_{a'} Q_i(s', a')], \quad (3)$$

where α in $(0,1)$ is the learning rate, δ in $[0,1)$ is the discount factor, π is the immediate profit from action a, π' is the profit earned in the subsequent 0period when the rival moves, and s' is the successor state. This two-step update reflects the sequential nature of the game: after firm i acts, firm j responds, and only then does firm i observe the full consequences of its action.

Action selection follows an ε-greedy policy. With probability ε_t, the firm explores the action space by picking a price uniformly at random from P. With probability $1 - \varepsilon_t$, it exploits its current knowledge by choosing the action that maximises $Q_i(s,\cdot)$. The exploration probability decays over time according to:

$$\varepsilon_t = (1 - \theta)^t, \quad (4)$$

where the decay parameter θ is calibrated so that $\varepsilon_0 = 1$ (full exploration at the start), $\varepsilon_{T/2} = 0.001$ (0.1% exploration at midpoint), and $\varepsilon_T = 10^{-6}$ (negligible exploration at the end).

The simulation parameters are set exactly as in Klein (2021): learning rate $\alpha = 0.3$, discount factor $\delta = 0.95$, and the total horizon of $T = 500{,}000$ periods. Both firms' Q-tables are initialised to zero. For each experimental condition, I run 1,000 independent simulations with distinct random seeds. The final $M = 50{,}000$ periods (10% of the horizon) are treated as the steady state and used for all the subsequent analysis.

3.3. Memory configurations

The defining feature of a Q-learning agent in this environment is its state representation – what information it conditions its pricing decision on. Under the Markov assumption that strategies depend only on payoff-relevant variables, the natural state for a firm about to move is the last price set by its competitor. This one-period memory structure is the starting point for most of the algorithmic collusion literature (Calvano et al., 2020; Klein, 2021). However, there is no reason that all firms in a market must use the same state representation. Differences in memory depth, i.e. how many periods of past prices an agent encodes in its state, constitute a natural and policy-relevant form of algorithmic heterogeneity. This paper examines three configurations that span a range from a complete symmetry to two qualitatively different types of memory asymmetry. The three memory configurations examined here are those introduced by Siegieda and Zawisza (2026), whose aggregate results provided the motivation for this cluster-based re-analysis. These are labelled:

- Scenario A: Symmetric one-period memory. Both firms condition their decisions on the competitor's most recent price. For firm i moving at time t, the state is $s_t = p_{j,t-1}$, where j denotes the competitor. The state space for each firm has the size $|S| = |P| = 7$. This is the canonical baseline studied by Klein (2021) and the configuration that yields the largest variety of pricing regimes;
- Scenario B: One-period vs. no memory. Firm 1 uses one-period memory ($s_t = p_{2,t-1}$), while Firm 2 has no memory at all – it observes a constant state regardless of the history of play ($|S_2| = 1$). A memoryless agent

cannot distinguish between different competitive situations; it effectively learns a single best price irrespective of what its competitor does. This configuration tests whether a sophisticated agent can exploit a naive one, or whether the naive agent nonetheless manages to sustain supra-competitive coordination;

- Scenario C: One-period vs. two-period memory. Firm 1 uses one-period memory ($s_t = p_{2,t-1}$), while Firm 2 uses two-period memory, conditioning on both the competitor's last price and its own last price: $s_t = (p_{2,t-1}, p_{1,t})$. The state space for Firm 2 expands to $|S_2| = |P|^2 = 49$, offering more strategic possibilities at the cost of substantially slower learning. This configuration examines whether deeper memory confers a strategic advantage and whether the expanded state space hampers convergence.

The three scenarios together allow a systematic assessment of how the memory heterogeneity – both its presence and its type – shapes the set of the resulting pricing regimes. For each scenario, I simulated 1,000 independent runs, each with a distinct random seed.

3.4. Performance metrics

To characterise the outcomes of each simulation run, I compute several metrics. The first is the average price, a direct measure of market outcomes and a proxy for consumer welfare. For each firm *i*:

$$\text{avg\_}p_i = (1/M) \cdot \textstyle\sum_{t=T-M+1}^{T} p_{i,t}. \tag{5}$$

I also report the market-wide average $\text{avg\_}p = (\text{avg\_}p_1 + \text{avg\_}p_2) / 2$.

Profitability measures the actual per-period profits obtained in the steady state:

$$\text{avg}_{\pi_i} = \left(\frac{1}{M}\right) \cdot \sum_{t=T-M+1}^{T} \pi_i\left(p_{i,t}, p_{j,t}\right). \tag{6}$$

Optimality captures how close each firm's learnt behaviour is to the best response given the competitor's strategy:

$$\Gamma_i = \frac{Q_i(p_{i,T}, s_T)}{\max_{p \in P} Q_i(p, s_T)}, \tag{7}$$

where Q_i is the firm's final Q-matrix and s_T is the state in the final period. The value of $\Gamma_i = 1$ (within a tolerance of 10^{-5}) indicates that the firm is playing the best response. A simulation run is classified as a Nash equilibrium if $\Gamma_1 = \Gamma_2 = 1$ simultaneously. Both optimality and the Nash criterion serve as supplementary diagnostics – they help interpret the economic nature of the regimes uncovered by clustering, but they are not the primary focus of the analysis. The optimality measure Γ_i and the Nash-equilibrium criterion are not used as input features for clustering. Clustering is based purely on the price features described in Section 3.5; the optimality measures appear only in the post-clustering analysis (Tables 1–3) to validate whether the identified fixed-price clusters correspond to the Nash equilibria, following Klein (2021).

These aggregate metrics are useful for establishing comparability with Klein (2021) and for providing a first summary of each scenario. However, as argued in Section 1, they collapse the rich heterogeneity of individual runs into single numbers. The cluster analysis described next is the core methodological contribution of this paper and is designed to overcome this limitation.

3.5. Cluster analysis methodology

The core methodological contribution of this paper is the application of cluster analysis to the steady-state pricing data. Rather than reducing the output of 1,000 simulation runs to a few means, I extract features that capture the structure of each run's final pricing behaviour and then use a clustering algorithm to partition the runs into groups that share similar characteristics.

The choice of features differs between the symmetric and asymmetric scenarios, reflecting a fundamental difference in the economic interpretation of firm labels.

The symmetric scenario (Scenario A: 1 vs. 1). In the symmetric configuration, both companies are ex ante identical – they use the same memory depth, the same learning parameters, and face the same environment. The assignment of the ‘Firm 1’ and ‘Firm 2’ labels is therefore arbitrary, and a given pricing pattern should be recognised as the same regime regardless of which firm happened to set the higher price. For example, a run where Firm 1 consistently charges 0.5 and Firm 2 charges 0.333 is economically distinct from the one where both charge 0.5, but it is the same regime regardless of which of the two companies set the high price. To achieve this label-invariance, I transform the raw price pair (p_1, p_2) into order statistics:

$$p_{min,t} = \min(p_{1,t}, p_{2,t}) \tag{8}$$

$$p_{max,t} = \max(p_{1,t}, p_{2,t}) \tag{9}$$

and then average these quantities over the steady-state periods to obtain two features per run:

$$\text{avg\_}p_{min} = (1/M) * \sum_{t=T-M+1}^{T} p_{min,t} \tag{10}$$

$$\text{avg\_}p_{max} = (1/M) * \sum_{t=T-M+1}^{T} p_{max,t} \,. \tag{11}$$

This two-dimensional representation summarises the steady-state behaviour of each run using two features. Runs where both firms set the same stable price cluster along the 45-degree line where avg_p_{min} ≈ avg_p_{max}. Runs with persistent price dispersion or large-amplitude cycles appear off the diagonal, with avg_p_{min} substantially below avg_p_{max}. The distance between these two values also proxies for the amplitude of price cycles when they are present.

Asymmetric scenarios (Scenarios B and C: 1 vs. 0 and 1 vs. 2). In the asymmetric configurations, the two companies differ in their memory depth and, as a consequence, in their strategic capabilities. The labels ‘Firm 1’ and ‘Firm 2’ are no longer arbitrary – they convey economically meaningful information about which agent has more memory and which has less. A run where the one-period-memory firm charges 0.5 and the zero-memory firm charges 0.333 is a different outcome from the one where the one-period firm charges 0.333 and the zero-memory firm charges 0.5, because the distribution of profits and the strategic roles are reversed. It is therefore appropriate to use the raw firm-specific mean prices directly as features:

$$\text{avg\_}p_1 = (1/M) * \sum_{t=T-M+1}^{T} p_{1,t} \tag{12}$$

$$\text{avg\_}p_2 = (1/M) * \sum_{t=T-M+1}^{T} p_{2,t} \,. \tag{13}$$

This preserves the identity of each firm and allows the clustering algorithm to identify regimes that differ according to which firm takes the lead in setting higher or lower prices. For Scenario B (1 vs. 0), Firm 1 has one-period memory and Firm 2 has none. For Scenario C (1 vs. 2), Firm 1 has one-period memory and Firm 2 has two-period memory.

For both the symmetric and the asymmetric scenarios, the clustering is based solely on the two price features described above, namely the order-statistic averages for Scenario A and the firm-specific averages for Scenarios B and C. This deliberately lean set of features captures the essential economic distinction between the fixed-price coordination (where the two features are equal or nearly equal) and cycling (where they differ), while keeping the dimensionality low enough for robust clustering. The more comprehensive statistics presented in Tables 1–3 are computed after the clusters are formed, and serve solely to characterise each regime and enable the reader to assess the economic interpretation of the groupings.

Standardisation and clustering algorithm. For all the three scenarios, the features are standardised to a zero mean and unit variance before clustering to ensure that each dimension contributes equally to the distance metric:

$$X_{scaled} = (X - \mu) / \sigma. \tag{14}$$

I use the K-means algorithm (MacQueen, 1967) with *n_init* = 50 random initialisations to mitigate the sensitivity to the centroid initialisation. The choice of *k* is guided by a combination of the elbow method (plotting within-cluster sum of squares against *k*) and the silhouette score (Rousseeuw, 1987), which measures how similar each point is to its own cluster compared to other clusters. These are standard tools in statistical learning for cluster validation (James et al., 2023, pp. 15–67). The final choice of *k* also takes into account a qualitative criterion: the resulting clusters must be economically interpretable. Across all the three scenarios, both the elbow and the silhouette diagnostics point toward *k* = 4 or *k* = 5 as the natural number of clusters. I select the value that yields the clearest separation into economically distinct regimes. The specific choice is explained and presented alongside the results for each scenario in the next section. The methodology is implemented in Python; analogous clustering procedures are widely available in other statistical computing environments, including R (Kamiński & Zawisza, 2012). Once each simulation run has been assigned to a cluster, I examine the within-cluster distributions of prices, profits, and dynamic patterns. For each cluster, I report the number of the assigned runs, the mean and standard deviation of prices, the mean profitability, and the share of periods in which both firms set equal prices.

4. Results

This section presents the results of the cluster analysis for each of the three memory configurations. For each scenario, I show the outcome of the K-means clustering and characterise the resulting groups in terms of average prices and profitability. Three qualitatively distinct types of regimes emerge: monopoly focal price (stable coordination on the joint-profit-maximising price of 0.5), collusive focal price (stable coordination on a supra-competitive price lower than 0.5), and incomplete Edgeworth cycles (persistent price undercutting with truncated, downward phases and supra-competitive averages).

Before examining the cluster compositions, let us note a general feature of the learning dynamics. The exploration probability decays to negligible levels by roughly the midpoint of the simulation horizon ($t \approx 250\,000$), so the steady-state behaviour is exploitation-driven. Which regime a run converges to depends on the stochastic early exploration phase; once the agents lock in, the pattern – whether a fixed price or a cycle – stays unchanged. In the cycling clusters, the cycles are regular and stable throughout the steady state. Thus, the identified clusters represent stable long-run outcomes rather than transient dynamics. Illustrative price and profit trajectories for Scenario A are presented in Figures A1 and A2 in the Appendix, where the cluster-level averages are plotted alongside the scenario-wide mean. These figures confirm that, as stated above, the clusters represent stable long-run outcomes rather than transient dynamics.

4.1. Scenario A: a symmetric one-period memory (1 vs. 1)

The symmetric one-period memory configuration is the canonical baseline studied by Klein (2021). The diagnostics – as described in section 3.5 – point to $k = 4$ clusters. As seen in Table 1, the clusters are clearly separated by price level and dispersion, and by profitability. Cluster 0 (n = 295) is the only group with substantial price dispersion and asymmetric profits, while in Clusters 1–3, near-perfect price equality and symmetric profits close to their theoretical grid-point levels could be observed. These differences reflect distinct pricing regimes that are characterised below.

Table 1. Scenario A: symmetric one-period memory – steady state (fixed or cyclical) prices and profits by cluster

Cluster	n	avg_p1	avg_p2	avg_p_min	avg_p_max	std_p_min	avg_pi	Interpretation
0	295	0.469	0.460	0.302	0.627	0.150	0.094	Incomplete collusive Edgeworth cycles
1	420	0.330	0.330	0.330	0.331	0.027	0.110	Partial collusive focal price (1/3)
2	252	0.500	0.500	0.499	0.501	0.015	0.125	Monopoly focal price (1/2)
3	33	0.667	0.667	0.667	0.667	0.000	0.111	Collusive focal price (2/3)

Source: author's calculation.

Cluster 0 (*n* = 295) is the only group with substantial price dispersion. The large discrepancy between avg_*p_min* (0.302) and avg_*p_max* (0.627), combined with the high standard deviation of *p_min* (0.150) indicates persistent price cycling with truncated downward phases. These cycles are incomplete in the sense that the undercutting phase stops well before prices reach the competitive floor, after which one firm resets the cycle with a large upward jump. The average profit of 0.094 is well above the competitive benchmark of 0.061, confirming that this regime, though dynamic, remains supra-competitive. This regime is labelled incomplete collusive Edgeworth cycles. Cluster 2 (n = 252) is the only cluster achieving the monopoly price (0.5) and maximum profits (0.125). This is the monopoly focal price regime. Clusters 1 (n = 420) and 3 (n = 33) are stable fixed-price regimes at non-monopoly grid points, specifically 0.333 (1/3) and 0.667 (2/3), with near-perfect price equality, very low standard deviations, and supra-competitive profits (0.110 and 0.111 respectively). They are labelled collusive focal price equilibria.

4.2. Scenario B: one-period memory vs. no memory (1 vs. 0)

In this asymmetric configuration, Firm 1 has one-period memory and Firm 2 has no memory. The diagnostics suggest $k = 4$ clusters. Table 2 presents the cluster characteristics. All four clusters show stable fixed-price outcomes with near-zero standard deviations; no Edgeworth cycles appear in this scenario. The most striking feature of Table 2 is the dominance of Cluster 0, which captures 74.2% of all runs and converges to the grid price 1/6 (0.167). This price yields the profit of 0.069, which is only marginally above the competitive benchmark of 0.061. The monopoly outcome (Cluster 3, with profit 0.125) emerges in only 8.4% of runs. Clusters 1 and 2 replicate the collusive focal prices at 1/3 and 2/3 observed in Scenario A, with profits of 0.111 in both cases. The absence of cycles is not a transient phenomenon: the inspection of the full price trajectories confirms that once the agents settle on a fixed price, they never depart from it during the steady state.

Table 2. Scenario B: asymmetric one-period memory vs no memory – steady state (fixed or cyclical) prices and profits by cluster

Cluster	n	avg_p1	avg_p2	avg_p_min	avg_p_max	std_p_min	avg_pi	Interpretation
0	742	0.167	0.167	0.167	0.167	0.000	0.069	Minimal supra-competitive focal price (1/6)
1	164	0.332	0.333	0.332	0.333	0.013	0.111	Collusive focal price (1/3)
2	10	0.667	0.667	0.667	0.667	0.000	0.111	Collusive focal price (2/3)
3	84	0.500	0.500	0.500	0.500	0.001	0.125	Monopoly focal price (1/2)

Source: author’s calculation.

The dominance of Cluster 0 is economically significant. Despite the presence of a memoryless agent, the market does not collapse into competitive pricing; instead, it stabilises at the lowest price on the grid that still yields supra-competitive profits. I label this regime the minimal supra-competitive focal price. The mechanism is collusive in the sense that both firms maintain a stable, equal price without undercutting, yet the resulting profit (0.069) is barely distinguishable from the competitive benchmark (0.061). This contrasts with other fixed-price clusters, where profits are substantially higher.

4.3. Scenario C: one-period memory vs. two-period memory (1 vs. 2)

In this configuration, Firm 1 has one-period memory and Firm 2 has two-period memory. The diagnostics support $k = 4$ clusters. Table 3 presents the cluster-wise summary. Cluster 0 (n = 512) is the largest and converges to an essentially monopolistic outcome (avg_*pi* = 0.124), although with a slightly higher variance than in Scenarios A and B, likely due to slower learning caused by Firm 2's larger state space. Cluster 1 (n = 406) reproduces the collusive focal price at 1/3 with symmetric profits (0.111). Cluster 3 (n = 32) is a collusive focal price near 2/3, although with moderately higher variance than its counterparts in Scenario A. Cluster 2 (n = 50) stands out. It consists of incomplete Edgeworth cycles with a striking profit asymmetry: the two-period memory agent (Firm 2) earns 0.132, while the one-period memory agent earns only 0.059. The average market profit of 0.096 remains well above the competitive benchmark. This is the only regime where the profit distribution is visibly unequal in favour of one company, and it illustrates how deeper memory can be leveraged to capture a disproportionate share of the supra-competitive surplus within a cycling regime. The cycles in this cluster are persistent throughout the entire steady-state window and do not show any systematic change in amplitude or frequency over time.

Table 3. Scenario C: aymmetric one-period memory vs two-period memory – steady state (fixed or cyclical) prices and profits by cluster

Cluster	n	avg_p1	avg_p2	avg_p_min	avg_p_max	std_p_min	avg_pi	Interpretation
0	512	0.496	0.501	0.494	0.503	0.044	0.124	Monopoly focal price (near 1/2)
1	406	0.333	0.333	0.333	0.334	0.016	0.111	Collusive focal price (1/3)
2	50	0.599	0.344	0.303	0.641	0.141	0.096*	Incomplete Edgeworth cycles
3	32	0.651	0.672	0.648	0.674	0.089	0.007	Collusive focal price (near 2/3)

Note. In Cluster 2, the two-period memory agent (Firm 2) earns an average profit of 0.132, while the one-period memory agent (Firm 1) earns 0.059. The table reports the average market profit (avg_*pi) of 0.096.*

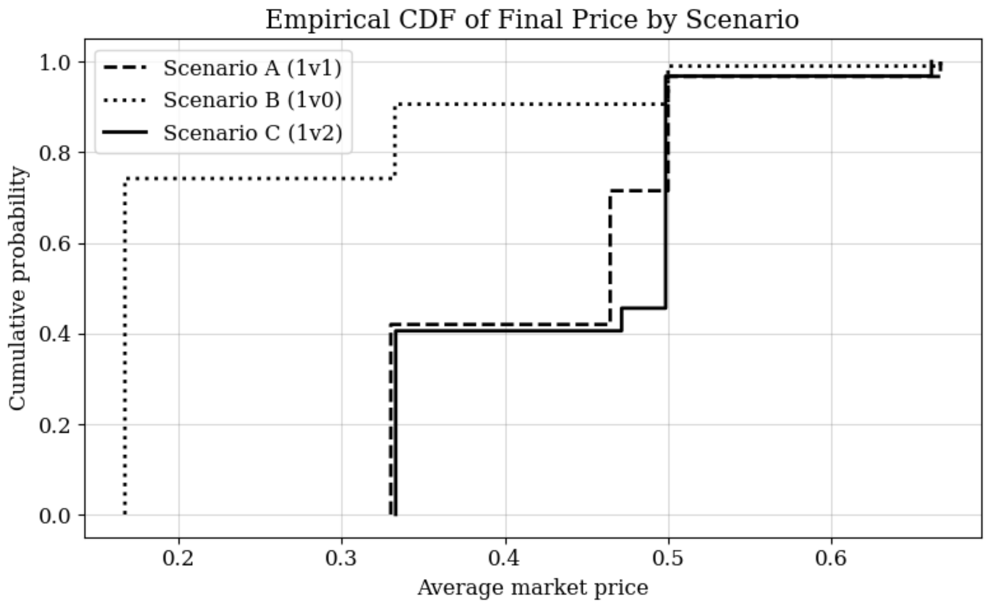
Source: author’s calculation.

4.4. Cross-Scenario Comparison

The tables in the preceding subsections characterise each cluster in isolation. To see how the overall landscape of pricing regimes changes with memory asymmetry, I turn to the full distribution of steady-state prices across all the 1,000 runs in each scenario. Figure 1 shows the empirical cumulative distribution function (CDF) of the average market price for each scenario. Figure 2 presents the corresponding side-by-side histograms.

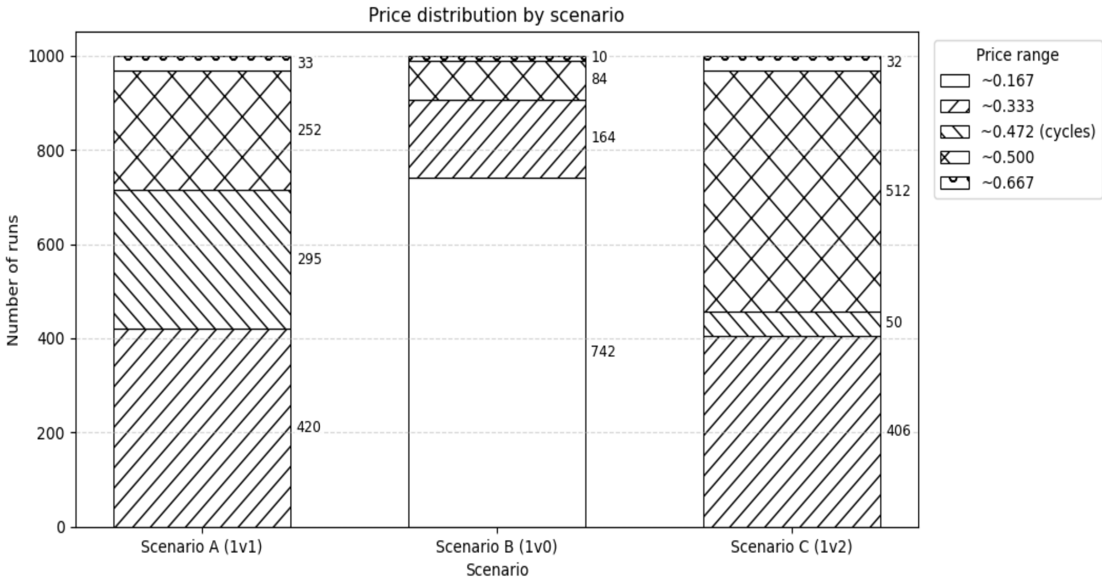
Under the specific parameterisation studied, the CDF of Scenario C lies almost everywhere below that of Scenario B, which in turn lies below that of Scenario A. This near-uniform ordering means that the price distribution in Scenario C first-order stochastically dominates (FOSD) that in Scenario B, and Scenario B stochastically dominates Scenario A. In other words, among the three configurations examined, prices are unambiguously higher in Scenario C than in Scenario B, and in Scenario B than in Scenario A – a ranking that a welfare-minded regulator would interpret as increasing consumer harm (Mas-Colell et al., 1995). The rightward shift of probability mass is equally clear in the histograms: moving from Scenario B (where the dominant outcome is the minimal supra-competitive price of 1/6) through Scenario A (where the supra-competitive price of 1/3 and incomplete cycles together account for over 70 % of runs) to Scenario C (where the monopoly price becomes the majority regime), the distribution visibly concentrates on higher prices.

Figure 1. Empirical CDF of steady-state market price by scenario



Source: author’s calculation.

Figure 2. Distribution of steady-state market prices by scenario



Source: author’s calculation.

Table 4 quantifies these shifts using the regime typology developed in Sections 4.1–4.3.

Table 4. Regime Distribution Across Scenarios (percentage of runs)

Regime	Scenario A (1v1)	Scenario B (1v0)	Scenario C (1v2)
Monopoly focal price (1/2)	25.2%	8.4%	51.2%
Collusive focal price (1/3)	42.0%	16.4%	40.6%
Collusive focal price (2/3)	3.3%	1.0%	3.2%
Collusive focal price (1/6)	–	74.2%	–
Incomplete Edgeworth cycles	29.5%	–	5.0%

Source: author’s calculation.

Three main findings are as follows. Firstly, the symmetric baseline (Scenario A) yields the greatest diversity of regimes: nearly 30% of runs fall into incomplete Edgeworth cycles, only a quarter achieve the monopoly price, and the remaining 45% are split between the supra-competitive prices of 1/3 and 2/3. The commonly reported aggregate statistics thus conceal a heterogeneous mix. Secondly, the memoryless agent in Scenario B eliminates cycles completely and anchors the market at the lowest supra-competitive price of 1/6 in almost three-quarters of the runs. Thirdly, introducing a two-period memory agent (Scenario C) more than doubles the probability of the monopoly outcome compared to the symmetric baseline (from 25.2% to 51.2%) and reduces the frequency of cycles from 29.5% to 5.0%. Memory asymmetry therefore does not merely preserve collusion – it systematically shifts the distribution of prices towards higher, more profitable outcomes, as the stochastic dominance ordering in Figure 1 and Figure 2 demonstrates.

5. Discussion

5.1. The Spectrum of Algorithmic Coordination

The cluster analysis shows that the algorithmic collusion is not a single outcome, but a spectrum of at least three pricing regimes. The first of them is the monopoly focal price, where both firms settle on the joint-profit-maximising price of 0.5. In the second, i.e. the collusive focal prices at non-monopoly grid points such as 1/3 or 2/3, the coordination is equally stable but profits are lower. The third one, incomplete Edgeworth cycles, consists in firms undercutting each other repeatedly, and yet the average price remaining close to the monopoly level. These regimes are all supra-competitive, but they differ fundamentally in their underlying dynamics and the consequences for the customers' wellbeing. In the symmetric baseline, only about a quarter of runs reach the monopoly outcome; the rest are split between other focal prices and cycles. The binary view of 'collusion or not' that dominates the literature thus misses the variety of ways in which algorithms can coordinate.

5.2. Memory asymmetry as a stabiliser of collusion

Across the three scenarios examined, memory asymmetry does not undermine algorithmic collusion; instead, it stabilises coordination. In the symmetric baseline, only about half of all the runs converge to a fixed-price equilibrium, while the rest follow cycles. When one firm has no memory (Scenario B), every run settles on a fixed price, eliminating cycles completely. When one firm has a two-period memory (Scenario C), fixed-price equilibria account for 95 % of runs, and the monopoly outcome alone captures over half of all runs – more than twice as many as in the symmetric case. Thus, within the parameter space studied, asymmetry pushes the market towards stationary collusion, and deeper memory on one side strongly favours the most profitable equilibrium.

5.3. Why are averages misleading

The standard practice of reporting only the average prices and profits masks the above-described entire structure. In a symmetric scenario, the global average profit of 0.109 is a blend of runs that yield 0.125, 0.110 and 0.094, each corresponding to a qualitatively different regime. A policymaker seeing only the aggregate number would not know if full monopoly collusion occurs in the minority of cases, or if cycling runs produce near-monopoly average prices through a competitive-looking process. Cluster analysis avoids this pitfall by letting the data reveal its own structure. It partitions the simulations into economically meaningful groups without imposing prior assumptions, using only the steady-state price patterns. This method provides a more comprehensive and actionable picture, and I suggest it become a standard robustness check in the future algorithmic-collusion studies.

5.4. Policy implications: detection and monitoring

For the competition authorities, the results imply that diagnosing algorithmic collusion requires looking beyond average price levels. Although authorities do not have access to thousands of repeated simulations, they can obtain historical time series of firm-level prices once an investigation begins – a single realisation of the price-generating process. By discretising the price grid and comparing the observed trajectory of the cluster prototypes identified here, they could assess whether the pattern resembles a stable focal-price regime, an Edgeworth cycle or a competitive benchmark. Different regimes produce different dynamic signatures, and cluster-based monitoring tools would thus allow more targeted interventions than simple price-level comparisons.

Two specific regimes warrant particular attention. One of them is incomplete Edgeworth cycles, which are deceptive in that prices change frequently, giving the impression of competition, yet the average price hovers near the monopoly level and the consumer harm is severe. The other is collusive focal prices at non-monopoly grid points, which are equally stable and harmful in relation to the competitive benchmark; disrupting one focal price may simply cause the algorithms to re-coordinate on another. Hence, the entire grid of possible focal prices – 1/6, 1/3, 1/2, 2/3 – should be of regulatory concern, not only the monopoly outcome.

5.5. Limitations and future research

This study is confined to a specific parameterisation: a coarse price grid with k = 6 intervals, fixed learning parameters, and a single horizon of 500,000 periods. Klein (2021) showed that finer grids (k = 24) amplify cycling dynamics. Whether the same regime typology and the stabilising effect of memory asymmetry hold under finer grids, alternative exploration schedules, or other learning algorithms (e.g., deep Q-networks) remains an open question. Additionally, the cluster analysis is descriptive; it does not explain why a given run converges to one regime rather than another. Understanding the determinants of regime selection, perhaps by analysing early-stage learning trajectories, is a promising direction for future work. Finally, a full consumer-wellbeing quantification of the different regimes, incorporating the deadweight loss from supra-competitive prices and the additional costs of price volatility would strengthen the policy conclusions.

6. Conclusions

This paper set out to look beyond the aggregate statistics that dominate the algorithmic collusion literature and to show the true heterogeneity of outcomes in a sequential pricing duopoly. Using cluster analysis on the steady-state pricing data of Q-learning agents, I found that algorithmic coordination is not a single monolithic phenomenon, but a spectrum of three distinct regimes: monopoly focal price, collusive focal price at other grid points, and incomplete Edgeworth cycles. Each regime is supra-competitive, yet each operates through a different mechanism and poses distinct challenges for detection and policy.

A key finding is that introducing memory asymmetry – a realistic form of between-firm heterogeneity – does not break collusion. On the contrary, it stabilises coordination, pushing the market towards fixed-price equilibria and, when one agent has deeper memory, strongly favouring the monopoly outcome. This result challenges the intuition that diverse algorithms automatically restore competition.

Methodologically, the paper demonstrates that cluster analysis is a simple yet powerful tool for revealing the structure hidden in simulation data. It avoids the information loss inherent in averaging and should be adopted as a standard complement to aggregate metrics in future research.

As regards competition policy, the findings suggest that market monitoring should move beyond price-level analysis. Tools that can classify pricing regimes based on their dynamic signatures could help authorities distinguish genuinely competitive markets from those in which algorithms autonomously coordinated. In an era of widespread algorithmic pricing, understanding what kind of collusion occurs in addition to whether it occurs at all will be essential.

References

Abada, I., Harrington, J. E., Lambin, X., & Meylahn, J. M. (2025). *Algorithmic Collusion: Where are we and where should we be going?* (Working paper). https://dx.doi.org/10.2139/ssrn.4891033.

Abada, I., & Lambin, X. (2020). *Artificial intelligence: Can seemingly collusive outcomes be avoided?* (Working paper). https://dx.doi.org/10.2139/ssrn.3559308.

Abada, I., Lambin, X., & Tchakarov, N. (2024). Collusion by mistake: Does algorithmic sophistication drive supra-competitive profits?. *European Journal of Operational Research*, *318*(3), 927–953. https://doi.org/10.1016/j.ejor.2024.06.006.

Bichler, M., Durmann, J., & Oberlechner, M. (2024). *Online optimization algorithms in repeated price competition: Equilibrium learning and algorithmic collusion*. https://doi.org/10.48550/arXiv.2412.15707.

den Boer, A. V., Meylahn, J. M., & Schinkel, M. P. (2022). *Artificial collusion: Examining supracompetitive pricing by Q-learning algorithms* (Amsterdam Law School Research Paper No. 25). https://dx.doi.org/10.2139/ssrn.4213600.

Byrne, D. P., & de Roos, N. (2019). Learning to coordinate: A study in retail gasoline. *The American Economic Review*, *109*(2), 591–619. https://doi.org/10.1257/aer.20170116.

Calvano, E., Calzolari, G., Denicolò, V., Harrington, J. E., & Pastorello, S. (2020). Protecting consumers from collusive prices due to AI. *Science*, *370*(6520), 1040–1042. https://doi.org/10.1126/science.abe3796.

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *The American Economic Review*, *110*(10), 3267–3297. https://doi.org/10.1257/aer.20190623.

Eckert, A. (2013). Empirical studies of gasoline retailing: A guide to the literature. *Journal of Economic Surveys*, *27*(1), 140–166. https://doi.org/10.1111/j.1467-6419.2011.00698.x.

Ezrachi, A., & Stucke, M. E. (2016). *Virtual Competition. The promise and perils of the algorithm-driven economy*. Harvard University Press.

Harrington, J. E. (2018). Developing competition law for collusion by autonomous artificial agents. *Journal of Competition Law & Economics*, *14*(3), 331–363. https://doi.org/10.1093/joclec/nhy016.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning with Applications in Python*. Springer. https://doi.org/10.1007/978-3-031-38747-0_2.

Kamiński, B., & Zawisza, M. (2012). *Receptury w R. Podręcznik dla ekonomistów*. Oficyna Wydawnicza SGH.

Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, *52*(3), 538–558. https://doi.org/10.1111/1756-2171.12383.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

Maskin, E., & Tirole, J. (1988). A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and Edgeworth cycles. *Econometrica*, *56*(3), 571–599. https://doi.org/10.2307/1911701.

Mehra, S. K. (2016). Antitrust and the Robo-Seller: Competition in the Time of Algorithms. *Minnesota Law Review*, *100*, 1323–1375.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Siegieda, K., & Zawisza, M. (2026). *Memory Arms Race in Q-learning: How Memory Asymmetry Shapes Algorithmic Collusion in Sequential Pricing* (Working paper). https://papers.ssrn.com/sol3/Delivery.cfm/6479701.pdf?abstractid=6479701&mirid=1.

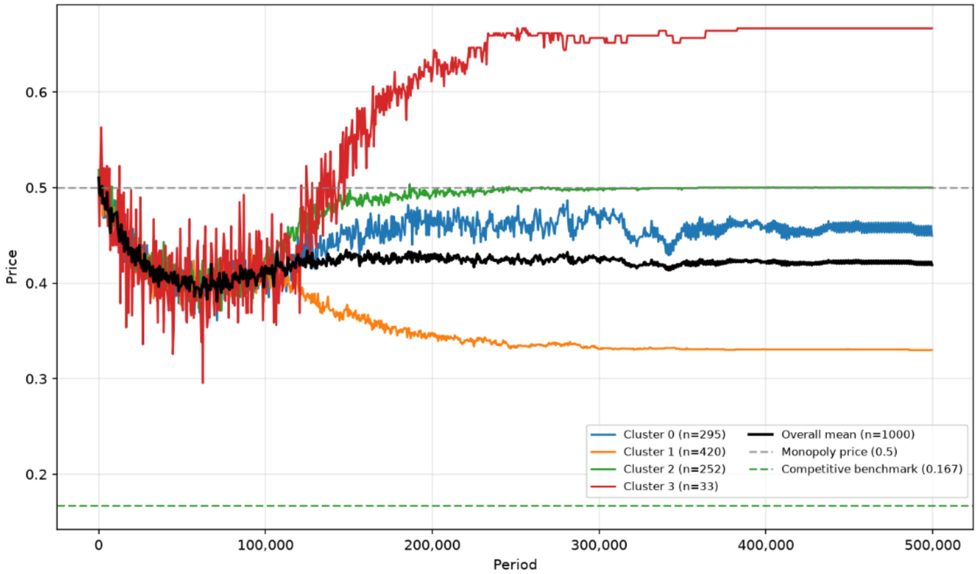
Waltman, L., & Kaymak, U. (2008). *Q*-learning agents in a Cournot oligopoly model. *Journal of Economic Dynamics & Control*, *32*(10), 3275–3293. https://doi.org/10.1016/j.jedc.2008.01.003.

Watkins, C. J. C. H., & Dayan, P. (1992). *Q*-learning. *Machine Learning*, *8*(3–4), 279–292. https://doi.org/10.1007/BF00992698.

Appendix

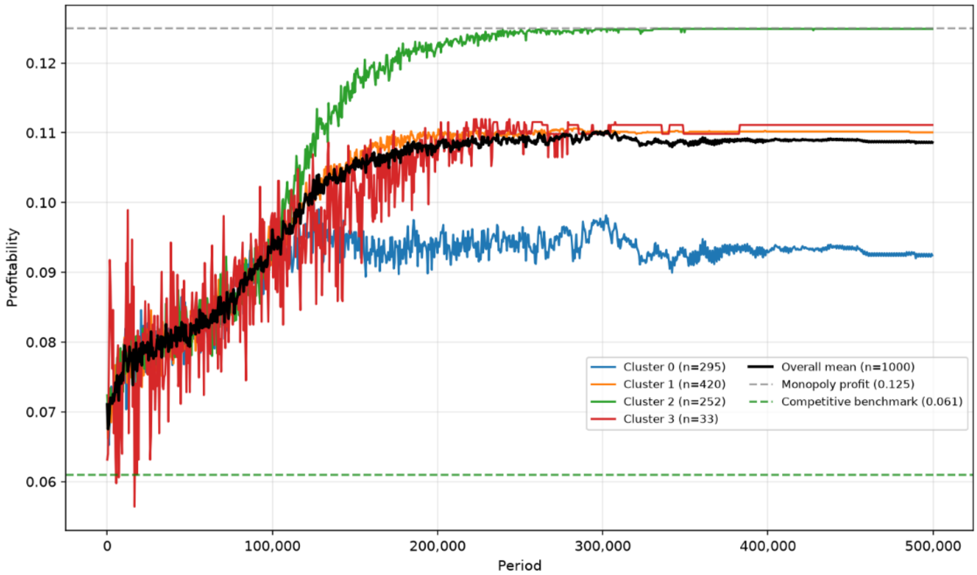
This appendix presents Figures A1 and A2, showing steady-state price and profit trajectories for the symmetric one-period memory configuration (Scenario A).

Figure A1. Price trajectories for Scenario A (1 vs. 1, symmetric one-period memory)



Source: author's calculation.

Figure A2. Profit trajectories for Scenario A (1 vs. 1, symmetric one-period memory)



Source: author's calculation.