Journal of Regulatory Economics (2022) 61:222–229
https://doi.org/10.1007/s11149-022-09449-4

ORIGINAL ARTICLE

Optimal pricing and investment for resources with alternative uses and capacity limits

Alessandro Avenali[1] · Tiziana D'Alfonso[1] · Pierfrancesco Reverberi[1]

Accepted: 15 June 2022 / Published online: 21 July 2022

Abstract
Airport runways, radio spectrum, and hospital beds are resources with capacity limits used to provide multiple services with specific capacity requirements in separate markets, which contribute to recover capacity investment costs. A welfare-maximizing and (possibly) budget-constrained firm, whose operating costs significantly increase as total capacity use presses against capacity, chooses prices and capacity. When the equilibrium capacity is reached, second-best Ramsey prices must be adjusted, and mark-ups on marginal costs may be higher for services with higher demand elasticities, if they intensively use capacity. Moreover, for a given output vector, the firm invests more than in first best. Instead, the equilibrium capacity may be first best when there is excess capacity to reduce operating costs and thus improve welfare. Our model can be used as a benchmark to evaluate the efficiency of market mechanisms for resource allocation and pricing, or when market mechanisms are not adopted.

Keywords Ramsey pricing · Capacity limit · Capacity requirements · Capacity investment

JEL Classification L51 · L90 · D42 · D45

1 Introduction

Airport runways, railway tracks, radio spectrum, and hospital beds are notable examples of resources with alternative uses that may be capacity constrained. This means that, if capacity is allocated to a given service, then it is not available to others, where

✉ Pierfrancesco Reverberi
pierfrancesco.reverberi@uniroma1.it

[1] Dipartimento di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti (DIAG), Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy

services have different capacity requirements.[1] Moreover, an increase in total capacity use that presses against available capacity imposes a substantial increase in operating costs on the resource owner to avoid that (perceived) service quality declines.[2] Finally, deploying these resources, or enjoying rights of use, involves huge investments that entail fixed sunk costs, and the resource owner may be budget constrained. Capacity investment costs are recovered by offering various services in separate markets with different demands.

In this framework, we study two related issues. First, which prices determine the optimal capacity allocation to each service? Second, what is the optimal capacity level consistent with such prices?

The main novelty of this paper is that there are multiple services with different capacity requirements that are competing to use the resource capacity, which may be limited, and are contributing to recover capacity costs. We develop a simple model where a welfare-maximizing and (possibly) budget-constrained firm invests in capacity to serve markets with independent demands.

There are two main findings. First, for a given capacity, the equilibrium prices of services may require combining the logic of the 'inverse elasticity rule' typical of Ramsey pricing (Baumol & Bradford, 1970) with the need to signal that capacity is fully used. In such a case, second-best Ramsey prices must be adjusted, and (relative) mark-ups over marginal operating costs may be higher for services with higher demand elasticities, as long as they use capacity more intensively than other services.

Second, for a given (feasible) output vector, when the equilibrium capacity is reached the firm invests more than in first best, at a level where the marginal cost of capacity is greater than the marginal decrease in operating costs as capacity rises. Instead, the equilibrium capacity level may be first best when it is not reached, and excess capacity serves to reduce operating costs and thus improve welfare.

Our model relies on the principles of Administered Incentive Pricing (AIP), which has been applied to induce an efficient use of radio spectrum (Ofcom, 2010). AIP calls for setting regulated charges that reflect the opportunity cost of the scarce resource, by taking account of alternative uses. To our knowledge, this paper is the first theoretical attempt to find efficient administered incentive prices.

This paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the model and analyzes the results. Section 4 concludes and discusses policy implications.

[1] At any given time, a capacity-constrained airport runway (railway track) serves one out of many possible markets (e.g. local versus long-distance trips), where the various aircrafts (trains) employed have different capacity needs. As to radio spectrum, technological innovation has made it possible for frequency bands traditionally reserved to terrestrial television to be adopted for delivering wireless broadband services. Hospital beds can be used for inpatient stays due to different medical treatments (e.g. there are critical care beds for heart attacks and intensive care beds for respiratory failure).

[2] In communications, as traffic grows energy consumption increases considerably due to retransmissions of lost multimedia packets, and since additional hardware devices must be turned on. Rail operators may have to add wagons to trains in response to a rise in travel demand. When an increase in hospital stays insists on intensive care units, extra machines must be rented, and personnel must work more overtime.

2 Literature review

Ramsey pricing has been widely used to allocate and regulate resources when it is critical to recover investment costs (i.e. marginal cost pricing creates a deficit), as in multiproduct natural monopolies (see e.g. Baumol & Bradford, 1970).[3] The basic framework has been extended to study the impact of external effects such as network externalities in communications (see e.g. Mitchell & Vogelsang 1991) and externality costs in transport, including congestion (see e.g. Oum & Tretheway, 1988).

While these generalizations relate to the demand side, we study how the Ramsey rule must be adjusted due to imperfections on the supply side (such as inelastic supply) to allocate resources with capacity limits to multiple services with different capacity requirements.[4]

Classical peak-load models consider capacity-constrained resources, but production takes place under constant marginal costs (see e.g. Crew et al., 1995). Mohring (1970) assumes increasing returns and finds Ramsey-like equilibrium prices in each demand period. Bailey and White (1974) show that the Ramsey peak price may be less than the off-peak price. Jordan (1983) finds that users with different capacity needs should pay specific capacity charges independent of the demand period. Despite apparent similarities, the peak-load problem is inherently different from ours, since peak and off-peak demands are not competing for available capacity.[5] As such, peak-load models ignore the issue of alternative uses of the capacity-constrained resource, which is the focus of our analysis.

Finally, a long-standing question in road pricing and investment (see e.g. Lindsey 2012) is that efficient road usage calls for marginal-cost pricing, while paying for roads requires pricing at average cost. Mohring (1970) notes that, with a deficit constraint and a single toll for road users, the equilibrium capacity might be a level at which marginal benefits are greater than marginal costs of capacity. Generally, in this literature the resource is not capacity constrained. We add to these papers by including an endogenous capacity limit.

3 Model and results

Consider a resource managed by a public (regulated private) firm that provides services in n ($n > 1$) markets with independent demands.[6] Let $p_i(q_i)$ be the inverse demand curve in market i ($i = 1, \ldots, n$), where p_i and q_i are respectively the price and quantity of service i. Let $S(q)$ be the gross consumer surplus for the output vector $q = (q_1, \ldots, q_n)$, with $\partial S(q)/\partial q_i = p_i$.

[3] The theory of public enterprise pricing has also been used to determine optimal prices for hospitals (see Harris, 1979).

[4] Likewise, Holguin-Veras and Jara-Diaz (1999) deal with priority systems in seaports, but they abstract from capacity investment.

[5] Indeed, there is one capacity constraint on the output level produced in each demand period, instead of a single capacity constraint on the total output produced in both periods.

[6] See e.g. Mohring (1970) on the form of utility functions that are consistent with the assumption of independent demands.

We assume that each service makes a constant marginal use of resource capacity K. Let ϕ_i ($\phi_i > 0$) be the marginal capacity use by service i ($i = 1, \ldots, n$), namely, the capacity units needed to obtain one unit of service i. Thus, $\sum_{i=1}^{n} \phi_i q_i$ is the total capacity use to supply the output vector q.[7] There is a capacity limit (endogenously determined) for capacity use, so that $\sum_{i=1}^{n} \phi_i q_i \leq K$. When capacity is reached, the resource is *capacity constrained*.

Let $C(q, K) = C^I(K) + C^O(q, K)$ be the total cost of providing the output vector q with capacity K, which is additively separable in two terms, namely, capacity investment costs $C^I(K)$ and operating costs $C^O(q, K)$.[8] Let $MC^I = \partial C^I(K)/\partial K$ be the marginal cost of capacity investment, with $MC^I > 0$. Let $MC_i^O = \partial C^O(q, K)/\partial q_i$ be the marginal operating cost of service i, with $MC_i^O > 0$.

We posit that an increase in total capacity use (due to a higher volume of services), which presses against available capacity, imposes significant additional operating costs on the firm to avoid that (perceived) service quality declines.[9] It follows that $\partial C^O(q, K)/\partial K < 0$. Moreover, the higher the total capacity use relative to available capacity, the larger the decrease in operating costs as capacity rises. A sufficient condition for this to occur is $\partial^2 C^O(q, K)/\partial K \partial q_i < 0$ ($i = 1, \ldots, n$).[10]

Let $W(q, K)$ and $\pi(q, K)$ respectively be the social welfare and profit functions. The firm chooses capacity and output levels to maximize welfare, subject to a budget and a capacity constraint:

$$\begin{aligned} &\max_{q,K} W(q, K) = S(q) - C(q, K) \\ &s.t. \, \pi(q, K) = \sum\nolimits_{i=1}^{n} p_i(q_i) q_i - C(q, K) \geq 0 \\ &\sum\nolimits_{i=1}^{n} \phi_i q_i \leq K \end{aligned} \tag{1}$$

Let λ_B and λ_K ($\lambda_B \geq 0$, $\lambda_K \geq 0$) be the Karush-Kuhn-Tucker multipliers for the budget and capacity constraints, respectively. The first-order conditions for (1) are ($i = 1, \ldots, n$):[11]

$$p_i - MC_i^O + \lambda_B \left(p_i + \tfrac{\partial p_i(q_i)}{\partial q_i} q_i - MC_i^O \right) - \lambda_K \phi_i = 0 \tag{2}$$

[7] This assumption simplifies the analysis but is not essential when capacity use is separable in outputs, namely, the total capacity use for providing q can be measured as $\sum_{i=1}^{n} K_i(q_i)$, where $K_i(q_i)$ is the capacity amount needed to produce q_i.

[8] This is a common assumption in the relevant literature (see e.g. Jordan, 1983).

[9] An increase in the volume of services may sometimes create congestion, which affects the demand side by reducing consumers' utility (e.g. transport users are delayed by a traffic growth). For simplicity, we do not consider this issue.

[10] An operating cost function with these properties is $C^O(q, K) = c \sum_{i=1}^{n} q_i + \alpha \left(\sum_{i=1}^{n} \phi_i q_i / K \right)$, where $\sum_{i=1}^{n} q_i$ is the total output (different from the total capacity use), and $\alpha > 0$. Thus, operating costs approach a finite (arbitrarily large) value when the total capacity use approaches the capacity limit, and this limit may be reached in equilibrium.

[11] We assume that the second-order conditions for a maximum are satisfied.

$$(1+\lambda_B)\left(-\frac{\partial C^O(q,K)}{\partial K} - MC^I\right) + \lambda_K = 0 \tag{3}$$

Let $\epsilon_i = -(\partial q_i/\partial p_i)(p_i/q_i)$ be the own price elasticity of demand for service i. From (2), we find:

$$p_i = \frac{(1+\lambda_B)\,MC_i^O + \lambda_K \phi_i}{1 + \frac{\lambda_B}{|\epsilon_i|}(|\epsilon_i| - 1)} \tag{4}$$

Then, from (4), for any given pair of services i and j $(i, j = 1, \ldots, n, i \neq j)$, we find:

$$\frac{p_i - \left(MC_i^O + \frac{\lambda_K}{1+\lambda_B}\phi_i\right)}{p_i}|\epsilon_i| = \frac{p_j - \left(MC_j^O + \frac{\lambda_K}{1+\lambda_B}\phi_j\right)}{p_j}\left|\epsilon_j\right| = \frac{\lambda_B}{1+\lambda_B} \tag{5}$$

As to the capacity choice, from (3) we have:

$$MC^I - \frac{\lambda_K}{1+\lambda_B} = -\frac{\partial C^O(q,K)}{\partial K} \tag{6}$$

Clearly, the equilibrium prices and capacity level depend on the shadow prices of the constraints. In what follows, we focus on the role played by the capacity limit.[12]

The benchmark case is the first-best outcome, which occurs when $\lambda_B = \lambda_K = 0$ holds at the optimum and (1) reduces to an unconstrained welfare maximization. From (4), first-best prices equal marginal operating costs. From (6), first-best capacity is such that the marginal cost of capacity investment equals the marginal value of capacity, which is reflected in lower operating costs as capacity rises. Note that the capacity level is efficient even if there is excess capacity at the optimum. Indeed, the social planner decides to hold idle capacity to reduce operating costs and thus improve welfare.

Assume that only the capacity constraint is binding in equilibrium (i.e. $\lambda_B = 0$, $\lambda_K > 0$). Inserting for $\lambda_B = 0$ in (4), we find that, despite cost recovery is not an issue, welfare maximization does not yield marginal-cost prices. Indeed, for such prices, the demand for services would exceed the capacity limit. Thus, equilibrium prices must include a mark-up over marginal operating costs. *Ceteris paribus*, the higher the capacity use by a service, and the higher the shadow price of the capacity constraint, the higher the mark-up, since the higher is the benefit of reducing the service volume.

Then, inserting for $\lambda_B = 0$ in (6), we find that the equilibrium capacity level is such that the marginal cost of capacity investment is greater than the marginal decrease in operating costs as capacity rises. Thus, *for a given (feasible) output vector*, the capacity

[12] Numerical simulations for the cases considered in this section are available from the authors on request.

level is higher than in first best,[13] and the opportunity cost of the capacity limit dictates how much capacity (and prices) must be raised to minimize the deadweight loss.

Assume now that both constraints are binding in equilibrium (i.e. $\lambda_B > 0, \lambda_K > 0$). Due to the need for cost recovery, *for a given capacity level* equilibrium prices are higher than marginal operating costs. Since the capacity limit is also an issue, then second-best Ramsey prices must be adjusted to signal that the resource is capacity constrained, so as to restrict the demand for services. From (5), *ceteris paribus*, equilibrium prices decrease with demand elasticities, but increase with marginal capacity uses, depending on the relative weight of the constraints (i.e. the opportunity costs of the capacity limit and budget threshold). We therefore obtain the following result.

Remark 1 Assume that both the budget and the capacity constraints are binding in equilibrium. Then, second-best Ramsey prices must be adjusted, and (relative) mark-ups may be higher for services with higher demand elasticities, if they use capacity more intensively than other services.

As to the capacity choice, from (6), the need to expand capacity is mitigated by the opportunity cost of the budget threshold, and thus by the parallel increase in prices (that reduces the demand for services) to satisfy the profit constraint, while minimizing the deadweight loss. *For a given output vector*, the capacity level is higher than in first best. We therefore obtain the following result.

Remark 2 For a given output vector, when the capacity constraint is binding in equilibrium the public (regulated private) firm invests more than in first best. Instead, when the capacity constraint is not binding in equilibrium, the capacity level may be first best, and the firm invests in excess capacity to reduce operating costs and thereby improve welfare.

4 Conclusions and policy implications

Resources with capacity limits are often used to provide multiple services with specific capacity requirements in separate markets with different demands, which contribute to recover capacity investment costs. Furthermore, operating costs significantly increase as the total capacity use presses against available capacity. We have developed a simple model for allocating and pricing resources with these features, such as airport runways, railway tracks, radio spectrum, and hospital beds.

We have found the equilibrium prices and capacity level of a welfare-maximizing and (possibly) budget-constrained firm, when multiple services compete for the same resource. We have shown that, when both cost recovery is an issue and the resource is capacity constrained, second-best Ramsey prices must be adjusted, and (relative) mark-ups over marginal operating costs may be higher for services with higher demand elasticities, if they use capacity more intensively than other services.

We have also shown that, for a given output vector, when capacity is limited in equilibrium the firm invests more than in first best, at a level where the marginal cost

[13] Indeed, we have assumed that $\partial C^O(q, K)/\partial K < 0$ and $MC^I > 0$.

of capacity is greater than the marginal decrease in operating costs as capacity rises. Instead, when the equilibrium capacity is not reached it may be first best, and excess capacity reduces operating costs thereby improving welfare.

Our model can be used to support existing resource allocation and pricing methods when market mechanisms are not implemented, such as for hospital services.[14] It suggests that, when different diseases and treatments compete for hospital beds with limited capacity (e.g. during a pandemic), prices for hospital services should be related to demand factors such as disease severity and treatment urgency (as proxies for demand elasticities), rather than simply reflect the supply side.

Our results also provide a benchmark to evaluate the efficiency of market mechanisms for allocating and pricing resources involved in the deregulation wave in network industries. Market mechanisms yield benefits relative to traditional *grandfather rights* (for airport slots) or *beauty contests* (for radio spectrum), but have been criticized on efficiency grounds (Avenali et al., 2015; Esö et al., 2010).[15] Our model pursues the optimal allocation of such resources, given that: (i) there is excess demand for alternative uses, but secondary market trading is not mature enough to secure efficient reallocation; (ii) coordination of multiple users is required, and the costs that would arise if these parties attempted to trade directly would be prohibitive; (iii) for a given capacity, sunk costs and/or regulatory provisions imply that changes of use are restricted in the short run.[16]

We have considered a vertically integrated monopoly that manages the resource and offers retail services. This assumption fits hospital beds, which are often ruled by a local authority. Airports and railway tracks are natural monopolies typically run by a public or regulated private firm, while radio spectrum is managed by the government. These entities usually offer resource access to downstream firms such as airlines or mobile companies that, in turn, deal with the end users of services.

As long as our results are referred to wholesale prices, they still hold with a vertically separated resource owner and a downstream market where firms do not have market power. The same results also hold with a regulated downstream monopoly, such as for rail services. In future work, we may study how downstream market power affects resource capacity allocation and pricing.

We may also study the effects of technological innovation on capacity investment. On the one hand, after innovation takes place, new services may compete for the resource (as for radio spectrum), which then becomes or remains capacity constrained over time. On the other hand, innovation may induce more efficient capacity uses by existing services.

---

[14] Hospital services are usually priced through the 'diagnosis-related group' (DRG) system, which determines hospital reimbursements based on estimated (average) treatment costs. If the hospital treats a patient for less (more) costs than the fixed amount it is paid for the patient's DRG, then it saves (loses) money on that hospitalization.

[15] In this framework, the equilibrium prices of our model could serve as reservation prices to run (simultaneous) auctions.

[16] For instance, mobile broadband access may suffer from scarcity of spectrum available for that use, while a similar portion of spectrum, which has been reserved for broadcasting television services in the same area, may be partially used.

Acknowledgements We received funding from Sapienza University of Rome – Progetto Ateneo RP11916B7A9EF5DB. We thank the Editor and an anonymous reviewer for helpful comments and suggestions on an earlier version of the manuscript.

References

Avenali, A., D'Alfonso, T., Leporelli, C., Matteucci, G., Nastasi, A., & Reverberi, P. (2015). An incentive pricing mechanism for efficient airport slot allocation in Europe. *Journal of Air Transport Management*, *42*, 27–36

Bailey, E. E., & White, L. J. (1974). Reversals in peak and offpeak prices. *The Bell Journal of Economics and Management Science*, *5*, 75–92

Baumol, W. J., & Bradford, D. F. (1970). Optimal departures from marginal cost pricing. *American Economic Review*, *60*, 265–283

Crew, M. A., Fernando, C. S., & Kleindorfer, P. R. (1995). The theory of peak-load pricing: A survey. *Journal of Regulatory Economics*, *8*, 215–248

Esö, P., Nocke, V., & White, L. (2010). Competition for scarce resources. *RAND Journal of Economics*, *41*, 524–548

Harris, J. E. (1979). Pricing rules for hospitals. *The Bell Journal of Economics*, *10*, 224–243

Holguin-Veras, J., & Jara-Diaz, S. (1999). Optimal pricing for priority service and space allocation in container ports. *Transportation Research Part B: Methodological*, *33*, 81–106

Jordan, W. J. (1983). Heterogeneous users and the peak-load pricing model. *Quarterly Journal of Economics*, *98*, 127–138

Lindsey, R. (2012). Road pricing and investment. *Economics of Transportation*, *1*, 49–63

Mitchell, B., & Vogelsang, I., I (1991). *Telecommunications Pricing*. Cambridge University Press

Mohring, H. (1970). The peak load problem with increasing returns and pricing constraints. *American Economic Review*, *60*, 693–703

Ofcom (2010). SRSP: The revised framework for spectrum pricing, available at: http://stakeholders.ofcom.org.uk/binaries/consultations/srsp/statement/srsp-statement.pdf

Oum, T. H., & Tretheway, M. W. (1988). Ramsey pricing in the presence of externality costs. *Journal of Transport Economics and Policy*, *22*, 307–317