SYM-OP-IS 2025
52nd Symposium on Operational Research
Palić, 7 - 10.9.2025.
UNIVERSITY OF BELGRADE
FACULTY OF ORGANIZATIONAL SCIENCES

Review article

DEMAND RATIONING RULES: BEYOND EFFICIENCY AND PROPORTIONALITY

BOJAN RISTIĆ[1], NIKOLA NJEGOVAN[2]

[1] University of Belgrade, Faculty of Economics and Business, Belgrade, bojan.ristic@ekof.bg.ac.rs, 0000-0002-9883-8914
[2] University of Belgrade, Faculty of Economics and Business, Belgrade, nikola.njegovan@ekof.bg.ac.rs, 0000-0002-9531-323

Abstract: *This paper examines demand rationing mechanisms in oligopoly models beyond the classical efficient and proportional rules, with a focus on capacity-constrained price competition. Efficient rationing assumes that consumers with the highest willingness to pay are served first by the lower-priced firm. In contrast, proportional rationing assigns all consumers an equal chance to buy. We introduce a general framework parameterised by α, encompassing both classical rules as special cases and enabling the modelling of switching-cost-based rationing. Such a framework captures the impact of consumer inertia, search frictions, or contractual obligations on residual demand. Our analysis reveals that different rationing rules have a significant effect on equilibrium outcomes and firms' strategic choices, particularly in terms of capacity and pricing. Departures from the efficient rule may weaken or amplify competitive intensity, thus reshaping market power. These insights directly affect competition protection policy, particularly in the areas of horizontal restrictive agreements, abuse of dominant position, and merger control. The findings advocate for a more nuanced integration of rationing rules into oligopoly models and policy tools.*

Keywords: *capacities, Bertrand-Edgeworth competition, rationing rules, switching costs.*

1. INTRODUCTION

Demand rationing rules are crucial in oligopoly models, particularly in the context of capacity-constrained price competition, as they determine how market demand is allocated among competing firms. This allocation has a direct impact on firms' strategic incentives, equilibrium outcomes, and overall market efficiency. The key question arises: What kind of demand does the more expensive firm face compared to its cheaper rival? Thus, at the heart of the matter lies the shape of the residual demand curve that a relatively more expensive firm faces.

Rationing mechanisms have long played a central role in models of imperfect competition where supply cannot meet demand due to binding capacity constraints. Classical Bertrand-Edgeworth frameworks typically assume that excess demand is allocated via proportional or efficient rationing, which is treated as exogenous to the model. In short, under efficient rationing, the cheaper firm sells to those consumers with the highest willingness to pay—at least up to the limit of its own capacity—while the more expensive firm is left to serve only the lower end of the market. On the other hand, under the proportional rationing rule, each consumer has an equal chance of being allocated to either the cheaper or the more expensive product. According to Dixon (1987), this makes the rule probabilistic, in contrast to the efficient rationing rule, which is deterministic.

This work builds on the classical literature on capacity-constrained competition, notably that of Edgeworth (1897), Levitan & Shubik (1972), and Kreps & Scheinkman (1983). These models typically examine the implications of Bertrand's competition under capacity constraints by treating rationing mechanisms as exogenous to the model. This paper poses the following question: What if rationing rules differ from the standard *efficient* or *proportional* mechanisms while remaining exogenous to the competition model? More precisely, what factors might shape such a rule? What would the market outcome be under competition between firms with capacity constraints, and what policy implications would this have? In the Bertrand-Edgeworth setting, where rationing becomes necessary due to excess demand, a price war does not necessarily lead to the Bertrand paradox. Indeed, in the seminal model of Kreps and Scheinkman (the KS model), a two-stage game involving capacity choice followed by price competition yields a Cournot-type equilibrium. However, this outcome crucially depends on the assumption of efficient rationing, and the model proves to be non-robust to changes in the rationing mechanism. Davidson & Deneckere (1986) offer a widely cited critique of the KS framework, demonstrating that the choice of rationing rule is central to its validity. If a different rationing rule replaces the efficient one, *ceteris paribus*, the outcome of the two-stage game no longer aligns

with that of the Cournot model. They prove this by using proportional rationing, which they associate with Edgeworth's original setup. Complementing this theoretical insight, Jacobs & Requate (2016) provide experimental evidence showing that the choice of rationing rule significantly affects the shape and outcome of residual demand, with proportional rationing leading to higher prices and profits, thereby challenging standard equilibrium predictions.

Furthermore, Davidson & Deneckere (1986) characterise these two rationing rules, efficient and proportional, as extremes, suggesting that a continuum of other, more nuanced rationing rules lies between them. In contrast, Herk (1993) introduces an additional assumption in shaping residual demand for high-priced firms: consumers face high switching costs when moving from one firm to another (Examples include: early termination fees as a form of contractual lock-in, loyalty programs, data or infrastructure incompatibility in tech industries, behavioural inertia of consumers caused by habitual usage or search costs, etc.). This introduces a new form of rationing rule that may result in residual demand that falls outside the range defined by the classical extremes. Consumer switching costs shape this rationing rule. Accordingly, we might refer to it as *switching-cost-based rationing*. In that case, it turns out that competitors' capacities and prices are not the only decisive factors in shaping the residual demand faced by the higher-priced firm.

Before attempting to answer the questions posed, it is essential to consider what constitutes efficient rationing and which alternative rules might plausibly shape the residual demand faced by the higher-priced firm. Accordingly, this paper is structured into three interrelated sections following the introductory remarks. The following two chapters discuss how alternative demand allocation rules might be formulated alongside efficient rationing, as the more commonly employed approach in capacity-constrained oligopoly models. The final chapter concludes the discussion, summarising the observed implications of different rationing mechanisms on economic policymaking, with a particular emphasis on competition protection policy.

2. LOGIC OF EFFICIENT RATIONING

In a capacity-constrained duopoly competition, if one firm sets a higher price than the other, its demand depends on consumer preferences and the market share that the lower-priced firm can and is willing to serve. The rationing rule determines how excess demand (beyond the capacity of the cheaper firm) is distributed, which becomes the residual demand faced by the other firm. The number of possible rationing rules is infinite. In this paper, in addition to the efficient rationing rule, the focus will also be on its two alternatives: the proportional rule, a classical example rooted in Edgeworth's work, and the so-called switching-cost rule. However, this chapter focuses on the economic meaning of efficient rationing. With efficient rationing, the higher-priced firm serves only consumers with a low willingness to pay, as the lower-priced firm captures the top portion of market demand up to available capacities, thereby maximising consumer welfare in such circumstances. Therefore, this indicates an entirely rational demand spillover in cases of limited capacity. Consider a situation where two direct competitors on a homogeneous product market (firms i and j) compete for market demand defined by $p = 1 - q$. The capacities of firms i and j (denoted k_i and k_j, respectively) are assumed to be constrained, such that neither firm can serve the entire market when in the position of the lower-priced seller. Thus, the limit on "cheaper buying" is effectively imposed by the capacity of the cheaper firm. If a firm i were to undercut its competitor, the sales it faces would be

$$q_i = \min\{k_i,\, q(p_i)\} = \min\{k_i,\, 1 - p_i\}\,, \tag{1}$$

and on the other hand, the sales of firm j, as the higher-priced competitor, would amount to

$$q_j = \min\left\{k_j,\, \max\left[0,\, q(p_j) - k_i\right]\right\} = \min\left\{k_j,\, \max\left[0,\, 1 - p_j - k_i\right]\right\}. \tag{2}$$

Finally, if the two firms were to charge identical prices, the sales of firm i would be

$$q_i = \min\left\{k_i,\, \left(\frac{k_i}{k_i + k_j}\right) q(p_i)\right\} = \min\left\{k_i,\, \left(\frac{k_i}{k_i + k_j}\right) (1 - p_i)\right\}. \tag{3}$$

Following expression (3), each firm would satisfy a share of total market demand at the same price proportional to its capacities relative to the total available capacities on the market, but without exceeding its individual capacity constraint. Note that in the case of equal capacities, expression (3) simplifies to $1/2(1 - p_i)$, indicating that demand is equally divided between the two competitors at a uniform price. Expression (2) is key when

discussing the rationing rule, as it determines the residual demand the higher-priced firm faces. Its graphical representation is given in Figure 1(a) (the blue line, shown as the demand function shifted in parallel by the capacity of the cheaper firm).

As Vives points out, the efficient rationing rule eliminates the possibility of resale, which is a logical feature since it assumes that individuals with the highest willingness to pay are served first at a lower price (Vives, 1999, p. 126). In contrast, Tirole insightfully notes that the residual demand defined by the efficient rule is precisely the one that would emerge if consumers could freely resell the good among themselves—i.e. if they could conduct costless arbitrage (Tirole, 1988, p. 214). This interpretation enables individuals with a lower reservation price, motivated by resale opportunities, to access the cheaper product, resell it, and reappear in the pool of consumers that the higher-priced firm faces. However, the resale mechanism that would realise the efficient rule presupposes a fully efficient exchange market, which Tirole considers an overly strong assumption (Tirole, 1988, p. 213).

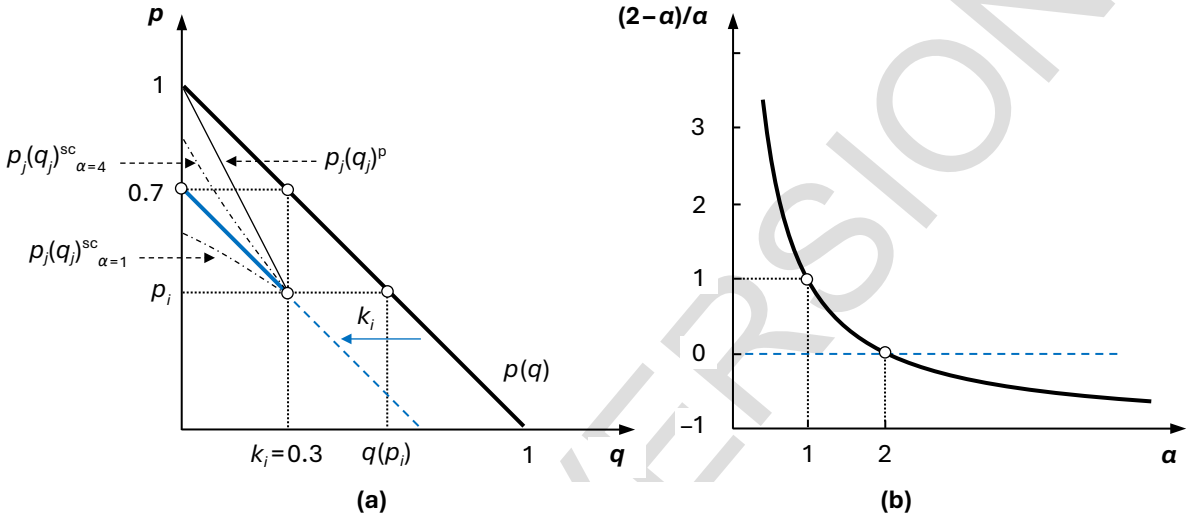

Figure 1: Alternative rationing rules

Consumers willing to pay more are served before those willing to pay less. Why? Does this necessarily mean that the seller must be able to distinguish between buyers based on their willingness to pay? These are some of the critical questions for understanding the context in which it is realistic to expect efficient rationing. Suppose a buyer's willingness to pay decreases with their geographical distance from the seller. In that case, those with the highest reservation prices would be located closest to the seller and, as a result, would be better informed about the timing and terms of sale. This would increase the likelihood that they are the first to be served when the product becomes available. In such a setting, the seller does not need to know which individual buyer has the highest reservation price—this will naturally be the one who shows up first at the shop door.

On the other hand, a buyer's position in the queue for the cheaper product may be interpreted as the result of individual market effort, and this effort need not be correlated with the consumer's reservation price. For example, if participating in the market requires sacrificing free time (commonly considered a normal good), it is plausible that individuals with lower income levels, and thus lower reservation prices, may exert greater market effort due to more available free time. As a result, individuals with lower reservation prices may get cheaper products sooner than those with higher ones. If resale is not the motive behind waiting in line, then the resulting allocation of goods will differ from that implied by the efficient rationing rule. As Dixon (1987) argues, under such conditions, the residual demand may deviate from that shaped by efficient rationing. This insight invites a further question: What types of residual demand might lie *above* or *below* the "blue line" from Figure 1(a), indicating weaker or stronger consumer responsiveness to price undercutting? Or, what are the alternatives to efficient rationing?

3. ALTERNATIVES TO EFFICIENCY

Let's begin this departure from efficiency by considering the *proportional rule* applied in the original Edgeworth (1925) model. The situations in which a firm is a lower-priced competitor or where prices are equal are interpreted in the same way as under the efficient rule—these correspond to expressions (1) and (3),

respectively. However, in the case where firm *j* is the higher-priced competitor, the residual demand it faces will take the form

$$q_j = \min\left\{k_j,\ \max\left[0,\ q(p_j) - k_i\frac{q(p_j)}{q(p_i)}\right]\right\}$$
$$= \min\left\{k_j,\ \max\left[0,\left(1-\frac{k_i}{q(p_i)}\right)q(p_j)\right]\right\} = \min\left\{k_j,\ \max\left[0,\left(1-\frac{k_i}{1-p_i}\right)(1-p_j)\right]\right\}. \quad (4)$$

The residual demand faced by firm *j* is obtained by multiplying the total market demand at the higher price ($p_j > p_i$) by a constant—that is, the proportionality factor $[1 - k_i/(1-p_i)]$. This factor reflects the share of total demand that *cannot* be served at the lower price and, therefore, remains available to be served at the higher price. The defining feature of the proportional rationing rule, and its principal distinction from the efficient one, is that every consumer with a sufficient willingness to pay can purchase from either firm based on the "first come, first served" principle. For example, when $k_i = 0.3$ and $p_i = 0.4$, the inverse residual demand curve shaped by the proportional rule for firm *j* is given by $p_j = 1 - 2q_j$, as shown in Figure 1(a) as $p_j(q_j)^{\mathrm{P}}$. It represents the original market demand rotated around the maximum willingness to pay, $p(0) = 1$.

Since the Edgeworth rule is not deterministic, as the efficient, the higher-priced firm cannot know with certainty which contingent of consumers will turn to it. However, it does know that they are not necessarily those with the lowest reservation prices. Consequently, the proportional rule provides the higher-priced firm with a more advantageous residual demand curve than the efficient one, allowing for softer competitive pressure in the oligopoly market structure.

In Ruebeck (2011), the formation of residual demand is critically examined through the lens of consumer choice, challenging traditional assumptions about its derivation. Thus, the residual demand is derived through a simulation of consumer arrivals at the firm's door based on a modified version of the proportional rationing rule. As in the case of the standard proportional rule, consumers are assumed to arrive at sellers randomly, while both the higher- and lower-priced firms have equal chances of being visited. This stems from the assumption that consumers lack information about which firm offers the lower price until they find out through experience, and they will purchase if their willingness to pay exceeds the price. In contrast, suppose that the willingness to pay is below the price. In that case, the individual will redirect to the other seller or forgo the purchase altogether if the second price also exceeds her willingness to pay. Naturally, introducing restrictions on the operation of the canonical Edgeworth rule would alter the nature of the resulting residual demand. It becomes clear that the issue of demand allocation is fundamentally empirical in nature. In such a context, introducing a rationing rule into the economic model is "a substitute for a complete analysis of consumer behaviour" (Tirole, 1988, p. 214).

Davidson & Deneckere (1986) suggest that efficient and proportional rationing represent boundary cases, framing a continuum of intermediate, more context-specific rules. In contrast to this view, Herk (1993) introduces an additional assumption: consumers in particular markets face significant switching costs when moving from one firm to another (a kind of "first come, first served", but with the burden of switching costs). Residual demand can then surpass the bounds defined by the classical extremes. In such cases, residual demand would depend on the intensity of switching costs (therefore, the *switching-cost rule*), which in turn defines the strategic room for manoeuvre available to the higher-priced firm. As in Herk (1993), for linear market demand, sales for higher-priced firms can be formalised as

$$q_j = \min\left\{k_j, \max\left[0,\ q(p_j) - k_i\left(\frac{q(p_i)}{q(p_j)}\right)^{\frac{2-\alpha}{\alpha}}\right]\right\},\ \text{for } \alpha \in (0,\infty). \quad (5)$$

As $\alpha \to 0$, the exponent in expression (5), $(2-\alpha)/\alpha$, tends toward infinity; conversely, as $\alpha \to \infty$, the exponent approaches -1. The function is positive for $\alpha < 2$ and negative for $\alpha > 2$. This behaviour of the exponent within the domain of definition of parameter α is illustrated in Figure 1(b). In this way, Herk (1993) used the parameter α to represent the impact of switching costs on the formation of residual demand, and consequently, their influence on the intensity of price competition between duopolists. In such a framework, when $\alpha = 2$, the rule corresponds to efficient rationing, while for $\alpha = \infty$, it reflects Edgeworth's proportional rule—as represented by expressions (2) and (4), respectively. For $\alpha = 1$, as in Herk (1993), the residual demand

is depicted in Figure 1(a) as $p_j(q_j)^{sc}{}_{\alpha=1}$. Given the capacity and price of the cheaper firm previously used (namely, $k_i = 0.3$ and $p_i = 0.4$), we obtain a nonlinear concave function: $q_j = (1 - p_j) - 0.3 \times [0.6/(1 - p_j)]$ which corresponds to $p_j(q_j)^{sc}{}_{\alpha=1}$, and lies below the residual demand curve associated with the efficient rule. Above the "blue line," for $\alpha = 4$, *ceteris paribus*, we would obtain a nonlinear convex function of the form: $q_j = (1 - p_j) - 0.3 \times [0.6/(1 - p_j)]^{-0.5}$.

In general, for $\alpha < 2$ within the domain of definition of α, one observes a family of concave residual demand functions that can be regarded as relatively less favourable for the higher-priced firm compared to the benchmark (efficient rationing rule). Conversely, for $\alpha > 2$, the corresponding family of convex residual demand functions is relatively more favourable to the higher-priced firm.

It is worthwhile to examine the market circumstances associated with different values of the parameter α, particularly in the context of consumer switching costs. Table 1 below provides a numerical example illustrating the behaviour of residual demand for values of α: 1, 2, 4 i ∞, based on discrete price increases by the higher-priced firm, using the same parameterisation as in the example shown in Figure 1(a).

Table 1: Consequences of different α

α	p_j: 0.4 → 0.45		p_j: 0.5 → 0.55	
	Δq_j	Arc ε	Δq_j	Arc ε
1	0.077	- 2.513	0.090	- 9.947
2	0.050	- 1.545	0.050	- 3.000
4	0.037	- 1.125	0.036	- 1.813
→ ∞	0.025	- 0.741	0.025	- 1.107

Based on the example presented in Table 1, it can be concluded that as the parameter α increases, the residual demand becomes more inert to price changes, as reflected in the values of the arc price elasticity of demand (Arc ε). Seemingly, higher values of α are more favourable to the higher-priced firm. When the magnitude of α is interpreted in the context of switching costs, lower values of the parameter α (below 2) indicate low switching costs—consistent with high price elasticity—whereas higher values (above 2) suggest the presence of significant switching costs. Therefore, α is governing the slope and position of the residual demand curve for firm j, thus affecting its pricing incentives and market power. The benchmark of full efficiency is related to $\alpha = 2$, where we have complete deterministic sorting by willingness to pay, where only the difference in prices matters.

4. CONCLUDING REMARKS

Our discussion on alternative demand rationing rules contributes to the field of oligopoly theory from at least three perspectives: first, in terms of the theoretical domain and functioning of economic models; second, by enhancing our understanding of how residual demand is formed for the higher-priced firm; and finally, by offering implications for the conduct of competition protection policy, given that oligopolistic markets are typically at the centre of competition authorities attention.

Thus, the considerations presented in this paper contribute to the discourse on demand-rationing rules beyond the framework defined by the classical rationing principles typically employed in models, namely, the efficient and proportional rules. This approach encompasses all key rationing forms as special cases of a more general form, thereby opening space for theoretical and empirical investigations of firm behaviour in oligopolistic markets. The focus of analyses of oligopolistic conduct, both from a modelling and an empirical perspective, could be situated within the context of the parameter α and the intensity of the lock-in caused by switching costs. Even in homogeneous product markets, the price differences and mode of customer rationing under capacity constraints need not be decisive for customer switching, which may significantly influence firms' strategic behaviour within competition models, such as the framework of the KS model.

Departures from the efficient rationing rule can invalidate the standard conclusion of the KS model—namely, that the outcome of the Cournot model necessarily emerges from a two-stage game where duopolists first choose capacities and then compete in prices. For instance, under any rationing rule more favourable to the higher-priced firm than the efficient one, if both firms hold Cournot-level capacities and one sets the Cournot price, the other would find it profitable to set a price above the Cournot level. Being aware of the probability that its rival will choose a price from the Edgeworth range above its own, each firm may hold some reserve capacities as a precautionary measure. Firms would do so because it allows them to increase output above the Cournot level and capture a larger market share by undercutting their competitors. A more favourable

residual demand compared to the efficient rule also implies greater room for price fluctuations for the higher-priced firm and, consequently, higher potential profits, which would induce firms to choose capacities exceeding the Cournot output level.

Conversely, in the case of residual demand that is less favourable to the high-priced firm than that predicted by the efficient rule, the market outcome is likely to be less competitive. Caution would dictate that firms in equilibrium hold lower capacity than what is expected under the Cournot equilibrium.

Finally, discussing the demand rationing rules may have essential implications for enforcing competition protection policy. The assessment of market power is particularly relevant in cases involving the abuse of dominance, as well as in the context of horizontal restrictive agreements and merger control.

Market power may persist due to hidden structural favouritism or behavioural frictions that are not readily observable through conventional indicators such as price-cost margins or the Herfindahl-Hirschman Index. Therefore, it would be helpful to develop empirical methods for estimating the value of the parameter α across different relevant markets. For instance, by comparing observed and predicted residual demand flows or by employing consumer surveys to assess brand loyalty, search frictions, and behavioural inertia. By logic, values of α above two may indicate significant market power necessary for claims related to abuse of dominant position, as well as greater cartel stability. A firm that undercuts the cartel price gains little additional demand, as many customers remain locked in with the higher-priced firm.

In the context of merger control, it is crucial to recognise that mergers may affect not only market concentration but also the structure of demand allocation. Accordingly, merger simulation models should incorporate rationing rules and switching costs while examining whether a merger may lead to implicit changes in α (through mechanisms such as shared loyalty programmes, unified branding strategies, contractual bundling, etc.). In any case, conditional merger approvals should include behavioural measures that ensure consumer choice neutrality based on the knowledge of the efficient rationing rule and its alternatives.

LITERATURE

Davidson, C. & Deneckere, R. (1986). Long-Run Competition in Capacity, Short-Run Competition in Price, and the Cournot Model. *The RAND Journal of Economics*, 17(3), 404-415. https://doi.org/10.2307/2555720

Dixon, H. (1987). The General Theory of Household and Market Contingent Demand. *The Manchester School*, 55(3), 287-304. https://doi.org/10.1111/j.1467-9957.1987.tb01303.x

Edgeworth, Y. F. ({1897} 1925). *The Pure Theory of Monopoly*. In: *Papers Relating to Political Economy*, Vol. I, Chapter E, (pp. 111-142). New York: Burt Franklin.

Farrell, J. & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. In: M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, Vol. 3, (pp. 1967-2072). Elsevier. https://doi.org/10.1016/S1573-448X(06)03031-7

Herk, F. L. (1993). Consumer Choice and Cournot Behavior in Capacity-Constrained Duopoly Competition. *The RAND Journal of Economics*, 24(3), 399-417. https://doi.org/10.2307/2555965

Jacobs, M. & Requate, T. (2016). Demand rationing in Bertrand-Edgeworth markets with fixed capacities: An experiment, Economics Working Paper, No. 2016-03, Kiel University, Department of Economics, Kiel. https://hdl.handle.net/10419/125820

Klemperer, P. (1987). Markets with consumer switching costs. *Quarterly Journal of Economics*, 102(2), 375-394. https://doi.org/10.2307/1885068

Kreps, M. D. & Scheinkman, A. J. (1983). Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes. *The Bell Journal of Economics*, 14(2), 326-337. https://doi.org/10.2307/3003636

Levitan, R. & Shubik, M. (1972). Price Duopoly and Capacity Constraints. *International Economic Review*, 13(1), 111-122. https://doi.org/10.2307/2525908

Ruebeck, S. C. (2011). Consumer Search, Rationing Rules, and the Consequence for Competition. In: J. Salerno, S. Jay Yang, D. Nau and Sun-Ki Chai (Eds.), *Social Computing, Behavioral-Cultural Modelling and Prediction* (pp. 155-162). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19656-0_24

Tirole, J. (1988). *The Theory of Industrial Organization*. The MIT Press, Cambridge Massachusetts.

Vives, X. (1999). *Oligopoly Pricing: Old Ideas and New Tools*. The MIT Press, Cambridge Massachusetts.