NORTHWESTERN UNIVERSITY

Analyzing and improving transportation infrastructure privatization in a competitive environment

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Civil and Environmental Engineering

By

Hang Shu

EVANSTON, ILLINOIS

September 2020

ABSTRACT

Analyzing and improving transportation infrastructure privatization in a competitive environment

Hang Shu

The transportation infrastructure privatization is an essential solution to the public fund's shortage of transportation investments. In the privatization practices, the private firms build, operate, and own the transportation infrastructure under a limited term. Due to the private sector's involvement, the efficiency of the transportation service is determined not only by the government's management but also by the strategic moves from private firms. In this dissertation, we analyze and improve the public-private partnership on the transportation infrastructure project considering the strategic interactions between the public sector and the private sector and among private firms themselves. In particular, we address two issues:

The first part of the dissertation aims to optimize the concessionaire selection and toll and capacity decision-making for the single road franchise under the cost information asymmetry. We systematically examine and analyze the existing contract methods under the general mechanism design framework. From the examination, we find out that the

most commonly used contracting approaches, the predetermined toll-capacity auction, and the competitive auctions, fail to limit the information rent and yield sub-optimal welfare for the public. To address this, we derive the optimal direct mechanism and its budget-balanced version as the theoretical benchmark. In addition, we study the implementation of these optimal mechanisms and invent a demand-based mechanism, which significantly improves the franchising mechanisms' practicality.

The second part of the dissertation analyzes the toll and capacity decisions of a single origin-destination private road network from the perspective of the government and multiple private firms. We extend the Pareto Efficiency analysis appearing in the literature for the single road case to the case with multiple roads owned by multiple firms. We also consider an oligopolistic market, where, (possibly) following bilateral negotiations with a government, self-interested, private operators set tolls and capacities independently. Characterization of the toll/capacity decisions from the bilateral negotiations suggests that the Pareto efficient decisions between the government and each firm are not sufficiently Pareto efficient in terms of the aggregate benefits of the whole network. We further analyze the situations and conditions for such Pareto inefficiency and provide Kaldor-Hicks improvements on the existing bilateral negotiation framework.

Acknowledgements

First of all, I would like to thank my advisor, Prof. Pablo L. Durango-Cohen. Without his support and guidance, I would not have made it. His mentorship helped me learn the essence of academic research: explorations, challenges, and perfections. I also want to thank Prof. Yu (Marco) Nie, who offered me great suggestions on the transportation network studies and general research directions. Prof. William Rogerson is anther great professor who I am very grateful to. He not only taught me the fundamentals of industrial organization but also helped me outline the mechanism design research that is present in this dissertation.

Also, I owe lots of thanks to my family, especially my parents and my grandparents. I am sorry that I have missed so many family gatherings. In addition, I would like to thank my girlfriend, Nancy, who has always been standing with me and supporting me through the good and tough times.

Table of Contents

List of Figures

CHAPTER 1

Introduction

In recent decades, the growing traffic demand and the limited public funds for transportation infrastructure intensify traffic congestion. In the last 30 years, vehicle miles traveled (VMT) in the US have increased by about 80%, while road mileage has only increased by approximately 5% (FHWA, 2013). Limited funding is one of the factors that restrict the development of public roads. Hagquist (2008) suggests that traditional transportation financing sources, such as the Highway Trust Fund, which rely on fuel taxes, severely restrict the development of transportation infrastructure. As an alternative, the federal government and numerous states are considering private participation in the development of transportation infrastructure. Thirty-two states and Puerto Rico have enacted legislation enabling Public-Private Partnerships (P3s) (Geddes and Wagner, 2013). In addition to the development of new roads, i.e., design, construction, operations, and management, governments can also franchise existing roads to private firms. For example, in 2005, the Skyway Corporation won a 99-year franchise for the Chicago Skyway and became the first privatization of an existing road in the US (Enright, 2006).

Private participation in transportation is also common in Europe, where limited resources also serve as the rationale. Albalate et al. (2009), for example, analyzed the level of private participation in tolled motorways in Europe. They found that 37% of roads (by length) are under concession agreements – most of them operated and maintained by private firms. The private sector plays a particularly important role in Southern Europe, where these trends are even more pronounced. There are also a number of (local) firms involved in the development of toll roads in China and elsewhere in Asia (Tan et al., 2010), where the implementation of P3s often takes the form of Build-Operate-Transfer agreements. In these agreements, firms transfer the roads back to the government at the end of a fixed-term (Yang and Meng, 2000). The Guangzhou-Shenzhen Super Highway Project, by a Hong Kong entrepreneur, is an emblematic example.

The typical life-cycle of the privatization of transportation infrastructure consists of five steps: project identification, planning/design, procurement, implementation, and transfer. In the beginning, the government evaluates the need for transportation infrastructure. Once the government decides to build transportation infrastructure, the project's technical details will be determined in the planning/design phase. Then, the government seeks private provision and investment. After the private concessionaires are determined, the government works with its private partners to implement the public-private contract. Commonly, after a fixed or flexible time period, the ownership of the project will be transferred to the government. Figure 1.1 summarizes the five steps. We focus on the intermediate three steps, planning/design, procurement, and implementation, in which the extensive strategic interactions from both the government and private agents are involved. Our topics include concessionaire selection, public-private negotiations, incentive creation for private participants, optimization of capacity and tolling decisions, and Pareto efficiency analysis of the public-private partnership.

Overall, this dissertation aims to analyze and improve the efficiency of transportation infrastructure privatization. In particular, we focus on the decision makings driven by strategic interactions between the public sector and firms in the highway franchise. The

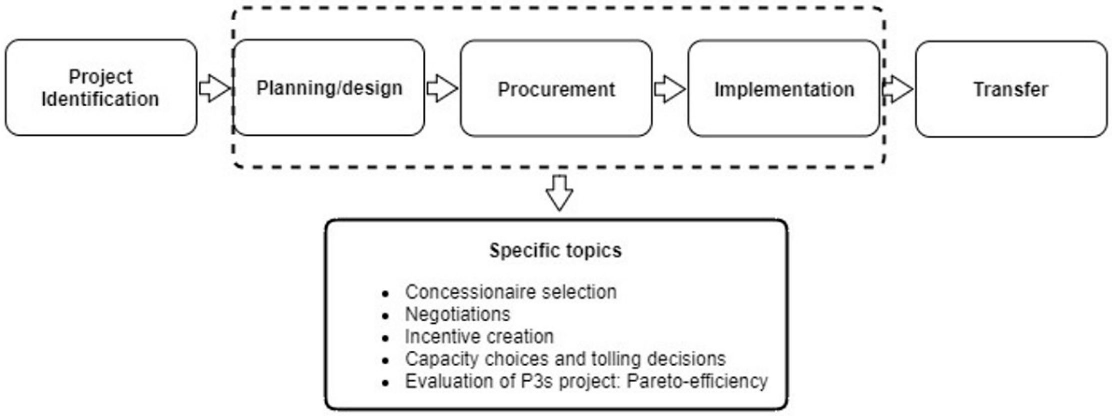


Figure 1.1. The life cycle of the transportation infrastructure privatization

remainder of this chapter discusses the motivation of our research, the research contributions, and the outline of this dissertation.

1.1. Motivation

Decision making in highway franchise has been addressed in the existing literature (de Palma and Lindsey, 2000; Verhoef, 2007; Tan et al., 2010). These researches focus on the interaction between the public sector and the private sector. However, very few of them attempt to address the complications stem from the competitions/interactions among private agents. In this vein, we think of two concerns regarding private interactions.

In the early period of the highway franchise, the government needs to select its private partner(s) and provides contracts for it(them). Commonly, there are multiple private firms competing for the franchise of a single road. Intuitively, the competition helps the government lower the franchise cost and select an efficient concessionaire. However, the government may or may not make use of the competition and optimize the welfare from privatization. In practice, the government's inability to optimize franchise is mainly due to the information asymmetry on the cost information. Thus, the government is unable to select the most efficient firm and to adjust the contracts, e.g., toll and capacity, according to the firm's productivity.

Another concern arises from the nature of transportation infrastructure: the network effects. When the government decides to privatize multiple road links in a highway network, the decision making on each link is mutually affected. Due to private ownership, the government is not able to manage the whole network in a centralized way. Instead, the government negotiates with each firm and set up contracts for each of them. Due to the self-interested nature of the private entity, such bilateral negotiations may not be Pareto efficient from the aggregate perspective, i.e., the benefits of the whole network.

In this dissertation, we address the aforementioned two concerns and optimize the franchising benefits with considerations on the competitive environment. Namely, as oppose to the public-private interaction, we focus on the realistic version of the franchise: the public-multiple-private interactions.

1.2. Contributions

The dissertation contributes to both the practices and theories of the highway franchise. The main contributions for the first part of the dissertation are:

(1) Apply the mechanism design theory to the highway franchise problem.

The concessionaire selection and toll/capacity optimization, under information asymmetry, are dependent on each other. By applying the mechanism design theory, we set up a generalized optimal control problem for the government to maximize welfare. In addition, we build on the classic model(Myerson, 1981) and decouple the concessionaire selection and toll/capacity optimization. As a result, we simplify the optimal franchising mechanism to the first-best information-symmetric optimization problem with distortion.

(2) Incorporate the budget balance constraint.

The general mechanism design framework makes the use of a contingent transfer from the government to the private firm. In order to address the public fund shortage, we incorporate a constraint on the transfer, so that the government never pays under any circumstances. The optimal franchising problem subject to budget constraint is proposed and solved. Further, we simplify the optimal budget-balanced franchising problem to the second-best information-symmetric optimization problem (Verhoef, 2007) with a distortion.

(3) Improve the implementations of the franchising mechanisms.

In addition to analyzing the theoretical optimal franchising mechanism, we also consider different ways to implement the optimal mechanism. We apply the scoring auction (Che, 1993) and analyze the deficiency of the naive scoring rule that is commonly used in the practice of competitive bidding. Moreover, we invent a novel demand-based mechanism, which significantly improves the transparency and convenience of the franchising process.

(4) Systematically examine and compare the current practices with the proposed mechanisms.

We build models for the current highway franchising practices, including competitive bidding, minimal cost auction, LPVR(least present value of revenue) auction, under the general mechanism design framework. In addition, we compare these practices with the proposed mechanisms and show a significant welfare increase from our proposed mechanisms.

The contributions for the second part of the dissertation are:

(1) Build the demand system to model the impact of toll/capacity decisions on the generalized road network.

We extend the analysis of toll/capacity decision in the highway from the single road(Tan et al., 2010) or two-link network(Verhoef, 2007) to a generalized road network, which shows how the substitution effects from the parallel links mix with the complement effects from the serial links. Our analytical results rely on a representation of the marginal demand system in matrix form, capturing the effects of changes in link tolls or capacities on either route or arc flows. These marginal effects are functions of congestion externalities, route structure, and demand elasticity. This characterization does not, as far as we are aware, appear elsewhere, and may be useful for comparative statics only superficially explored herein.

(2) Develop the optimizing framework from the perspective of each firm, the collection of firms and the government.

While the existing literature focuses on the social welfare gain from the road, we also examine the link profit and aggregate profit of the network. We not only extend the optimal v/c ratio results in the literature(Tan et al., 2010), but also reduce the dimensions of the optimization problems by applying the optimal v/c ratio.

(3) Define and analyze two versions of Pareto efficiency.

We define Aggregate Pareto efficiency and Decentralized Pareto efficiency to model two scenarios: the trade-off between the government and the collection of firms, and the trade-off between the government and each individual firms. We show that the Decentralized Pareto efficient toll/capacity decisions, resulting from the bilateral negotiations, may or may not be Aggregate Pareto efficient.

(4) Demonstrate the (possible)Pareto inefficiency of the bilateral negotiations and analyze its relationship with the network structure.

We further examine the bilateral negotiations on the toll/capacity decisions between the government and each private firm. We show the Pareto inefficiency only occurs under conditions that are related to both the demand side and the network structure. Moreover, as an example, we derive an inequality condition to identify if there is potential Pareto inefficiency.

1.3. Dissertation outline

The remainder of the dissertations is organized as follows:

Chapter 2 discusses our first topic on the single road franchising problem under cost information asymmetry. We focus on the early period of franchising: the government needs to select a concessionaire from multiple private bidders and optimize the franchising contract at the same time. We build on the generalized mechanism design framework and improve the implementations of the franchising approaches.

Chapter 3 presents the problem of oligopolistic competition under a generalized private road network. We evaluate the highway benefits from the perspective of the government, each individual firm, and the collection of firms. In addition, we evaluate the Pareto efficiency of the contracts set up by the bilateral negotiations between the government and each private firm.

The conclusion and further research directions are present in Chapter 4. Supplementary materials are included in the appendices at the end of each chapter, and the references are shown at the end of the dissertation.

CHAPTER 2

On the design of optimal auctions for concessions of private roads: Firm selection, government payments, toll and capacity levels under information asymmetry

2.1. Introduction

Private participation in the development of transportation infrastructure is ubiquitous, and is expected to grow further. In the US, for example, 32 states and Puerto Rico have enacted legislation enabling Public-Private Partnerships (Geddes and Wagner, 2013) . In addition to development of new roads, i.e., design, construction, operations and management, governments can also franchise existing roads to private firms. For example, in 2005, the Skyway Corporation won a 99-year franchise for the Chicago Skyway, and became the first privatization of an existing road in the US (Enright, 2006). Private participation in transportation is also common in Europe. Albalate et al. (2009), for example, analyzed the level of private participation in tolled motorways in Europe. They found that 37% of roads (by length) are under concession agreements – most of them operated and maintained by private firms. The private sector plays a particularly important role in Southern Europe, where these trends are even more pronounced. There are also a number of (local) firms involved in the development of toll roads in China and elsewhere in Asia (Tan et al., 2010), where the implementation of P3s often takes the form of Build-Operate-Transfer agreements. In these agreements firms transfer the roads back to the government at the end of a fixed term (Yang and Meng, 2000). The Guangzhou-Shenzhen Super Highway Project, by a Hong Kong entrepreneur, is an emblematic example.

In addition to supplementing public funding, privatization of roads may have other advantages. Blom-Hansen (2003), for example, points out that that lack of competition for the public sector may reduce its incentives to perform efficiently. Small et al. (1989) mentions that the general public is often more accepting of paying tolls charged by private operators, as opposed to by public agencies.

At the same time, private provision of transportation infrastructure does raise significant concerns. This is due, in part, to the misalignment of incentives between firms seeking to maximize profit, and governments seeking to maximize social welfare. In particular, as explained in Small et al. (1989), both the public sector, if it decides to toll, and firms have the incentive to absorb congestion externalities. Firms, however, charge a mark-up that depends on the demand elasticity, and that reduces social welfare. These concerns are reflected in the literature, where papers in the last 2 decades have analyzed the effect of privatization on project outcomes, i.e., tolling and capacity levels, and on social welfare in different settings: network topologies and ownership regimes. For example, the work of de Palma and Lindsey (2000) compares the social welfare of a 2-segment parallel road network, under 4 different ownership regimes: public-free, private-free, private-private and public-private. They find that the public-private duopoly results in lower social welfare than does the private-private one. Xiao et al. (2007) extend the analysis to the case of multiple parallel roads, each operated by a different firm, and show how additional competition affects the toll and capacity choices, and increases social welfare. Mun and Ahn (2008) study serial networks and find that the sum of the mark-ups set by independent, profit-maximizing operators exceeds the markup that would be charged by a single firm, i.e., a monopolist, operating the entire network. This means that decentralized, private operation of a pure serial network leads to reductions in both social welfare and in total operator profit! This phenomenon is referred to as double marginalization. In the same vein, Yang and Meng (2000, 2002) present a bi-level optimization framework to conduct the analysis on general networks with multiple OD pairs. They apply the framework to an example where they analyze the concession of a two-directional link in an inter-city expressway network in the Pearl River Delta Region of South China. They split the toll-capacity plane into 4 regions depending on whether the concession is profitable or not, and whether or not it increases social welfare vis-à-vis the network without the link.

In addition to identifying market structures that might be more amenable to privatization, a number of approaches have been studied in the literature to improve the welfare generated by road concessions. Among them, we note contract provisions restricting toll rates, toll revenues, capacity, rates of return (Tsai and Chu, 2003; Tan et al., 2010), government payments/subsidies or tax incentives (Zhang and Durango-Cohen, 2012), and, most relevant to the work herein, the use of auctions to award concessions.

At a high level, 2 types of approaches have appeared in the literature in the analysis of auctions for road franchising: the first, largely inspired by practice, where firms bid for a road concession where the quality variables, i.e., toll and capacity levels, are predetermined by the government; and the second, where quality variables are determined by the firms as part of the bidding strategy.

In the context of auctions with predetermined quality variables, minimum-cost auctions, where bidders submit cost estimates and the bidder with least cost wins, are the simplest and most widely-studied and used framework. Extensions include *A+B bidding*, where the cost associated with the time to complete the project, B, is added to in the project costs, A, to obtain the total costs of the the bid (Herbsman, 1995). The bidder with minimum $A+B$ cost wins the franchise. El-Rayes and Kandil (2005) further extends the it A+B cost measure to include a quality measure C. Other approaches considered in the literature include adding revenue metrics to the cost measure. Engel et al. (1997), for example, introduce the *least present value of revenue* (LPVR) auction, where bids are for the present value of toll revenue that they propose to collect over the duration of the franchise agreement. The firm that bids the LPVR wins the franchise. From the government's perspective, LPVR is appealing because firms compete. From the firm perspective, LPVR is attractive because it mitigates risks associated with demand uncertainty, and other factors, because franchise agreements extend until the specified LPVR is realized. Nombela and De Rus (2004) extend the LPVR to include costs. They refer to the criterion as the *least present value of net revenue*, where franchises are awarded on the basis of profit/income.

The simplicity and flexibility of these auctions explain their appeal. The papers above analyze their performance in settings involving uncertainty, scope changes, etc. Contractual provisions, e.g., government guarantees, renegotiation opportunities, to address these issues in implementations are common. Disadvantages that serve as motivation for the work herein include:

(1) The design burden, i.e., determining (optimal) toll and capacity levels, falls on the government. This is unappealing because governments may lack the technical expertise, or importantly, the information that is necessary. In particular,

welfare maximizing toll and capacity levels depend on a firm's production efficiency as reflected in its construction costs, which are unlikely to be known to the government in advance of the auction.

(2) The selection criteria are not aligned with the government's objective to maximize social welfare. In particular, firms, even those with a competitive advantage due to their high production efficiency, do not have incentive to build roads that exceed the minimum capacity or quality requirements because costs increase, thereby reducing the probability of winning.

Auctions with predetermined quality variables are discussed further in Section 2.3. In terms of auctions where firms determine the levels of quality variables as part of their bidding strategies, Verhoef (2007); Ubbels and Verhoef (2008) are the seminal studies in the context of road concessions. Their focus is to understand how the criterion used to select the winning bid, i.e., the scoring function, impacts the welfare generated by the project in different settings, i.e., roads in isolation, roads interacting with an untolled complement or substitute, and networks in general. They explain that the choice of criterion is critical in designing auctions where the bidders' strategies are contingent on private information. Because the objective is to assess and compare scoring functions, rather than developing a model that captures the information structure, they assume that the auctions are perfectly competitive, which allows them to determine the project outcomes by optimizing the aforementioned criteria subject to the constraint that firms earn normal/zero profits. The constraint ensures that firms have incentive to participate in the auction. Among the interesting results, they find that using demand/patronage maximization as a criterion maximizes social welfare, i.e., it yields the first-best solution for roads in isolation, and the second-best solution for roads in networks when spillovers exist.

In this paper, we relax the assumption of perfectly competitive auctions, which is predicated on having a large number of homogeneous bidders. Indeed, studies, such as Jofre-Bonet and Pesendorfer (2000), have shown that, in practice, auctions for road concessions often attract small numbers of firms that may differ greatly in terms of their capabilities. Following the bargaining model presented in Shi et al. (2016), in our model, heterogeneity is manifested in the firms' marginal construction costs, which is assumed to be private information. Also, to structure the model, each bidders' variable construction costs are assumed to be *iid* random variables drawn from a probability distribution function that is common knowledge to the firms and the government. Other important assumptions follow Verhoef (2007); Ubbels and Verhoef (2008). Here, we apply the framework of Myerson (1981) to design optimal road concession auctions yielding Nash Equilibrium bids, consisting of toll and capacity strategies, that maximize a project's expected social welfare. The framework relies on the Revelation Principle where, given the opponents' strategies, firms find it in their interest to reveal their private information, i.e., their marginal costs, as part of the bids (Myerson, 1979). In turn, this allows governments to allocate projects to the most efficient firms. We characterize optimal bidding strategies and observe that firms have incentive to exploit their private information by proposing roads with lower capacity and higher tolls than those obtained under perfect information or in the case of perfectly competitive auctions. The assumption of complete information and the Revelation Principle mean that, in turn, the government can extract profits that may stem from this distortion. Project outcomes depend on the realization of the marginal construction costs, but we observe that they approach those obtained under perfectly competitive auctions when there are a large number of bidders.

In terms of implementation and following Che (1993), we derive a scoring function corresponding to the optimal direct mechanism. The scoring function consists of 2 parts: the social welfare and a term that accounts for bidders' incentives to exploit their private information. We also compare and derive analytical bounds for the performance of 2 simple, but sub-optimal alternatives: a naïve scoring auction where the scoring function corresponds to the government's welfare maximization objective, but does not account for distortions; and, inspired by the Patronage Maximizing Auction of Verhoef (2007), we present a Demand Pricing Mechanism where firms select bundles consisting of a demand level and an associated government compensation. The Demand Pricing Mechanism is motivated by the complexity and potential lack of transparency associated with evaluating scoring functions. A numerical example shows that the Demand Pricing Mechanism performs almost as well as the Optimal Direct Mechanism because it partially account for distortions.

To conclude this section, we note that there is a synergistic literature in applied economics examining the design of procurement auctions with product characteristics, referred to as quality variables. McAfee and McMillan (1987); Riordan et al. (1986); Laffont and Tirole (1986) analyze the optimal revealed procurement mechanism with quality variables. In those mechanisms, each bidder submits their private information and the government specifies the choices of quality variable, price, and the winner. Che (1993) considers an application in the procurement of weapon systems, and develops the fundamentals of scoring auctions, where each bidder submits multi-dimensional production specifications and the government specifies a scoring function of their specifications. Scoring auctions usually follow the form of first or second price auctions, in which the score replaces the price. The scoring auction may not constitute the optimal procurement mechanism, but it has great practical value due to its simplicity to bidders. $A+B$ bidding is an example of a scoring auction, where the score is the sum of construction cost and time cost. This paper follows the theoretical development of the procurement auction, while emphasizing the application to highway franchising, where toll and capacity are the quality variables of interest.

The remainder of the paper is organized as follows. In Section 2.2, we present a model of production/cost function heterogeneity, and social optimal tolling and capacity choices. We then discuss basics of Mechanism Design and present our proposed public optimal direct mechanism along with its budget-balanced version in Section 2.3. In Section 2.4, we evaluate the scoring auction and propose a novel mechanism: Demand purchase in order to facilitate the direct mechanism. Section 2.6 provides some thoughts and implications from our proposed mechanism. Section 2.7 presents a numerical comparison between proposed mechanism and some commonly-used mechanisms. Conclusions and potential further developments appear in Section 2.8.

2.2. Toll-capacity choice under symmetric information

We consider a government wishing to franchise the construction and operation of a highway. Construction costs include materials, labor, and capital costs. We build on Small et al. (1989), who explains that road construction technology is such that costs can be reasonably approximated by the sum of a fixed cost term, and a linear variable cost term that is a function of road capacity. Empirical evidence, see e.g., Levinson and Gillen (1998), suggests that (fixed and variable) costs associated with equipment and material procurement do not vary widely across firms, but that there are significant differences across firms in variable costs associated with labor and financing/borrowing costs. Thus, as shown in (2.1), we consider a specification of firm construction costs where the fixed costs are homogeneous across firms, and the per-unit of capacity cost is firm-dependent. This is similar to Shi et al. (2016).

$$C_i^c(K) = c_0 + c_i \cdot K \tag{2.1}$$

where $c_0 > 0$ is the fixed cost, $c_i \geq 0, i = 1, \ldots, I$ is the ith firm's marginal cost. K denotes road capacity, and is measured in number of vehicles per hour. We assume that the fixed costs are *common knowledge*, and, in later sections, that the marginal costs are *private information*. Further, we assume the set of marginal costs are *iid* random variables drawn from the probability density function, $f(\cdot)$, with finite support, i.e., $c_i \in \mathcal{C} = [c^{min}, c^{max}]$. Letting $\boldsymbol{c} \equiv [c_1, \ldots, c_I]$ denote the collection of I firms' marginal costs, we can write the joint probability density function of $\boldsymbol{c}$ as

$$\phi(\boldsymbol{c}) = \prod_{i=1}^{I} f(c_i) \tag{2.2}$$

Having specified the construction cost function, and following Verhoef (2007), the social welfare or aggregate surplus, and profit associated with awarding a concession to firm i to build a road of capacity K are given as follows:

$$SW(N,K) = \int_{n=0}^{n=N} P(n)dn - N \cdot C(N,K) - C_i^c(K) \tag{2.3}$$

$$\pi_i(\tau, N, K) = N \cdot \tau - C_i^c(K) \tag{2.4}$$

The variables τ and N, which, respectively, represent the toll charge and the number of users, i.e., the demand/traffic, traveling through the road segment. $P(n)$ is the inverse demand function and corresponds to nth user's willingness to pay for travel, i.e., the nth user's utility of travel. $C(N,K)$ is the average travel cost per user. As is done elsewhere, we refer to $C(N,K)$ as the per-user average congestion cost function. It relates travel time to travel costs. Equation (2.3) excludes a term for toll revenues/payments, $N \cdot \tau$, because these transactions correspond to transfers between the parties.

We make the following assumptions for model tractability:

Assumption 1. *Demand and cost functions: First and second order conditions*

(1) *The inverse demand function, $P(n)$, is twice differentiable, decreasing, and concave, i.e., $P'(n) = \frac{\partial P(n)}{\partial n} < 0$ and $P''(n) = \frac{\partial^2 P(n)}{\partial n^2} \leq 0$.*

(2) *The congestion cost function is given as, $C(N,K) = g(\frac{N}{K})$, which is homogeneous of degree 0. Further, we let $\mu = \frac{N}{K}$, and assume that $g(\mu)$ is twice differentiable, increasing, and convex, i.e., $g'(\mu) = \frac{\partial g(\mu)}{\partial \mu} > 0$ and $g''(\mu) = \frac{\partial^2 g(\mu)}{\partial \mu^2} \geq 0$.*

μ is referred to as the *volume-to-capacity* or *v/c* ratio. Here, we assume that the relationship among operator's decision variables, τ and K, and the demand, N, is given by the *user equilibrium condition*, described in Assumption 2 below:

Assumption 2. *User Equilibrium*

The Nth user's willingness to pay equals the total generalized travel cost; that is,

$$P(N) = \tau + C(N, K)$$

which allows us to write the demand as a function of the decision variables as $N(\tau, K)$.

The above specification and assumptions lead to the following social welfare maximization problem, the First-Best, **FB**, problem, reflecting the government's *ex-post* objective subsequent to awarding the road concession to firm i:

$$\textbf{FB} \qquad \max_{\tau, K} \ SW(N, K) = \int_{n=0}^{n=N} P(n)dn - N \cdot C(N, K) - C_i^c(K)$$
$$\text{subject to:} \quad P(N) = \tau + C(N, K)$$

From the first-order conditions, we have

$$\begin{gathered} \tau^* = N \cdot C_N \\ K^* : -N \cdot C_K\big|_{K=K^*} = c_i \end{gathered} \tag{2.5}$$

where $C_N \equiv \frac{\partial C(N,K)}{\partial N}$ and $C_K \equiv \frac{\partial C(N,K)}{\partial K}$. The optimal toll, τ^*, internalizes the congestion externality, $N \cdot C_N$. This is the well-known result of Pigou (1932). We also note that the optimal capacity depends on a firm's construction efficiency, captured in c_i. Allowing for the congestion cost specification in Assumption 1.1,

$$\mu^* : g'(\mu) \cdot \mu^2\big|_{\mu=\mu^*} = c_i \tag{2.6}$$

Equation (2.6) implies that the optimal v/c ratio only depends on $g(\cdot)$ and c_i. Once the optimal v/c ratio, μ^*, is determined, the optimal values of the decision variables are given as follows:[1]

$$
\begin{aligned}
\tau^* &= \mu^* \cdot g'(\mu^*) = \frac{c_i}{\mu^*} \\
N^* &= P^{-1}\left(\tau^* + g(\mu^*)\right) \\
K^* &= \frac{N^*}{\mu^*}
\end{aligned}
$$

Letting $\mu^*(c_i)$ represent the optimal solution for marginal cost c_i, it follows from Assumption 1.3 that $\mu^*(c_i)$ is increasing in c_i, i.e., $\frac{\partial \mu^*(c_i)}{\partial c_i} > 0$. This means higher marginal construction costs lead to more congestion. Similarly, optimal tolls, capacities, and the induced demand are, respectively, increasing, decreasing, and decreasing in c_i. Again, letting the variables be defined as functions of c_i, $\frac{\partial \tau^*(c_i)}{\partial c_i} > 0$, $\frac{\partial K^*(c_i)}{\partial c_i} < 0$, and $\frac{\partial N^*(c_i)}{\partial c_i} < 0$. It follows that social welfare is decreasing in c_i.

In terms of the firm's profit, Equation (2.4), we note that $\pi_i(\tau^*, N^*, K^*) = -c_0 < 0$. That is, implementation of the optimal solution to **FB** does not allow firms to recover fixed costs, c_0, and thus, hinders firms' participation. Solutions analyzed in the literature include (i) subsidy mechanisms to compensate firms for the loss (or to ensure a minimum profit); and (ii) deriving tolls, capacities, and the induced demand via solution to the Ramsey Problem, **RP**, introduced by Ramsey (1927), and presented below:

[1] Assumption 1 and $P(0) > \frac{c_i}{\mu^*} g(\mu^*)$ ensure that $N^* > 0$, i.e., that the solution is interior, and that τ^*, N^*, K^* is the unique optimal solution to **FB**.

$$
\textbf{RP} \qquad \max_{\tau,K} \; SW(N,K) = \int_{n=0}^{n=N} P(n)dn - N \cdot C(N,K) - C_i^c(K)
$$

$$
\text{subject to:} \quad P(N) = \tau + C(N,K)
$$

$$
\pi_i(\tau, N, K) \geq 0
$$

The solution to **RP** yields the maximum social welfare that ensures viability from firm i's standpoint.

Agreement between a government and a firm on the implementation of solutions to **FB** or **RP** requires information on c_i, which we assume is unknown to the government. In a general setting with multiple firms, this highlights difficulties in the process of awarding a concession, where it is desirable to select the most efficient firm. In the following section, we present a mechanism where firms have the incentive to reveal their marginal cost, and set tolls and capacities to the solutions of **FB** or **RP**, depending on the situation.

2.3. Analysis of concessions via auctions

Integrated models considering firm screening/selection and toll-capacity decisions have appeared in the literature (Yang and Meng, 2000; Ubbels and Verhoef, 2008). Most analyses rely on the assumption of complete information regarding the firm's cost structure. Shi et al. (2016) is a notable exception, but, rather than firm selection, the focus is contracting with a given firm. In practice, the process of firm selection is generally separated from toll-capacity decisions. That is, the government selects a firm through an auction and then negotiates with the winner for an optimal contract.

There are two major concerns on such a process:

(1) Common auctions, such as first-price and second-price auction, have to be adapted for the new type of private information, i.e., production efficiency.

(2) Toll and capacity, determined after the auction, affect bidder's profitability from the highway, and therefore the bidding strategy in the auction. We mainly address them in this section.

Here, we consider a government interested in maximizing public welfare, PW, which is defined as follows:

$$PW = CS - M = \int_0^N P(n)dn - N \cdot C(N, K) - N\tau - M \tag{2.7}$$

where M denotes a direct government payment/subsidy to the firm. PW, therefore, corresponds to the part of the social welfare realized by the users/consumers and the government, i.e., it is the welfare that is left after the operating profit, π, and direct payment, M, are excluded. Mathematically, $SW = CS - M + N\tau - C_i^c(K) + M = PW + \pi + M$. The last two terms represent the firm's total profit: operating profit and direct payment. The inclusion of the direct payment does not change SW, because it is a transfer between the parties, but it does change PW. When M is set to ensure that the firm's total profit is non-negative, i.e., when the constraint $\pi + M \geq 0$ is imposed in the PW maximization problem, we note that the constraint is binding, $M = -\pi$, and that the PW maximization problem is equivalent to **FB**. In the remainder of this section, we consider the design of public welfare optimizing mechanisms, combining the ideas of firm selection and toll-capacity optimization.

2.3.1. Predetermined toll-capacity bidding

In conventional auctions, governments set requirements for toll and capacity levels. For example, they may optimize welfare based on an estimate of the marginal construction cost, $\bar{c}$, which they may use to set toll and capacity, $\tau^*(\bar{c})$ and $K^*(\bar{c})$.[2] The government payment, M, is the key bid element. Minimum cost, *A+B*, *LPVR*, and *LPVNR* auctions all fall into this category. Given a predetermined toll and capacity parametrized by $\bar{c}$, firm i's operating profit is:

$$
\begin{aligned}
\pi(c_i; \tau^*(\bar{c}), N^*(\bar{c}), K^*(\bar{c})) &= N^*(\bar{c})\tau^*(\bar{c}) - C_i^c(K^*(\bar{c})) \\
&= N^*(\bar{c})\tau^*(\bar{c}) - (c_0 + \bar{c} \cdot K^*(\bar{c})) + (c_0 + \bar{c} \cdot K^*(\bar{c})) \\
&\quad - (c_0 + c_i \cdot K^*(\bar{c})) \\
&= N^*(\bar{c}) \cdot \frac{\bar{c}}{\mu^*(\bar{c})} - c_0 - \bar{c} \cdot \frac{N^*(\bar{c})}{\mu^*(\bar{c})} + (c_0 + \bar{c} \cdot K^*(\bar{c})) \\
&\quad - (c_0 + c_i \cdot K^*(\bar{c})) \\
&= (\bar{c} - c_i)K^*(\bar{c}) - c_0
\end{aligned}
\tag{2.8}
$$

as defined earlier, $N^*(\bar{c})$ is the demand induced by **FB** maximizing toll and capacity levels for marginal cost $\bar{c}$, $N^*(\bar{c}) \equiv N^*(\tau^*(\bar{c}), K^*(\bar{c}))$. Firm i's revenues depend on $\bar{c}$, but its (construction) costs are based on its actual marginal cost, c_i. We observe that $\pi(c_i; \tau^*(\bar{c}), N^*(\bar{c}), K^*(\bar{c}))$ decreases linearly with c_i. Having specified the probability density function for firm i's marginal costs, $f(c_i)$, for a given $\bar{c}$, the corresponding probability density function of firm i's profit, $h(y)$, is given by

[2] $\bar{c}$ may, for example, be obtained by calculating the expected marginal cost, $E[c_i]$.

$$h(\bar{c}; y) = f(\pi^{-1}(y; \tau(\bar{c}), N(\bar{c}), K(\bar{c}))),$$
$$y \in [\pi^{min} = \pi(c^{max}; \tau(\bar{c}), N(\bar{c}), K(\bar{c})), \pi^{max} = \pi(c^{min}; \tau(\bar{c}), N(\bar{c}), K(\bar{c}))] \tag{2.9}$$

where $\pi^{-1}(\cdot)$ is the profit function's *inverse*. This shows that considering a firm's profit as its private information is equivalent to considering its marginal cost. In minimum cost auctions with predetermined toll-capacity levels, bidder i submits its requested government payment, M_i. Thus, the optimal bidding strategies maximize

$$\{\pi(c_i; \tau^*(\bar{c}), N^*(\bar{c}), K^*(\bar{c})) + M_i\} \cdot P(\text{winning with } M_i)$$

. In LPVR (Enright, 2006) or LPVNR (De Rus and Nombela, 2000) auctions, firms bid on the revenue.[3] Thus, bidding strategies maximize

$$\begin{aligned}[M_i - C_i^c(K^*(\bar{c})]P(\text{winning with } M_i) &= P(\text{winning with } M_i)[M_i \\ &\quad + \pi(c_i; \tau^*(\bar{c}), K^*(\bar{c})) - N^*(\bar{c}) \cdot \tau^*(\bar{c})]\end{aligned}$$

Both types of auctions lead to the same bids because $N^*(\bar{c}) \cdot \tau^*(\bar{c})$ is constant. The outcomes of predetermined toll-capacity bidding are presented below.

Proposition 1. *Outcomes of predetermined toll-capacity bidding*

[3] In LPVNR auctions, the net revenue excludes the initial construction costs.

(1) *Bidder i's interim private surplus, i.e., the total profit consisting of operating profit and subsidy, is*

$$\int_{\pi^{min}}^{\upsilon_i} H^{I-1}(y)dy = K^*(\bar{c}) \int_{c_i}^{c^{max}} F^{I-1}(x)dx \tag{2.10}$$

where $H(\cdot)$ represents the cumulative distribution function corresponding to density $h(\cdot)$, and $\upsilon_i = \pi(c_i; \tau^*(\bar{c}), N^*(\bar{c}), K^*(\bar{c}))$ denotes firm i's ex-post profit.

(2) *The ex-ante public welfare is*

$$E_x \left[CS(\bar{c}) + \pi(x; \bar{c}) - \frac{F(x)}{f(x)} \cdot K^*(\bar{c}) \right] \tag{2.11}$$

where E_x denotes the expectation over the winning firm's marginal cost, x.

Proof. Since both auctions are first-price auction, (2.10) simply follows existing results from Riley and Samuelson (1981). A derivation of (2.11) is in 2.9.

The sum of the first two terms inside the expectation of (2.11) is the *ex-post* public welfare when the winning firm has a unit cost x. The third term is a distortion related to the distribution, which will be extensively discussed in the following sections. Besides the distortion term, the choice of toll-capacity is optimized according to the given $\bar{c}$ instead of the true type.

2.3.2. Direct mechanism with toll-capacity screening

Before presenting the optimal mechanism, we introduce the general form of a mechanism (W, Π, A), which consists of the following: I heterogeneous bidders; a set of messages W_i from each bidder i to the auctioneer; an allocation rule $\Pi : \boldsymbol{W} \mapsto \Delta$, where Δ is the probability distribution of winning over all bidders; Payment rule $A : \boldsymbol{W} \mapsto \mathbb{R}^I$. Notice that the aforementioned predetermined cases, in which the payment is the message itself, fit into the general framework as well. Since our focus is not the dynamics of bidding but static results at equilibrium. We can simplify our analysis by studying equivalent mechanisms with the same equilibrium, based on the Revelation Principle (Myerson, 1979):

Theorem 1. *(Revelation Principle) Given a mechanism and an equilibrium for that mechanism, there exists a direct mechanism in which: each bidder i truthfully report their own private information c_i; allocations and payments are contingent on the reported information; the outcomes are the same as in the given equilibrium of the original mechanism. Therefore, a direct mechanism can be defined as (Q, M), where $Q : \mathcal{C}^I \mapsto \Delta$ is allocation rule and $M : \mathcal{C}^I \mapsto \mathbb{R}^I$ is payment rule.*

The Revelation Principle simplifies the Mechanism Design Problem because it is only necessary to consider the optimization of truth-telling mechanism. Under the direct mechanism, each firm i truthfully reports its unit cost c_i, receives $M_i(\boldsymbol{c})$ as compensation, and is awarded the franchise with probability $q_i(\boldsymbol{c})$, which follows an allocation rule $Q(\boldsymbol{c})$. It is worth noting that the payment could be positive, negative or zero depending on the profitability of the road. To simplify the allocation rule, we impose the following assumption,

Assumption 3. *For each collection of (reported) marginal costs $\boldsymbol{c}$, the franchise is awarded to exactly one firm. That is, $Q(\boldsymbol{c}) \equiv [q_1(\boldsymbol{c}), \ldots, q_I(\boldsymbol{c})]$ is such that $q_i(\boldsymbol{c}) \in \{0, 1\}, i = 1, \ldots, I$, and $\sum_{i=1}^{I} q_i(\boldsymbol{c}) = 1$.*

A tie-breaking rule may be needed in cases where multiple firms report the same marginal cost. Toll and capacity schedules are specified as follows:

Definition 1. *Toll and capacity schedules*
Define toll and capacity schedule as function $T : \mathcal{C} \mapsto \mathbb{R}^+$, and function $L : \mathcal{C} \mapsto \mathbb{R}^+$, respectively. A bidder who reports c_i has to set the toll at $T(c_i)$ and capacity at $L(c_i)$ if it wins the franchise. The mapping T and L is published by the government prior to the bidding.

Recall that we specify the first-best toll and capacity function $\tau^*(\cdot)$ and $K^*(\cdot)$ under symmetric information. This schedule is supposed to specify the optimal toll-capacity matching the winning firm's production efficiency. In addition, this schedule would further screen the concessionaire. That is, the firm's toll-capacity choices reveal its production efficiency. Ideally, under a carefully designed mechanism, a more efficient firm tends to set up a higher capacity and lower toll. We refer to such a mechanism as the toll-capacity screening mechanism. We summarize the process as follows:

(1) The government publish the subsidy (or payment) rule, M, toll and capacity schedule, T and L, and allocation rule Q, to private firms, which specify how much a firm expect to receive(or pay if negative), what proposed toll and capacity is and how to determine the winner, based on the reported production efficiency measures $\boldsymbol{c}$.
(2) Private firms(truthfully) report their production efficiency, induced by the direct mechanism.

(3) The government determines the winner of franchise based on their reported production efficiency.

(4) winner receives M, and builds the road with toll-capacity specified by T and L.

The keys to designing direct mechanisms are to (i) induce bidders to report their true marginal costs; and (ii) ensuring that bidders have the incentive to participate. The first condition is referred to as *Incentive Compatibility* (IC) condition, and the second condition is referred to as *Individual Rationality* (IR). Relevant versions of the 2 conditions are presented below.

Definition 2. *Bayesian Incentive Compatibility*

For a toll and capacity screening mechanism (Q, M, T, L) bidder i's expected utility of reporting $\tilde{c}_i$, $U_i(\cdot)$, is defined as

$$U_i(\tilde{c}_i, c_i) = E_{\boldsymbol{c}_{-i}}\left[q_i(\tilde{c}_i, \boldsymbol{c}_{-i}) \cdot \pi(c_i; T(\tilde{c}_i), N(\tilde{c}_i), L(\tilde{c}_i)) + M_i(\tilde{c}_i, \boldsymbol{c}_{-i})\right] \tag{2.12}$$

The expectation is taken over the marginal costs of the $I-1$ firms other than firm i, denoted $\boldsymbol{c}_{-i}$. Further, (Q, M, T, L) is said to be incentive compatible if and only if

$$V_i(c_i) \equiv U_i(c_i, c_i) \geq U_i(\tilde{c}_i, c_i) \quad \tilde{c}_i \in \mathcal{C}, i = 1, \ldots, I. \tag{2.13}$$

where $V_i(c_i)$ is the expected utility under the truth-telling strategy.

Notice the above condition is a special case of Bayesian Incentive Compatibility because we assume that each bidder's belief always follows the common prior distribution. It guarantees that, at the Bayesian Nash Equilibrium, each firm i's strategy is to report its true unit cost. Equation (2.12) shows that the reported marginal cost affects a firm's chance of winning, toll and capacity outcomes, and the payment it receives. It does not, however, impact its cost structure. Essentially, a screening mechanism satisfying incentive compatibility punishes firms that report false marginal costs.

Another requirement of the mechanism is to guarantee enough incentive to participants. In the context of complete information, as Verhoef (2007) presents, the government optimizes social welfare under the condition that each bidder can earn at least zero profit. For the case of asymmetric information, the government has to guarantee that any firm with any possible unit cost earns at least as much as their outside option. Formally, we refer to such property as Individual Rationality (IR):

Definition 3. *Interim Individual Rationality*
Each bidder's expected utility of participating is higher than the utility of the outside option.

$$V_i(c_i) \geq 0 \quad i = 1, \ldots, I \tag{2.14}$$

Situations where the expected value of the *do-nothing* alternative are not equal to zero can be handled by adjusting the right-hand-side of the above inequality. We observe that if the expected utility decreases with type, c_i, then the IR condition is equivalent to imposing a non-negative (expected) profit constraint on the firm with the highest marginal cost. That is, IR boils down to the participation constraint of the critical type, i.e., to $V_i(c^{max}) \geq 0$.

Notice that *interim* IR does not guarantee a non-negative pay-off for all opponents' profiles, $\boldsymbol{c}_{-i}$. That is, there may be cases where a firm may regret to participate in a mechanism under *interim* IR. A stronger version of individual rationality is defined below,

Definition 4. *Ex-post Individual Rationality*

We define the ex-post utility of a firm given own type c_i *and others' types* $\boldsymbol{c}_{-i}$ *, under mechanism* (Q, M, T, L) *as,*

$$u_i(c_i, \boldsymbol{c}_{-i}) = q_i(c_i, \boldsymbol{c}_{-i}) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) + M_i(c_i, \boldsymbol{c}_{-i}) \tag{2.15}$$

Ex-post individual rationality holds if and only if,

$$u_i(c_i, \boldsymbol{c}_{-i}) \geq 0 \quad c_i \in \mathcal{C}, \boldsymbol{c}_{-i} \in \mathcal{C}^{I-1} \tag{2.16}$$

Thus, *ex-post* IR ensures that every firm has the incentive to participate in the auction without regard to its own or its opponents' marginal costs. *Ex-post* IR eliminates the risk of ending with a loss. However, it does not work well with the Bayesian IC, which is defined at the *interim* level. In the following sections, we firstly relax the IR to *interim* level and then discuss approaches to tighten the constraints. Hereafter, IR without specification refers to *interim* IR.

2.3.3. Optimal Mechanism Design

Given the Revelation Principle, the task is to find an optimal allocation rule Q, toll-capacity schedule T, L, and payment rule M that satisfy IC and IR. Following Myerson (1981), the Generalized Payoff Equivalence presented below provides a tractable form of IC constraint.

Proposition 2. *Generalization of Payoff Equivalence*

Consider a toll-capacity screening mechanism (Q, M, T, L)*. Let* $\pi(c_i; T(\tilde{c}_i), N(\tilde{c}_i), L(\tilde{c}_i))$ *denote the profit of a winning firm with productivity* c_i *claiming* $\tilde{c}_i$*. Define function* $G_i : \mathcal{C}^2 \mapsto \mathbb{R}$ *as,*

$$G_i(\tilde{c}_i, c_i) = E_{\boldsymbol{c}_{-i}}\left[q_i(\tilde{c}_i, \boldsymbol{c}_{-i}) \cdot \pi'(c_i; T(\tilde{c}_i), N(\tilde{c}_i), L(\tilde{c}_i))\right] \tag{2.17}$$

where $\pi'(\cdot)$ *is the partial derivative of profit function with respect to the true marginal cost,* $\pi'(\cdot) \equiv \frac{\partial \pi(\cdot)}{\partial c_i}$*. That is,* $G_i(\cdot)$ *is the partial derivative of* $U_i(\cdot)$ *with respect to the true type,* c_i*. Let* $S_i(c_i)$ *be the value of* $G_i(\cdot)$ *when firm* i *claims* c_i*, i.e.,* $S(c_i) \equiv G_i(c_i, c_i)$*.* (Q, M, T, L) *satisfies IC if and only if,*

(1) S_i *is the gradient of* V_i*,*

$$V_i(c_i) = V_i(x_1) + \int_{x_1}^{c_i} S_i(x)dx \quad \forall x_1 \in \mathcal{C}, c_i \in \mathcal{C} \tag{2.18}$$

(2) $G_i(\cdot)$ *satisfies the Generalized Monotonicity Condition presented below:*

$$\int_{x_1}^{c_i} \left(G_i(c_i, x) - G_i(x, x)\right) dx \geq 0, \quad \forall x_1 \in \mathcal{C}, c_i \in \mathcal{C} \tag{2.19}$$

Thus, the conditions are equivalent. □

The Generalized Monotonicity Condition is, essentially, a second-order requirement for IC. As is shown in Myerson (1981); Jehiel et al. (1999) in the context of the private value auctions

$$\int_{c_i'}^{c_i} (G_i(c_i, x) - G_i(x, x))\, dx \geq 0 \Leftrightarrow \int_{c_i'}^{c_i} (\mathbb{P}_i(c_i) - \mathbb{P}_i(x))\, dx \geq 0 \tag{2.20}$$

where $\mathbb{P}_i(c_i) \equiv E_{\boldsymbol{c}_{-i}}[q_i(\boldsymbol{c})] = E_{\boldsymbol{c}_{-i}}[q_i(c_i, \boldsymbol{c}_{-i})]\,, \quad c_i \in \mathcal{C}, i = 1, \ldots, I$, and corresponds to firm i's *interim* probability of winning, i.e., the probability of winning given the firm's own type, c_i.[4] Also, $\mathbb{P}_i(c_i)$ is monotonically-decreasing in c_i.

The mechanism design problem, **MD**, involves PW maximization, subject to user equilibrium, IC, and IR. Since the government is not able to observe the realization of $\boldsymbol{c}$ before setting the rules, the objective is written *ex-ante*. That is, the government optimizes the expected PW, based on the prior distribution.

$$\textbf{MD} \qquad \max_{Q,T,L,M} \qquad E_{\boldsymbol{c}}\left[\sum_{i=1}^{I} (q_i(\boldsymbol{c}) \cdot CS(T(c_i), N(c_i), L(c_i))) - M_i(\boldsymbol{c})\right]$$

subject to:

$$\begin{aligned}
P(N(c_i)) &= C(N(c_i), L(c_i)) + T(c_i) && (2.21)\\
V_i(c_i) &= E_{\boldsymbol{c}_{-i}}[q_i(\boldsymbol{c}) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) + M_i(\boldsymbol{c})] && (2.22)\\
\frac{\partial V_i(c_i)}{\partial c_i} &= S_i(c_i) && (2.23)\\
S_i(c_i') &\leq S_i(c_i), \quad c_i' \leq c_i && (2.24)\\
V_i(c^{max}) &\geq 0 && (2.25)\\
\forall c_i, c_i' &\in \mathcal{C}, \quad \forall i = 1, \ldots I
\end{aligned}$$

[4]Due to the independence assumption, it is not necessary to consider the conditional expectation, $E_{\boldsymbol{c}_{-i}|c_i}[q_i(\boldsymbol{c})]$.

where CS is the consumer surplus function, defined earlier. The objective is to maximize the expected PW, where Equations (2.21)–(2.25) are, respectively, the user equilibrium, indirect utility, Envelope Theorem, Monotonicity, and IR conditions. Also, we note that $G_i(\tilde{c}_i, c_i) = -\mathbb{P}_i(\tilde{c}_i) \cdot L(\tilde{c}_i)$, which means that $S_i(c_i) = -\mathbb{P}_i(c_i) \cdot L(c_i)$.

MD can be re-written by substituting (2.21) and (2.22) in the objective function. As is done in Basov (2006) and shown in 2.9, integration by parts and other steps are used to obtain the following simplified version of the problem, **MD-S**.

$$\textbf{MD-S} \quad \max_{Q,T,L} \qquad E_{\boldsymbol{c}}\left[\sum_{i=1}^{I} q_i(\boldsymbol{c}) \cdot \left(SW(c_i; T(c_i), N(c_i), L(c_i)) - \frac{F(c_i)}{f(c_i)} L(c_i) - V_0\right)\right]$$

subject to:

$$\begin{aligned}
\mathbb{P}_i(c_i) \cdot L(c_i) &\leq \mathbb{P}_i(c_i') \cdot L(c_i'), \quad c_i' \leq c_i && (2.26)\\
V_0 = V_i(c^{max}) &\geq 0 && (2.27)\\
\forall c_i, c_i' &\in \mathcal{C}, \quad \forall i = 1, \ldots I
\end{aligned}$$

where $SW(c_i; T(c_i), N(c_i), L(c_i))$ is the social welfare, given c_i, the toll schedule, $T(\cdot)$, the capacity schedule, $L(\cdot)$, and the induced demand, $N(c_i)$. We define firm i's virtual surplus as

$$R(c_i; T(c_i), L(c_i)) = SW(c_i) - \frac{F(c_i)}{f(c_i)} \cdot L(c_i) \tag{2.28}$$

Notice that $R(\cdot)$ captures the ith firm's contribution to the expected public welfare given $T(\cdot), L(\cdot)$. To see this, we rewrite the objective as

$$E_{\boldsymbol{c}}\left[\sum_{i=1}^{I} q_i(\boldsymbol{c}) \cdot R(c_i; T(c_i), L(c_i))\right]$$

Since there is only one winner, i.e., $\sum_{i=1}^{I} q_i(\boldsymbol{c}) = 1$. **MD-S** boils down to the maximization of virtual surplus received from the winner. One should notice that, even though we can equivalently take the virtual surplus as the expected contribution to the public welfare, the virtual surplus is different from the actual surplus, i.e., the actual public welfare given the bidder's type. The optimal allocation rule, therefore, is to award the franchise to the bidder with the highest virtual surplus. However, $R(\cdot)$ depends on the prior distribution f, and on the capacity schedule $L(\cdot)$. The monotonicity constraint further complicates the relationship among Q, T and L, when maximizing the virtual surplus. Below, we analyze the optimal rules and discuss the relevant assumptions.

2.3.4. Optimal allocation, toll and capacity levels

The definition of the virtual surplus function allows for separation of the optimization of each $R(c_i; T, L)$ with respect to the toll and capacity schedules, from the concessionaire selection captured in the allocation rule, Q. Expanding the virtual surplus function, we obtain

$$\begin{aligned} R(c_i;T,L) &= \int_0^{N(T(c_i),L(c_i))} P(n)dn - N(T(c_i),L(c_i)) \cdot C(N(T(c_i),L(c_i)),L(c_i)) \\ &\quad - C^c(c_i,L(c_i)) - \frac{F(c_i)}{f(c_i)} L(c_i) \\ &= \int_0^{N(T(c_i),L(c_i))} P(n)dn - N(T(c_i),L(c_i)) \cdot C(N(T(c_i),L(c_i)),L(c_i)) \\ &\quad - c_0 - (c_i + \frac{F(c_i)}{f(c_i)})L(c_i) \end{aligned} \tag{2.29}$$

which shows that the objective is to maximize social welfare with a distorted/virtual marginal cost, $\gamma_i(c_i)$, where $\gamma_i(c_i) \equiv c_i + \frac{F(c_i)}{f(c_i)}$. One should notice that there is no distortion when the unit cost is at the lower bound because $F(c^{min}) = 0$. This is the relevant version of the "no distortion at the top" result. A firm with marginal cost at the lower bound, c^{min}, always has the lowest virtual unit cost, but in general, a lower marginal cost does not imply a lower virtual marginal cost. Whereas in the case of complete information, the optimal allocation rule is to award the concession to most efficient firm, distortions arising in the case of asymmetric information mean that less efficient firms can generate higher virtual surpluses. Thus, to simplify the analysis, we consider a relevant variation of the *Regularity Assumption* commonly appearing in the Mechanism Design literature (Che, 1993; Myerson, 1981).

Assumption 4. *Regularity*

The virtual unit cost, $\gamma_i(\cdot)$, is non-decreasing in c_i, i.e., $\frac{\partial \gamma_i(c_i)}{\partial c_i} \geq 0$.

This assumption ensures the *ex-post* optimality of awarding the franchise to the firm with the lowest virtual unit cost.[5] The solution to the problem of maximizing the virtual surplus, Equation (2.29), is obtained from the results to **FB** with the distorted marginal cost. Namely, $T^*(c_i) = \tau^*(\gamma(c_i))$ and $L^*(c_i) = K^*(\gamma(c_i))$. Since $\gamma_i(c_i) \geq c_i$, we observe that the distortions from information asymmetry lead to (i) higher optimal tolls than in **FB**, i.e., $T^*(c_i) \geq \tau^*(c_i)$, (ii) lower optimal capacity, (iii) lower optimal demand, and (iv) more congestion, i.e., higher optimal v/c ratios. Also, following the discussion in Section 2.2 and Assumption 4 guarantee that $T^*(c_i)$ and $L^*(c_i)$ are increasing and decreasing in c_i, i.e., $\frac{\partial T^*(c_i)}{\partial c_i} > 0$ and $\frac{\partial L^*(c_i)}{\partial c_i} < 0$.

Finally, we observe that Assumption 4 ensures that constraint (2.26) in **MD-S** is always satisfied. Firms with higher marginal costs, c_i, have higher virtual marginal costs, $\gamma_i(c_i)$, and therefore, smaller chance to win the franchise, i.e, $\mathbb{P}_i(c_i)$ is non-increasing with c_i. Also, as discussed above, $L^*(c_i)$ decreases with c_i. Thus, $\mathbb{P}_i(c_i) \cdot L(c_i)$ is non-increasing with c_i.

2.3.5. Optimal government payments

Next, we examine the optimal payment rule and IR conditions. From (2.22) and (2.23) in **MD-S**, the *interim* or conditional expected payment is given by

[5]We do note that ironing techniques have been developed to address cases where Assumption 4 does not hold (Rochet and Choné, 1998). We also note that the assumption is not as strong as it seems because distributions with non-increasing probability density, such as the uniform and exponential distributions are sufficiently regular. Certain instances of normal and truncated normal distributions satisfy Assumption 4, as well.

$$\bar{M}_i(c_i) \equiv E_{\boldsymbol{c}_{-i}}[M_i(c_i, \boldsymbol{c}_{-i})] = V_0 - \int_{c^{max}}^{c_i} \mathbb{P}_i(x)L(x)dx - \mathbb{P}_i(c_i) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) \tag{2.30}$$

Since the IR constraint is binding at optimality, i.e., $V_0 = 0$, the first term can be removed from the above expression.

The constraints in **MD-S**, therefore, allow for flexibility in selecting *ex-post* payment rules as long as their conditional expectations satisfy (2.30). For example, one may set the *ex-post* payments to be the same as the interim payments. While attractive because a firms' payment only depends on its own c_i (but not others), such a rule violates *ex-post* IR, i.e., the winner may not have sufficient subsidy, while some firms may receive a free payment. Thus, we propose the following *ex-post* payment rule for firm i:

$$M_i^*(\boldsymbol{c}) = -q_i(\boldsymbol{c}) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) - \int_{c^{max}}^{c_i} q_i(x, \boldsymbol{c}_{-i})L(x)dx \tag{2.31}$$

where $E_{\boldsymbol{c}_{-i}}[M_i^*(\boldsymbol{c})] = \bar{M}_i(c_i)$. $M_i^*(\boldsymbol{c})$ is *ex-post* IR follows from

$$\begin{aligned} u_i(c_i, \boldsymbol{c}_{-i}) &= q_i(\boldsymbol{c}) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) + M_i^*(\boldsymbol{c}) \\ &= -\int_{c^{max}}^{c_i} q_i(x, \boldsymbol{c}_{-i})L(x)dx \geq 0 \end{aligned}$$

where the inequality is due to the fact that $c_i \leq c^{max}$ and $q_i(x, \boldsymbol{c}_{-i})L(x)$ is non-negative. The integral term in the expression is referred to as the information rent. Notice that, for firm i, the integral's argument is zero, i.e., $q_i(x, \boldsymbol{c}_{-i}) = 0$, until x becomes the best bid.

Because exactly one firm wins, the *ex-post* information rent, i.e., once all marginal costs are submitted, is

$$-\int_{c^{max}}^{c_w} q_w(x, \boldsymbol{c}_{-i}) \cdot L(x)dx = \int_{c_w}^{c_l} L(x)dx \tag{2.32}$$

where firm w reports the lowest marginal cost, and firm l reports the second-lowest marginal cost. That is, it is a generalization of the well-known Vickery Auction (or the second-price auction), where the highest bidder wins, but pays a price equal to the second highest bid, and realizes a value equal to the difference in the bids. The optimal solution to **MD-S** is summarized below:

(1) Q^p is given by $q_i^p(\boldsymbol{c}) = \begin{cases} 1 & ; \quad c_i = \min\{c_1, c_2, \ldots, c_I\} \\ 0 & ; \quad \text{otherwise} \end{cases}$.

A tie-breaking rule may be necessary.

(2) $T^p(c_i) = \tau^*(c_i + \frac{F(c_i)}{f(c_i)})$.

(3) $L^p(c_i) = K^*(c_i + \frac{F(c_i)}{f(c_i)})$.

(4) $M_i^p(\boldsymbol{c}) = -q_i(\boldsymbol{c}) \cdot \pi(c_i; T(c_i), N(c_i), L(c_i)) - \int_{c^{max}}^{c_i} q_i(x, \boldsymbol{c}_{-i})L(x)dx$.

2.3.6. Budget balance

The shortage of public funds is one of the motivations to consider road concessions. Thus far, we have not incorporated budget restrictions on the payment, M, in the analysis. The optimal mechanism is the asymmetric-information structured counterpart of **FB**. In this section, we present a budget-balanced optimal mechanism that is the asymmetric-information structured counterpart of **RP**. From the results in the previous section, the *ex-post* government payment is

$$\begin{aligned} M_w^p(\boldsymbol{c}) &= -\int_{c^{max}}^{c_w} q_w(x, \boldsymbol{c}_{-w}) L^p(x)dx - q_w(\boldsymbol{c}) \cdot \pi(c_w; T^p(c_w), N^p(c_w), L^p(c_w)) \\ &= \int_{c_w}^{c_l} L^p(x)dx - \pi(c_w; T^p(c_w), N^p(c_w), L^p(c_w)) \end{aligned} \tag{2.33}$$

where w and l denote the bidders with lowest and second lowest marginal costs. The first term in the above expression is the profitability difference, and the second term is winner's *ex-post* operating profit given by:

$$\begin{aligned} \pi(c_w; T^p(c_w), N^p(c_w), L^p(c_w)) &= N^*(\gamma_w(c_w)) \cdot \tau^*(\gamma_w(c_w)) - C^c(c_w; K^*(\gamma_w(c_w))) \\ &= N^*(\gamma_w(c_w)) \cdot \tau^*(\gamma_w(c_w)) - C^c(\gamma_w(c_w); K^*(\gamma_w(c_w))) \\ &\quad + \frac{F(c_w)}{f(c_w)} \cdot K^*(\gamma_w(c_w)) \\ &= \frac{F(c_w)}{f(c_w)} \cdot K^*(\gamma_w(c_w)) - c_0 = \frac{F(c_w)}{f(c_w)} \cdot L(c_w) - c_0 \end{aligned} \tag{2.34}$$

where the last line follows the results presented in Section 2.2, where we show that the solution to **FB** leaves the firm with a loss of c_0. As $\frac{F(c_w)}{f(c_w)}$ increases with c_w, the distortion profit increases with the marginal cost. Plugging (2.34) into (2.33) yields the *ex-post* government payment

$$M_w^p(\boldsymbol{c}) = c_0 + \int_{c_w}^{c_l} L(x)dx - \frac{F(c_w)}{f(c_w)} \cdot L(c_w) \tag{2.35}$$

Thus, in order to make firm participation viable, from an *ex-post* standpoint, the government must compensate the firm for the fixed costs, c_0, as well as for the difference between the information rent and the distortion profit. Essentially, the government collects the distortion profit uses it to recover the information rent and the fixed cost. Since the profitability difference depends on the realizations of c_w and c_l, it is not possible to establish whether the subsidy is higher or lower than c_0. When the distortion profit is large enough, the total transfer could be negative. The limitation here is that the government cannot rely on the *ex-post* analysis in the process of designing the mechanism because the realization of $\boldsymbol{c}$ is unknown in advance. This serves as motivation for studying the optimal mechanism subject to the additional constraint of *ex-post* budget balance. Due to the Revelation Principle, the only adjustment needed is to impose an additional budget constraint on the set of incentive compatible mechanisms. The new problem is

$$\textbf{BBMD} \quad \max_{Q,T,L} \quad E_{\boldsymbol{c}}\left[\sum_{i=1}^{I} q_i(\boldsymbol{c}) \cdot \left(SW(c_i, T(c_i), L(c_i)) - \frac{F(c_i)}{f(c_i)} L(c_i)\right) - V_0\right]$$

subject to:

$$\begin{aligned}
\mathbb{P}_i(c_i) \cdot L(c_i) &\leq \mathbb{P}_i(c_i') \cdot L(c_i'), \quad c_i' \leq c_i && (2.36)\\
E_{\boldsymbol{c}_{-i}}[M_i(c_i, \boldsymbol{c}_{-i})] &= V_0 - \int_{c^{max}}^{c_i} \mathbb{P}_i(x) L(x) dx - \mathbb{P}_i(c_i)\pi(c_i; T(c_i), N(c_i), L(c_i) && (2.37)\\
\sum_{i=1}^{I} M_i(\boldsymbol{c}) &\leq 0 && (2.38)\\
V_0 = V_i(c^{max}) &\geq 0, && (2.39)\\
\forall c_i, c_i' &\in \mathcal{C}, \quad \forall i = 1, \ldots I
\end{aligned}$$

Constraint (2.37) combines (2.22) and (2.23) from **MD**, and provides an explicity definition of the government transfers. Constraint (2.38) restricts the *ex-post* transfers to being non-positive. The proposition below presents a solution to **BBMD** that relies on decomposing the problem into two parts: the allocation rule, and the toll-capacity schedule.

Proposition 3. *Optimal budget-balanced mechanism*

*Under a regular, common prior, the following direct mechanism optimizes **BBMD**:*

(1) *Q^B is given by* $q_i(\boldsymbol{c}) = \begin{cases} 1 & ; \quad c_i = \min\{c_1, c_2, \ldots, c_I\} \\ 0 & ; \quad otherwise \end{cases}$.

(2) *Transfer is given by,*

$$M_i^B(\boldsymbol{c}) = \begin{cases} -q_i(\boldsymbol{c}) \cdot [\pi(c_i; T^B(c_i), N^B(c_i), L^B(c_i)) + \frac{\int_{c^{max}}^{c_i} \mathbb{P}_i(x)L(x)dx}{\mathbb{P}_i(c_i)}] & ; \quad \mathbb{P}_i(c_i) \neq 0 \\ 0 & ; \quad otherwise \end{cases}$$

(3) *For each firm i, $T^B(c_i)$ and $L^B(c_i)$ solve the following Ramsey Problem* **RP-R***:*

$$\mathbf{RP-R} \qquad \max_{\tau,K} \qquad \int_0^N P(n)dn - N \cdot C(N,K) - C^c(\gamma_i(c_i); K)$$

subject to:

$$\begin{aligned} P(N) &= \tau + C(N,K) \\ \mathbb{P}_i(c_i) \cdot (N \cdot \tau - C^c(c_i; K)) &\geq -\int_{c^{max}}^{c_i} \mathbb{P}_i(x) L^B(x) dx \end{aligned}$$

Proof. We want to show that the mechanism (Q^B, L^B, T^B, M^B) as presented in the proposition is the optimal solution to problem **BBMD**. Following a similar logic to the one used in analyzing **MD-S**, we can verify that the mechanism is IC, and *ex-post* IR, i.e., that constraints (2.36), (2.37), and (2.39) are satisfied. To verify that the mechanism is budget balanced, we note that the *ex-post* total payment is given by

$$\sum_{i=1}^{I} M_i^B(\boldsymbol{c}) = -\pi(c_w; T^B(c_w), N^B(c_w), L^B(c_w)) - \frac{\int_{c^{max}}^{c_w} \mathbb{P}_w(x) L^B(x)dx}{\mathbb{P}_w(c_w)}$$
$$= \frac{-\mathbb{P}_w(c_w)\cdot \pi(c_w; T^B(c_w), N^B(c_w), L^B(c_w)) - \int_{c^{max}}^{c_w} \mathbb{P}_w(x) L^B(x)dx}{\mathbb{P}_w(c_w)} \leq 0$$

where the inequality follows from the second constraint in **RP-R** and w is the winner. Thus, the mechanism is a feasible solution to **BBMD**.

We proceed by contradiction to verify that the mechanism is optimal. Assuming there is another IC, *ex-post* IR, and *ex-post* budget balanced mechanism, (Q', T', L', M'), yielding a strictly higher *ex-ante* public welfare than (Q^B, T^B, L^B, M^B). Recall that the *ex-ante* public welfare is the expectation of the virtual surplus over the bidders. Thus, there exists at least one firm, j, with marginal cost, c_j, such that $R(c_j, T', L') > R(c_j, T^B, L^B)$. Noticing that the objective function in **RP-R** is the virtual surplus, it must be the case that the profitability constraint in **RP-R** is violated.[6] Thus,

$$\mathbb{P}_j(c_j)\cdot \pi(c_j; T'(c_j), N'(c_j), L'(c_j)) < -\int_{c^{max}}^{c_j} \mathbb{P}_j(x) L'(x)dx$$

Now, from the IC constraint.

$$E_{\boldsymbol{c}_{-j}}[M_j'(\boldsymbol{c})] = -\mathbb{P}_j(c_j)\cdot \pi(c_j; T'(c_j), N'(c_j), L'(c_j)) - \int_{c^{max}}^{c_j} \mathbb{P}_j(x) L'(x)dx > 0$$

[6]User equilibrium condition has to be satisfied in any feasible mechanism.

When the conditional expectation of the payment is positive, there must be a positive *ex-post* payment, for some $\hat{\boldsymbol{c}}_{-w}$. Due to the *ex-post* IR, losing firms don't pay, i.e., $M_i'(\boldsymbol{c}) \geq 0 \quad \forall i \neq w$. Thus, j must be the winning firm, and the total payment $\sum_{i=1}^{I} M_i'(\boldsymbol{c}) \geq M_j'(\boldsymbol{c}) > 0$, which contradicts to the *ex-post* budget balance constraint. $\square$

Recall that for **MD** we propose an *ex-post* payment rule, M^p, which is analogous to the Vickery Auction. The payment M^p depends not only on the winner's marginal cost but also on the second-lowest marginal cost. In contrast, M^B, is not contingent on other bidder's types. The reason is related to the fact that *ex-post* budget balance requires the total payment to be non-positive for any realization of $\boldsymbol{c}$. This is equivalent to restricting the payment that the winner receives. Given the General Pay-off Equivalence, the *interim* payment, i.e., the conditional expectation given the winner's type, is fixed. Thus, to restrict the winner's *ex-post* payment, we use the constraint in **RP-R** to set all of the *ex-post* payments to their expectation. The difficulty arises, however, from the computing of the conditional expectation, since it not only depends on the prior but also on the capacity rule $L^B(\cdot)$. In practice, **BBMD** can be solved recursively, i.e., alternating between optimal toll/capacity schedules and the constraint in **(RP-R)**. Further, we note that **BBMD** is not just restricted to zero-budget case. We can apply it to any fixed budget (by adjusting c_0).

2.4. Implementation

In the preceding section, we present a direct mechanism that optimizes public welfare, where firms reveal their private information, and where the government uses this information to award a franchise. However, in many situations it may be difficult or unappealing for firms to evaluate or disclose their production efficiency, e.g., they may want to keep their business practices private. The analysis in this section builds on the Taxation Principle presented in Guesnerie (1981); Rochet (1985), stating that direct mechanisms can be implemented by having firms/bidders select from bundles consisting of technical characteristics, e.g., toll and capacity levels, and government compensation. In particular, we consider three implementations that circumvent the need to disclose private marginal costs.

First, we analyze an Optimal Scoring Auction, where based on their marginal costs, firms/bidders select a virtual surplus level, which serves as a proxy for public welfare, and a compensation level. By constructing a scoring function that accounts for distortions, the Optimal Scoring Auction corresponds to an implementation the Optimal Direct Mechanism. We then analyze a Naive Scoring Auction where the scoring function corresponds to the public welfare, i.e., the government's objective. While intuitive, the implementation is sub-optimal because it does not account for distortions. Finally, we present a novel, Demand Pricing Mechanism where firms select bundles consisting of a demand level and an associated government compensation. The Demand Pricing Mechanism is motivated by the complexity and potential lack of transparency associated with evaluating scoring functions. Because the mechanism does partially account for distortions, we are able to show (analytically) that it performs at least as well as the Naive Scoring Auction.

2.4.1. Optimal Scoring Auction

Scoring auctions are analogous to First-Price or Second-Price Auctions. Rather than price, the criterion to evaluate bids consists of a function that captures tradeoffs along multiple dimensions. Scoring auctions are widely used in competitive bidding for public projects, with the aforementioned $A+B$ bidding being a representative example. Scoring auctions are categorized into first-score, second-score, and second-preferred auction. The first two are analogous to the first-price and second-price auction, where the winning firm, the bidder with the highest score, is required to set the technical parameters, e.g., toll, capacity, payment, to match the best or the second-best score, respectively. The second-preferred auction requires the winning firm to set the parameters to those appearing in the second-best bid. Che (1993) shows that Optimal Direct Mechanisms can be implemented as first or second score auctions with different compensation strategies. This, however, is not possible with second-preferred auctions. In this section, we present an implementation of the Optimal Direct Mechanism as a first-score auction relying on a scoring function structured as described below in Assumption 5.

Assumption 5. *Structure of scoring function*
The scoring function is assumed to have the form $a(\tau, K, M) = \alpha(\tau, K) - M$*, where the arguments correspond to the toll, capacity, and government payment.* $\alpha(\cdot)$ *is referred to as the quality scoring function, and is independent of the compensation.*

Each bidder trades off the potential profit with the probability of winning based on their score. For the first-score auction, each bidder solves

$$\begin{aligned} &\max_{\tau,K,M} \quad \{\pi(c_i;\tau,N(\tau,K),K)+M\}\cdot P\left(\text{winning with } a(\tau,K,M)\right) \Leftrightarrow \\ &\max_{\tau,K,M} \quad \{\pi(c_i;\tau,N(\tau,K),K)+M\}\cdot \left(F_a(a(\tau,K,M))\right)^{I-1} \end{aligned} \tag{2.40}$$

where F_a is the cumulative distribution of scores at equilibrium. To understand F_a, we consider the symmetric Bayesian Nash Equilibrium, at which each firm applies asymmetric strategy that maps its marginal cost to a score. Because production efficiencies are *iid*, and the equilibrium strategy is symmetric, the distribution of scores is also *iid*. The probability that a firm wins is the joint probability that its $I-1$ opponents have lower scores, i.e., higher unit costs, and is given by F_a^{I-1}.

Desirable specifications of the scoring function lead to (i) (equilibrium) scores that increase with production efficiency, i.e., $a(\tau, K, M)$ decrease with c_i; and (ii) induce toll and capacity levels that maximize the public welfare. Lemma 1 is an intermediate result that we use in the specification of scoring functions that satisfy the second condition.

Lemma 1. *Optimal toll and capacity levels in Scoring Auctions*
With a scoring function $a(\tau, K, M) = \alpha(\tau, K) - M$, firm i chooses toll and capacity levels, τ and K, that maximize $\alpha(\tau, K) + \pi(c_i; \tau, N(\tau, K), K)$.

Proof. Suppose firm i selects the prescribed τ, K, M that maximize

$$\{\pi(c_i; \tau, N(\tau, K), K) + M\} \cdot P\,(\text{winning with } a(\tau, K, M))$$

, but there is another set, τ', K', such that, $\alpha(\tau', K') + \pi(c_i; \tau', N(\tau', K'), K') > \alpha(\tau, K) + \pi(c_i; \tau, N(\tau, K), K)$. Letting $M' = \alpha(\tau', K') + M - \alpha(\tau, K)$, we have $a(\tau, K, M) = \alpha(\tau, K) - M = \alpha(\tau', K') - M' = a(\tau', K', M')$, i.e., the bid with τ', K', M' yields the same score and probability of winning as that of firm i's preferred bid. However, it also yields a total profit that is higher than the equilibrium bid as

$$\pi(c_i;\tau',N(\tau',K'),K')+M' = \pi(c_i;\tau',N(\tau',K'),K')+\alpha(\tau',K')+M-\alpha(\tau,K)$$
$$> \pi(c_i;\tau,N(\tau,K),K)+a(\tau,K,M)+M-a(\tau,K,M)$$
$$= \pi(c_i;\tau,N(\tau,K),K)+M$$

contradicting to the optimality of the bid with τ, K, M. $\square$

Lemma 1 provides a framework to design (quality) scoring functions, $\alpha(\cdot)$, that induce firms to bid (based on) toll and capacity levels that maximize public welfare, $T^p(c_i) = \tau^*(\gamma_i(c_i))$ and $L^p(c_i) = K^*(\gamma_i(c_i))$. Proposition 4 follows Che (1993) and relies on the observation that the toll and capacity levels that optimize the bidding strategy also maximize the virtual surplus function for a given marginal cost, c_i.

Proposition 4. *Optimal Scoring Function*

A first-score auction with the quality scoring function below results in the implementation of the Optimal Direct Mechanism.

$$\alpha^p(\tau,K) \quad = \quad CS(\tau,N(\tau,K),K) - \int_0^K \frac{F(c^p(k))}{f(c^p(k))}dk$$

where $c^p(\cdot)$ is the inverse of the public optimal capacity schedule, $L^p(c_i)$.[7] *Furthermore, adding the constraint $M \leq 0$ and adjusting the inverse schedule result in an implementation of the Budget-Balanced Public-Optimal Mechanism.*

Proof. See 2.9.

[7]The inverse function exists due to the strict monotonicity of the optimal capacity schedule.

Given the specification of $\alpha^p(\tau, K)$, $\alpha^p(\tau, K)+\pi(c_i;\tau, N(\tau, K), K) = SW(K, N(\tau, K))$ $-\int_0^K \frac{F(c(k))}{f(c(k))}dk$, which is similar to the virtual surplus function, $R(\cdot)$ – both contain the social welfare and a term of the inverse hazard rate. Further, we note that both functions have the same derivatives with respect to τ and K. Thus, given a firm's marginal cost, c_i, toll and capacity levels that maximize $R(\cdot)$ also maximize $\alpha^p(\cdot) + \pi(c_i;\cdot)$. Because $\frac{F(c_i)}{f(c_i)} > 0$, the firm's objective overestimates the cost of capacity, which is consistent with the solution to **MD-S**.

2.4.2. Naive Scoring Auction

Intuitively, considering a naive quality scoring function, where $\alpha^n(\tau, K) = CS(\tau, N(\tau, K),$ $K)$, seems appealing. That is, a rule where each firm's score is a function of the consumer surplus generated by its plan. The naive rule, where awards are exclusively based on the government's preference, is widely-used in procurement auctions. In contrast to $\alpha^p(\cdot)$, $\alpha^n(\cdot)$ ignores the distortion term. Again, following Lemma 1, each firm maximizes

$$\begin{aligned}\alpha^n(\tau, K) + \pi(c_i;\tau, N(\tau, K), K) &= CS(\tau, N(\tau, K), K) + \pi(c_i;\tau, N(\tau, K), K)\\ &= SW(c_i;\tau, N(\tau, K), K)\end{aligned}$$

From the analysis under symmetric information, it follows that non-distorted toll-capacity schedules yield higher *ex-post* social welfare than do the distorted schedules. Under asymmetric information, however, the non-distorted schedules yield losses in the expected public welfare. Following Che (1993):

Lemma 2. *Ex-ante public welfare of scoring auction*

Under first-score auction, the expected public welfare is given by,

$$E_w\left[SW^e(c_w;\tau_w,K_w)-\frac{F(c_w)}{f(c_w)}K_w\right]$$

where τ_w, K_w are the toll and capacity choice of the winner w.

Proof. The proof is similar to the proof of Proposition 1.

While under the optimal scoring rule each firm optimizes the virtual surplus it generates, under the naive rule each firm optimizes the SW it generates. Letting $R^*(x)$ denote the optimal SW generated by a firm with marginal cost x, i.e.,

$$R^*(x)=\max_{N,K}\int_0^N P(n)dn-C(N,K)-C^c(x;K)$$

Thus, for firm i, the respective virtual surplus from the naive and optimal scoring rules are $R^*(c_i)-\frac{F(c_i)}{f(c_i)}K^*(c_i)$, and $R^*(\gamma_i)$. The difference is given by,

$$\begin{aligned}\Delta_n &= R^*(\gamma_i)-R^*(c_i)+\frac{F(c_i)}{f(c_i)}K^*(c_i)=\int_{c_i}^{\gamma_i}R^{*\prime}(h)dh+\frac{F(c_i)}{f(c_i)}K^*(c_i)\\ &= \frac{F(c_i)}{f(c_i)}K^*(c_i)-\int_{c_i}^{\gamma_i}K^*(h)dh\end{aligned}$$

where the last line follows from the Envelope Theorem. Notice the loss is always non-negative because $K^*(\cdot)$ is a strictly decreasing function. Due to "no distortion at the bottom", there is no loss of virtual surplus when firm i has the lowest possible marginal cost. In general, the loss comes from the variation of capacity choices for different marginal cost. Thus, the loss is at minimal if the capacity schedule is not sensitive to changes in marginal cost, and the distortion difference, $\frac{F(c_i)}{f(c_i)}$, is small. The loss stems from the payment, M. Once the toll and capacity choice is given according to the quality scoring function, the bidder then chooses the optimal payment from (2.40). Different quality scoring functions have different equilibrium payments. The Naive Scoring Auction, in particular, results in an excessive M, which reduces the public welfare even though the road maximizes social welfare. On the other hand, the Optimal Scoring Rule results in a better balance between social welfare and (the losses from) the information rent.

2.5. A Demand Pricing Mechanism

In this section, we present a novel, demand-based mechanism, which addresses possible lack of transparency or fairness associated with the implementation of scoring auctions. The first concern arises from the complexity of the social welfare function and the distortion term. The issue of fairness arises because bidders can exploit the information they may have about other firms. The proposed mechanism is based on the observation that the *public welfare* can be written as follows:

$$
\begin{aligned}
PW &= CS - M = \int_0^N P(n)dn - N \cdot C(N,K) - N\tau - M \\
&= \int_0^N P(n)dn - N \cdot (C(N,K) + \tau) - M = \int_0^N P(n)dn - N \cdot P(N) - M
\end{aligned}
$$

where the last expression follows from the user equilibrium condition. In turn, this shows that, no matter how a firm sets the toll and capacity, the public welfare is determined by the government payment, M, and the demand, N.[8] Thus, the government's task is to

[8]With a suitably-specification of the demand function, the argument can be extended further to establish irrelevance of other road characteristics, such as quality or safety.

assess the tradeoffs between demand and direct payment, whereas the firm is tasked with operating efficiency. In particular, the *user equilibrium* condition reduces the firm's task to optimizing the capacity level to satisfy a given demand. This is seen by rewriting the firm's operating profit as

$$\pi = N\tau - C^c(K) = N \cdot (P(N) - C(N,K)) - C^c(K)$$

For a fixed demand level, the first-order condition is given by

$$\begin{aligned} \frac{\partial \pi}{\partial K} &= -N \cdot C_K - c_i = 0 \Rightarrow \\ & \mu^* : g'(\mu)\mu^2\big|_{\mu=\mu^*} = c_i \end{aligned} \tag{2.41}$$

which coincides with the result for **(FB)** in Equation (2.6). Thus, if bids consist of demand and payment combinations, firms are competing by adjusting the capacity level to the optimal v/c ratio given in Equation (2.41). For a given N, firm i's profit is

$$\pi^d(c_i; N) \equiv N \cdot (P(N) - g(\mu^*(c_i))) - c_0 - c_i \frac{N}{\mu^*(c_i)} \tag{2.42}$$

where $\mu^*(c_i)$ is the v/c ratio for the (undistorted) marginal cost c_i.

Inspired by the irrelevance of toll and capacity levels on the public welfare, we present a simple mechanism, in which a government publishes a compensation policy contingent on demand levels submitted by bidders, and denoted $Y : \mathbb{R}^+ \mapsto \mathbb{R}$. Firm selection is based on demand offers.

Understanding that the government's response is given by Y, we define firm i's bidding strategy as the mapping $N^d : \mathcal{C} \mapsto \mathbb{R}^+$. Then, for firm i with marginal cost c_i, its equilibrium strategy is to bid the demand at $N^d(c_i)$, and receive payment $Y(N^d(c_i))$ if it wins. Following the Revelation Principle, we first look into the equivalent direct mechanism, where each firm reports its type c_i, thereby committing to a demand and payment, respectively, following N^d and $M^d : \mathcal{C} \mapsto \mathbb{R}$. Here, M^d is a function of the bidder's own type, c_i, as opposed the set, $\boldsymbol{c}$. That is, the payment is assumed to be independent from other bidders' types as the demand pricing schedule, Y, is a function of the bidder's demand proposal only. Repeating our construction for the Direct Mechanism yields the following optimal demand pricing problem:

$$\textbf{(DP)} \quad \max_{Q,N,M^d} \quad E_{\boldsymbol{c}}\left[\sum_{i=1}^{I} q_i(\boldsymbol{c}) \cdot \big(CS(N(c_i)) - M^d(c_i)\big)\right]$$

subject to:

$$V_i(c_i) = \mathbb{P}(c_i) \cdot \big(\pi^d(c_i; N(c_i)) + M^d(c_i)\big) \tag{2.43}$$

$$\tilde{K}(c_i) = \frac{N(c_i)}{\mu^*(c_i)} \tag{2.44}$$

$$\frac{\partial V_i(c_i)}{\partial c_i} = S_i(c_i) = -\mathbb{P}(c_i) \cdot \tilde{K}(c_i) \tag{2.45}$$

$$S_i(c_i') \leq S_i(c_i), \quad c_i' \leq c_i \tag{2.46}$$

$$V_i(c^{max}) \geq 0 \tag{2.47}$$

$$\forall c_i, c_i' \in \mathcal{C}, \quad \forall i = 1, \ldots I$$

(DP) exhibits 4 important differences compared to the original toll-capacity screening problem, **(MD)**:

(1) Firms select the capacity schedule, $\tilde{K}$, through the non-distorted v/c ratio.

(2) The toll schedule is removed because it is determined by the capacity and demand levels.

(3) The payment in the indirect utility expression, Equation (2.43), appears inside the parenthesis, which means that the payment rule is restricted to the winning firm.

(4) Critically, we note that the monotonicity condition, Equation (2.46), depends on each firm's optimal v/c ratio, $\mu^*(c_i)$.

Assumption 4 ensures the monotonicity of the capacity schedule, and therefore that the monotonicity constraint in **(MD)** is satisfied. In **(DP)**, however, the Regularity Condition is not sufficient to ensure the monotonicity of $\tilde{K}$, and thereby the monotonicity of S_i.[9] Below, we first derive the optimal demand schedule, i.e., optimal bidding strategies. We then present a condition that ensures the monotonicity of the solution.

As in our solution to **(MD)**, we solve for the payment rule in constraints (2.43), and substitute into the objective function. This allows us to consider firm selection and virtual surplus maximization separately. Firm i's virtual surplus is given by

$$R^d(c_i) \;\; = \;\; \int_0^N P(n)dn - N \cdot g(\mu^*(c_i)) - C^c(\gamma_i; K^d(c_i))$$

[9]When there is only one bidder, i.e., $I = 1$ and $\mathbb{P} = 1$, S_i always equals to $\tilde{K}$. Thus, the monotonicity of $\tilde{K}$ is a necessary condition.

The optimal demand level, $N^d(c_i)$, is obtained by evaluating the first-order condition.

$$N^d(c_i) : P(N)|_{N=N^d(c_i)} = \frac{\gamma_i}{\mu^*(c_i)} + g(\mu^*(c_i)) \tag{2.48}$$

where we observe that the distortion on choices of N does not affect $\mu^*(c_i)$. The expression for the distorted profit at the optimal choice of demand, $N^d(c_i)$ is

$$\begin{aligned} \pi^d(c_i; N^d(c_i)) &= N^d(c_i) \cdot (P(N^d(c_i)) - g(\mu^*(c_i))) - C^c(c_i; \frac{N^d(c_i)}{\mu^*(c_i)}) \\ &= \frac{N^d(c_i)}{\mu^*(c_i)}(\gamma_i(c_i) - c_i) - c_0 \\ &= K^d(c_i) \cdot \frac{F(c_i)}{f(c_i)} - c_0 \end{aligned}$$

The first term in the final expression is the product of the adjusted capacity and the distorted marginal cost. This is similar to the result for the **(MD)** problem.

From (2.48), we observe that a government always prefers the firm with the lowest $\gamma_i(c_i)$ because it yields the highest virtual surplus, even when multiple firms serve the same demand level. Unfortunately, even with the Regularity Assumption, optimal demand schedules, satisfying Condition (2.48), are not always monotonic in c_i. To see this, we write N^d using the demand function, $P^{-1}(\cdot)$, i.e., $N^d = P^{-1}\left(g(\mu^*) + \frac{\gamma_i}{\mu^*}\right)$, and take the derivative with respect to c_i:

$$\begin{aligned} \frac{\partial N^d}{\partial c_i} &= \frac{1}{P'} \cdot \left(g'\mu' + \frac{\gamma_i'\mu^* - \gamma_i\mu'}{(\mu^*)^2}\right) = \frac{1}{P' \cdot (\mu^*)^2} \cdot \left(g'\mu'(\mu^*)^2 + \gamma_i'\mu^* - \gamma_i\mu'\right) \\ &= \frac{1}{P' \cdot (\mu^*)^2} \cdot \left(\gamma_i'\mu^* - \frac{F(c_i)}{f(c_i)}\mu'\right) \end{aligned} \tag{2.49}$$

where the second line follows from (2.41). The sign of the derivative, which depends on the prior distribution and on the congestion function, is inconclusive. Thus, to arrive at a regularity condition that ensures monotonicity of the demand, we assume that the condition presented in Lemma 3 below holds for all firms.

Lemma 3. *Regularity Condition for Demand Pricing Mechanism*
For any congestion function, $g(\cdot)$, N^d is decreasing with respect to c_i if and only if

$$2c_i\gamma_i'(c_i) \;\; > \;\; \frac{F(c_i)}{f(c_i)}, \quad \forall c_i \in \mathcal{C}, i = 1, \ldots, I$$

Proof. See 2.9.

We note that the condition specified in Lemma 3 is not necessary for some congestion functions. For example, a uniform distribution from c^{min} to c^{max}, where $c^{min} \geq 0$, the left-hand-side in the condition equals $4 \cdot c_i$, while the right-hand-side equals $c_i - c^{min}$. Thus, any uniform distribution with non-negative support ensures the monotonicity of N^d. We also note that the condition in Lemma 3 is is stronger than Assumption 4. The latter requires γ_i' to be non-negative, whereas the former restricts γ_i' to be larger than a positive number. We do observe that only prior distributions with unrealistically high kurtosis violate the stronger condition. The following proposition shows that the result ensures that the monotonicity constraint in **(DP)** holds.

Proposition 5. *N^d decreasing with respect to c_i ensures that the monotonicity condition in* **(DP)**, *Equation* (2.46), *is satisfied.*

Proof. Assumption 4 implies the probability of winning, $\mathbb{P}(c_i) \geq 0$, decreases with c_i. The assumption that N^d is decreasing with respect to c_i means that $K^d(c_i) = \frac{N^d(c_i)}{\mu^*(c_i)}$ is also decreasing with c_i because $\mu^*(c_i) > 0$ is increasing with c_i. Thus, $S_i(c_i) = -\mathbb{P}(c_i) \cdot K^d(c_i)$ is non-decreasing in c_i. $\square$

The optimal payment rule is obtained by substituting the Envelope Theorem conditions, Equation set (2.45), into the Indirect Utility constraints, Equations (2.43), as shown below:

$$\begin{aligned} M^d(c_i) &= \frac{V_i(c_i)}{\mathbb{P}_i(c_i)} - \pi^d(c_i; N^d(c_i)) = \frac{\int_{c_i}^{c^{max}} \mathbb{P}(x)\tilde{K}(x)dx}{\mathbb{P}(c_i)} - \pi^d(c_i; N^d(c_i)) \\ &= \frac{\int_{c_i}^{c^{max}} \mathbb{P}(x)\tilde{K}(x)dx}{\mathbb{P}(c_i)} - \frac{N^d(c_i)}{\mu^*(c_i)} \frac{F(c_i)}{f(c_i)} + c_0 \end{aligned}$$

where the last expression follows from Equations (2.48) and (2.42).

In the context of **(DP)**, the Taxation Principle ensures that the Revealed Mechanism can be implemented through a non-linear pricing model, as long as all bids at a given demand level commit to the same payment (without regard to the underlying c_i) (Rochet, 1985). If N^d is strictly decreasing in c_i, then each demand level corresponds to a unique marginal cost, which satisfies the condition. If, however, N^d is not strictly decreasing, there could be different payments for the same demand level. In particular, firms with lower marginal costs would request higher payments to match less efficient firms, which would preclude implementation of the Optimal Demand Pricing. This, in turn, explains the importance of Lemma 3 to enable governments to screen firms based on their demand proposals. The payment, M^d, as a function of the demand, N^d is written as

$$Y(N) = M^d(c^d(N^d(c_i))) = \frac{\int_{c^d(N^d(c_i))}^{c^{max}} \mathbb{P}(x)K^d(x)dx}{\mathbb{P}(c^d(N^d(c_i)))} - K^d(c_i) \cdot \frac{F(c^d(N^d(c_i)))}{f(c^d(N^d(c_i)))} + c_0 \quad (2.50)$$

where $c^d(\cdot)$ is the inverse of the monotonically decreasing demand function, N^d. The first, second, and third terms, respectively, represent the information rent, the excess distortion profit, and the fixed cost. Thus, the payment does not cover the variable construction cost but rather offsets the difference between the distortion profit and information rent. Figure 2.1 presents an example of pricing schedules. The parameters used in the example are presented in Section 2.7.

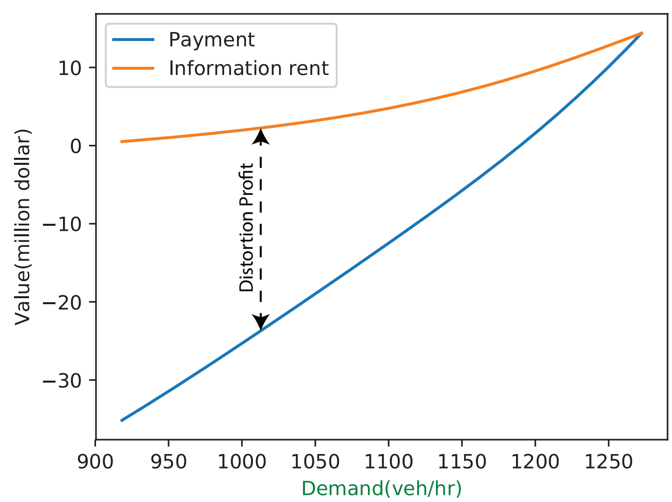


Figure 2.1. Optimal payment schedule

The information rent is always positive and increasing with demand. Since both the distortion and demand decrease with the marginal cost, the distortion profit decreases with the demand. As a result, the payment, which offsets the distortion profit, increases with the demand. One should notice that it is not a coincidence that the payment curve in the figure transitions from negative to positive. The reason is that, for the lowest demand and therefore the highest marginal cost, the distortion profit is at its maximum, but the information rent is zero. The payment rule requires the firm to pay back all the excessive profit. For the highest demand, and equivalently lowest marginal cost, the distortion profit is zero, but the information rent is at its maximum. In this case, the government compensates the firm at an amount equal to its information rent. Because the demand schedule N^d is monotonically decreasing, the most efficient firm submits the highest demand proposal.

In summary, the optimal Demand-Pricing Mechanism is given as

Proposition 6. *Optimal Demand-Pricing Mechanism*
The government implements the optimal Demand-Pricing Mechanism, **(DP)***, by offering* Y*, as specified from Equation* (2.50)*, and selecting the proposal with the highest demand.*

Proof. See 2.9.

Among the attributes of the Demand Pricing Mechanism, we note:

(1) The mechanism is straightforward and simple to the bidders because the only information they need to send is the demand proposal, instead of a detailed construction plan.

(2) The payment schedule is homogeneous across bidders. Firms bidding the same demand can expect to receive the same payment. Thus, it is impossible for the insider to intentionally change the payment to a specific bidder.

(3) The determination of winner is simple and transparent, thereby avoiding complications evaluating complicated functions in scoring auctions, for example.

There is, however, a caveat in the Demand Pricing Mechanism that stems from the stronger regularity condition. Specifically, we note demand is distorted, but the v/c ratios are not, and as a result, the mechanism does not optimize the virtual surplus. To see this, we define $R^{\diamond}(x, y)$ as the optimal surplus, SW, associated with an optimal v/c ratio for a marginal cost of x, and for a virtual marginal cost y, i.e.,

$$R^{\diamond}(x,\gamma) \;\equiv\; \max_{N} \int_0^N P(n)dn - N \cdot g(\mu^*(x)) - C^c\left(\gamma, \frac{N}{\mu^*(x)}\right)$$

In **(MD)**, firm i maximizes the virtual surplus that it generates, $R^{\diamond}(\gamma_i, \gamma_i)$, whereas in **(DP)**, each firm maximizes $R^{\diamond}(c_i, \gamma_i)$. From the Envelope Theorem

$$\begin{aligned}\frac{\partial R^{\diamond}(x,\gamma_i)}{\partial x} &= \left(\frac{\gamma_i}{\mu^{*2}(x)} - g'(\mu^*(x))\right) \cdot N^{\diamond}(x,\gamma_i) \cdot \mu^{*\prime}(x) \\ &= \left(\gamma_i - g'(\mu^*(x)) \cdot \mu^{*2}(x)\right) \cdot \frac{N^{\diamond}(x,\gamma_i) \cdot \mu^{*\prime}(x)}{\left(\mu^*(x)\right)^2}\end{aligned}$$

where $N^{\diamond}(x, \gamma)$ is the argument that maximizes $R^{\diamond}(x, \gamma)$. From (2.6), we see that γ_i is the unique point where $\left.\frac{\partial \bar{R}(x,\gamma_i)}{\partial x}\right|_{x=\gamma_i} = 0$. Because both $\mu^*(x)$ and $g'(\mu^*(x))$ are increasing, $\frac{\partial R^{\diamond}(x,\gamma_i)}{\partial x} > 0$ for $x < \gamma_i$, and $\frac{\partial R^{\diamond}(x,\gamma_i)}{\partial x} < 0$ for $x > \gamma_i$. Thus,

$$\Delta_d = R^{\diamond}(\gamma_i,\gamma_i) - R^{\diamond}(c_i,\gamma_i) = \int_{c_i}^{\gamma_i} \left(\gamma_i - g'(\mu^*(h)) \cdot \mu^{*2}(h)\right) \cdot \frac{N^{\diamond}(h,\gamma_i)\mu^{*\prime}(h)}{\mu^*(h)^2} dh > 0 \quad (2.51)$$

which shows that there is always a loss from not distorting the v/c ratio. However, in most cases, the loss is small. A bound for the loss is derived from (2.51) and (2.6)

$$
\begin{aligned}
\Delta_d &= \int_{c_i}^{\gamma_i} (\gamma_i - h) \cdot \frac{N^\diamond(h,\gamma_i)\cdot \mu^{*\prime}(h)}{\mu^*(h)^2} dh < (\gamma_i - c_i) \cdot \int_{c_i}^{\gamma_i} \frac{N^\diamond(h,\gamma_i)\cdot \mu^{*\prime}(h)}{\mu^*(h)^2} dh \\
&= \frac{F(c_i)}{f(c_i)} \cdot \int_{\mu^*(c_i)}^{\mu^*(\gamma_i)} \frac{N^\diamond\left(\mu^{*-1}(m),\gamma_i\right)}{m^2} dm < \frac{F(c_i)}{f(c_i)} \cdot N^\diamond(c_i,\gamma_i) \left(\frac{1}{\mu^*(c_i)} - \frac{1}{\mu^*(\gamma_i)}\right)
\end{aligned}
$$

That is, the closer the inverse of two v/c ratios are, the tighter bound on the virtual surplus loss. Hence, Δ_d is expected to be small when the distortion $\frac{F(c_i)}{f(c_i)}$ is small or the optimal v/c ratios are large.

To conclude this section, we observe that the virtual surplus loss from the implementation of the Demand Pricing Mechanism is always less than the loss stemming from the implementation of the commonly-used Naive Scoring Rule, i.e., $\Delta_n - \Delta_d \geq 0$. The intuition is that the partial distortion, i.e., the demand distortion in the Demand Pricing Mechanism, is better than the "no distortion" situation with the Naive Scoring Rule. Distorting demand is not sufficient to fully optimize the government's *ex-ante* benefits, but the distortion is second-best under the constraint that bidders self-select the capacity levels. The numerical examples, in Section 2.7, show that Δ_d is close to 0 and significantly lower than the loss from the naive scoring rule, Δ_n.

2.6. Discussion

In this section, we discuss the conditions for our proposed mechanisms as well as their practical advantages and disadvantages. In particular, we discuss our common prior information structure, summarize the number-of-bidder-independent property, and evaluate the *ex-post* Pareto inefficiency from the distortion.

2.6.1. Bayesian vs prior-free

The Bayesian Assumption, i.e., both the government and private firm know the prior distribution, plays a fundamental role in this paper. Despite its analytical advantages, the assumption may be too strong in practice because it depends on the details of the *common knowledge*, the distribution of types, which the principal and the agent may not know (Loertscher and Marx, 2015). Prior-free mechanisms deal with such an issue by imposing a stronger version of incentive compatibility, the Dominant Strategy Incentive Compatibility (DSIC). That is, firms find telling the truth always optimal given any combination of types for the other firms, as opposed to a distribution of types. The most famous example of the prior-free mechanism is the VCG mechanism, but the objective of VCG mechanism is the *ex-post* measure, i.e., the optimality of decisions given a profile of bidders, which differs from the objective considered herein, i.e., optimality of expected surplus. In general, the prior-free mechanism is not *ex-ante* optimal as the IC constraints are stronger. Based on the prior distribution, one may improve the expected surplus by relaxing the IC in some cases while tightening the IC in other cases.

One should also notice that there is no absolute optimal mechanism since the mechanism is always set *ex-ante*. One mechanism may yield a high surplus to one profile of agents but yield a low surplus to another profile of agents. The prior-free mechanism solves the optimization problem given the profile, but it is not able to optimize the *ex-ante* surplus across different profiles. One exception relates to the efficient mechanism, whose goal is simply optimizing the allocation. However, it is not able to guarantee the auctioneer's pay-off is maximized. Essentially, the principal needs to tradeoff among the profiles of agents. As explained in McAfee and Vincent (1992), sometimes the auctioneer even sets a reservation price to block some types of agents to maximize the *ex-ante* surplus. To address the unclear prior distribution issue, the government may find out the approximate optimal mechanism in worst-case over the possible distributions (Hartline, 2013).

In practice, bidder collusion (in detriment of the government) is another possible concern. Under the prior-free mechanism, bidders may take huge advantages over the government by colluding. For example, under the minimal cost procurement auction, bidders may work together to bid higher than their true values, and the winner realizes an additional government payment. While, in our proposed Bayesian framework, especially the Demand Pricing Mechanism, the subsidy can be independent of other bidders' type. The payment under the Bayesian framework depends on the government's estimation of types rather than the actual types. Essentially, the Bayesian optimal mechanism helps the government utilize the limited information they have about the bidders.

2.6.2. Number of bidders

So far, we have not discussed the number of bidders for highway franchising. Intuitively, an increasing number of bidders intensifies competition, thereby limiting the bidder's surplus. One may guess that the number of bidders would change toll-capacity decisions, as it does in most oligopolistic markets. However, the proposed mechanism has a different property.

Property 1. *The toll-capacity schedule for the public optimal mechanism is independent of the number of bidders. Moreover, the demand schedule for the optimal demand pricing is independent of the number of bidders.*

Notice that the above property even applies when we consider the monopoly case, i.e., $I = 1$. In fact, the optimal T and L for public optimal mechanism solves Shi et al. (2016)'s single bidder contracting problem. The independence is straightforward, as we can observe that the optimal schedule is essentially the first-best toll-capacity schedule with the distorted marginal cost, which only depends on the prior. This property demonstrates a potential advantage: bidders' winning strategy and the optimal proposal would not change with the number of bidders. Thus, there is no room for fraudulent schemes like bid-rigging.

However, the number of bidders does affect the mechanisms in various ways. First, the number of bidders may[10] affect the *ex-ante* payment made by the government. The public optimal mechanism requires the government to make the payment conditional on the winner's marginal cost as well as other bidder's marginal cost. An analogous example is that, under second-price auction, the winner is less likely to receive a large surplus against a large number of bidders, as the second-highest bidder is more likely to have a private value close to the winner. From Equation (2.33), under the public optimal mechanism, the winner x receives, $\int_{c_x}^{c_l} L(h)dh - \pi$, where l is the second-best firm[11]. We note the payment increases with the production efficiency difference between the winner and the next-best firm. Namely, the competition is not reflected by the toll-capacity choice but the government's payment. The exception is that the number of bidders affects the budget-balance constraint, which further affects the toll and capacity schedule in **(BBMD)**.

[10]However, we can also construct a payment rule that is independent of the number of bidders.

[11]Notice, when $I = 1$, there is no next-best user. Thus, c_l is the production efficiency of the government's outside option.

Second, under the *iid* assumption, the increase in the number of bidders also changes the distribution of the winner's marginal cost. That is, given a large number of bidders, the winner is more likely to have a lower marginal cost. In a limiting case of infinite bidders, the winner's marginal cost approaches the lower bound. Recall the "no distortion at the bottom" conclusion. For the low marginal cost, the public optimal toll-capacity is, barely distorted. Thus, the number of bidders does not change the distortion for a given bidder, but to increase the chance of having a better bidder in the auction. Ultimately, the government benefits from the increase of the number of bidders, not through the response of the firms, but the self-adjustment of the mechanism.

2.6.3. Analysis on distortions

In the previous sections, we derive the distortion on the (public) optimal mechanism. In order to limit the information rent, the government has to make an *ex-post* sub-optimal decision, i.e., distorting the first-best decisions. In the context of the highway franchise, there are two decisions to be made: one is the concessionaire selection decision, and the other in the planning decision, i.e., the toll-capacity choice. We examine these two types of decisions separately.

Recall our discussions on the regularity condition, the distorted marginal cost, in general, does not always increase with the marginal cost. Thus, the most efficient firm may not provide the highest virtual surplus, which the government intends to maximize. The interpretation is that efficient firms know their superiority over less efficient firms, so they increase the mark-up to the government. If the increase in the mark-up is greater than the production advancement, the government is better to award the less-efficient firm for a less expensive payment. This phenomenon only occurs in extreme situations, e.g., when the prior has unrealistically high kurtosis. In this paper, we exclude the possibility of selecting a sub-optimal concessionaire by imposing Assumption 4. Thus, in regular cases, information asymmetry is not sufficient to distort the concessionaire selection.

On the other hand, the public optimal mechanism, however, always[12] takes the *ex-post* sub-optimal toll-capacity choice. In particular, the toll is over-priced, the capacity is under-built, and the road ends up with a higher v/c ratio than optimum, $\mu^*(\gamma_i) > \mu^*(c_i)$, because $\gamma_i > c_i$. One may argue that such distortions are Pareto inefficient, since Tan et al. (2010) show that Pareto-efficient contract requires the v/c ratio to be optimal, i.e., at $\mu^*(c_i)$. Such a claim is confusing if we consider the trivial case of contracting with a given concessionaire, i.e., $I = 1$. When there are two parties, the government and the given concessionaire, contracting on the toll-capacity choices, the outcome is supposed to be Pareto efficient.

The seeming controversy is explained by considering the process of revealing the marginal cost. After the revelation of true marginal costs, the government is indeed willing to set up the contract, which yields the optimal v/c ratio, $\mu(c_i)$, higher consumer surplus, and link profit. That is, from the *ex-post* perspective, the distorted toll-capacity choice is not Pareto-efficient. However, before the revelation of true marginal costs, the government has to play the sequential game shown in Figure 2.2.

Essentially, the game is non-cooperative and results a Pareto-inefficient Nash Equilibrium (CS_2, π_2), as $CS_2 < CS_4$ and $\pi_2 < \pi_4$. Therefore, the toll-capacity distortion is *ex-ante* Pareto efficient, i.e., there is no better prior-bidding strategy making the expected

[12]We ignore the trivial case of c_i hitting the lower bound here.

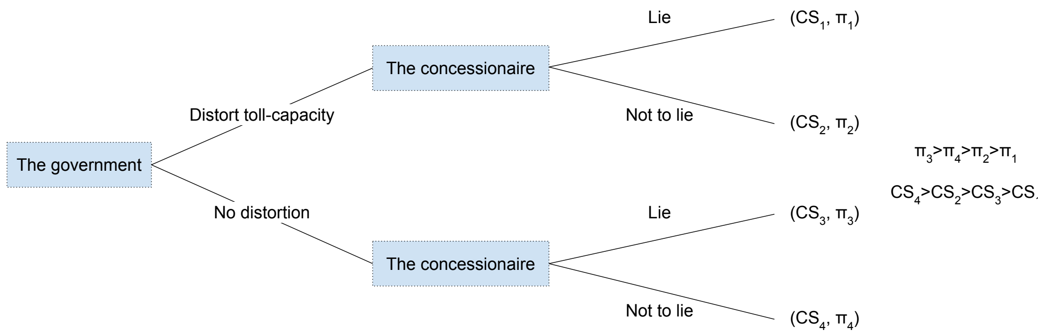


Figure 2.2. Sequential game of reporting marginal cost for toll-capacity choice

consumer surplus and profit higher at the same time. One may argue that a post-auction renegotiation may bring Pareto improvement on distorted toll-capacity. However, the renegotiation changes the bidder's pay-off and violates the incentive compatibility. If the bidder knows there is a renegotiation after bidding, it may deviate its reporting strategy and decide to lie.

However, the demand distortion, used in Demand Pricing Mechanism, is *ex-post* Pareto-efficient. This conclusion can be explained in two-fold. First, the demand distortion retains the optimal v/c ratio, which follows the Pareto efficiency condition in Tan et al. (2010). Second, under demand pricing, the bidder is allowed to choose toll-capacity level freely. Thus, the link profit is maximized, given the demand level. There is no deviation improving the profit unless the demand level is decreased. As the decrease of demand always reduces consumer surplus, there is no deviation improving profit and consumer surplus at the same time, even after the government selects the concessionaire. In conclusion, the demand distortion results in an *ex-post* Pareto efficient contract but an *ex-ante* sub-optimum for the government. In contrast, the toll-capacity distortion results in an *ex-post* Pareto inefficient contract but the *ex-ante* optimum for the government.

2.7. Numerical examples

Having discussed the analytic properties of mechanisms, we present numerical examples to further compare them. To generate heterogeneous firms, we randomly sample firms' production efficiency based on a truncated normal prior. In particular, the support of $f(\cdot)$ is between \$20,000 per veh/hr to \$40,000 per veh/hr, the fixed cost is set to \$2 million. We include the Optimal Mechanism, the Demand Pricing Mechanism, the predetermined toll-capacity mechanism, and the undistorted mechanism, i.e., the naive scoring auction. The congestion function is assumed to take the form of BPR function, specifically, that $C(N, K) = 0.5 * 20(1 + (\frac{N}{K})^4)$. The inverse demand is assumed to be linear as $P(N) = 80 - 0.05N$.

2.7.1. Virtual surplus and actual surplus

In the preceding sections, we define the virtual surplus as a proxy of the *ex-ante* performance. The Optimal Mechanism always yields the highest *ex-ante* public welfare as it optimizes the virtual surplus for any type. It is worth noting, however, that such comparisons are inconclusive when one mechanism yields a higher virtual surplus in some cases but a lower virtual surplus in other cases. Figure 2.3 shows the comparisons of virtual surplus among the proposed mechanisms.

We see that for (almost) all values of the marginal cost, the virtual surplus is ranked from high to low as the Optimal Mechanism, Demand Pricing, no distortion (the Naive

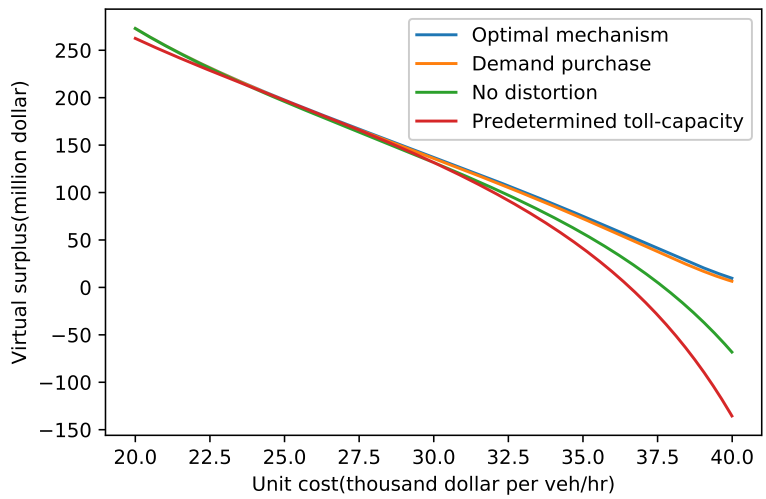
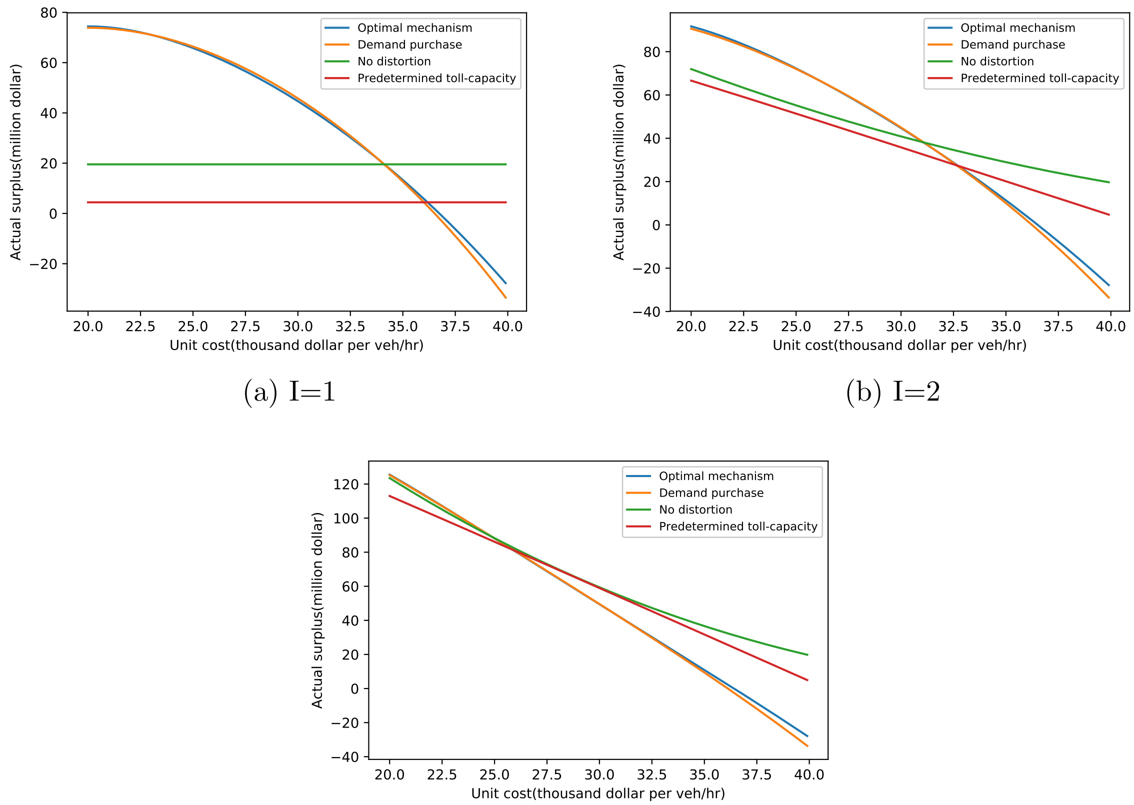


Figure 2.3. Comparison of virtual surplus

Scoring Rule), and predetermined toll-capacity bidding. Also, the curve of demand pricing overlaps the curve of optimal mechanism almost everywhere, which implies the welfare loss from Demand Pricing is small. But the virtual surplus loss from the no distortion and predetermined cases are significant. These observations are consistent with Lemma 4. However, we need to clarify that the only purpose for comparing virtual surplus is to evaluate the *ex-ante* performance. There is no point in comparing virtual surplus for a specific marginal cost. For example, the overlaps of all four mechanisms at $25k per veh/hr do not imply that the public welfare, given the winner's marginal cost close to $25k per veh/hr, is the same among them. To demonstrate the comparison case by case, we need to calculate the actual surplus, i.e., the actual public welfare given a certain type of winner. From Section 2.6, the payment is affected by the number of bidders, so does the actual surplus. Thus, we demonstrate actual surplus over the different number of bidders, shown in the Figure 2.4.



(a) I=1

(b) I=2

(c) I=8

Figure 2.4. Actual surplus varying number of bidders

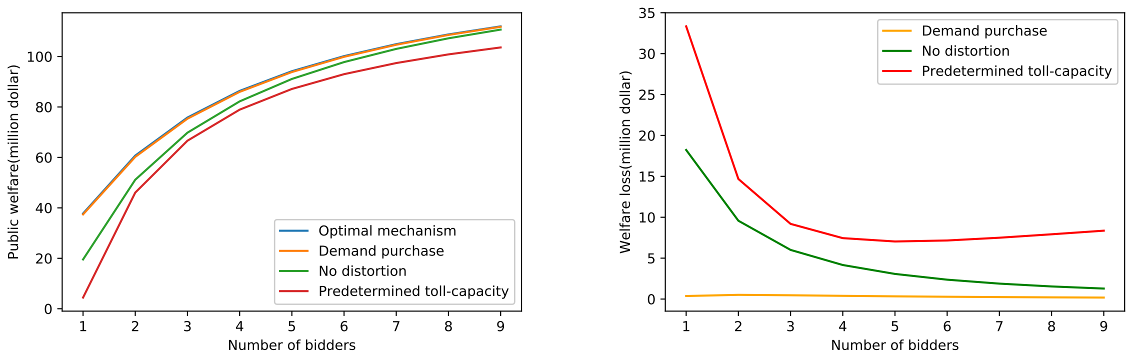
For the single bidder case, the undistorted mechanism and predetermined mechanism, yield a horizontal curve. This is related to the bargaining between the government and the only bidder. The outside option for the government is to build the highway with a constant but higher marginal cost. As a result, the unique bidder optimizes its benefit with the constraint that the government receives at least the surplus that it would receive from the outside option. Hence, the more efficient firm asks for more payment, and the less efficient firm asks for less payment. Eventually, the government's actual surplus is constant over types. Since the predetermined mechanism is not able to optimize with the type, it lies below that of "no distortion". The actual surplus curve of the optimal mechanism and the demand pricing is very close and vary with the marginal cost. Both mechanisms yield higher actual surplus when the marginal cost is low but lower actual surplus when the marginal cost is high than that of undistorted mechanism and predetermined mechanism. This is a demonstration that no mechanism is absolutely optimal. Essentially, the actual surplus at a very high marginal cost is sacrificed for the actual surplus at low marginal cost, because it is less likely for the winner to have very high marginal cost, e.g., higher than \$33k per veh/hr. Such observations are more apparent as we increase the number of bidders: The threshold, at which the undistorted mechanism starts to outperform the optimal mechanism, is decreasing with the number of bidders. That is, the optimal mechanism attempts to sacrifice the actual surplus at a wider range of types as the number of bidders increases. Because when the number of bidders increases, the winner is more selective and more likely to have a low marginal cost.

2.7.2. Ex-ante performance

Having compared the mechanisms case by case, we move to the ultimate measures of the performance: the *ex-ante* public welfare.

Figure 2.5a shows the comparison of *ex-ante* public welfare. Similar to the preceding comparisons, the optimal mechanism and demand pricing are close, and they outperform undistorted mechanisms for any number of bidders. "No distortion" mechanism gradually converges to the distorted mechanisms as the number of bidders rises.

The convergence is due to the "No distortion at the bottom" argument. When the number of bidders increases, the winner is more likely to have a marginal cost close to



(a) Ex-ante public welfare

(b) Ex-ante welfare loss

Figure 2.5. Ex-ante public welfare and loss

the lower bound. Thus, the distortion effects are vanishing. However, the predetermined mechanism is not able to converge due to its inability to adjust toll and capacity. Figure 2.5b shows a better demonstration of the differences in public welfare by benchmarking the optimal mechanism. In summary, we verify that the optimal mechanism always has the best *ex-ante* performance and that demand pricing performs very close to the optimal mechanism. "No distortion" does not perform well when the number of bidders is small, i.e., $I < 4$, but it improves very fast as the number of bidders increases. Predetermined mechanism always yields the lowest *ex-ante* public welfare, and its welfare loss stops decreasing very early, at $I = 4$.

2.8. Conclusion

We study the problem of designing mechanisms for highway franchising to address the issue of cost information asymmetry. Specifically, we analyze two issues caused by such information asymmetry: concessionaire selection, and the toll-capacity setting. Instead of ignoring the interactions between these two issues, we build a toll-capacity screening framework and combine concessionaire selection and toll-capacity optimization. It turns out the expected public welfare is governed by the virtual surplus and we make the use of virtual surplus to decouple the interactions.

Given the optimal benchmark, we provide practical implementations to fix information asymmetry by distorting toll, capacity or demand schedule. In particular, the public optimal scoring system effectively limits the private firms' information rent and significantly improve the *ex-ante* public welfare. The demand pricing yields very close performance as well as a great improvement on the transparency and convenience. Our results show that the distorted mechanism yields great practical value when

(1) The number of bidders is low.

(2) The prior distribution of productivity is common knowledge.

(3) The support of the prior has a wide range, i.e., there is a wide range of possible marginal costs.

This work can be extended in two directions. First, we may consider the dynamics of the contracts. In practice, there are always demand risks for the bidder. Thus, the government may allow the bidders to flexibly adjust the toll and capacity decisions. It is interesting to investigate how to incorporate the concerns of demand risk into the dynamic decision-making process while limiting the information rent. Second, it is interesting to consider the franchise of a road network instead of a single road. The interdependencies among links affect the bidder's decision. For example, if there are two parallel links in the network, wining two links together gives additional synergy benefit to the concessionaire. Essentially, such consideration is related to the combinatorial auction.

Further, we note that the analysis may apply to other infrastructure systems, such as the auction for power transmission or telecommunications.

2.9. Proofs and Simplifications

Proposition 1 *Outcomes of predetermined toll-capacity bidding*

(1) *Bidder i's interim private surplus, i.e., the sum of profit and subsidy, is*

$$\int_{\pi^{min}}^{\upsilon_i} H^{I-1}(y)dy \quad = \quad K^*(\bar{c}) \int_{c_i}^{c^{max}} F^{I-1}(x)dx$$

where $H(\cdot)$ represents the cumulative distribution function corresponding to h.

(2) *The ex-ante public welfare is*

$$E_x \left[CS(\bar{c}) + \pi(x;\bar{c}) - \frac{F(x)}{f(x)} \cdot K^*(\bar{c}) \right]$$

where E_x denotes the expectation over the winning firm's marginal cost, x.

Proof:

(1) Following Riley and Samuelson (1981), firm i's reservation value, $\upsilon_i \equiv \pi(c_i; \tau(\bar{c}), N(\bar{c}), K(\bar{c}))$. Under equilibrium bidding strategies for first-price auctions, a firm with reservation value υ_i receives an (interim) expected payment, $M(\upsilon_i)$, given by

$$M(\upsilon_i) \quad = \quad \int_{\pi^{min}}^{\upsilon_i} H^{I-1}(y)dy - \upsilon_i \cdot H^{I-1}(\upsilon_i)$$

Notice $M(v_i)$ is not the bidding price, but the expected payment in advance of the auction, i.e., before the private bidder knows if it is the winner. Thus, the *interim* private surplus is given by

$$\begin{aligned}
M_i(v_i) + v_i \cdot P(\text{winning with } v_i) &= \int_{\pi^{min}}^{v_i} H^{I-1}(y)dy - v_i \cdot H^{I-1}(v_i) + v_i \cdot H^{I-1}(v_i) \\
&= \int_{\pi^{min}}^{v_i} H^{I-1}(y)dy \\
&= K^*(\bar{c}) \int_{c_i}^{c^{max}} F^{I-1}(x)dx
\end{aligned}$$

The last expression is from a variable change from profit to marginal cost, including noting that, from (2.8), $dy = -K^*(\bar{c})dx$.

(2) Thus, the *ex-ante* private surplus for a random bidder is:

$$\begin{aligned}
\int_{\pi^{min}}^{\pi^{max}} \left[\int_{\pi^{min}}^{w} H^{I-1}(y)dy \right] h(w)dw &= \int_{\pi^{min}}^{\pi^{max}} \left[\int_{y}^{\pi^{max}} h(w)dw \right] H^{I-1}(y)dy \\
&= \int_{\pi^{min}}^{\pi^{max}} [1 - H(y)]\, H^{I-1}(y)dy \\
&= \int_{\pi^{min}}^{\pi^{max}} \frac{1 - H(y)}{I \cdot h(y)} I \cdot H^{I-1}(y) \cdot h(y)dy
\end{aligned}$$

where the first step comes from changing the integration order. Because that the cumulative probability distribution of the winner's type H^I. Thus, the corresponding density function is $\frac{dH^I(y)}{dy} = I \cdot H^{I-1}(y) \cdot h(y)$, which is in the integrand. This, means that each of the bidders' contribution to the private surplus is

$$E_y\left[\frac{1-H(y)}{I\cdot h(y)}\right] \;=\; \frac{1}{I}E_y\left[\frac{1-H(y)}{h(y)}\right]$$

where E_y is the expectation following the winner's profit distribution. Since there are I bidders and the public welfare is the social welfare minus private surplus, the *ex-ante* public welfare is given by

$$E_y\left[SW(y;\bar{c})-\frac{1-H(y)}{h(y)}\right] \;=\; E_x\left[CS(\bar{c})+\pi(x;\bar{c})-\frac{F(x)}{f(x)}\cdot K^*(\bar{c})\right]$$

where E_x is the expectation following the winner's unit cost distribution.

□

Simplification of problem MD

The expected public welfare can be rewritten as follows:

$$\begin{aligned}
&E_{\boldsymbol{c}}\left[\sum_{i=1}^{I}\left(q_i(\boldsymbol{c})\cdot CS(T(c_i),N(c_i),L(c_i))\right)-M_i(\boldsymbol{c})\right]\\
&=\sum_{i=1}^{I}\left\{E_{\boldsymbol{c}}\left[q_i(\boldsymbol{c})\cdot CS(T(c_i),N(c_i),L(c_i))\right]-E_{c_i}\left[E_{\boldsymbol{c}_{-i}|c_i}\left[M_i(c_i,\boldsymbol{c}_{-i})\right]\right]\right\}\\
&=\sum_{i=1}^{I}\{E_{\boldsymbol{c}}\left[q_i(\boldsymbol{c})\cdot CS(T(c_i),N(c_i),L(c_i))\right]\\
&\quad+E_{c_i}\left[\mathbb{P}_i(c_i)\cdot\pi(c_i;T(c_i),N(c_i),L(c_i))-V_i(c_i)\right]\}\\
&=\sum_{i=1}^{I}\left\{E_{\boldsymbol{c}}\left[q_i(\boldsymbol{c})\cdot\left(CS(T(c_i),N(c_i),L(c_i))+\pi(c_i;T(c_i),N(c_i),L(c_i))\right)\right]-E_{c_i}\left[V_i(c_i)\right]\right\}
\end{aligned}$$

where the second expression comes from the definition of indirect utility in (2.13), $V_i(c_i)$, and the third expression comes from the definition of conditional expectation $\mathbb{P}(c_i)$. We can also use the Envelope Theorem and apply integration by parts to obtain an expression for the last set of terms in the above expression:

$$
\begin{aligned}
E_{c_i}\left[V_i(c_i)\right] &= \int_{c^{min}}^{c^{max}} V_i(x)f(x)dx \\
&= F(x)V_i(x)|_{c^{min}}^{c^{max}} - \int_{c^{min}}^{c^{max}} \left(-\mathbb{P}_i(x)\cdot L(x)\right)F(x)dx \\
&= V_i(c^{max}) + \int_{c^{min}}^{c^{max}} \left(\mathbb{P}_i(x)\cdot L(x)\right)F(x)dx \\
&= V_i(c^{max}) + \int_{c^{min}}^{c^{max}} \left(\mathbb{P}_i(x)\cdot L(x)\cdot\frac{F(x)}{f(x)}\right)f(x)dx \\
&= V_i(c^{max}) + E_{\boldsymbol{c}}\left[q_i(\boldsymbol{c})\cdot L(c_i)\frac{F(c_i)}{f(c_i)}\right]
\end{aligned}
$$

Thus, the expected public welfare is

$$
\begin{aligned}
& \sum_{i=1}^{I} E_{\boldsymbol{c}}\Bigg[q_i(\boldsymbol{c})\cdot\left(CS(T(c_i),N(c_i),L(c_i)) + \pi(c_i;T(c_i),N(c_i),L(c_i)) - L(c_i)\frac{F(c_i)}{f(c_i)}\right) \\
& \quad -V_i(c^{max})\Bigg] \\
= \;& E_{\boldsymbol{c}}\Bigg[\sum_{i=1}^{I} q_i(\boldsymbol{c})\cdot\left(CS(T(c_i),N(c_i),L(c_i)) + \pi(c_i;T(c_i),N(c_i),L(c_i)) - L(c_i)\frac{F(c_i)}{f(c_i)}\right) \\
& \quad -V_i(c^{max})\Bigg]
\end{aligned}
$$

Proposition 4 *Optimal Scoring Function*

A first-best scoring auction with the scoring function appearing below results in an implementation of the Optimal Direct Mechanism.

$$\alpha^p(\tau, K) \quad = \quad CS(\tau, N(\tau, K), K) - \int_0^K \frac{F(c^p(k))}{f(c^p(k))} dk$$

where $c^p(\cdot)$ is the inverse of the public optimal capacity schedule, $L^p(c_i)$. Furthermore, adding the constraint $M \leq 0$ and adjusting the inverse schedule result in an implementation of the Budget-Balanced Public-Optimal Mechanism.

Proof: From Lemma 1, each firm maximizes

$$\alpha^p(\tau, K) + \pi(c_i; \tau, N(\tau, K), K) = SW(N(\tau, K), K) - \int_0^K \frac{F(c^p(k))}{f(c^p(k))} dk$$

, which corresponds to an adjusted social welfare function with construction cost function $\tilde{C}^c(c_i, K) = c_0 + c_i \cdot K + \int_0^K \frac{F(c^p(k))}{f(c^p(k))} dk$. From the Regularity Assumption, i.e., Assumption 4, $\tilde{C}^c(\cdot)$ is a convex function in τ and K. Along with Assumption 1, we have that a point satisfying the conditions below results in an optimal bid.

$$\frac{\partial SW(\cdot)}{\partial \tau} = 0 \quad \text{and} \quad \frac{\partial SW(\cdot)}{\partial K} - \frac{F(c^p(K))}{f(c^p(K))} = 0$$

which we note are the same conditions for a point optimizing virtual surplus. We can turn to the constrained optimization problem by claiming $M \leq 0$ beforehand, and apply the similar trick to implement the budget-balanced version of the optimal mechanism.

□

Lemma 3 *Regularity Condition for Demand Pricing Mechanism*

For any congestion function, $g(\cdot)$, N^d is decreasing with respect to c_i if and only if

$$2c_i\gamma_i'(c_i) \quad > \quad \frac{F(c_i)}{f(c_i)}, \quad c_i \in \mathcal{C}, i = 1, \ldots, I$$

Proof: From Equation (2.49), N^d strictly decreases, if and only if for any c_i, $\gamma_i'\mu^* - \frac{F}{f}\mu^{*\prime} > 0$. Differentiating (2.6) with respect to c_i gives

$$g''\mu^{*2}\mu' + 2g'\mu^*\mu^{*\prime} \quad = \quad 1$$

Noticing that $\mu^*(c_i) > 0$, $\mu^*(c_i)' > 0$, and that $g''(\cdot) \geq 0$, we have

$$2g'\mu^*\mu^{*\prime} \leq 1 \Rightarrow 2g'\mu^{*2}\mu^{*\prime} \leq \mu^* \Rightarrow 2c_i \leq \frac{\mu^*}{\mu^{*\prime}} \tag{2.52}$$

where the last line comes from (2.6). Now, we revisit the conditions on the derivatives,

$$\gamma_i'\mu^* - \frac{F}{f}\mu^{*\prime} > 0 \Rightarrow \gamma_i'\frac{\mu^*}{\mu^{*\prime}} - \frac{F}{f} > 0$$

In terms of the arguments for the proof:

Sufficiency: If $2c_i\gamma_i' > \frac{F}{f}$, then $\gamma_i'\frac{\mu^*}{\mu^{*\prime}} - \frac{F}{f} \geq 2c_i\gamma_i' - \frac{F}{f} > 0$.

Necessity: When $g'' = 0$, (2.52) becomes equality. Then, $2c_i\gamma_i'$ is equivalent to $\gamma_i'\frac{\mu^*}{\mu^{*\prime}}$ and the inequality, $2c_i\gamma_i' > \frac{F}{f}$, has to be satisfied. □

Proposition 6 *The government implements the optimal Demand-Pricing Mechanism,* **(DP)***, by offering Y, as specified from Equation* (2.50)*, and selecting the proposal with the highest demand.*

Proof: From the Revelation Principle, the demand pricing is optimal as long as, bidder with each c_i always selects the optimal demand $N^d(c_i)$ and the government selects the firm with the lowest c_i. Assume the bidder i selects a pair $(N', Y(N'))$, different from the optimal revealed decision, $(N^d(c_i), Y(N^d(c_i)))$. Then, due to unique mapping, $(N', Y(N'))$ must correspond to another marginal cost, $(N^d(c_i'), Y(N^d(c_i')))$. Due to the IC constraints, the expected utility from $(N', Y(N'))$ is lower than that from $(N^d(c_i), Y(N^d(c_i)))$. Thus, each bidder i will bid the optimal demand corresponding to their true marginal cost, $N^d(c_i)$. Next, as N^d is strictly decreasing, the highest demand corresponds to the lowest marginal cost. Thus, the bidder biding the highest demand is the most efficient firm. Putting both parts together, awarding the franchise to the bid with the highest demand, and offering Y yield an implementation of the Optimal Demand Pricing Mechanism. □

Lemma 4. *Performance Bound*

The virtual surplus loss from implementation of the Demand Pricing Mechanism is always less than the loss stemming from implementation of the commonly-used Naive Scoring Rule, i.e., $\Delta_n - \Delta_d \geq 0$.

Proof: From the definitions,

$$\begin{aligned}
\Delta_n - \Delta_d &= \frac{F(c_i)}{f(c_i)}K^*(c_i) - \int_{c_i}^{\gamma_i} K^*(h)dh - \int_{c_i}^{\gamma_i}(\gamma_i - h)\frac{N^\diamond(h,\gamma_i)\cdot\mu^{*\prime}(h)}{\mu^{*2}(h)}dh \\
&\geq \frac{F(c_i)}{f(c_i)}K^*(c_i) - \int_{c_i}^{\gamma_i}\frac{N^\diamond(h,h)}{\mu^*(h)}dh - \int_{c_i}^{\gamma_i}(\gamma_i - h)\frac{N^\diamond(h,h)\cdot\mu^{*\prime}(h)}{\mu^{*2}(h)}dh \\
&\geq \frac{F(c_i)}{f(c_i)}K^*(c_i) - N^\diamond(c_i,c_i)\cdot\left[\int_{c_i}^{\gamma_i}\frac{\gamma_i\cdot\mu^{*\prime}(h)}{\mu^{*2}(h)}dh + \int_{c_i}^{\gamma_i}\frac{\mu^*(h) - h\cdot\mu^{*\prime}(h)}{\mu^{*2}(h)}dh\right] \\
&= \frac{F(c_i)}{f(c_i)}K^*(c_i) - N^\diamond(c_i,c_i)\cdot\left[\frac{\gamma_i}{\mu^*(h)} - \frac{h}{\mu^*(h)}\right]_{\gamma_i}^{c_i} \\
&= \frac{F(c_i)}{f(c_i)}K^*(c_i) - N^\diamond(c_i,c_i)\cdot\left[\frac{\gamma_i}{\mu^*(c_i)} - \frac{\gamma_i}{\mu^*(\gamma_i)} + \frac{\gamma_i}{\mu^*(\gamma_i)} - \frac{c_i}{\mu^*(c_i)}\right] \\
&= \frac{F(c_i)}{f(c_i)}K^*(c_i) - N^\diamond(c_i,c_i)\cdot\frac{\gamma_i - c_i}{\mu^*(c_i)} \\
&= \frac{F(c_i)}{f(c_i)}K^*(c_i) - \frac{N^\diamond(c_i,c_i)}{\mu^*(c_i)}\frac{F(c_i)}{f(c_i)} = 0
\end{aligned}$$

$\square$

CHAPTER 3

Pareto Efficiency analysis of oligopolistic private road network

3.1. Introduction

Growing traffic demand and limited transportation infrastructure contribute to congestion. In the last 30 years, vehicle miles traveled (VMT) in the US have increased by about 80%, while road mileage has only increased by approximately 5% (FHWA, 2013). Limited funding is one of the factors that restrict the development of public roads. Hagquist (2008) suggests that traditional transportation financing sources, such as the Highway Trust Fund, which rely on fuel taxes, severely restrict the development of transportation infrastructure. As an alternative, the federal government and numerous states are considering private participation in the development of transportation infrastructure. Thirty-two states and Puerto Rico have enacted legislation enabling Public-Private Partnerships (P3s) (Geddes and Wagner, 2013). In addition to the development of new roads, i.e., design, construction, operations, and management, governments can also franchise existing roads to private firms. For example, in 2005, the Skyway Corporation won a 99-year franchise for the Chicago Skyway and became the first privatization of an existing road in the US (Enright, 2006).

Private participation in transportation is also common in Europe, where limited resources also serve as the rationale. Albalate et al. (2009), for example, analyzed the level of private participation in tolled motorways in Europe. They found that 37% of roads (by length) are under concession agreements – most of them operated and maintained by private firms. The private sector plays a particularly important role in Southern Europe, where these trends are even more pronounced. There are also a number of (local) firms involved in the development of toll roads in China and elsewhere in Asia (Tan et al., 2010), where the implementation of P3s often takes the form of Build-Operate-Transfer agreements. In these agreements, firms transfer the roads back to the government at the end of a fixed term (Yang and Meng, 2000). The Guangzhou-Shenzhen Super Highway Project, by a Hong Kong entrepreneur, is an emblematic example.

In addition to supplementing public funding, privatization of roads may have other advantages. Blom-Hansen (2003), for example, argue that private operation may outperform public operation because the public sector may lack incentives to perform efficiently. The lack of incentives, in part, is traced to the lack of competition. As opposed to private firms, public agencies may not have to maximize production or resource-allocation efficiency. Privatization, therefore, may have a positive impact on production efficiency in road construction and management. Finally, as an additional advantage of privatization, Small et al. (1989) mentions the general public's greater acceptance to pay tolls charged by private operators, as opposed to by public agencies.

At the same time, the private provision of transportation infrastructure does raise significant concerns. This is due, in part, to the misalignment of incentives between firms and governments (Mohring and Harwitz, 1962). These concerns are reflected in the literature, where papers in the last 2 decades have analyzed the reduction of social welfare due to privatization in different settings: network topologies and ownership regimes. For example, the seminal work of de Palma and Lindsey (2000) compares the social welfare of a 2-segment parallel road network, under 4 different ownership regimes: public-free, private-free, private-private, and public-private. Among other interesting results, they find that the competition between a private firm and a public agency yields greater social welfare than does competition between 2 private firms. Using a similar setup, Verhoef (2007) examines toll pricing when a firm interacts with a public road as either a substitute when the 2 roads are in parallel, or as a complement when the 2 roads are in series. The work shows that both the public sector, if it decides to toll, and firms have the incentive to absorb congestion externalities. Firms, however, charge a mark-up, which reduces social welfare. Interestingly, in the case of serial networks, the sum of the mark-ups set by independent, profit-maximizing operators exceeds the mark-up that would be charged by a single firm, i.e., a monopolist, operating the entire network. This means that decentralized, private operation of a pure serial network leads to reductions in both social welfare and in total operator profit! This phenomenon is referred to as (double) marginalization and is described by Mun and Ahn (2008) in the context of transportation networks. In an analogous fashion, Xiao et al. (2007) study social welfare in parallel networks. They provide an analytical bound on the mark-ups, and ensuing welfare loss as a function of the number of links, i.e., the level of competition.

Mechanisms to reduce or eliminate social welfare losses associated with mark-ups have been analyzed in the literature. An intuitive approach is to regulate toll and capacity choices, i.e., forcing firms to adopt welfare-maximizing toll and capacity levels. As is noted in the literature, depending on the structure of the construction cost function, this approach leads to operator losses in cases where there are *economies of scale*, i.e., in cases where marginal costs, which determine welfare-maximizing tolls, are less than average costs. This problem derives from the *Self-Finance Theorem* of Mohring and Harwitz (1962), which was extended to dynamic and to network settings by Arnott et al. (1993) and Yang and Meng (2000), respectively. It is relevant to note that this problem is of practical importance as empirical evidence suggests that road construction cost functions exhibit economies of scale stemming from significant fixed costs related to their equipment, technology, and capital costs (Small et al., 1989). We also note that certain situations, i.e.., network structure and ownership regimes, may exacerbate losses from setting tolls at marginal costs. Verhoef (2007) shows that, when a private road competes with a free, public substitute, welfare-maximizing tolls may not even recover variable construction costs.

Other mechanisms aimed at limiting welfare losses that have been considered include *second-best pricing*, auctions, and tax refunds. Second-based pricing involves setting tolls and capacities by solving social-welfare maximization problems subject to the constraint of non-negative operator profits (Verhoef, 2007). Ubbels and Verhoef (2008); Nombela and De Rus (2004) are important examples of economic analysis of auctions to franchise elements in transportation systems. Zhang and Durango-Cohen (2012) considers extending tax refunds to private operators. In addition to issues related to the implementation of these approaches, an important criticism is that they are predicated on the assumption of perfect competition, where road operators are price (and capacity) takers leading to a zero-profit outcome. In short, mechanisms and strategies aimed at reducing or eliminating social welfare losses may not provide sufficient incentive for private participation, and in turn, motivate the need for a different approach to benchmark interactions between governments and private firms.

Recognizing that, due to market structure and other factors, in many instances, (negotiated) outcomes are likely to fall in between each party's preferences, Tan et al. (2010) conduct a Pareto Analysis to characterize efficient combinations social welfare and operator profit. Interestingly, for the case of single roads, they find that social welfare and profit maximization lead to optimal concession periods of the same duration, and capacity levels leading to the same congestion/utilization/level-of-service. Most importantly, they show that any convex combination of social-welfare and profit-maximizing tolls can be Pareto Efficient, which suggests that efficient outcomes can arise naturally from negotiations between a government and a private operator. We review these results further in Section 3.2.

In terms of contributions, we not only extend the work of Tan et al. (2010) to road networks, but also examine the strategic effects of self-interested private firms. As performance benchmarks achieved under the government's planning, we first consider Aggregate Pareto Efficiency, which refers to the trade-off between social welfare and total/aggregate profits. This analysis leads to direct extensions of the results for single roads. Namely, optimal capacity levels ensure that the service quality, i.e., the utilization/volume-to-capacity (v/c) ratio, for each road in the network satisfies the same condition as that in the single road case. We also show that any convex combination of social-welfare and profit-maximizing tolls can be Aggregate Pareto Efficient. For the case of decentralized control, i.e., where operators set tolls and capacities independently with the objective to maximize their individual profits, we adapt the concept of Decentralized Pareto Efficiency from the work of Feldman (1973). This concept applies to toll and capacity decisions that may arise from the bilateral negotiations between each private operator and the government. We show that Decentralized Pareto Efficient capacity levels are set as in the centralized case and that efficient toll levels are bounded. The bounds, in turn, provide a necessary condition when decentralized tolling strategies are Aggregate Pareto Efficient. Further, we show an analytical example where the bounds, which indicate when the government's interventions/coordination, such as Kaldor-Hicks transfers, can bring a Pareto improvement, are determined by the network structure and total demand.

Features that distinguish our technical approach from others appearing in the literature are that (i) we do not restrict our analysis to specific network structures, but instead consider general networks under the user equilibrium condition; (ii) we extend the independence of the optimal v/c ratio to a general network and prove there is no network effect of (other links')toll/capacities on the optimal v/c ratio; (iii) our analytical results rely on a representation of the *marginal demand system* in matrix form, capturing the effects of changes in link tolls or capacities on either route or arc flows. These marginal effects are functions of congestion externalities, route structure, and demand elasticity. This characterization does not, as far as we are aware, appear elsewhere, and may be useful for comparative statics only superficially explored herein.

In terms of organization, the paper consists of 2 main sections: Section 3.2 presents the analysis for single roads, and Section 3.3 presents the extensions to the case of general networks. Each section begins with definitions of the social welfare, profit functions, and characterizations of the corresponding marginal demand systems. We then consider the social welfare and profit maximization problems. Finally, we present our Pareto Efficiency analysis. Each section includes an example. The example for road networks presented in Subsection 3.3.5 includes a few analytical results that apply to simple stage networks.

3.2. Analysis of the single road case

In this section, we present the analysis for the case of single roads. The objectives are to introduce notation, highlight results appearing in the literature, and provide the reader with a benchmark to compare the assumptions that are required to extend the analysis to the case of networks consisting of multiple roads. We begin by defining the elements and presenting expressions for the social welfare and operator profit functions. We then derive the marginal effects of toll and capacity choices on traffic volume under the assumption that demand is determined by the user equilibrium condition. We use the marginal effects to specify conditions satisfied by social-welfare and profit-maximizing toll and capacity levels and extend the analysis to characterize Pareto-Efficient strategies. We conclude the section with a numerical example.

3.2.1. Social welfare and firm profit

Following Verhoef (2007), for the single road case, expressions for the social welfare, i.e., aggregate surplus, and firm profit are given as follows:

$$SW(K,N) = \int_{n=0}^{n=N} P(n)dn - N \cdot C(N,K) - C^c(K) \tag{3.1}$$

$$\pi(\tau,K,N) = N \cdot \tau - C^c(K) \tag{3.2}$$

The variables τ, K, and N, respectively, represent the toll charge, road capacity, and the number of users. $P(n)$ denotes the inverse demand function, which corresponds to nth user's willingness to pay for travel, i.e., nth user's utility of travel. $C(N,K)$ is the average travel cost per user, which is the function of travel time weighted by the users' value of time. As is done elsewhere, we refer to $C(N,K)$ as the per-user average congestion cost function. $C^c(K)$ is the road construction cost function. Equation (3.1) excludes a term for toll revenues/payments, $N \cdot \tau$, because these transactions correspond to transfers between the parties.[1] We note that the above expressions apply to private or government construction and operation of tolled and untolled roads.

Assumption 6. *Demand and cost functions: First and second order conditions*

(1) *The inverse demand function, $P(n)$, is twice differentiable, decreasing, and concave, i.e., $P'(n) = \frac{\partial P(n)}{\partial n} < 0$ and $P''(n) = \frac{\partial^2 P(n)}{\partial n^2} \leq 0$.*

(2) *The construction cost function, $C^c(K)$, is linear, i.e., $C^c(K) = c_0 + c \cdot K$, where c_0 and c are, respectively, the fixed and per-unit-of-capacity cost.*

(3) *The congestion cost function is given as, $C(N,K) = g(\frac{N}{K})$, which is homogeneous of degree 0. Further, we let $\mu = \frac{N}{K}$, and assume that $g(\mu)$ is twice differentiable, increasing, and convex, i.e., $g'(\mu) = \frac{\partial g(\mu)}{\partial \mu} > 0$ and $g''(\mu) = \frac{\partial^2 g(\mu)}{\partial \mu^2} \geq 0$.*

μ is referred to as the *volume-to-capacity* or *v/c* ratio. Equations relating an operator's decision variables, τ and K, to the demand, N, are referred to as a *demand system*, which we denote $N(\tau, K)$. Here, we assume that the relationship among the variables is given by the *user equilibrium condition*, described in Assumption 7 below:

[1] Under the assumption of additive social welfare, consumer surplus corresponds to the difference between transport utility and the generalized travel cost: congestion cost plus toll payments. Producer surplus, i.e., operator profit, corresponds to the difference between toll revenue and construction cost.

Assumption 7. *User Equilibrium: Single Road*

The Nth user's willingness to pay equals the total generalized travel cost; that is,

$$P(N) = \tau + C(N, K)$$

In turn, Assumption 7 allows us to write the social welfare and profit functions in terms of the operator's decision variables, τ and K: $SW\left(K, N(\tau, K)\right)$ and $\pi\left(\tau, K, N(\tau, K)\right)$.

3.2.2. Marginal effects: Toll and capacity effects on demand

In general, it can be difficult to find an expression for $N(\tau, K)$. We can, however, use the equilibrium condition to derive the marginal effects of decisions, τ and K, on the demand. In particular, taking the derivative of the condition in Assumption 7 with respect to τ we get

$$P'(N)\frac{\partial N}{\partial \tau} = 1 + C_N \frac{\partial N}{\partial \tau}$$

where $P'(N)$ is the derivative of the inverse demand function evaluated at $n = N$, and C_N is the partial derivative of the average congestion cost function with respect to number of users evaluated at $N, K,$, i.e., $C_N \equiv \frac{\partial C(N,K)}{\partial N}$. C_N represents the *average congestion externality* induced by the Nth user. Rewriting the above expression, we obtain

$$\frac{\partial N}{\partial \tau} = \frac{1}{P'(N) - C_N} \tag{3.3}$$

Repeating the analysis with respect to capacity, we have:

$$\begin{aligned} P'(N)\frac{\partial N}{\partial K} &= C_N\frac{\partial N}{\partial K} + C_K, \text{ and} \\ \frac{\partial N}{\partial K} &= \frac{C_K}{P'(N) - C_N} \end{aligned} \tag{3.4}$$

where $C_K \equiv \frac{\partial C(N,K)}{\partial K}$ is the marginal congestion cost reduction per unit of capacity at N, K.

Assumption 6 ensures expressions (3.3) and (3.4) are well-defined. It also follows that $P'(N) - C_N < 0$, and that the overall effects are as follows: $\frac{\partial N}{\partial \tau} < 0$ and $\frac{\partial N}{\partial K} > 0$. That is, under Assumptions 6 and 7, increasing the toll leads to fewer users, and increasing capacity leads to additional users. Interestingly, we also note that the marginal effects of toll and capacity on demand are sensitive to congestion externalities. Large (average) congestion externalities reduce demand sensitivity to changes in tolls or capacities.

Combining (3.3) and (3.4),

$$\frac{\partial \tau}{\partial K} = \frac{\partial C(N,K)}{\partial K} = C_K \tag{3.5}$$

Recalling that $C_K < 0$, this means that a 1 unit capacity increase (reduction) offsets a -\$$C_K$ increase (reduction) in the toll, i.e., making the 2 prescribed changes leads to the same generalized travel costs and demand. -\$$C_K$ is the equilibrium-preserving *marginal rate of substitution* between toll charge and capacity.

3.2.3. Social welfare and profit maximization

We begin our analysis by considering the toll and capacity levels that maximize social welfare under the assumption that the demand system is given by the equilibrium condition. The first-order condition for social-welfare-maximizing toll level is:

$$\begin{aligned}
\frac{\partial SW(K,N(\tau,K))}{\partial \tau} &= \frac{\partial SW(K,N(\tau,K))}{\partial N}\cdot\frac{\partial N(\tau,K)}{\partial \tau} = 0 \Rightarrow \qquad (3.6)\\
&\frac{\partial SW(K,N(\tau,K))}{\partial N} = 0 \text{ because } \tfrac{\partial N}{\partial \tau} < 0\\
&\frac{\partial}{\partial N}\left(\int_{n=0}^{n=N} P(n)dn - N\cdot C(N,K) - C^c(K)\right) = 0\\
&P(N) - C(N,K) - N\cdot C_N = 0 \Rightarrow\\
\tau^*_{SW} &= N\cdot C_N \text{ follows from Assumption 7} \qquad (3.7)
\end{aligned}$$

Thus, the optimal toll, τ^*_{SW}, is set to the congestion cost imposed on the system by the Nth user, i.e., the marginal congestion externality. The first-order condition for the optimal capacity level is:

$$\frac{\partial SW(K,N(\tau,K))}{\partial K} = -N\cdot C_K - \frac{\partial C^c(K)}{\partial K} + \frac{\partial SW(K,N(\tau,K))}{\partial N}\cdot\frac{\partial N(\tau,K)}{\partial \tau}\cdot\frac{\partial \tau}{\partial K} = 0$$

$$\begin{aligned}
&-N\cdot C_K - c = 0 \text{ follows from (3.6)}\\
&-N\frac{dg(\mu)}{d\mu}\frac{\partial \mu}{\partial K} - c = 0, \text{ follows from Assumption 6}\\
&-N\cdot g'(\mu)\cdot\frac{-N}{K^2} = c\\
&\mu^* : \mu^2\cdot g'(\mu)\big|_{\mu=\mu^*} = c \qquad (3.8)
\end{aligned}$$

Thus, for a given demand, N, the optimal capacity level, $K^*(N)$, is set so that the v/c ratio satisfies condition (3.8). Since the congestion externalities , $N \cdot C_N = \mu \cdot g'(\mu)$, can be purely determined by the v/c ratio, (3.8) implies that the optimal v/c ratio leads to the constant level of congestion externalities, $\frac{c}{\mu^*}$, regardless of the toll or demand level. Rewrite expressions for the optimal demand and toll, N^*_{SW} and τ^*_{SW}:

$$N^*_{SW} : P(N) - g(\mu^*) - \frac{c}{\mu^*}\bigg|_{N=N^*_{SW}} = 0 \tag{3.9}$$

$$\tau^*_{SW} = \frac{c}{\mu^*} \tag{3.10}$$

The first-order conditions for the profit maximization problem are:

$$\frac{\partial\pi(\tau, K, N(\tau, K))}{\partial\tau} = N + \frac{\partial\pi(\tau, K, N(\tau, K))}{\partial N} \cdot \frac{\partial N(\tau, K)}{\partial\tau} = 0 \tag{3.11}$$

and

$$\begin{aligned}
\frac{\partial\pi(\tau, K, N(\tau, K))}{\partial K} &= -c + \frac{\partial\pi(\tau, K, N(\tau, K))}{\partial N} \cdot \frac{\partial N(\tau, K)}{\partial K} = 0 \\
&\quad -c + \frac{\partial\pi(\tau, K, N(\tau, K))}{\partial N} \cdot \frac{\partial N(\tau, K)}{\partial\tau} \cdot \frac{\partial\tau}{\partial K} = 0 \\
&\quad -c + \frac{\partial\pi(\tau, K, N(\tau, K))}{\partial N} \cdot \frac{\partial N(\tau, K)}{\partial\tau} \cdot C_K = 0 \text{ follows from (3.5)} \\
&\quad -c - N \cdot C_K = 0 \text{ follows from (3.11)}
\end{aligned} \tag{3.12}$$

Thus, the solution to the system of equations (3.11) and (3.12) is such that the optimal capacity level is set so that the v/c ratio satisfies condition (3.8). Rewriting (3.11) yields the following expressions for the profit maximizing demand and toll levels:

$$N_\pi^*: \qquad P(N) - g(\mu^*) - \frac{c}{\mu^*} + N \cdot P'(N)\Big|_{N=N_\pi^*} = 0 \tag{3.13}$$

$$\tau_\pi^* = N \cdot C_N - N \cdot P'(N) = \frac{c}{\mu^*} - N_\pi^* \cdot P'(N_\pi^*) = \tau_{SW}^* - N_\pi^* \cdot P'(N_\pi^*) \tag{3.14}$$

In summary, for the single road case, we show that social welfare and profit-maximizing capacity levels are set so that the ensuing v/c ratio, μ^*, satisfies condition (3.8). Such results for the single road case already appear in the literature(Tan et al., 2010), but we extend and build on them in the following sections. Essentially, the optimization of capacity choice is the result of the trade-off between congestion cost and capacity expansion cost. Moreover, the optimization of v/c ratio is independent of demand and tolling decisions in either profit optimization or social welfare optimization. To see this, we solve optimal v/c ratio by (3.8) and then determine optimal demand by (3.13) or (3.9) and optimal tolling choice by (3.14) or (3.10). Note that the profit-maximizing toll is greater than the welfare-maximizing toll, i.e., $\tau_\pi^* - \tau_{SW}^* = -N_\pi^* \cdot P'(N_\pi^*) > 0$. That is, the government and private firms have the same incentive to balance the cost between congestion and capacity investment but have different incentives to set up toll and demand. The private sector prefers a higher toll and lower demand than the government does. Second-order conditions are presented in 3.5.14.

3.2.4. Pareto Efficiency analysis

Given that the objectives of social welfare maximization and profit maximization are not aligned, it is of interest to characterize outcomes in terms of their Pareto Efficiency (Tan et al., 2010; Guo and Yang, 2009). We briefly visit their results and then characterize the Pareto Efficiency Frontier by presenting parametric solutions to the corresponding Scalarization Problems (Geoffrion, 1968).

Definition 5. *Pareto Efficiency*
For the single road case, the set of variables, $\{\tau^+, K^+, N^+\}$, is said to be strictly Pareto Optimal/Efficient, if and only if there is no combination $\{\hat{\tau}, \hat{K}, \hat{N}\}$ such that $SW(\hat{K}, \hat{N}) \geq SW(K^+, N^+)$ and $\pi(\hat{\tau}, \hat{K}, \hat{N}) \geq \pi(\tau^+, K^+, N^+)$, with at least one of the 2 inequalities being strict.

The set of all Pareto Efficient toll, capacity, and demand combinations is referred to as the *Pareto Frontier*. A deviation from $\{\bar{\tau}, \bar{K}, \bar{N}\}$ to $\{\hat{\tau}, \hat{K}, \hat{N}\}$, where $SW(\hat{K}, \hat{N}) \geq SW(\bar{K}, \bar{N})$, $\pi(\hat{\tau}, \hat{K}, \hat{N}) \geq \pi(\bar{\tau}, \bar{K}, \bar{N})$, with at least one inequality being strict, is referred to as a *Pareto Improvement*. The Pareto Frontier, therefore, includes all combinations of outcomes where there is no Pareto Improvement, i.e., no party can improve their outcome without adversely affecting others.

For a given $0 \leq \alpha \leq 1$, the Scalarization Problem for the single road, joint social welfare and profit-maximizing problem, SP(α), is presented below.

$$\max_{\tau,K,N} \quad (1-\alpha)\cdot SW(K,N) + \alpha\cdot\pi(\tau,K,N) \tag{3.15}$$

subject to:

$$P(N) = \tau + C(N,K) \tag{3.16}$$

where the parameter α represents the weight or bargaining power for the profit maximizer relatively to the government. As shown by Geoffrion (1968), the Pareto Frontier is given by the union of the solutions to the Scalarization Problems obtained by allowing the parameter α to vary in the $[0,1]$ interval. Using the similar optimization approaches[2], we show that the optimal solution to SP(α) is given by:

$$N^+(\alpha): \quad P(N) - g(\mu^*) - \frac{c}{\mu^*} + \alpha\cdot N\cdot P'(N)\Big|_{N=N^+(\alpha)} = 0 \tag{3.17}$$

$$K^+(\alpha) = \frac{N^+(\alpha)}{\mu^*} \tag{3.18}$$

$$\tau^+(\alpha) = \frac{c}{\mu^*} - \alpha\cdot N^+(\alpha)\cdot P'(N^+(\alpha)) \tag{3.19}$$

Second-order conditions ensuring the optimality of the above solution are considered in 3.5.14. The Pareto Frontier is given by $\{\tau^+(\alpha), K^+(\alpha), N^+(\alpha)\}$ where α takes on all values in $[0,1]$. Assumption 6 ensures the Pareto Frontier is continuous with the outcome variables taking on values in-between the social welfare and profit maximizing levels. Due to the independence of v/c ratio , Pareto Efficient contracts always follow the optimal v/c

[2]See 3.5.1 for detailed derivations.

ratio. While Pareto Optimal tolls are monotonically-increasing with α, i.e., the weight for profit., whereas Pareto Optimal demand and capacity levels are monotonically-decreasing.

3.2.5. Numerical example for single road case

To make the above results tangible, we consider a numerical example where the inverse demand, construction cost, and congestion cost functions are as follows:

$$P(N) = 42 - 0.02 \cdot N \tag{3.20}$$

$$C^c(K) = 4 \cdot K \tag{3.21}$$

$$C(N, K) = 10 \cdot \left[1 + 0.15 \cdot \left(\frac{N}{K}\right)^4\right] \tag{3.22}$$

where τ is measured in \$ per vehicle, K and N are measured in vehicles per hour. $SW(K, N)$ and $\pi(\tau, K, N)$ correspond to the welfare and profit rates in \$ per hour. Thus, $P(N)$ and $C(N, K)$ are in \$ per vehicle, and $C^c(K)$ is in \$ per hour. Here, we assume the inverse demand function, $P(N)$, is linear, that the fixed construction costs, c_0, are \$0 per hour, and that the congestion cost function $C(N, K)$, follows the BPR specification (USBPR, 1964). The \$10 per hour may, for instance, correspond to an average value of time of \$0.5 per minute per vehicle for a road with a free-flow travel time of 20 minutes.

Figure 3.1(a) presents the Pareto Efficient combinations of demand and capacity. The slope of the straight line connecting the social welfare and profit-maximizing levels corresponds to the optimal v/c ratio, μ^*. The profit-maximizing capacity and demand levels are less than corresponding social welfare maximizing levels. Figure 3.1(b) shows how social welfare and profit function change for the range of possible Pareto Efficient tolls. We observe that social welfare is maximized at a toll of approximately \$4 per vehicle. The welfare is decreasing with the toll. On the other hand, the profit function increases with the toll. The maximizing toll level is approximately \$16 per vehicle.

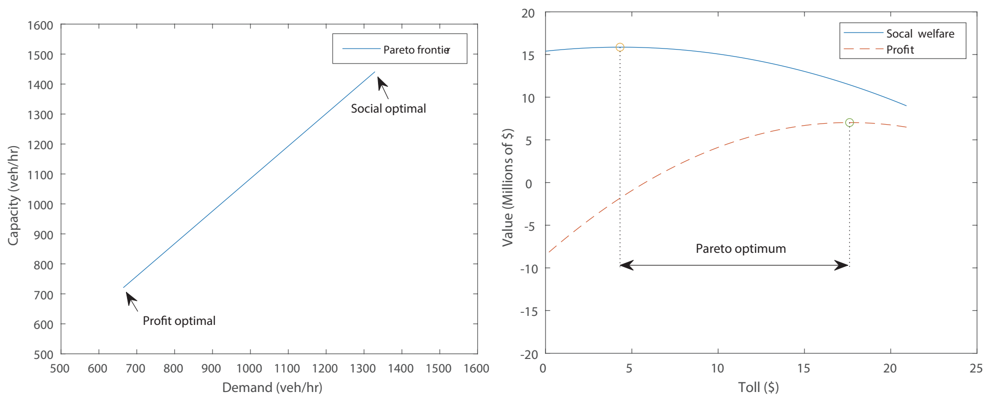


Figure 3.1. (a) Single Road Pareto Frontier: Demand (N) vs. Capacity (K); (b) Social welfare, $SW(K,N)$, and profit, $\pi(\tau,K,N)$, as a function of toll, τ.

3.3. Analysis of the network case

Here, we consider extensions of the results presented in the previous section to the case of transportation networks consisting of L road links, each with its own (private) operator. The description of the network case follows Small and Verhoef (2007), although, for simplicity, we consider the single market, i.e., single-origin single-destination. This assumption is common in the literature. It seems applicable in the analysis of inter-city networks, as opposed to urban networks, and is intended to highlight the effect of the network structure/topology, e.g., to study the effect of competition in networks comprised of parallel road segments (Xiao et al., 2007; de Palma and Lindsey, 2000).

3.3.1. Social welfare and firm profits

Extensions of the social welfare and profit functions, Equations (3.1) and (3.2), for the network case are as follows:

$$SW(\mathbf{K},\mathbf{N}) = \int_{n=0}^{n=N} P(n)dn - \sum_{l=1}^{L} N_l \cdot C(N_l, K_l) - \sum_{l=1}^{L} C^c(K_l) \tag{3.23}$$

$$\pi_l(\mathbf{T},\mathbf{K},\mathbf{N}) = N_l \cdot \tau_l - C^c(K_l),\ l = 1, \ldots, L \tag{3.24}$$

Here, $\mathbf{T},\mathbf{K},\mathbf{N}$ are L-dimensional vectors, with components τ_l, K_l, N_l, respectively, representing the toll, capacity, and traffic in road l. $\pi_l(\mathbf{T},\mathbf{K},\mathbf{N})$ corresponds to road l operator's profit. We continue to rely on the 3 elements of Assumption 6 for the inverse demand function, as well as the construction and congestion cost functions that apply to each road, l. To extend Assumption 7, we follow the development (and notation) in Small and Verhoef (2007), where the demand, N, is split among R possible routes, i.e., $N = \sum_{r=1}^{R} v_r$. v_r denotes the flow, i.e., the number of users, traversing the network along route r, where the route flows can be collected in the R-dimensional vector, $\mathbf{V}$.

The user equilibrium assumption for the network case is stated as follows Wardrop (1952):

Assumption 8. *User Equilibrium: Network Case*

For a single OD problem with total throughput N, used routes, i.e., those with $v_r > 0$, have generalized travel costs equal to the Nth user's willingness to pay, $P(N)$; and unused routes, i.e., those with $v_r = 0$, have generalized travel costs greater than $P(N)$.

Mathematically:

$$
\begin{aligned}
v_r &\geq 0, \ r = 1, \ldots, R && (3.25)\\
v_r \cdot \left[\sum_{l=1}^{L} \delta_{rl} \cdot (C(N_l, K_l) + \tau_l) - P(N) \right] &= 0, \ r = 1, \ldots, R && (3.26)\\
\sum_{l=1}^{L} \delta_{rl} \cdot (C(N_l, K_l) + \tau_l) - P(N) &\geq 0, \ r = 1, \ldots, R && (3.27)
\end{aligned}
$$

δ_{rl} is an indicator variable with value 1, if road l is on route r; and value 0, otherwise. The $R \times L$ matrix, $\Delta = [\delta_{rl}]$, is referred to as the route-link incidence matrix. Thus, $N_l = \sum_{r=1}^{R} \delta_{rl} \cdot v_r$ and $\boldsymbol{N} = \Delta^T \boldsymbol{V}$.

3.3.2. Marginal effects: Toll and capacity effects on demand

Here, we extend the analysis in Section 3.2.2 to the network case. We begin writing the conditions that describe the demand system and consider the associated marginal effects of the operators' decision variables, $\mathbf{T}$ and $\mathbf{K}$, on demand. We also address the marginal rates of substitution between tolls and capacity charges.

First, we observe that the equilibrium condition in the network case, Equation (3.26), yields (up to) R conditions that must be satisfied by the aggregate demand, N, for a given set of tolls and capacities, $\mathbf{T}$ and $\mathbf{K}$.[3] These conditions, referred to as the *demand system* for the network case, allow us to write the route flows as a function of the tolls

[3]Under the assumption that $v_r > 0, r = 1, \ldots, R$, i.e., that unused routes, including those that contain cycles, are removed from Δ, (3.26) yields R conditions:

$$\sum_{l=1}^{L} \delta_{rl} \cdot (C(N_l, K_l) + \tau_l) = P(N), r = 1, \ldots, R \qquad (3.28)$$

One of the implications of removing unused routes from Δ is that unused links are also removed, i.e., every column of Δ has at least one 1.

and capacities, i.e., $v_r(\mathbf{T}, \mathbf{K}), r = 1, \ldots, R$. We also recall that the aggregate demand, N, and the link flows, $N_l, l = 1, \ldots, L$, can be determined from the route flows. As in Section 3.2.2, we consider marginal changes in demand with respect to the operators' decision variables. In particular, from (3.28), we have:

$$
\begin{aligned}
\frac{\partial}{\partial \tau_i}\left(P(N)\right) &= \frac{\partial}{\partial \tau_i}\left(\sum_{l=1}^{L} \delta_{rl} \cdot \left[C(N_l, K_l) + \tau_l\right]\right), \ r = 1, \ldots, R, i = 1, \ldots, L \\
P'(N)\frac{\partial N}{\tau_i} &= \sum_{l=1}^{L} \delta_{rl} \cdot \frac{\partial C(N_l, K_l)}{\partial N_l} \cdot \left(\sum_{s=1}^{R} \frac{\partial N_l}{\partial v_s} \cdot \frac{\partial v_s}{\partial \tau_i}\right) \\
&\quad + \sum_{l=1}^{L} \delta_{rl} \cdot \frac{\partial \tau_l}{\partial \tau_i}, \ r = 1, \ldots, R, i = 1, \ldots, L \\
P'(N) \cdot \left(\sum_{s=1}^{R} \frac{\partial v_s}{\partial \tau_i}\right) &= \sum_{l=1}^{L} \delta_{rl} \cdot C_{N_l} \cdot \left(\sum_{s=1}^{R} \delta_{sl} \cdot \frac{\partial v_s}{\partial \tau_i}\right) + \delta_{ri}, \ r = 1, \ldots, R, i = 1, \ldots, L
\end{aligned}
$$

Defining the $R \times L$ Jacobian matrix of the route flows with respect to the tolls, $\mathcal{J}_\tau^v \equiv \left[\frac{\partial v_s}{\partial \tau_i}\right]$, we have that entry (s, i) in the matrix corresponds to $\frac{\partial v_s}{\partial \tau_i}$, and allows us to rewrite the above set of $R \times L$ equations in matrix form as follows:

$$
\mathbf{1}_{R\times R}\mathcal{J}_\tau^v P'(N) = \Delta \mathcal{C}_N \Delta^T \mathcal{J}_\tau^v + \Delta \tag{3.29}
$$

where $\mathcal{C}_N$ is a diagonal matrix consisting of the marginal congestion costs associated with demand increases in each link; that is, $\mathcal{C}_N = diag\left\{C_{N_1}, C_{N_2}, \ldots, C_{N_L}\right\}$. Similarly, the marginal congestion costs, i.e., reductions, associated with capacity expansions can be collected in the matrix $\mathcal{C}_K = diag\left\{C_{K_1}, C_{K_2}, \ldots, C_{K_L}\right\}$, which allows us to write

$$\mathbf{1}_{R\times R}\mathcal{J}_K^v P'(N) = \Delta\mathcal{C}_N\Delta^T\mathcal{J}_\tau^v + \Delta\mathcal{C}_K \tag{3.30}$$

Equations (3.29) and (3.30) are analogous to the single road counterparts.[4] The associated marginal effects are unique when $\left[\mathbf{1}_{R\times R}P'(N) - \Delta\mathcal{C}_N\Delta^T\right]$ is invertible. In general, this condition is too strong, and thus, there can be multiple solutions to the above systems of equations.

We can also derive Jacobian matrices for the link flows, $\mathbf{N}$, where $\mathcal{J}_\tau^N = \left[\frac{\partial N_i}{\partial \tau_l}\right] = \Delta^T\mathcal{J}_\tau^v$ and $\mathcal{J}_K^N = \left[\frac{\partial N_i}{\partial K_l}\right] = \Delta^T\mathcal{J}_K^v$. Proposition 7 extends the results in Section 3.2.2 to the link flows on a network of roads. Lemma 5 is an intermediate result.

Lemma 5. *There exists a solution to the linear system of equations given by* $\Delta\mathbf{x} = \mathbf{1}_{R\times 1}$.

Proof See 3.5.6.

Lemma 5 provides a convenient way to connect link flows to total throughput, even though Δ may not be invertible. In particular, it follows that $\mathbf{x}^T\mathbf{N} = \mathbf{x}^T\Delta^T\mathbf{V} = \mathbf{1}_{R\times 1}^T\mathbf{V} = N$, which states that the total traffic departing the origin is equal to the sum of traffic along all routes, which, in turn, is equal to the total demand/throughput, N. Now, noticing that $\mathbf{1}_{R\times 1}^T\mathcal{J}_\tau^v = \mathbf{x}^T\mathcal{J}_\tau^N$, we can rewrite the marginal route flow systems, (3.29) and (3.30), as follows:

[4]Two special cases of networks discussed later are the (pure) parallel and (pure) serial networks. The corresponding incidence matrices are $\Delta_p = \mathcal{I}_{L\times L}$ and $\Delta_s = \mathbf{1}_{1\times L}$, respectively. $\mathcal{I}_{L\times L}$ is and $L\times L$ identity matrix, and $\mathbf{1}_{1\times L}$ is a row vector where all components are 1.

$$\mathcal{X}\mathcal{J}_{\tau}^{N}P'(N) = \Delta\mathcal{C}_{N}\mathcal{J}_{\tau}^{N} + \Delta \tag{3.31}$$

$$\mathcal{X}\mathcal{J}_{K}^{N}P'(N) = \Delta\mathcal{C}_{N}\mathcal{J}_{K}^{N} + \Delta\mathcal{C}_{K} \tag{3.32}$$

where $\mathcal{X} \equiv [\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}]^T$.

Proposition 7. *Effect of decision variables on the network's link flows*

(1) *The solutions to marginal link flow system, $\mathcal{J}_{\tau}^{N}$ and $\mathcal{J}_{K}^{N}$, are unique.*

(2) *The capacity effect on link flow is the product of toll effect and marginal return of capacity, i.e., for any links i and l, $\frac{\partial N_i}{\partial K_l} = \frac{\partial N_i}{\partial \tau_l} C_{K_l}$.*

(3) *The flow on a link decreases with its toll and increases with its capacity, i.e., for any link i, $\frac{\partial N_i}{\partial \tau_i} < 0$ and $\frac{\partial N_i}{\partial K_i} > 0$.*

Proof See 3.5.2.

Proposition 7 shows that the elements of Assumption 6 allow us to extend the results to the link flows, including the toll-capacity marginal rate of substitution for a given link. We do note, however, that the aggregate effects of toll and capacity changes are inconclusive. Braess et al. (2005) presents an example of a counter-intuitive situation where adding capacity results in a higher travel cost and lower total demand. The proof of Proposition 9, relies on the assumption, as stated below.

Assumption 9. *Non-Braess networks*

(1) *Increasing toll on any road does not increase overall network demand, i.e.,* $\frac{\partial N}{\partial \tau_l} \leq 0, l = 1, \ldots, L$.

(2) *Increasing capacity on any road does not decrease overall demand, i.e.,* $\frac{\partial N}{\partial K_l} \geq 0, l = 1, \ldots, L$.

3.3.3. Social welfare and profit maximization

Again, we extend the analysis to the network case. For profit maximization, we consider both the problem of maximizing aggregate profit, as well as individual operator profit maximization. The former leads to a direct extension of the single road results, whereas the latter applies to situations where the operators are self-interested agents. Both cases are important building blocks for the Pareto Efficiency analysis presented in the following section. The link i's profit optimization problem is given by,

$$\max_{\tau_i, K_i} \quad \pi_i(\tau_i, K_i, N_i(\mathbf{V}(\tau_i, \mathbf{T}_{-i}, K_i, \mathbf{K}_{-i}))) = \tau_i \cdot N_i - C^c(K_i)$$

where $\mathbf{T}_{-i}$ and $\mathbf{K}_{-i}$ are tolls and capacities for the other $L-1$ operators, respectively. The aggregate profit optimization problem is given by,

$$\max_{\mathbf{T}, \mathbf{K}} \quad \Pi(\mathbf{T}, \mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) = \sum_{l=1}^{L} \tau_l \cdot N_l - \sum_{l=1}^{L} C^c(K_l)$$

The social welfare maximization problem expressed as a function of the operators' decision variables is as follows:

$$\max_{\mathbf{T}, \mathbf{K}} \quad SW(\mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) = \int_0^N P(n)dn - \sum_{l=1}^{L} N_l \cdot C(N_l, K_l) - \sum_{l=1}^{L} C^c(K_l)$$

Before solving the above three optimization problems, we generalize the independence of v/c ratio discussed in the single road case.

Proposition 8. *The optimal v/c ratio of an arbitrary link i in an arbitrary road network is independent of:*

(1) *whether the goal is link profit, aggregate profit or social welfare;*

(2) *its own tolling choice and other link's tolling/capacity choices.*

Proof See 3.5.3.

The first independence is the generalization of the single road result shown in Section. 3.2.3. The second independence implies that there are no network effects on the optimal choice of v/c ratio. It is worth noting that Verhoef (2007) and Yang and Meng (2000) show that the first-best and self-balance of a road link require the optimality of other links. However, we show the independence of the optimal v/c ratio for arbitrary choices of toll and capacities of other links in a network. The independence raises an issue when there are routes with different lengths. Since the congestion level is constant, the longer route incurs significantly higher travel costs. The optimal planning strategy is then to eliminate the longer route and put more resources on shorter routes. Thus, we make the following assumptions:

Assumption 10. *Homogeneity of routes:*

(1) *All links have the same congestion cost function and construction cost function.*

(2) *All route traverses the same number of links. Let M be the number of links, then $\sum_{l=1}^{L} \delta_{rl} = M, r = 1, \ldots, R.$*

For any of the three optimization goals, we can reduce the toll/capacity decision making to choosing volumes or tolls with capacities choices adjusted by the link volume. The above homogeneity assumptions further simplifies[5] the analysis, as the route congestion, is homogeneous. We start with social welfare optimization. The first-order condition for the social optimum of a network is:

$$N^*_{SW}: \qquad P(N) - \sum_{l=1}^{L} \delta_{sl} \cdot \left[g(\mu^*_l) + \frac{c}{\mu^*_l} \right] \Bigg|_{N=N^*_{SW}} = 0, s = 1, \ldots, R \tag{3.33}$$

$$\tau^*_{SW}(s) \;=\; \sum_{l=1}^{L} \delta_{sl} \cdot \frac{c}{\mu^*_l}, s = 1, \ldots, R \tag{3.34}$$

where μ^*_l stands for the optimal v/c ratio in link l. Note that (3.33) can only be solved in the case of homogeneous routes. If otherwise, the weighted average of the left-hand-side of R equations in (3.33) equals to 0. We use the conditions to obtain closed-form expressions (3.34) for the optimal route tolls, $\tau^*_{SW}(r)$, given by $\sum_{l=1}^{L} \delta_{rl} \cdot \tau_l$. Analogous to the single road case, the social optimal route tolls are given by the sum of congestion externalities imposed by each route's marginal user. However, the link toll does not have to internalize the link congestion externalities. That is, as long as the route toll equals to the (weighted average)route congestion externalities, any combinations of links tolls are socially optimal.

[5]When routes are not homogeneous, we can still reduce the dimensions to link volumes. In that case, the optimal condition requires the weighted average route congestion externalities to be internalized. Essentially, we simplify the calculation of the weighted average in the homogeneous case.

Next, we move to the aggregate profit optimal decisions. Prior to first order considerations, we observe that:

$$
\begin{aligned}
\Pi\left(\mathbf{T},\mathbf{K},\mathbf{N}\right) - SW\left(\mathbf{K},\mathbf{N}\right) &= \sum_{l=1}^{L} \tau_l \cdot N_l - \sum_{l=1}^{L} C^c(K_l) \\
&\quad - \left[\int_{n=0}^{n=N} P(n)dn - \sum_{l=1}^{L} N_l \cdot C(N_l, K_l) - \sum_{l=1}^{L} C^c(K_l) \right] \\
&= \sum_{l=1}^{L} \tau_l \cdot N_l - \int_{n=0}^{n=N} P(n)dn + \sum_{l=1}^{L} N_l \cdot C(N_l, K_l) \\
&= \sum_{l=1}^{L} N_l \cdot [\tau_l + C(N_l, K_l)] - \int_{n=0}^{n=N} P(n)dn \\
&= \sum_{l=1}^{L} \sum_{r=1}^{R} v_r \cdot \delta_{rl} \cdot [\tau_l + C(N_l, K_l)] - \int_{n=0}^{n=N} P(n)dn \\
&= \sum_{r=1}^{R} v_r \cdot \sum_{l=1}^{L} \delta_{rl} \cdot [\tau_l + C(N_l, K_l)] - \int_{n=0}^{n=N} P(n)dn \\
&= \sum_{r=1}^{R} v_r \cdot P(N) - \int_{n=0}^{n=N} P(n)dn \\
&= NP(N) - \int_{n=0}^{n=N} P(n)dn
\end{aligned}
$$

Note that the difference between aggregate profit and social welfare is purely a function of total demand. There is no disagreement on the optimal traffic assignment between the aggregate profit maximizer and the social welfare maximizer. When the total demand is inelastic, profit optimum coincides with social optimum. While in the elastic demand case, the disagreement rises from the choice of total demand:

$$
\begin{aligned}
N_\Pi^* : \qquad & P(N) - \sum_{l=1}^{L} \delta_{sl} \cdot \left[g(\mu_l^*) + \frac{c}{\mu_l^*} \right] - NP'(N) \bigg|_{N=N_\Pi^*} = 0, s = 1, \ldots, R \quad (3.35) \\
\tau_\Pi^*(r) \quad = \quad & \tau_{SW}^*(r) - NP'(N)|_{N=N_\Pi^*}, r = 1, \ldots, R \quad (3.36)
\end{aligned}
$$

Because $NP'(N)|_{N=N_\Pi^*} < 0$, we see that aggregate-profit-maximizing route tolls are greater than the corresponding social-welfare-maximizing route tolls, and that social-welfare optimizing total demand levels exceed the profit-maximizing levels, i.e., $\tau_\Pi^*(r) > \tau_{SW}^*(r)$ and $N_\Pi^* < N_{SW}^*$ for $r = 1, \ldots, R$. Essentially, for the aggregate goals, i.e., social welfare and aggregate profit, the optimization problem on network case is analogous to that on single road case.

We also consider operator i's profit maximization problem, given tolls and capacities for the other $L-1$ operators, $\mathbf{T}_{-i}$ and $\mathbf{K}_{-i}$, respectively, the first-order conditions are as follows:

$$
\begin{aligned}
\frac{\partial \pi_i(\cdot)}{\partial \tau_i} &= N_i + \tau_i \cdot \frac{\partial N_i}{\partial \tau_i} = 0 \quad (3.37) \\
\frac{\partial \pi_i(\cdot)}{\partial K_i} &= -c + \tau_i \cdot \frac{\partial N_i}{\partial K_i} = 0 \quad (3.38) \\
& \quad -c + \tau_i \cdot \frac{\partial N_i}{\partial \tau_i} \cdot \frac{\partial \tau_i}{\partial K_i} = 0 \\
& \quad -c - N_i \cdot \left(\frac{\partial C(N_i, K_i)}{\partial K_i} \right) = 0, \text{follows from Proposition 7.2 and (3.37)} \\
& \quad -c + \mu_i^2 \cdot g'(\mu_i) = 0 \Rightarrow \\
\mu_i^* : \qquad & \mu_i^2 \cdot g'(\mu_i)\big|_{\mu_i = \mu_i^*} = c \quad (3.39)
\end{aligned}
$$

Once again, the optimal capacity level induces the conclusion in Proposition 8. In addition, the link profit optimal toll follows the general condition for the profit-maximizing price, i.e., unit elastic demand.

3.3.4. Pareto Efficiency analysis

We consider 2 approaches to extend the single road analysis to a network with L roads. In the first approach, we consider the trade-off between social welfare and the operators' aggregate profit. In the second approach, we consider the trade-off between social welfare and individual operators' profits.

Definition 6. *Aggregate Pareto Efficiency*
For the case of a network consisting of L roads, the set of variables, $\{\mathbf{T}^+, \mathbf{K}^+, \mathbf{N}^+\}$, is said to be strict Pareto Optimal/Efficient, if and only if, there is no set $\{\hat{\mathbf{T}}, \hat{\mathbf{K}}, \hat{\mathbf{N}}\}$ such that $SW(\hat{\mathbf{K}}, \hat{\mathbf{N}}(\cdot)) \geq SW(\mathbf{K}^+, \mathbf{N}^+(\cdot))$ and $\Pi(\hat{\mathbf{T}}, \hat{\mathbf{K}}, \hat{\mathbf{N}}(\cdot)) \geq \Pi(\mathbf{T}^+, \mathbf{K}^+, \mathbf{N}^+(\cdot))$, with at least one of the 2 inequalities being strict.

Under Assumption 8, the corresponding scalarization problem is as follows:

$$\max_{\mathbf{T},\mathbf{K}} \quad (1-\alpha)\cdot SW(\mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) + \alpha \cdot \Pi(\mathbf{T}, \mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) \tag{3.40}$$

Building on the results presented in the previous section, the optimal solution to the scalarization problem is given by:

$$
\begin{aligned}
N^{+}(\alpha): \qquad & P(N) - \sum_{l=1}^{L} \delta_{rl} \cdot \left[g(\mu_l^*) + \frac{c}{\mu_l^*} \right] \\
& + \alpha \cdot NP'(N)|_{N=N^{+}(\alpha)} = 0, r = 1, \ldots, R \qquad (3.41) \\
K_i(\alpha) \quad = \quad & \frac{1}{\mu_i^*} N_i^{+}(\alpha), i = 1, \ldots, L \qquad (3.42) \\
\tau^{+}(r, \alpha) \quad = \quad & \sum_{l=1}^{L} \delta_{rl} \cdot \frac{c}{\mu_l^*} - \alpha \cdot N^{+}(\alpha) P'(N^{+}(\alpha)), r = 1, \ldots, R \qquad (3.43)
\end{aligned}
$$

where μ_i^* is obtained from (3.58). The Pareto Frontier is constructed by letting α vary in $[0, 1]$ and finding all solutions that satisfy the above conditions. While we cannot determine the effect of α on link flows, optimal link tolls, or capacities, we observe that aggregate demand, $N^{+}(\alpha)$, decreases monotonically with α from N_{SW}^* to N_{Π}^*, and that tolls for route r, $\tau^{+}(r, \alpha)$, increase monotonically with α from $\tau_{SW}^*(r)$ to $\tau_{\Pi}^*(r)$. The most important implication is that the trade-off between aggregate profit and social welfare is solely on the aggregate demand or the route toll, which is independent of the network structure. Once the aggregate demand or the route toll is determined, both the government and the union of private firms have the same optimal toll/capacity assignment on links.

While Aggregate Pareto Efficiency provides the benchmark where the firms are cooperative, it is of interest to consider a framework allowing for the characterization of outcomes where operators are self-interested. Feldman (1973) define Pareto Efficiency for a set of decentralized agents, where considerations are made for the benefits of a variety of coalitions. We build on their work to consider a framework where we assume that tolling and capacity decisions arise from bilateral negotiations between each operator and the government. Formally, we apply the following definition of Decentralized Pareto Efficiency:

Definition 7. *Decentralized Pareto Efficiency*
For the case of a network consisting of L roads, the set of variables $\{\mathbf{T}^{\#}, \mathbf{K}^{\#}, \mathbf{N}^{\#}\}$ is said to be Decentralized Pareto Optimal/Efficient if and only if, for every operator $i = 1, 2, ..., L$, fixing other operators' decisions $\{\mathbf{T}^{\#}_{-i}, \mathbf{K}^{\#}_{-i}\}$, there is no set $\{\hat{\tau}_i, \hat{K}_i\}$ such that $SW(\hat{K}_i, \mathbf{K}^{\#}_{-i}, \mathbf{N}(\cdot)) \geq SW(K^{\#}_i, \mathbf{K}^{\#}_{-i}, \mathbf{N}(\cdot))$, and $\pi_i(\hat{\tau}_i, \mathbf{T}^{\#}_{-i}, \hat{K}_i, \mathbf{K}^{\#}_{-i}, \mathbf{N}(\cdot)) \geq \pi_i(\tau^{\#}_i, \mathbf{T}^{\#}_{-i}, K^{\#}_i, \mathbf{K}^{\#}_{-i}, \mathbf{N}(\cdot))$, with at least one of the 2 inequalities being strict.

Decentralized Pareto Efficiency ensures that, for a given set of link tolls and capacities, no operator can unilaterally improve both social welfare and its own profit. That is, given set of tolls and capacities for other $L-1$ roads, $\{\mathbf{T}_{-i}, \mathbf{K}_{-i}\}$, firm$i$ selects $\tau^{\diamond}_i$ and $K^{\diamond}_i$ that are Pareto non-dominated, i.e., no other toll and capacity yields a higher social welfare and individual profit.[6] The set of Pareto non-dominated tolls and capacities for operator i, $\{\tau^{\diamond}_i(\alpha_i, \mathbf{T}_{-i}, \mathbf{K}_{-i}), K^{\diamond}_i(\alpha_i, \mathbf{T}_{-i}, \mathbf{K}_{-i})\}$ can be obtained by the following scalarization problem with $\alpha_i \in [0, 1]$:

$$
\begin{aligned}
\max_{\tau_i, K_i} \quad & h(\tau_i, K_i) \\
\equiv \quad & (1-\alpha_i) \cdot SW(K_i, \mathbf{K}_{-i}, \mathbf{N}(\mathbf{V}(\tau_i, \mathbf{T}_{-i}, K_i, \mathbf{K}_{-i})) \\
& \quad + \alpha_i \cdot \pi_i(\tau_i, K_i, N_i(\mathbf{V}(\tau_i, \mathbf{T}_{-i}, K_i, \mathbf{K}_{-i})) \qquad (3.44)
\end{aligned}
$$

[6]Here, the dominance is defined as the superiority over both individual profit and social welfare, as opposed to superiority over all counter-strategies that are commonly used in the context of Game Theory.

Following Proposition 8, we can show that the solution for firm i requires optimal[7] v/c ratio in link i. However, the firm may not have enough incentive to adjust other links' v/c ratio. Since, in addition to the trade-off on aggregate demand, the private firm also makes a trade-off between its share of the total profit and the welfare loss from the distortion on other links' v/c ratio. Thus, the Pareto non-dominated problem has no simplified general solution as that in the aggregate case. However, we can derive bounds on Pareto non-dominated tolls for individual operators. Prior to stating the conclusion, we present an intermediate result related to the generalized cost of travel.

Lemma 6. *The generalized cost of travel in each link increases with the toll in the link. That is,* $\frac{\partial}{\partial \tau_i}\left(\tau_i + C(N_i, K_i)\right) > 0, i = 1, \ldots, L$.

Proof See 3.5.9.

The proposition essentially shows that, for a link, the toll increase always outweighs the reduction in congestion due to toll increase. Notice that this conclusion may or may not hold in terms of the generalized route travel cost.

Proposition 9. *Pareto non-dominated tolls in a road network,* $\tau_i^\diamond, i = 1, \ldots, L$, *are set so that route tolls,* $\tau^\diamond(s), s = 1, \ldots, R$, *where* $\tau^\diamond(s) \equiv \sum_{l=1}^{L} \delta_{sl} \cdot \tau_l^\diamond$, *are greater than or equal to the total route congestion externality. That is,*

$$\tau^\diamond(s) \quad \geq \quad \sum_{l=1}^{L} \delta_{sl} \cdot N_l \cdot C_{N_l}, \;\; s = 1, \ldots, R$$

Proof See 3.5.11.

That is, Pareto non-dominated tolls always internalize the congestion externalities, i.e.,

[7]The first-order condition for the Pareto non-dominated problem is shown in 3.5.13

each user at least pays the congestion it induced. It also implies that, any competitive behaviors among private firms are not enough to lower the tolls below the social optimal level. In turn, the set of Decentralized Pareto Efficient strategies consists of the set of mutually Pareto non-dominated strategies for all firms. Applying Proposition 9 to each link, we arrive at the following conclusion.

Lemma 7. *At Decentralized Pareto Optimum, social welfare always (weakly)decreases with link toll, i.e.,* $\frac{\partial SW}{\partial \tau_i}|_{\tau_i=\tau_i^{\#}} \leq 0$.

Proof See 3.5.12.

The lemma implies that the government always wants to lower the link toll in a bilateral negotiation with a private firm. Intuitively, one can believe the upper bound of the Decentralized Pareto Optimum is at where the scalarization problem is purely private-driven, which can be reflected by the weights. The weights in the scalarization problems reflect a government's bargaining power in relation to each of the firms. When $\alpha_i = 0, i = 1, \ldots, L$, the set of Pareto non-dominated solutions correspond to the social welfare-maximizing tolls and capacities presented earlier. When $\alpha_i = 1, i = 1, \ldots, L$, each firm has full bargaining power over the government, and the set of Pareto non-dominated solutions correspond to the set of Nash Equilibria. We note that, in general, Nash Equilibrium strategies do not maximize social welfare, nor aggregate profit. Altogether, the following proposition establishing upper bounds on the link tolls:

Proposition 10. *Nash Equilibrium link tolls,* τ_i^e*, are upper bounds of Decentralized Pareto Efficient link tolls, i.e.,* $\tau_i^{\#} \leq \tau_i^e, i = 1, \ldots, L$.

Proof See 3.5.8.

The above result also provides an upper bound for Decentralized Pareto Efficient route tolls, where $\tau^{\#}(s) \leq \tau^{e}(s) = \sum_{l=1}^{L} \delta_{sl} \cdot \tau_{i}^{e},\ s = 1, \ldots, R$.

To conclude, we note that Decentralized Pareto Efficient strategies can be Aggregate Pareto Efficient. Both concepts yield the same v/c ratio, μ^{*}. Hence, a necessary condition is for route tolls to fall within the bounds for the case of Aggregate Pareto Efficiency. Further, as shown in Proposition 9, route tolls in the Decentralized case internalize the route congestion externalities, and thus, it is only necessary to compare the upper bounds, i.e., Decentralized Pareto Efficient strategies can be Aggregate Pareto Efficient when $\tau^{e}(s) \leq \tau_{\Pi}^{*}(s), s = 1, \ldots, R$.

3.3.5. Example for network case

In this section, we present an example to highlight the results presented in the previous section and to further develop insights. In particular, we are interested in benchmarking the effect of network structure/topology on road operators' toll and capacity decisions, and for this, we analyze a simple stage network consisting of M segments, each with L_m links arranged in parallel. The simple stage network is represented in Figure 3.2 below.

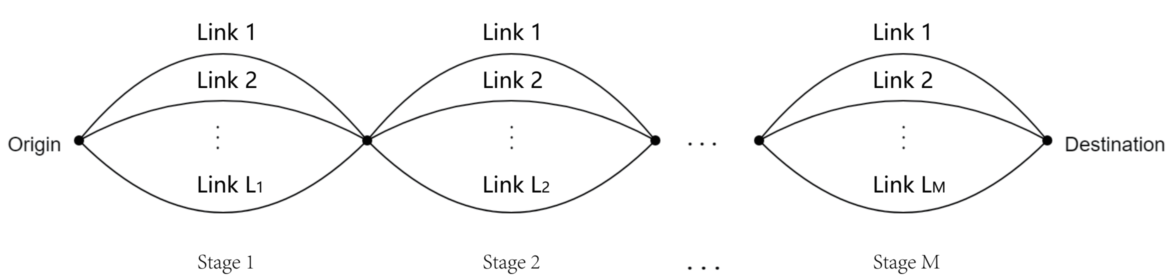


Figure 3.2. Simple stage network

The structure of the simple stage network is such that every route includes exactly one road/link in every stage, and thus, there are $\prod_{m=1}^{M} L_m$ possible routes. The simple stage network is, perhaps, the simplest generalization of serial and parallel networks, which are widely studied in the literature (cf. Mun and Ahn (2008); De Borger and Van Dender (2006); de Palma and Lindsey (2000); Xiao et al. (2007)). They allow for simultaneous consideration of the 2 relevant relationships among links: within each stage, firms compete with each other, thereby splitting the market power and profit; across segments, firms complement/depend on each other, thereby benefiting from each others' capital investments and being adversely affected by their tolling decisions.[8] We begin this section by presenting closed-form expressions for the marginal effects of operators' toll and capacity decisions on link flows and aggregate demand. We then discuss the effect of competition in attenuating double-marginalization, and, in turn, in ensuring that Decentralized Pareto Efficient strategies are Aggregate Pareto Efficient. Finally, we present a numerical example that highlights some of the insights.

3.3.5.1. Closed-form results for simple stage networks

We begin by characterizing the effect of toll and capacity decisions on demand for pure parallel networks, where $R = L$, and the incidence matrix is given by $\mathcal{I}_{L\times L}$. Thus, (3.29) becomes:

$$\mathbf{1}_{L\times L}\mathcal{J}_{\tau}^{N} P'(N) = \mathcal{C}_N \mathcal{J}_{\tau}^{N} + \mathcal{I} \tag{3.45}$$

and observe that:

[8] While not modeled herein, firms across segments share risks.

(1) The square matrix $[\mathbf{1}_{L\times L}P'(N) - \mathcal{C}_N]$ is invertible.[9] Hence, $\mathcal{J}_\tau^N$ is uniquely given by $[\mathbf{1}_{L\times L}P'(N) - \mathcal{C}_N]^{-1}$.

(2) Letting $b = P'(N)\mathrm{tr}\left(\mathcal{C}_N^{-1}\right) - 1$, where $\mathrm{tr}(\cdot)$ is the trace operator, $\mathcal{J}_\tau^N$ is given by

$$\mathcal{J}_\tau^N \quad = \quad \frac{P'(N)}{b} \cdot \mathcal{C}_N^{-1}\mathbf{1}_{L\times L}\mathcal{C}_N^{-1} - \mathcal{C}_N^{-1}$$

with element-wise representation

$$\frac{\partial N_i}{\partial \tau_l} \quad = \quad \begin{cases} \frac{P'(N)}{b\cdot C_{N_i}\cdot C_{N_i}} - \frac{1}{C_{N_i}} \text{ , if } i = l \\ \frac{P'(N)}{b\cdot C_{N_i}\cdot C_{N_l}} \text{ , otherwise} \end{cases} \quad i, l = 1, \dots, L$$

(3) The marginal effect of tolling on the aggregate demand is given by, $\frac{\partial N}{\partial \tau_l} = \frac{1}{b\cdot C_{N_l}}, l = 1, \dots, L$.

(4) Finally, the marginal effects of capacity on demand are given by $\mathcal{J}_K^N = \mathcal{J}_\tau^N\mathcal{C}_K$.

The above observations follow from results in linear algebra. Importantly, we use them as building blocks to analyze (vertical) relationships among operators in simple stage networks. The intuition is to obtain independent inverse demand functions, $P_m(N)$, for each stage, m, and then apply the results for the pure parallel networks. In particular, for each stage m, we have the following L_m conditions:

---

[9] $\mathbf{1}P'(N)$ is negative semi-definite, and $\mathcal{C}_N$ is positive definite. Thus, $\mathbf{1}P'(N) - \mathcal{C}_N$ is negative definite and invertible.

$$C(N_{mi}, K_{mi}) + \tau_{mi} \quad = \quad P_m(N),\ i = 1, \ldots, L_m \tag{3.46}$$

$$\sum_{i=1}^{L_m} N_{mi} \quad = \quad N \tag{3.47}$$

Fixing toll and capacity choices for stages other than m, means that

$$P_m(N) = P(N) - \sum_{q=1,\ldots,M:q\neq m} P_q(N) \tag{3.48}$$

The equation implies that, for fixed tolls and capacities in stages other than m, $P_m(N)$ only depends on total demand N. Thus, each of the stages is an independent pure parallel network. Hence, the applicability of the earlier observations by replacing $P'(N)$ with $P'_m(N)$ for links within segment m. For links in other segments, the marginal effect on demand is transferred through the changes to the total demand. These effects, i.e., the demand system for the simple stage network, are detailed in Proposition 11 below.

Proposition 11. *For a simple stage network with M stages:*

(1) *The derivative of the segment inverse demand function is given by*

$$P'_m(N) \quad = \quad P'(N) - \sum_{q=1,\ldots,M:q\neq m} \frac{1}{\sum_{l=1}^{L_q} \frac{1}{C_{N_{ql}}}}$$

(2) *The marginal effects of tolls and capacities on link flows in the same segment are as follows:*

$$\begin{aligned}
\mathcal{J}_{\tau,m}^{N} &= \frac{P_m'(N)}{b_m} \cdot \mathcal{C}_{N,m}^{-1} \mathbf{1}_{L_m \times L_m} \mathcal{C}_{N,m}^{-1} - \mathcal{C}_{N,m}^{-1} \text{ and} \\
\mathcal{J}_{K,m}^{N} &= \mathcal{J}_{\tau,m}^{N} \mathcal{C}_{K,m}
\end{aligned}$$

where $\mathcal{J}_{\tau,m}^{N}$, $\mathcal{J}_{K,m}^{N}$, b_m, $\mathcal{C}_{N,m}$ *and* $\mathcal{C}_{K,m}$ *are analogous to the elements introduced in the description of pure parallel network.*

(3) *The marginal effects of tolls and capacities on link flows in different segments are as follows:*

$$\begin{aligned}
\frac{\partial N_{mi}}{\partial \tau_{ql}} &= \frac{1}{b_m \cdot C_{N_{ql}}} \cdot \frac{1}{C_{N_{mi}} \cdot \left(\sum_{j=1}^{L_m} \frac{1}{C_{N_{mj}}} \right)}, i = 1, \ldots, L_m, l = 1, \ldots, L_q, m \neq q = 1, \ldots, M \\
\frac{\partial N_{mi}}{\partial K_{ql}} &= \frac{C_{K_{ql}}}{b_m \cdot C_{N_{ql}}} \cdot \frac{1}{C_{N_{mi}} \cdot \left(\sum_{j=1}^{L_m} \frac{1}{C_{N_{mj}}} \right)}, i = 1, \ldots, L_m, l = 1, \ldots, L_q, m \neq q = 1, \ldots, M
\end{aligned}$$

Proof See 3.5.7.

Building on Proposition 11, yields the results summarized below for the marginal effects of tolls and capacity decisions on aggregate and link demands for simple stage networks:

Proposition 12. *Comparative statics for simple stage network*

(1) *Simple stage networks are non-Braess, i.e., increasing toll (capacity) in a link decreases (increases) total throughput. That is,* $\frac{\partial N}{\partial \tau_{ml}} < 0, \frac{\partial N}{\partial K_{ml}} > 0, l = 1, \ldots, L_m, m = 1, \ldots, M$.

(2) *Increasing toll (capacity) in a link decreases (increases) traffic in all links that are in other segments. Mathematically,* $\frac{\partial N_{mi}}{\partial \tau_{ql}} < 0$ *and* $\frac{\partial N_{mi}}{\partial K_{ql}} > 0, i = 1, \ldots, L_m, l = 1, \ldots, L_q, m, q = 1, \ldots, M : m \neq q$.

(3) *Increasing toll (capacity) in a link increases (decreases) traffic in all links within the same segment. Mathematically,* $\frac{\partial N_{mi}}{\partial \tau_{ml}} > 0, \frac{\partial N_{mi}}{\partial K_{ml}} < 0, i, l = 1, \ldots, L_m : i \neq l, m = 1, \ldots, M$.

Proof As each route includes exactly 1 link on each segment, the aggregate demand equals the sum of link flows in a segment, i.e.,

$$\frac{\partial N}{\partial \tau_{mi}} = \sum_{l=1}^{L_m} \frac{\partial N_{ml}}{\partial \tau_{mi}} = \frac{1}{b_m \cdot C_{N_{mi}}}$$

Because $b_m < 0$, $\frac{\partial N}{\partial \tau_{mi}} < 0$, and similarly $\frac{\partial N}{\partial K_{mi}} > 0$. The signs for the other link-level effects is follows from Proposition 11 and the fact that $b_m < 0$. □

In Proposition 12, the first result ensures Assumption 9 holds, and thus, Proposition 9 applies to the simple stage network. Results 2 and 3 confirm the intuition that links on different segments exhibit cooperative relationships, whereas links in the same segment exhibit competitive relationships.

To conclude this section, we use the simple stage network to highlight the relationship between network structure and the link tolls. Intuitively, competition lowers the route tolls. However, the simple stage network also includes complement relationships among links in different stages. The complement relationship leads to the well-known *marginalization* effect (Gaudet and Van Long, 1996), yielding higher tolls than those that would be charged by a single operator. Mun and Ahn (2008) examine the marginalization effect among two serial private links. While in simple stage networks, the marginalization effect interacts with the competition among parallel links. To illustrate the effect of decentralization on route tolls, we consider a special case, where there are infinite number of links, i.e., equal substitutes, in each segment, $L_m = \hat{L} \to \infty, m = 1, \ldots, M$. In this case, the Nash Equilibrium tolls for each link converge to the congestion externality, i.e., $\tau^e_{mi} \to \frac{c}{\mu^*_{mi}}$. Thus, the Nash Equilibrium route tolls converge to the route congestion externality, which is less than the upper bound of Aggregate Optimality. That is, perfect competition eliminates the mark-up, regardless of marginalization effects. In the same spirit, Proposition 13 shows that a sufficiently large $\hat{L}$ guarantees that Decentralized Pareto Efficient tolls are also Aggregate Optimal.

Proposition 13. *For a homogeneous simple stage network with M stages and $\hat{L}$ links in each segment, Decentralized Pareto Efficient link tolls are guaranteed to be Aggregate Optimal, if and only if,*

$$\hat{L} \quad \geq \quad \frac{P'_m(N)}{P'(N)} = 1 - \frac{(M-1)\cdot c}{N \cdot P'(N) \cdot \mu^*}$$

Proof See 3.5.10

Several observations can be made. First, inequality always holds when $M = 1$. Also, $\hat{L}$ increases with M, i.e., increased competition is needed to counter the marginalization effect. Second, as a result of the concavity of $P(N)$, $\hat{L}$ decreases with total demand, N. The interpretation is that the link mark-up stems from the derivative of segment inverse demand $P'_m(N)$, while the aggregate profit optimal mark-up stems from the derivative of total inverse demand $P'(N)$. When total demand increases, the absolute value of $P'_m(N)$

increases slower than $P'(N)$. Thus, it requires fewer links, to offset $P'_m(N)$, and thus keep the total marginalization mark-up below the aggregate profit optimal mark-up.

When the number of competing links is not large enough to control the marginalization effect, Decentralized Pareto Efficient tolling strategies may not be Aggregate Optimal. That is, bilateral/distributed negotiations may lead to an outcome that is not Aggregate Optimal. An implication is that in these cases, it is possible to compensate the operators to improve social welfare or aggregate profit without adversely affecting the other measure. Such improvement by compensation is called *Kaldor-Hicks Improvement* (Stringham, 2001). In the context of toll-road markets, this compensation can be implemented with a subsidy or tax break. This and other issues are discussed further in the following numerical examples.

3.3.5.2. Numerical example for simple stage network

The parameters used to generate the examples are identical to those used in Section 3.2.5[10]. We consider a network with 2 segments, each of which has 2 links, i.e.., $M = 2$ and $\hat{L} = 2$. We assume that all links are homogeneous in terms of the congestion and construction cost functions. As discussed, under the optimal v/c ratio, social welfare and aggregate profit only depend on the route toll. Thus, we plot these two aggregate measures versus the route toll, shown in Figure 3.3. Similar to the single road case, changes to the route tolls decrease either social welfare or aggregate profit.

As far as Decentralized Pareto Efficient tolling strategies on the same network, we recall that Propositions 9 and 10 only provide lower and upper bounds on the tolls. Thus, we find Decentralized Pareto Efficient tolls numerically over a discrete grid. In order to

[10]The construction and congestion function for each link is the same as that for the preceding single road example. Also, the overall inverse demand function is the same.

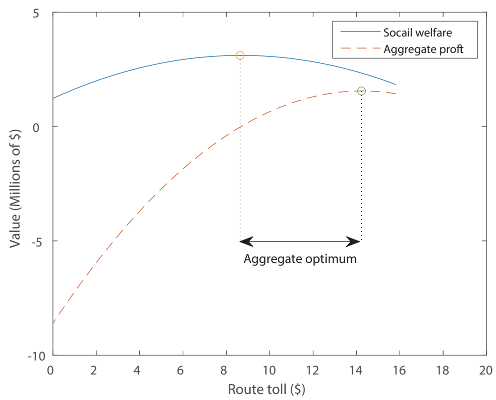


Figure 3.3. Social welfare, $SW(\mathbf{K}, \mathbf{N})$, and aggregate profit, $\Pi(\mathbf{T}, \mathbf{K}, \mathbf{N})$, as a function of route tolls, $\tau(r), r = 1, \ldots, R$

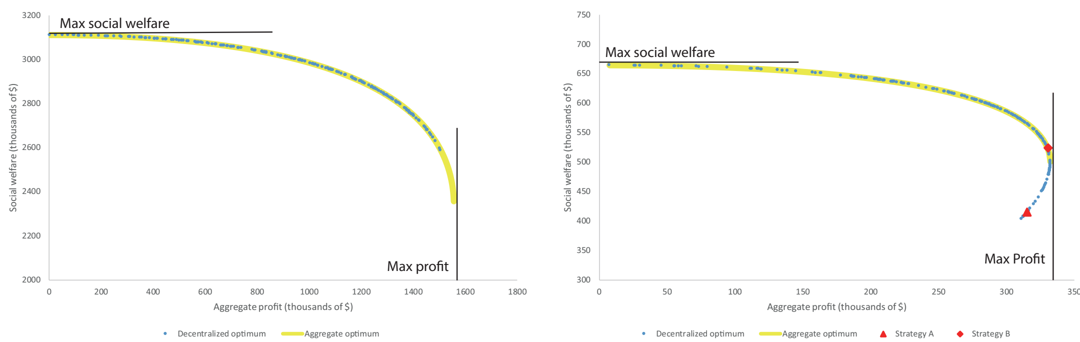
account for the demand effect on marginalization, we use two inverse demand functions with the same slope but different intercepts.[11] Figure 3.4 shows the Decentralized Pareto Efficient tolling decisions. In the two-segment network, the route toll is the sum of tolls in the 2 segments. Hence, the 45-degree downward sloping line corresponds to the bounds derived for the Aggregate Optimal route tolls. These bounds divide the 2-dimensional tolling space into three parts: overcharge area(where the route toll exceeds the upper bound), undercharge area(where the route toll is less than lower bound), and Aggregate Optimal area.

For both cases, all the Decentralized Pareto Efficient tolls are below the 45-degree line crossing the Nash Equilibrium solution, which shows that Nash Equilibrium solutions give the highest route tolls among Decentralized Pareto Efficient tolls. Also, under low demand, Nash Equilibrium tolls, as well as other Decentralized Pareto Efficient strategies lie in the overcharge area. While, under high demand, all Decentralized Pareto Efficient

[11]The inverse demand function for high demand case is, $P(N) = 42 - 0.02N$, same as the preceding examples. While the inverse demand function for low demand case is, $P(N) = 36 - 0.02N$.

(a) Low demand case (b) High demand case

Figure 3.4. Decentralized Pareto Efficient segment tolls



(a) Low demand case (b) High demand case

Figure 3.5. Social welfare v. Aggregate profit

tolls, including the Nash Equilibrium tolls, are Aggregate Optimal. $\hat{L} = 2$ is less than the number of links needed under low demand but is larger than that under high demand. Figure 3.5 provides additional evidence. We observe that in the case of low demand, some Decentralized Pareto Efficient decisions yield lower social welfare gain than the profit optimal decision does. Under high demand, the Nash Equilibrium tolls lead to Aggregate Optimal solution.

In the low demand case, we may apply Kaldor-Hicks Improvements. Consider the Decentralized Pareto Efficient strategy A, marked in Figure 3.5. It lies off the aggregate optimal frontier, under low demand. Another strategy B, marked in the aggregate optimal

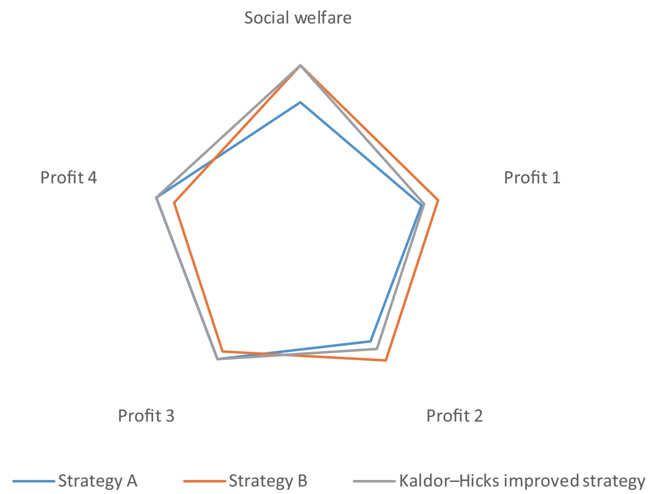


Figure 3.6. Operator's profit and social welfare

frontier, yields higher aggregate profit and social welfare than strategy A does. Figure[12] 3.6 shows that the deviation from decision A to decision B decreases operators' 3 and 4 profit, but increases social welfare and other firms' profit. Thus, such deviation does not amount to a Decentralized Pareto improvement. However, we can compensate operators 3 and 4, by some of the increased profits of 1 and 2, which results in the grey polygon in Figure 3.6. After such adjustment, operators 1 and 2 still have higher profit, while 3 and 4 suffer no profit reduction.

[12]The figure is scaled. See 3.5.15 for details on the values.

3.4. Conclusions

Tolling and capacity decisions in transportation networks have been widely-studied with a number of approaches. At a strategic level, the motivation is to understand the effects of network structure, of decisions by self-interested operators on the performance of such systems, and to shed light on the need for government interventions. Interest in these analyses appears to be increasing as P3s and other agreements to develop or operate (roads within) transportation networks are gaining ground around the world.

In the first part of the paper, we consider tolling and capacity decisions for individual roads. The objective is to introduce notation and assumptions, as well as to review results appearing in the literature (cf. Tan et al. (2010)). Under the assumption that demand/traffic is determined endogenously by the user equilibrium conditions, our model yields closed-form expressions for social welfare and profit-maximizing toll and capacity decisions. In particular, we show that the social welfare-maximizing toll is set to the marginal congestion externality. The profit-maximizing toll includes a mark-up that depends on the elasticity of the demand function. This higher toll means that the profit-maximizing system serves a smaller demand. Both the profit and social welfare maximizing capacity levels are set to keep the road utilization, i.e., the volume-to-capacity ratio, constant. We build on the aforementioned results to characterize Pareto Efficient strategies and provide expressions that describe how tolls, capacities, and demand vary as the weight parameter shifts between profit and social welfare maximization.

In the second part of the paper, we consider 3 cases related to the problem of setting tolls and capacities within a road network serving a single market, i.e., with a single-origin and a single-destination. In addition to the problem of maximizing social welfare, we consider both the aggregate profit maximization problem, where tolls and capacities are set to maximize the sum of profits, and the individual-operator profit maximization problem, under the assumption that decision-makers are self-interested, by rational agents that make decisions independently.

We show that the results and insights obtained for the single road problem can be extended to the social welfare and aggregate profit maximization problems. In particular, the welfare-maximizing route tolls are set to the sum of the marginal congestion externalities on each of the links, and that optimal route tolls for the aggregate profit maximization problem include a mark-up. Moreover, the independence of the optimal v/c ratio for the single road case, shown in the exiting literature, can be applied to any networks with any toll/capacity levels. Again, we provide closed-form expressions for Aggregate Pareto Optimal/Efficient strategies as a function of the weight parameter as well as the Pareto frontier.

Even though the independent optimal v/c ratio also applies in the decentralized case, the optimality conditions, do not lead to closed-form expressions for the tolls. To benchmark optimal tolling strategies for this problem, i.e., understand outcomes that may arise from bilateral negotiations between the government and self-interested private operators, we follow the work of Feldman (1973) and define the concept of Decentralized Pareto Efficiency. We show that efficient route tolls are bounded below by social welfare maximizing route tolls. We also explain that Nash Equilibrium tolls provide an upper bound on (Pareto)efficient tolls. These bounds provide a necessary condition when Decentralized Pareto Efficient strategies are not Aggregate Pareto Efficient, i.e., when there may be (government) interventions that could simultaneously improve social welfare and (aggregate)profit.

To illustrate the above results, we consider the analysis of homogeneous, simple stage networks. As described, such networks are, perhaps, the simplest generalization of pure parallel or serial networks, which are widely used in the literature. Our analysis of this example allows us to characterize how much competition is needed to attenuate (multiple) marginalizations, and importantly when there are opportunities for government interventions, i.e., transfers, to improve social welfare or individual operator profit without adversely affecting other players/stakeholders.

3.5. Proofs and Conditions

3.5.1. First order conditions for Pareto Efficiency of single road

To facilitate evaluation of the optimality conditions, we represent the problem in terms of τ, μ, N instead of the original variables, and use Assumption 7 to write τ as a function of μ and N. Thus, under Assumptions 6 and 7, the solution to SP(α) can be obtained by solving the following optimization problem:

$$\begin{aligned}\max_{\mu,N} \quad & \mathcal{L}(\alpha,\mu,N) \equiv (1-\alpha)\left[\int_{n=0}^{n=N} P(n)dn - N\cdot g(\mu) - c_0 - c\cdot\frac{N}{\mu}\right] \\ & +\alpha\left[N\cdot(P(N) - g(\mu)) - c_0 - c\cdot\frac{N}{\mu}\right]\end{aligned}$$

The first-order conditions are:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\cdot)}{\partial \mu} &= (1-\alpha)\left[-N \cdot g'(\mu) + c \cdot \frac{N}{\mu^2}\right] + \alpha\left[-N \cdot g'(\mu) + c \cdot \frac{N}{\mu^2}\right] = 0 \\
&= N\left[\frac{c}{\mu^2} - g'(\mu)\right] = 0 && (3.49) \\
\frac{\partial \mathcal{L}(\cdot)}{\partial N} &= (1-\alpha)\left[P(N) - g(\mu) - \frac{c}{\mu}\right] + \alpha\left[P(N) + N \cdot P'(N) - g(\mu) - \frac{c}{\mu}\right] = 0 \\
&= P(N) - g(\mu) - \frac{c}{\mu} + \alpha \cdot N \cdot P'(N) = 0 && (3.50)
\end{aligned}
$$

3.5.2. Proof for Proposition 7

(1) The solutions to marginal link flow system, $\mathcal{J}_\tau^N$ and $\mathcal{J}_K^N$, are unique.

(2) The capacity effect on link flow is the product of toll effect and marginal return of capacity, i.e., for any links i and l, $\frac{\partial N_i}{\partial K_l} = \frac{\partial N_i}{\partial \tau_l} C_{K_l}$.

(3) The flow on a link decreases with its toll and increases with its capacity, i.e., for any link i, $\frac{\partial N_i}{\partial \tau_i} < 0$ and $\frac{\partial N_i}{\partial K_i} > 0$.

Proof

(1) For each column in Equation (3.31), we have

$$
\left[\mathcal{X} P'(N) - \Delta \mathcal{C}_N\right] \mathcal{J}_{\tau,i}^N = \Delta_i \tag{3.51}
$$

where $\mathcal{J}_{\tau,i}^N$ and Δ_i are, respectively, the ith columns of $\mathcal{J}_\tau^N$ and Δ. We show the result by contradiction, assuming there there are 2 solutions to $\mathcal{J}_{\tau,i}^N$, y_1 and y_2, such that $y_1 = \Delta^T z_1$, $y_2 = \Delta^T z_2$ and $y_1 \neq y_2$. z_1 and z_2 are corresponding marginal effects on route flows. Plugging y_1 and y_2 into (3.51) and subtracting

$$[\mathcal{X}P'(N) - \Delta\mathcal{C}_N]\,\Delta^T(z_1 - z_2) = \mathbf{0} \Rightarrow$$

$$(z_1 - z_2)^T\,[\mathcal{X}P'(N) - \Delta\mathcal{C}_N]\,\Delta^T(z_1 - z_2) = \mathbf{0} \Rightarrow$$

$$(z_1 - z_2)^T\mathcal{X}P'(N)\Delta^T(z_1 - z_2) - (z_1 - z_2)^T\Delta\mathcal{C}_N\Delta^T(z_1 - z_2) = \mathbf{0} \Rightarrow$$

$$(z_1 - z_2)^T\mathbf{1}_{R\times R}P'(N)(z_1 - z_2) - (z_1 - z_2)^T\Delta\mathcal{C}_N\Delta^T(z_1 - z_2) = \mathbf{0}$$

Because $\mathbf{1}_{R\times R}$ and $P'(N) < 0$, the first term in the last expression is non-positive. $\mathcal{C}_N$ is positive definite, which means the second term is strictly negative. Thus, the left-hand-side of the last expression is strictly negative. This means that $\Delta^T(z_1 - z_2) = y_1 - y_2 \neq \mathbf{0}$, which contradicts the assumption that both y_1 and y_2 solve the system of equations (3.51). An analogous argument applies to the capacity effects, $\mathcal{J}_K^N$.

(2) We observe that $\mathcal{J}_K^N = \mathcal{J}_\tau^N C_K$ is the unique solution of the system of equations (3.32). Componentwise, $\frac{\partial N_i}{\partial K_l} = \frac{\partial N_i}{\partial \tau_l} C_{K_l}$.

(3) Consider

$$\left[\mathcal{J}_{\tau,i}^v\right]^T \mathbf{1}_{R\times R}\mathcal{J}_{\tau,i}^v P'(N) \;=\; \left[\mathcal{J}_{\tau,i}^v\right]^T \Delta\mathcal{C}_N\Delta^T\mathcal{J}_{\tau,i}^v + \left[\mathcal{J}_{\tau,i}^v\right]^T \Delta_i$$

Notice that both $\mathbf{1}_{R\times R}$ and $\Delta\mathcal{C}_N\Delta^T$ are positive semi-definite, and $P'(N) < 0$. Thus,

$$\begin{aligned}\left[\mathcal{J}_{\tau,i}^v\right]^T \mathbf{1}_{R\times R}\mathcal{J}_{\tau,i}^v P'(N) &\leq 0 \text{ and}\\ \left[\mathcal{J}_{\tau,i}^v\right]^T \Delta\mathcal{C}_N\Delta^T\mathcal{J}_{\tau,i}^v &\geq 0\end{aligned}$$

which implies $\left[\mathcal{J}^v_{\tau,i}\right]^T \Delta_i = \sum_{r=1}^R \delta_{ri} \cdot \frac{\partial v_r}{\partial \tau_i} = \frac{\partial N_i}{\partial \tau_i} \leq 0$.

Since $\mathcal{C}_N$ is positive definite, the equality, $\frac{\partial N_i}{\partial \tau_i} = 0$, only holds when $\mathbf{1}_{R\times R}\mathcal{J}^v_{\tau,i} = \mathbf{0}$ and $\Delta^T \mathcal{J}^v_{\tau,i} = \mathbf{0}$. However, in such case the ith column of (3.29) yields $\Delta_i = \mathbf{0}$, i.e., link i is not on any route, which does not hold. Hence, $\frac{\partial N_i}{\partial \tau_i} < 0$. From the previous result and Assumption 6, $\frac{\partial N_i}{\partial K_i} = \frac{\partial N_i}{\partial \tau_i} C_{K_i} > 0$.

□

3.5.3. Proof for Proposition 8

The optimal v/c ratio of an arbitrary link i in an arbitrary road network is independent of:

(1) whether the goal is link profit, aggregate profit or social welfare;

(2) its own tolling choice and other link's tolling/capacity choices.

Proof Since user equilibrium balances generalized travel costs, i.e., congestion cost plus toll, among routes, the equilibrium is preserved as long as the generalized travel cost in each link is fixed. That is, we may change toll and capacity in a link i while preserving the equilibrium, as long as $\tau_i + C(N_i, K_i)$ is fixed. Hence, we are able to define a set $D_i(P_i, N_i)$ of toll-capacity decisions, in which toll-capacity decisions preserve the equilibrium with link flow N_i and generalized link cost P_i. That is, given $(\tau_i, K_i) \in D_i(P_i, N_i)$, we have $\tau_i + C_i(N_i, K_i) = P_i$. Apparently, the set $D_i(P_i, N_i)$ is not singleton if $P_i > 0$ and $N_i > 0$. Thus, we can write the equilibrium-preserving link profit, i.e., subject to $D_i(P_i, N_i)$, as,

$$\pi^p_i(K_i; N_i, P_i) = N_i \tau_i(K_i) - C^c(N_i, K_i) = N_i(P_i - C(N_i, K_i)) - C^c(N_i, K_i) \qquad (3.52)$$

Notice the uni-dimensional function π_i^p is concave since its second order derivative, $-N_i C_{KK_i}$, is negative. The first order condition,

$$\frac{d\pi_i^p}{dK_i} = -N_i C_{K_i} - c = 0 \tag{3.53}$$

which leads to (3.8), determines the optimal solution and μ^*. Similarly, the equilibrium-preserving social welfare function is given by,

$$SW^p(K_i; \boldsymbol{N}, \boldsymbol{K}_{-i}) = \int_0^N P(n)dn - \sum_l^L N_l C(N_l, K_l) - \sum_l^L C^c(K_l) \tag{3.54}$$

where $\boldsymbol{K}_{-i}$ represents the collection of capacity choices other than K_i. Then, the corresponding first order condition is,

$$\frac{dSW^p}{dK_i} = -N_i C_{K_i} - c = 0 \tag{3.55}$$

same as the first order condition for optimal π_i^p. Given that both SW^p and π_i^p have continuous and strictly decreasing derivatives, they have an unique optimal v/c ratio at μ^*, where first order condition is satisfied.

It implies that if the link's v/c ratio is not at μ^*, there is always a better solution within the subset $D_i(P_i, N_i)$. That is, if the collection of decisions $(\boldsymbol{\tau}, \boldsymbol{K})$ leads to an non-optimal v/c ratio, we can construct another collection $(\boldsymbol{\tau'}, \boldsymbol{K'})$ which preserves the equilibrium and increases social welfare, link profit(for the non-optimal link) and aggregate profit(since other links' profit is unaffected).

In conclusion, the optimal capacity for link profit, aggregate profit and the social welfare must yield the optimal v/c ratio in each(for aggregate optimization) or the optimizing link(for individual optimization). $\square$

3.5.4. First order condition for social optimum

The first-order optimality conditions are:

$$\begin{aligned} \frac{\partial SW(\cdot)}{\partial \tau_i} &= \sum_{s=1}^{R} \frac{\partial SW(\cdot)}{\partial v_s} \cdot \frac{\partial v_s}{\partial \tau_i} = 0, i = 1, \ldots, L \\ \frac{\partial SW(\cdot)}{\partial K_i} &= -N_i \cdot \frac{\partial C(N_i, K_i)}{\partial K_i} - c + \sum_{s=1}^{R} \frac{\partial SW(\cdot)}{\partial v_s} \cdot \frac{\partial v_s}{\partial K_i} = 0, i = 1, \ldots, L \end{aligned} \tag{3.56}$$

Equation (3.56) can be rewritten as

$$\begin{aligned} \frac{\partial SW(\cdot)}{\partial K_i} &= -N_i \cdot \frac{\partial C(N_i, K_i)}{\partial K_i} - c + \sum_{s=1}^{R} \frac{\partial SW(\cdot)}{\partial v_s} \cdot \frac{\partial v_s}{\partial \tau_i} \cdot \frac{\partial \tau_i}{\partial K_i} = 0, i = 1, \ldots, L \\ \frac{\partial SW(\cdot)}{\partial K_i} &= -N_i \cdot \frac{\partial C(N_i, K_i)}{\partial K_i} - c + \frac{\partial \tau_i}{\partial K_i} \cdot \sum_{s=1}^{R} \frac{\partial SW(\cdot)}{\partial v_s} \cdot \frac{\partial v_s}{\partial \tau_i} = 0, i = 1, \ldots, L \\ \frac{\partial SW(\cdot)}{\partial K_i} &= -N_i \cdot \frac{\partial C(N_i, K_i)}{\partial K_i} - c, i = 1, \ldots, L \end{aligned} \tag{3.57}$$

Thus, the social welfare maximizing capacity levels are such that the optimal v/c ratios, μ_i^*, satisfy

$$\mu_i^* : \mu_i^2 \cdot g'(\mu_i)\big|_{\mu_i = \mu_i^*} = c, i = 1, \ldots, L \tag{3.58}$$

Recalling from the single road analysis that the optimal toll and capacity follow from the first-order condition satisfied by the social welfare maximizing demand level, we construct a set of tolls by considering the case of social welfare maximizing route flow levels, i.e., $\frac{\partial SW(\cdot)}{\partial v_s} = 0, s = 1, \ldots, R$. In this case,

$$\begin{aligned}\frac{\partial SW(\cdot)}{\partial v_s} &= P(N) \cdot \frac{\partial N}{\partial v_s} - \sum_{l=1}^{L} \left[C(N_l, K_l) + N_l \cdot \frac{\partial C(N_l, K_l)}{\partial N_l} \right] \cdot \frac{\partial N_l}{\partial v_s} = 0, s = 1, \ldots, R \\ & P(N) - \sum_{l=1}^{L} [g(\mu_l) + \mu_l \cdot g'(\mu_l)] \cdot \delta_{sl} = 0, s = 1, \ldots, R\end{aligned}$$

Note that the above derivation relies on $\frac{\partial N}{\partial v_r} = 1$, which follows from $N = \sum_{r=1}^{R} v_r$, and on $\frac{\partial N_l}{\partial v_r} = \delta_{rl}$, which follows from $N_l = \sum_{r=1}^{R} \delta_{rl} v_r$.

3.5.5. First order condition for aggregate profit optimum

Under Assumption 8, the optimality conditions for the aggregate profit maximization problem are, for each firm i:

$$\begin{aligned}&\frac{\partial \Pi(\mathbf{T}, \mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K})))}{\partial \tau_i} = \frac{\partial}{\partial \tau_i} \left(SW(\mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) + NP(N) - \int_0^N P(n)dn \right) \\ &= \sum_{s=1}^{R} \frac{\partial}{\partial v_s} \left(SW(\mathbf{K}, \mathbf{N}(\mathbf{V}(\mathbf{T}, \mathbf{K}))) + NP(N) - \int_0^N P(n)dn \right) \cdot \frac{\partial v_s}{\partial \tau_i} = 0 \qquad (3.59)\end{aligned}$$

$$\frac{\partial \Pi(\mathbf{T},\mathbf{K},\mathbf{N}(\mathbf{V}(\mathbf{T},\mathbf{K})))}{\partial K_i} = \frac{\partial}{\partial K_i}\left(SW(\mathbf{K},\mathbf{N}(\mathbf{V}(\mathbf{T},\mathbf{K}))) + NP(N) - \int_{n=0}^{n=N} P(n)dn\right)$$
$$= -N_i \cdot \frac{\partial C(N_i,K_i)}{\partial K_i} - c + \sum_{s=1}^{R} \frac{\partial}{\partial v_s}\left(SW(\mathbf{K},\mathbf{N}(\mathbf{V}(\mathbf{T},\mathbf{K}))) + NP(N) - \int_0^N P(n)dn\right) \cdot \frac{\partial v_s}{\partial K_i}$$
$$= \sum_{s=1}^{R} \frac{\partial}{\partial v_s}\left(SW(\mathbf{K},\mathbf{N}(\mathbf{V}(\mathbf{T},\mathbf{K}))) + NP(N) - \int_0^N P(n)dn\right) \cdot \frac{\partial v_s}{\partial \tau_i} \cdot \frac{\partial \tau_i}{\partial K_i}$$
$$-N_i \cdot \frac{\partial C(N_i,K_i)}{\partial K_i} - c = 0 \tag{3.60}$$

From condition (3.60), we have that the aggregate profit maximizing capacity levels are such that the v/c ratios, μ_i^*, satisfy $\mu_i^* : \mu_i^2 g'(\mu_i)|_{\mu_i=\mu_i^*} = c, i = 1, \ldots, L$, which is identical to condition (3.58) for social welfare maximization. To construct a solution, we consider the case of profit maximizing route flows, i.e.,

$$\begin{aligned}
\frac{\partial}{\partial v_r}\Pi\left(\mathbf{K}^*\left(\mathbf{N}(\mathbf{V})\right),\mathbf{N}(\mathbf{V})\right) &= \frac{\partial}{\partial v_r}\Bigg(SW\left(\mathbf{K}^*\left(\mathbf{N}(\mathbf{V})\right),\mathbf{N}(\mathbf{V})\right) \\
&\qquad + NP(N) - \int_{n=0}^{n=N} P(n)dn\Bigg) = 0, r = 1,\ldots,R \\
&= P(N) - \sum_{l=1}^{L} \delta_{rl} \cdot \left[g(\mu_l^*) + \frac{c}{\mu_l^*}\right] + \left[P(N) + NP'(N)\right] \cdot \frac{\partial N}{\partial v_r} \\
&\qquad - P(N) \cdot \frac{\partial N}{\partial v_r} = 0, r = 1,\ldots,R \\
&= P(N) - \sum_{l=1}^{L} \delta_{rl} \cdot \left[g(\mu_l^*) + \frac{c}{\mu_l^*}\right] \\
&\qquad + NP'(N)|_{N=N_\Pi^*} = 0, r = 1,\ldots,R \\
&= \sum_{l=1}^{L} \delta_{rl} \cdot \left[\tau_l(\mathbf{N},\mathbf{K}) - \frac{c}{\mu_l^*}\right] + NP'(N)\Bigg|_{N=N_\Pi^*} = 0, r = 1,\ldots,R
\end{aligned} \tag{3.61}$$

3.5.6. Proof for Proposition 5

There exists a solution to the linear system of equations given by $\Delta\mathbf{x} = \mathbf{1}_{R\times 1}$.

Proof Letting the components of $\mathbf{x}$ be such that:

$$x_l = \begin{cases} 1 & l \text{ connects the origin to another node in the network.} \\ 0 & \text{otherwise.} \end{cases}$$

Recall that because the graph is acyclic, each route has exactly one link that departs from the origin. This means that the product of the rth row of Δ, $\Delta(r,:)$, and $\mathbf{x}$, $\Delta(r,:)\mathbf{x} = 1$, and the overall result follows. $\square$

3.5.7. Proof for Proposition 11

(1) The derivative of the segment inverse demand function is given by

$$P'_m(N) = P'(N) - \sum_{q=1,\ldots,M:q\neq m} \frac{1}{\sum_{l=1}^{L_q} \frac{1}{C_{N_{ql}}}}$$

(2) The marginal effects of tolls and capacities on link flows in the same segment are as follows:

$$\mathcal{J}^N_{\tau,m} = \frac{P'_m(N)}{b_m} \cdot \mathcal{C}^{-1}_{N,m}\mathbf{1}_{L_m\times L_m}\mathcal{C}^{-1}_{N,m} - \mathcal{C}^{-1}_{N,m} \text{ and}$$

$$\mathcal{J}^N_{K,m} = \mathcal{J}^N_{\tau,m}\mathcal{C}_{K,m}$$

where $\mathcal{J}^N_{\tau,m}$, $\mathcal{J}^N_{K,m}$, b_m, $\mathcal{C}_{N,m}$ and $\mathcal{C}_{K,m}$ are analogous to the elements introduced in the description of pure parallel network.

(3) The marginal effects of tolls and capacities on link flows in different segments are as follows:

$$\begin{aligned}
\frac{\partial N_{mi}}{\partial \tau_{ql}} &= \frac{1}{b_m \cdot C_{N_{ql}}} \cdot \frac{1}{C_{N_{mi}} \cdot \left(\sum_{j=1}^{L_m} \frac{1}{C_{N_{mj}}}\right)}, i = 1, \ldots, L_m, l = 1, \ldots, L_q, m \neq q = 1, \ldots, M \\
\frac{\partial N_{mi}}{\partial K_{ql}} &= \frac{C_{K_{ql}}}{b_m \cdot C_{N_{ql}}} \cdot \frac{1}{C_{N_{mi}} \cdot \left(\sum_{j=1}^{L_m} \frac{1}{C_{N_{mj}}}\right)}, i = 1, \ldots, L_m, l = 1, \ldots, L_q, m \neq q = 1, \ldots, M
\end{aligned}$$

Proof Recall that a link's congestion cost is defined as a function of its v/c ratio. Thus, the v/c ratio is given by the inverse of the congestion cost function, i.e., $\mu_i(g) = g^{-1}(\mu_i), i = 1, \ldots, L$.

(1) Thus, for each stage m and adapting the notation for the simple stage network, (3.46) and (3.47) can be written as

$$\begin{aligned}
g(\mu_{mi}) + \tau_{mi} &= P_m(N),\ i = 1, \ldots, L_m \\
\sum_{i=1}^{L_m} \mu_{mi}(g) \cdot K_{mi} &= N
\end{aligned}$$

Taking the derivative with respect to N, we have

$$\begin{aligned}
\frac{\partial g(\mu_{mi})}{\partial N} &= P'_m(N),\ i = 1, \ldots, L_m \Rightarrow \\
\frac{\partial g(\mu_{mi})}{\partial N} &= \frac{\partial g(\mu_{mj})}{\partial N},\ i \neq j = 1, \ldots, L_m
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{L_m} \frac{\partial \mu_{mi}(g)}{\partial g} \cdot \frac{\partial g(\mu_{mi})}{\partial N} \cdot K_{mi} &= 1 \\
\frac{\partial g(\mu_{mi})}{\partial N} \cdot \sum_{l=1}^{L_m} \frac{\partial \mu_{ml}(g)}{\partial g} \cdot K_{ml} &= 1, i = 1, \ldots, L_m
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial g(\mu_{mi})}{\partial N} &= \frac{1}{\sum_{l=1}^{L_m} \frac{\partial \mu_{ml}(g)}{\partial g} \cdot K_{ml}}, i = 1, \ldots, L_m \\
&= \frac{1}{\sum_{l=1}^{L_m} \frac{\partial g^{-1}(\mu_{ml})}{\partial g} \cdot K_{ml}}, i = 1, \ldots, L_m \\
&= \frac{1}{\sum_{l=1}^{L_m} \frac{1}{\frac{\partial g(\mu_{ml})}{\partial \mu_{ml}}} \cdot K_{ml}}, i = 1, \ldots, L_m \text{ follows from the Inverse Function Theorem} \\
&= \frac{1}{\sum_{l=1}^{L_m} \frac{1}{C_{N_{ml}}}}, i = 1, \ldots, L_m \qquad (3.62)
\end{aligned}$$

Along with (3.48)

$$P'_m(N) = P'(N) - \sum_{q=1,\ldots,M:q\neq m} \frac{1}{\sum_{l=1}^{L_q} \frac{1}{C_{N_{ql}}}}$$

(2) These results are extensions of the observations that apply to the pure parallel network.

(3) From (3.62), the marginal effect on traffic volume is

$$
\begin{aligned}
\frac{\partial N_{mi}}{\partial N} &= \frac{\partial}{\partial N}\left(\mu_{mi}(g)\cdot K_{mi}\right) \\
&= K_{mi}\cdot\frac{\partial \mu_{mi}(g)}{\partial g}\cdot\frac{\partial g(\mu_{mi})}{\partial N} \\
&= \frac{K_{mi}}{\frac{\partial g(\mu_{mi})}{\partial \mu_{mi}}}\frac{1}{\sum_{l=1}^{L_m}\frac{1}{C_{N_{ml}}}} = \frac{1}{C_{N_{mi}}\cdot\sum_{l=1}^{L_m}\frac{1}{C_{N_{ml}}}}
\end{aligned}
$$

For different segments, the tolling effect is captured by changes in the aggregate demand N. Thus,

$$
\begin{aligned}
\frac{\partial N_{mi}}{\partial \tau_{ql}} &= \frac{\partial N}{\partial \tau_{ql}}\cdot\frac{\partial N_{mi}}{\partial N} \\
&= \frac{1}{b_m\cdot C_{N_{mi}}}\cdot\frac{1}{C_{N_{mi}}\cdot\sum_{l=1}^{L_m}\frac{1}{C_{N_{ml}}}}
\end{aligned}
$$

The capacity effect is analogous.

$\square$

3.5.8. Proof for lemma Proposition 10

Nash Equilibrium link tolls are upper bounds on Decentralized Pareto Efficient link tolls, i.e., $\tau_i^{\#} \leq \tau_i^e, i = 1,\ldots,L$.

Proof First order condition of the scalarization problem implies that, $(1-\alpha)\frac{\partial SW}{\partial \tau_i} + \alpha\frac{\partial \pi_i}{\partial \tau_i} = 0$. From lemma 7, the partials of profit, $\frac{\partial \pi_i}{\partial \tau_i}$ has to be non-negative. That is, $N_i + \frac{\partial N_i}{\partial \tau_i}\cdot\tau_i\Big|_{\tau_i=\tau_i^{\#}} \geq 0,\ i = 1,\ldots,L$ with equality holding in the case of Nash Equilbria when $\alpha_i = 1,\ i = 1, \ldots, L$. Mathematically,

$$N_i^{\#} + \frac{\partial N_i^{\#}}{\partial \tau_i^{\#}} \cdot \tau_i^{\#} \geq N_i^e + \frac{\partial N_i^e}{\partial \tau_i^e} \cdot \tau_i^e = 0,\ i = 1, \ldots, L \tag{3.63}$$

where τ_i^e and N_i^e are, respectively, Nash Equilibrium tolls and demands for road i. We assume there is an i where $\tau_i^{\#} > \tau_i^e$, and proceed by contradiction. From Proposition 7.3, we note that this assumption implies that $N_i^{\#} < N_i^e$ and

$$N_i^{\#} + \frac{\partial N_i^{\#}}{\partial \tau_i^{\#}} \cdot \tau_i^{\#} \quad < \quad N_i^e + \frac{\partial N_i^e}{\partial \tau_i^e} \cdot \tau_i^e$$

which is inconsistent with (3.63). Thus, we conclude $\tau_i^{\#} \leq \tau_i^e, i = 1, \ldots, L$. $\square$

3.5.9. Proof for Lemma 6

The generalized cost of travel in each link increases with the toll in the link. That is, $\frac{\partial}{\partial \tau_i}\left(\tau_i + C(N_i, K_i)\right) > 0, i = 1, \ldots, L$.

Proof We construct a new diagonal matrix, $\tilde{\mathcal{C}}_N$, where

$$\tilde{\mathcal{C}}_{N_j} \quad = \quad \begin{cases} \mathcal{C}_{N_i} + \frac{1}{\frac{\partial N_i}{\partial \tau_i}} & \text{if } i = j \\ \mathcal{C}_{N_j} & \text{otherwise} \end{cases}$$

That is, we only update the ith element from $\mathcal{C}_N$. Building on (3.52), we have

$$\left[\mathcal{J}_\tau^v(i)\right]^T \Delta \tilde{\mathcal{C}}_N \Delta^T \mathcal{J}_\tau^v(i) P'(N) \quad = \quad \left[\mathcal{J}_\tau^v(i)\right]^T \Delta \tilde{\mathcal{C}}_N \Delta^T \mathcal{J}_\tau^v(i) + \left[\mathcal{J}_\tau^v(i)\right]^T \Delta(i)$$

Assuming $\tilde{C}_{N_i} \geq 0$, means $\tilde{\mathcal{C}}_N$ is positive semi-definite, and thus, $[\mathcal{J}_\tau^v(i)]^T \Delta\tilde{\mathcal{C}}_N \Delta^T \mathcal{J}_\tau^v(i) \geq 0$. However, the left hand side, $[\mathcal{J}_\tau^v(i)]^T \Delta\tilde{\mathcal{C}}_N \Delta^T \mathcal{J}_\tau^v(i) P'(N) < 0$, leading to contradiction.

Thus, $\tilde{\mathcal{C}}_{N_i} = C_{N_i} + \frac{1}{\frac{\partial N_i}{\partial \tau_i}} = \frac{C_{N_i}\frac{\partial N_i}{\partial \tau_i}+1}{\frac{\partial N_i}{\partial \tau_i}} < 0$. As $\frac{\partial N_i}{\partial \tau_i} < 0$, we have $C_{N_i}\frac{\partial N_i}{\partial \tau_i} + 1 = \frac{\partial}{\partial \tau_i}\left(\tau_i + C(N_i, K_i)\right) > 0$. $\square$

3.5.10. Proof for Proposition 13

For a homogeneous simple stage network with M stages and $\hat{L}$ links in each segment, Decentralized Pareto Efficient link tolls are guaranteed to be Aggregate Optimal, if and only if,

$$\hat{L} \geq \frac{P_m'(N)}{P'(N)} = 1 - \frac{(M-1)\cdot c}{N \cdot P'(N) \cdot \mu^*}$$

Proof According to Proposition 10, operator i sets tolls bounded above by τ_{mi}^e. In homogeneous simple stage networks, average congestion externalities, tolls and traffic are equalized (on all used links). Hence, $b_m = \frac{\mu^* \cdot N \cdot P_m'(N)}{c} - 1$, and $N = \sum_{i=1}^{L_m} N_{mi} = \hat{L} \cdot N_{mi}$. We have

$$\begin{aligned}
\tau_{mi}^e &= \frac{c}{\mu^*} \cdot \left(1 + \frac{P_m'(N)}{b_m \cdot C_{N_{mi}} - P_m'(N)}\right) \\
&= \frac{c}{\mu^*} \cdot \left(1 + \frac{P_m'(N)}{\frac{b_m \cdot c}{N_{mi} \cdot \mu^*} - P_m'(N)}\right) \\
&= \frac{c}{\mu^*} \cdot \left(1 + \frac{N_{mi} \cdot P_m'(N)}{(\hat{L}-1) \cdot N_i \cdot P_m'(N) - \frac{c}{\mu^*}}\right)
\end{aligned}$$

Along with that $P'_m(N) = P'(N) - \frac{(M-1)\cdot c}{N\cdot\mu^*}$, and the symmetric structure of segments, we just need to evaluate the inequality

$$
\begin{aligned}
\tau^e(r) \;\; = \;\; & M\cdot\tau_i^e \le \tau_\Pi^*(r) \Rightarrow \\
& \frac{c}{\mu^*}\frac{M\cdot N_i\cdot P'_m(N)}{(\hat{L}-1)\cdot N_i\cdot P'_m(N) - \frac{c}{\mu^*}} + N\cdot P'(N) \le 0 \Rightarrow \\
& M\cdot N_i\cdot P'_m(N)\cdot\frac{c}{\mu^*} + (\hat{L}-1)\cdot N_i\cdot P'_m(N)\cdot N\cdot P'(N) - N\cdot P'(N)\cdot\frac{c}{\mu^*} \ge 0 \Rightarrow \\
& M\cdot N_i\cdot P'_m(N)\cdot\frac{c}{\mu^*} - N_i\cdot P'_m(N)\cdot N\cdot P'(N) \\
& +\hat{L}\cdot N_i\cdot P'_m(N)\cdot N\cdot P'(N) - N\cdot P'(N)\cdot\frac{c}{\mu^*} \ge 0 \Rightarrow \\
& N_i\cdot P'_m(N)\cdot\left(M\cdot\frac{c}{\mu^*} - N\cdot P'(N)\right) + N\cdot P'(N)\cdot\left(N\cdot P'_m(N) - \frac{c}{\mu^*}\right) \ge 0 \Rightarrow \\
& (N\cdot P'(N) - N_i\cdot P'_m(N))\cdot\left(N\cdot P'(N) - M\frac{c}{\mu^*}\right) \ge 0 \Rightarrow \\
& N\cdot P'(N) - N_i\cdot P'm(N) \le 0 \Rightarrow \\
& \hat{L} = \frac{N}{N_i} \ge \frac{P'_m(N)}{P'(N)} = 1 - \frac{(M-1)\cdot c}{N\cdot P'(N)\cdot\mu^*}
\end{aligned}
$$

□

3.5.11. Proof for Proposition 9

Pareto non-dominated tolls in a road network, $\tau_i^\diamond, i = 1,\ldots,L$, are set so that route tolls, $\tau^\diamond(s), s = 1,\ldots,R$, where $\tau^\diamond(s) \equiv \sum_{l=1}^{L}\delta_{sl}\cdot\tau_l^\diamond$, are greater than or equal to the total route congestion externality. That is,

$$\tau^{\diamond}(s) \;\; \geq \;\; \sum_{l=1}^{L} \delta_{sl} \cdot N_l \cdot C_{N_l} = \sum_{l=1}^{L} \delta_{sl} \cdot \frac{c}{\mu_l^*}, \;\; s = 1, \ldots, R$$

Proof First, we observe that $\tau_i < N_i \cdot C_{N_i} = \frac{c}{\mu_i^*}$ implies

$$\begin{aligned} N_i + \frac{\partial N_i}{\partial \tau_i}\tau_i &= \frac{N_i C_{N_i} + C_{N_i}\frac{\partial N_i}{\partial \tau_i}\tau_i}{C_{N_i}} \\ &> \frac{\tau_i + C_{N_i}\frac{\partial N_i}{\partial \tau_i}\tau_i}{C_{N_i}} \\ &= \frac{\tau_i(1 + \frac{\partial N_i}{\partial \tau_i}C_{N_i})}{C_{N_i}} > 0, \text{ follows from Lemma 6} \end{aligned}$$

We proceed by contradiction to prove the result. Assume there is a route, r, where $\tau^{\diamond}(r) < \sum_{l=1}^{L} \delta_{rl} \cdot \frac{c}{\mu_l^*}$. This means that there exists a link i along route r where $\tau_i^{\diamond} < \frac{c}{\mu_i^*}$. We have that

$$\frac{\partial h(\cdot)}{\partial \tau_i} \;\; = \;\; (1 - \alpha_i) \cdot \left(\tau(r) - \sum_{l=1}^{L} \delta_{sl} \cdot \frac{c}{\mu_l^*} \right) \cdot \frac{\partial N}{\partial \tau_i} + \alpha_i \cdot \left(N_i + \tau_i \cdot \frac{\partial N_i}{\partial \tau_i} \right)$$

When evaluated at $\tau(r) = \tau^{\diamond}(r)$, the first term in the above expression is (strictly) positive. $\frac{\partial N}{\partial \tau_i} \leq 0$ follows from Assumption 9, which excludes Braess networks. Thus, for $\frac{\partial h(\cdot)}{\partial \tau_i} = 0$, the second term, evaluated at $\tau_i^{\diamond}$, would have to be negative. The earlier observation, however, is that $\tau_i^{\diamond} < \frac{c}{\mu_i^*}$ implies $\left(N_i + \tau_i \cdot \frac{\partial N_i}{\partial \tau_i} \right)\Big|_{\tau_i = \tau_i^{\diamond}} > 0$, which is inconsistent. Thus, we conclude $\tau^{\diamond}(r) \geq \sum_{l=1}^{L} \delta_{rl} \cdot \frac{c}{\mu_l^*}, r = 1, \ldots, R$. □

3.5.12. Proof for Lemma 7

At decentralized optimum, social welfare always (weakly)decreases with link toll, i.e., $\frac{\partial SW}{\partial \tau_i}|_{\tau_i=\tau_i^{\#}} \leq 0$.

Proof Take partials of the social welfare function,

$$\begin{aligned}
\frac{\partial SW}{\partial \tau_i} &= P(N)\frac{\partial N}{\partial \tau_i} - \sum_{l=1}^{L}\frac{\partial N_l}{\partial \tau_i}(N_l \cdot C_{N_l} + C(N_l, K_l)) \\
&= P(N)\sum_{r=1}^{R}\frac{\partial v_r}{\partial \tau_i} - \sum_{l=1}^{L}(\sum_{r=1}^{R}\delta_{rl}\cdot\frac{\partial v_r}{\partial \tau_i})(N_l \cdot C_{N_l} + C(N_l, K_l)) \\
&= \sum_{r=1}^{R}\frac{\partial v_r}{\partial \tau_i}(P(N) - \sum_{l=1}^{L}\delta_{rl}\cdot C(N_l, K_l) - \sum_{l=1}^{L}\delta_{rl}\cdot N_l \cdot C_{N_l}) \\
&= \sum_{r=1}^{R}\frac{\partial v_r}{\partial \tau_i}(\tau(r) - \sum_{l=1}^{L}\delta_{rl}\cdot N_l \cdot C_{N_l})
\end{aligned}$$

At decentralized optimum, each link has the same level of v/c ratio. Given the homogeneity assumption, all routes have the same route toll and number of links. Thus,

$$\frac{\partial SW}{\partial \tau_i} = \sum_{r=1}^{R}\frac{\partial v_r}{\partial \tau_i}(\tau^{\#}(s) - M\frac{c}{\mu^*}) = \frac{\partial N}{\partial \tau_i}(\tau^{\#}(s) - M\frac{c}{\mu^*}) \leq 0 \tag{3.64}$$

The last line comes from Proposition 9 and the non-Braess assumption. □

3.5.13. First order condition for Pareto non-dominated problem

$$
\begin{aligned}
\frac{\partial h(\cdot)}{\partial \tau_i} &= (1-\alpha_i)\cdot\frac{\partial SW(\cdot)}{\partial \tau_i} + \alpha_i \cdot \frac{\partial \pi_i(\cdot)}{\partial \tau_i} = 0 && (3.65)\\
& \quad (1-\alpha_i)\cdot\sum_{s=1}^{R}\frac{\partial SW(\cdot)}{\partial v_s}\cdot\frac{\partial v_s}{\partial \tau_i} + \alpha_i\cdot\left(N_i + \tau_i\cdot\frac{\partial N_i}{\partial \tau_i}\right) = 0 \\
\frac{\partial h(\cdot)}{\partial K_i} &= (1-\alpha_i)\cdot\frac{\partial SW(\cdot)}{\partial K_i} + \alpha_i \cdot \frac{\partial \pi_i(\cdot)}{\partial K_i} = 0 && (3.66)\\
& \quad (1-\alpha_i)\cdot\left(-N_i\cdot\frac{\partial C(N_i,K_i)}{\partial K_i} - c + \sum_{s=1}^{R}\frac{\partial SW(\cdot)}{\partial v_s}\cdot\frac{\partial v_s}{\partial K_i}\right) + \\
& \quad \alpha_i\cdot\left(-c + \tau_i\cdot\frac{\partial N_i}{\partial K_i}\right) = 0 \\
& \quad (1-\alpha_i)\cdot\left(\mu_i^2\cdot g'(\mu_i) - c + \frac{\partial C(N_i,K_i)}{\partial K_i}\cdot\sum_{s=1}^{R}\frac{\partial SW(\cdot)}{\partial v_s}\cdot\frac{\partial v_s}{\partial \tau_i}\right) \\
& \quad +\alpha_i\cdot\left(-c + \frac{\partial C(N_i,K_i)}{\partial K_i}\cdot\tau_i\cdot\frac{\partial N_i}{\partial \tau_i}\right) = 0 \\
& \quad (1-\alpha_i)\cdot\left(\mu_i^2\cdot g'(\mu_i) - c + \frac{\partial C(N_i,K_i)}{\partial K_i}\cdot\sum_{s=1}^{R}\frac{\partial SW(\cdot)}{\partial v_s}\cdot\frac{\partial v_s}{\partial \tau_i}\right) \\
& \quad +\alpha_i\cdot\left(-c - N_i\cdot\frac{\partial C(N_i,K_i)}{\partial K_i} + \frac{\partial C(N_i,K_i)}{\partial K_i}\cdot\left(N_i + \tau_i\cdot\frac{\partial N_i}{\partial \tau_i}\right)\right) = 0 \\
& \quad \mu_i^2\cdot g'(\mu_i) - c + \frac{\partial C(N_i,K_i)}{\partial K_i} && (3.67)\\
& \quad \cdot\left[(1-\alpha_i)\cdot\sum_{s=1}^{R}\frac{\partial SW(\cdot)}{\partial v_s}\cdot\frac{\partial v_s}{\partial \tau_i} + \alpha_i\cdot\left(N_i + \tau_i\cdot\frac{\partial N_i}{\partial \tau_i}\right)\right] = 0 \\
\mu_i^* : & \quad \mu_i^2\cdot g'(\mu_i)\big|_{\mu_i=\mu_i^*} = c && (3.68)
\end{aligned}
$$

firm i set capacity at the optimal v/c ratios, i.e., condition (3.58), in all mutually Pareto non-dominated strategies.

3.5.14. Optimality of solutions to the first-order conditions in the single road and aggregate cases

We observe that when $N > 0$, there is at most one solution, $\mu^+(\alpha), N^+(\alpha)$, to the first-order conditions (3.49) and (3.50). $P(0) > g(\mu^*) + \frac{c}{\mu^*}$ ensures $\frac{\partial \mathcal{L}(\cdot)}{\partial N}\Big|_{N=0} > 0$, and that the solution to SP(α) is such that $N > 0$. To show that the unique stationary point of $\mathcal{L}(\cdot)$, $\mu^+(\alpha), N^+(\alpha)$ is a (global) maximum, we consider the Hessian matrix:

$$H(\mathcal{L}) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}(\cdot)}{\partial \mu^2} & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial \mu \partial N} \\ \frac{\partial^2 \mathcal{L}(\cdot)}{\partial N \partial \mu} & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial N^2} \end{bmatrix} = \begin{bmatrix} -N \cdot g''(\mu) - \frac{2 \cdot c \cdot N}{\mu^3} & -g'(\mu) + \frac{c}{\mu^2} \\ -g'(\mu) + \frac{c}{\mu^2} & (1+\alpha)P'(N) + \alpha \cdot N \cdot P''(N) \end{bmatrix}$$

We notice that $\frac{\partial^2 \mathcal{L}(\cdot)}{\partial \mu^2} < 0$, that $\frac{\partial^2 \mathcal{L}(\cdot)}{\partial N^2} < 0$, and that when evaluated at $\mu = \mu^*$, $\frac{\partial^2 \mathcal{L}(\cdot)}{\partial \mu \partial N} = 0$. Thus, $H(\mathcal{L})$ is negative definite, and we conclude that the stationary point is a local maximum. As noted earlier, the boundary point $N = 0$ is suboptimal, which means that $\mu^+(\alpha), N^+(\alpha)$ is a solution to the optimization problem. We do not have a closed-form expression for the disaggregate case, so we cannot verify local concavity.

3.5.15. Social welfare and profits for the numerical example in Section 3.3.5.2

The table below shows the comparison among strategy A, strategy B and the Kaldor-Hicks improvement.

Value in thousands of $	Social welfare	Profit 1	Profit 2	Profit 3	Profit 4
Strategy A	1036.80	739.82	700.15	828.04	890.06
Strategy B	1310.15	844.37	844.37	776.26	776.26
Kaldor–Hicks improved strategy	1310.15	761.58	761.58	828.04	890.06

CHAPTER 4

Conclusion

This dissertation analyzes the competition of private firms in the highway franchise from two perspectives: one is the private firms competing for the single road franchise, another is the oligopolistic competition from multiple road owners in a road network. We show that both of these two types of competition have significant impacts on the public-private partnership for the transportation infrastructure franchise. In this chapter, we summarize the results for each of the two topics and present high-level conclusions of the dissertation and future research directions.

4.1. Competition for the franchise

The government always benefits from the competition for the franchising bidding. However, we show in the dissertation that the increase of the number of bidders does not directly affect the toll/capacity decision making under the optimal mechanism design. In fact, under the Bayesian setting, the government does not have to change any rules of the optimal mechanism for the different number of bidders. Notice that this conclusion depends on the common prior assumption. Since the government has prior information, the optimal strategy is to have the firms compete against a expected firms following the prior distribution rather than have the firms compete against each other.

However, the competition does improve the *ex-ante* public welfare. The increase in the number of bidders changes the distribution of the winner's type. Since the government selects the most efficient firm, the best among the bidders is expected to be better as the number of bidders increases. That is, the competition makes the winner selective. Moreover, the competition also reduces private information rent. Under the optimal mechanism, the information rent is the integral of the product of the capacity and the *interim* probability of winning, which decreases with the number of bidders.

The effectiveness of the competition also depends on how the government implements the mechanism. We show that competitive bidding is not fully competitive, as the efficient firm is still able to charge an excessive premium on the bidding price of its proposal. Moreover, when the government regulates the demand level instead of toll/capacity decision, the firm still has room to extract excessive information surplus through self-selecting the toll and capacity choices. But in practice, the government is still able to regulate the demand to achieve an approximation of the optimum and simplify the bidding process.

4.2. Oligopoly in the private road network

The private road competition in a network is essentially a Bertrand oligopoly with a sophisticated demand system, which is determined by the network structure. We have drawn some conclusions on such a demand system. First, the capacity decision is not an act of competition but an auxiliary decision to adjust the volume in the road. That is, the firm chooses the capacity for the optimal v/c ratio, not for stealing customers from other road links. On the contrary, the tolling decision is competitive. Firms would lower the price for more volume when a competitor comes. Also, as in the vertical market, the firms in a network may collaborate for the profit. Thus, there might be double marginalization effects, i.e., multiple firms' interaction results in a higher price than the monopoly price, in the network. The double marginalization effects can be reduced or even eliminated by the competition in the parallel links. And we show that how these canceling effects work depends on the total demand and demand elasticity as well.

Moreover, when the parallel competition is not able to cancel out the double marginalization effects, the bilateral negotiations between the government and each individual firm may not be Pareto efficient in terms of the aggregate benefits of the whole network. That is, the private competition in the network is non-cooperative. The efforts put into negotiations with private firms always help increase the welfare but may not be sufficient to solve the Pareto inefficiency. In such circumstances, the government has to coordinate all the private firms towards higher social welfare and aggregate profit. In practice, a Kaldor-Hick improvement implemented by the transfers between the government and firms is able to facilitate such coordination.

4.3. Future works

This dissertation not only addresses the concerns in the highway franchise but also builds up a framework for further studies on the relevant topics. In particular, the dissertation can be extended in the two major ways:

(1) We can build on the demand pricing mechanism and further improves the practicality of the implementations: reducing the demand risk

So far, we ignore the uncertainty in demand. In fact, the firms have a great concern about the demand risk due to the massive upfront investment. In addition, firms may be more sensitive to demand change when the incentive is created based on the demand. LPVR auction(Engel et al., 1997) effectively eliminates the demand risk for the private firm by providing a flexible term extension if the expected revenue is not met at the end of the original term. However, the term extension eliminates the incentives of maintaining a high demand level and providing better service. An interesting and challenging research direction is to make the use of our demand-related incentive framework while taking into account the demand risk for private firms.

Essentially, in addition to the principal-agent problem discussed in this dissertation, there is an asymmetry on the risk aversion between the government and private firms. The government is almost an expected utility maximizer, while the private firms are very risk-averse, especially for the highway project, which involves massive investment. LPVR transfers the demand risk from the private concessionaire to the government so that demand affects the public welfare, via the length of private ownership, instead of the firm's revenue in the present value.

Our ongoing research project addresses the demand risk and toll-capacity optimization by the dynamic mechanism design. That is, the government provides dynamic contacts to the firms, which are contingent not only on the firms' initial proposal but also on the demand performance in the subsequent periods. It is worth noting that such a dynamic environment has another source of private information: the demand shock. When the volume in the road is low, the reason might be the exogenous demand shock or the firm's inadequate service. The firm observes the demand shock directly, but the government is not able to. Thus, we need to capture the information asymmetry on the demand shock and make sure the payment schedule is based on the efforts the firms put into the project rather than the actual demand level. Essentially, we are designing a dynamic demand pricing mechanism that updates the payment schedule after each period of the realization of demand. Further, once the dynamic mechanism reveals the demand shock, in addition to removing its impact on the payment schedule, the government may adjust the concession term to ensure the private revenue is not dramatically affected by the demand shocks. In that way, the government plays the role of the principal and an issuer. Put them together, we aim to extend the dissertation by designing a dynamic mechanism in which the government optimizes toll/capacity choices, concessionaire selection, and mitigates the demand risk for private firms.

(2) Quadratic subsidy: a prior-free implementation.

Another ongoing work is related to the implementation of a prior-free mechanism. When there is no common knowledge on the prior distribution, we may optimize the public welfare without requiring the prior distribution. We can still build on a similar mechanism design framework while imposing a stricter incentive compatibility rule: the dominant strategy incentive compatibility(DSIC), which ensures the firm always truthfully reports without even guessing other firm's strategy. The corresponding solution is essentially a VCG mechanism which induces the optimal allocation and decision making. Moreover, the mechanism can be approximately implemented by a subsidy schedule in the form of a quadratic function of demand. That is, the private concessionaire is able to earn a bonus or a tax credit for each unit of traffic. By the increasing per-unit reward function, the private firm has the same objective function as the government and therefore has the incentive to set up public optimal decisions. One should notice two major differences from the demand pricing mechanism: This subsidy function does not depend on the prior; This subsidy is designed for creating incentives for the firms to set up social optimal decisions rather than screening the private firms.

(3) Combine the analysis of oligopolistic private road network and the franchising mechanism.

We can think of bidding for multiple roads in a network instead of a single road. The allocation rule is, therefore, not only to determine the winners but also to assign the links to the set of winners. In practice, there might be multiple road links owned by the same firm. Thus, the government assigns the ownership not only by the firm's efficiency but also by the road link's position in the network. For example, the government should not allocate two parallel links to the same firm because the shared ownership eliminates the competition from the parallel links.

References

Albalate, D., Bel, G., Fageda, X., 2009. Privatization and regulatory reform of toll motorways in Europe. Governance 22, 295–318.

Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. The American Economic Review 83, 161–179.

Basov, S., 2006. Multidimensional screening. volume 22. Springer Science & Business Media.

Blom-Hansen, J., 2003. Is private delivery of public services really cheaper? Evidence from public road maintenance in Denmark. Public Choice 115, 419–438.

Braess, D., Nagurney, A., Wakolbinger, T., 2005. On a paradox of traffic planning. Transportation Science 39, 446–450.

Che, Y.K., 1993. Design competition through multidimensional auctions. The RAND Journal of Economics , 668–680.

De Borger, B., Van Dender, K., 2006. Prices, capacities and service levels in a congestible bertrand duopoly. Journal of Urban Economics 60, 264–283.

de Palma, A., Lindsey, R., 2000. Private toll roads: Competition under various ownership regimes. The Annals of Regional Science 34, 13–35.

De Rus, G., Nombela, G., 2000. Least present value of net revenue: a new auction-mechanism for highway concessions .

El-Rayes, K., Kandil, A., 2005. Time-cost-quality trade-off analysis for highway construction. Journal of construction Engineering and Management 131, 477–486.

Engel, E., Fisher, R., Galetovic, A., 1997. Revenue-based auctions and unbundling infrastructure franchises .

Enright, D., 2006. The Chicago Skyway sale: An analytical review. NW Financial Group, LLC .

Feldman, A.M., 1973. Bilateral trading processes, pairwise optimality, and Pareto optimality. The Review of Economic Studies 40, 463–473.

FHWA, H., 2013. Highway statistics. US Department of Transportation .

Gaudet, G., Van Long, N., 1996. Vertical integration, foreclosure, and profits in the presence of double marginalization. Journal of Economics & Management Strategy 5, 409–432.

Geddes, R.R., Wagner, B.L., 2013. Why do US states adopt public-private partnership enabling legislation? Journal of Urban Economics 78, 30–41.

Geoffrion, A.M., 1968. Proper efficiency and the theory of vector maximization. Journal of Mathematical Analysis and Applications 22, 618–630.

Guesnerie, R., 1981. On taxation and incentives: Further reflections on the limits to redistribution. Inst. für Ges.-und Wirtschaftswiss., Wirtschaftstheoretische Abt., Univ.

Guo, X., Yang, H., 2009. User heterogeneity and bi-criteria system optimum. Transportation Research Part B: Methodological 43, 379–390.

Hagquist, R., 2008. Higher gas efficiency equals lower fuel revenues. Public Roads 72. URL: `https://www.fhwa.dot.gov/publications/publicroads/08nov/03.cfm`.

Hartline, J.D., 2013. Mechanism design and approximation. Book draft. October 122.

Herbsman, Z.J., 1995. A+B bidding method—Hidden success story for highway construction. Journal of construction Engineering and Management 121, 430–437.

Jehiel, P., Moldovanu, B., Stacchetti, E., 1999. Multidimensional mechanism design for auctions with externalities. Journal of economic theory 85, 258–293.

Jofre-Bonet, M., Pesendorfer, M., 2000. Bidding behavior in a repeated procurement auction: A summary. European Economic Review 44, 1006–1020.

Laffont, J.J., Tirole, J., 1986. Using cost observation to regulate firms. Journal of political Economy 94, 614–641.

Levinson, D.M., Gillen, D., 1998. The full cost of intercity highway transportation. Transportation Research Part D: Transport and Environment 3, 207–223.

Loertscher, S., Marx, L.M., 2015. Prior-free Bayesian optimal double-clock auctions. Technical Report. working paper.

McAfee, R.P., McMillan, J., 1987. Competition for agency contracts. The RAND Journal of Economics , 296–307.

McAfee, R.P., Vincent, D., 1992. Updating the reserve price in common-value auctions. The American Economic Review 82, 512–518.

Mohring, H., Harwitz, M., 1962. Highway benefits: An analytical framework. Northwestern University Press, Evanston, USA.

Mun, S.I., Ahn, K.J., 2008. Road pricing in a serial network. Journal of Transport Economics and Policy (JTEP) 42, 367–395.

Myerson, R.B., 1979. Incentive compatibility and the bargaining problem. Econometrica: journal of the Econometric Society , 61–73.

Myerson, R.B., 1981. Optimal auction design. Mathematics of operations research 6, 58–73.

Nombela, G., De Rus, G., 2004. Flexible-term contracts for road franchising. Transportation Research Part A: Policy and Practice 38, 163–179.

Pigou, A.C., 1932. The economics of welfare, 1920. McMillan&Co., London, UK .

Ramsey, F.P., 1927. A contribution to the theory of taxation. The Economic Journal 37, 47–61.

Riley, J.G., Samuelson, W.F., 1981. Optimal auctions. The American Economic Review 71, 381–392. URL: `http://www.jstor.org/stable/1802786`.

Riordan, M.H., Sappington, D.E.M., et al., 1986. Awarding monopoly franchises. Bell Communications Research, Incorporated.

Rochet, J.C., 1985. The taxation principle and multi-time Hamilton-Jacobi equations. Journal of Mathematical Economics 14, 113–128.

Rochet, J.C., Choné, P., 1998. Ironing, sweeping, and multidimensional screening. Econometrica , 783–826.

Shi, S., Yin, Y., Guo, X., 2016. Optimal choice of capacity, toll and government guarantee for build-operate-transfer roads under asymmetric cost information. Transportation Research Part B: Methodological 85, 56–69.

Small, K.A., Verhoef, E.T., 2007. The economics of urban transportation. Routledge, New York, USA.

Small, K.A., Winston, C., Evans, C.A., 1989. Road work: A new highway pricing and investment policy. Brookings Institution Press, Washington, D.C., USA.

Stringham, E.P., 2001. Kaldor-Hicks efficiency and the problem of central planning. Quarterly Journal of Austrian Economics 4, 41–50.

Tan, Z., Yang, H., Guo, X., 2010. Properties of Pareto-efficient contracts and regulations for road franchising. Transportation Research Part B: Methodological 44, 415–433.

Tsai, J.F., Chu, C.P., 2003. The analysis of regulation on private highway investment under a build-operate-transfer scheme. Transportation 30, 221–243. doi:10.1023/A:1022503428001.

Ubbels, B., Verhoef, E.T., 2008. Auctioning concessions for private roads. Transportation Research Part A: Policy and Practice 42, 155–172.

USBPR, 1964. Traffic assignment manual. Technical Report. US Bureau of Public Roads. Washington, DC.

Verhoef, E.T., 2007. Second-best road pricing through highway franchising. Journal of Urban Economics 62, 337–361.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers 1, 325–378.

Xiao, F., Yang, H., Han, D., 2007. Competition and efficiency of private toll roads. Transportation Research Part B: Methodological 41, 292–308.

Yang, H., Meng, Q., 2000. Highway pricing and capacity choice in a road network under a build–operate–transfer scheme. Transportation Research Part A: Policy and Practice 34, 207–222.

Yang, H., Meng, Q., 2002. A note on "highway pricing and capacity choice in a road network under a build-operate-transfer scheme". Transportation Research Part A: Policy and Practice 36, 659–663.

Zhang, Z., Durango-Cohen, P., 2012. A strategic model of public–private partnerships in transportation: Effect of taxes and cost structure on investment viability. Research in Transportation Economics 36, 9–18.