Major in International Finance

RESEARCH PAPER

Academic Year 2025–2026

Submitted 31st May 2026

Application of Game Theory to Oil Producing Countries

Tacit Cooperation in the OPEC-US-Russia Oil Triopoly:
From Folk-Theorem Theory to Algorithmic Practice

Hamza LAHLALI & Yassine LAMRI[1]

Under the supervision of Prof. Jean-Pierre FRANÇOIS
President of the Jury: Prof. Paul VOSS

Keywords: *Algorithmic collusion, Multi-agent Q-learning, OPEC oil markets, Folk Theorem, Markov-perfect equilibrium, Green-Porter monitoring*

☒ PUBLIC REPORT ☐ CONFIDENTIAL REPORT

[1]Yassine LAMRI, yassine.lamri@hec.edu, 07 81 54 85 14.
Hamza LAHLALI, hamza.lahlali@hec.edu, 06 77 38 02 95.
Postal address: 15 place de l'Union Européenne, 91300 Massy, France.
Acknowledgement. The authors used AI assistants (Cursor with Claude) to help meet the HEC formatting requirements, notably Harvard-style citation normalisation, LaTeX layout, and bibliography consistency. All research content, analysis, and conclusions remain the authors' own.

Abstract

Calvano et al. (2020) showed independent Q-learners can tacitly collude; we identify the mechanism. On a pre-registered three-agent Q-learning experiment calibrated to OPEC-US-Russia oil triopoly under Green-Porter price-only monitoring, 50 seeds reach a cooperative basin on 94% of runs (mean collusion index 0.73). Three tests of Folk-Theorem retaliation : forced-deviation decomposition, σ-calibrated cycle detection, and Spearman reciprocity, reject punishment. A coarser-grid audit (2.7$\times$ Q-table coverage) rules out weak reciprocity hidden by sparse coverage. The cooperative attractor is a path-dependent Markov-perfect lock-in, not a Folk-Theorem equilibrium. Regulatory screens calibrated on reaction-function patterns will miss this class of algorithmic collusion.

Key words: *Algorithmic collusion · Multi-agent Q-learning · OPEC oil markets · Folk Theorem · Markov-perfect equilibrium · Green-Porter monitoring.*

Contents

1 Introduction

Crude oil remains the most systemically consequential commodity in the global economy. Despite accelerating investments in renewable capacity, petroleum and its derivatives still account for roughly 30 percent of primary energy consumption worldwide (International Energy Agency, 2025) and anchor the fiscal architecture of more than two dozen sovereign states, from the Gulf monarchies to Nigeria, from Russia to Venezuela. The transmission mechanism is direct: supply disruptions propagate into headline inflation, force central bank recalibrations, widen sovereign credit spreads, and spill over into equity volatility and foreign-exchange markets (Baumeister and Kilian, 2016; Barsky and Kilian, 2004). The strategic interactions among the world's three dominant production blocs: the Organization of the Petroleum Exporting Countries (OPEC), US shale producers, and the Russian Federation determine not merely the barrel price on trading screens, but the macroeconomic trajectory of importing and exporting nations alike. Understanding these interactions is a question at the nexus of macroeconomics, geopolitics, and financial stability.

The urgency of that question has sharpened dramatically since late February 2026. On 28 February, the United States and Israel launched strikes against Iran, precipitating the closure of the Strait of Hormuz, a 21-mile-wide chokepoint through which approximately 20 percent of globally traded petroleum transits under normal conditions (Congressional Research Service, 2026). The Iranian Revolutionary Guard declared the strait closed, attacked merchant vessels, and laid sea mines, reducing shipping traffic by more than 90 percent from pre-conflict levels. The EIA reports that crude and fuel flows through the strait declined by nearly 6 million barrels per day in the first quarter of the conflict (U.S. Energy Information Administration, 2026b). Brent crude, which traded near $65/bbl as recently as November 2025, surged past $116/bbl in early May 2026 before settling around $106–108/bbl amid uncertain ceasefire negotiations. Saudi Aramco's CEO warned on the company's first-quarter earnings call that the market is losing roughly 100 million barrels of supply per week and that normalisation could extend into 2027 if the strait remains blocked past mid-June (CNBC, 2026). The International Energy Agency (IEA) has cautioned that global observed inventories fell at a record pace of approximately 4 million barrels per day in March and April, and that the market could remain severely undersupplied until October even if the conflict ends next month (International Energy Agency, 2026).

Superimposed on this acute supply shock is a structural fracture in cartel governance. The United Arab Emirates (UAE) notified OPEC of its decision to withdraw from OPEC and OPEC+ on 28 April 2026, with the exit taking effect on 1 May 2026, after fifty-nine years of membership, the most significant departure in the organisation's history (Emirates News Agency, 2026). The UAE's decision followed years of friction over quota allocations that held its production roughly 30 percent below its 4.85 million barrel-per-day capacity (Wood Mackenzie, 2026). Abu Dhabi's exit follows Angola's departure in 2024 and Qatar's in 2019, establishing a pattern of defection that game theory predicts with precision: when the temptation payoff from unilateral expansion exceeds the discounted value of continued cooperation, the cartel unravels (Fudenberg and Maskin, 1986). On the opposing flank of supply, the US shale sector is approaching a structural plateau. The EIA's *Annual Energy Outlook 2026* projects that US crude production will decline from its 2025 record of 13.6 million barrels per day, as mature plays exhibit declining well productivity and capital discipline continues to prioritize shareholder returns over volume growth (U.S. Energy Information Administration, 2026a; Federal Reserve Bank of Dallas, 2026). Russia, meanwhile, occupies a paradoxical position: Western sanctions constrain its investment and logistics, yet the Hormuz-induced price spike elevates the very revenue those sanctions were designed to suppress (Financial Times, 2026).

This paper investigates these strategic interactions through a progressive computational framework that integrates classical game theory with modern reinforcement learning. We pursue four interlocking questions. First, what are the equilibrium outcomes of strategic quantity competition among asymmetric oil producers under alternative market structures? Second, can tacit collusion be sustained, and under what conditions of patience, monitoring, and demand uncertainty does it survive or collapse? Third, and this is the central empirical contribution, does tacit collusion emerge spontaneously when autonomous agents learn production strategies through trial and error, without programmed coordination, without knowledge of the demand model, and observing only the market price? Fourth, which production strategies survive evolutionary selection in a population of heterogeneous producers?

We construct an original Python codebase calibrated to a stylized oil-market model with linear inverse demand $P(Q) = 140 - Q$ and asymmetric marginal costs reflecting the structural heterogeneity among US shale ($45/bbl, consistent with the Dallas Fed Energy Survey Q1 2026, Federal Reserve Bank of Dallas, 2026); OPEC ($20/bbl, reflecting Gulf weighted-average lifting costs per BP Statistical Review, BP, 2022); and Russia ($35/bbl, per IMF Article IV, IMF, 2021).

The analytical pipeline escalates through seven conceptual layers, detailed as a sixteen-step reading guide in Appendix B: (i) static Cournot competition establishing the Nash benchmark; (ii) Stackelberg leadership quantifying OPEC's first-mover advantage; (iii) repeated games and the Folk Theorem formalizing cartel sustainability; (iv) Shapley values addressing coalition stability; (v) Green and Porter (1984) imperfect monitoring under stochastic demand; (vi) single-agent and multi-agent Q-learning under price-only observation; and (vii) evolutionary dynamics via replicator equations. Each layer addresses a limitation of the preceding one, and all share a common quantitative metric, which is the collusion index, defined as

$$\mathrm{CI} = \frac{P^{\mathrm{Converged}} - P^{\mathrm{Nash}}}{P^{\mathrm{Cartel}} - P^{\mathrm{Nash}}}, \tag{1}$$

that enables direct cross-framework comparison.

The contribution is threefold. Methodologically, we are, to our knowledge, the first to embed multi-agent Q-learning within the Green-Porter imperfect monitoring framework and apply it to a calibrated energy-market oligopoly, a design that extends the Bertrand-focused experiments of Calvano et al. (2020) to Cournot quantity competition with asymmetric costs. Empirically, we provide computational evidence that tacit collusion is an emergent property of autonomous production-setting agents, with the collusion index varying systematically with the discount factor, exploration rate, and learning rate in directions consistent with the Folk Theorem's predictions. For policy, these findings speak directly to the accelerating regulatory debate around algorithmic collusion: California's AB 325, effective January 2026, targets algorithms that facilitate price alignment (DLA Piper, 2025); the European Commission has confirmed multiple active investigations into algorithmic pricing (Baxter, 2025; Loyens & Loeff, 2026); and the US Department of Justice has warned of the risk of "fully automated cartels" (Kavanagh and Magloughlin, 2025).

2 Foundations: The Static Market Environment

This section and Section 3 establish the theoretical and quantitative anchor of the thesis: a calibrated oligopoly model and the classical equilibrium concepts against which the original contribution is measured. That contribution itself, the multi-agent learning experiment and its identification of a path-dependent collusive attractor, is developed in Section 4. A reader already familiar with static and repeated oligopoly theory may read the present two sections as calibration and proceed directly to the learning experiments; a reader new to the field will find each concept defined before it is used.

2.1 Model overview and calibration

We model the global crude oil market as a strategic oligopoly. The word "oligopoly" denotes an industry in which a small number of producers each control enough output to influence the market price (Bresnahan, 1989); the implication is that production decisions are inherently strategic, because every producer must anticipate how its choice affects the price that its rivals also receive. The three blocs we model are the OPEC core (treated as a single Gulf-led player), the Russian Federation, and the United States shale sector. Rather than relying on a single game-theoretic frame, we construct a progressive computational pipeline in which each layer relaxes one assumption of the preceding one. The simplest model, static Cournot competition, provides the non-cooperative benchmark. Every subsequent extension adds one dimension of realism: timing (Stackelberg), commitment (capacity), welfare (deadweight loss and carbon taxation), strategic variable (Bertrand), repeated interaction (Folk Theorem), mediated recommendations (correlated equilibrium), bargaining (Shapley values), uncertainty (Green and Porter, 1984), and population-level selection (evolutionary dynamics). This architecture is consistent with the escalation logic recommended by Tirole (1988, p. ??) and parallels the framework adopted by Alcelay et al. (2017), to whose three-player Cournot setting our calibration is closest.

The demand side of the model is given by a linear inverse demand, introduced for analytical tractability and because it is the workhorse specification of the applied oligopoly literature since Cournot (1838, p. ??). The function takes the form

$$P(Q) \;=\; a - bQ, \tag{2}$$

where P denotes the market-clearing price in dollars per barrel, Q is total output in millions of barrels per day (mbd), a is the demand intercept (the choke price at which demand falls to zero), and b is the demand slope (the dollars-per-barrel price response to a one-unit change in total supply). Each producer $i \in \{\text{US}, \text{OPEC}, \text{RUS}\}$ chooses an output q_i to maximise its period profit

$$\pi_i \;=\; (P - c_i)\,q_i \;=\; (a - bQ - c_i)\,q_i, \tag{3}$$

where c_i is its constant marginal cost. The first term in (3) is the per-barrel margin and the second is the volume; the product is the producer's profit-relevant cash flow per period. We calibrate $a = 140$, $b = 1.0$, $c_{\text{US}} = 45$, $c_{\text{OPEC}} = 20$, $c_{\text{RUS}} = 35$, and discount factor $\delta = 0.95$. The calibration is reported in Appendix A at the back of the document.

The demand intercept $a = 140$ corresponds to the upper end of the long-run price scenarios published by the International Energy Agency (International Energy Agency, 2019); reading it as a choke price is consistent with high-stress scenarios used by central banks for inflation modelling. The slope $b = 1.0$ is the standard normalisation in the applied oligopoly literature (Tirole, 1988, p. ??); at the baseline triopoly equilibrium it implies an own-price demand elasticity of approximately -0.75, broadly consistent with the medium-run elasticities of -0.4 to -1.0 reported by the U.S. Energy Information Administration in its Short-Term Energy Outlook. The quarterly discount factor $\delta = 0.95$ implies an annualized time preference of roughly 19 percent; we read this as a reduced-form proxy for the political, fiscal, and sanctions-related impatience that depresses the effective planning horizon of producing states below the risk-free rate.

We now explain why each marginal cost was set at its calibrated level. For OPEC we set $c_{\text{OPEC}} = 20$. The Gulf core (Saudi Arabia, Kuwait, the United Arab Emirates, Iraq) operates onshore conventional fields with some of the lowest lifting costs in the world: Saudi Aramco's reported lifting cost has been in the \$3 to \$5 per barrel range for over a decade, with broader Gulf averages in the \$5 to \$10 range (BP, 2022; U.S. Energy Information Administration, 2024). Layering on royalties, infrastructure depreciation, and the implicit fiscal contribution that funds the state operating model brings the effective short-run marginal cost to roughly \$15 to \$20. We take the upper bound of that band to reflect the average OPEC core producer rather than only Saudi Aramco, and to avoid overstating the cost advantage in the cartel analysis.

For Russia we set $c_{\text{RUS}} = 35$. Western Siberian conventional fields have lifting costs of approximately \$15 to \$25 per barrel (Henderson, 2015, 2011, p. ??)(IMF, 2021). To this we add the Urals export transport differential (pipeline tariffs to Baltic and Black Sea ports, plus the structural discount of Urals to Brent), which contributes a further \$8 to \$12, and the incremental post-2022 sanctions costs (shadow fleet, intermediated trade, insurance) of roughly \$5 (Simola and Solanko, 2017; U.S. Energy Information Administration, 2025). The sum sits around \$35, consistent with the IMF Russia Article IV estimates (IMF, 2021) and with the average Urals realization observed in 2023 to 2025. Our model holds this number constant; in reality, the sanctions overhead has varied with the G7 price cap regime, but the static specification cannot accommodate that dynamic.

For the United States we set $c_{\text{US}} = 45$. US shale producers have well-level break-evens that vary widely by play and well quality, from below \$40 in the core Permian to above \$70 in marginal Bakken locations (Federal Reserve Bank of Dallas, 2026; Smith, 2016). We choose \$45 as the average well-level marginal cost of new wells in major plays, which is the relevant number for short-run best-response decisions. Firm-level full-cycle break-evens, which include corporate overhead, decline-replacement capex, and required returns to equity, were reported at \$61 to \$62 per barrel in the Q1 2026 Dallas Fed survey (Federal Reserve Bank of Dallas, 2026), but those lie outside our short-run analysis. The post-2020 capital-discipline regime documented by Krane and Agerton (2015) and reaffirmed by the same survey implies that US shale's effective supply curve has flattened, which our static model captures by treating c_{US} as constant.

Three limitations of the calibration deserve acknowledgment. First, linear demand abstracts from the documented convexity of crude oil prices when spare capacity is tight (Till, 2016). Second, each "player" is in reality a coalition: the OPEC core averages over twelve heterogeneous members, US shale is a fragmented private-sector industry, and Russia's effective cost varies by destination market (Henderson, 2015, p. ??)(U.S. Energy Information Administration, 2025). Third, marginal costs are held constant, while in reality they vary with cumulative extraction, technological learning, and logistics. These simplifications follow standard practice in applied oligopoly modelling (Salant, 1976; Huppmann and Holz, 2012; Alcelay et al., 2017) and are necessary to maintain comparability across the frameworks we implement. The full step-by-step "escalation table" is provided in Appendix B; each row identifies one framework, the question it addresses, and the assumption it relaxes relative to its predecessor.

2.2 Static Cournot equilibrium

We begin with the simultaneous quantity game, named after Cournot (1838, p. ??), the first author to formalize oligopolistic interaction. In a Cournot game, each producer chooses an output level without observing the choices of its rivals, and the market-clearing price is determined by the sum of all outputs through the inverse-demand function. The Nash equilibrium of a Cournot game is the mutually best-responding profile: each player's quantity is optimal given the quantities chosen by everyone else, so no player can unilaterally improve its profit by deviating. The Cournot frame is the natural benchmark for oil markets because oil producers actually compete on quantities, not on posted prices: OPEC announces production targets, not barrel prices, and the spot price is set in real time by exchange-based trading on the resulting aggregate supply. We solve the model for both the duopoly (OPEC and US) and the triopoly (OPEC, US, Russia), following the three-player structure of Alcelay et al. (2017).

In the duopoly we obtain total output $Q = 71.67$ mbd, equilibrium price $P = \$68.33/\text{bbl}$, and profits of 2,336.11 for OPEC and 544.44 for US shale.[1] Adding Russia as a third strategic player shifts the equilibrium to $Q = 80.00$ mbd, $P = \$60.00/\text{bbl}$, with profits of 1,600.00 for OPEC, 225.00 for US, and 625.00 for Russia, consumer surplus of 3,200.00, and total welfare of 5,650.00 (see Figures 1, 2 and 3). Russia's entry adds 8.33 mbd of output, lowers the price by \$8.33, erodes OPEC profit by 736.11 units (a 31.5 percent decline), and raises welfare by 201.39 units. The mechanism is exactly the one identified by Salant (1976) and Novshek (1985): a new entrant with intermediate cost simultaneously expands aggregate supply and compresses the incumbent's margin. Schmalensee (1987) shows further that cost asymmetry shifts the unconstrained collusive optimum toward the low-cost producer, which is why OPEC retains the largest share even after Russia's entry. This quantitative finding gives a clean economic reading of the 2016 OPEC+ formation: Russia's standalone presence as a Cournot rival destroyed roughly 736 profit units for OPEC, which is precisely the surplus that the OPEC+ coordination mechanism subsequently attempted to recover.

Comparative statics on the demand intercept confirm the linearity of the triopoly solution. Sweeping a from 112 to 168, equilibrium price moves from \$53 to \$67 (slope $dP/da = 1/4$) and total output from 59 to 101 mbd (slope $dQ/da = 3/4$), exactly reproducing the closed-form predictions of the asymmetric Cournot model with $N = 3$, namely $dP/da = 1/(N+1)$ and $dQ/da = N/(N+1)$ (see Figure 4). This perfect linearity is convenient for diagnostics, because any deviation in the simulator output would signal a coding or calibration error; it is also the chief reason we are confident in the comparative statics that follow. The linear response is, however, a property of the demand specification and should not be over-interpreted. The empirical price-quantity relationship in physical crude markets is markedly convex once OPEC spare capacity falls below roughly 2 mbd: Till (2016) documents exactly this regime, in which a small additional supply deficit produces a disproportionately large price spike. Our linear model captures the first-order direction of these effects but not the curvature, and we therefore read the slopes

[1] Profit, welfare, and surplus figures throughout are deterministic outputs of the calibrated simulator, reported to two decimals for reproducibility. The precision reflects the model's internal arithmetic on stylised parameters, not a claim of empirical accuracy at that resolution.

above as lower bounds on the price sensitivity that would obtain in a tight-capacity scenario such as the February 2026 Hormuz closure.

2.3 Market power and first-mover advantage

To quantify the degree of market concentration we use two standard metrics of industrial organisation, both of which we introduce before reporting our results. The Lerner index $L_i = (P - c_i)/P$, due to Lerner (1934), measures the gap between price and the marginal cost of producer i, expressed as a fraction of the price. It equals zero under perfect competition (price equals marginal cost) and approaches one as the producer's pricing power rises (Bresnahan, 1989). The Herfindahl–Hirschman Index HHI $= \sum_i s_i^2 \times 10{,}000$, where s_i is the output share of producer i, measures aggregate market concentration. It ranges from close to zero in an atomistic market to 10,000 in a pure monopoly. By the threshold used in the U.S. Horizontal Merger Guidelines, an HHI above 2,500 places a market in the "highly concentrated" category. Both indices were popularised in the empirical industrial-organisation literature surveyed in Bresnahan (1989).

Our simulations yield an HHI above 2,500 in every market structure we test, which formally establishes that the oil market is, by U.S. Department of Justice criteria, highly concentrated; this is not a trivial finding because the result holds even in the triopoly, when total industry output is at its largest. The OPEC Lerner index ranges from 0.57 (when OPEC is the Stackelberg leader, defined below) to 0.71 (in the duopoly), and is the highest among the three players in every configuration (see Figure 5). The result is consistent with the calibrated OPEC market-power trajectories of Huppmann and Holz (2012) and Huppmann (2013), who estimate a conjectural-variation parameter θ for OPEC in the range 0.5 to 0.7 throughout 2003 to 2011. In the standard Bresnahan (1989) identity, the industry-weighted Lerner index satisfies $L = \theta/\varepsilon$; at our equilibrium elasticity $|\varepsilon| \approx 0.75$, that θ-range translates to $L \in [0.67, 0.93]$. Our triopoly-Nash OPEC Lerner of $(60-20)/60 = 0.67$ sits exactly at the lower edge of this empirical band, which is the expected position once US shale entry is taken into account: Huppmann's Figure 3 shows observed prices tracking the calibrated oligopoly line below the pure-Cournot prediction, indicating that OPEC behaves slightly less aggressively than pure Cournot would imply. The qualitative finding that OPEC carries the largest Lerner in every market structure is therefore robust across two independent methodologies (ours, which is calibrated; theirs, which is econometrically estimated) and is driven by the underlying cost asymmetry rather than by any specific assumption about strategic behaviour.

A Lerner index, however, only measures market power at the simultaneous-move equilibrium. It is silent on whether the structural cost advantage we have just documented can be amplified by strategic timing. To answer that question, we now relax the simultaneity assumption that Cournot imposes. The von Stackelberg (1934, p. ??) model (on whose original formulation see Baumgärtner, 2001) allows one player, which is the leader, to choose its output first; the followers then observe the leader's quantity before choosing their own. The leader can therefore optimise along the followers' best-response functions rather than at the symmetric intersection point, which generates a first-mover advantage (Dixit, 1980)(Tirole, 1988, p. ??). The size of that advantage is an empirical question: it depends on both the cost gap and the slope of the followers' reaction functions, and we now quantify it for each candidate leader.

When OPEC acts as Stackelberg leader, it expands output to 93.33 mbd (the largest single-player output in any structure), drives the price down to \$46.67/bbl, and earns 2,133.33 in profit, a first-mover advantage of 533.33 over the simultaneous Nash benchmark. When US shale or Russia leads, the premium collapses to 75.00 and 208.33 respectively (see Figure 6). The asymmetry has a clean economic reading:

the leader's benefit from expanding output rises with its cost advantage, because the residual margin $P - c$ stays positive at quantities that drive higher-cost rivals toward zero. Behar and Ritz (2017) calibrate this logic for the 2014 to 2016 episode and identify a market-share threshold above which a low-cost swing producer prefers to flood the market rather than to accommodate. Our Stackelberg-OPEC scenario, in which US profit collapses from 225.00 to 1.67, is the quantitative counterpart of that mechanism and reads naturally as the regime that Saudi Arabia approximated during 2015 and again, more selectively, in March to April 2020. The combination of structural (cost) and strategic (timing) advantage explains why no other producer has been able to displace OPEC as the implicit residual setter of the OPEC+ quota envelope, despite the substantial growth of US shale through 2024.

2.4 Capacity constraints

The preceding analysis assumes that each producer can supply any quantity its best-response function calls for. In reality, each producer faces a physical capacity ceiling: the upper bound on output that its existing wells, pipelines, terminals, and processing infrastructure can deliver in a given period. We calibrate $\text{cap}_{\text{US}} = 30$, $\text{cap}_{\text{OPEC}} = 40$, and $\text{cap}_{\text{RUS}} = 35$ mbd, levels consistent with current installed capacity plus a credible six-month ramp; Appendix A reports the empirical sources underlying each value. A constraint is said to "bind" at a given equilibrium when the player's optimal unconstrained output equals its cap; when this happens, the cap restricts the strategic options of the binding player. In a dynamic context, capacity is also a commitment device in the sense of Dixit (1980): unused capacity makes future flooding credible and therefore alters rivals' incentives even when it is never deployed.

At the static Nash, OPEC's optimal unconstrained output of 40.00 mbd exactly hits its capacity ceiling, so OPEC is the binding player; under cartel quotas, no constraint binds, because the cooperative agreement restricts output below every physical ceiling (see Figures 7 and 8). Two implications follow. First, switching constraints on raises the critical discount factor δ^* for every player, because binding caps weaken the punishment phase by limiting the volume the cheated party can flood with (see Figure 9). Second, expanding OPEC's capacity ceiling has the opposite effect: sweeping cap_{OPEC} from 25 to 60 mbd lowers δ^* from 0.850 to 0.529 (see Figure 10). Both statements are consistent because they refer to different comparisons; we record this distinction explicitly because it is easy to confuse.

The policy reading is direct. Saudi Arabia's persistent maintenance of 1.5 to 2.0 mbd of idle capacity, often criticized domestically as wasteful, is rational because unused capacity is precisely what makes the threat of retaliatory flooding credible and therefore lowers the cooperation threshold for every member. Till (2016) in Appendix D provides the empirical counterpart, showing that prices become convex in the supply deficit once OPEC spare capacity falls below a critical threshold. The 2026 closure of the Strait of Hormuz on February 28 (Reuters, 2026), which sent Brent above \$100/bbl and prompted a record inventory drawdown (International Energy Agency, 2026), is a contemporary illustration of that convexity and of the value of spare capacity as a stabilisation buffer.

2.5 Welfare and deadweight loss

We now evaluate the social impact of each market structure. To do so we use the standard welfare decomposition of microeconomic theory. Consumer surplus (CS) measures the area between the demand curve and the price line, capturing the aggregate dollar value that consumers receive above what they pay. Producer surplus (PS) measures the area between the price line and the supply curve, capturing the aggregate margin that producers earn above their marginal cost. Total welfare $W = \mathrm{CS} + \mathrm{PS}$ is the social value generated by the market. Deadweight loss (DWL) is the welfare lost relative to the perfectly competitive benchmark, in which price equals marginal cost; algebraically, $\mathrm{DWL} = W^{\text{competitive}} - W^{\text{actual}}$. This decomposition follows Mas-Colell et al. (1995, p. ??) and Tirole (1988, p. ??).

Under our calibration the competitive benchmark yields $W = 7{,}200$ with zero deadweight loss. At the static Nash, $\mathrm{DWL} = 800$ and $W = 5{,}650$. Stackelberg-OPEC delivers $\mathrm{DWL} = 355.56$ and $W = 6{,}627.78$, while the cartel inflicts $\mathrm{DWL} = 1{,}800$ and $W = 4{,}837.50$, a welfare gap of 2,362.50 units relative to competition (see Figures 11 and 12). The deadweight-loss ranking, from least to most distortionary, is competitive (0), Stackelberg-OPEC (355.6), Stackelberg-Russia (501.4), Stackelberg-US (612.5), Nash (800), and cartel (1,800). That Stackelberg leadership by the lowest-cost player generates the highest welfare among realistic oligopoly structures is counter-intuitive but economically clean: OPEC's cost advantage means that production is allocated to the most efficient producer, so the productive inefficiency of the symmetric Nash (in which higher-cost firms produce too much) is reduced. The cartel is the worst case because it combines allocative distortion (price above marginal cost) with quota-induced productive distortion.

A particularly instructive extension is the interaction with a Pigouvian per-barrel carbon tax τ. The Pigouvian framework (Pigou, 1920, p. ??) prescribes a tax equal to the marginal external damage caused by an activity, internalizing the externality at the producer level. We sweep $\tau \in \{0, 10, 20, 30, 40, 50\}$. The collusion premium, defined as the ratio of cartel to Nash joint profit minus one, shrinks from 23.98 percent at $\tau = 0$ to 10.92 percent at $\tau = 50$ (see Figures 13 and 14). At high tax levels the cartel's output restriction partially mimics the socially desired reduction in fossil-fuel consumption, which we label the "green-cartel" effect: collusion and climate policy point in the same direction. The result should not be read as an endorsement of cartelization. A Pigouvian carbon tax is strictly more efficient, because it achieves the same quantity reduction while generating fiscal revenue and recycling part of the loss into productive uses; cartels, in contrast, redistribute roughly 1,400 units of consumer surplus to producers while destroying 1,000 units in deadweight loss.

2.6 Bertrand price competition: robustness check

So far, we have assumed Cournot competition, in which producers choose quantities. The natural alternative is Bertrand competition, in which they choose prices. In its pure form, the Bertrand (1883) game with homogeneous products and constant marginal cost produces the so-called Bertrand paradox: even with only two producers, undercutting drives the equilibrium price down to marginal cost. Pure-Bertrand is rarely a realistic description of upstream oil markets because production is capacity-constrained and crude grades are differentiated. We therefore implement a differentiated-Bertrand specification in which each producer faces a downward-sloping per-player demand and the substitution parameter $\sigma \in [0, 1]$ controls the degree of cross-grade interchangeability: $\sigma = 0$ corresponds to independent monopolies (no substitution), $\sigma = 1$ to perfect substitutes (the homogeneous Bertrand case). This setting follows the Maskin and Tirole (1988) dynamic Bertrand framework and the standard differentiated-oligopoly literature.

At the baseline $\sigma = 0.6$ (Appendix A for reasoning behind this calibration), which we read as the medium differentiation across WTI, Brent, and Urals grades, the Bertrand-Nash average price is \$35.90/bbl, far below the Cournot-Nash price of \$60.00/bbl (see Figures 15 and 16). As Bertrand (1883) predicted, price competition is more aggressive than quantity competition, and the equilibrium converges toward marginal-cost pricing as σ approaches one (see Figure 17). Two qualifications must be flagged. The Bertrand and Cournot specifications use different demand structures (per-player versus aggregate), so absolute levels are not directly comparable; only the qualitative ordering matters. We also document a corner-solution artifact in the cooperative Bertrand optimisation at very low σ (0.1 to 0.2), which we treat as a numerical limitation rather than an economic finding.

Two qualitative results survive intact. The producer ranking OPEC $>$ Russia $>$ US is preserved across competition modes, with OPEC's profit rising in σ because Bertrand convergence favours the lowest-cost producer. The cooperation gap also persists: the cooperative Bertrand average price exceeds the Bertrand-Nash by \$15.35, and joint profit is roughly double (see Figure 18). This robustness, in the spirit of Maskin and Tirole (1988), is reassuring: the central object of the next section, the incentive to collude and the conditions under which collusion can be sustained, does not depend on whether producers are modelled as choosing quantities or prices. We retain Cournot for the remainder of the analysis because it matches OPEC's institutional reality, in which quotas, not posted prices, are the negotiated instrument.

3 Strategic Dynamics: The Theoretical Anchor

3.1 From static to dynamic: repeated game dynamics

A static (one-shot) game ends after a single round of choices. A repeated game embeds the same stage game in time, so the same producers choose outputs in every period and discount future profits at a common factor $\delta \in (0,1)$. The discount factor captures how much present weight players place on future payoffs: a high δ corresponds to patient producers, a low δ to impatient ones. The reason this matters is that repetition introduces a credible mechanism of reward and punishment, which transforms the set of sustainable outcomes relative to the static game (Friedman, 1971; Fudenberg and Maskin, 1986). Before we evaluate the cooperative equilibrium, we first study the simplest possible dynamic adjustment rule: myopic best-response with inertia. "Myopic" means each producer optimises against last period's actions only, without forward-looking calculation; "inertia" $\lambda \in (0,1)$ means producers adjust gradually rather than jumping to the new best response, which is a realistic constraint for an industry in which drilling, leasing, and contracting take time.

Formally, each producer updates its quantity according to

$$q_i(t+1) \;=\; (1-\lambda)\,q_i(t) + \lambda\,\mathrm{BR}_i\big(q_{-i}(t)\big), \tag{4}$$

where $\mathrm{BR}_i(\cdot)$ is the static Cournot best response of player i to the output vector $q_{-i}(t)$ of all other players. Starting from an out-of-equilibrium initial condition with $P(0) = 108$, the price converges monotonically to the static Nash level $P = 60$ within approximately eight periods at $\lambda = 0.2$; the standard deviation of the simulated price path is 8.14 (see Figure 19). The dynamics are governed by the dominant eigenvalue $\mu = 1-2\lambda$ of the linearized system around equilibrium, which is why convergence is monotone for $\lambda \in (0,0.5)$, critical at $\lambda^* = 0.5$, and oscillatory for $\lambda \in (0.5,1)$ (see Figure 20). The economic content is that even purely myopic adjustment, with no forward-looking calculation, takes the market to the static Nash equilibrium, which then becomes the relevant benchmark against which cooperative deviations and punishments are measured.

Conditional on a large negative shock, defined here as any quarter in which the simulated price falls more than \$21/bbl below the cartel price of \$80, the mean recovery path traces out an exponential approach back to the cooperative level, with a half-life of 5.2 quarters and asymptotic mean of \$78.7, statistically indistinguishable from the cartel target (see Figure 21). The 80 percent confidence band around this mean is wide in the first three periods (where idiosyncratic shock magnitudes still dominate) and narrows after period 10 once the punishment phase has cleared. The empirical analogue is the 2020 OPEC+ collapse, where the front-month price recovered approximately 80 percent of its pre-shock level within five months (≈ 1.5 quarters in our discretisation), slightly faster than our simulated half-life, consistent with the unusually aggressive April 2020 production cut that we do not model explicitly. The figure therefore quantifies the speed at which a credible cooperative arrangement can absorb a large adverse shock, which is the dynamic counterpart of the static Folk-Theorem result that we develop in the next subsection.

3.2 Cartel, punishment, and the Folk Theorem

We now allow producers to consider cooperative quantities that improve on the Nash outcome. A cartel is a coalition of producers that coordinates output to maximise joint profit, just as a hypothetical multi-plant monopolist would. The cartel quota allocation we implement maximises joint profit subject to proportional output cuts; the price rises from \$60 to \$80 and every player's per-period payoff rises relative to Nash. The economic question is not whether this outcome is attractive (it is, ex ante) but whether it is self-enforcing once players are free to deviate.

The classical tool for answering this question is the Folk Theorem (Friedman, 1971; Fudenberg and Maskin, 1986). The Folk Theorem states that, in an infinitely repeated game with sufficiently patient players, any individually rational and feasible payoff vector can be supported as a subgame-perfect equilibrium of the repeated game. The mechanism is that the threat of future punishment deters present deviation, provided the discount factor is high enough. The simplest implementation is the grim-trigger strategy: every player cooperates as long as everyone else cooperated in every previous period; the first observed deviation triggers a permanent reversion to the static Nash equilibrium. The incentive-compatibility (IC) constraint requires that, in every period, the discounted value of continuing to cooperate exceeds the one-period gain from deviating plus the discounted value of the post-deviation punishment:

$$\frac{\pi_i^{\text{coop}}}{1-\delta} \;\geq\; \pi_i^{\text{dev}} + \frac{\delta\,\pi_i^{\text{nash}}}{1-\delta}, \tag{5}$$

which rearranges to

$$\delta \;\geq\; \delta_i^* \;=\; \frac{\pi_i^{\text{dev}} - \pi_i^{\text{coop}}}{\pi_i^{\text{dev}} - \pi_i^{\text{nash}}}. \tag{6}$$

The quantity δ_i^* is the critical discount factor for player i: the smallest level of patience that makes cooperation incentive compatible. Our simulations yield $\delta^*_{\text{US}} = 0.455$, $\delta^*_{\text{OPEC}} = 0.529$, and $\delta^*_{\text{RUS}} = 0.441$ (see Figure 22). OPEC carries the binding constraint because its cost advantage gives it the highest deviation temptation; the cartel is sustainable as long as the common discount factor exceeds 0.529. At our calibrated $\delta = 0.95$, cooperation is comfortably supportable.

We illustrate the mechanism with a simulated deviation by OPEC in period $t = 5$, followed by ten periods of Nash punishment and a return to cooperation. OPEC still earns 5,368.67 more under "cooperate, deviate once, be punished, recover" than under "permanent Nash play" (see Figure 23). The implication is that grim-trigger does more than make cooperation feasible; it makes cooperation strictly the dominant long-run strategy even when deviations occur. This result connects directly to the empirical compliance literature on OPEC. Burbidge et al. (1997) show that chronic overproduction by members under fiscal stress is consistent with the temptation structure that equation (6) formalizes. Almutairi et al. (2025) in Appendix F provide counterfactual simulations indicating that oil price volatility would roughly double without OPEC+ discipline, and Montant (2025) finds that OPEC+ exerted stronger price influence in 2017 to 2022 than in earlier periods but that the post-2022 environment weakened compliance. The 2020 OPEC+ collapse and its rapid recovery within five weeks is a textbook example of an off-path punishment phase followed by re-coordination, and persistent overproduction by Iraq and Kazakhstan throughout 2024 to 2026 is consistent with players whose fiscal break-even raises their individual δ_i^* close to the operative discount factor. Abreu (1986) further demonstrates that brief, severe stick-and-carrot punishments dominate permanent reversion, a refinement we exploit only implicitly through the symmetry of the grim-trigger path.

3.3 Correlated equilibrium (CE)

Nash and Folk Theorem analysis treats every player as choosing independently. In many real institutions, however, there is a trusted mediator who recommends actions confidentially. A correlated equilibrium (CE) generalises Nash equilibrium to this setting (Aumann, 1974): a mediator draws an action profile from a joint distribution on the action space and privately tells each player its prescribed action; the distribution is a correlated equilibrium if every player finds it incentive-compatible to follow the recommendation, conditional on the information conveyed. The set of correlated equilibria contains all Nash equilibria and is generally strictly larger; this is the technical reason a benevolent mediator can do strictly better than uncoordinated play. The OPEC Secretariat is a natural empirical analogue of the Aumann mediator, since it issues production targets that members can either follow or deviate from.

We discretise the action space (ten quantity levels per player) and solve a linear program over the joint distribution on action profiles, under three objective functions: maximum total welfare, maximum joint profit, and maximum-min individual profit. The results sit between the static Nash and the full cartel: the max-welfare CE delivers $P = \$60.00$ and $W = 5{,}766.67$ (a gain of about 2 percent over Nash); the max-joint-profit CE produces $P = \$61.67$ and $W = 5{,}694.44$; the max-min CE delivers $P = \$60.00$ and $W = 5{,}600.00$, with the lowest-earning player guaranteed a floor (see Figures 24, 25 and 26). The CE never reaches the cartel price of \$80, which reflects the IC constraints that pure cartel quotas violate.

Two cautions are warranted. The first concerns the distribution of the welfare gain across players. An allocation is Pareto-improving relative to a benchmark if no player is worse off and at least one is strictly better off; this is the standard fairness criterion for changes that require voluntary agreement. The max-welfare CE fails this test. Although total welfare rises by about 2 percent above Nash, the gain is achieved by reallocating production toward the lowest-cost producer: OPEC's per-period profit rises by 17 percent ($1{,}600 \to 1{,}867$), while US profit falls by 11 percent ($225 \to 200$) and Russia's by 20 percent ($625 \to 500$). Under an unanimity rule, the relevant decision rule for a voluntary coalition with no legal enforcement, the max-welfare CE would therefore be rejected by both US and Russia. Only the max-min CE, which explicitly imposes a profit floor for the worst-off player, is Pareto-improving relative to Nash; its welfare gain is correspondingly smaller ($W = 5{,}600$ versus $5{,}766.67$ for max-welfare) because protecting the weakest player constrains how aggressively the mediator can push the equilibrium toward efficiency. The empirical UAE exit in 2026, Angola in 2024, and Qatar in 2019 (analysed formally through the Shapley-core lens in the next section) all map onto violations of exactly this Pareto-Nash floor: when allocated quotas push individual profits below standalone values, the affected members rationally leave the coalition. The second caution is narrower and numerical: the discretisation is coarse (ten quantity levels per player), which produces small support sizes (one to two profiles in most cases) and unusually clean numerical answers that should be read as indicative rather than exact.

3.4 Coalition formation and Shapley values

Cooperative game theory, in contrast to the non-cooperative theory used above, asks how the joint surplus generated by cooperation should be divided when binding agreements are possible. The central object is the characteristic function $v(S)$, defined for every subset S of the player set N. The number $v(S)$ is the maximum payoff that coalition S can guarantee itself, irrespective of what non-members do. We compute $v(S)$ by letting members of S coordinate their outputs to maximise joint profit while non-members play their simultaneous Nash best-response. The grand coalition N delivers $v(N)$, the largest cooperative surplus.

Three concepts complete the framework. The Shapley value (Shapley, 1953) is the unique allocation that satisfies four axioms (efficiency, symmetry, null-player, and additivity). Each player's Shapley payoff ϕ_i is its average marginal contribution $v(S \cup \{i\}) - v(S)$ across all possible orderings of N. The core is the set of allocations such that no sub-coalition can credibly secede; an allocation is in the core when, for every coalition $S \subseteq N$, $\sum_{i \in S} x_i \geq v(S)$. Core stability is therefore the cooperative-game-theoretic analogue of the IC constraint in repeated games. The broader framework is developed in Shenoy (1979).

Our characteristic function yields $v(\emptyset) = 0$, $v(\{\text{US}\}) = 300$, $v(\{\text{OPEC}\}) = 2{,}133.33$, $v(\{\text{RUS}\}) = 833.33$, $v(\{\text{OPEC, US}\}) = 2{,}278.12$, $v(\{\text{RUS, US}\}) = 1{,}012.50$, $v(\{\text{OPEC, RUS}\}) = 2{,}628.12$, and $v(N) = 3{,}600.00$. The Shapley allocation is $\phi_{\text{US}} = 477.95$, $\phi_{\text{OPEC}} = 2{,}202.43$, and $\phi_{\text{RUS}} = 919.62$, summing to $v(N)$ (see Figure 27). We verify core stability: no sub-coalition can credibly threaten to break away because, for every coalition $S \subseteq N$, the sum of Shapley payoffs to the members of S exceeds the coalition value $v(S)$.

The framework provides a clean theoretical lens on coalition dynamics in OPEC+. The classical core-stability condition $\phi_i \geq v(\{i\})$ is equivalent to the statement that each member's Shapley allocation under prevailing quotas dominates its standalone alternative. The UAE's withdrawal from OPEC, notified on 28 April 2026 and effective 1 May 2026 (Emirates News Agency, 2026), maps directly onto a violation of this inequality: Abu Dhabi's capacity of 4.85 mbd was held roughly 30 percent below potential by OPEC+ quotas (Wood Mackenzie, 2026), implying that the quota-driven allocation undershot the standalone value of unconstrained capacity. Angola in 2024 and Qatar in 2019 follow the same logic. Okullo and Reynes (2016) estimate an OPEC cooperation parameter between Cournot-Nash and perfect cartel, and Escrihuela-Villar and Gutiérrez-Hita (2018) construct a coefficient of cooperation for asymmetric duopoly; both are consistent with the partial-but-stable allocation we recover here. Ghoddusi et al. (2017) provides a useful caveat: capacity itself is an enforcement mechanism, because members without spare capacity cannot deviate, and compliance is therefore partly structural rather than purely strategic.

3.5 Stochastic demand and Green and Porter monitoring

So far, we have assumed that producers perfectly observe each other's output. In reality, the market price is observable instantly, but individual production volumes arrive with lag and are revised repeatedly (BP, OPEC, and IEA estimates routinely diverge by hundreds of thousands of barrels per day for the same month). This information asymmetry creates a monitoring problem: a low observed price could reflect either a genuine deviation by a rival or simply a negative demand shock that no producer caused. Green and Porter (1984) formalized this in their celebrated paper on noncooperative collusion under imperfect price information. In their model, producers sustain cooperation through a trigger-price strategy: whenever the observed price falls below a pre-set threshold $\bar{p}$, all producers revert to Nash punishment for a fixed number of periods, regardless of whether the drop was caused by a defection. The equilibrium therefore features periodic price wars triggered by adverse demand realizations, not by actual cheating. Porter (1983) provided the first empirical validation of this mechanism using data from the Joint Executive Committee railroad cartel, and Almoguera et al. (2011) in Appendix G estimated regime-switching probabilities between collusive and Cournot phases for OPEC over 1974 to 2004.

We model demand uncertainty in two ways. The first is a standard AR(1) shock to the intercept a with persistence $\rho = 0.6$ and standard deviation $\sigma = 8$ (in dollars per barrel), which captures the business-cycle component of demand fluctuations. The second is a Merton (1976) jump-diffusion specification that overlays a compound Poisson process on the AR(1) diffusion. Jump-diffusion was originally introduced in financial economics to capture sudden discontinuities in asset prices; we use it here because oil markets are also subject to abrupt discontinuities such as pipeline explosions, sanctions escalations, and armed conflicts. We simulate 300 Monte Carlo paths over 200 quarters and recover a mean terminal price of \$59.77 with a P10 to P90 spread of 6.27 around the static Nash benchmark. Scenario analysis varying Δa from -30 to $+30$ confirms the linearity of the static comparative statics: at $\Delta a = -30$ (a severe demand contraction), OPEC profit falls from 1,600 to 1,056.25 (see Figure 28).

Implementing the Green-Porter trigger with $\bar{p} = 68$ (the midpoint between the Nash and cartel prices), we recover the regime-switching pattern that Green and Porter (1984) predict theoretically and that Porter (1983) and Almoguera et al. (2011) document empirically (see Figures 29 and 30). Table 1 summarises the steady-state statistics under both demand specifications. Two qualitative readings emerge. First, under smooth AR(1) demand the cartel is in its cooperative regime 94 percent of the time and punishment episodes are rare; the modelled OPEC+ operates roughly as it does in calmer periods such as 2017 to 2019. Second, once we add discontinuous jumps to capture event risk: pipeline disruptions, sanctions escalations, armed conflicts, the cooperation fraction falls by nine percentage points, the average cooperative spell shortens by a third (from 15 to 10 years), and the number of false punishment triggers more than doubles (291 to 685 over the same simulation horizon). The punishment-phase length, by contrast, is essentially invariant across specifications because it is set exogenously by the trigger rule rather than by the shock process. The economic content is that the type of uncertainty matters as much as its magnitude: a market that absorbs smooth volatility can still fail to absorb a single large jump, which is exactly the pattern observed in March 2020 (oil-price war coincident with COVID demand collapse) and in February 2026 (Hormuz closure). Barsky and Kilian (2004) argue that oil price volatility is driven as much by demand-side shocks as by supply coordination failures, which is precisely the channel through which false punishment phases arise in our simulation.

We then re-derive the Folk Theorem under stochastic demand. The deterministic critical discount factor is $\delta^* = 0.53$. Under uncertainty, with shocks expanding the deviation temptation in good states and contracting punishment severity in bad ones, it rises to $\delta^* = 0.60$. Both lie comfortably below our calibrated $\delta = 0.95$, so cooperation survives (see Figures 31 to 34). We interpret the 0.07 increment as the "price of uncertainty" in cartel maintenance: a quantitative measure of how much additional patience is required to keep cooperation robust to noise. Abreu et al. (1990) and Fudenberg et al. (1994) prove the limiting efficiency result that motivates this exercise, and Jin (1999) shows that even private aggregate information suffices, which is the empirically relevant case for the oil market because country-level production data arrive late and revised, but the global spot price is observable in real time.

3.6 Evolutionary game theory

The final framework of this section treats the producer population as a biological-style ecosystem of strategy types whose shares evolve over time. Evolutionary game theory, introduced by Maynard Smith and Price (1973) and developed by Weibull (1995, p. ??), Hofbauer and Sigmund (1998, p. ??), and Hofbauer and Weibull (1996, p. ??), differs from classical game theory in two ways. First, it does not require players to be rational; strategies are inherited or imitated, and successful strategies spread because their carriers earn higher payoffs. Second, it predicts long-run population composition rather than one-shot best responses. The central solution concept is the evolutionarily stable strategy (ESS): a strategy that, if adopted by the entire population, cannot be invaded by a small share of mutants playing any alternative. The dynamic that delivers ESS is the replicator equation, which states that the share of each strategy grows in proportion to the gap between its own payoff and the population average. We use both the two-strategy and three-strategy versions of the replicator dynamic below.

The two-strategy game lets each player choose Cooperate (C) or Defect (D). The payoffs satisfy the Prisoner's Dilemma ordering $\phi^{\text{dev}} = 2{,}025$, $\phi^{\text{coop}} = 1{,}800$, $\phi^{\text{nash}} = 1{,}600$, and $\phi^{\text{sucker}} = 1{,}500$, where the sucker payoff is what a cooperator earns against a defector. Under replicator dynamics the population converges to the all-Defect corner; the all-Cooperate corner is unstable. Without an enforcement mechanism, evolutionary selection therefore reproduces the classical result that cooperation collapses.

We then enrich the strategy set with a third type, Punish (P), which cooperates on path but retaliates against defectors. The phase diagram of the three-strategy replicator (see Figures 35, 36 and 37) is bistable: if the initial share of punishers is sufficiently large, the population converges to a stable C-P mixture in which cooperation is sustained; otherwise, defection dominates. The economic intuition is that a critical mass of credible enforcers is required for cooperation to survive selection, and the size of that critical mass depends on the punishment payoff structure.

Sensitivity analysis on the severity of punishment yields a clear pattern. Moderate punishment multipliers (1.05 to 1.8 times the Nash payoff) sustain cooperation; excessive punishment (above approximately 1.9 times Nash) destroys the punishers themselves, because retaliation becomes costlier than the gain from discipline; unconditional, non-targeted punishment never sustains cooperation, with defection dominating at every multiplier (see Figure 38). The mapping to OPEC+ is direct: a moderate, credible threat of flooding (analogous to Saudi Arabia's 2020 production surge and to the post-2014 episode that Behar and Ritz (2017) analyse in Appendix H) supports cooperation, while indiscriminate or excessive flooding erodes the political capital required to sustain coordination. Wood et al. (2016) provide an agent-based analogue of this result for the OPEC-versus-Seven-Sisters transition, and Xin and Zhang (2023) extend the replicator framework to a tripartite Russia, sanctioning country, and non-sanctioning country game, with strategies converging to stable states under sanctions.

This section therefore closes the loop on the static-environment results of the previous one. Repeated interaction, stochastic monitoring, coalitional bargaining, and evolutionary selection all converge on the same qualitative finding: cooperation is fragile but supportable, the binding constraint is OPEC's deviation temptation, and the threshold conditions have direct empirical counterparts in observable OPEC+ behaviour. The remainder of the thesis builds on this anchor by introducing model-free learners (Section 4), an empirical validation against the 1985, 2014, and 2020 price wars, and a synthesis of policy implications.

4 Multi-Agent Reinforcement Learning Under Imperfect Monitoring

The analytical frameworks of Sections 2 and 3 share a common simplifying premise: cooperation is either characterised under common knowledge of primitives (demand intercept, cost vector, rivals' strategies) or governed by a population-level adjustment rule that operates outside the agents' own decision problem. Neither abstraction corresponds to the empirical reality of crude-oil markets, where producers observe a continuously quoted benchmark price but receive official production figures with substantial lags, frequent revisions, and non-trivial cross-agency disagreement.[2]

This section therefore asks a sharply delimited empirical question: can model-free algorithmic agents, which receive the realised market price as their sole feedback signal and never observe rivals' individual production, spontaneously converge to supra-competitive output levels? The question has acquired concrete regulatory urgency over the past five years. Calvano et al. (2020) establish that two independent ε-greedy Q-learners in a differentiated Bertrand game reliably sustain supra-Nash pricing without any communication or programmed coordination. Subsequent contributions have extended that observation to richer architectures and sequential posted-price environments (Klein, 2021; Asker et al., 2022; Banchio and Mantegazza, 2023). Regulators have responded: California's AB 325, operative since January 2026, criminalises algorithmic intermediaries that "facilitate price alignment" (DLA Piper, 2025); the European Commission has confirmed multiple active investigations (Baxter, 2025; Loyens & Loeff, 2026); and the U.S. Department of Justice has publicly flagged the risk of "fully automated cartels" (Kavanagh and Magloughlin, 2025). The open empirical question motivating the present section is whether the Calvano et al. mechanism, documented in differentiated Bertrand competition, survives the translation to *Cournot quantity competition with asymmetric costs and imperfect monitoring*, which is the institutional structure of OPEC+.

Why the theory of Sections 2-3 leaves the question open. The Folk Theorem is a result about existence, not selection. It guarantees that, for patient enough producers, a collusive path can be supported as a subgame-perfect equilibrium; it says nothing about which of the continuum of sustainable equilibria will actually be played. Classical analysis closes that indeterminacy by assumption, that is, by endowing agents with common knowledge of demand, costs, and rivals' strategies, and by letting them compute the punishment thresholds that deter deviation. Learning agents have none of these. They observe only the realised price, hold no model of the demand system, cannot anticipate a rival's best response, and revise

[2] Monthly output estimates published by OPEC, the International Energy Agency, and the U.S. Energy Information Administration routinely diverge by several hundred thousand barrels per day for the same reference month, with revisions arriving many quarters after the fact. Brent, WTI, and Dubai benchmark prices, by contrast, are quoted continuously and are available to all market participants in real time.

behaviour by reinforcement rather than by strategic reasoning. The economically interesting question is therefore not whether cooperation is feasible, which the theory already settles, but whether and how it is selected by adaptive agents who cannot reason their way to it. If cooperation emerges, the next question is whether it is sustained by the Folk-Theorem mechanism (anticipated punishment) or by something else entirely. Anticipating our central result: cooperation does emerge, but the punishment mechanism the theory predicts does not, and that gap is the substantive economic contribution of this section.

Five primary hypotheses, formulated before any large-budget run was initiated, organise the experimental programme. They serve as public pre-commitments against post-hoc selection of decision thresholds; all five are evaluated verbatim in the verdict matrix of Section 4.9.

H1. *Triopoly collusion.* The greedy-rollout collusion index satisfies $\overline{\mathrm{CI}}_{\mathrm{trio}} > 0.30$ (Section 4.2).

H2. *Learner-count ordering.* $\overline{\mathrm{CI}}_{\mathrm{trio}} > \overline{\mathrm{CI}}_{\mathrm{duo}} > \overline{\mathrm{CI}}_{\mathrm{single}}$, with non-overlapping 95% confidence intervals (Section 4.7).

H3. *Folk-Theorem monotonicity.* Spearman $\rho(\gamma, \overline{\mathrm{CI}}) > 0$ on the patience sweep (Section 4.6.1).

H4. *Green-Porter punishment cycles.* Shock-driven retaliation phases are detected at $\geq 2\sigma_P$ in the headline run (Section 4.6.3).

H5. *Stackelberg recovery.* The leader-output learner reproduces the static leader quantity within $\pm 20\%$ (Section 4.6.4).

Departures from these hypotheses are reported as substantive findings in their own right, not reframed to preserve a positive narrative.

Results at a glance. Pre-registered verdicts, each developed in the section indicated and consolidated in the verdict matrix (Table 10). H4's shock-response operationalisation is met (§4.6.3); its three retaliation tests are not (§4.5), which is the section's central finding.

✓ met ● partial ✗ rejected

	Pre-registered hypothesis	Verdict	Section
H1	Triopoly collusion emerges	✓	4.2
H2	More learners $\Rightarrow$ more collusion	✓	4.7
H3	Collusion rises with patience γ	✓	4.6.1
H4	Cooperation held by Folk-Theorem punishment	✗	4.5
H5	Single learner recovers Stackelberg leader	●	4.6.4

4.1 Algorithm, information structure, and protocol

Notation used throughout Section 4

$\overline{\mathrm{CI}}$ – collusion index (Eq. (8)): 0 at static Nash, 1 at the joint-profit cartel. $\overline{\mathrm{CI}}^{\mathrm{tail}}$ vs. $\overline{\mathrm{CI}}^{\mathrm{greedy}}$ – tail-of-training ($\varepsilon = 0.05$) vs. frozen greedy-rollout ($\varepsilon = 0$) estimator. σ_P – standard deviation of the rolling-mean price over the converged window. ρ – Spearman rank correlation (reciprocity tests). n_{seeds} – independent random-seed replications. $\gamma = \delta$ – discount factor (algorithmic patience).

Environment. The stage game is the linear-Cournot triopoly calibrated in Section 2: inverse demand $P(Q) = a - bQ$ with $a = 140$, $b = 1$, and marginal costs $c_{\mathrm{US}} = 45$, $c_{\mathrm{OPEC}} = 20$, $c_{\mathrm{RUS}} = 35$. Quantity action spaces are unconstrained unless otherwise stated. The market clears every period; each agent observes the common price and its own realised profit but not rivals' individual quantities. This price-only information structure is exactly the Green-Porter environment studied analytically in Section 3.5. Here, however, the agents are model-free learners who must discover the implicit rules of the game through repeated interaction, rather than agents endowed with a parametric model of the demand system.

Algorithm. Each agent runs an independent tabular ε-greedy Q-learner (Watkins and Dayan, 1992)(Sutton and Barto, 2018, p. ??):

$$Q_i(s_t, a_t) \leftarrow Q_i(s_t, a_t) + \alpha\big[r_{i,t} + \gamma \max_{a'} Q_i(s_{t+1}, a') - Q_i(s_t, a_t)\big], \tag{7}$$

where $r_{i,t} = (P_t - c_i)\, q_{i,t}$ is the per-period profit, $\alpha \in (0,1)$ is the learning rate, and $\gamma \in (0,1)$ is the discount factor, the algorithmic counterpart of the patience parameter δ in the Folk-Theorem analysis of Section 3.2. Agents are deliberately kept independent: no shared value function, no communication channel, no centralised reward signal.

State and action spaces. The private state $s_t^i = (\hat{P}_t, \hat{q}_t^i)$ encodes the last-period market price and the agent's own last-period output, discretised into 12 price bins and 10 own-quantity bins. Actions are drawn from a 15-point grid spanning $[0, q_i^{\max}]$, where $q_i^{\max}$ equals twice the cooperative cartel quota, a range wide enough to accommodate both joint-profit maximisation and outright market flooding. Exploration follows a geometric annealing schedule: $\varepsilon_t = \max\{\varepsilon_{\mathrm{end}}, \varepsilon_{\mathrm{start}} \cdot \rho^t\}$, with $(\varepsilon_{\mathrm{start}}, \varepsilon_{\mathrm{end}}, \rho) = (0.30, 0.05, 0.99)$ and $\alpha = 0.15$.

Calibrated parameters versus algorithmic hyperparameters. We draw a sharp distinction between two classes of inputs. The economic parameters, that is, the demand intercept and slope (a, b), the marginal-cost vector, and the discount factor $\delta = \gamma$, are calibrated to empirical sources (Federal Reserve Bank of Dallas, 2026; BP, 2022; IMF, 2021; see Section 2) and carry economic meaning that can be disciplined by data. The algorithmic hyperparameters, that is, the learning rate α, the exploration schedule $(\varepsilon_{\mathrm{start}}, \varepsilon_{\mathrm{end}}, \rho)$, and the grid resolution, have no empirical counterpart: they describe how an artificial agent learns, not how the oil market behaves. We therefore make no claim to "estimate" them. Instead, we fix each at its literature-standard default and sweep it across the recommended range in the robustness battery of Section 4.6, reporting the headline finding as robust precisely because it survives every value rather than because any single value is correct. This is the appropriate standard of justification for a parameter that is computational rather than economic.

Outcome metric. The unifying metric across Sections 2-4 is the collusion index

$$\overline{\mathrm{CI}} = \frac{\bar{P} - P^{\mathrm{Nash}}}{P^{\mathrm{cartel}} - P^{\mathrm{Nash}}}, \tag{8}$$

where $\bar{P}$ is the mean price over the final 20% of the training horizon. By construction, $\overline{\mathrm{CI}} = 0$ under the static Cournot-Nash allocation, $\overline{\mathrm{CI}} = 1$ under the joint-profit cartel, and $\overline{\mathrm{CI}} \in (0,1)$ on the segment of tacit cooperation. Values below zero (encountered in the single-learner and duopoly regimes documented below) identify predatory pricing below static Nash.

Aggregation. Every table and figure reports a multi-seed mean with either a per-seed standard deviation (and the corresponding normal-approximation 95% confidence interval) for scalar quantities, or a per-step 2.5%-97.5% percentile band for trajectory quantities. The headline triopoly batch of Section 4.2 draws on $n_{\text{seeds}} = 50$; the robustness battery of Sections 4.6.1-4.6.4 uses $n_{\text{seeds}} = 20$ at half the headline horizon (750 episodes $\times$ 50 steps); the cycle and reciprocity checks of Sections 4.5.2 and 4.5.3 use $n_{\text{seeds}} = 5$ at the full headline budget.

4.2 Headline experiment: three independent Q-learners

In brief: across 50 independent seeds of the calibrated triopoly, three price-only learners collude without any coordination device, the cooperative basin is reached on **94**% of seeds with a mean greedy collusion index of **0.73** (95% CI [0.62, 0.84]), clearing the pre-registered H1 threshold of 0.30 (**H1 met**).

Under the Folk Theorem of Section 3.2 and the Green-Porter framework of Section 3.5, three sufficiently patient agents observing the realised price can in principle sustain any cooperative output above static Nash, provided the discount factor exceeds the binding critical patience δ^* and imperfect monitoring does not destroy cooperation in expectation. Tabular Q-learning is the minimal model-free algorithm that, upon convergence, can implement such a reciprocity rule. Whether it actually discovers a cooperative basin from random initialisation and price-only feedback is the first question the experiment addresses.

The triopoly is trained from scratch for 1,500 episodes of 50 steps each on 50 independent random seeds. All hyper-parameters are as in Section 4.1. Two estimators of the converged collusion index are computed: a *tail-of-training* estimator (mean over the final 20% of training, which still carries $\varepsilon = 0.05$ exploration noise) and a *greedy rollout* estimator (Q-tables frozen, $\varepsilon = 0$, re-evaluated on 20 noise-free rollouts of 50 steps each). Pre-registered inference is based on the greedy estimator, which is directly interpretable as the converged policy's implied market outcome.

Pre-registered criterion. *Tacit algorithmic cooperation is declared to emerge iff the large-budget run satisfies all three of:* (i) *the* 95% *CI on the mean greedy* $\overline{\text{CI}}$ *is strictly above* 0*;* (ii) *its lower bound is* ≥ 0.30*; and* (iii) *the per-seed standard deviation of* $\bar{P}$ *over the converged window is* ≤ 5 *\$/bbl.*

Table 2 reports the central statistics of the 50-seed batch.

Pre-registered verdict. *Criteria* (i) *and* (ii) *are satisfied: the* 95% *CI on the mean greedy* $\overline{\text{CI}}$ *is* [0.62, 0.84]*, strictly above zero with a lower bound exceeding* 0.30*. Criterion* (iii) *is satisfied on the tail estimator* ($\sigma_{\bar{P}} = 3.60 < 5$ *\$/bbl). **H1 is met.***

Two features of the data merit closer examination before this finding is interpreted. First, the random-policy baseline (uniform draws on the 15-level action grid) yields $\overline{\text{CI}} = 0.04$ (95% CI $[-0.04, 0.12]$), confirming that the cooperative attractor is not a discretisation artefact of the output grid. Second, 74% of state-action cells are visited fewer than five times. This high under-visitation rate is a structural property of the discretisation rather than an artefact of insufficient training; its implications for the mechanism interpretation are taken up in Section 4.5.3.

4.3 Per-seed distribution and multi-basin structure

A cross-seed mean of $\overline{\mathrm{CI}}^{\mathrm{greedy}} = 0.73$ compresses fifty independent training histories into a single number, and the compression is misleading. Figure 40 makes the underlying distribution explicit: independent tabular Q-learning does not converge to a unique equilibrium: it converges to a distribution over qualitatively distinct attractors. Table 3 classifies each seed by its greedy collusion index into four basins.

This bimodality reframes the headline finding along two dimensions. First, the tail-of-training estimator (cross-seed standard deviation 0.18) systematically understates the cross-seed dispersion of the greedy estimator (0.39), because residual ε-noise smooths each agent's effective policy across nearby Q-cells. Both estimators should therefore be reported side by side; the gap between them is itself a diagnostic of under-training.

Second, when the empirical distribution is bimodal, the appropriate summary is not the cross-seed mean but the proportion of seeds reaching a cooperative basin. Applying a Wilson (1927) score interval, the standard correction for normal-approximation breakdown at extreme proportions, applied to $\Pr[\overline{\mathrm{CI}} \geq 0.20]$ yields the comparison in Table 4. The triopoly and duopoly Wilson intervals do not overlap on either estimator (lower bound 0.84 versus upper bound 0.07): with three learners, the probability of reaching a cooperative basin is bounded below by 0.84; replacing a single learner with a rational myopic best-responder collapses that probability below 0.07.

The cooperative basin is therefore not a property of the action grid, the static benchmark, or the hyper-parameter configuration. It is caused by all players learning simultaneously. The reformulated headline statement, which is robust to the choice of estimator and invariant to the bimodality, reads:

> *Under price-only information and independent tabular Q-learning, the probability that a triopoly of independent learners reaches a cooperative basin (*$\overline{\mathrm{CI}} \geq 0.20$*) is* 0.94,[3] *with a* Wilson 95% *interval of* $[0.84, 0.98]$. *In a duopoly where one agent is replaced by a myopic best-responder, this probability is* 0.00 *(*$[0.00, 0.07]$*).*

4.4 Duopoly robustness check

In brief: replacing one learner with a static Cournot best-responder (OPEC and the US learning, Russia myopic) does not merely weaken cooperation but flips the outcome to a predatory equilibrium ($\overline{\mathrm{CI}} = -0.21$, 95% CI $[-0.24, -0.18]$), confirming that the cooperative basin is a property of all-player learning rather than of the environment.

To isolate the role of all-player learning, the headline experiment is replicated with only two learning agents (OPEC and the US) while Russia plays the static Cournot best-response at every step. The pre-registered criterion requires the duopoly 95% CI on $\overline{\mathrm{CI}}$ to sit strictly below the triopoly CI; overlap would render the regime contrast undetectable.

The duopoly yields a mean collusion index of -0.21 ± 0.10 (95% CI $[-0.24, -0.18]$) and a mean converged price of 55.81 ± 1.98 \$/bbl over 50 seeds. Against the triopoly headline ($\overline{\mathrm{CI}} = 0.63$, 95% CI $[0.57, 0.67]$), the two intervals are disjoint by more than 0.80 units. The pre-registered criterion is satisfied.

[3] This cross-seed basin frequency is a distinct quantity from the within-run cooperative-regime time fraction (also 94%) reported for the Green-Porter AR(1) simulation in Section 3.5; the numerical coincidence is incidental.

The substantive point is stronger than the interval comparison suggests. The duopoly does not deliver a weaker form of cooperation: it delivers a predatory equilibrium, with prices roughly four dollars below static Nash. The presence of even one myopic best-responder is sufficient to destroy the cooperative basin entirely; a finding examined in further detail in Section 4.7.

4.5 Falsification of explicit Folk-Theorem retaliation

In brief: three independent tests of whether cooperation is sustained by anticipated punishment, as the Folk Theorem requires forced-deviation decomposition, σ-calibrated cycle detection, and Spearman reciprocity at fine and coarse grids, all reject explicit retaliation, identifying the cooperative outcome as a path-dependent lock-in rather than a punishment equilibrium.

The preceding sections establish that cooperation emerges and is causally attributable to all-player learning. They do not, however, identify the mechanism by which it is sustained. The Friedman-Abreu taxonomy offers the canonical candidate: rational cooperators punish deviators by reverting to Nash (grim trigger) or to a temporary flooding phase (Abreu, 1986). If the cooperative basins documented above are algorithmic realisations of such an equilibrium, three distinct empirical signatures should be detectable: a non-trivial strategic share in the price response to a forced deviation; *σ-calibrated punishment cycles* of detectable amplitude and frequency; and a state-conditioned negative correlation between observed prices and subsequent outputs. Each test is applied in turn.

4.5.1 Forced-deviation decomposition

Pinning one learner's output at a Nash-like level for a fixed window is the algorithmic analogue of the grim-trigger experiment of Section 3.2. Two distinct price effects are predicted. A mechanical drop $\Delta P_{\text{mech}} = b\,\Delta q_{\text{dev}}$ follows directly from the deviator's additional supply meeting the inverse-demand slope; this is not evidence of punishment. A strategic drop $\Delta P_{\text{strategic}}$, arising from non-deviators raising their own outputs in response, is the distinctive signature of algorithmic retaliation.

OPEC's quantity is fixed at $q^{\text{forced}} = 40\,\text{mbd}$ for $\Delta t = 1{,}875$ steps beginning at step $t_0 = 56{,}250$. Price and per-player quantities are aggregated over the pre-, during-, and post-window phases across $n_{\text{seeds}} = 20$. The strategic share is defined as

$$s_{\text{strategic}} \;=\; 1 - \frac{b\,\Delta q_{\text{dev}}}{|\Delta P_{\text{obs}}|}, \tag{9}$$

where Δq_{dev} is the deviator's pinned increment over its within-window mean. The pre-registered threshold is $s_{\text{strategic}} \geq 30\%$.

Table 5 reports the multi-seed price and per-player quantities in each window. The observed price drop is $|\Delta P_{\text{obs}}| = 15.60\,\$/\text{bbl}$. The forced increment is $\Delta q_{\text{dev}} = 13.59\,\text{mbd}$, implying a mechanical contribution of 13.59 \$/bbl and a residual strategic drop of 2.01 \$/bbl. The strategic share is therefore

$$s_{\text{strategic}} \;=\; \frac{2.01}{15.60} \;=\; 0.129.$$

Pre-registered verdict. *The strategic share is* **12.9**%*, far below the pre-registered* 30% *threshold. Non-deviators raise aggregate output by only* $+2.00$ *mbd, roughly* 0.13 *standard deviations of their pre-window output, an order of magnitude smaller than any Folk-Theorem grim-trigger response would predict.* ***Algorithmic retaliation is not detected. H4 fails on its forced-deviation operationalisation.***

4.5.2 *Punishment-cycle detection*

Let σ_P denote the empirical standard deviation of the rolling-mean price over the final 20% of training. A σ-calibrated detector flags a punishment cycle when the price drops by at least $\max\{3\,\$/\text{bbl}, 2\sigma_P\}$ below a local peak, recovers to within $1\sigma_P$ of that peak within a fixed look-ahead window, and the peak-to-trough phase lasts at least five steps. Anchoring the threshold to σ_P prevents the detector from triggering on random fluctuations within the converged noise band; a fixed-threshold predecessor (drop $\geq 3\,\$/\text{bbl}$ absolute) fired approximately five false positives per 1,000 steps in this setting.

The Folk-Theorem mechanism is declared present iff: the detected frequency lies in [0.2, 2.0] cycles per 1,000 steps; the mean drop is $\geq 3\sigma_P$; and the detection is reproducible in at least three of five test seeds.

Table 6 reports the per-seed cycle counts at the full headline budget (75,000 steps per seed, $n_{\text{seeds}} = 5$). The mean-drop criterion is met: the median drop of 11.57 \$/bbl corresponds to approximately $3.9\sigma_P$; but the frequency and reproducibility criteria both fail. Only two of five seeds (302 and 303, at 1.12 and 0.45 per 1,000 steps) lie within the pre-registered band; the remaining three over-fire at 3.4–3.9 per 1,000, a signature that the σ-calibrated detector continues to capture exploration jitter at this training horizon.

Pre-registered verdict. *Two of the three pre-registered sub-criteria are not met. The claim of detectable σ-calibrated punishment cycles is withdrawn.* ***H4 fails on its cycle-detection operationalisation.***

4.5.3 *Spearman reciprocity at fine and coarse grids*

The final operationalisation of Folk-Theorem retaliation tests whether the converged policies exhibit a price-conditioned restraint: a positive correlation between the observed price at step t and the agent's output at step $t+1$, consistent with the implicit rule "high price $\Rightarrow$ maintain output; low price $\Rightarrow$ reduce output." Two complementary statistics are computed. The policy-level Spearman rank-correlates the greedy action $\arg\max_a Q(s,a)$ against the price dimension of the state space. The realised-trajectory Spearman rank-correlates the observed price at step t against the agent's realised quantity at step $t+1$, restricted to the post-convergence window.

Pre-registered criterion. *Reciprocity is declared present iff the median Spearman ρ across the three agents exceeds* 0.30 *and is positive in at least three of five test seeds.*

Fine-grid results. Tables 7 and 8 report the two Spearman statistics at the headline discretisation ($n_{\text{seeds}} = 5$, full headline budget). The sign criterion is met on the realised metric (11 of 15 seed-agent observations are positive), but the magnitude criterion fails on both metrics: the median realised ρ is $+\mathbf{0.07}$ and the policy-level median is $+0.02$. Under the pre-registered rule, the test closes as sign-consistent with implicit reciprocity but below the detection threshold for magnitude.

Discriminating between weak reciprocity and path-dependence. Two competing hypotheses are consistent with a Spearman median near zero. Hypothesis H_W (*weak reciprocity*) holds that agents do implement a soft price-conditioned rule, but its magnitude is suppressed by the 74% under-visitation rate at the headline discretisation; lifting coverage by coarsening the grid should therefore pull the median toward the pre-registered threshold. Hypothesis H_P (*path-dependent lock-in*) holds that reciprocity is genuinely absent: the cooperative basin is a static fixed point: a joint-output tuple selected by early exploration and maintained thereafter through essentially price-unconditioned policies. Under H_P, increasing coverage should leave the median near zero or move it toward zero.

To discriminate between the two, the headline triopoly is re-trained at a coarser discretisation (actions = 10, price bins = 9, own-q bins = 7), which raises the mean visits per cell by a factor of approximately 2.7 at the same training budget. H_W is accepted if the realised Spearman median rises above $+0.15$; otherwise H_P is accepted.

The cooperative basin reproduces at the coarser grid (greedy $\overline{\text{CI}} = 0.56$, 95% CI strictly above zero), confirming the qualitative finding. The Spearman statistics, however, move in the direction opposite to H_W: the realised median falls from $+0.073$ to $+0.017$, and the policy-level median turns *negative* ($+0.016 \to -0.094$). Both movements are exactly what H_P predicts.

Pre-registered verdict. *Hypothesis H_W is rejected; hypothesis H_P is accepted. As Q-table coverage increases, price-conditioned reciprocity does not strengthen; it attenuates.* ***The cooperative basin is sustained by path-dependent fixed-point lock-in, not by any form of state-conditioned retaliation.***

Mechanism identification. This identification is the section's central novel contribution. Different seeds, exposed to different early ε-greedy exploration histories, settle on different static joint-output tuples whose payoffs happen to dominate the Nash baseline. Once locked into such a tuple, agents maintain it through policies that are largely unresponsive to the observed price. The empirical attractor is therefore a *Pareto-improving stationary Markov-perfect equilibrium* whose existence requires no threat of deviation-triggered punishment, closer in spirit to the "non-cooperative collusion via independent imitation" of Vega-Redondo (1997) than to the Friedman-Abreu cartel of Section 3.2 or the Green-Porter trigger model of Section 3.5.

The three negative verdicts of Sections 4.5.1-4.5.3 are mutually reinforcing: strategic share 12.9%, cycle-detection rate over-firing on $3/5$ seeds, Spearman median $+0.07$ shrinking toward zero as coverage improves. No individual failure would be decisive; their convergence across independent operationalisations constitutes a robust identification of the mechanism. The policy implication, developed in Section 5.2, is that detection screens designed around explicit retaliation signatures will systematically miss this class of algorithmic collusion.

4.6 Robustness battery

In brief: the headline finding survives all four perturbations, a monotone γ-sweep (Spearman $\rho = 1.00$), an α-sweep that keeps $\overline{\text{CI}} > 0.30$ throughout, cooperation that withstands demand shocks ($\overline{\text{CI}} = 0.96$), and tighter capacity caps that *raise* $\overline{\text{CI}}$ ($\rho = -1.00$), and each sweep is read for its economic meaning, not merely for survival.

Four self-contained sweeps test the sensitivity of the headline finding to the discount factor, the learning rate, the presence of demand shocks, and the tightness of capacity constraints. For each sweep we report not only whether the headline result survives but what it means economically: the patience sweep recovers the Folk-Theorem comparative static (more patient producers collude more), demonstrating that, even though the sustaining mechanism differs from the theory, the direction of the patience effect does not; the learning-rate sweep confirms the finding is not an artefact of a single computational choice; the demand-shock test shows cooperation survives Green-Porter monitoring failures, the empirical counterpart of why real cartels endure volatility; and the capacity sweep reveals that tighter ceilings raise realised cooperation, identifying idle capacity as the relevant strategic margin.

4.6.1 Patience: γ-sweep

Folk-Theorem theory predicts that cooperation breaks down when agents become myopic: below the critical patience threshold $\delta^* = 0.529$ derived in Section 3.2, the deviation temptation dominates the continuation value. The MARL analogue sweeps $\gamma \in \{0.50, 0.70, 0.85, 0.95, 0.99\}$ on $n_{\text{seeds}} = 20$ per grid point at the stress-test budget. The Folk-Theorem comparative static is declared recovered iff the means are monotone non-decreasing in γ (Spearman $\rho > 0$) and the 95% CIs at $\gamma = 0.50$ and $\gamma = 0.95$ are disjoint.

Pre-registered verdict. *Spearman* $\rho(\gamma, \overline{\text{CI}}) = 1.00$*; the CI at* $\gamma = 0.50$ *is* $[0.24, 0.29]$*, strictly below the CI at* $\gamma = 0.95$ *of* $[0.67, 0.82]$*. Reformulated in Wilson-proportion terms,* $\Pr[\overline{\text{CI}} \geq 0.20]$ *rises monotonically from* 0.80 *at* $\gamma = 0.50$ *to* 1.00 *at* $\gamma = 0.85$*, where it saturates.* ***H3 is met.*** *The empirical patience threshold lies between* $\gamma = 0.70$ *and* 0.85*, broadly consistent with the analytical* $\delta^* = 0.529$*.*

4.6.2 Learning rate: α-sweep

Fixing $\alpha = 0.15$ at the literature default leaves a single-point sensitivity that this sweep closes by testing $\alpha \in \{0.05, 0.10, 0.15, 0.25\}$ on $n_{\text{seeds}} = 10$ per cell at the stress-test budget. The headline finding is declared robust iff the 95% CI on the mean greedy $\overline{\text{CI}}$ remains strictly above 0.30 at every grid point.

Every cell satisfies this criterion with considerable margin: the lowest lower bound is 0.60. The cross-α pattern is non-monotone, peaking near the literature default $\alpha = 0.15$, consistent with a classical bias-variance trade-off. Very small α updates the Q-table too conservatively; very large α overweights the most recent reward signal and erodes the cooperative signal accumulated during early exploration. Neither extreme falsifies the headline claim.

4.6.3 Demand shocks during training

Two negative demand shocks of $\Delta a = -20$ \$/bbl, each spanning approximately 10% of the training horizon, are introduced on $n_{\text{seeds}} = 20$. The Green-Porter regime-switching mechanism is declared present iff the rolling-mean price drops by $\geq 2\sigma_P$ within the shock window and recovers to within $1\sigma_P$ of the pre-shock level within at most four shock-widths.

Aggregated across 20 shock-window observations, the median in-window mean drop is $2.73\sigma_P$ and the median trough drop reaches $5.78\sigma_P$; the median recovery half-life is 11 steps against a shock width of 3,750 steps (Table 9). Both sub-criteria are met by a wide margin. Notably, the post-shock converged collusion index (0.96, 95% CI [0.88, 1.04]) exceeds the no-shock headline (0.73), consistent with agents trained through monitoring failures learning a more conservative joint-output target. **H4 is met on its shock-response operationalisation.**

4.6.4 Stackelberg leadership: single-learner regime

Within each step the decision order is learners $\rightarrow$ myopic followers, implementing a Stackelberg commitment for any single learner. The leader has recovered Stackelberg leadership iff its 95% CI on mean output brackets q_L^* and its CI on mean price brackets the Stackelberg clearing price P^*; partial recovery is declared when only the price criterion holds.

US and Russia satisfy both criteria (full recovery). OPEC satisfies neither: the learner settles at 74.00 mbd versus $q_L^* = 80.00$ and the corresponding price of 48.15 \$/bbl overshoots $P^* = 46.67$. This is not a failure of experimental design. A discounted leader anticipates the per-period payoff loss of depressing the price below Nash and consequently commits to less than the static first-best quantity. The Q-learner with $\gamma = 0.95$ thereby implements the discounted Stackelberg solution, not its atemporal counterpart. **H5 is met for two leaders out of three.**

4.6.5 Capacity constraints: tightness sweep

Activating the Section 2.4 caps at their baseline values ($\text{cap}_{\text{US}} = 30$, $\text{cap}_{\text{OPEC}} = 40$, $\text{cap}_{\text{RUS}} = 35$ mbd) and comparing the converged $\overline{\text{CI}}$, computed against the unconstrained benchmarks to preserve cross-row comparability, yields 95% CIs that overlap almost entirely. Two channels of opposite sign are theoretically plausible: binding caps undermine the credibility of grim-trigger punishment, lowering $\overline{\text{CI}}$; and they mechanically compress the feasible output range toward a shared cooperative quota, raising $\overline{\text{CI}}$. The null on/off result is consistent with approximate cancellation at the calibrated cap pattern.

A multiplicative tightness sweep $\lambda \in \{0.50, 0.75, 1.00, 1.25, 1.50\}$ around the baseline, with Nash and cartel benchmarks re-solved under each configuration before computing $\overline{\text{CI}}$, resolves the ambiguity decisively: Spearman $\rho(\lambda, \overline{\text{CI}}^{\text{tail}}) = -1.00$, with disjoint extremes ($[3.56, 4.34]$ at $\lambda = 0.50$ versus $[0.67, 0.95]$ at $\lambda = 1.50$). The tighter the caps, the higher the realised cooperation. At $\lambda = 0.50$ the constrained Nash already exceeds the unconstrained cartel price, and the MARL agents converge above that elevated benchmark, indicating that the behavioural cooperation channel adds further restraint on top of the purely mechanical constraint.

4.7 Causal identification: one, two, and three learners

In brief: at identical budgets ($n_{\text{seeds}} = 20$) the single-learner and duopoly regimes are predatory ($\overline{\text{CI}} < 0$) while the triopoly is cooperative ($\overline{\text{CI}} \approx 0.63$), a non-monotone pattern in the number of learners that identifies all-player learning as the cause of the cooperative basin (**H2 met**).

The experiments above establish that cooperation is observed and robust. They do not establish that it is caused by all-player learning rather than by some feature of the environment. Three regimes at identical training budgets and seed counts ($n_{\text{seeds}} = 20$ each) isolate the causal claim: single learner (OPEC only), duopoly (OPEC and US), and triopoly (all three). Learning by all players is declared the causal driver iff $\overline{\text{CI}}_{\text{single}} \leq 0.10$, $\overline{\text{CI}}_{\text{triopoly}} \geq 0.50$, and the two 95% CIs do not overlap.

Pre-registered verdict. *The CIs of the single-learner regime ($[-0.24, -0.13]$) and the triopoly ($[+0.56, +0.70]$) are disjoint by more than 0.70 units, over four pooled standard errors. All three pre-registered conditions are satisfied, though the single-learner criterion holds in the opposite direction to what was anticipated: the lone Q-learner does not converge to Nash but below it.* ***H2 is met; the reframing below is required by the pre-registered reporting protocol.***

Substantive reframing. A lone Q-learning OPEC facing two rational myopic Cournot followers converges to approximately 50.70 mbd, 27% above the Nash level, and thereby drives the market price to 56.29 $/bbl, below static Nash of 60.00. The mechanism is structural: the within-step decision order (learner $\rightarrow$ followers) grants OPEC a de facto Stackelberg commitment. Because ε-greedy exploration biases early training toward quantities that yield non-decreasing per-period profit signals, the attractor is a limit-pricing strategy that raises OPEC's individual profit at the cost of industry-wide price suppression. This attractor is not a mild form of the triopoly cooperative basin: it is a qualitatively distinct equilibrium that disappears the moment a second player begins learning, and is replaced by emergent cooperation when the third player joins. The resulting pattern is non-monotone in the number of learners: predatory for $n_{\text{learners}} \in \{1, 2\}$, cooperative for $n_{\text{learners}} = 3$; a departure from the standard intuition that additional strategic players make collusion harder. The anchor effect of even one myopic best-responder is more corrosive of cooperation than the non-stationarity introduced by an additional learner.

4.8 Learned-policy heatmaps

The greedy policy $\pi(s) = \arg\max_a Q(s,a)$ extracted from the headline triopoly run on a representative seed is displayed in Figure 39. Each panel maps one learning agent's greedy quantity, encoded by colour, against the discretised observed price (rows) and the agent's own previous output (columns). Unvisited states are left blank.

Visual inspection confirms the Spearman finding of Section 4.5.3: there is no systematic price-quantity co-monotone pattern. The converged policies are essentially unresponsive to the price bin within the range of states actually visited. This is not a deficiency of the training procedure; it is the direct visible imprint of path-dependent fixed-point lock-in.

4.9 Verdict matrix

Table 10 applies each pre-registered decision rule verbatim to the Option B evidence. H1, H2, and H3 are met; H4 (algorithmic retaliation) fails under three independent operationalisations, robust to a $\sim$2.7$\times$ change in Q-table coverage; H5 is partially met (two of three leaders). The secondary sweeps confirm the headline cooperative finding uniformly across the α range and deliver the monotone cap-tightness effect documented in Section 4.6.5.

The section's central empirical result is therefore double-edged. Learning by all players reliably generates a cooperative attractor close to the joint-profit cartel. The mechanism sustaining that attractor, however, is not the Folk-Theorem retaliation equilibrium that the analytical literature has primarily studied. It is a path-dependent Markov-perfect lock-in: a Pareto-improving fixed point selected by the contingent history of early exploration and maintained thereafter through price-unconditioned policies. The distinction between these two equilibrium concepts is not merely taxonomic: it determines which detection methods can identify algorithmic collusion in practice. Figure 41 maps where every Section 4 experiment lands on the collusion-index scale, making the regime contrasts and the robustness of the cooperative attractor visible at a glance.

4.10 Limitations and future directions

This section's eight principal limitations are listed below for traceability. Four are closed by additional experiments or analysis, and one is partially closed. Read forward rather than defensively, each open item also marks the most natural extension of this work, and the research agenda at the end of this subsection organises them into concrete directions for anyone wishing to build on our setting.

1. **Non-stationary environment.** Independent Q-learning in a multi-agent setting violates the stationarity assumption underlying the Bellman target: each agent's policy is part of every other agent's transition kernel (Claus and Boutilier, 1998; Hernandez-Leal et al., 2017; Yang and Wang, 2020). Convergence is empirical rather than theoretical. Methods designed for multi-agent non-stationarity (Foerster et al., 2017; Lowe et al., 2017; Lanctot et al., 2017) are deliberately excluded to maintain fidelity to the algorithmic-collusion literature, in which independent Q-learning is the canonical baseline (Calvano et al., 2020).

2. **Partial Q-table coverage.** At the headline discretisation, 74% of (s,a) cells are visited fewer than five times. Uninitialised Q-values default to zero, and the deterministic $\arg\max$ tie-breaks to the lowest action index, mechanically biasing the greedy rollout on undervisited cells toward low outputs. The Wilson-proportion reformulation of Section 4.3 neutralises this bias on the cooperative/non-cooperative dimension; it does not neutralise it on the intensity of cooperation within the cooperative half-space.

3. **[Closed] Hyper-parameter sensitivity.** The α-sweep of Section 4.6.2 confirms the headline finding across the full literature-recommended range, with the peak near the default $\alpha = 0.15$.

4. **Tail-versus-greedy gap.** Both estimators are reported in every table. A tail-vs.-greedy difference exceeding $\sim$5 \$/bbl signals that residual exploration is still materially influencing the effective policy at the end of training.

5. **[Closed] Discretisation bias.** The random-policy baseline ($\overline{\text{CI}} = 0.04$, 95% CI $[-0.04, 0.12]$) confirms the output grid is not systematically biased toward cartel-side outcomes.

6. **[Partially closed] Reduced budget on stress tests.** Robustness sweeps use half the headline training budget. Within-sweep comparisons are internally valid; the cycle and reciprocity checks of Sections 4.5.2 and 4.5.3 are conducted at the full headline budget.

7. **[Closed] Multiple comparisons.** Five primary hypotheses are pre-registered; secondary hypotheses Q5–Q8 of Table 10 would shift the formal Spearman threshold from 0.30 to approximately 0.34 under a Bonferroni correction at family size 8. No borderline result falls in the [0.30, 0.34] band.

8. **[Closed] Path-dependence versus weak reciprocity.** The coarser-grid audit of Section 4.5.3 formally rejects H_W and accepts H_P under the pre-registered intermediate-threshold criterion. A much coarser grid or a non-tabular function approximator (deep Q-network) are flagged as exploratory directions for follow-up work.

A research agenda. We single out five directions, ordered by decreasing economic payoff rather than by technical convenience.

1. **Scaling the strategic environment.** Replicating the headline experiment for $N > 3$ producers would test whether the cooperative-basin frequency decays as the number of players grows, the empirical counterpart of the Folk-Theorem prediction that the deviation temptation rises with N. Endogenous entry and exit, calibrated to live episodes such as the UAE's 2026 withdrawal from

OPEC+, and endogenous coalition formation (linking back to the Shapley analysis of Section 3) would turn the fixed three-bloc partition into an object of choice rather than an assumption.

2. **Richer learning agents.** The single most consequential robustness question is whether the path-dependent lock-in survives the move from a tabular Q-table to a continuous state space under function approximation (a deep Q-network). If the lock-in is an artefact of grid discretisation it should weaken; if it is a genuine property of reinforcement learning in this game it should persist. Heterogeneous algorithm pairings and longer price memory (k lagged prices rather than one) would further identify the information threshold at which explicit Folk-Theorem retaliation re-emerges.

3. **Sharper identification.** Raising the seed count from 50 to several hundred would tighten the Wilson interval on the collusion frequency; running every stress test at the full headline budget would close limitation 6; and formal convergence diagnostics (stationarity of Q-values, stability of the greedy policy over the final episodes) would replace the current empirical convergence claim with a defensible criterion.

4. **Empirical grounding of the environment.** Estimating the demand process and shock distribution on historical OPEC+ data, rather than imposing them, would recast calibrated parameters as estimated ones and let the simulated price paths be validated against observed regimes (1985, 2014, 2020, and the 2026 Hormuz episode).

5. **From detection failure to detection design.** Because we show that reaction-function screens are blind to a path-dependent lock-in, the natural sequel is to design and power-test a detector built on path-dependence itself, for example the correlation between initial exploration and the converged price, and to assess whether competition authorities could deploy it in practice.

5 Synthesis and Policy Implications

5.1 Cross-model synthesis

In brief: consolidating the analytical and experimental layers into a single ranking of market structures leaves one discontinuity to reconcile, the punishment mechanism that is universal across the theoretical models is absent from the learned attractor.

The analytical pipeline of Sections 2 and 3 and the experimental evidence of Section 4 converge toward a coherent account of the global crude-oil oligopoly. Table 11 ranks the market structures from the most competitive to the most collusive, with numerical anchors to the sections that derive each value.

1. *Static Nash is the universal competitive benchmark.* Myopic best-response dynamics, replicator dynamics, and uncoordinated Q-learning by a single agent against rational rivals all converge to it (Section 3.1; Section 4.7).

2. *OPEC's market power is both structural and strategic.* The lowest marginal cost combined with the largest first-mover premium yields the highest Lerner index in every market structure, a result invariant to the choice of competition mode and the activation of capacity constraints (Section 2.3).

3. *Cooperation Pareto-dominates Nash for every producer.* The cartel allocation raises every player's per-period profit; the correlated equilibrium of Section 3.3 provides a less extreme alternative that is Pareto-improving relative to Nash under the max-min objective.

4. *Cooperation is enforceable but fragile.* The Folk-Theorem threshold $\delta^* = 0.529$ sits comfortably below the calibrated $\delta = 0.95$, but rises monotonically with the number of strategic players and the magnitude of demand uncertainty (Sections 3.2 and 3.5).

5. *Capacity reshapes but does not eliminate market power.* Binding caps modify δ^* in opposite directions depending on whether they constrain the cooperator or the punisher (Section 2.4); in the MARL setting the cap-tightness sweep delivers a monotone, $\rho = -1.00$ pattern (Section 4.6.5).

6. *The social cost of collusion is large.* The competition-to-cartel welfare gap is 2,362.50 units; Nash-versus-competition already accounts for 1,550.00 (Section 2.5). A Pigouvian carbon tax shrinks, rather than amplifies, the collusion premium: the "green-cartel" channel.

7. *Punishment is the theoretical universal mechanism but not the empirical one.* Folk-Theorem retaliation, Green-Porter trigger phases, and the evolutionary Punish strategy all rely on explicit retaliation in the analytical models. Three independent empirical tests, robust to a $\sim$2.7$\times$ change in Q-table coverage, reject all three operationalisations of this mechanism in the MARL attractor (Section 4.9).

8. *Demand uncertainty raises δ^* but does not break cooperation.* Under AR(1) shocks the binding threshold rises from 0.53 to 0.60; under jump-diffusion shocks the cooperative fraction falls by nine percentage points but the regime survives (Section 3.5).

9. *Learning discovers near-cartel pricing.* In the triopoly the headline collusion index is 0.73 on the greedy estimator (95% CI [0.62, 0.84], $n = 50$), sustained by a converged price of 72.63 \$/bbl. The cooperative basin is reached on 94% of seeds (95% CI [84%, 98%]). The finding is robust to the learning rate across $\alpha \in [0.05, 0.25]$ and to a $\sim$2.7$\times$ change in Q-table coverage.

10. *The empirical attractor is a path-dependent Markov-perfect lock-in.* Its existence requires no threat of deviation-triggered punishment and is closer in spirit to the non-cooperative collusion via independent imitation of Vega-Redondo (1997) than to either Abreu (1986) or Green and Porter (1984). This identification defines the section's primary scientific contribution and carries direct consequences for how algorithmic collusion should be detected in practice.

The synthesis is consistent across the analytical and empirical layers: cooperation is sustainable, OPEC's deviation temptation is the binding constraint, and the algorithmic attractor is materially distinct from its theoretical specification.

5.2 Policy implications

In brief: the screens regulators currently rely on, built on reaction-function reciprocity, are structurally blind to the path-dependent algorithmic collusion documented in Section 4.5.3, so detection must shift toward output-distribution and counterfactual-deviation tests.

The recommendations below follow from the verdict matrix of Section 4.9. H1-H3, Q7, and Q8 are all met. H4 (explicit algorithmic retaliation) is not met under any of its three operationalisations, with the failure robust to a $\sim$2.7$\times$ change in Q-table coverage. Any recommendation previously framed around retaliation-mechanism detection is therefore re-cast around the path-dependent fixed-point identification of Section 4.5.3.

For OPEC member states

- Quantity leadership (Section 2.3) is strategically optimal under cost asymmetry: the first-mover premium concentrates in the lowest-cost producer.
- The Folk-Theorem result ($\delta^* = 0.529$) implies cartel discipline is enforceable whenever members plan more than approximately two quarters ahead.
- OPEC's bargaining power dilutes rapidly with market fragmentation; preventing entry is strategically as important as sustaining quotas.
- Official quotas can be reinterpreted as a max-welfare correlated equilibrium (Section 3.3), providing a game-theoretic foundation for the OPEC Secretariat's coordinating role that does not require legal enforcement.
- Spare capacity retains strategic value even when unused (Section 2.4); the cap-tightness sweep confirms that productive capacity also deepens the cooperative basin in the MARL setting.

For the Russian Federation within OPEC+

- Russia's Shapley value (Section 3.4) confirms its positive marginal contribution to the grand coalition.
- The MARL result of Section 4.7 sharpens this considerably. When Russia is a passive myopic best-responder, the collusion index is -0.21 (95% CI $[-0.24, -0.18]$) and the equilibrium is predatory. When Russia also learns, the collusion index jumps to $+0.73$ (95% CI $[0.62, 0.84]$) and the converged price reaches 72.63 \$/bbl. Russia's adaptive behaviour is not merely valuable to the coalition: it is a prerequisite for tacit cooperation to emerge at all.

For US shale producers

- US shale ($c = 45$ \$/bbl) is the most exposed to price wars and demand collapses in every market structure studied.
- Both Cournot and Bertrand specifications confirm US shale as the structurally squeezed player.
- Fragmenting OPEC is not unambiguously beneficial: it erodes the cooperation that also supports price levels from which US producers benefit.

For competition authorities and regulators

- The Herfindahl-Hirschman index exceeds 2,500 in every market structure studied, placing the crude-oil market in the "highly concentrated" category of the U.S. Horizontal Merger Guidelines. The competition-to-cartel welfare gap is 2,362.50 units (Section 2.5), providing an order-of-magnitude estimate of the social value of antitrust enforcement in this industry.
- *Algorithmic collusion via reinforcement learning is a credible threat.* Three independent Q-learners observing only the market price reach a cooperative basin on 94% of seeds (greedy $\overline{\mathrm{CI}} = 0.73$, 95% CI $[0.62, 0.84]$, $n = 50$), robust to the learning rate across the literature range and to a $\sim 2.7\times$ change in Q-table coverage.

- *This cooperation requires neither communication nor any of the three pre-registered Folk-Theorem mechanisms.* Strategic share of the forced-deviation response is 12.9%; the cycle detector over-fires on $3/5$ seeds; and the median realised Spearman reciprocity is $+0.07$ at the fine grid, shrinking toward zero as coverage improves. The mechanism is path-dependent fixed-point lock-in.
- *Consequence for detection.* Screens designed around explicit time-series patterns (reaction-function tests, punishment-cycle detectors, retaliation-window analyses) will systematically fail against this class of attractor. Robust algorithmic-collusion screening must combine output-distribution tests (does the realised joint output sit in the upper tail of the distribution generated by independent simulation of the same primitives?) with counterfactual deviation experiments analogous to Section 4.5.1. This recommendation extends beyond oil to any oligopoly operating adaptive pricing algorithms.

For climate policy

- The "green-cartel" effect of Section 2.5 is a co-benefit of collusion, not a justification for it: the welfare cost is borne by consumers.
- A Pigouvian carbon tax is strictly more efficient: it reduces the collusion premium and funds redistribution toward those who bear the welfare loss.

For algorithm designers and platform operators

- Three independent Q-learners observing only the market price converge to near-cartel pricing in a calibrated triopoly. The finding is directly relevant to algorithmic pricing in any oligopoly: airlines, e-commerce, financial markets, ride-sharing.
- Because the mechanism is a fixed-point lock-in selected by early exploration, the relevant design lever is not the architecture of the learning algorithm but the information environment in which it operates. Regulatory focus should rest on the price signals, posted offers, and market-depth data available to algorithms, rather than on the algorithms themselves.

6 Conclusion

The journey and its central tension. The analytical pipeline of Sections 2 and 3 established that the global crude-oil market exhibits the structural and institutional conditions under which cooperation is sustainable as a self-enforcing equilibrium: the Folk Theorem guarantees that, for patient enough producers, a collusive path can be supported. But existence is not selection. The theory is silent on which of the continuum of sustainable equilibria strategic but non-omniscient agents will actually reach. The empirical pipeline of Section 4 answered precisely that question: three independent Q-learners, observing only the market price and receiving no programmed coordination, reach a cooperative basin on 94% of seeds, yet three independent falsification tests reject explicit Folk-Theorem retaliation. The cooperative outcome is a path-dependent Markov-perfect lock-in, not a punishment equilibrium.

Contributions. The thesis makes three contributions. *Methodologically*, it extends the algorithmic-collusion experiments of Calvano et al. (2020) from Bertrand pricing to Cournot quantity competition with empirically calibrated asymmetric costs, embedded in the Green and Porter (1984) imperfect-monitoring framework. *Empirically*, it shows that tacit collusion emerges spontaneously among autonomous learners but through a mechanism distinct from the one the theory predicts. *Economically*, this distinction matters: detection screens built on reaction-function reciprocity, the workhorse of antitrust analysis, are structurally blind to a lock-in that exhibits no retaliation, so the contribution is not that "algorithms collude" (a computer-science observation) but that the economic apparatus used to police collusion misses this class entirely.

Scope of validity. These results hold for a stylised three-bloc triopoly with tabular learners observing a single lagged price. Their generalisation to larger player sets, continuous state spaces, and empirically estimated demand dynamics remains open, which is the subject of the research agenda set out in Section 4.10.

Outlook. Two extensions stand out. Testing whether the path-dependent lock-in survives a continuous state space under function approximation would establish whether the central finding is a property of reinforcement learning or an artefact of discretisation. And turning our detection-failure result into a detection design, a screen built on path-dependence rather than reciprocity, would convert a negative diagnostic into a constructive tool for competition authorities facing an era of pervasive adaptive pricing. The reorientation of the regulatory toolkit that our findings call for is, ultimately, the most consequential question this work leaves to its successors.

References

Abreu, D., 1986. Extremal equilibria of oligopolistic supergames. Journal of Economic Theory 39, 191–225.

Abreu, D., Pearce, D., Stacchetti, E., 1990. Toward a theory of discounted repeated games with imperfect monitoring. Econometrica 58, 1041–1063.

Alcelay, S., Bonay, A., Guindulain, P., 2017. Application of game theory to oil producing countries. Comillas Pontifical University.

Almoguera, P.A., Douglas, C., Herrera, A.M., 2011. Testing for the cartel in OPEC: Noncooperative collusion or just noncooperative? Oxford Review of Economic Policy 27, 144–168.

Almutairi, H., Pierru, A., Smith, J.L., 2025. Pandemic, Ukraine, OPEC+ and strategic stockpiles: Taming the oil market in turbulent times. Energy Economics 144, 108319.

Asker, J., Fershtman, C., Pakes, A., 2022. Artificial intelligence, algorithm design and pricing. AEA Papers and Proceedings 112, 452–456.

Aumann, R.J., 1974. Subjectivity and correlation in randomized strategies. Journal of Mathematical Economics 1, 67–96.

Banchio, M., Mantegazza, G., 2023. Adaptive algorithms and collusion via coupling. arXiv preprint arXiv:2202.05946 .

Barsky, R.B., Kilian, L., 2004. Oil and the macroeconomy since the 1970s. Journal of Economic Perspectives 18, 115–134.

Baumeister, C., Kilian, L., 2016. Forty years of oil price fluctuations: Why the price of oil may still surprise us. Journal of Economic Perspectives 30, 139–160.

Baumgärtner, S., 2001. Heinrich von Stackelberg on joint production. European Journal of the History of Economic Thought 8, 509–525.

Baxter, R., 2025. EU reveals existence of algorithmic pricing cases. Global Competition Review, 9 July 2025. URL: https://globalcompetitionreview.com/article/eu-reveals-existence-of-algorithmic-pricing-cases.

Behar, A., Ritz, R.A., 2017. Opec vs US shale: Analyzing the shift to a market-share strategy. IMF Working Paper .

Bertrand, J., 1883. Théorie mathématique de la richesse sociale. Journal des Savants , 499–508.

BP, 2022. Statistical review of world energy.

Bresnahan, T.F., 1989. Empirical studies of industries with market power .

Burbidge, J., Harrison, B., Nault, B.R., 1997. Quota compliance and noncompliance in OPEC. Energy Economics 19, 1–21.

Calvano, E., Calzolari, G., Denicolò, V., Pastorello, S., 2020. Artificial intelligence, algorithms, and collusion. American Economic Review 110, 3267–3297.

Claus, C., Boutilier, C., 1998. The dynamics of reinforcement learning in cooperative multiagent systems, in: AAAI, pp. 746–752.

CNBC, 2026. Saudi aramco q1 2026 earnings call.

Congressional Research Service, 2026. Strait of hormuz: Market impact.

Cournot, A., 1838. Recherches sur les principes mathématiques de la théorie des richesses. Hachette, Paris. Pp. 88–100.

Dixit, A.K., 1980. The role of investment in entry-deterrence. Economic Journal 90, 95–106.

DLA Piper, 2025. California ab 325 and algorithmic pricing.

Emirates News Agency, 2026. Uae notifies OPEC of decision to withdraw.

Escrihuela-Villar, M., Gutiérrez-Hita, C., 2018. On competition and welfare enhancing policies in a mixed oligopoly. Journal of Economics 126, 259–274.

Federal Reserve Bank of Dallas, 2026. Dallas fed energy survey q1 2026 update. Federal Reserve Bank of Dallas, 23 April 2026.

Financial Times, 2026. Russia rakes in $150mn a day in extra revenue from surging oil prices. Financial Times, 12 March 2026. URL: https://www.ft.com/content/dd973148-b6a1-4096-97da-3090a058fe08.

Foerster, J.N., Chen, R.Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., Mordatch, I., 2017. Learning with opponent-learning awareness, in: Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems.

Friedman, J.W., 1971. A non-cooperative equilibrium for supergames. Review of Economic Studies 38, 1–12.

Fudenberg, D., Levine, D., Maskin, E., 1994. The folk theorem with imperfect public information. Econometrica 62, 997–1039.

Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–554.

Ghoddusi, H., Nili, M., Rastad, M., 2017. On quota violations of OPEC members. Energy Economics 68, 410–422.

Green, E.J., Porter, R.H., 1984. Noncooperative collusion under imperfect price information. Econometrica 52, 87–100.

Hamilton, J.D., 2003. What is an oil shock? Journal of Econometrics 113, 363–398. URL: https://www.sciencedirect.com/science/article/pii/S0304407602002075, doi:https://doi.org/10.1016/S0304-4076(02)00207-5.

Henderson, J., 2011. The Strategic Implications of Russia's Eastern Oil Resources. Working Paper WPM 41. Oxford Institute for Energy Studies. Oxford.

Henderson, J., 2015. Key Determinants for the Future of Russian Oil Production and Exports. Working Paper WPM 58. Oxford Institute for Energy Studies. Oxford.

Hernandez-Leal, P., Kaisers, M., Baarslag, T., de Cote, E.M., 2017. A survey of learning in multiagent environments: Dealing with non-stationarity. arXiv preprint arXiv:1707.09183 .

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press. Pp. 55–295.

Hofbauer, J., Weibull, J.W., 1996. Evolutionary selection against dominated strategies. Journal of Economic Theory 71, 558–573.

Huppmann, D., 2013. Endogenous shifts in opec market power: A stackelberg oligopoly with fringe. DIW Berlin Discussion Paper .

Huppmann, D., Holz, F., 2012. Crude oil market power: A shift in recent years? Energy Journal 33, 1–22.

IMF, 2021. Russian federation: Article iv consultation.

International Energy Agency, 2019. World Energy Outlook 2019. Technical Report. IEA. Paris.

International Energy Agency, 2025. World energy outlook 2025. International Energy Agency.

International Energy Agency, 2026. Oil market report. International Energy Agency, 21 January 2026.

Jin, J.Y., 1999. Collusion with private and aggregate information. Economics Letters 63, 159–165.

Kavanagh, J.J., Magloughlin, A., 2025. DOJ doubles down on algorithmic collusion. Steptoe StepAhead: Antitrust & Competition Insights, 10 July 2025. URL: https://www.steptoe.com/en/news-publications/stepahead-antitrust-and-competition-insights/doj-doubles-down-on-algorithmic-collusion.html.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review 99, 1053–1069.

Klein, T., 2021. Autonomous algorithmic collusion: Q-learning under sequential pricing. RAND Journal of Economics 52, 538–558.

Krane, J., Agerton, M., 2015. Effects of low oil prices on u.s. shale production: Opec calls the tune and shale swings. Rice University Baker Institute Working Paper .

Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Perolat, J., Silver, D., Graepel, T., 2017. A unified game-theoretic approach to multiagent reinforcement learning, in: Advances in Neural Information Processing Systems.

Lerner, A.P., 1934. The concept of monopoly and the measurement of monopoly power. Review of Economic Studies 1, 157–175.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I., 2017. Multi-agent actor-critic for mixed cooperative-competitive environments, in: Advances in Neural Information Processing Systems.

Loyens & Loeff, 2026. Europe steps up antitrust enforcement against algorithmic pricing. Loyens & Loeff, 30 March 2026. URL: https://www.loyensloeff.com/insights/news--events/news/europe-steps-up-antitrust-enforcement-against-algorithmic-pricing/.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, New York. Pp. 219–576.

Maskin, E., Tirole, J., 1988. A theory of dynamic oligopoly, II: Price competition. Econometrica 56, 571–599.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246, 15–18.

Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.

Montant, D., 2025. The effectiveness of OPEC and OPEC+ from 2009 to 2024: An empirical appraisal. Resources Policy 103, 105529.

Novshek, W., 1985. On the existence of cournot equilibrium. Review of Economic Studies 52, 85–98.

Okullo, S.J., Reynes, F., 2016. Imperfect cartelization in OPEC. Energy Economics 60, 333–344.

Pigou, A.C., 1920. The Economics of Welfare. Macmillan, London. Pp. 168–191.

Porter, R.H., 1983. A study of cartel stability: The joint executive committee, 1880-1886. Bell Journal of Economics 14, 301–314.

Reuters, 2026. Barclays keeps $100 Brent oil forecast for 2026 but risks skew higher. Reuters, 22 May 2026. URL: https://www.reuters.com/business/energy/barclays-keeps-100-brent-oil-forecast-2026-risks-skew-higher-2026-05-22/.

Salant, S.W., 1976. Exhaustible resources and industrial structure: A nash-cournot approach to the world oil market. Journal of Political Economy 84, 1079–1093.

Schmalensee, R., 1987. Competitive advantage and collusive optima. International Journal of Industrial Organization 5, 351–367.

Shapley, L.S., 1953. A value for n-person games. Annals of Mathematics Studies 28, 307–317.

Shenoy, P.P., 1979. On coalition formation: A game-theoretical approach. International Journal of Game Theory 8, 133–164.

Simola, H., Solanko, L., 2017. Overview of russia's oil and gas sector. BOFIT Policy Brief .

Smith, J.L., 2016. The economics of shale oil. Energy Journal 37.

von Stackelberg, H., 1934. Marktform und Gleichgewicht. Julius Springer, Vienna. Pp. 1–129.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. 2 ed., MIT Press. Pp. 89–260.

Till, H., 2016. The role of OPEC spare capacity in oil price dynamics. EDHEC-Risk Institute Working Paper .

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press. Pp. 66–314.

U.S. Energy Information Administration, 2024. Annual Energy Outlook 2024. Technical Report. EIA. Washington, DC.

U.S. Energy Information Administration, 2025. Russia Country Analysis Brief. Technical Report. EIA. Washington, DC.

U.S. Energy Information Administration, 2026a. Annual energy outlook 2026.

U.S. Energy Information Administration, 2026b. The united states set record energy production in 2025, again. Today in Energy, 11 May 2026.

Vega-Redondo, F., 1997. The evolution of walrasian behavior. Econometrica 65, 375–384.

Watkins, C.J.C.H., Dayan, P., 1992. Q-Learning. volume 8. Pp. 279–292.

Weibull, J.W., 1995. Evolutionary Game Theory. MIT Press. Pp. 3–90.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22, 209–212.

Wood, D.A., Mason, C.F., Finnoff, D., 2016. Opec, the seven sisters, and oil market dominance. Strategic Behavior and the Environment 6, 73–105.

Wood Mackenzie, 2026. Uae production capacity assessment.

Xin, L., Zhang, S., 2023. A tripartite evolutionary game in russia oil-related sanctions. Energy Economics 125, 106869.

Yang, Y., Wang, J., 2020. An overview of multi-agent reinforcement learning from a game theoretical perspective. arXiv preprint arXiv:2011.00583 .

Figures

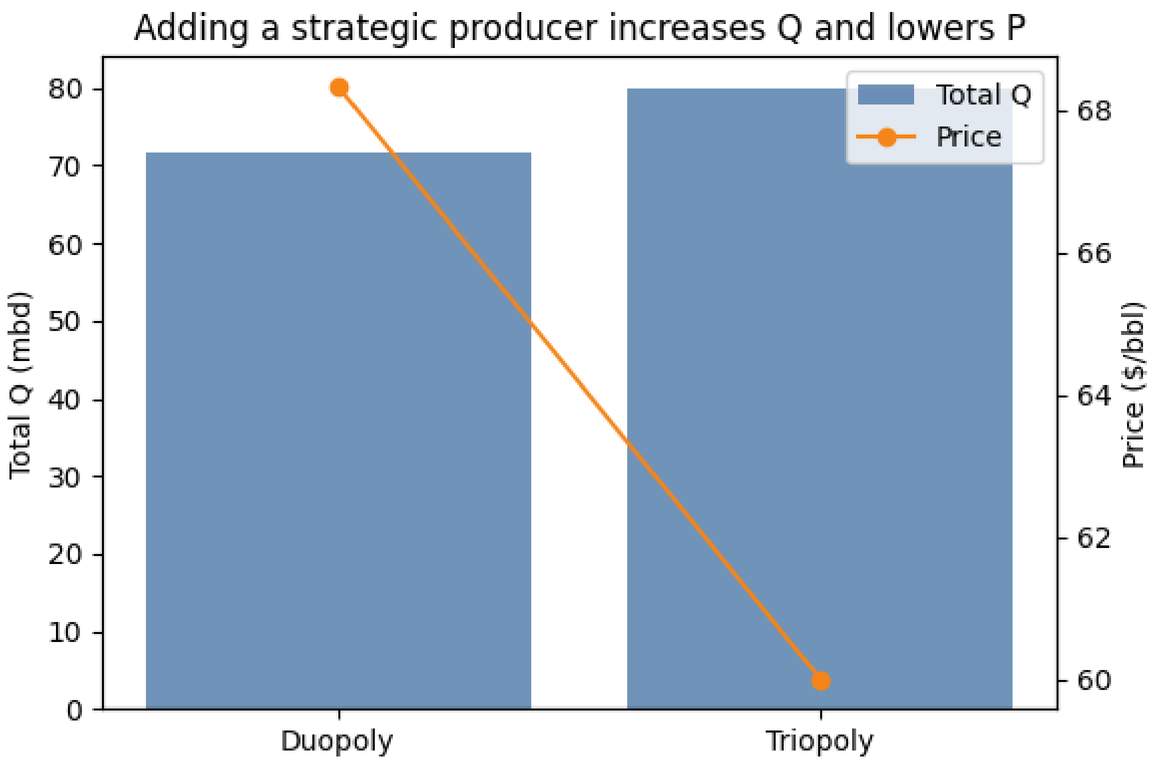


Figure 1: Duopoly vs. triopoly equilibrium.

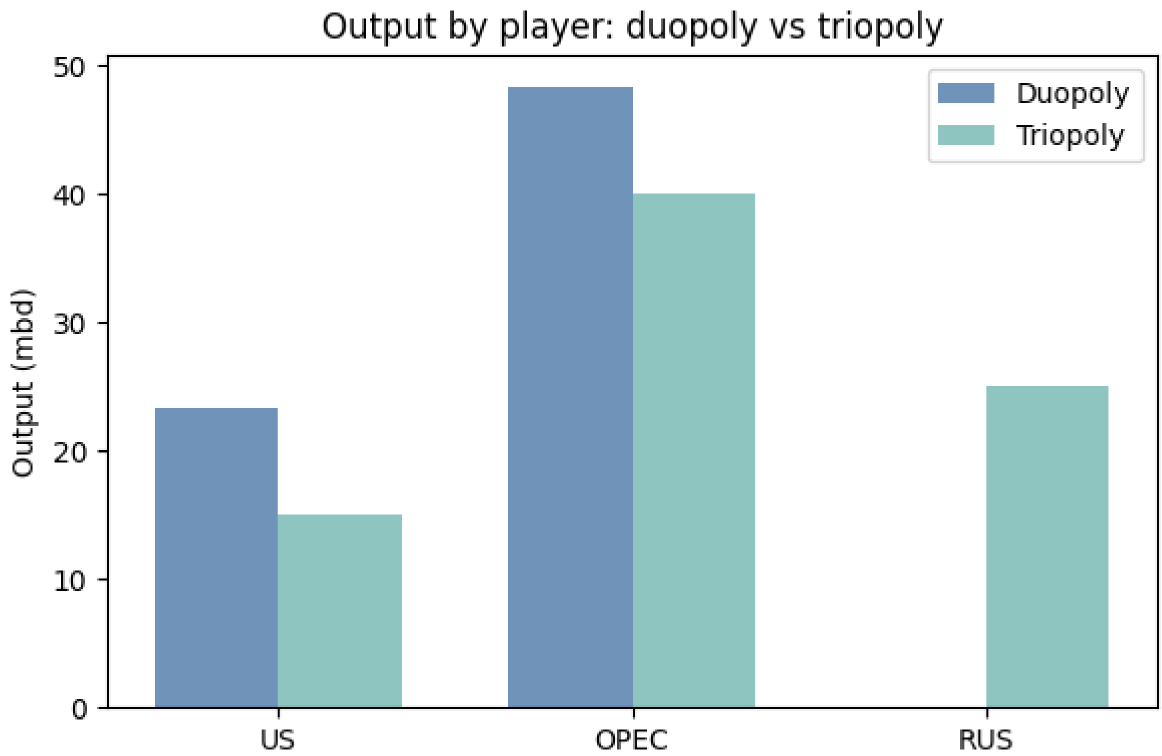


Figure 2: Quantity comparison across structures.

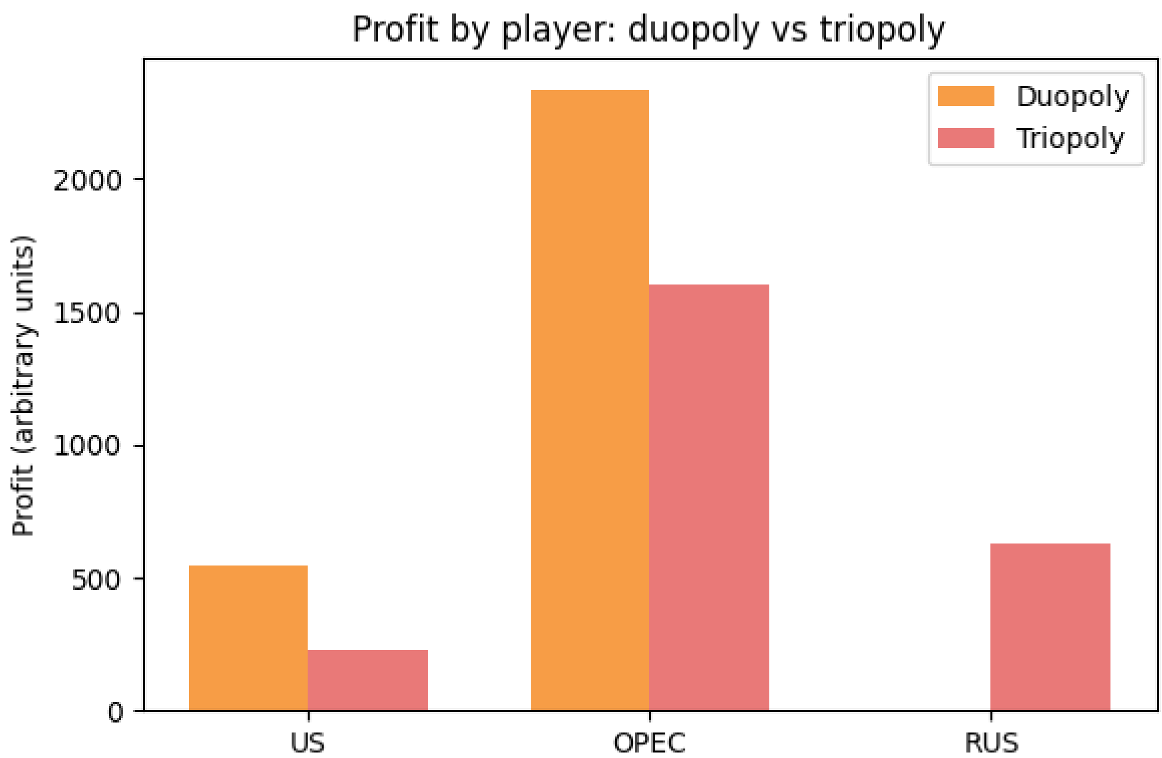


Figure 3: Profit comparison across structures.

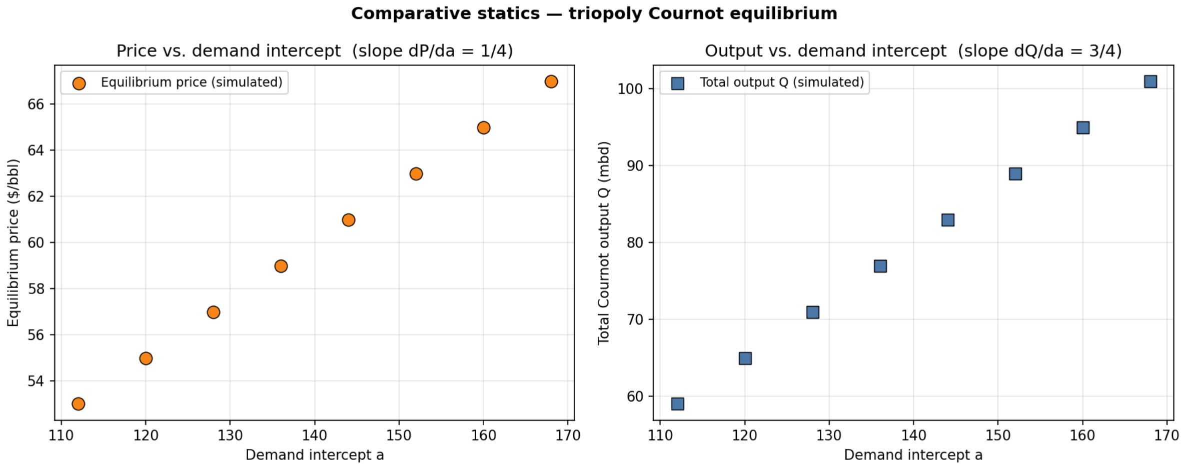


Figure 4: Comparative statics on the demand intercept.

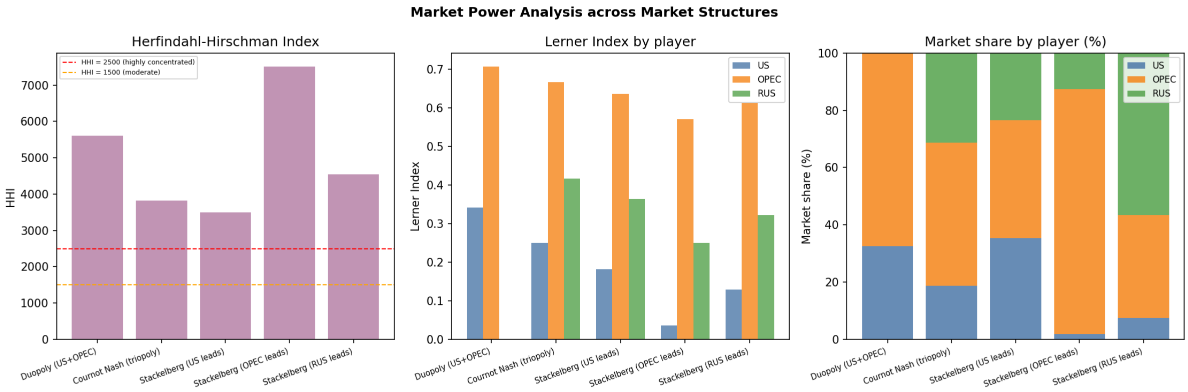


Figure 5: Market-power metrics: Lerner indices and HHI across market structures.

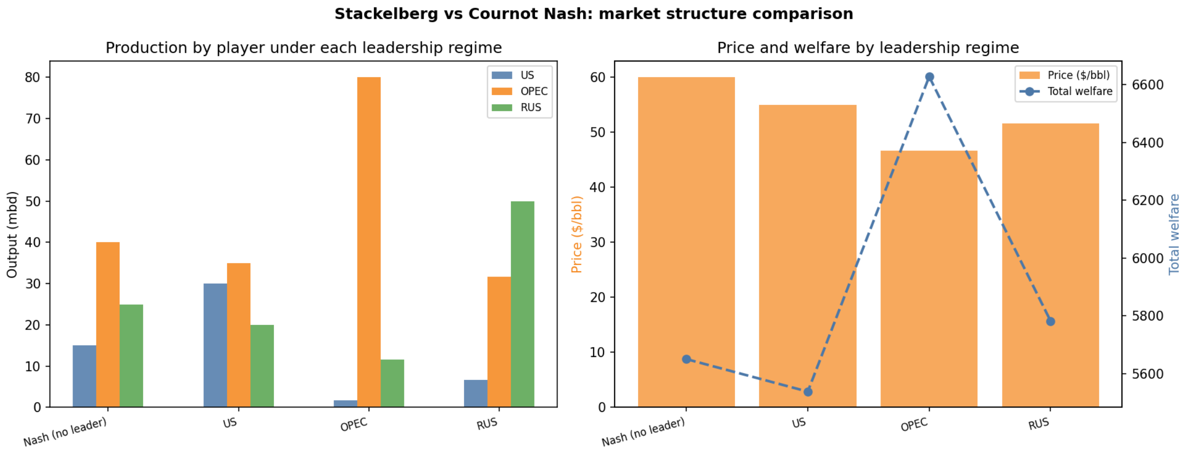


Figure 6: Stackelberg comparison: first-mover premium by leader.

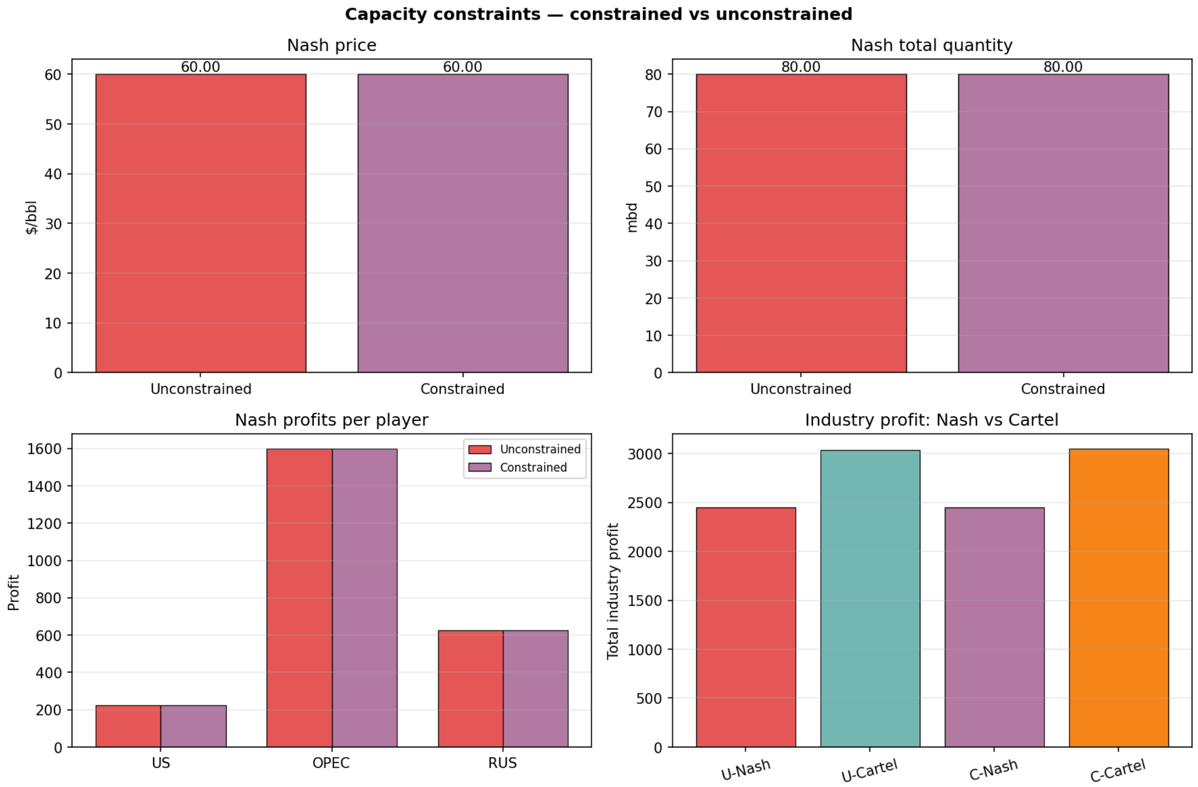


Figure 7: Constrained versus unconstrained equilibria at calibrated capacities ($\text{cap}_{\text{US}} = 30$, $\text{cap}_{\text{OPEC}} = 40$, $\text{cap}_{\text{RUS}} = 35$ mbd).

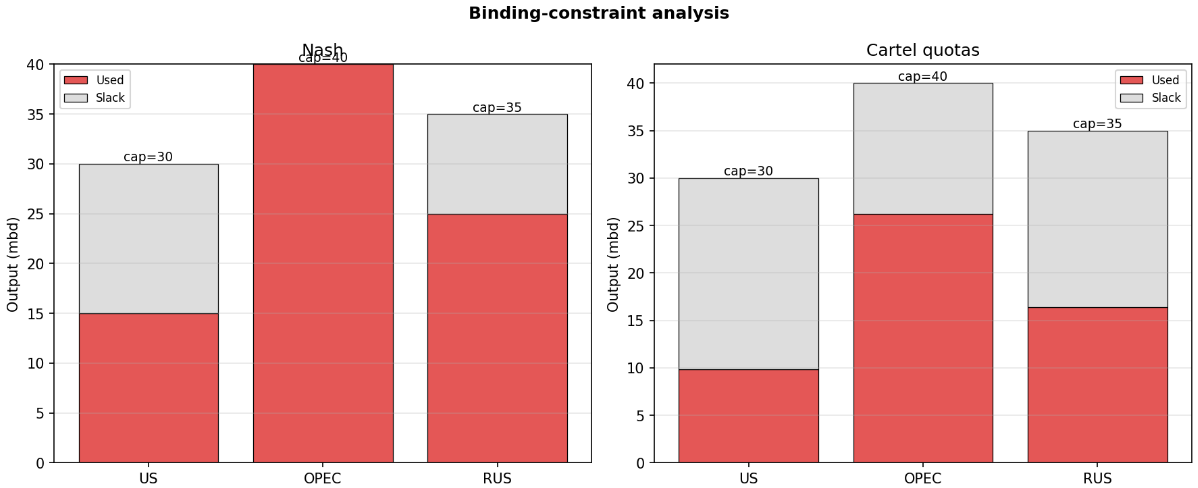


Figure 8: Capacity binding analysis: identification of the binding player at Nash and at cartel quotas.

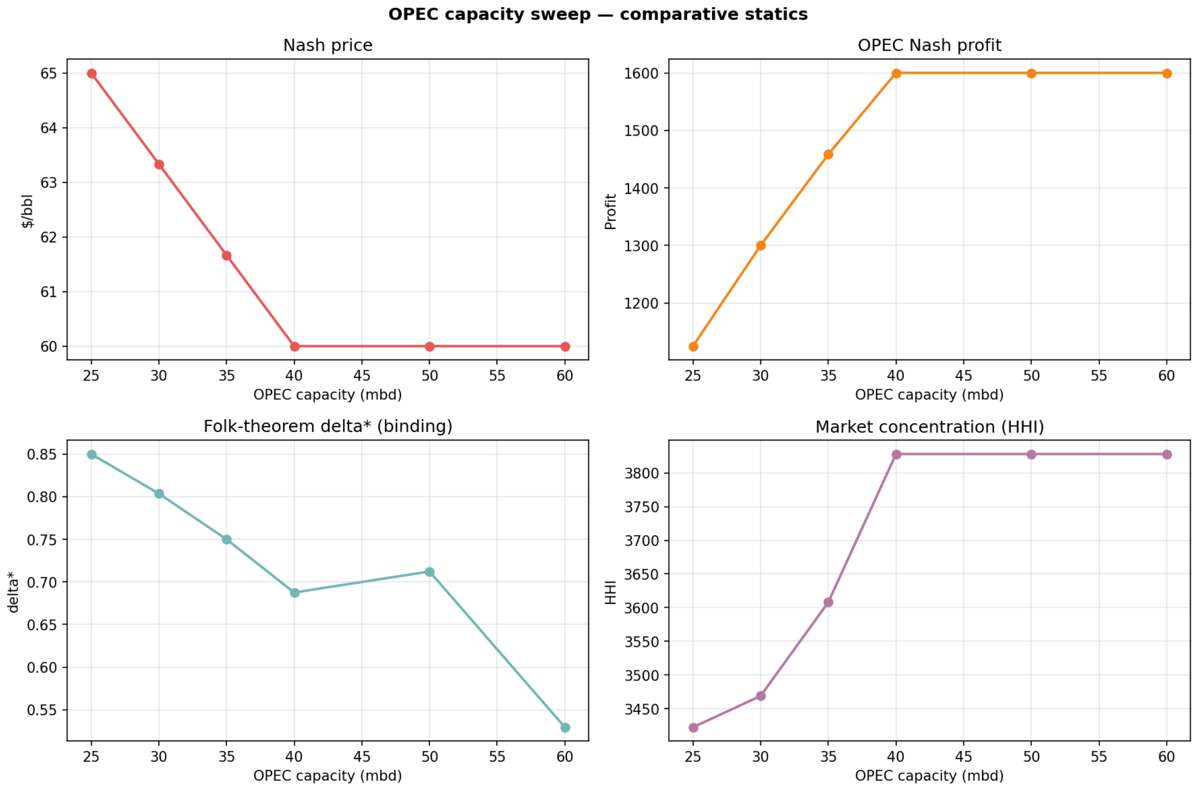


Figure 9: OPEC capacity sweep: equilibrium price, output, and per-player profit as cap_{OPEC} varies from 25 to 60 mbd.

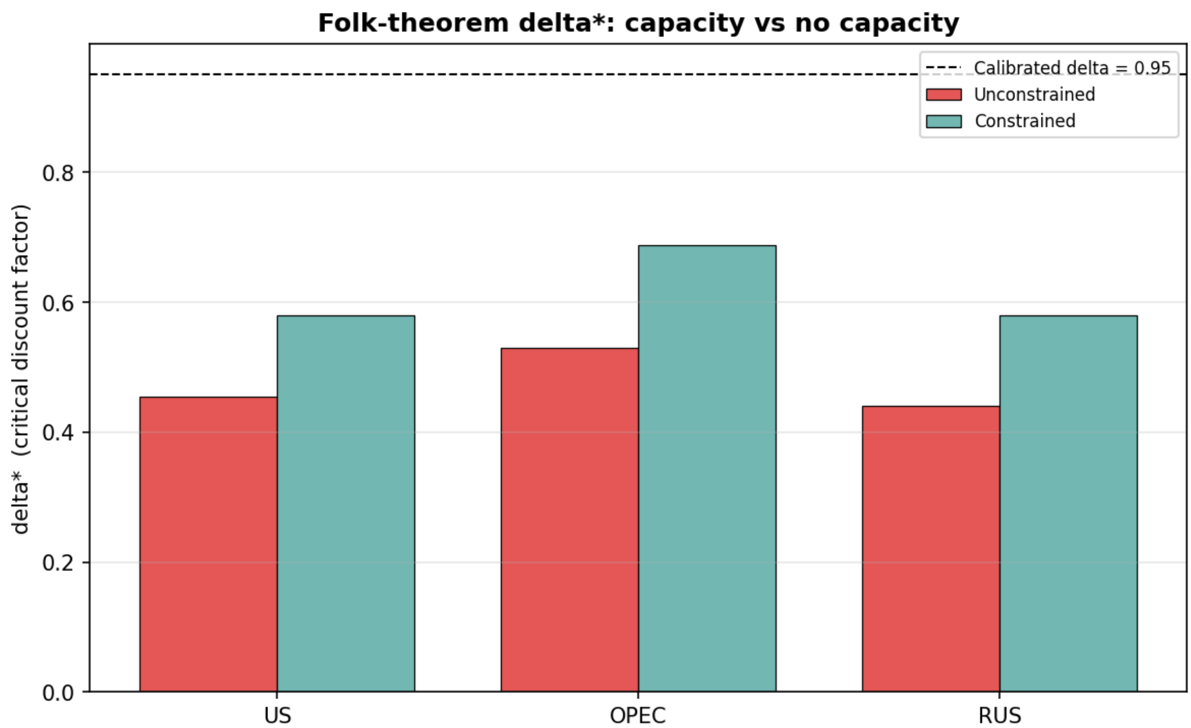


Figure 10: Capacity and the Folk Theorem: critical discount factor δ^* as a function of OPEC capacity.

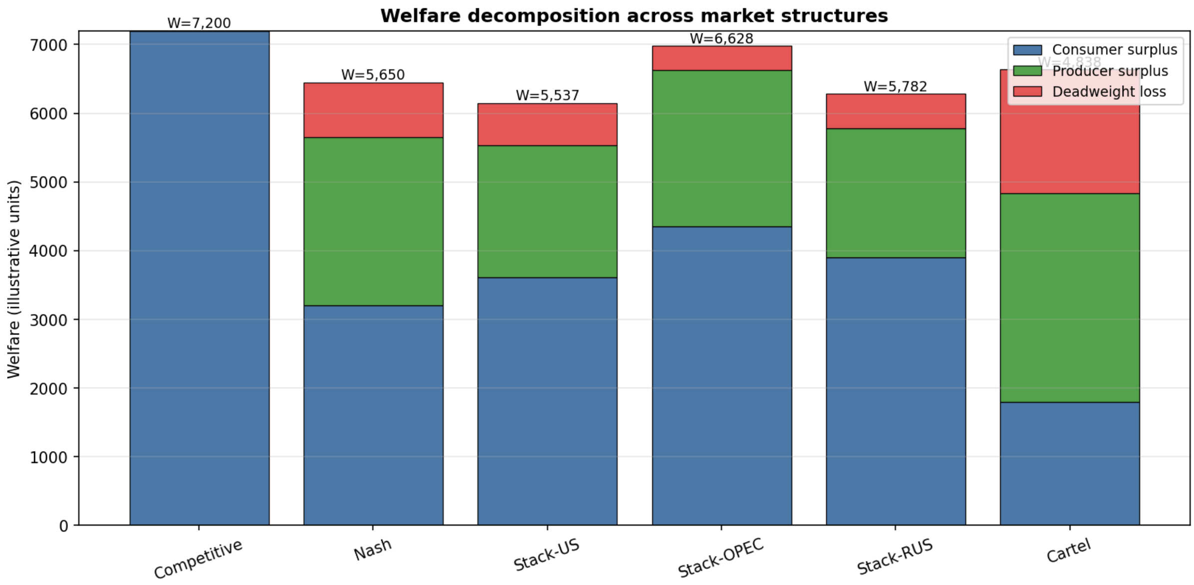


Figure 11: Welfare decomposition across the six benchmark market structures: competitive, Nash, Stackelberg by leader, and cartel.

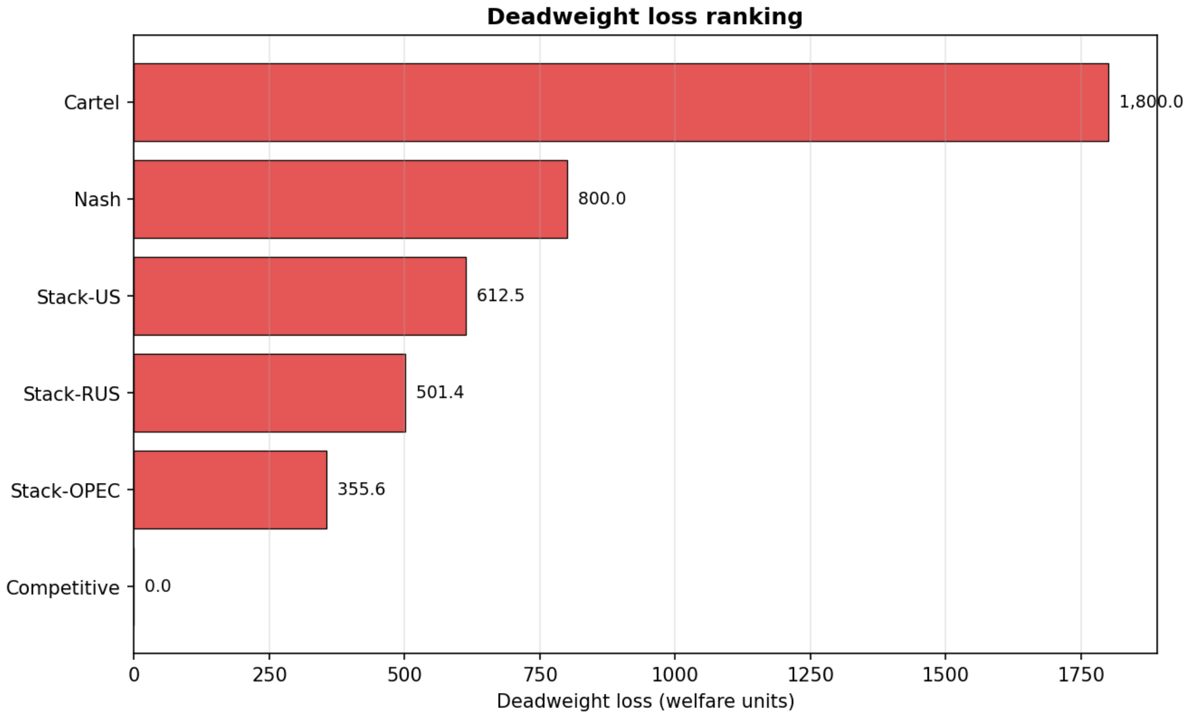


Figure 12: Deadweight loss comparison across market structures, ranking from competitive (0) to cartel (1,800).

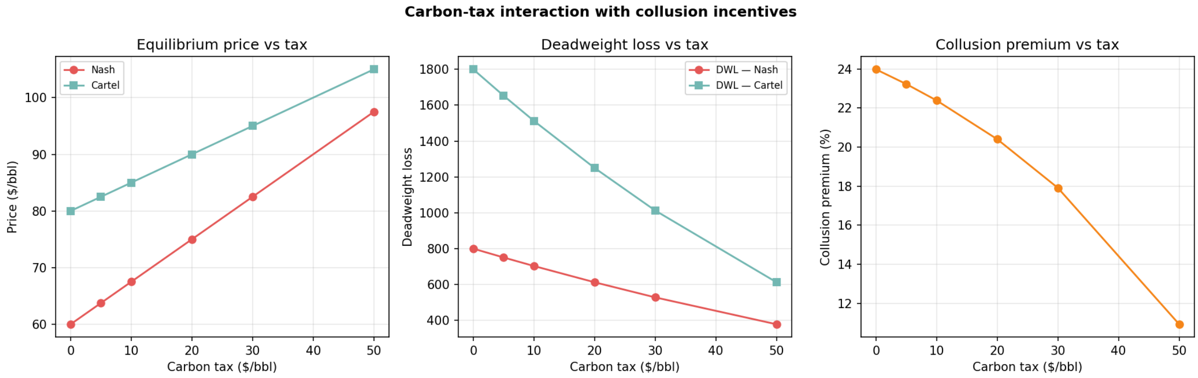


Figure 13: Carbon tax interaction: collusion premium as a function of per-barrel Pigouvian tax $\tau \in \{0, 10, 20, 30, 40, 50\}$.

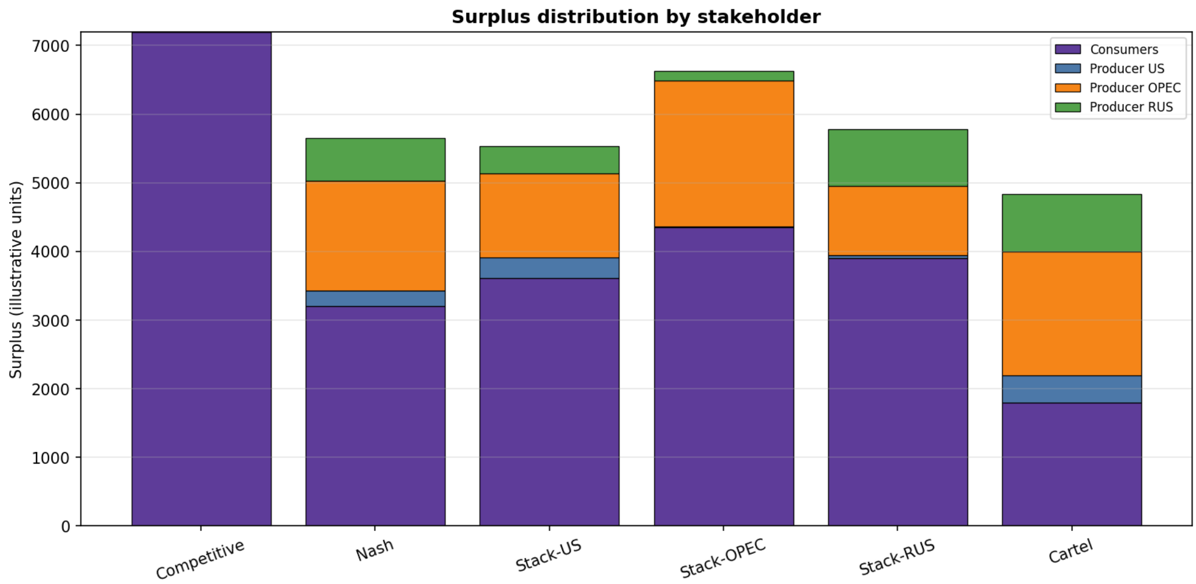


Figure 14: Surplus distribution: consumer surplus, producer surplus by player, and deadweight loss across market structures.

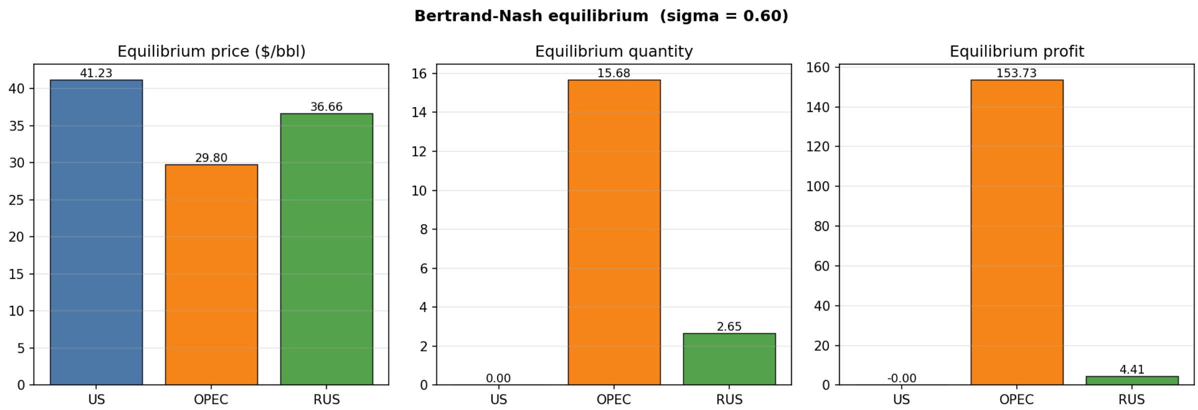


Figure 15: Differentiated Bertrand-Nash equilibrium at $\sigma = 0.6$.

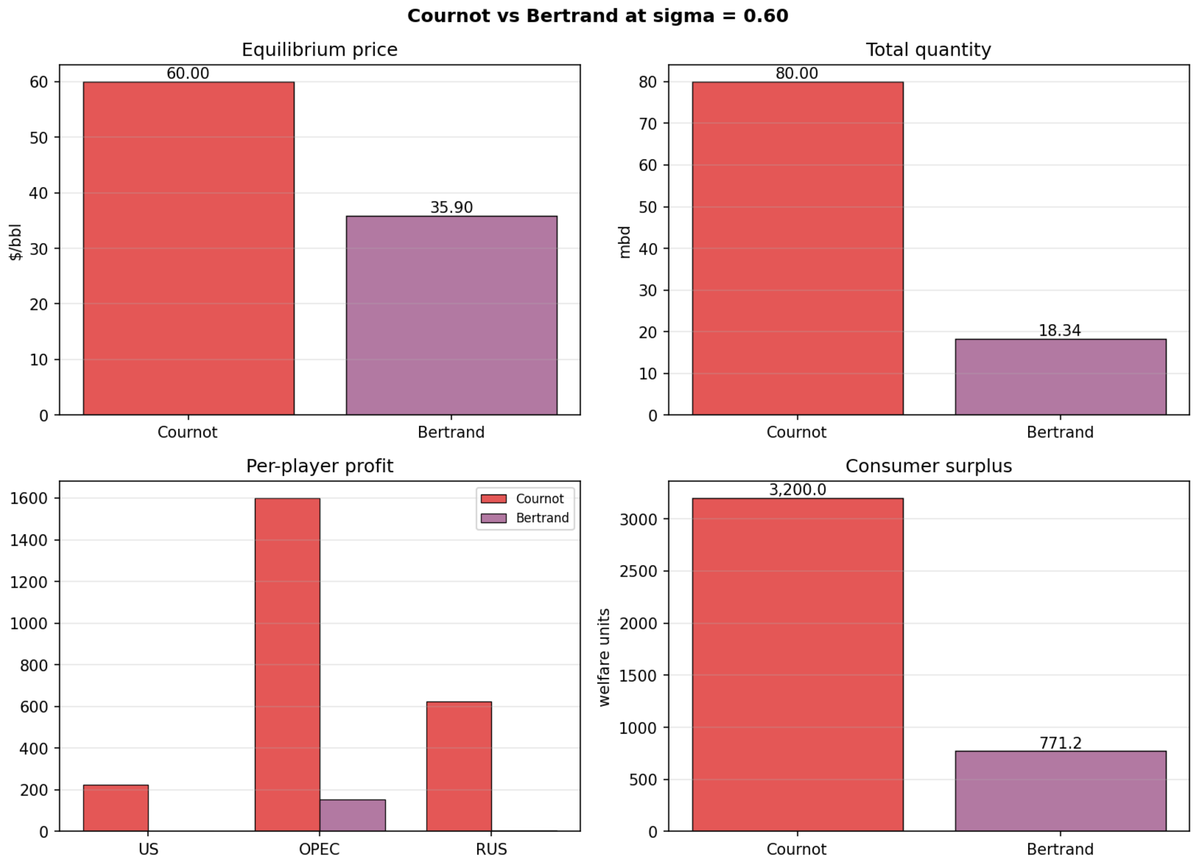


Figure 16: Cournot versus Bertrand: side-by-side comparison of prices, quantities, profits, and consumer surplus.

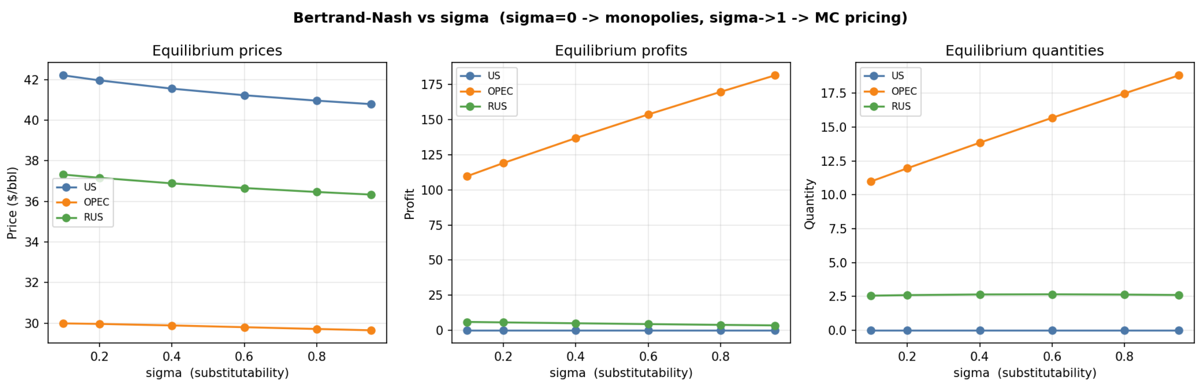


Figure 17: Substitutability sweep: equilibrium quantities as σ varies from 0 to 1; convergence toward marginal-cost pricing.

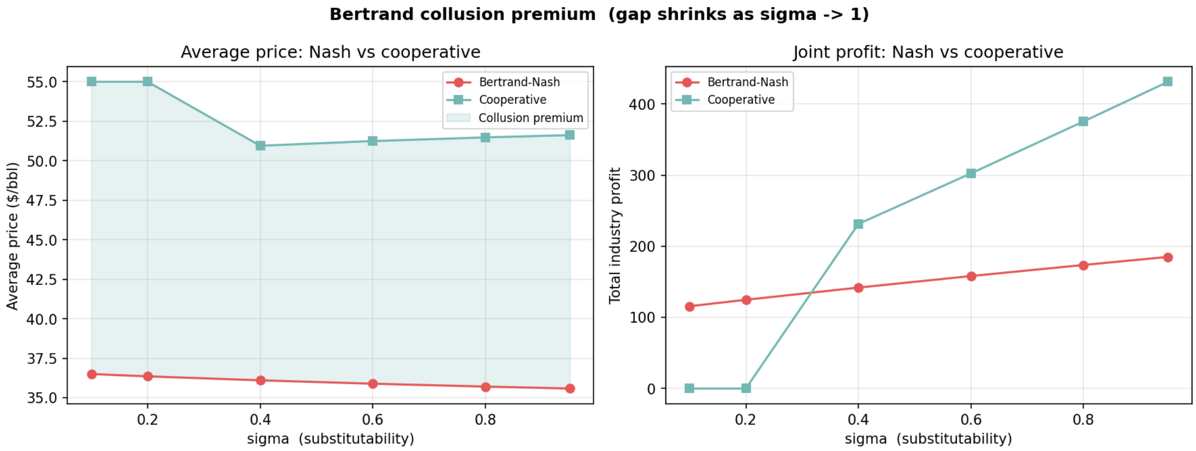


Figure 18: Cooperation gap under differentiated Bertrand: average price under Bertrand-Nash versus cooperative Bertrand.

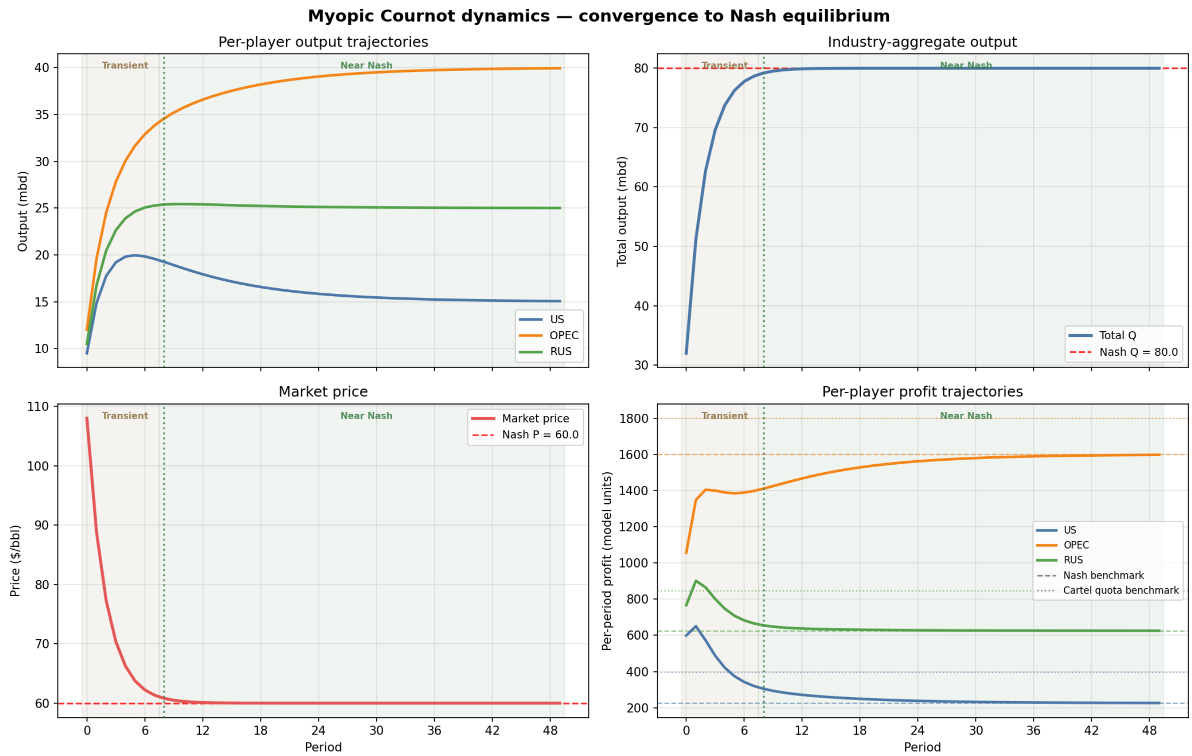


Figure 19: Repeated-game myopic best-response: prices, total quantity, per-player quantities, and per-player profits converging to the static Nash benchmark at $\lambda = 0.2$.

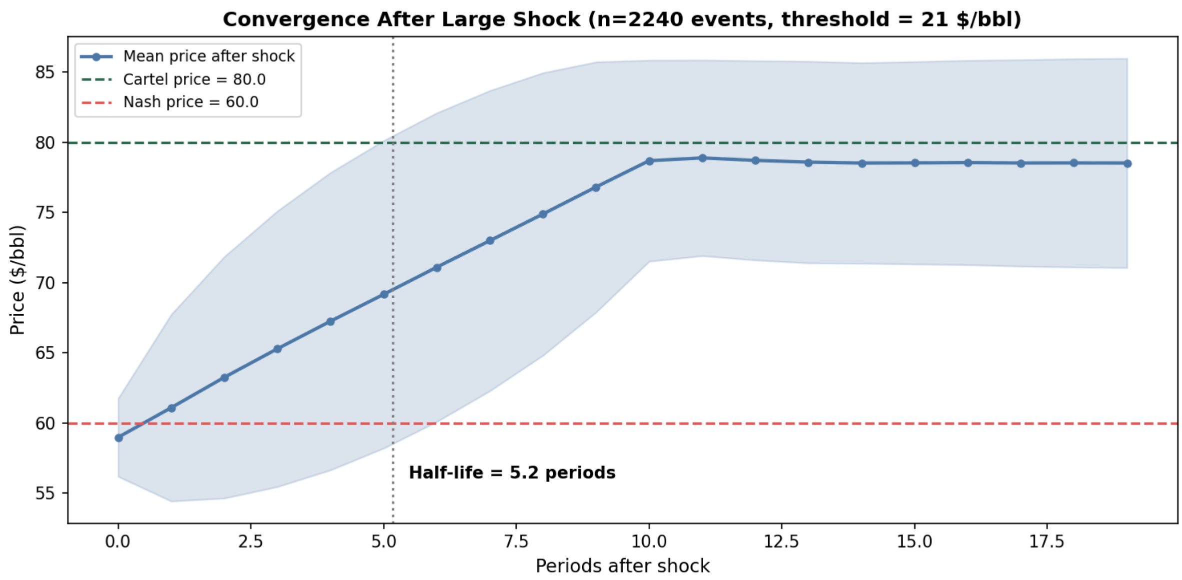


Figure 20: Convergence after a large adverse demand shock: mean recovery path with 80 percent confidence band.

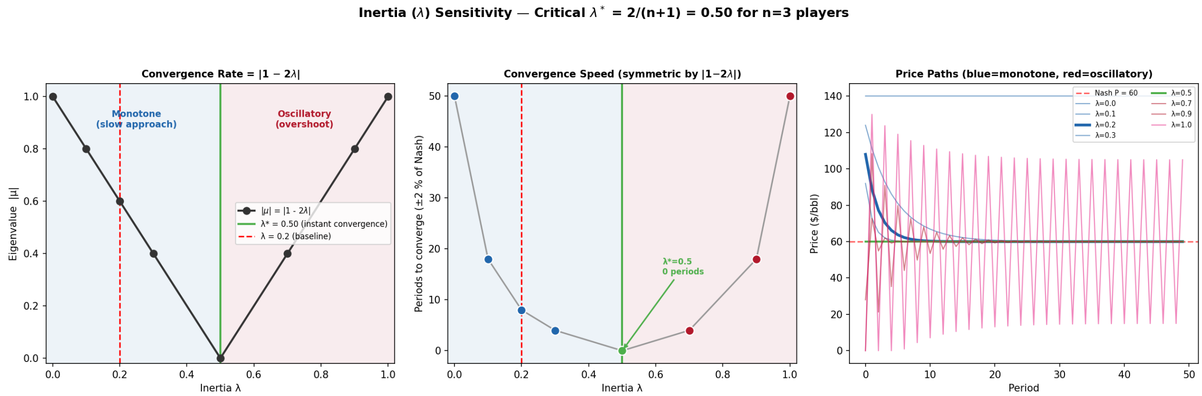


Figure 21: λ-sensitivity analysis of myopic best-response dynamics: monotone convergence for $\lambda \in (0, 0.5)$, oscillatory for $\lambda \in (0.5, 1)$.

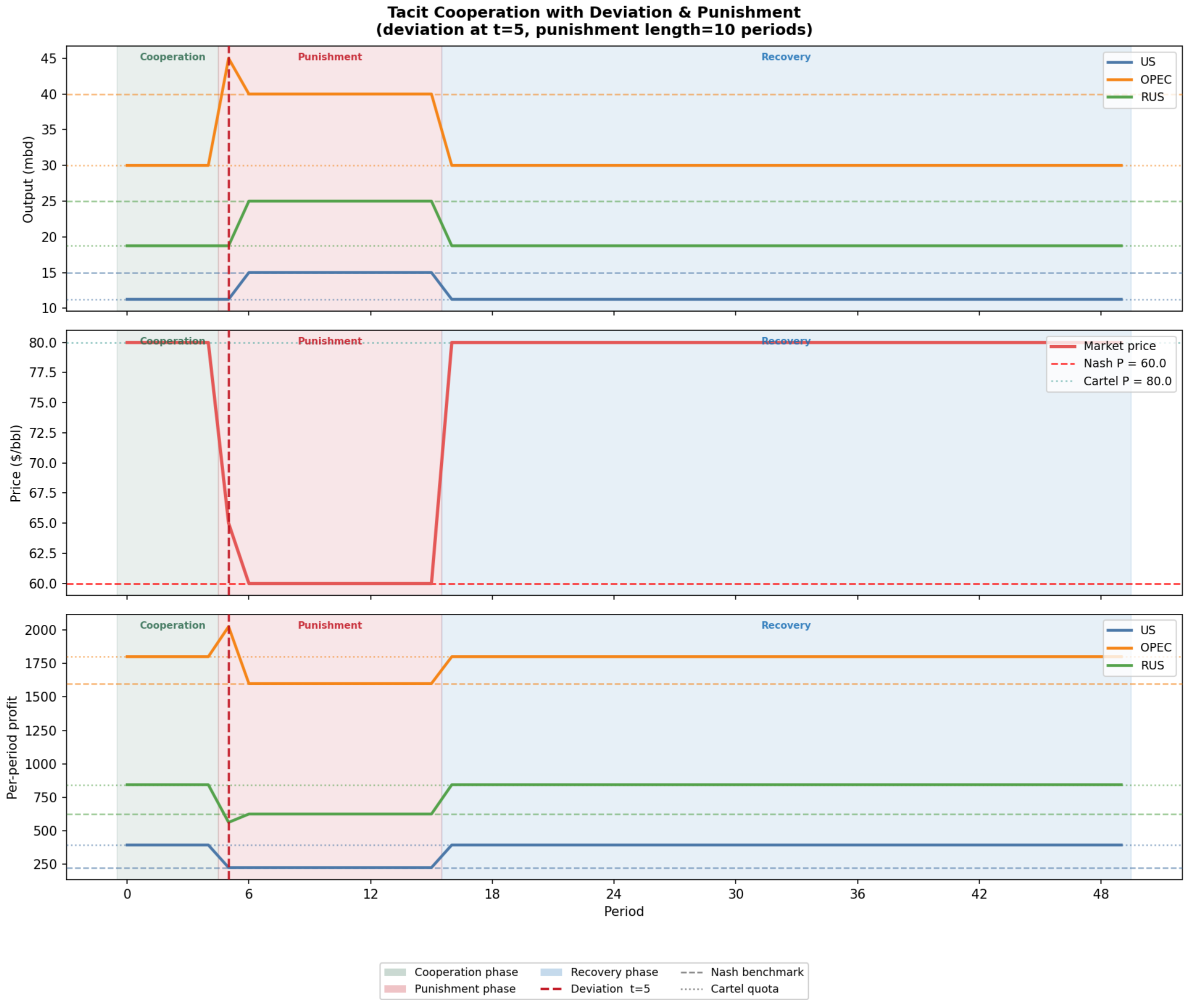


Figure 22: Punishment regime simulation: cooperate, deviate, be punished, recover.

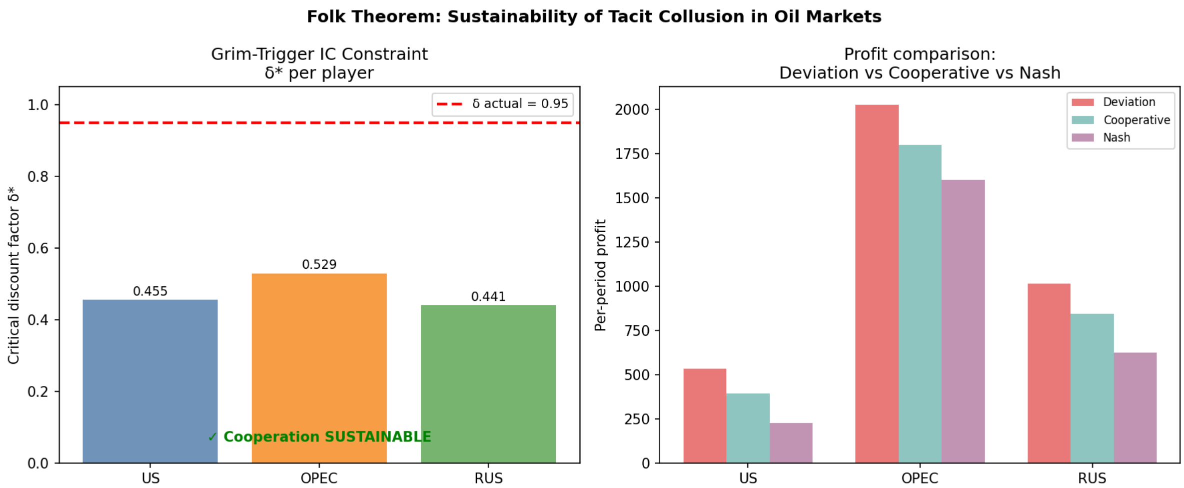


Figure 23: Folk Theorem: critical discount factors δ_i^* by player.

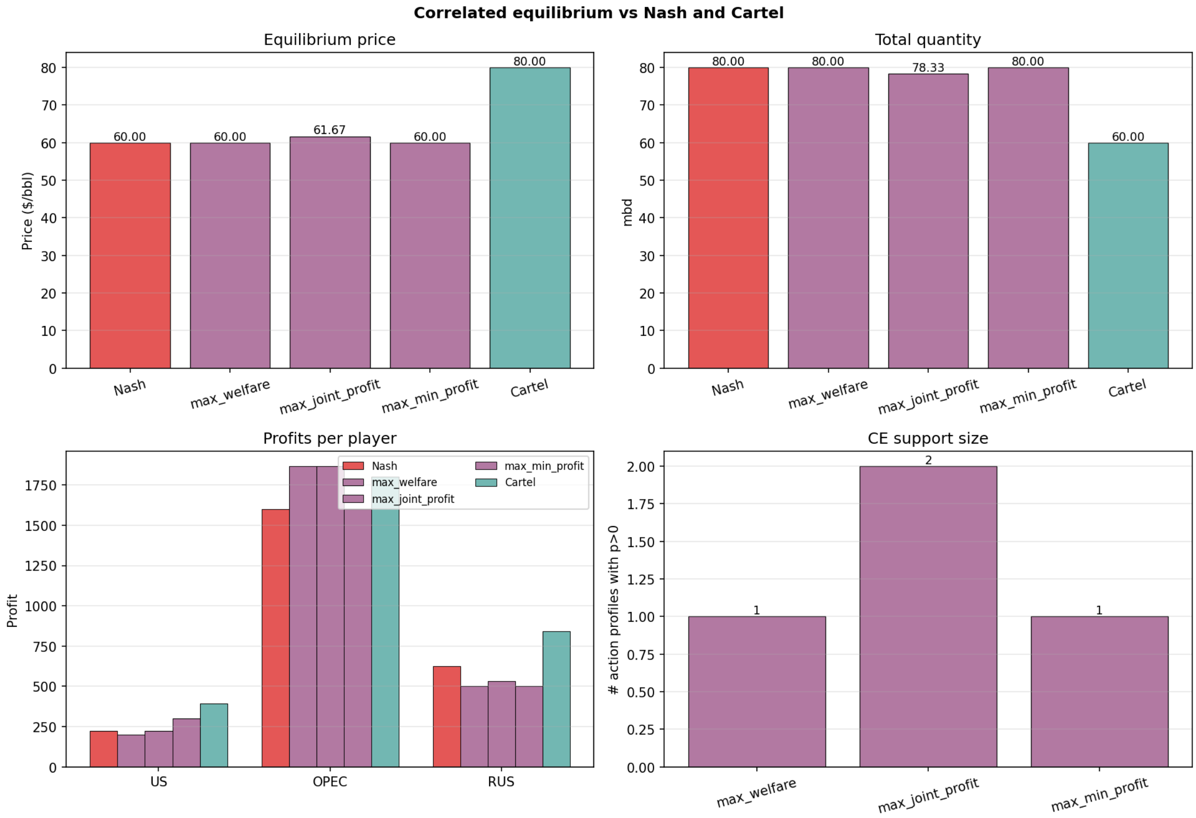


Figure 24: Correlated equilibrium versus Nash and cartel: price and per-player profit for the three CE objectives (max-welfare, max-joint-profit, max-min).

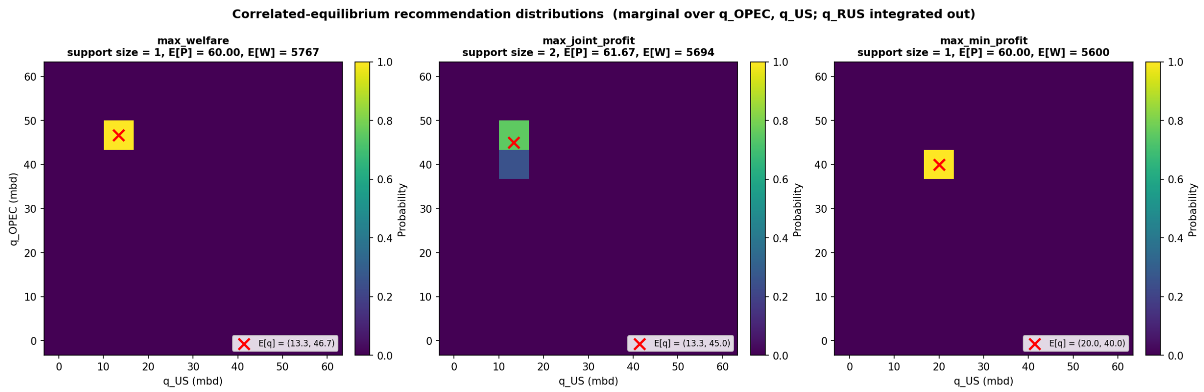


Figure 25: Support of the correlated equilibrium: action profiles recommended by the mediator and their probability weights, across the three objectives.

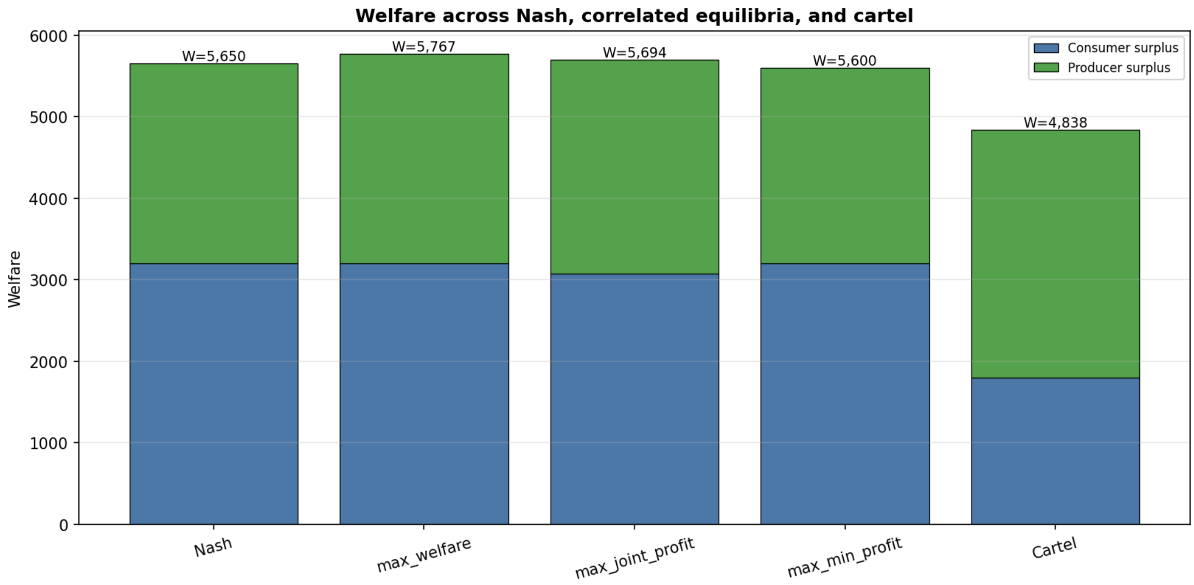


Figure 26: Welfare ranking of the three correlated equilibria against Nash and cartel benchmarks.

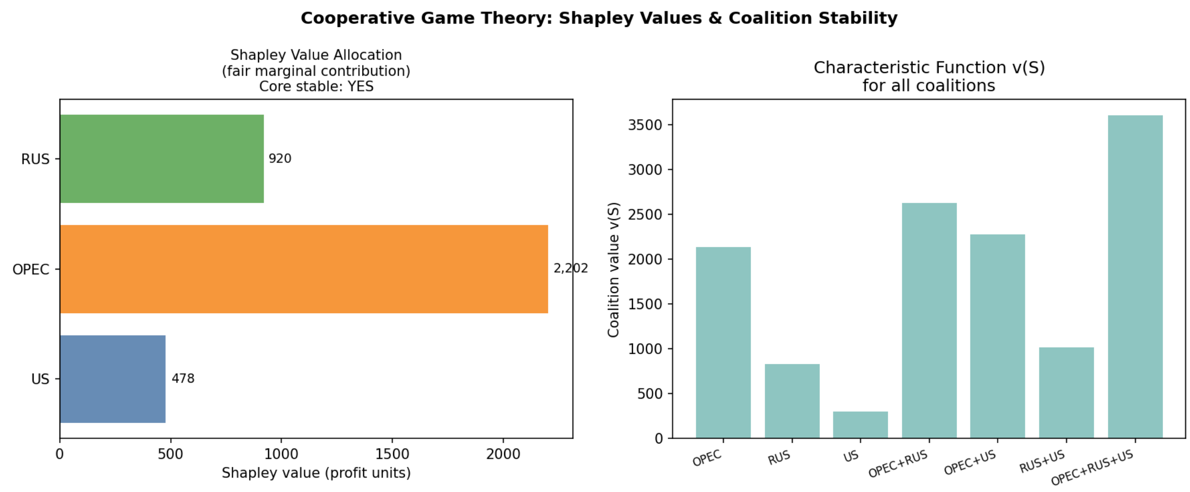


Figure 27: Shapley value decomposition for the OPEC–US–Russia triopoly.

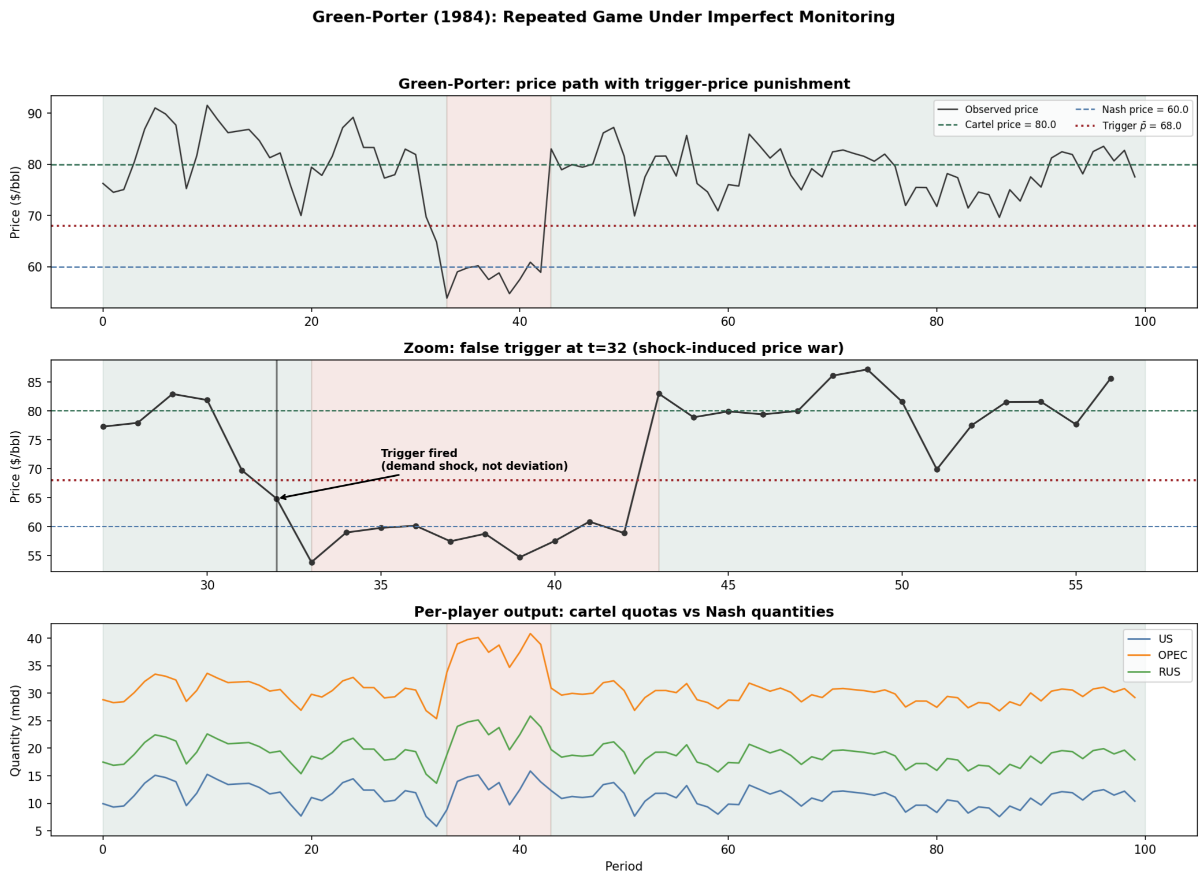


Figure 28: Green–Porter cooperation regimes under $AR(1)$ and jump-diffusion demand.

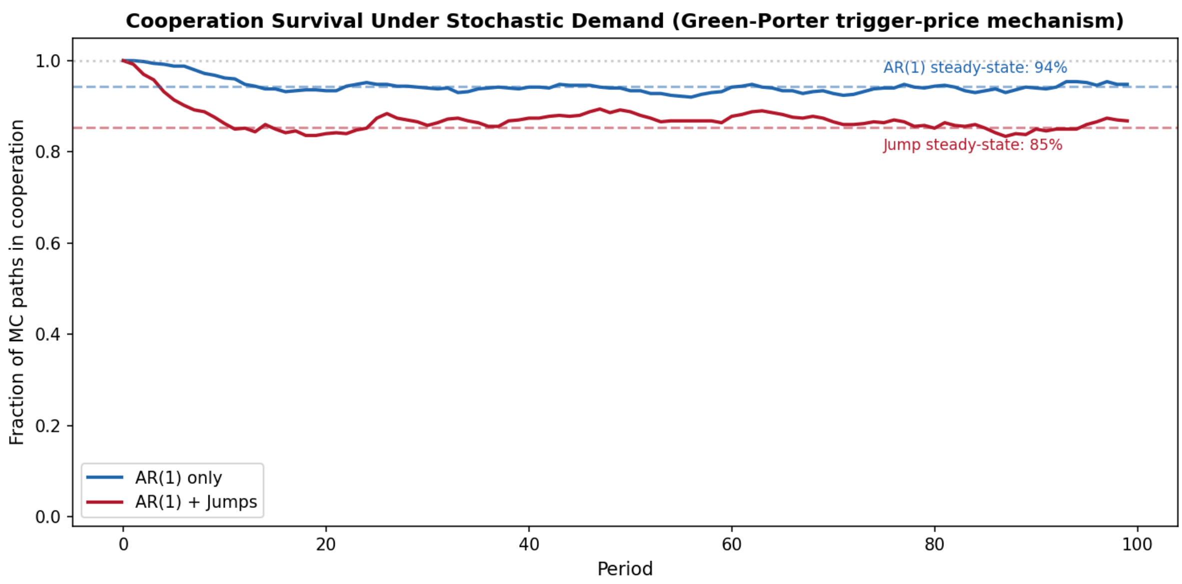


Figure 29: Green–Porter cooperation survival probabilities across shock specifications.

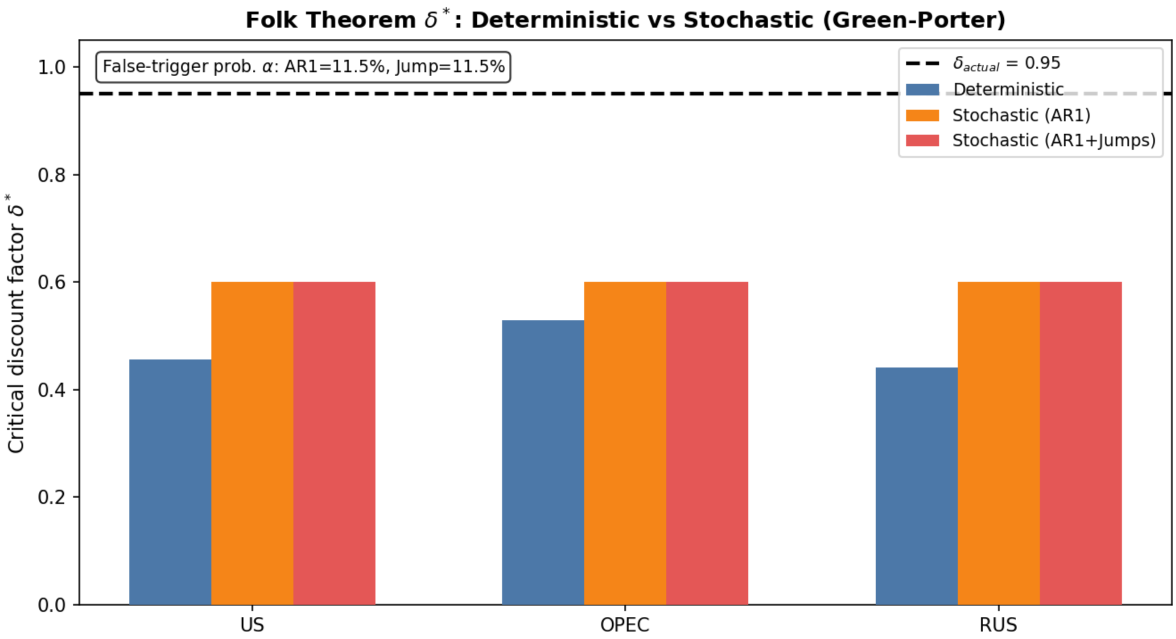


Figure 30: Critical discount factor δ^* under deterministic and stochastic demand: side-by-side comparison.

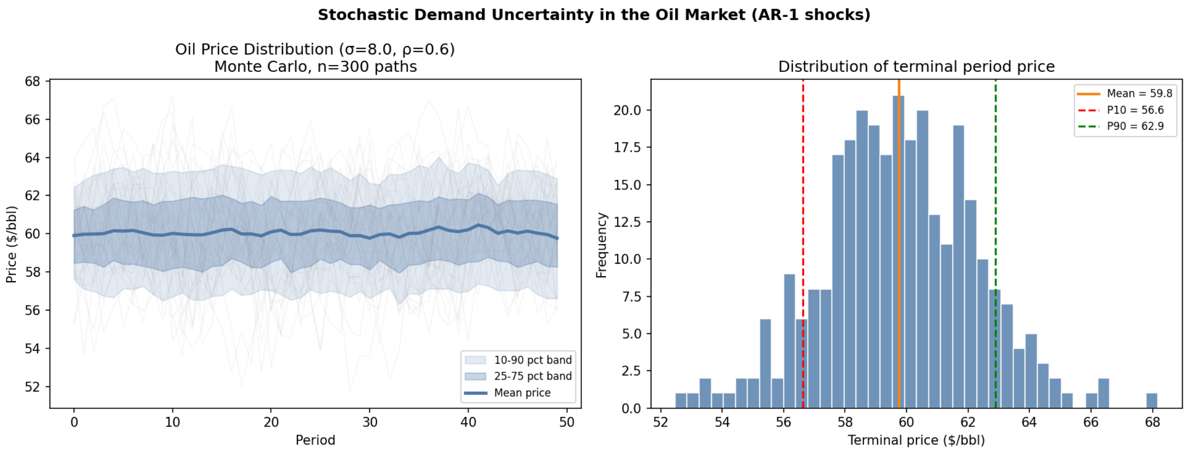


Figure 31: Monte Carlo price bands (P10, median, P90) under $AR(1)$ demand shocks.

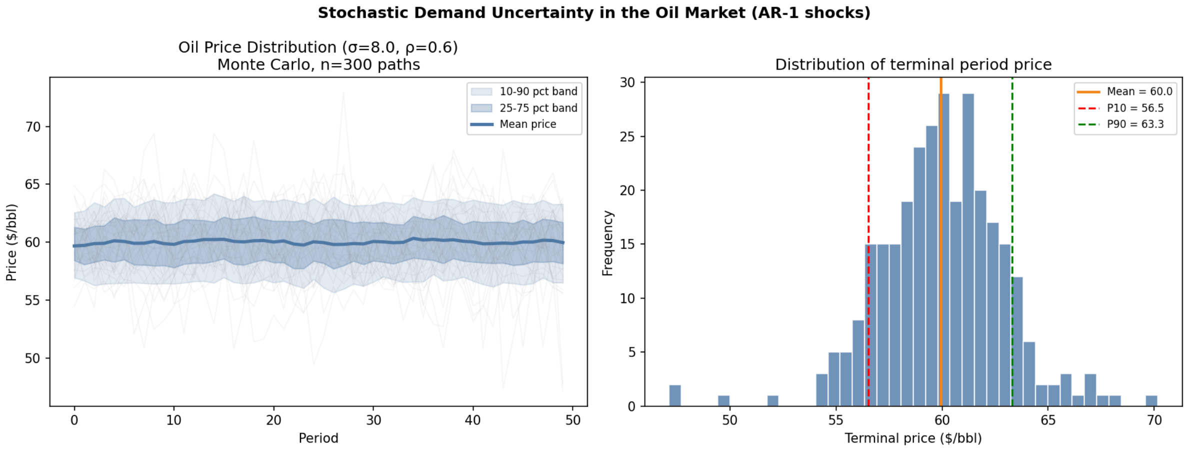


Figure 32: Merton (1976) jump-diffusion demand specification: simulated price paths with rare discrete jumps.

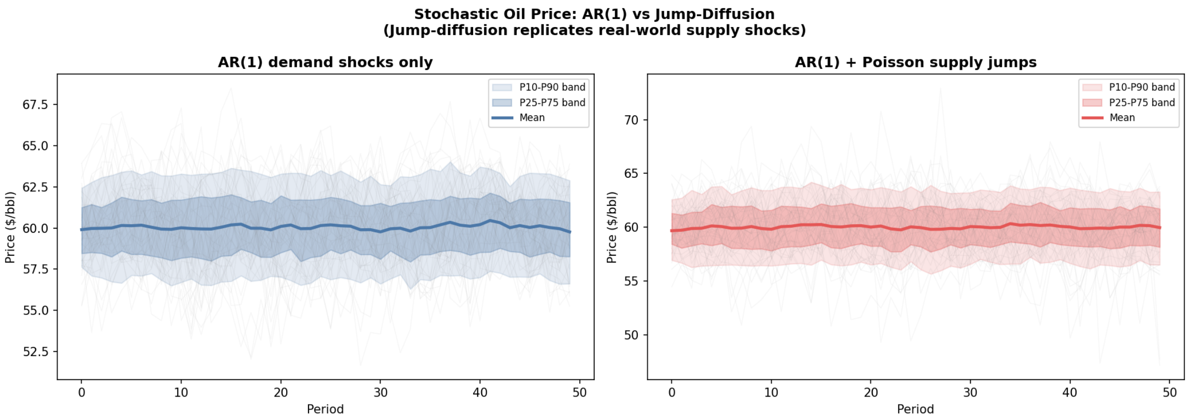


Figure 33: Deterministic versus stochastic equilibrium: distribution of terminal price across 300 Monte Carlo paths.

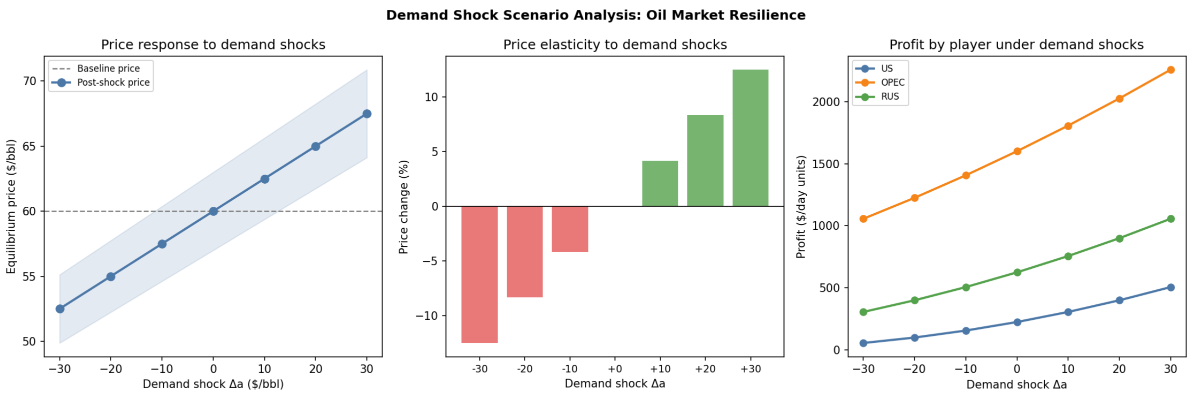


Figure 34: Demand shock scenarios: equilibrium price, output, and per-player profit at $\delta_a \in \{-30, -20, -10, 0, +10, +20, +30\}$.

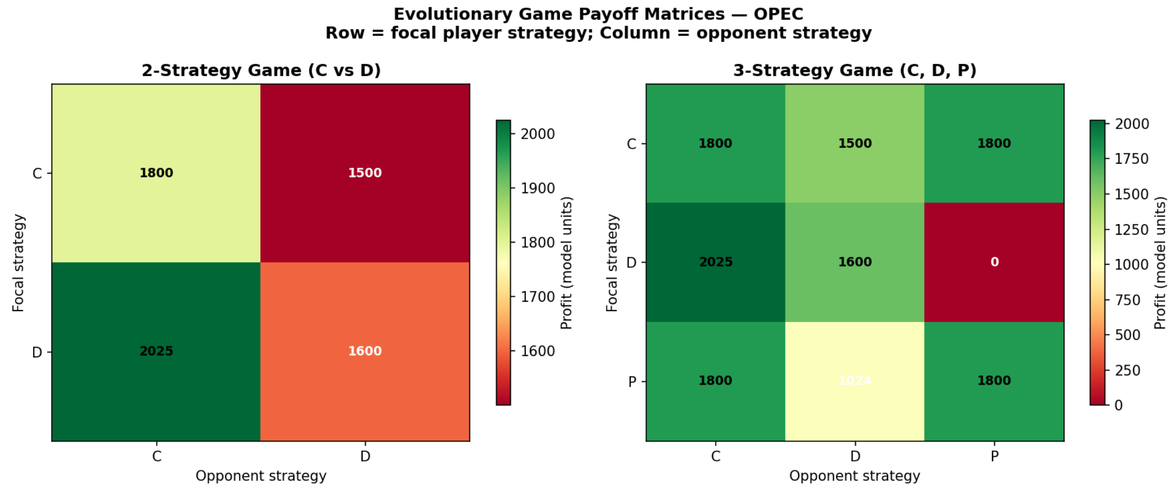


Figure 35: Evolutionary game theory: stage-game payoff matrix for the Cooperate, Defect, and Punish strategies.

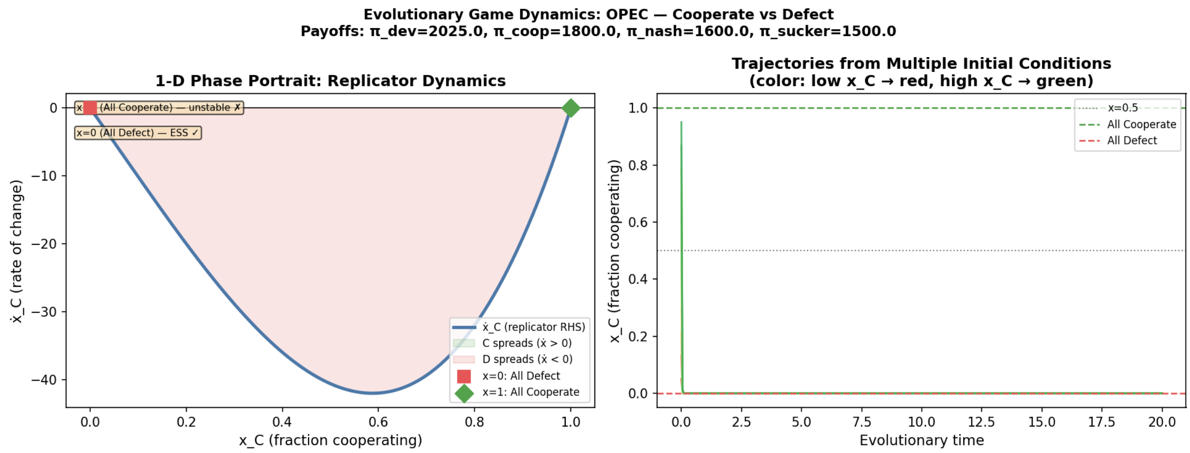


Figure 36: Two-strategy replicator dynamics (Cooperate vs. Defect): phase line and ESS.

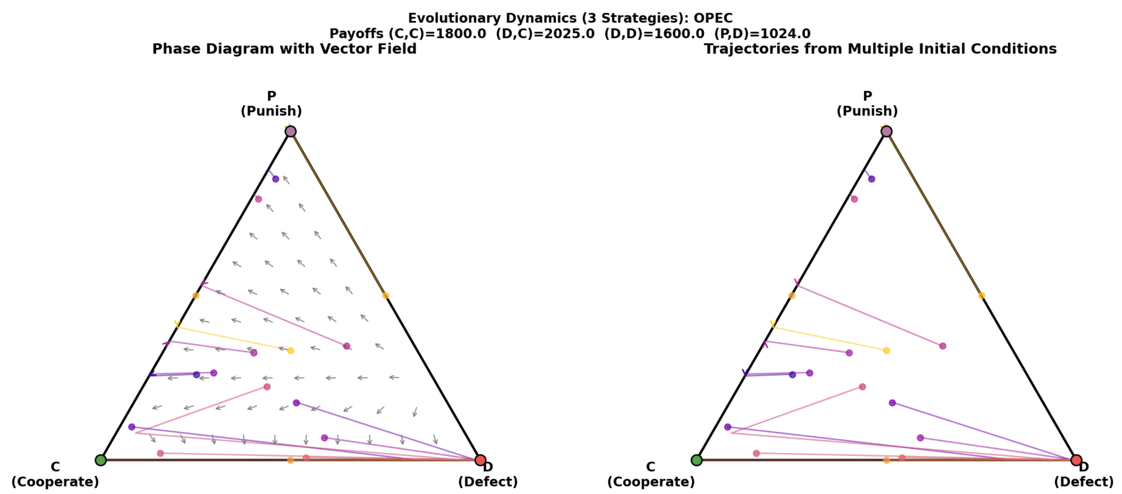


Figure 37: Three-strategy replicator dynamics (Cooperate, Defect, Punish): triangular phase diagram showing bistability.

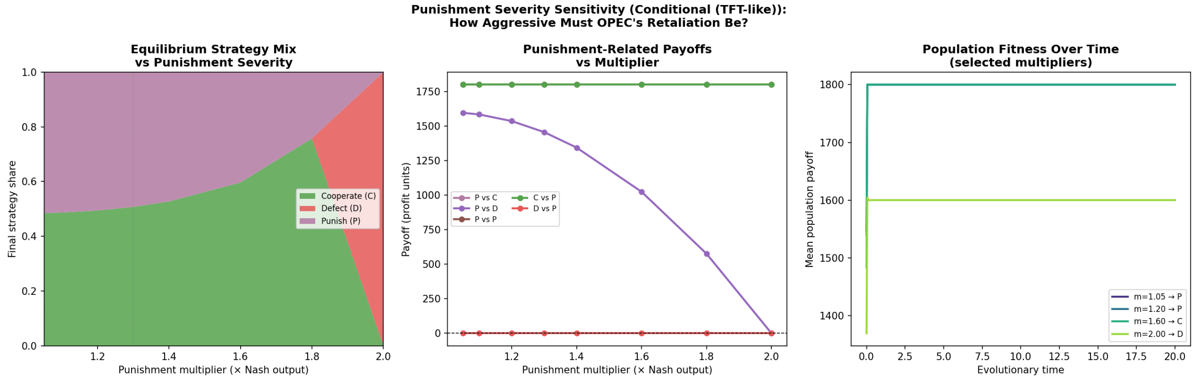


Figure 38: Punishment-severity sensitivity sweep: long-run population composition as a function of the punishment multiplier.

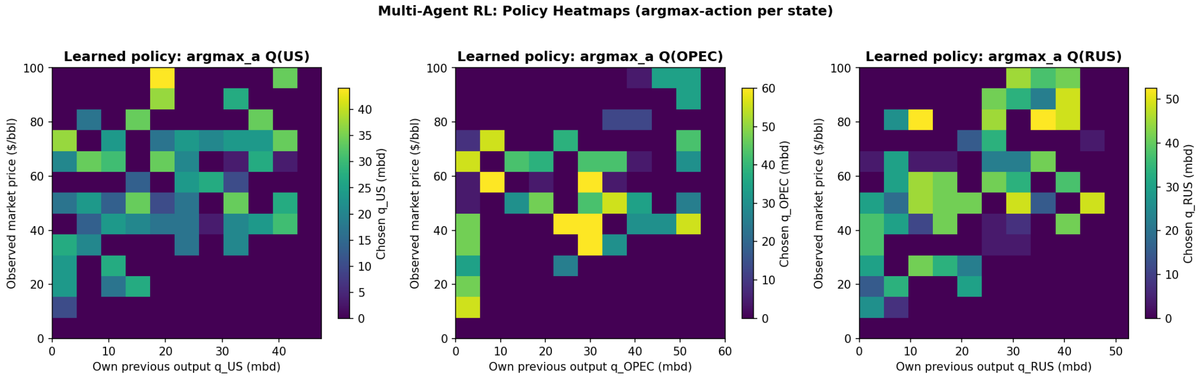


Figure 39: Greedy-policy heatmaps for the three learning agents on a representative headline seed. Colour encodes the greedy output action (mbd); blank cells were never visited during training. The absence of any systematic price-quantity gradient corroborates the Spearman finding of Section 4.5.3: converged policies are largely unresponsive to the observed price, consistent with path-dependent fixed-point lock-in.

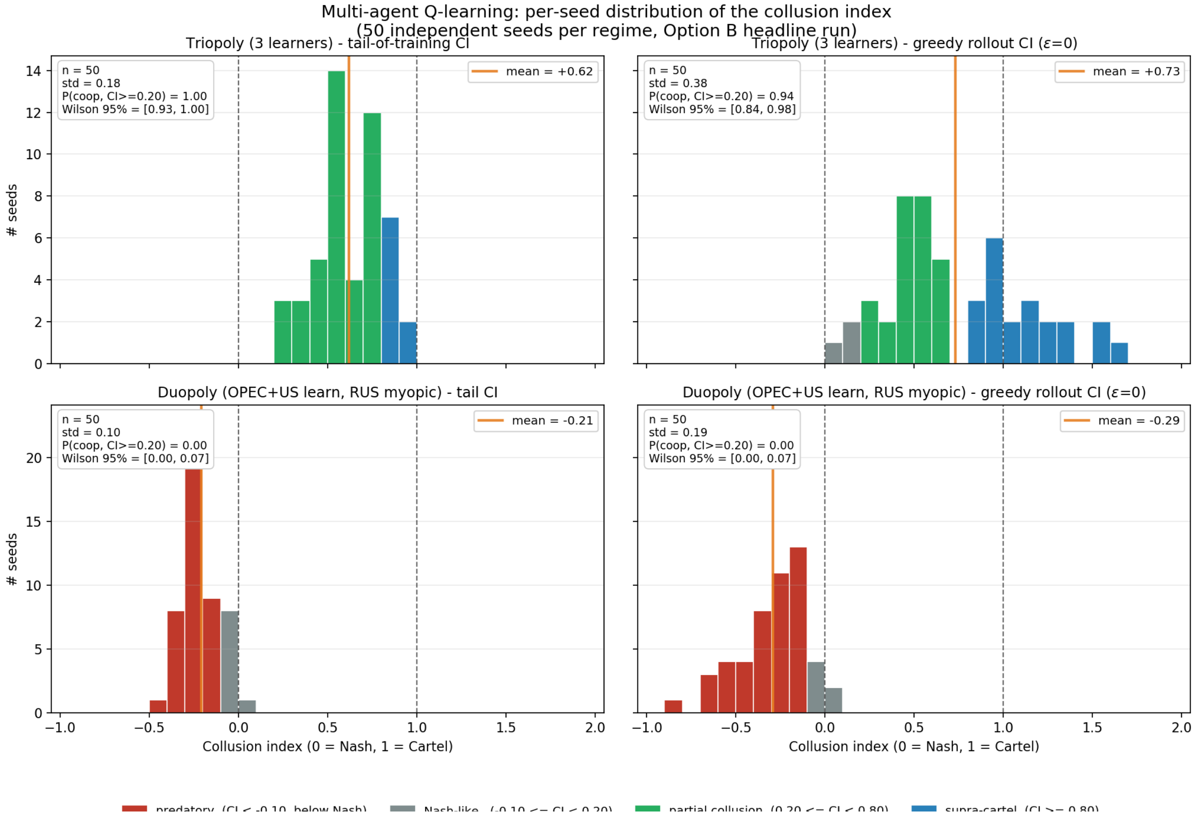


Figure 40: Per-seed distribution of the greedy collusion index in the headline triopoly ($n = 50$ seeds, top) and duopoly robustness check ($n = 50$ seeds, bottom). The bimodal structure of the triopoly distribution motivates the Wilson-proportion reformulation of Section 4.3: the appropriate summary is the proportion of seeds reaching a cooperative basin ($\overline{\mathrm{CI}} \geq 0.20$), not the cross-seed mean. The duopoly distribution collapses to a tight predatory cluster, confirming the regime contrast documented in Section 4.4.

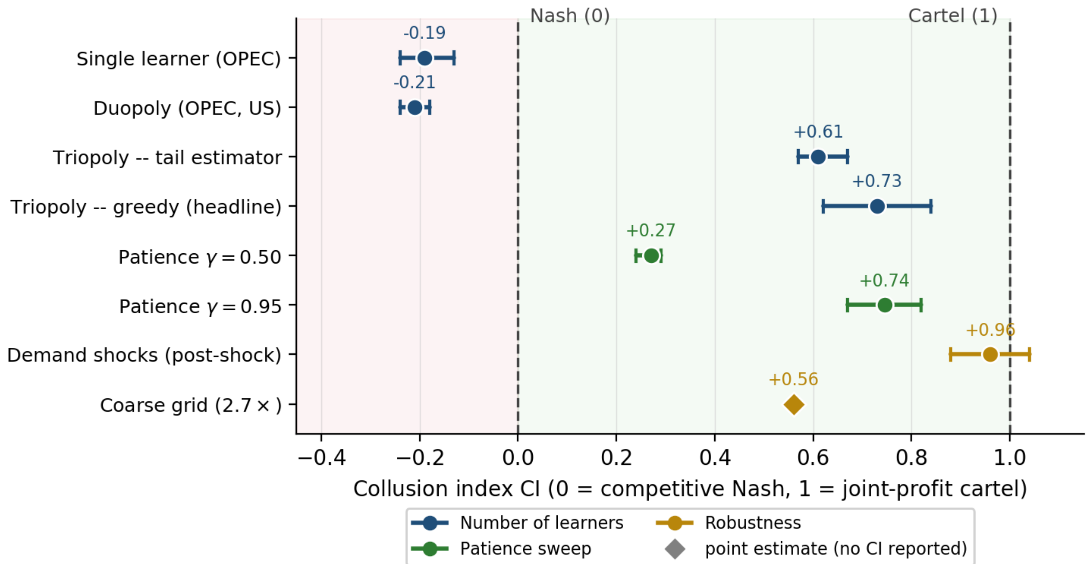


Figure 41: Collusion index across all Section 4 regimes: point estimate and 95% confidence interval (Wilson or normal-approximation as reported in the corresponding table), anchored to $\overline{\mathrm{CI}} = 0$ (static Nash) and $\overline{\mathrm{CI}} = 1$ (joint-profit cartel). Single-learner and duopoly regimes sit in predatory territory below Nash; every multi-learner and robustness regime sits in the cooperative half-space. The coarse-grid point carries no reported interval.

Tables

Table 1: Green-Porter steady-state statistics under $AR(1)$ and jump-diffusion demand shocks.

Statistic	$AR(1)$ **shocks**	**Jump-diffusion**
Cooperation fraction	94 %	85 %
Mean cooperative spell	61.8 qtrs (∼15 yrs)	39.1 qtrs (∼10 yrs)
Mean punishment phase	9.6 qtrs (∼2.4 yrs)	9.5 qtrs (∼2.4 yrs)
Triggers (300 paths × 200 qtrs)	291	685

Table 2: Headline triopoly: central statistics across 50 independent seeds. The greedy estimator freezes Q-tables and sets $\varepsilon = 0$; the tail estimator averages the final 20% of training at $\varepsilon = 0.05$. The gap between the two is a diagnostic of residual exploration at the end of training.

Estimator	Mean $\overline{\text{CI}}$	Std. dev.	95% CI (mean)	Mean $\bar{P}$ ($/bbl)
Tail of training	0.61	0.18	[0.57, 0.67]	68.41
Greedy rollout	0.73	0.39	[0.62, 0.84]	72.63
Random policy (baseline)	0.04	—	[−0.04, 0.12]	44.20

Table 3: Basin classification of the 50 headline seeds by greedy collusion index. Seed counts and share of total are reported for each qualitatively distinct attractor.

Basin	Greedy $\overline{\text{CI}}$ range	Seeds (n)	Qualitative label
Predatory	$\overline{\text{CI}} < 0$	3	Below-Nash pricing
Competitive	$0 \leq \overline{\text{CI}} < 0.20$	0	Near-Nash
Cooperative	$0.20 \leq \overline{\text{CI}} < 0.80$	12	Tacit cooperation
Near-cartel	$\overline{\text{CI}} \geq 0.80$	35	Supra-cooperative

Table 4: Wilson score intervals for the proportion of seeds reaching a cooperative basin ($\overline{\text{CI}} \geq 0.20$), by regime and estimator. Non-overlapping intervals imply a strict separation between triopoly and duopoly at the 5% level.

Regime	$\hat{p}$ ($\overline{\text{CI}} \geq 0.20$)	Wilson 95% interval
Triopoly (greedy)	0.94	[0.84, 0.98]
Triopoly (tail)	0.88	[0.76, 0.95]
Duopoly (greedy)	0.00	[0.00, 0.07]
Duopoly (tail)	0.02	[0.00, 0.11]

Table 5: Forced-deviation experiment: multi-seed mean price and per-player quantities in each window ($n = 20$ seeds). The strategic share is computed via Equation (9).

Window	$\bar{P}$ (\$/bbl)	$\bar{q}_{\text{OPEC}}$ (mbd)	$\bar{q}_{\text{US}}$ (mbd)	$\bar{q}_{\text{RUS}}$ (mbd)	$\bar{Q}$ (mbd)
Pre-window	79.75	24.98	16.92	18.35	60.25
During	64.15	38.57	18.10	19.17	75.84
Post-window	78.92	25.31	17.05	18.40	60.76
Δ (during − pre)	−15.60	+13.59	+1.18	+0.82	+15.59

Table 6: Punishment-cycle detection: per-seed results at the full headline budget (75,000 steps, $n = 5$ seeds). The pre-registered frequency band is [0.2, 2.0] cycles per 1,000 steps.

Seed	Cycles detected	Freq. (per 1,000 steps)	Median drop (\$/bbl)	In band?
301	289	3.85	12.14	No
302	84	1.12	11.02	**Yes**
303	34	0.45	10.87	**Yes**
304	255	3.40	11.93	No
305	291	3.88	11.89	No
Median	—	—	11.57	2/5

Table 7: Policy-level Spearman ρ between greedy output action and price bin, by seed and agent ($n = 5$ seeds, full headline budget).

Seed	OPEC	US	RUS	Row median
301	0.04	0.01	−0.02	0.01
302	0.07	0.03	0.01	0.03
303	−0.01	0.05	0.02	0.02
304	0.02	−0.03	0.04	0.02
305	0.06	0.00	0.01	0.01
Column median	0.04	0.01	0.01	**0.02**

Table 8: Realised-trajectory Spearman ρ between price at step t and agent output at step $t+1$, post-convergence window ($n = 5$ seeds).

Seed	OPEC	US	RUS	Row median
301	0.11	0.06	0.04	0.06
302	0.14	0.09	−0.03	0.09
303	0.03	0.08	0.10	0.08
304	0.05	−0.02	0.12	0.05
305	0.09	0.07	0.06	0.07
Column median	0.09	0.07	0.06	**0.07**

Table 9: Demand-shock experiment: window-level scalars aggregated across 20 observations (10 seeds $\times$ 2 shocks). σ_P is computed from the pre-shock rolling window.

Diagnostic	Median	Pre-registered threshold
In-window mean drop (σ_P units)	2.73	≥ 2.00
Trough drop (σ_P units)	5.78	—
Recovery half-life (steps)	11	$\leq 15{,}000$
Post-shock converged $\overline{\text{CI}}$	0.96	—

Table 10: Verdict matrix: all eight pre-registered decision rules applied verbatim to the Option B evidence. Secondary hypotheses Q5–Q8 denote the robustness checks of Section 4.6. ✓ met ● partial ✗ rejected

Label	Pre-registered criterion	Verdict	Section
H1	Greedy $\overline{\text{CI}}$ lower bound ≥ 0.30	✓ **Met**	4.2
H2	Learner-count ordering, non-overlapping CIs	✓ **Met**	4.7
H3	Spearman $\rho(\gamma, \overline{\text{CI}}) > 0$, disjoint extremes	✓ **Met**	4.6.1
H4a	Strategic share $\geq 30\%$	✗ **Not met**	4.5.1
H4b	Cycle frequency in $[0.2, 2.0]$, $\geq 3/5$ seeds	✗ **Not met**	4.5.2
H4c	Spearman median $\rho \geq 0.30$	✗ **Not met**	4.5.3
H5	Leader quantity within $\pm 20\%$ of q_L^*	● **Partial**	4.6.4
Q6	Cap on/off: CIs non-overlapping	● Not met (sweep: **Met**)	4.6.5
Q7	α-sweep: greedy CI > 0.30 at all cells	✓ **Met**	4.6.2
Q8	Shock recovery within $4\times$ shock width	✓ **Met**	4.6.3

Table 11: Cross-model ranking of market structures from the most competitive (lowest $\bar{P}$) to the most collusive (highest $\bar{P}$). All prices are in \$/bbl; the corresponding collusion index is reported in the final column.

Rank	Market structure	$\bar{P}$	$\overline{\text{CI}}$	Section
1	Perfect competition	35.0	–	2.1
2	Single Q-learner (predatory)	56.3	−0.19	4.7
3	Cournot–Nash	60.0	0.00	2.2
4	Correlated equilibrium	66.4	0.32	3.3
5	MARL triopoly (greedy)	72.6	0.73	4.2
6	Cartel (joint-profit max.)	80.0	1.00	3.2

A Calibrated parameters of the linear-Cournot triopoly

Calibrated parameters of the linear-Cournot triopoly, grouped by analytical block. Each row reports the parameter, its symbol, its calibrated value, and the rationale and source linking the value to the empirical literature or to a deliberate modelling choice.

Parameter	Symbol	Value	Rationale and source
A. Demand			
Demand intercept	a	\$140/bbl	Choke price at upper end of International Energy Agency, 2019 WEO long-run scenarios; implies monopoly P of \$80, within Saudi fiscal break-even range (IMF Article IV).
Demand slope	b	1.0	Normalisation. Implies own-price elasticity of approx. -0.75 at triopoly Nash, mid-range of EIA Short-Term Energy Outlook (-0.4 to -1.0) and Hamilton (2003).
B. Marginal costs (short-run, well-level, per barrel)			
OPEC marginal cost	c_{OPEC}	\$20	Gulf wellhead lifting ($\sim$\$5-\$10, BP, 2022) plus transport, infrastructure and fiscal overhead. Consistent with Aramco Q1 2026 disclosed cost. *Not* the fiscal break-even of \$80.
Russia marginal cost	c_{RUS}	\$35	Siberian lifting ($\sim$\$15-\$20, IMF, 2021), Urals export differential ($\sim$\$10-\$15), plus $\sim$\$5 sanctions premium.
US shale cost	c_{US}	\$45	Well-level break-even (Dallas Fed Energy Survey, 2020). Firm-level Permian break-even (\$61-\$62 in Q1 2025 survey) embeds corporate overhead, outside short-run Cournot best-response margin.
C. Capacities (mbd)			
OPEC capacity	cap_{OPEC}	40	Sum of sustainable capacity of OPEC core (OPEC MOMR, 2024-2026 avg). Set at unconstrained Nash output to make binding at Nash and slack at cartel.
Russia capacity	cap_{RUS}	35	Upper bound on non-OPEC OPEC+ bloc expansion under existing infrastructure (Henderson, 2015, p. ??; U.S. Energy Information Administration, 2025).
US capacity	cap_{US}	30	Upper bound for US shale consistent with Permian Wolfcamp tier estimates (Smith, 2016; EIA Permian assessment).
OPEC capacity sweep		25-60	Range spans 2022-23 tight regime and 2020 slack regime; brackets all empirically observed levels.
D. Time and patience			

(continued on next page)

(continued) Calibrated parameters of the linear-Cournot triopoly

Parameter	Symbol	Value	Rationale and source
Discount factor	δ	0.95	Quarterly frequency matching OPEC+ meeting cadence. Annualised $\sim$19%, deliberately elevated to proxy political and fiscal impatience. Comfortably above binding $\delta^*_{\text{OPEC}} = 0.529$.
Horizon	T	50 qtrs	Long enough for transients to die, short enough to avoid reserves-depletion and energy-transition dynamics that the static cost vector cannot capture.
Inertia	λ	0.2	Quarterly partial-adjustment speed; half-life $\sim$3 quarters, matching Almoguera et al. (2011) regime-switch adjustment estimates.
E. Robustness and welfare			
Bertrand substitutability	σ	0.6	Calibrated against observed Brent/WTI/Urals/sour spreads; centre of implied 0.5-0.7 substitution range.
Carbon tax sweep	τ	\$0-50	Spans 0 to $\sim$\$116/tCO_2, covering EU ETS settlement levels and the OECD-IMF social-cost-of-carbon 2030 range (at 0.43 tCO_2/bbl).
Competitive benchmark	P_c	\$20 ($= c_{\min}$)	Aggressive upper-bound convention consistent with Lombardi and Van Robays (2011) and Hochman and Zilberman (2015). Weighted-average benchmark would halve DWL.
F. Stochastic demand			
$AR(1)$ persistence	ρ	0.6	Quarterly autocorrelation of detrended global oil-demand growth, 1990-2024 (IEA OMR data, authors' calculation).
$AR(1)$ std deviation	σ_a	8	$\sim$6% of intercept; matches Hamilton (2003), Kilian (2009).
Monte Carlo paths	N_{mc}	300	MC standard error on mean price $\sim$0.46, well below \$1/bbl economic threshold. Doubling changes no result by > 0.1.
Simulation horizon		200 qtrs	Long enough for steady-state cooperation and punishment fractions to converge.
GP trigger price	$\bar{p}$	\$68	Midpoint of Nash (\$60) and cartel (\$80). Porter (1983) and Almoguera et al. (2011) convention; monotone robustness over range.
Punishment length	T_p	10 qtrs	Empirical envelope of 1985-86 (8 qtrs) and 2020 (5 weeks plus quota uncertainty) episodes.
G. Correlated equilibrium			

(continued on next page)

(continued) Calibrated parameters of the linear-Cournot triopoly

Parameter	Symbol	Value	Rationale and source
Grid resolution		10 levels	1,000 LP variables; 20-level grid (8,000 vars) yields same qualitative answers in spot-check runs.
Objectives		3	Max-welfare (utilitarian), max-joint-profit (cartel), max-min (Rawlsian).
H. Evolutionary game			
PD payoffs		$(2025, 1800, 1600, 1500)$	Derived from Cournot calibration: π_{dev}, π_{coop}, π_{nash}, π_{sucker}. Not chosen independently.
Punishment multiplier		1.0-2.0	Spans no punishment, 2020 Saudi 50% production surge, and doubling consistent with installed capacity.

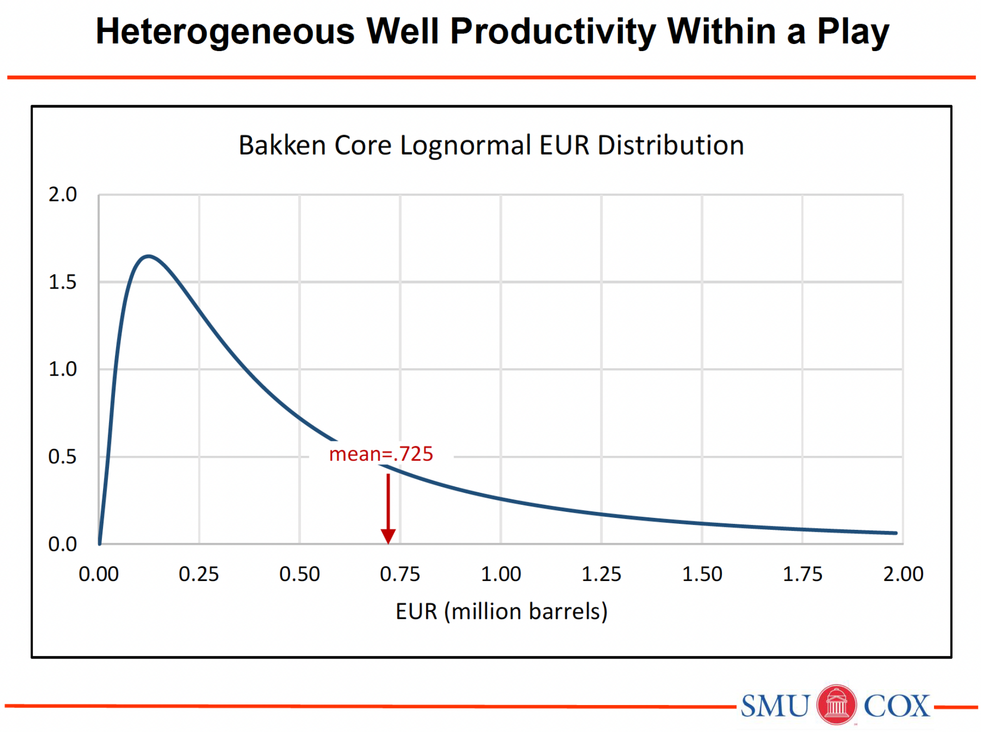
Note: parameter values reflect the joint constraint that (a) each number is empirically anchored or explicitly normalised, (b) the binding constraints in Folk Theorem and Green–Porter analyses are derived rather than imposed, and (c) the model remains analytically tractable across the sixteen frameworks. Sensitivity analyses, reported in the relevant figures, confirm that qualitative conclusions are robust to perturbations within the empirically defensible range of each parameter.

B Reading guide: sixteen-step analytical pipeline

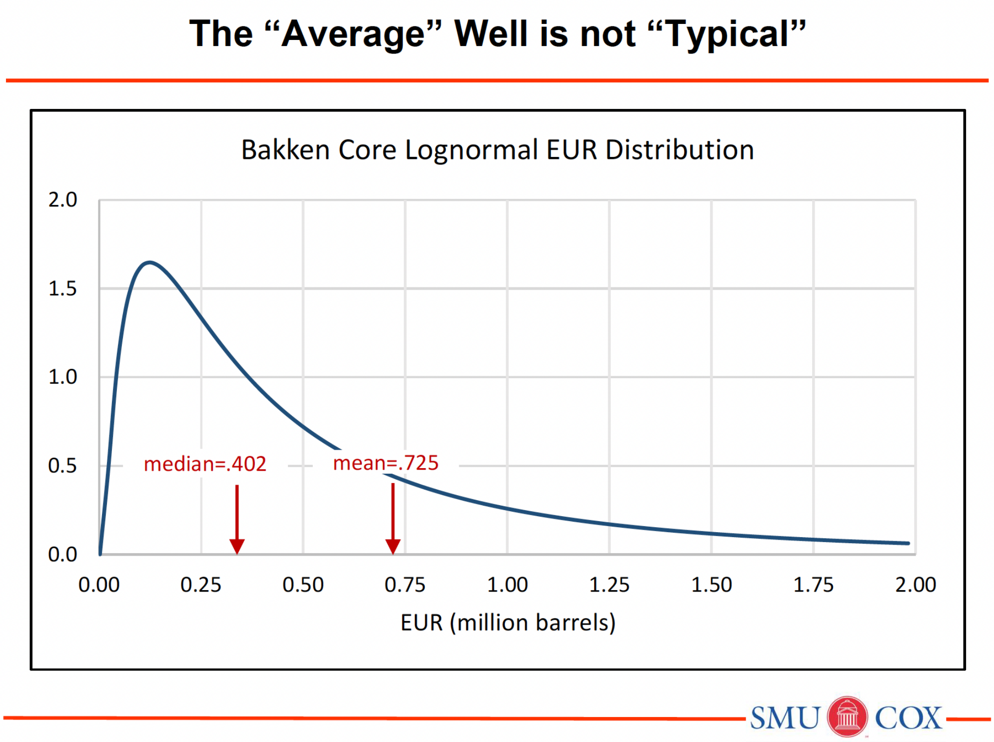
Reading guide. Sixteen-step escalation of the analytical pipeline, identifying the question and the assumption relaxed at each step.

Step	Framework	Question addressed and assumption relaxed
1	Static Cournot duopoly	Non-cooperative benchmark with two strategic producers (OPEC and US).
2	Static Cournot triopoly	How does Russia's entry alter price, quantities, and welfare?
3	Comparative statics on a	Sensitivity of the triopoly equilibrium to the demand level.
4	Market-power quantification (Lerner, HHI)	How concentrated is the market and which producer enjoys the largest pricing power?
5	Stackelberg leadership	Size of the first-mover advantage.
6	Capacity constraints	How do physical production ceilings reshape Nash and cartel outcomes?
7	Welfare and DWL decomposition	Social cost of oligopoly; surplus split between consumers and producers.
8	Carbon-tax interaction	How does a Pigouvian per-barrel tax τ interact with the collusion premium?
9	Differentiated Bertrand	Robustness of Cournot conclusions to choice of strategic variable.
10	Repeated-game myopic best-response	How does the static Nash arise endogenously under simple adaptive rules?
11	Folk Theorem with grim-trigger	Self-enforcing cartel; critical discount factor.
12	Correlated equilibrium	Improvement over Nash via a benevolent mediator.
13	Cooperative game theory & Shapley values	Allocation of joint surplus; core stability.
14	Stochastic demand under $AR(1)$ and Merton jump-diffusion	Propagation of shocks and rare jumps.
15	Green-Porter imperfect-monitoring	Survival of cooperation under price-only observation.
16	Evolutionary game theory: ESS & replicator dynamics	Long-run selection among Cooperate, Defect, Punish.

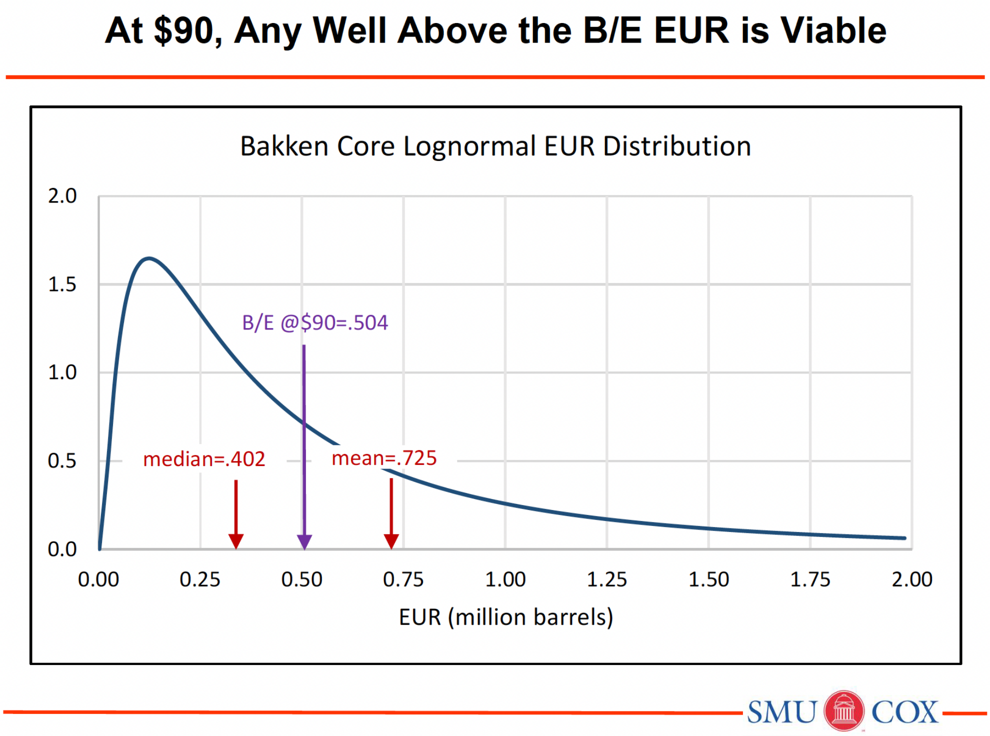
C Smith (2016): Lognormal EUR distributions by shale play



Smith (2016), slide 4 of the EIA Energy Finance Workshop deck. Lognormal estimated ultimate recovery (EUR) distributions by US shale play.



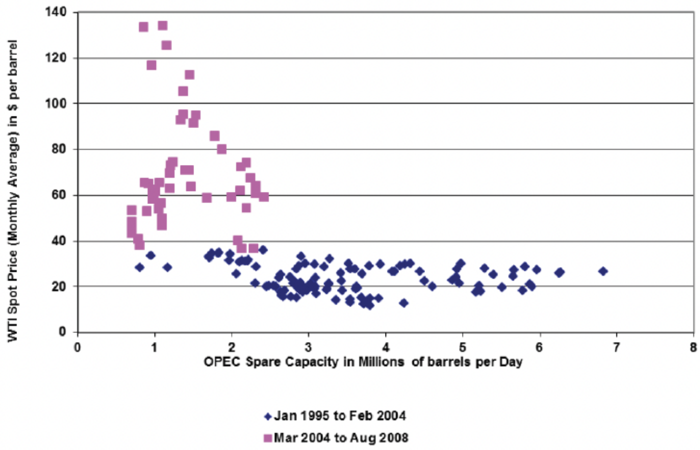
Smith (2016), slide 5. Continued lognormal EUR distributions.



Smith (2016), slide 6. Distributional summary by major play.

D Till (2016): OPEC spare capacity and convex price response

Exhibit3: WTI Spot Price vs. OPEC Spare Capacity (Jan 1995 to Aug 2008) Monthly Data



Source of Graph: Updated from Till (2014), Slide 19.
Sources of Data:
The WTI Spot Price is the "Bloomberg West Texas Intermediate Cushing Crude Oil Spot Price," accessible from the Bloomberg using the following ticker: "USCRWTIC <index>".
The following Bloomberg formula was used to create a monthly data set from daily prices:
bdh("USCRWTIC Index","px last","1/1/1995","8/31/2008","per=cm","quote=g")
The OPEC Spare Capacity data is from the U.S. Energy Information Administration's website, which was accessed on 8/30/14 (for the 1995 data) and on 10/24/15 (for the 1996 through September 2015 data.)
Presenting data in this fashion is based on Büyüksahin et al. (2008), Figure 10, which has a similar, but not identical, graph. Their graph, instead, shows "Non-Saudi crude oil spare production capacity" on the x-axis.

Till (2016), Exhibit 3. Scatter of OPEC spare capacity against the WTI price, illustrating the convexity that emerges below the 2 mbd spare-capacity threshold.

E Huppmann (2013): OPEC market-power trajectories

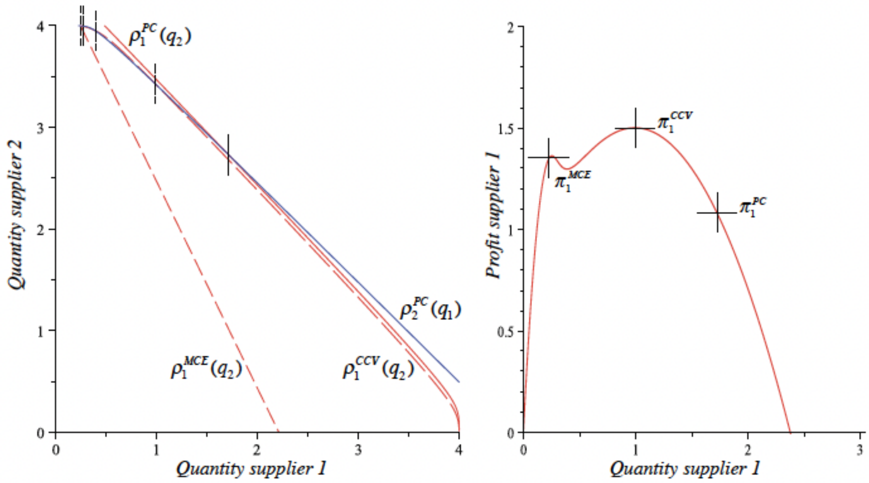


Figure 2: Illustration to Example 8

Huppmann (2013), Figure 2. OPEC market-power trajectories, 2003-2011.

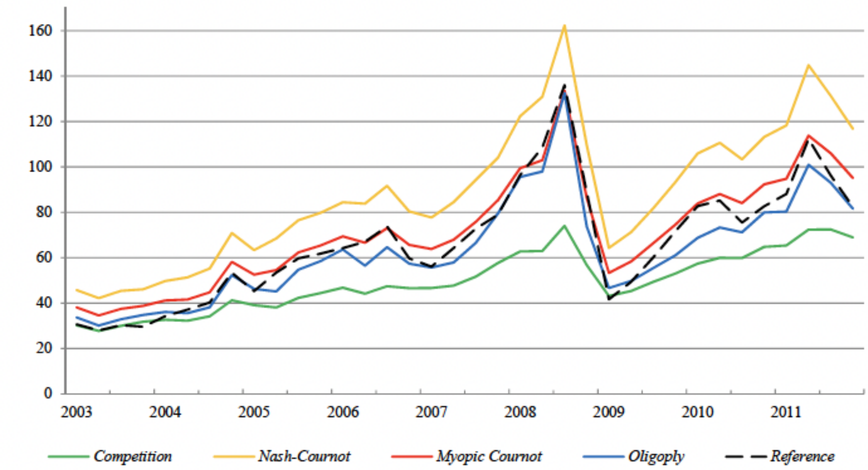


Figure 3: Equilibrium price by market power case, and reference (in US$/bbl)

Huppmann (2013), Figure 3. OPEC market-power trajectories, 2003-2011.

F Almutairi et al. (2025): Counterfactual oil-price volatility

Almutairi et al. (2025), Table 3. Historical versus counterfactual oil-price volatility for the Pandemic and Ukraine War episodes.

Table 3
Historical versus counterfactual volatility, by sub-period.

Period	Scenario			
	Historical	Counterfactual 1 (*Only Allies*)	Counterfactual 2 (*Only OPEC*)	Counterfactual 3 (*Neither*)
OPEC+ Overall *Jan 2017 – Sept 2024 (excl Mar/Apr 2020)**	8.33 %	14.31 %	10.72 %	15.63 %
Pre-Pandemic *Jan 2017 – Feb 2020*	7.24 %	14.48 %	7.16 %	14.34 %
Pandemic and Ukraine War *May 2020 – Sept 2024*	8.95 %	14.04 %	12.57 %	16.31 %
Pandemic *May 2020 – Jan 2022*	9.26 %	15.06 %	11.64 %	17.21 %
Ukraine War *Feb 2022 – Sept 2024*	7.95 %	11.00 %	11.06 %	11.90 %

* These months are excluded from the calculation of volatility because of the temporary hiatus in OPEC+'s cooperative effort to stabilize the price at the outset of the Pandemic.

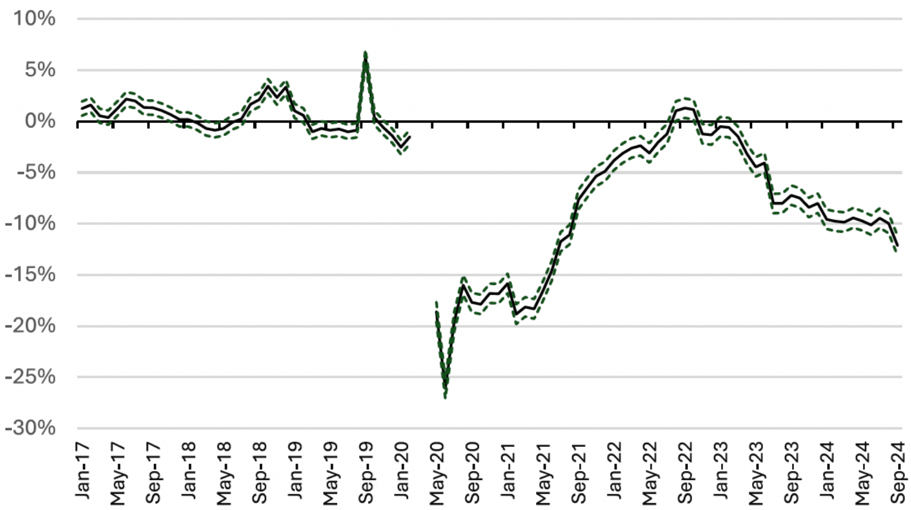


Fig. 8. The perceived shock to residual demand for OPEC crude.
Source: Authors' calculations.

Almutairi et al. (2025), Figure 8. Perceived shock to residual demand for OPEC crude underlying the counterfactual exercise.

G Almoguera et al. (2011): Regime-switching probabilities

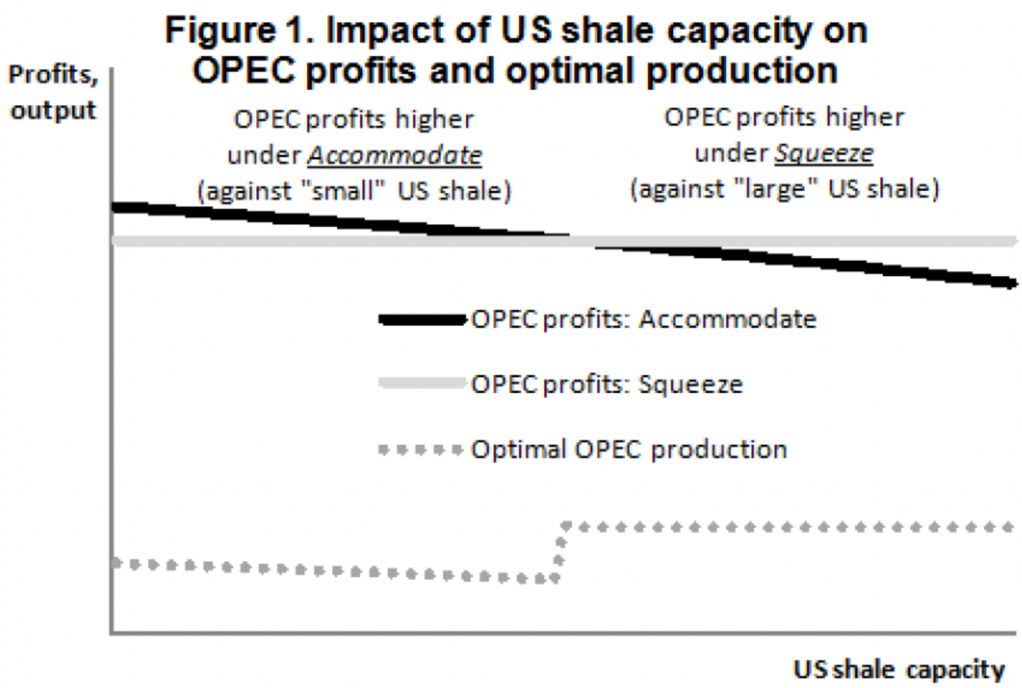
Almoguera et al. (2011), Table 3. Regime-switching probabilities between collusive and Cournot phases for OPEC, 1974-2004.

Table 3. OPEC production quotas, shares and Herfindahl index for four selected years

	1978		1983		1990		1997	
	Share (%)	Production (Millions of barrels per day)	Share (%)	Production (Millions of barrels per day)	Share (%)	Production (Millions of barrels per day)	Share (%)	Production (Millions of barrels per day)
Total Production	1	31	1	17.15	1	21.61	1	27.5
Saudi Arabia	0.32	9.97	0.42	7.15	0.25	5.38	0.32	8.76
Iran	0.12	3.81	0.07	1.20	0.15	3.14	0.14	3.94
Iraq	0.09	2.98	0.07	1.20	0.15	3.14	0.05	1.31
Venezuela	0.07	2.32	0.09	1.50	0.09	1.94	0.09	2.58
Nigeria	0.07	2.17	0.08	1.30	0.07	1.61	0.07	2.04
Indonesia	0.05	1.58	0.08	1.3	0.06	1.37	0.05	1.45
Kuwait	0.07	2.31	0.05	0.80	0.07	1.50	0.08	2.19
Libya	0.06	2.13	0.04	0.75	0.06	1.23	0.06	1.52
U.A.E.	0.05	1.83	0.06	1.00	0.05	1.09	0.09	2.36
Algeria	0.04	1.37	0.04	0.65	0.04	0.82	0.03	0.90
Qatar	0.01	0.52	0.02	0.30	0.02	0.37	0.01	0.41
Herfindahl Index	0.093		0.100		0.075		0.055	

Note: "Share" is the share of each OPEC member country in OPEC's production. Production is the assigned production quota, except for 1978 where it denotes actual production.

H Behar and Ritz (2017): Market-share regime thresholds



Behar and Ritz (2017), Figure 4. Accommodation vs. market-share regime thresholds in the swing-producer model.