

Cournot competition under a binding price ceiling: equilibrium structure and exact pass-through

Tumendelger Lkhagvasuren

tumendelger.l@ufe.edu.mn

University of Finance and Economics

Research Article

Keywords: Cournot competition, Price ceiling, Nondifferentiable payoffs, Equilibrium multiplicity, Aggregative games, Nikaidō–Isoda gap function

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11113875/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Cournot competition under a binding price ceiling: equilibrium structure and exact pass-through

Tumendelger Lkhagvasuren ^{1*}

¹Graduate School, University of Finance and Economics, Ulaanbaatar,
Mongolia.

Corresponding author(s). E-mail(s): tumendelger.l@ufe.edu.mn;

Abstract

We study Cournot competition under an administered price ceiling that truncates inverse demand from above, with firms bounded by capacities. The ceiling makes payoffs non-differentiable exactly where the equilibrium lies, so first-order conditions do not characterise best replies. We show that concavity nonetheless survives: for any inverse demand generating concave revenue, capped revenue is the *minimum* of a linear and a concave function, so each payoff stays concave in own output and a pure-strategy equilibrium exists. A price floor produces a maximum instead and the argument fails — an asymmetry between the instruments. Where the ceiling binds through the kink, a profile is an equilibrium if and only if the market clears at the ceiling and each firm's output lies between its Cournot reply and its competitive supply at the ceiling price. The equilibrium set is therefore a polytope whose dimension and coordinate ranges we compute: the aggregate is unique, the allocation in general is not. Multiplicity is confined to ceilings strictly between a supply-maximizing ceiling and the uncapped equilibrium price, and fills that interval whenever two or more firms are constrained neither by capacity nor by cost. Two exact comparative statics follow: a public supply passes through into total supply at a rate of exactly zero or one, never between; and total supply is single-peaked in the ceiling. On a calibrated instance damped best-response iteration converges from every start, is certified by a Nikaidô–Isoda gap function, and yet returns different equilibria from different starts — all inside the predicted segment.

Keywords: Cournot competition, Price ceiling, Nondifferentiable payoffs, Equilibrium multiplicity, Aggregative games, Nikaidô–Isoda gap function

JEL Classification: C72 , D43 , L13 , L51

MSC Classification: 91A10 , 91A80 , 91B26 , 90C26

1 Introduction

An administered price ceiling turns a smooth Cournot game into a nonsmooth one. The ceiling enters inverse demand as a minimum, so each payoff is piecewise quadratic with a kink, and the first-order conditions cease to characterise a best reply at that kink — which, when the ceiling is tight enough to hold the firms there, is exactly where the equilibrium sits. The usual apparatus therefore fails at the only point that matters. Existence cannot be read off the stationarity conditions. Uniqueness cannot be read off a monotonicity argument applied to a gradient that does not exist. An iteration built on best replies has no reason to stop at a solution. This paper asks what survives, and finds that a good deal does, in a sharper form than the smooth theory offers.

The economics of price caps in oligopoly is well studied and the qualitative picture is understood: a cap placed between marginal cost and the unconstrained oligopoly price *raises* output, one placed below cost removes it, and effects of both signs are accordingly reported (Earle et al, 2007; Grimm and Zöttl, 2010; Reynolds and Rietzke, 2018; Roques and Savva, 2009; Nie et al, 2023). The sharpest of those results assume constant marginal cost (Earle et al, 2007; Grimm and Zöttl, 2010); with increasing marginal cost, as here, a welfare-improving cap need not exist at all (Reynolds and Rietzke, 2018). What that literature supplies is comparative statics under assumptions that keep the problem smooth, or numerical illustration where it does not. What it does not supply is the structure of the equilibrium set under a binding ceiling — for n firms, with capacities, and with necessary and sufficient conditions describing it. That is what this paper supplies, and the structure turns out to be the source of everything else.

Four results are established, none of which follows from the smooth theory. The ceiling preserves concavity of the payoff in own output, and hence existence, for any demand under which uncapped revenue is concave, because capping takes a minimum where flooring would take a maximum — which is why no analogue of the result is available for a price floor (Proposition 1 and Remark 1). Where the ceiling binds through the kink, a profile is an equilibrium if and only if the market clears at the ceiling and each firm's output lies between its Cournot reply and its competitive supply at the ceiling price, so the equilibrium set is a nonempty compact convex polytope whose dimension and coordinate ranges we compute: the aggregate is determined, the split in general is not (Proposition 2 and Corollary 1). The mechanism is a segment of slope exactly -1 in the best-reply correspondence, on which a firm does nothing but fill the gap to the kink (Lemma 2). A public supply placed into the same market passes through into total supply at a rate of exactly 0 or exactly 1, never between, with the switch at an explicitly computable point (Proposition 3 and Corollary 2). And, treating the ceiling itself as the parameter, total supply is single-peaked in $\bar{p}$ and is maximized at an explicitly computable $\bar{p}^\circ$, which is exactly the lower endpoint of the interval of ceilings outside which the equilibrium is unique and, whenever that interval is nonempty, the largest ceiling that maximizes supply (Propositions 4 and 5). Multiplicity is therefore not an inconvenience sitting to one side of the policy problem; once two firms are free of their capacity and cost constraints, it occupies exactly the range of caps that are too loose to be supply-maximizing and too tight to be slack.

We also report a computational experiment, which is what first drew attention to the structure above. Where the equilibrium lies at the kink, damped best-response iteration converges from every start, passes its own stopping test, is certified by the Nikaidô–Isoda gap function (Nikaidô and Isoda, 1955) — and returns different equilibria from different starts. On a calibrated instance 100 random starts yield 50 distinct equilibria, agreeing on the aggregate to nine significant figures and disagreeing on the split by up to 5.8 per cent of total supply; every one lies inside the segment Corollary 1 predicts, and run sequentially rather than simultaneously the iteration covers that segment entire (Section 4). A solver run once therefore returns an equilibrium and no indication that a six-per-cent band of splits is equally one. What the gap function supplies is a certificate, not a correction; what identifies the solution set is Proposition 2. The one thing the iteration cannot do is contract, and we locate that failure exactly: on the slope- (-1) branch the damped simultaneous map multiplies the aggregate error by $1 - \lambda|\mathcal{F}|$, so it converges if and only if $\lambda < 2/|\mathcal{F}|$, and the undamped map never does (Remark 6).

The paper sits at the junction of three lines of work.

Existence and uniqueness in Cournot games. That Cournot equilibria exist under conditions weaker than concavity of the payoff has been known since Roberts and Sonnenschein (1976) and Novshek (1985), and uniqueness under increasing marginal cost is classical (Szidarovszky and Yakowitz, 1977; Kolstad and Mathiesen, 1987). The route taken here is the concave-game route of Debreu (1952), Fan (1952) and Glicksberg (1952), in the form closest to the present setting due to Nikaidô and Isoda (1955). The contribution is not a new existence theorem but the observation that the standard sufficient condition — concavity of revenue in own output — is preserved exactly by capping and destroyed exactly by flooring. Existence under genuinely discontinuous payoffs requires machinery of a different kind (Reny, 1999); here the payoffs are continuous but nondifferentiable, and it is quasiconcavity, not continuity, that is at risk.

Concave, aggregative and supermodular games. The capped game is a concave game in the sense of Rosen (1965), but Rosen’s diagonal strict concavity fails — Proposition 2 exhibits a continuum, so it must — and what survives is uniqueness of the aggregate rather than of the profile. The game is also aggregative, each payoff depending on own output and on the total, hence a nice aggregative game in the sense of Acemoglu and Jensen (2013), whose aggregate comparative statics is the natural setting for Proposition 3; see also Corchón (1994), Jensen (2010) and the share-function treatment of Cornes and Hartley (2012). The feature of interest here is not that the aggregate responds predictably but that over an entire regime it does not respond at all. Cournot competition is a game of strategic substitutes and submits to the lattice-theoretic treatment of Amir (1996), following Milgrom and Roberts (1990), Vives (1990) and Monderer and Shapley (1996); the ceiling is what flattens best replies, and Lemma 2 isolates the flat piece. That flat piece leaves the game aggregative with a non-increasing best-reply selection, so it remains a pseudo-potential game in the sense of Dubey et al (2006), and in its finite versions sequential best-reply dynamics is acyclic (Kukushkin, 2004) — a point we return to in Section 4.4, because it is what separates the sequential iteration, which behaves, from the simultaneous one, which has

an exact stability threshold. Capacity-constrained quantity competition is itself classical (Kreps and Scheinkman, 1983), though here the bounds are exogenous physical limits rather than first-stage choices.

Price caps and capacity in oligopoly. The closest structural analogue we are aware of is Tao et al (2018), where multiplicity in a duopoly arises from a binding *joint capacity* constraint rather than from a price cap. The mechanism is related — a constraint shared across firms flattens best replies and leaves the split undetermined — but the instrument is different, and the characterisation here is for n firms with individual capacities and with necessary and sufficient conditions. We have not found the equilibrium set under a binding cap characterised elsewhere.

Computational reformulations. The paper belongs, computationally, to the line that recasts an equilibrium problem as a single optimisation problem in a Nikaidô–Isoda function (Nikaidô and Isoda, 1955; Facchinei and Kanzow, 2007; von Heusinger and Kanzow, 2009; Panicucci et al, 2009). We use that function not as an algorithm but as a *certificate*, in the role gap functions play for variational inequalities (Fukushima, 1992) and merit functions play in complementarity problems (Facchinei and Pang, 2003).

The game arises from an application, described here only far enough to motivate the structure. A landlocked economy imports refined fuel through a small number of licensed firms at an administered retail price, each firm bounded by what it can physically bring across the border in a month; the state holds a reserve and may release y from it. Proposition 3 says what such a release accomplishes, in a form that does not depend on the calibration: while the ceiling binds, nothing at all.

The paper is organised as follows. Section 2 sets out the model and Section 3 establishes the four results: existence (Section 3.1), the structure of the equilibrium set (Section 3.2), the pass-through of a public supply (Section 3.3), and the comparative statics in the ceiling (Section 3.4). Section 4 checks the theory against a calibrated instance and reports the certification experiment, and Section 5 concludes.

2 Model

Let $n \geq 2$ firms choose outputs $q_i \geq 0$, $i = 1, \dots, n$, and write $Q = \sum_j q_j$. Firm i is bounded above by a capacity $\kappa_i > 0$, so its strategy set is $[0, \kappa_i]$. Let $y \geq 0$ be a public supply placed into the market and taken as given by the firms. Demand is linear and truncated by an administered ceiling $\bar{p}$:

$$P(Q + y) = \min\{\bar{p}, a - b(Q + y)\}, \quad a > 0, \bar{p} > 0, b > 0. \quad (1)$$

A ceiling at or above a never binds; it is allowed so that Section 3.4 can vary $\bar{p}$ freely. Firm i has a linear-quadratic cost with constant $c_i \geq 0$ and curvature $\gamma_i > 0$, so its payoff is

$$\pi_i(q_i, q_{-i}) = q_i P(Q + y) - c_i q_i - \frac{1}{2} \gamma_i q_i^2. \quad (2)$$

Proposition 1 below is proved for general inverse demand; everything after it uses Eq. (1).

Five quantities recur. The first is the total at which the ceiling stops binding,

$$Q^{\text{kink}}(y) = \frac{a - \bar{p}}{b} - y, \quad (3)$$

so that $P = \bar{p}$ for $Q \leq Q^{\text{kink}}(y)$ and $P < \bar{p}$ above it. The next two are individual quantities evaluated at the ceiling price. Firm i 's *competitive* supply at $\bar{p}$ is what it would produce facing $\bar{p}$ parametrically,

$$\bar{q}_i = \min \left\{ \kappa_i, \frac{(\bar{p} - c_i)_+}{\gamma_i} \right\}, \quad (4)$$

and its *Cournot* reply at $\bar{p}$ is what it would produce if it took account of its own effect on the price,

$$\underline{q}_i = \min \left\{ \kappa_i, \frac{(\bar{p} - c_i)_+}{b + \gamma_i} \right\}, \quad (5)$$

where $(\cdot)_+$ denotes the positive part. Since $b > 0$ we have $\underline{q}_i \leq \bar{q}_i$, with equality exactly when firm i is pinned by capacity ($\kappa_i \leq (\bar{p} - c_i)_+ / (b + \gamma_i)$) or excluded by cost ($\bar{p} \leq c_i$). Write the two aggregates

$$Q^{\text{free}} = \sum_{i=1}^n \bar{q}_i, \quad Q^{\text{cour}} = \sum_{i=1}^n \underline{q}_i, \quad Q^{\text{cour}} \leq Q^{\text{free}}. \quad (6)$$

Neither depends on y , a fact Proposition 3 turns into a comparative static. Both are nondecreasing and continuous in $\bar{p}$, while $Q^{\text{kink}}(y)$ is strictly decreasing in $\bar{p}$; Section 3.4 rests on nothing more than that.

Assumption 1. The ceiling binds at the public supply under consideration:

$$Q^{\text{cour}} \leq Q^{\text{kink}}(y). \quad (7)$$

Assumption 1 is a restriction on primitives, and it is the only form in which the hypothesis is ever used: every proof in Sections 3.2 to 3.4 consumes Eq. (7) and nothing else. Its reading is the expected one. Q^{cour} is the total the firms would place on the market if each took account of its own effect on the price while facing $\bar{p}$; $Q^{\text{kink}}(y)$ is the total the market absorbs at the ceiling. Equation (7) says the first falls short of the second, so the ceiling is still active once the firms have optimised against it, and $P = \bar{p}$ at equilibrium. This is the case in which the smooth theory fails; when the ceiling is slack the game is the standard one.

Lemma 3 locates the threshold: Eq. (7) holds if and only if $\bar{p} \leq p^u$, where p^u is the price prevailing in the same game with the ceiling removed. Stating the assumption on primitives rather than through that auxiliary game is not only tidier, it is what keeps every proof below independent of that game. The uncapped game does have a unique equilibrium — its pseudogradient Jacobian is $-(b\mathbf{1}\mathbf{1}^\top + \text{diag}(b + \gamma_i))$, whose symmetric part is negative definite, so Rosen's diagonal strict concavity holds (Rosen, 1965) — but no result below needs that fact. Lemma 3 gets uniqueness of p^u from

strict monotonicity of $Q^{\text{cour}} - Q^{\text{kink}}$, and part (ii) of Proposition 5, which is the only statement here about a ceiling that does *not* bind, is proved from a contraction property of the clamped Cournot reply. The ceiling is exactly what destroys diagonal strict concavity; Remark 4 returns to the point.

3 Results

The four results of this section are stated for the game of Section 2 and proved from it. Section 3.1 establishes existence, and locates the asymmetry between a ceiling and a floor. Section 3.2 characterises the equilibrium set. Section 3.3 derives the exact pass-through of a public supply, and Section 3.4 the comparative statics in the ceiling itself.

3.1 Existence

At the kink π_i is not differentiable, and under Assumption 1 the equilibrium may lie there. Concavity in own strategy is nevertheless preserved, and existence follows from it. We prove this for general inverse demand, because the mechanism has nothing to do with linearity.

Proposition 1 (Existence). *Let $P_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ be continuous and such that for every i and every $Z \geq 0$ the uncapped revenue $q \mapsto q P_0(q + Z)$ is concave on $[0, \kappa_i]$, and let each cost function C_i be continuous and convex. Then for every ceiling $\bar{p}$ and every $y \geq 0$ the game with payoffs*

$$\pi_i(q_i, q_{-i}) = q_i \min\{\bar{p}, P_0(Q + y)\} - C_i(q_i)$$

on $\prod_i [0, \kappa_i]$ has a pure-strategy Nash equilibrium. In particular this holds for the game defined by Eqs. (1) and (2).

Proof. Each strategy set $[0, \kappa_i]$ is nonempty, compact and convex, and π_i is continuous in the profile, being a composition of continuous functions. Fix q_{-i} and write $Z_i = \sum_{j \neq i} q_j + y \geq 0$ for the volume firm i faces from all other sources, private and public. Because $q_i \geq 0$, multiplication by q_i preserves the order of the two arguments of the minimum, so

$$R_i(q_i) = q_i \min\{\bar{p}, P_0(q_i + Z_i)\} = \min\{\bar{p} q_i, q_i P_0(q_i + Z_i)\}. \quad (8)$$

The first argument is linear in q_i , hence concave; the second is concave by hypothesis. A pointwise minimum of concave functions is concave, so R_i is concave on $[0, \kappa_i]$. Subtracting the convex C_i preserves concavity, so π_i is concave in q_i for every q_{-i} . Existence of a pure-strategy equilibrium follows from the Debreu–Glicksberg–Fan theorem (Debreu, 1952; Fan, 1952; Glicksberg, 1952), or in the form closest to the present setting from Nikaidô and Isoda (1955).

For Eqs. (1) and (2), take $P_0(x) = a - bx$, so that $q P_0(q + Z) = q(a - bZ) - bq^2$ is a concave quadratic, and $C_i(q_i) = c_i q_i + \frac{1}{2} \gamma_i q_i^2$, which is convex because $\gamma_i > 0$. $\square$

Remark 1 (Ceilings versus floors). Equation (8) is what does the work, and it is the direction of the truncation that puts a minimum there rather than a maximum. A ceiling truncates the price from *above*, removing revenue at high own-output and leaving it untouched at low own-output, so capped revenue is a minimum of two concave functions and the kink bends the revenue function downward. A price *floor* $\underline{p}$ gives $q_i \max\{\underline{p}, P_0(q_i + Z_i)\} = \max\{\underline{p}q_i, q_i P_0(q_i + Z_i)\}$, a maximum of two concave functions, and concavity genuinely fails. One instance settles it. Take $P_0(x) = 2 - x$, $\underline{p} = 1$ and $Z_i = 0$, so that the two branches are q and $q(2 - q)$ and they cross at $q = 1$. The revenue $q \mapsto q \max\{1, 2 - q\}$ is not concave on $[0, 2]$: at $q = 1$ it takes the value 1, strictly below the chord joining $q = \frac{1}{2}$ and $q = \frac{3}{2}$, which passes through $\frac{9}{8}$. Nor does it become *convex* at the crossing, which would at least have left a convex-analytic route open: on $[\frac{1}{2}, 1]$, immediately to the left of the crossing, the maximum is still the strictly concave branch, and at $q = \frac{3}{4}$ it takes the value $\frac{15}{16}$, strictly *above* the chord joining the endpoints, which passes through $\frac{7}{8}$. The same instance with a ceiling, $q \mapsto q \min\{1, 2 - q\}$, is concave on all of $[0, 2]$. The floored revenue is still quasiconcave, being nondecreasing on $[0, 2]$, but a cost removes even that. With $C(q) = 0.9q$ the floored payoff is $q \mapsto \max\{0.1q, 1.1q - q^2\}$: it rises to 0.3025 at $q = 0.55$, falls to 0.1 at $q = 1$ and climbs again to 0.2 at $q = 2$, so its upper level set at 0.18 is $[0.2, 0.9] \cup [1.8, 2]$, which is not convex. The capped payoff with the same cost, $q \mapsto \min\{0.1q, 1.1q - q^2\}$, is concave. Under a floor the payoff need not be even quasiconcave in own output, so neither Proposition 1, nor Debreu–Glicksberg–Fan in its quasiconcave form, nor any argument routed through Rosen (1965) applies. Continuity is not the obstacle — payoffs remain continuous under a floor, so the discontinuous-games machinery of Reny (1999) is not what is missing either; quasiconcavity is. We do not claim nonexistence under a floor, only that the route taken here is closed, and closed for a reason that is visible in one line.

Remark 2 (What the linear case adds). For Eq. (1) the concavity in Proposition 1 can be seen quantitatively as well. Let $\hat{q} = (a - \bar{p})/b - Z_i$ be the own quantity at which the ceiling stops binding. If $\hat{q} > 0$, revenue is linear with slope $\bar{p}$ on $[0, \hat{q}]$ and a concave quadratic on $[\hat{q}, \infty)$; the left derivative at $\hat{q}$ is $\bar{p}$ and the right derivative is $a - bZ_i - 2b\hat{q} = \bar{p} - b\hat{q}$, so the derivative jumps *down* by exactly $b\hat{q} \geq 0$. The size of the jump, $b\hat{q}$, is the marginal revenue a firm forgoes on its inframarginal units the instant it pushes the market past the kink, and it is what makes the kink an attractor. A smooth peak is attained at a single quantity; a downward kink is attained on an interval, because the left slope can still be positive where the right slope is already negative. That interval is exactly $[\underline{q}_i, \bar{q}_i]$ of Proposition 2, and it is at most $b(\bar{p} - c_i)_+ / (\gamma_i(b + \gamma_i))$ wide — exactly that wide when firm i 's competitive supply at the ceiling is set by price rather than by capacity, narrower when capacity binds, and a single point when capacity binds hard enough. Either way it collapses as $b \downarrow 0$, along with the jump that opened it.

3.2 The equilibrium set

Uniqueness fails under Assumption 1. This section says exactly how. We first record the aggregate, which is unique, and then the set of profiles delivering it, which in general is not.

3.2.1 The aggregate

Lemma 1 (Aggregate representation). *Fix $y \geq 0$ and suppose Assumption 1 holds. Then every equilibrium of the game defined by Eqs. (1) and (2) has the same total output,*

$$Q^*(y) = \min \{ Q^{\text{free}}, Q^{\text{kink}}(y) \}. \quad (9)$$

If $Q^{\text{free}} \leq Q^{\text{kink}}(y)$ the equilibrium is moreover unique and equal to the profile $\bar{q} = (\bar{q}_1, \dots, \bar{q}_n)$ of competitive supplies at $\bar{p}$.

Proof. The argument compares payoffs directly, uses one-sided deviations only, and never passes through the uncapped game: the sole consequence of Assumption 1 it consumes is Eq. (7).

Fix an equilibrium q , write $Q = \sum_j q_j$ and $Z_i = Q - q_i + y$. Two observations do the work, each a comparison of one concave quadratic at two points.

(a) *No equilibrium has total above the kink.* Suppose $Q > Q^{\text{kink}}(y)$. The prevailing price $p = a - b(Q + y)$ is then strictly below $\bar{p}$, and if $q_i > 0$ then for t slightly below q_i the total is still above the kink, so on that range $\pi_i(t, q_{-i}) = (a - bZ_i - c_i)t - \frac{1}{2}(\gamma_i + 2b)t^2$: a concave quadratic with peak at $(a - bZ_i - c_i)/(\gamma_i + 2b)$. Such a function fails to decrease to the left of q_i only if its peak is at or above q_i , so equilibrium forces $(\gamma_i + 2b)q_i \leq a - bZ_i - c_i$. Since $a - bZ_i = p + bq_i$, this is $(b + \gamma_i)q_i \leq p - c_i < \bar{p} - c_i$, whence $q_i < (\bar{p} - c_i)_+ / (b + \gamma_i)$ and, with $q_i \leq \kappa_i$, $q_i \leq \bar{q}_i$, which holds trivially when $q_i = 0$. Summing over i gives $Q \leq Q^{\text{cour}} \leq Q^{\text{kink}}(y)$ by Eq. (7), which contradicts $Q > Q^{\text{kink}}(y)$. Hence $Q \leq Q^{\text{kink}}(y)$.

(b) *At a profile with $Q \leq Q^{\text{kink}}(y)$ the price is $\bar{p}$, and it stays $\bar{p}$ for every $t \leq q_i$,* so $\pi_i(t, q_{-i}) = (\bar{p} - c_i)t - \frac{1}{2}\gamma_i t^2$ there. The same peak comparison, read on the other side, gives $\gamma_i q_i \leq (\bar{p} - c_i)_+$, that is $q_i \leq \bar{q}_i$ for every i ; hence $Q \leq Q^{\text{free}}$. If moreover $Q < Q^{\text{kink}}(y)$, each firm can raise t above q_i by a positive amount without carrying the market past the kink, and on that range the payoff is the same quadratic; not gaining forces $\gamma_i q_i \geq (\bar{p} - c_i)_+$ or $q_i = \kappa_i$, that is $q_i \geq \bar{q}_i$.

Now conclude. If $Q^{\text{kink}}(y) \leq Q^{\text{free}}$, then $Q < Q^{\text{kink}}(y)$ is impossible: (b) would give $q = \bar{q}$ and hence $Q = Q^{\text{free}} \geq Q^{\text{kink}}(y)$. With (a) this leaves $Q = Q^{\text{kink}}(y) = \min\{Q^{\text{free}}, Q^{\text{kink}}(y)\}$. If instead $Q^{\text{free}} \leq Q^{\text{kink}}(y)$, then $Q \leq Q^{\text{free}}$ by (b). Should $Q < Q^{\text{kink}}(y)$, (b) gives $q = \bar{q}$ outright. Should $Q = Q^{\text{kink}}(y)$, then $Q^{\text{free}} \leq Q^{\text{kink}}(y) = Q \leq Q^{\text{free}}$ forces $Q = Q^{\text{free}}$, and coordinates bounded above by the $\bar{q}_i$ with that sum must equal them, so $q = \bar{q}$ again. Either way $q = \bar{q}$ and $Q = Q^{\text{free}} = \min\{Q^{\text{free}}, Q^{\text{kink}}(y)\}$. $\square$

3.2.2 The profiles

Proposition 2 (Characterisation of the equilibrium set). *Fix y , suppose Assumption 1 holds, and suppose the ceiling binds through the kink rather than through capacity, that*

is $Q^{\text{free}} \geq Q^{\text{kink}}(y)$. Then q is a Nash equilibrium if and only if

$$\sum_{i=1}^n q_i = Q^{\text{kink}}(y) \quad \text{and} \quad \underline{q}_i \leq q_i \leq \bar{q}_i \quad \text{for every } i, \quad (10)$$

with $\underline{q}_i$ and $\bar{q}_i$ as in Eqs. (4) and (5). Total output is therefore the same at every equilibrium, while the allocation across firms need not be.

Proof. Both directions come from the definition of equilibrium by the same two comparisons, and no derivative is taken. Write $Z_i = \sum_{j \neq i} q_j + y$, $A_i = a - bZ_i$, and record the identity

$$\left[\alpha s - \frac{\beta}{2} s^2 \right] - \left[\alpha t - \frac{\beta}{2} t^2 \right] = (s - t) \left(\alpha - \frac{\beta}{2} (s + t) \right), \quad (11)$$

which replaces every first-order argument below.

Sufficiency. Suppose $\sum_j q_j = Q^{\text{kink}}(y)$ and $\underline{q}_i \leq q_i \leq \bar{q}_i$ for every i . Then $q_i + Z_i = (a - \bar{p})/b$, so $A_i = \bar{p} + bq_i$. Fix i and $t \in [0, \kappa_i]$.

If $t \leq q_i$ the total stays at or below the kink, the price is $\bar{p}$, and $\pi_i(t, q_{-i}) = (\bar{p} - c_i)t - \frac{1}{2}\gamma_i t^2$. Apply Eq. (11) with $\alpha = \bar{p} - c_i$, $\beta = \gamma_i$, $s = t$: the factor $s - q_i$ is nonpositive, and $\alpha - \frac{\beta}{2}(t + q_i) \geq \alpha - \beta q_i \geq 0$ because $q_i \leq \bar{q}_i$ gives $\gamma_i q_i \leq (\bar{p} - c_i)_+$. (If $\bar{p} \leq c_i$ then $\bar{q}_i = 0$, so $q_i = 0$ and there is no such $t < q_i$.) Hence $\pi_i(t) \leq \pi_i(q_i)$.

If $t \geq q_i$ the total is at or above the kink, the price is $A_i - bt$, and $\pi_i(t, q_{-i}) = (A_i - c_i)t - \frac{1}{2}(\gamma_i + 2b)t^2$. Apply Eq. (11) with $\alpha = A_i - c_i = \bar{p} + bq_i - c_i$ and $\beta = \gamma_i + 2b$: the factor $t - q_i$ is nonnegative, and $\alpha - \frac{\beta}{2}(t + q_i) \leq \alpha - \beta q_i \leq 0$ as soon as $\bar{p} - c_i \leq (b + \gamma_i)q_i$, which is exactly $q_i \geq \underline{q}_i$ unless $q_i = \kappa_i$, in which case there is no such $t > q_i$. Hence $\pi_i(t) \leq \pi_i(q_i)$, and q_i is a best reply for every i .

Necessity. Lemma 1 together with $Q^{\text{free}} \geq Q^{\text{kink}}(y)$ gives $\sum_j q_j = Q^{\text{kink}}(y)$. Step (b) of its proof gives $q_i \leq \bar{q}_i$. And the $t \geq q_i$ comparison, read backwards, gives $q_i \geq \underline{q}_i$: if $q_i < \underline{q}_i$ then $q_i < \kappa_i$ and $\alpha > \beta q_i$, so Eq. (11) makes a sufficiently small increase strictly profitable.

The two conditions are jointly attainable because $\underline{q}_i \leq \bar{q}_i$. $\square$

Equation (10) has a reading that makes the multiplicity intelligible. Under a binding ceiling every firm sells at the same price $\bar{p}$ whatever it does, up to the moment the market as a whole crosses the kink. A firm producing less than its Cournot reply $\underline{q}_i$ would find the gap to the kink large enough that expanding through it pays, despite the price it would then break; a firm producing more than its competitive supply $\bar{q}_i$ would rather cut back and sell fewer units at the unchanged price $\bar{p}$. Anything between the two is a best reply, and the market simply has to clear at the kink. *The equilibrium set is the set of ways of allocating kink demand among firms so that each sits between its Cournot and its competitive supply at the ceiling price.* The strategic and the competitive benchmarks bracket the equilibrium instead of selecting it.

Corollary 1 (The equilibrium set is a polytope). *Under the hypotheses of Proposition 2 the equilibrium set is*

$$\mathcal{E}(y) = \left\{ q \in \prod_i [\underline{q}_i, \bar{q}_i] : \sum_i q_i = Q^{\text{kink}}(y) \right\},$$

a nonempty compact convex polytope. The set so defined is nonempty if and only if

$$Q^{\text{cour}} \leq Q^{\text{kink}}(y) \leq Q^{\text{free}}, \quad (12)$$

and if both inequalities in Eq. (12) are strict then, writing $\mathcal{F} = \{i : \underline{q}_i < \bar{q}_i\}$ for the set of firms pinned neither by capacity nor by cost,

$$\dim \mathcal{E}(y) = |\mathcal{F}| - 1. \quad (13)$$

In particular the equilibrium is unique whenever $|\mathcal{F}| \leq 1$ or either inequality in Eq. (12) holds with equality, and is a continuum otherwise. Firm i 's output ranges over exactly

$$\left[\max \{ \underline{q}_i, Q^{\text{kink}}(y) - (Q^{\text{free}} - \bar{q}_i) \}, \min \{ \bar{q}_i, Q^{\text{kink}}(y) - (Q^{\text{cour}} - \underline{q}_i) \} \right] \quad (14)$$

as q ranges over $\mathcal{E}(y)$. When $|\mathcal{F}| = 2$, say $\mathcal{F} = \{i, j\}$, the set is the segment $\{(q_i, q_j) : q_i + q_j = S, q_i \in [\alpha, \beta]\}$ with $S = Q^{\text{kink}}(y) - \sum_{k \notin \mathcal{F}} \bar{q}_k$, $\alpha = \max\{\underline{q}_i, S - \bar{q}_j\}$ and $\beta = \min\{\bar{q}_i, S - \underline{q}_j\}$, of Euclidean length $\sqrt{2}(\beta - \alpha)$.

Proof. The description of $\mathcal{E}(y)$ is Eq. (10) verbatim: the intersection of the box $\prod_i [\underline{q}_i, \bar{q}_i]$ with the hyperplane $\sum_i q_i = Q^{\text{kink}}(y)$, hence compact, convex and polyhedral. Such an intersection is nonempty if and only if the target lies between the sums of the endpoints, which is Eq. (12); its left inequality is Assumption 1 itself.

For the dimension, consider the space of *chords*, $\text{span}\{x - x' : x, x' \in \mathcal{E}(y)\}$. Every chord vanishes at each coordinate outside $\mathcal{F}$, where the box is a single point, and sums to zero, because both endpoints sum to $Q^{\text{kink}}(y)$; so the chord space is contained in

$$D_{\mathcal{F}} = \left\{ v \in \mathbb{R}^n : v_i = 0 \text{ for } i \notin \mathcal{F}, \sum_i v_i = 0 \right\},$$

which for $\mathcal{F} \neq \emptyset$ has dimension $|\mathcal{F}| - 1$: it is the kernel of a surjective linear functional on the $|\mathcal{F}|$ coordinates it occupies, so rank-nullity gives $\dim D_{\mathcal{F}} + 1 = |\mathcal{F}|$. Conversely, if both inequalities in Eq. (12) are strict, the profile $q_i = \underline{q}_i + \theta(\bar{q}_i - \underline{q}_i)$ with $\theta = (Q^{\text{kink}}(y) - Q^{\text{cour}})/(Q^{\text{free}} - Q^{\text{cour}}) \in (0, 1)$ lies in $\mathcal{E}(y)$ and is interior to the box at every $i \in \mathcal{F}$. Moving from it by $\delta(e_i - e_{i_0})$ for small $\delta > 0$ keeps both the box constraints and the clearing constraint, so each $e_i - e_{i_0}$ with $i, i_0 \in \mathcal{F}$ is a chord; and those generate $D_{\mathcal{F}}$, since any $v \in D_{\mathcal{F}}$ equals $\sum_{i \in \mathcal{F}} v_i(e_i - e_{i_0})$ once $\sum_i v_i = 0$. The two spaces therefore coincide and $\dim \mathcal{E}(y) = |\mathcal{F}| - 1$.

For Eq. (14), clearing gives $q_i = Q^{\text{kink}}(y) - \sum_{j \neq i} q_j$ while $\sum_{j \neq i} q_j$ ranges over $[Q^{\text{cour}} - \underline{q}_i, Q^{\text{free}} - \bar{q}_i]$; the lower end of that range is attained by putting every other firm at $\underline{q}_j$, the upper end by putting every one at $\bar{q}_j$, and every value between by interpolating linearly, which keeps each q_j inside its own interval. Intersecting with $q_i \in [\underline{q}_i, \bar{q}_i]$ gives Eq. (14), and the interpolation shows every point of it is attained.

For the segment, if $\mathcal{F} = \{i, j\}$ the remaining coordinates are pinned at $q_k = \bar{q}_k$ and $q_j = S - q_i$; feasibility of both coordinates is $q_i \in [\underline{q}_i, \bar{q}_i]$ and $S - q_i \in [\underline{q}_j, \bar{q}_j]$, which is $q_i \in [\alpha, \beta]$ — Eq. (14) written out. $\square$

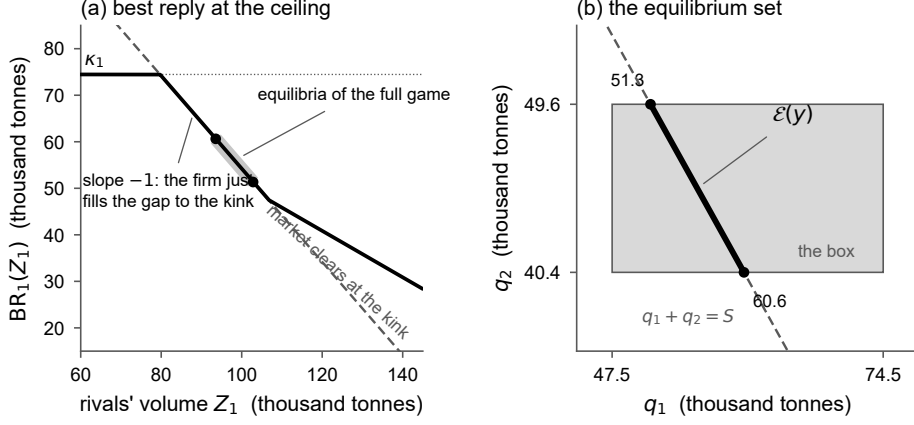


Fig. 1 The mechanism and its consequence, on the calibrated instance of Section 4.1. **a** Firm 1's best reply as a function of the volume Z_1 it faces from all other sources, from Eq. (15): flat at capacity, then a branch of slope exactly -1 on which the firm does nothing but fill the gap to the kink, then the uncapped Cournot reply. The dashed line is the locus on which the market clears at the kink, $q_1 = (a - \bar{p})/b - Z_1$; the middle branch *coincides* with it, which is why the fixed point is not isolated. The shaded stretch and its two endpoints are the profiles that are equilibria of the full four-firm game. **b** The same set in the (q_1, q_2) plane: the box $[\underline{q}_1, \bar{q}_1] \times [\underline{q}_2, \bar{q}_2]$ intersected with the clearing hyperplane $q_1 + q_2 = S$, $S = 100,981.35$, firms 3 and 4 being pinned at capacity. Quantities in thousands of tonnes per month

Lemma 2 (Best replies and the source of the continuum). *Fix q_{-i} , write $Z_i = \sum_{j \neq i} q_j + y$ and $\hat{q}_i = [(a - \bar{p})/b - Z_i]_+$ for the own output at which the ceiling stops binding. The best reply is single-valued and equals*

$$BR_i(Z_i) = \begin{cases} \min\{\kappa_i, (\bar{p} - c_i)_+/\gamma_i\} & \text{if } (\bar{p} - c_i)_+/\gamma_i \leq \hat{q}_i, \\ \min\{\kappa_i, \hat{q}_i\} & \text{if } v_i(Z_i) \leq \hat{q}_i \leq (\bar{p} - c_i)_+/\gamma_i, \\ \min\{\kappa_i, v_i(Z_i)\} & \text{if } \hat{q}_i \leq v_i(Z_i), \end{cases} \quad (15)$$

where $v_i(Z_i) = (a - bZ_i - c_i)_+/(2b + \gamma_i)$ is the uncapped Cournot best reply. On the middle branch, wherever $0 < \hat{q}_i \leq \kappa_i$, $BR_i = \hat{q}_i = (a - \bar{p})/b - Z_i$ is affine in Z_i with slope exactly -1 : the firm does nothing but absorb whatever the others leave between their total and the kink.

Proof. Write $A_i = a - bZ_i$ for the price at zero own output absent the ceiling, $p_i^0 = \min\{\bar{p}, A_i\}$ for the price actually prevailing there, and $f(\alpha, \beta; t) = \alpha t - \frac{\beta}{2}t^2$. For $t \geq 0$ capped revenue is a minimum by Eq. (8), and subtracting the cost leaves the payoff a

minimum of two concave quadratics:

$$\pi_i(t, q_{-i}) = \min \left\{ f(p_i^0 - c_i, \gamma_i; t), f(A_i - c_i, \gamma_i + 2b; t) \right\}. \quad (16)$$

The two branches agree at $t = \hat{q}_i$, because $p_i^0 + b\hat{q}_i = A_i$ in both regimes; the first is the smaller for $t \leq \hat{q}_i$ and the second for $t \geq \hat{q}_i$, so Eq. (16) says the payoff follows the price-taking branch below the kink and the Cournot branch above it.

A concave quadratic $f(\alpha, \beta; \cdot)$ with $\beta > 0$ attains its maximum over $[0, \kappa]$ at $\min\{\kappa, \alpha_+/\beta\}$ and at no other point, being strictly increasing to the left of that point and strictly decreasing to its right — this is Eq. (11) again, with the sign of the second factor determined by the position of $\frac{1}{2}(s+t)$ relative to α/β . If $\hat{q}_i = 0$ the ceiling is already slack at zero own output, so $p_i^0 = A_i$ and the second branch of Eq. (16) is the smaller for every $t \geq 0$, since $\gamma_i + 2b > \gamma_i$; the payoff is that branch throughout and its clamp is the third case of Eq. (15). If $\hat{q}_i > 0$ then $p_i^0 = \bar{p}$, the first branch peaks at $(\bar{p} - c_i)_+/\gamma_i$ and the second at $v_i(Z_i)$, and the three cases of Eq. (15) are the three positions of $\hat{q}_i$ relative to those two peaks.

If $(\bar{p} - c_i)_+/\gamma_i \leq \hat{q}_i$, the payoff — the first branch on $[0, \hat{q}_i]$ — rises to $(\bar{p} - c_i)_+/\gamma_i$ and falls after it, and the second branch is decreasing on $[\hat{q}_i, \kappa_i]$ because $\gamma_i \hat{q}_i \geq (\bar{p} - c_i)_+$ and $A_i = \bar{p} + b\hat{q}_i$ together give $v_i(Z_i) \leq \hat{q}_i$. If $\hat{q}_i \leq v_i(Z_i)$, the second branch is still rising at $\hat{q}_i$ and the first is rising on all of $[0, \hat{q}_i]$, so the maximizer is $v_i(Z_i)$. In the remaining case the first branch is still rising at $\hat{q}_i$ while the second is already falling there, so the junction itself is the maximizer. Strictness of the two monotonicities makes it unique in each case, and the capacity bound truncates from above throughout.

On the middle branch the maximizer is $\min\{\kappa_i, \hat{q}_i\}$. Where $0 < \hat{q}_i \leq \kappa_i$ this is $\hat{q}_i = (a - \bar{p})/b - Z_i$, affine in Z_i with slope -1 ; where capacity binds the branch is flat at κ_i , and where $\hat{q}_i = 0$ it is flat at 0.

Finally BR_i is nonincreasing in Z_i , so the game is one of strategic substitutes: $\hat{q}_i$ and v_i both fall as Z_i rises, and $\hat{q}_i < (\bar{p} - c_i)_+/\gamma_i$ forces $v_i(Z_i) \leq (\bar{p} - c_i)_+/\gamma_i$, so the three branches are traversed in decreasing order and the value never rises. $\square$

Remark 3 (Why the fixed point is not isolated). Lemma 2 identifies the mechanism. In a smooth Cournot game best replies have slope strictly between -1 and 0 , the aggregate map is a contraction, and the fixed point is isolated. Under a binding ceiling the middle branch of Eq. (15) has slope exactly -1 below capacity, so along it the aggregate map is the identity: at a fixed point where two firms sit on that branch, shifting output from one to the other leaves both at their best replies and every other firm's reply unchanged, so fixed points come in continua. Substituting the equilibrium condition $Z_i = Q^{\text{kink}}(y) + y - q_i$ into the middle branch of Eq. (15) recovers $\hat{q}_i = q_i$, and its two side conditions become exactly $\underline{q}_i \leq q_i \leq \bar{q}_i$, which is Proposition 2 again. The continuum is not a degeneracy of the linear specification: it is what a truncated price does to the slope of a best reply, and any inverse demand flat at the ceiling produces it.

Remark 4 (Relation to concave and aggregative games). The game is concave in own strategy, hence a concave game in the sense of Rosen (1965); his diagonal strict concavity condition, however, does not hold, since Proposition 2 exhibits a continuum.

It is also aggregative, each payoff depending on own output and on the total, and so is a nice aggregative game in the sense of [Acemoglu and Jensen \(2013\)](#), whose aggregate comparative statics is the natural setting for [Proposition 3](#); the feature of interest here is not that the aggregate responds predictably but that over an entire regime it does not respond at all. Uniqueness of the aggregate suffices for any application depending on the profile only through Q^* , as a model of physical shortage does; one whose objective depended on which firm supplied would require an equilibrium selection. [Corollary 1](#) makes that selection problem finite-dimensional and convex: $\mathcal{E}(y)$ is nonempty and compact, so any continuous selection criterion attains its optimum on it.

3.3 Pass-through

Because Q^* is piecewise linear in y and kinked where the regime changes, we define the pass-through of a public supply into total supply $Q^*(y) + y$ over an interval rather than at a point:

$$\Lambda(y', y'') = \frac{[Q^*(y'') + y''] - [Q^*(y') + y']}{y'' - y'}, \quad y'' > y'. \quad (17)$$

Proposition 3 (Exact pass-through). *Let $y' < y''$ and suppose [Assumption 1](#) holds at every $y \in [y', y'']$. Then on that interval $Q^*(y) = \min\{Q^{\text{free}}, Q^{\text{kink}}(y)\}$ and $\Lambda(y', y'') \in [0, 1]$. Moreover the two bounds are attained and nothing between them is, on any interval that stays within one regime:*

- (i) *if $Q^{\text{free}} + y'' \leq (a - \bar{p})/b$, then $Q^* \equiv Q^{\text{free}}$ on $[y', y'']$ and $\Lambda = 1$. In particular $\Lambda = 1$ whenever $\bar{p} \leq c_i$ for every i , where $Q^{\text{free}} = 0$ and the firms have exited;*
- (ii) *if $Q^{\text{free}} \geq Q^{\text{kink}}(y)$ at both endpoints, then $Q^*(y) + y = (a - \bar{p})/b$ at each, and $\Lambda = 0$: the public supply is displaced one for one.*

Proof. The representation is [Lemma 1](#). For (i), $Q^{\text{free}} + y'' \leq (a - \bar{p})/b$ says $Q^{\text{free}} \leq Q^{\text{kink}}(y)$ at every $y \leq y''$, so [Lemma 1](#) makes the equilibrium unique and equal to $\bar{q}$, the price-taking optimum at $\bar{p}$, in which y does not appear; so $Q^* = Q^{\text{free}}$ is constant in y and $Q^* + y$ rises one for one. For (ii), [Lemma 1](#) pins the aggregate at $Q^{\text{kink}}(y)$, so $Q^*(y) + y = (a - \bar{p})/b$ takes the same value at y' and y'' . For the bounds, [Eq. \(9\)](#) gives

$$Q^*(y) + y = \min \{ Q^{\text{free}} + y, (a - \bar{p})/b \},$$

which is nondecreasing in y and increases at rate at most one. $\square$

Corollary 2. *The regime switches at $y^\circ = (a - \bar{p})/b - Q^{\text{free}}$. For $y < y^\circ$ pass-through is 1; for $y > y^\circ$ it is 0. An interval straddling y° returns an average of the two, and any reported pass-through strictly between 0 and 1 under [Assumption 1](#) is such an average and not a marginal rate.*

Under [Assumption 1](#) a release is therefore either wholly displaced by private withdrawal or wholly additional, according to whether the firms' own optimum reaches the kink. The dichotomy is a statement about the aggregate only, and by [Lemma 1](#) the

aggregate is unique even where the profile is not — so the comparative static is well defined on the whole of the continuum, without an equilibrium selection.

3.4 The ceiling as a parameter

So far $\bar{p}$ has been fixed. Treating it as the regulator's instrument turns Lemma 1 into a complete description of how supply and the equilibrium set respond to the cap. Throughout this section y is fixed and we write $Q^{\text{free}}(\bar{p})$, $Q^{\text{cour}}(\bar{p})$, $q_i(\bar{p})$, $\bar{q}_i(\bar{p})$ and $Q^{\text{kink}}(\bar{p}) = (a - \bar{p})/b - y$ to show the dependence. Recall from Section 2 that Q^{free} and Q^{cour} are nondecreasing and continuous in $\bar{p}$ while Q^{kink} is strictly decreasing, so each of $Q^{\text{free}} - Q^{\text{kink}}$ and $Q^{\text{cour}} - Q^{\text{kink}}$ is continuous and strictly increasing and has at most one zero.

Lemma 3 (The binding ceiling, on primitives). *$Q^{\text{cour}} - Q^{\text{kink}}$ is continuous and strictly increasing in $\bar{p}$. Hence Assumption 1 is downward closed: if it holds at a ceiling it holds at every lower one. If moreover $Q^{\text{cour}}(\bar{p}) \leq Q^{\text{kink}}(\bar{p})$ at some $\bar{p}$ and $Q^{\text{cour}}(\bar{p}) \geq Q^{\text{kink}}(\bar{p})$ at some larger one, there is a unique p^u with $Q^{\text{cour}}(p^u) = Q^{\text{kink}}(p^u)$, and Assumption 1 holds at $\bar{p}$ if and only if $\bar{p} \leq p^u$. That p^u is the equilibrium price of the same game with the ceiling removed.*

Proof. $Q^{\text{cour}}(\bar{p}) = \sum_i \min\{\kappa_i, (\bar{p} - c_i)_+ / (b + \gamma_i)\}$ is a finite sum of continuous non-decreasing functions of $\bar{p}$, and $Q^{\text{kink}}(\bar{p}) = (a - \bar{p})/b - y$ is continuous and strictly decreasing, so the difference is continuous and strictly increasing. A strictly increasing function has at most one zero and the set on which it is nonpositive is an initial segment, which is the downward closure and the equivalence $\bar{p} \leq p^u$. Existence of the zero on the bracketing interval is the intermediate value theorem.

For the identification of p^u : in the uncapped game payoffs are smooth and concave, and firm i 's first-order condition at price p reads $p - c_i - (b + \gamma_i)q_i = 0$ where interior, so $q_i = \min\{\kappa_i, (p - c_i)_+ / (b + \gamma_i)\} = \underline{q}_i(p)$; market clearing is $p = a - b(\sum_i q_i + y)$, that is $\sum_i q_i = Q^{\text{kink}}(p)$. The uncapped equilibrium price is therefore a root of $Q^{\text{cour}} - Q^{\text{kink}}$, and there is only one. $\square$

Proposition 4 (Supply is single-peaked in the ceiling). *Fix $y \geq 0$ and suppose $Q^{\text{free}}(\bar{p}) \leq Q^{\text{kink}}(\bar{p})$ for some $\bar{p} \leq a$. At $\bar{p} = a$ we have $Q^{\text{kink}} = -y \leq 0 \leq Q^{\text{free}}$, so by the intermediate value theorem and strict monotonicity of $Q^{\text{free}} - Q^{\text{kink}}$ the root*

$$\bar{p}^\circ : \quad Q^{\text{free}}(\bar{p}^\circ) = Q^{\text{kink}}(\bar{p}^\circ) \quad (18)$$

exists and is unique; and since $Q^{\text{cour}} \leq Q^{\text{free}}$, the same two ceilings bracket a zero of $Q^{\text{cour}} - Q^{\text{kink}}$, so p^u of Lemma 3 exists as well. (The hypothesis is mild: Q^{free} vanishes at every $\bar{p} \leq \min_i c_i$, so it holds as soon as $\min_i c_i < a - by$, which is just the requirement that some firm be willing to supply at a price the market will pay.) Then

$\bar{p}^\circ \leq p^u$, and total supply as a function of the ceiling is

$$Q^*(\bar{p}) + y = \begin{cases} Q^{\text{free}}(\bar{p}) + y, & \bar{p} \leq \bar{p}^\circ, \\ (a - \bar{p})/b, & \bar{p}^\circ \leq \bar{p} \leq p^u, \\ (a - p^u)/b, & \bar{p} \geq p^u. \end{cases} \quad (19)$$

It is therefore nondecreasing on $(0, \bar{p}^\circ]$, strictly decreasing on $[\bar{p}^\circ, p^u]$ and constant thereafter, and it attains its maximum over all ceilings at $\bar{p} = \bar{p}^\circ$, where its value is $(a - \bar{p}^\circ)/b$. Which ceilings share that maximum turns on whether any firm is still flexible at $\bar{p}^\circ$, and the two cases differ in kind. Writing $\mathcal{M}$ for the set of supply-maximizing ceilings:

(a) if $Q^{\text{cour}}(\bar{p}^\circ) < Q^{\text{free}}(\bar{p}^\circ)$ — equivalently $\bar{p}^\circ < p^u$ — then

$$\mathcal{M} = \{ \bar{p} \leq \bar{p}^\circ : Q^{\text{free}}(\bar{p}) = Q^{\text{free}}(\bar{p}^\circ) \}, \quad (20)$$

an interval with upper endpoint $\bar{p}^\circ$; so $\bar{p}^\circ$ is the largest supply-maximizing ceiling, and the maximum strictly exceeds the supply $(a - p^u)/b$ prevailing with no ceiling at all;

(b) if $Q^{\text{cour}}(\bar{p}^\circ) = Q^{\text{free}}(\bar{p}^\circ)$ — equivalently $\bar{p}^\circ = p^u$, every firm being pinned by capacity or excluded by cost at $\bar{p}^\circ$ — then supply is flat from $\bar{p}^\circ$ upward,

$$\mathcal{M} = \{ \bar{p} \leq \bar{p}^\circ : Q^{\text{free}}(\bar{p}) = Q^{\text{free}}(\bar{p}^\circ) \} \cup [\bar{p}^\circ, \infty),$$

which is unbounded above: capping buys nothing that not capping does not buy, and $\bar{p}^\circ$ is not the largest maximizer.

In case (a), $\mathcal{M}$ degenerates to the single point $\bar{p}^\circ$ if and only if Q^{free} is strictly increasing at $\bar{p}^\circ$ from the left — equivalently, if and only if $c_i < \bar{p}^\circ$ and $(\bar{p}^\circ - c_i)/\gamma_i \leq \kappa_i$ for some i , that is, if and only if some firm's competitive supply at $\bar{p}^\circ$ is positive and set by price rather than held down by capacity. The inequality is weak: a firm whose competitive supply exactly reaches capacity at $\bar{p}^\circ$ is still expanding immediately to the left of it.

Proof. $Q^{\text{cour}} \leq Q^{\text{free}}$ by Eq. (6), so $Q^{\text{cour}} - Q^{\text{kink}} \leq Q^{\text{free}} - Q^{\text{kink}}$ pointwise; both are strictly increasing, so the root of the first is at least the root of the second, that is $p^u \geq \bar{p}^\circ$. For $\bar{p} \leq p^u$ the ceiling binds by Lemma 3 and Lemma 1 gives

$$Q^*(\bar{p}) + y = \min \{ Q^{\text{free}}(\bar{p}) + y, (a - \bar{p})/b \}. \quad (21)$$

A minimum of a nondecreasing and a strictly decreasing function is single-peaked, which is the shape; the two arguments cross at $\bar{p}^\circ$, which gives the first two branches of Eq. (19). For $\bar{p} > p^u$ the ceiling is slack, and part (ii) of Proposition 5 below, whose proof uses nothing from this one, shows that every equilibrium is $\underline{q}(p^u)$, whatever the ceiling. Its total is $Q^{\text{cour}}(p^u) = Q^{\text{kink}}(p^u)$, so supply is constant at $(a - p^u)/b$, the

value of the second branch at p^u ; Eq. (19) is therefore continuous, and monotonicity on each branch is immediate.

For the maximizer set, strict monotonicity of $Q^{\text{free}} - Q^{\text{kink}}$ gives $Q^{\text{free}}(\bar{p}) \leq Q^{\text{kink}}(\bar{p})$ for every $\bar{p} \leq \bar{p}^\circ$, so Eq. (21) reads $Q^*(\bar{p}) + y = Q^{\text{free}}(\bar{p}) + y$ there, and among those ceilings the maximizers are exactly the $\bar{p}$ with $Q^{\text{free}}(\bar{p}) = Q^{\text{free}}(\bar{p}^\circ)$ — an interval because Q^{free} is nondecreasing, closed at its lower end because Q^{free} is continuous. On $(\bar{p}^\circ, p^u]$ supply is $(a - \bar{p})/b < (a - \bar{p}^\circ)/b$, strictly below the maximum. Above p^u it is *constant* at $(a - p^u)/b$, which is strictly below the maximum if $\bar{p}^\circ < p^u$ and equal to it if $\bar{p}^\circ = p^u$; this last branch is the one that must not be compared with $(a - \bar{p})/b$, which it exceeds. That is the dichotomy, and its trigger is $Q^{\text{cour}}(\bar{p}^\circ)$ against $Q^{\text{free}}(\bar{p}^\circ)$: since $Q^{\text{free}}(\bar{p}^\circ) = Q^{\text{kink}}(\bar{p}^\circ)$, the gap $Q^{\text{cour}} - Q^{\text{kink}}$ is negative at $\bar{p}^\circ$ exactly when $Q^{\text{cour}}(\bar{p}^\circ) < Q^{\text{free}}(\bar{p}^\circ)$, and by strict monotonicity that is exactly $\bar{p}^\circ < p^u$.

For the degeneracy criterion, $\mathcal{M} = \{\bar{p}^\circ\}$ in case (a) exactly when $Q^{\text{free}}(\bar{p}) < Q^{\text{free}}(\bar{p}^\circ)$ for every $\bar{p} < \bar{p}^\circ$. Every summand of $Q^{\text{free}}(\bar{p}) = \sum_i \min\{\kappa_i, (\bar{p} - c_i)_+/\gamma_i\}$ is nondecreasing, so this holds exactly when some summand is strictly increasing to the left of $\bar{p}^\circ$. Summand i is *not*: it is identically zero if $c_i \geq \bar{p}^\circ$, and identically κ_i on $[c_i + \gamma_i \kappa_i, \infty) \ni \bar{p}^\circ$ if $(\bar{p}^\circ - c_i)/\gamma_i > \kappa_i$. Otherwise $c_i < \bar{p}^\circ$ and $(\bar{p}^\circ - c_i)/\gamma_i \leq \kappa_i$, the summand equals $(\bar{p}^\circ - c_i)/\gamma_i$ at $\bar{p}^\circ$, and at any $\bar{p} < \bar{p}^\circ$ it is at most $(\bar{p} - c_i)_+/\gamma_i < (\bar{p}^\circ - c_i)/\gamma_i$. Finally $\bar{p}^\circ < p^u$ gives $(a - \bar{p}^\circ)/b > (a - p^u)/b$. $\square$

That a suitably chosen cap raises output is the qualitative conclusion of the price-cap literature (Earle et al, 2007; Grimm and Zöttl, 2010). Proposition 4 makes it exact and, more to the point, locates the optimum: the supply-maximizing ceiling is the one at which the firms' competitive supply exactly clears kink demand, so that the two constraints on supply — willingness and demand at the capped price — bind simultaneously. Below $\bar{p}^\circ$ the cap suppresses willingness to supply; above it, the cap is slack enough that the firms hold output down to the kink themselves. The optimum is unique only when some firm is still expanding its competitive supply at $\bar{p}^\circ$. When every firm's competitive supply is already held at capacity there — as in the calibrated instance of Section 4.1 — tightening the cap below $\bar{p}^\circ$ leaves supply unchanged until the cap reaches $\max_i(c_i + \gamma_i \kappa_i)$, and a whole interval of ceilings is supply-maximizing. Whether $\bar{p}^\circ$ is the *largest* of them is a separate question with a separate answer, and it is the one Proposition 5 turns on. If some firm is flexible at $\bar{p}^\circ$ — $\underline{q}_i(\bar{p}^\circ) < \bar{q}_i(\bar{p}^\circ)$ for some i , which is case (a) — then $\bar{p}^\circ < p^u$ and $\bar{p}^\circ$ is the last ceiling at which supply has not yet begun to fall and the first at which the equilibrium set is about to open up. If no firm is, the two thresholds coincide, supply is flat from $\bar{p}^\circ$ upward, and every higher ceiling is supply-maximizing too — for the uninteresting reason that the cap constrains nobody. The regulator's choice is then between capping at $\bar{p}^\circ$ and not capping at all, and the two deliver the same output.

Proposition 5 (Multiplicity occupies exactly one interval). *Under the hypotheses of Proposition 4:*

- (i) for $\bar{p} \leq \bar{p}^\circ$ the equilibrium is unique and equals $\bar{q}(\bar{p})$;
- (ii) for $\bar{p} \geq p^u$ the equilibrium is unique and equals $\underline{q}(p^u)$, the uncapped equilibrium;

(iii) for $\bar{p} \in (\bar{p}^\circ, p^u)$ the equilibrium set is the polytope of Corollary 1, of dimension $|\mathcal{F}(\bar{p})| - 1$ where $\mathcal{F}(\bar{p}) = \{i : \underline{q}_i(\bar{p}) < \bar{q}_i(\bar{p})\}$. It is a continuum if and only if at least two firms are pinned neither by capacity nor by cost at $\bar{p}$, that is $|\mathcal{F}(\bar{p})| \geq 2$.

Consequently, when at least two firms are unpinned at $\bar{p}^\circ$ and at every ceiling from there to p^u , the set of ceilings at which the equilibrium is not unique is exactly the open interval $(\bar{p}^\circ, p^u)$, which is then nonempty; its lower endpoint is, by case (a) of Proposition 4, the largest supply-maximizing ceiling.

Proof. (i) $\bar{p} \leq \bar{p}^\circ$ gives $Q^{\text{free}}(\bar{p}) \leq Q^{\text{kink}}(\bar{p})$ by strict monotonicity of $Q^{\text{free}} - Q^{\text{kink}}$, and Lemma 1 then delivers uniqueness at $\bar{q}(\bar{p})$.

(iii) For $\bar{p} \in (\bar{p}^\circ, p^u)$ strict monotonicity gives $Q^{\text{cour}}(\bar{p}) < Q^{\text{kink}}(\bar{p}) < Q^{\text{free}}(\bar{p})$, so both inequalities of Eq. (12) are strict and Eq. (13) applies. The dimension is at least one exactly when $|\mathcal{F}(\bar{p})| \geq 2$; when $|\mathcal{F}(\bar{p})| \leq 1$ the coordinates outside $\mathcal{F}(\bar{p})$ are fixed and the clearing condition then fixes the remaining one.

(ii) Let $\bar{p} \geq p^u$, so $Q^{\text{kink}}(\bar{p}) \leq Q^{\text{cour}}(\bar{p})$ by Lemma 3. An equilibrium exists, and without Proposition 1: $\underline{q}(p^u)$ is the equilibrium of the uncapped game (Lemma 3), its price is $p^u \leq \bar{p}$, so the cap leaves its payoffs as they are, while capping can only lower the payoff of a deviation. Let q be any equilibrium, with total Q and prevailing price $p = a - b(Q + y)$. Step (b) of Lemma 1 gives $Q \geq Q^{\text{kink}}(\bar{p})$.

If $Q = Q^{\text{kink}}(\bar{p})$ then $q_i \geq \underline{q}_i(\bar{p})$ for every i by the necessity half of the proof of Proposition 2, which uses the clearing condition $Q = Q^{\text{kink}}(\bar{p})$ and not Assumption 1. So $Q \geq Q^{\text{cour}}(\bar{p}) \geq Q^{\text{kink}}(\bar{p}) = Q$ and the three are equal; coordinates bounded below by the $\underline{q}_i(\bar{p})$ with that sum must equal them, so $q = \underline{q}(\bar{p})$ and $Q^{\text{cour}}(\bar{p}) = Q^{\text{kink}}(\bar{p})$, that is $\bar{p} = p^u$.

If $Q > Q^{\text{kink}}(\bar{p})$ then step (a) of Lemma 1 gives $q_i \leq \underline{q}_i(\bar{p})$ for every i . Also $\hat{q}_i \leq q_i$, since $\hat{q}_i = [Q^{\text{kink}}(\bar{p}) - Q + q_i]_+$ and $Q \geq Q^{\text{kink}}(\bar{p})$; hence $(b + \gamma_i)\hat{q}_i \leq (b + \gamma_i)q_i \leq (\bar{p} - c_i)_+$. Since $A_i = \bar{p} + b\hat{q}_i$ whenever $\hat{q}_i > 0$, this says the Cournot branch of Eq. (16) peaks at or above the junction $\hat{q}_i$; its clamped maximizer therefore lies in $[\hat{q}_i, \kappa_i]$, where capped and uncapped payoffs agree, so it also maximizes the capped payoff, and by the uniqueness in Lemma 2 it equals q_i . Written out, and using $A_i = p + bq_i$,

$$q_i = \min \left\{ \kappa_i, \frac{(p + bq_i - c_i)_+}{\gamma_i + 2b} \right\}. \quad (22)$$

The right-hand side of Eq. (22), read as a map of q_i , is nondecreasing and Lipschitz with constant $b/(\gamma_i + 2b) < 1$, because $t \mapsto (t)_+$ and $t \mapsto \min\{\kappa_i, t\}$ are both 1-Lipschitz; so it has at most one fixed point. And $\underline{q}_i(p)$ is one: if $\underline{q}_i(p) = \kappa_i$ then $(b + \gamma_i)\kappa_i \leq (p - c_i)_+$ forces $(\gamma_i + 2b)\kappa_i \leq p + b\kappa_i - c_i$, and if $\underline{q}_i(p) < \kappa_i$ then $(b + \gamma_i)\underline{q}_i(p) = (p - c_i)_+$ gives $p + b\underline{q}_i(p) - c_i = (\gamma_i + 2b)\underline{q}_i(p)$. Hence $q_i = \underline{q}_i(p)$ for every i , so $Q = Q^{\text{cour}}(p)$, while $p = a - b(Q + y)$ says $Q = Q^{\text{kink}}(p)$; the price is therefore the root of $Q^{\text{cour}} - Q^{\text{kink}}$, which is p^u .

Either way $q = \underline{q}(p^u)$: the equilibrium is unique and does not move with the ceiling once the ceiling is slack. No appeal to diagonal strict concavity is made. $\square$

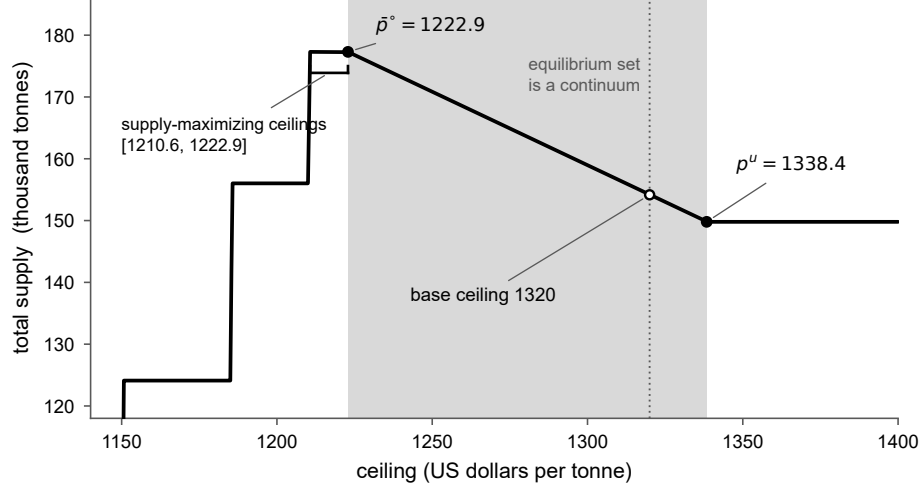


Fig. 2 Total supply as a function of the ceiling, on the calibrated instance of Section 4.1, from Eq. (19). Supply rises in steps as each firm’s competitive supply reaches its capacity, peaks where that competitive supply exactly clears kink demand, then falls along $(a - \bar{p})/b$ until the ceiling ceases to bind at p^u , and is flat thereafter. The shaded band is the interval $(\bar{p}^o, p^u)$ on which the equilibrium is a continuum (Proposition 5); the base ceiling of 1320 sits inside it. Because every firm’s competitive supply is held at capacity at $\bar{p}^o$ in this instance, the supply-maximizing ceilings form the bracketed interval $[1210.6, 1222.9]$ rather than a single point, and $\bar{p}^o$ is its upper endpoint. Supply in thousands of tonnes per month

Remark 5 (The two endpoints are different kinds of point). At $\bar{p}^o$ the polytope collapses because Q^{kink} meets the *upper* face of the box, forcing $q = \bar{q}$: every firm is at its competitive supply and the cap is exactly tight enough to absorb it. At p^u it collapses because Q^{kink} meets the *lower* face, forcing $q = \underline{q}$: every firm is at its Cournot reply and the cap is exactly on the point of ceasing to bind. Between the two faces the clearing hyperplane cuts the interior of the box, and it is the thickness of the box — the wedge $\bar{q}_i - \underline{q}_i$, which Remark 2 traces to the derivative jump at the kink — that the equilibrium set inherits. Both vanish with b , at a rate one can write down:

$$\bar{q}_i - \underline{q}_i \leq \frac{b(\bar{p} - c_i)_+}{\gamma_i(b + \gamma_i)}, \quad p^u - \bar{p}^o \leq b(Q^{\text{free}}(p^u) - Q^{\text{cour}}(p^u)),$$

so the two thresholds are $O(b^2)$ apart. As $b \downarrow 0$ the box degenerates, $\bar{p}^o$ and p^u merge, and multiplicity disappears along with the strategic effect that produced it.

4 Numerical example

This section does two things. It checks the theory of Sections 3.2 and 3.4 against an instance drawn from the application sketched in Section 1, where the predictions are sharp enough to be falsified by a number. And it reports what happens when equilibria of that instance are computed rather than derived, which is the observation that prompted the theory.

4.1 The instance

The instance is stated in tonnes per month and US dollars per tonne, and has $n = 4$. Demand at the base ceiling $\bar{p}_1 = 1320$ is $D = 154,170$ and the firms' shares of it are $(0.42, 0.28, 0.18, 0.12)$; the demand slope is $b = 0.0042$ and the intercept is $a = \bar{p}_1 + bD = 1967.514$, so that $Q^{\text{kink}}(0) = D$ exactly at the base ceiling. Base marginal costs are $(1120, 1150, 1185, 1210)$, escalated as $c_{i,z} = c_i(1 + e_z)$ with $e_z \in \{0, 0.10, 0.25\}$ across three regimes, against a ceiling $\bar{p}_z = 1320(1 + \frac{1}{2}e_z)$, so $\bar{p} = (1320, 1386, 1485)$. Cost curvature is $\gamma_i = 0.55/(Ds_i)$ and capacity is $\kappa_i = \theta_z Ds_i$ with $\theta = (1.15, 0.75, 0.35)$. The public supply is $y = 0$ throughout. Every number in Sections 4.2 and 4.3 follows from these primitives in closed form, and every count in Section 4.4 is computed from them; nothing is fitted.

In the first regime $\sum_i \kappa_i = 177,296$ exceeds $Q^{\text{kink}} = 154,170$, so the ceiling binds through the kink and Proposition 2 applies. In the remaining two the capacity sum falls below the kink and the total is set by capacity; in the third the ceiling has in addition fallen below firm 4's cost, and that firm exits.

4.2 The polytope, checked

In regime 1 the bounds of Eqs. (4) and (5) are

$$\begin{aligned}\bar{q} &= (74,464.11, 49,642.74, 31,913.19, 21,275.46), \\ \underline{q} &= (47,522.94, 40,353.77, 31,913.19, 21,275.46),\end{aligned}$$

so $Q^{\text{cour}} = 141,065.36 \leq Q^{\text{kink}} = 154,170 \leq Q^{\text{free}} = 177,295.50$: both inequalities of Eq. (12) hold strictly, and the equilibrium set is a nonempty polytope. Firms 3 and 4 have $\underline{q}_i = \bar{q}_i = \kappa_i$ — they are pinned at capacity even at their Cournot reply — so $\mathcal{F} = \{1, 2\}$ and Eq. (13) predicts $\dim \mathcal{E} = 1$: a segment. Solving for it, firms 3 and 4 supply 53,188.65 between them, firms 1 and 2 share the residual 100,981.35, and

$$q_1 \in [51,338.61, 60,627.58], \quad q_2 \in [40,353.77, 49,642.74],$$

a segment of length 9,288.97 in each coordinate — 6.03 per cent of total supply — and Euclidean diameter 13,136.58. This is the exact prediction against which the computational results below are read.

4.3 The supply-maximizing ceiling

Regime 1 also makes Propositions 4 and 5 concrete. Solving Eq. (18) gives $\bar{p}^\circ = 1222.87$, at which the equilibrium is unique and total supply is 177,295.50; the uncapped equilibrium price of Lemma 3 is $p^u = 1338.38$, with supply 149,794.43. The base ceiling of 1320 lies inside $(\bar{p}^\circ, p^u) = (1222.87, 1338.38)$, which is why regime 1 exhibits a continuum at all — and it lies well above $\bar{p}^\circ$, so tightening the cap from 1320 to 1222.87 would raise equilibrium supply by 23,126 tonnes per month, or 15.0 per cent, while simultaneously restoring uniqueness.

The instance is of the degenerate kind Proposition 4 allows for, and is worth displaying as such. Here $\gamma_i \kappa_i = 0.55 \times 1.15 = 0.6325$ for every firm, so $c_i + \gamma_i \kappa_i =$

(1120.63, 1150.63, 1185.63, 1210.63) and every firm’s competitive supply is held at capacity at every ceiling above 1210.63. No firm is *pinned* at $\bar{p}^\circ$ in the sense of Section 2, however: each firm’s Cournot reply there is below its capacity, so the instance is in case (a) of Proposition 4, with $\bar{p}^\circ < p^u$. Hence Q^{free} is constant on $[1210.63, \bar{p}^\circ]$ and supply is maximal, at 177,295.50, at *every* ceiling in that interval: the set Eq. (20) is $[1210.63, 1222.87]$, not the point $\bar{p}^\circ$. What distinguishes $\bar{p}^\circ$ among them is that it is the largest, and so the only one at which the ceiling is exactly tight — which is the property Proposition 5 uses. The instance therefore sits in the interior of the region Proposition 5 identifies: a cap loose enough to leave supply on the table and tight enough to leave the split undetermined. We record this as an illustration of the propositions and not as a policy recommendation; total supply is not welfare, and the calibration is not the subject of this paper.

4.4 Certification

Equilibria of the game are found here by maximizing the Nikaidô–Isoda gap function (Nikaidô and Isoda, 1955)

$$F(q) = \sum_{i=1}^n \left[\pi_i(q) - \max_{q'_i \in [0, \kappa_i]} \pi_i(q'_i, q_{-i}) \right] \leq 0, \quad (23)$$

which vanishes exactly on the equilibrium set. This is the continuous form of the reduction of Enkhbat et al (2017), which extends Mills (1960) to n players and makes a Nash equilibrium the global solution of a single nonconvex maximization; global optimality conditions for problems of this class are given by Strekalovsky (1998) and multistart schemes for them by Gornov and Zarodnyuk (2014). The inner maximum in Eq. (23) is one-dimensional and available in closed form on each side of the kink — it is Lemma 2 — so F is evaluated exactly rather than approximately, and the value reported for a candidate profile is an exact error bound and not an estimate. Every figure in this subsection is produced from the primitives of Section 4.1 and nothing else, by two scripts available from the author on request: one in Python and one in R, sharing no code. Random starts are drawn not from either language’s generator but from the MINSTD recurrence $x \mapsto 16807x \bmod (2^{31} - 1)$, whose every product is below 2^{53} and therefore exact in double precision. The stream is thus the same in both — and reconstructible in three lines in any language, which is what makes the counts below reproducible rather than merely reported. The two agree on every figure in this subsection except one, and that exception is itself an observation; we come to it.

Recasting an equilibrium problem as the optimisation of a Nikaidô–Isoda function is standard for generalised Nash equilibrium problems, in which the constraints are shared (Facchinei and Kanzow, 2007); regularised variants give smooth reformulations (von Heusinger and Kanzow, 2009), and Panicucci et al (2009) solve the resulting problem by global minimisation, as we do here. Our game is simpler in that its constraints are not shared, and harder in that the payoff is nonsmooth exactly where the equilibrium sits.

The use we make of Eq. (23) is that of a merit function rather than of an algorithm. A merit function vanishes precisely on the solution set and is computable away

from it, so its value at a candidate point is an error bound certifying, or refusing to certify, that the point solves the problem; this is the role gap functions play for variational inequalities (Fukushima, 1992) and, more generally, the role of merit functions and error bounds in complementarity and equilibrium problems (Facchinei and Pang, 2003). We evaluate F at the limit returned by each solver run and report it, rather than relying on the stopping test of the iteration that produced the point.

Table 1 Certification on the calibrated instance of Section 4.1. “BR” is damped simultaneous best response with $\lambda = 1/2$ from 100 random starts, stopped on a relative step test of 10^{-10} ; “CMS” is multistart maximization of Eq. (23) from 20 random starts. “Cert.” counts starts whose limit has relative gap $|F|/|\sum_i \max_{q'_i} \pi_i(q'_i, q_{-i})|$ below 10^{-7} . “Distinct” counts how many different profiles the 100 BR runs return, rounded to 10^{-4} . The two spread columns report the range of the total, and the largest range of a single firm’s output, across the 100 limits, each as a percentage of total supply; the last column is the width Corollary 1 predicts. The two CMS entries in the first row are the counts returned by the two reference implementations — SciPy’s L-BFGS-B and R’s `optim` with method L-BFGS-B. They differ; the 20/20 of the other two rows does not. That contrast is the third observation below.

Regime	BR		spread			CMS cert.
	cert.	Distinct	in Q	in q_i	predicted	
Ceiling binds at the kink	100/100	50	$5 \times 10^{-9}\%$	5.83%	6.03%	1/20, 8/20
Capacity binds, all active	100/100	1	$8 \times 10^{-9}\%$	$4 \times 10^{-9}\%$	0	20/20
Capacity binds, one exited	100/100	1	$1 \times 10^{-8}\%$	$5 \times 10^{-9}\%$	0	20/20

Table 2 Sensitivity of best response to the damping, in the regime where the ceiling binds through the kink. “Stop test” counts the starts at which the iteration’s own relative step test fires within 10^5 sweeps; “Cert.” counts those whose limit is an equilibrium. Two firms are unpinned here, so $|\mathcal{F}| = 2$ and Remark 6 puts the stability threshold at $\lambda = 2/|\mathcal{F}| = 1$. The last row is the undamped map.

λ	$ 1 - \lambda \mathcal{F} $	Stop test	Cert.	median rel. $ F $
1/4	0.50	100/100	100/100	4.2×10^{-11}
1/2	0	100/100	100/100	1.9×10^{-11}
3/4	0.50	100/100	100/100	7.9×10^{-12}
1	1	0/100	0/100	3.3×10^{-1}

Table 1 reports the outcome in the three configurations the calibration passes through, and Table 2 the sensitivity of the iteration to the damping. Three observations follow.

First, best response does not fail here — and that is the point. Damped simultaneous best response certifies all 100 random starts in every regime. In regimes 2 and 3 it returns the *same* profile from every start: the capacity sum falls below the kink, Lemma 1 gives a unique equilibrium at $\bar{q}$, and the iteration finds it. In regime 1 it

returns 50 different ones. Those 50 profiles agree on the total to nine significant figures — a spread of 5×10^{-9} per cent of supply, which is the arithmetic of the stopping test and not an economic quantity — and disagree on the split by up to 5.83 per cent of supply. Every one of them lies inside the segment Corollary 1 predicts, and together they cover 97 per cent of it. Run sequentially rather than simultaneously, with firms replying one at a time, the iteration converges in a median of ten sweeps and its limits cover the predicted segment *entire*, to 6.025 per cent against the predicted 6.025.

An equilibrium solver run once on this instance therefore returns an equilibrium, passes its own convergence test, is certified by Eq. (23), and gives no indication whatever that a six-per-cent band of splits is equally an equilibrium. Which member of the band it returns is a function of where it started. This is Corollary 1 observed rather than a defect of any method, and it is the sense in which the characterisation, and not the computation, is what identifies the solution set.

That sequential best response converges is what the game theory leads one to expect, though on a continuum it does not guarantee it. The capped game is aggregative and its best reply is single-valued and nonincreasing (Lemma 2), hence a pseudo-potential game in the sense of Dubey et al (2006). In a finite game of that kind Cournot tâtonnement admits no improvement cycles and therefore stops (Kukushkin, 2004), which is what Remark 7 observes on a lattice; with interval strategy sets the absence of cycles does not by itself deliver convergence, and we claim no more than the experiment shows. The slope-(-1) branch does not obstruct sequential play; what it obstructs is the *simultaneous* iteration, and exactly to the following extent.

Remark 6 (The stability threshold of simultaneous best response). Let $\mathcal{F}$ be the set of firms sitting on the middle branch of Eq. (15) below capacity, $0 < \hat{q}_i \leq \kappa_i$, along the iteration, and consider the damped simultaneous map $q \mapsto (1 - \lambda)q + \lambda \text{BR}(q)$. On that branch $\text{BR}_i = (a - \bar{p})/b - Z_i = Q^{\text{kink}}(y) - \sum_{j \neq i} q_j$, so each $i \in \mathcal{F}$ moves by $\lambda(Q^{\text{kink}}(y) - Q)$ and every other firm stays put. Hence

$$Q^{\text{kink}}(y) - Q^+ = (1 - \lambda|\mathcal{F}|)(Q^{\text{kink}}(y) - Q),$$

so the iteration contracts the aggregate error if and only if $0 < \lambda < 2/|\mathcal{F}|$, is marginally stable at $\lambda = 2/|\mathcal{F}|$, and diverges above it. Two consequences. The undamped map, $\lambda = 1$, fails for every $|\mathcal{F}| \geq 2$ — there is no contraction to be had from best replies of slope -1. And the admissible window shrinks like $1/|\mathcal{F}|$, so any damping fixed in advance fails once enough firms sit on the branch. Where the iteration does converge, summing the geometric series gives $q_i^\infty = q_i^0 + (Q^{\text{kink}}(y) - Q^0)/|\mathcal{F}|$ for $i \in \mathcal{F}$: the equilibrium reached is an affine function of the starting point, which is why 100 starts give 50 answers.

Second, that threshold is exactly what Table 2 measures. Two firms are unpinned in regime 1, so $|\mathcal{F}| = 2$ and the predicted threshold is $\lambda = 1$. Damping at $1/4$, $1/2$ and $3/4$ certifies every start; the undamped map certifies none. What it does instead is cycle, and exactly as predicted: at $\lambda = 1$ with $|\mathcal{F}| = 2$ the multiplier $1 - \lambda|\mathcal{F}|$ is -1, so the aggregate error flips sign at every step and is restored after two. All 100 starts are drawn into the *same* period-two orbit — every firm at capacity, alternating with every firm's reply to that — and it swings the aggregate over 38.9 per cent of supply,

straddling Q^{kink} . The failure is loud rather than silent: the iteration’s own step test never fires, within 10^5 sweeps, at any of the 100 starts, so no limit is on offer to be mistaken for a solution. Across all four damping values and all three regimes, the step test fired at a profile that was not an equilibrium exactly zero times.

Third, the merit function certifies exactly, but is hard to maximize on a flat set — and this is the one place the two implementations part company. Multistart maximization of Eq. (23) certifies 1 of 20 starts in regime 1 under SciPy’s L-BFGS-B and 8 of 20 under R’s. In regimes 2 and 3, where the solution is a point, both certify 20 of 20. That the count disagrees exactly where the equilibrium set is a segment, and agrees exactly where it is a point, is not an embarrassment of the experiment but its sharpest reading: on a set with empty interior, attaining $|F| = 0$ is a property of the optimiser’s path and not of the problem. Every profile either implementation certifies has total $Q^{\text{kink}} = 154,170$ to nine significant figures and a first-firm output inside the predicted interval $[51,338.61, 60,627.58]$ — the single SciPy one at 51,343.10, the eight from R spread over $[57,119.88, 60,626.49]$. The uncertified runs stop at relative gaps of order 10^{-4} to 10^{-3} , and it is Eq. (23), evaluated exactly, that tells them apart from the ones that are equilibria. Regimes 2 and 3 are the control, and confirm that the difficulty in regime 1 is caused by the geometry Proposition 2 describes and not by either solver.

Remark 7 (On lattice discretisation). The theorem of Enkhbat et al (2017) is stated for finite games, which suggests placing each q_i on the lattice $\{0, h, 2h, \dots\} \cap [0, \kappa_i]$ and applying it directly. That works, and what it produces is worth recording. A lattice game here is finite and aggregative, and it inherits strategic substitutes from the payoff itself: π_i has decreasing differences in (q_i, Z_i) , so against a larger Z_i no output above a maximizer against the smaller one does strictly better — whatever the strategy set, interval or lattice. The smallest best reply is therefore nonincreasing in Z_i , and so is the best reply itself wherever it is single-valued, which on a lattice fails only at exact ties. With that, the lattice game is a pseudo-potential game (Dubey et al, 2006) whose Cournot tâtonnement is acyclic (Kukushkin, 2004); on the instance of Section 4.1, sequential discrete best reply terminated at an exact pure equilibrium — discrete gap exactly zero — from every one of 200 random starts at each of $h = 1000, 100, 10$ and 1 tonne. What refinement does not do is remove the indeterminacy: it discretises it. Any lattice point of $\mathcal{E}(y)$ would be a lattice equilibrium, since shrinking the deviation set can only make a profile harder to upset. On this instance there is none, because firms 3 and 4 must sit at capacities that are not multiples of h , but the lattice equilibria line up along the segment all the same. An exact enumeration — the payoff is concave in own output, so a lattice point is a lattice best reply as soon as neither neighbour pays more — finds 3 pure equilibria at $h = 1000$, 94 at $h = 100$, 930 at $h = 10$ and 9,290 at $h = 1$: once h is small, their number grows like the width of the segment divided by h . Discretisation therefore replaces a continuum by a large finite set, and a finite-game algorithm applied to it returns one member of that set, selected by its initialisation exactly as in Section 4.4.

5 Conclusion

Four properties of Cournot competition under a binding price ceiling have been established. The ceiling preserves concavity of the payoff in own output, and hence existence, for any demand under which uncapped revenue is concave, because capping takes a minimum where flooring would take a maximum. Where the ceiling binds through the kink, the equilibrium set is a polytope: the market clears at the ceiling and each firm sits between its Cournot and its competitive supply at the ceiling price, so the aggregate is determined and the split in general is not, with an exact dimension and exact coordinate ranges. A public supply added to such a market passes through into total supply at a rate of exactly zero or exactly one, with the switch at an explicitly computable point; intermediate values are averages across that switch. And, as a function of the ceiling, total supply is single-peaked and is maximized at a ceiling $\bar{p}^\circ$ sitting exactly at the lower endpoint of the interval of ceilings outside which the equilibrium is unique — and, whenever that interval is nonempty, $\bar{p}^\circ$ is the largest ceiling that maximizes supply.

The last of these is the result we would most like to see tested outside the linear specification. It says that the multiplicity is not incidental to the policy problem but coextensive with the range of caps a regulator would actually be choosing among, and it suggests that an equilibrium-selection assumption is doing quiet work in any applied model of a capped market that reports a unique outcome.

Three questions remain open. The first is whether uniqueness of the aggregate is a general property of Cournot games under a binding ceiling with truncated strategy sets, or an artefact of linear demand. Lemma 2 suggests it is not an artefact — the slope- (-1) branch appears whenever inverse demand is flat at the ceiling, whatever it does elsewhere — but the argument in Lemma 1 does use linearity above the kink, and a general answer would say when the regulator’s problem of choosing $\bar{p}$ or y — a bilevel problem with this game as its lower level — is well posed without an equilibrium selection. The second is whether the polytope of Corollary 1 carries a selection with a game-theoretic justification rather than a computational one; the profiles at which all but one firm sits at a bound are the natural candidates, which is suggestive but not an argument. The third is computational: whether the global optimality conditions of Strekalovsky (1998) extend to that bilevel problem, with Eq. (23) as its lower-level condition, and whether the gap function can serve as a penalty in a single-level reformulation.

Statements and Declarations

Funding. The author received no funding for this work.

Competing interests. The author has no competing interests to declare that are relevant to the content of this article.

Ethics approval and consent to participate. Not applicable: no human participants, animal subjects or personal data.

Originality and prior publication. This paper is not under consideration by any other journal.

Author contributions. The paper has a sole author, who is responsible for the whole of it.

Use of generative artificial intelligence. The author used a generative AI assistant for language editing and for preparing the manuscript, including drafting and checking parts of the arguments. The author reviewed and verified all results and takes full responsibility for the content of this paper.

Data availability. No datasets were generated or analysed. The calibrated instance of Section 4.1 is stated in full in the text, and every quantity reported from it in Sections 4.2 and 4.3 follows from those primitives in closed form.

Code availability. All code underlying this paper is available from the author on formal request.

References

- Acemoglu D, Jensen MK (2013) Aggregate comparative statics. *Games and Economic Behavior* 81:27–49. <https://doi.org/10.1016/j.geb.2013.03.009>
- Amir R (1996) Cournot oligopoly and the theory of supermodular games. *Games and Economic Behavior* 15(2):132–148. <https://doi.org/10.1006/game.1996.0062>
- Corchón LC (1994) Comparative statics for aggregative games: the strong concavity case. *Mathematical Social Sciences* 28(3):151–165. [https://doi.org/10.1016/0165-4896\(94\)90001-9](https://doi.org/10.1016/0165-4896(94)90001-9)
- Cornes R, Hartley R (2012) Fully aggregative games. *Economics Letters* 116(3):631–633. <https://doi.org/10.1016/j.econlet.2012.06.024>
- Debreu G (1952) A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences* 38(10):886–893. <https://doi.org/10.1073/pnas.38.10.886>
- Dubey P, Haimanko O, Zapechelnyuk A (2006) Strategic complements and substitutes, and potential games. *Games and Economic Behavior* 54(1):77–94. <https://doi.org/10.1016/j.geb.2004.10.007>
- Earle R, Schmedders K, Tatur T (2007) On price caps under uncertainty. *The Review of Economic Studies* 74(1):93–111. <https://doi.org/10.1111/j.1467-937X.2007.00415.x>
- Enkhbat R, Batbileg S, Tungalag N, et al (2017) A computational method for solving N-person game. *The Bulletin of Irkutsk State University, Series “Mathematics”* 20:109–121. <https://doi.org/10.26516/1997-7670.2017.20.109>

- Facchinei F, Kanzow C (2007) Generalized Nash equilibrium problems. *4OR* 5(3):173–210. <https://doi.org/10.1007/s10288-007-0054-4>
- Facchinei F, Pang JS (2003) *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, <https://doi.org/10.1007/b97543>
- Fan K (1952) Fixed-point and minimax theorems in locally convex topological linear spaces. *Proceedings of the National Academy of Sciences* 38(2):121–126. <https://doi.org/10.1073/pnas.38.2.121>
- Fukushima M (1992) Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming* 53:99–110. <https://doi.org/10.1007/BF01585696>
- Glicksberg IL (1952) A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society* 3(1):170–174. <https://doi.org/10.2307/2032478>
- Gornov A, Zarodnyuk T (2014) Computing technology for estimation of convexity degree of the multiextremal function. *Machine Learning and Data Analysis* 1(10):1345–1353
- Grimm V, Zöttl G (2010) Price regulation under demand uncertainty. *The BE Journal of Theoretical Economics* 10(1). <https://doi.org/10.2202/1935-1704.1626>
- von Heusinger A, Kanzow C (2009) Optimization reformulations of the generalized Nash equilibrium problem using Nikaido–Isoda-type functions. *Computational Optimization and Applications* 43(3):353–377. <https://doi.org/10.1007/s10589-007-9145-6>
- Jensen MK (2010) Aggregative games and best-reply potentials. *Economic Theory* 43(1):45–66. <https://doi.org/10.1007/s00199-008-0419-8>
- Kolstad CD, Mathiesen L (1987) Necessary and sufficient conditions for uniqueness of a Cournot equilibrium. *The Review of Economic Studies* 54(4):681–690. <https://doi.org/10.2307/2297489>
- Kreps DM, Scheinkman JA (1983) Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics* 14(2):326–337. <https://doi.org/10.2307/3003636>
- Kukushkin NS (2004) Best response dynamics in finite games with additive aggregation. *Games and Economic Behavior* 48(1):94–110. <https://doi.org/10.1016/j.geb.2003.06.007>
- Milgrom P, Roberts J (1990) Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* 58(6):1255–1277. <https://doi.org/>

- Mills H (1960) Equilibrium points in finite games. *Journal of the Society for Industrial and Applied Mathematics* 8(2):397–402. <https://doi.org/10.1137/0108026>
- Monderer D, Shapley LS (1996) Potential games. *Games and Economic Behavior* 14(1):124–143. <https://doi.org/10.1006/game.1996.0044>
- Nie Py, Wen Hx, Wang C (2023) Price cap regulation under duopoly. *Managerial and Decision Economics* 44(3):1376–1382. <https://doi.org/10.1002/mde.3766>
- Nikaidô H, Isoda K (1955) Note on non-cooperative convex game. *Pacific Journal of Mathematics* 5(S1):807–815. <https://doi.org/10.2140/pjm.1955.5.807>
- Novshek W (1985) On the existence of Cournot equilibrium. *The Review of Economic Studies* 52(1):85–98. <https://doi.org/10.2307/2297471>
- Panicucci B, Pappalardo M, Passacantando M (2009) On solving generalized Nash equilibrium problems via optimization. *Optimization Letters* 3:419–435. <https://doi.org/10.1007/s11590-009-0122-0>
- Reny PJ (1999) On the existence of pure and mixed strategy Nash equilibria in discontinuous games. *Econometrica* 67(5):1029–1056. <https://doi.org/10.1111/1468-0262.00069>
- Reynolds SS, Rietzke D (2018) Price caps, oligopoly, and entry. *Economic Theory* 66(3):707–745. <https://doi.org/10.1007/s00199-016-0963-6>
- Roberts J, Sonnenschein H (1976) On the existence of Cournot equilibrium without concave profit functions. *Journal of Economic Theory* 13(1):112–117. [https://doi.org/10.1016/0022-0531\(76\)90069-7](https://doi.org/10.1016/0022-0531(76)90069-7)
- Roques FA, Savva N (2009) Investment under uncertainty with price ceilings in oligopolies. *Journal of Economic Dynamics and Control* 33(2):507–524. <https://doi.org/10.1016/j.jedc.2008.07.002>
- Rosen JB (1965) Existence and uniqueness of equilibrium points for concave n -person games. *Econometrica* 33(3):520–534. <https://doi.org/10.2307/1911749>
- Strekalovsky AS (1998) Global optimality conditions for nonconvex optimization. *Journal of Global Optimization* 12:415–434. <https://doi.org/10.1023/A:1008277314050>
- Szidarovszky F, Yakowitz S (1977) A new proof of the existence and uniqueness of the Cournot equilibrium. *International Economic Review* 18(3):787–789. <https://doi.org/10.2307/2525963>

- Tao Y, Wang X, Yang C (2018) Duopoly models with a joint capacity constraint. *Journal of Economics* 125(2):159–172. <https://doi.org/10.1007/s00712-018-0597-1>
- Vives X (1990) Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics* 19(3):305–321. [https://doi.org/10.1016/0304-4068\(90\)90005-T](https://doi.org/10.1016/0304-4068(90)90005-T)