Journal of Urban Economics 101 (2017) 45–56

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

Cost recovery of congested infrastructure under market power☆

Erik T. Verhoef[a,b,1]

[a] Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[b] Visiting Professor, Department of Economics, University of Gothenburg

ARTICLE INFO

Article history:
Received 27 November 2013
Revised 5 June 2017
Available online 20 June 2017

JEL codes:
R41
R48
D62

Keywords:
Congestion pricing
Capacity choice
Self-financing infrastructure
Market power
Airport congestion

ABSTRACT

The Mohring–Harwitz (1962) theorem states that the degree of self-financing of congested infrastructure is equal to the elasticity of the capacity cost function in the optimum, so that under neutral scale economies exact self-financing applies. The theorem breaks down when the infrastructure is used by operators with market power, the case in point often being airlines at a congested airport. This paper proposes a regulatory scheme that restores self-financing in such cases; partially so in general, and perfectly so under specific circumstances that include (1) the satisfaction of a particular proportionality condition, and (2) either the isolation of budgets needed for subsidies to counter demand-related mark-ups, or perfectly elastic demands so that such mark-ups are zero. Moreover, exact self-financing applies in this scheme independent of the elasticity of the capacity cost function, and occurs for both parametric and "manipulable" congestion pricing.

1. Introduction

The supply of public infrastructure is one of the main tasks of governments, and may involve a substantial share of public budgets. Consequently, the pricing of its use is in practice not only motivated by the efficiency argument of getting prices equal to marginal costs, but also as a means of raising revenues to finance infrastructure supply. The celebrated theorem on self-financing of infrastructure capacity, due to Mohring and Harwitz (1962), identifies a remarkable connection between these two possible roles of pricing. It states that under certain technical conditions, the degree of self-financing of optimally priced and dimensioned congested infrastructure is equal to the elasticity of the capacity cost function. Hence, under neutral scale economies, where this elasticity is unity, the revenues from optimal pricing – following Pigou's (1920) prescription that the congestion toll be equal to the marginal external cost – will be exactly equal to the capital cost associated with the supply of the optimal capacity.

Subsequent research has shown that this theorem, originally derived in the context of a static model for a single road with homogeneous users, survives various extensions that make the setting more realistic. For reviews of this literature, see for example De Palma and Lindsey (2007), and Small and Verhoef (2007). In particular, the theorem remains true for networks rather than single facilities (Yang and Meng, 2002); with heterogeneous users (Arnott and Kraus, 1998a); with dynamic congestion technologies, such as the bottleneck model of Vickrey (1969) (Arnott et al., 1993), but also more generally (Arnott and Kraus, 1998a); when including maintenance, and wear and tear (Newbery, 1989); when allowing for variable prices for inputs such as land (Small, 1999); and in present value terms when considering the long run (Arnott and Kraus, 1998b). But the theorem breaks down in other circumstances. One example is where capacity is lumpy, not continuous. Another one, under consideration in this paper, is when infrastructure users possess market power.

A burgeoning literature, most of which concerns aviation, has discussed how optimal congestion tolls for actors with market power may be considerably lower than what is implied by the conventional Pigouvian prescription. One reason is that under Nash-Cournot behaviour, operators with market power internalize congestion imposed upon their own services (Daniel, 1995; Brueckner, 2002), so that the optimal congestion toll should include only marginal congestion effects on other firms' services. Quite intuitively, Brueckner (2002) finds that for uniform values of time and

---

☆ Financial support from ERC (AdG Grant #246969 OPTION) is gratefully acknowledged. I thank Vincent van den Berg, Hugo Silva, and two anonymous reviewers as well as the journal's handling editor Vernon Henderson for very helpful comments on an earlier version of this paper.

E-mail address: e.t.verhoef@vu.nl

[1] Affiliated to the Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam.

http://dx.doi.org/10.1016/j.jue.2017.06.004

uniform marginal congestion effects, this implies firm-specific tolls that are a fraction $(1 - s_i)$ of the conventional marginal external congestion cost, with s_i being the firm's market share at the airport. All else equal, larger firms should then face lower tolls. Because the optimal investment rule for capacity remains unchanged compared to the usual model of congestion with atomistic users, while the optimal pricing rule is adapted, the Mohring–Harwitz rule breaks down.

As a consequence, the prospects for fully self-financing infrastructure are reduced for such markets, compared to the case of atomistic users. This is reinforced when the tolls take into account the fact that profit-maximizing Cournot competitors apply a demand-related mark-up, which increases when demand becomes less elastic and vanishes only in the limiting case of perfectly elastic demand.[2] This calls for tolls that are adjusted downward further compared to the tolls proposed by Brueckner (2002), and that even may become negative when the (negative) demand-related market power component in the toll outweighs the (positive) congestion component (Pels and Verhoef, 2004). This is consistent with Buchanan's (1969) rule for optimal environmental taxation of a monopolist.

Later contributions have identified market configurations that would give better prospects for self-financing under market power, as optimal tolls move closer to atomistic levels. For example, this was found for Stackelberg competition as opposed to Nash competition (Brueckner and Van Dender, 2008); for differentiated Bertrand competition as opposed to Cournot competition (Silva and Verhoef, 2013); and for dynamic bottleneck congestion as opposed to static congestion (Silva, Verhoef and Van den Berg, 2014). Obviously, these findings are of less help, the more accurately the market form is best described by the conventional Cournot–Nash model with static flow congestion. Basso and Zhang (2006) and Zhang and Zhang (2006) included the role of airports in modelling congestion between airlines at airports. Zhang and Zhang (2006) study how a budget constraint to achieve a balanced budget would then affect the behaviour of a welfare-maximizing airport – surprisingly finding that there will be overinvestment in capacity at the margin.

This paper addresses the problem of financing congested infrastructure when operators have market power from a different perspective. A regulatory scheme is proposed that does not involve budget constraints or lump-sum payments in the optimal pricing problem, but that nevertheless restores self-financing for congested infrastructure under market power; partially so in general, and perfectly so under specific circumstances to be spelled out in detail below. These circumstances include (1) the satisfaction of a particular proportionality condition, and, quite naturally, (2) either the isolation of budgets needed for subsidies to counter demand-related mark-ups, or perfectly elastic demands so that such mark-ups are zero anyway. Under the proposed scheme, exact self-financing then applies independent of the elasticity of the capacity cost function, which is markedly different from the atomistic case of road tolls. What is more, a balanced net budget then not only applies at the aggregate level (*i.e.*, for the facility), but also for each operator individually; a result that follows from the fact that operators pay a congestion toll but receive a subsidy to stimulate voluntary contributions to the provision of capacity by operators. This result remains true when operators treat the tolls parametrically, but also for so-called "manipulable" tolls as proposed by Brueckner and Verhoef (2010). These are tolls that are designed to allow for the fact that operators with market power can be expected to be exploit the fact that tolls are not truly parametric but instead depend on their own behaviour, providing them with an incentive to manipulate the toll.[3]

The finding that exact self-financing holds independent of the elasticity of the capacity cost function is likely to make application of the scheme only more attractive in reality. That is, substantial net surpluses or deficits from, for example, airport operations seem more likely to cause political and social opposition than cases where such an airport is close to breaking even. The same can be said of balanced budgets for each operator individually. The result that with market power optimal congestion tolls are inversely proportional to the operator's size, is easily interpreted as "unfair", making its application less attractive from the political perspective. A scheme that leads to a balanced budget in each operator's contribution to infrastructure finances seems, from that perspective, more attractive.

The basic idea behind the scheme is simple. It exploits the notion that non-atomistic operators have an incentive to contribute voluntarily to the supply of capacity. This incentive is not socially optimal, so subsidization of capacity provision is necessary to achieve the first-best. Larger firms, apart from facing a lower optimal toll, have a larger incentive to contribute to capacity, and therefore need a lower optimal per-unit-of-capacity subsidy to do so. The balanced budget result means that the toll revenues are just sufficient to cover the subsidies on capacity provision that are required to make the firms collectively supply the first-best aggregate capacity. The budgets will be balanced exactly only if a particular "equi-proportionality" condition is fulfilled: each firm's share in the total capacity cost should be equal to its share in the total output. It will be shown that this condition, although intuitively plausible and natural, will not be satisfied spontaneously in the decentralized optimum. For exact self-financing to apply, the regulator would therefore have to make sure the condition is exactly satisfied; obviously, for smaller violations the relative deficits or surpluses are more modest than for larger deviations.

The paper has the single main objective of presenting the theoretical result, which is illustrated in the Appendix by a simple numerical exercise. Practical implementation issues will, however, also be addressed when they arise, albeit briefly. Arguably, the most pressing challenge in practice would be to cope with the non-stationarity of market equilibria, where capacity investments are typically long-lasting and irreversible. This would make firms unwilling to contribute to the cost of capacity, insofar as this concerns sunk costs from earlier investments. A promising way to circumvent this in practice would be to tie the supply of capacity to a short-run flexible complementary service, for which it is credible that it can be reduced in size if smaller contributions are made, so that effective capacity remains flexible. In the context of aviation, airport staff would be a good example. Indeed, although the analysis applies to the general case of a congestible facility used by operators with market power, aviation is an important example and the one that motivated this paper in the first place. Still, the principle applies more broadly; other examples may include con-

---

[2] Both Oum and Zhang (1993) and Brander and Zhang (1990) conclude that actual airline pricing behaviour appears to be closer to Cournot than to Bertrand behaviour.

[3] I thank a reviewer for pointing out that it is important to distinguish clearly between the different meanings of "market power" in the context of this paper. A first is that operators are so large relative to the facility that there is a non-negligible degree of self-imposed congestion that the operator internalizes without interference of a regulator. This mechanism motivates the central question of this paper, and is present throughout the analyses to be discussed. A second is that operators may face a downward sloping demand and hence find it profitable to apply a demand-related mark-up in their pricing. This mechanism is excluded from the analysis by assuming that corrective subsidization to restore efficiency would be financed form a separate budget, a budget that can subsequently be ignored by assuming that firms face perfectly elastic demands so that the mark-ups would fall to zero. The third is that operators are large enough to be able to exploit the fact that taxes and subsidies are set in response to their own actions, and can thus be manipulated by them. This aspect is, again, considered explicitly, in Section 4.