



HAL
open science

Investment timing and scale of operation under private information about demand forecast

Daniel Danau

► **To cite this version:**

Daniel Danau. Investment timing and scale of operation under private information about demand forecast. 2005. halshs-00004668

HAL Id: halshs-00004668

<https://shs.hal.science/halshs-00004668v1>

Preprint submitted on 20 Sep 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Investment timing and scale of operation under private information about demand forecast

Daniel Danau*

(Université Lyon 2)

8 Septembre, 2005

Abstract

We study the regulatory policy of a monopoly facing stochastic demand for the service it provides after performing an irreversible investment in infrastructure. We find that under uncertainty, bundling the decisions about investment timing and scale of operation is beneficial. When public-private cooperation is necessary to cover the investment cost and accumulation of public funds is costly, waiting longer is traded-off against rationing consumers. As soon as informational asymmetries about traffic forecast appear, the regulator enlarges the waiting period even more but sets the quantity closer to the first best level, as compared to the second best environment.

JEL classification: D6; D8; L5

Keywords: Stochastic demand; investment timing; regulation of monopoly, public-private financing

*Laboratoire d'Economie des Transports, ISH, 14 Av. Berthelot, F-69363, Lyon;
Tel: 33 (0)4 72 72 65 17; Fax: 33 (0)4 72 72 64 48; Email: danieldanau@let.ish-lyon.cnrs.fr

1 Introduction

In the present paper we examine the design of the optimal regulatory policy when the regulated firm firstly shares a sunk cost investment in infrastructure together with public authorities and subsequently operates the service which can be provided by using the infrastructure. The decision variables, namely the timing of investment and the operation scale, are both determined at the evaluation stage of the project.

The basic study of optimal timing for investment in an irreversible project in presence of uncertainty is that of McDonald and Siegel (1986)). They show that there exists an optimal delay until the expectation of marginal productive capital is high enough to limit the risk of its further decline. The firm decides when to invest in an indivisible unit of capital that would be offered as output on a competitive market. A specific orientation of subsequent studies is that of the correspondence between uncertainty and optimal length of the capacity to be installed, when such decision is irreversible. Firstly, many authors consider the firm's decision in building subsequent incremental units, according to the dynamic of demand for capacity. The pioneer works of such kind are those of Pindyck (1988) and Berthola (1988). Secondly, Sodal (2001), shows that in a single investment case, a monopoly can choose the size of one shot investment, together with its timing. Like this, the firm benefits from two options: when to invest and at which scale to operate on the supply market.

The decision of investing in capacity units is equivalent in such studies with the supply decision. However, different cases can be relevant in practice. In transportation for instance, a new physical infrastructure is indivisible. The decision maker can choose the number of lines of a railway or the broad of a highway but not the number of kilometers for covering the demand between two specific areas. Subsequently, the service market is not competitive, but rather provided by a monopolist that is regulated. The regulation of the price is according to the cost of production and the demand for the service, not for capacity. The optimal price may clear the market at a demand lower than the capacity installed.

The decision maker benefits in the evaluation of the project from two instruments: investment timing in an indivisible capacity and the size of operation. As both decisions depend on the uncertain demand, there exists an optimal bundling of such decisions. The aim of this study is to show that such an optimum exists and to examine its relevance for the optimal regulation in the transportation sector. Differently from the majority of studies about public-private partnerships in transportation, our approach is not an incomplete contract one. We assume that a regulatory contract for both investment and operation is signed out under full commitment so that any renegotiation issue is ruled out.

Contracts between public and private sector for such investments are very present today. In literature, examples are provided by Szymanski (1991), who investigates both socially and privately optimal timing of investment in transportation, but does not look at a regulatory mechanism which is able to align the two perspectives. His analysis is performed under the assumption that demand is deterministic and perfectly observable.

In reality the demand for transportation service is not deterministic. Traffic provision is affected by forecast errors, which should be accounted for in the regulatory design process. It is argued that errors follow from technical as well as incentive reasons.

For instance, Quinet (1998) distinguishes three sources of inaccuracy: model structure, current data and future value of exogenous variables. Small and Winston (1999) underline that uncertainty over future traffic originates, on one side, from exogenous factors like economic and technological conditions, which are difficult to properly forecast; on the other side, uncertainty is due to transferability of data across time and space. However, the Authors mainly attribute the forecast failure to incentive problems. In their empirical comparison between roads and railroads, Skamris and Flybjerg (1997) show that forecast about railroad traffic seems to be technically more problematic than the one about roads; nevertheless, they agree that errors are associated also to incentive problems. Some pieces of work, belonging to the domain of literature about concession contracts (Crampes and Estache (1997))

as well as about privatization (Trujillo et al. (2000)), analyze the incentive issue in provision of transportation service. No study examine the optimal investment timing when such an issue arises.

In our investigation, we allow for both technical and incentive errors in traffic prevision. The Real Option techniques have recently been exploited to study the managers / shareholders relationships beset by informational asymmetries. Grenadier and Wang (2004) focus on the case of managers who have to engage in a project whose value exhibits two essential components: a constant one, that is unobservable, and a stochastic one, that is observable. In their turn, Antle *et Al.* (2004) investigate the optimal investment timing when the unobservable component of the project (in their setup, the cost), is stochastic. Both papers achieve the conclusion that managers display greater inertia in their investment behavior than shareholders do; hence, the latter adopt a compensation strategy which reduces incentives to postpone investments beyond the optimal time.

Starting from the studies previously illustrated, we explore how the incentive mechanism for appropriate investment period in transportation infrastructure should be structured. In line with the Real Option literature (see Dixit and Pindyck, 1994), we assume that the demand is partially deterministic and partially stochastic. Moreover, as Antle *et Al.* (2004), we allow the firm to privately observe the current realization of the stochastic variable in each period; the assumption that the latter is stochastic and that realizations are only privately observable is backed by real-world transportation studies. As the size of initial investment requires that the market for the service be monopolistic (at least for a reasonable period), we propose a regulatory mechanism which bundles investment time and service operation. Then the authority disposes of two instruments for controlling one sector of activity.

The economic model on which we are mainly based is the one of regulation when subventions are allowed and the collection of public funds is costly. We show that, when the cost of investment needs to be shared between public funds and firm's funds so that the later's budget constraint is to be satisfied, it is necessary to trade-off the investment delaying (with respect to the first-best one) against consumers'

rationing in each period of operation. As soon as informational asymmetries are introduced, the new agency cost changes the optimal waiting period/quantity pair: the latter becomes longer but larger output is provided.

The paper is organized as follows. In Section 2, the model and the first-best benchmark are presented. In Section 3, we characterize the second-best environment where in symmetric information context the private investor needs to be ensured his budget balance through public participation in investment. In Section 4, the regulator and the agent are assumed to have different demand previsions because only the latter can observe the current realization of the stochastic demand before investing; we study the arising trade-off which arises between agency cost and efficiency. Section 5 concludes.

2 The model and first-best benchmark

We consider a project which requires the investor to bear a significant sunk cost I for building an essential infrastructure. The social planner offers to the firm a contract for investing and operating the transportation service over an unlimited horizon¹. The participation of the firm is ensured through a share of sunk cost investment from public funds. For simplicity of analysis, both social planner and the operator are assumed to be risk neutral. The crucial decisions to be made consist in the time of investment, as determined according to the stochastic shift of the demand, and the scale of service operation.

2.1 Consumer surplus and demand

We assume that the demand for the transportation service provided on the market under scrutiny changes unpredictably when random shocks occur. Time is continuous and indexed by $t \in [0, \infty)$. The inverse demand function at time $t \geq 0$ is given by

$$p_t = p(Q) y_t. \tag{1}$$

¹This may be seen as a sufficiently long, though finite, period.

As (1) shows, p_t is composed by two parts. $p(Q)$ is the non-stochastic component of the demand function, which directly depends on the quantity Q of service. Furthermore, y_t shifts the aggregate willingness to pay between any two periods and for any quantity of the good, according to a shock affecting the industry. We assume that the shock y_t follows a geometric Brownian motion with drift α and volatility σ , so that

$$\frac{dy_t}{y_t} = \alpha dt + \sigma dz_t. \quad (2)$$

y_0 is the value it takes at the current date $t_0 = 0$, while $z_t = \epsilon\sqrt{d_t}$ is a simple Brownian motion. As in real options literature (see Dixit and Pindyck (1994)), we assume that the dynamic of market demand is driven by consumers' tastes for the output and not by replication of initial consumers.

We hereafter make two assumptions which help perform the subsequent analysis. Neither of them is particularly strong, rather they are in line with the hypotheses usually adopted by the literature.

- *Assumption 1:* $p(Q)$ on the domain \mathbb{R}_+ is strictly decreasing, continuously differentiable and integrable, and it is independent of the current shock y_t . Moreover, the mapping $Q \rightarrow Qp(Q)$ is strictly concave and $p(0)$ satisfies the inequalities $0 < p(0) < \infty$.
- *Assumption 2:* We denote the discounting rate of the problem by r . The drift α , which we introduced in (2), and the discount rate r satisfy the following conditions: $0 < \alpha < r < 1$. In particular, the inequality $\alpha < r$ means that the unit expected increase in demand is lower than the unit discount rate.

Assumption 1 ensures that in absence of production cost, the optimal scale of operation, provided the value of y_t , exists and is unique. Assumption 2 states the convergence condition which is standard in the Real Option literature and guarantees that the optimal waiting period is finite.

Denote $s(Q) = \int_0^Q p(s) ds$. The consumer surplus at any instant t , is given by

$$s(Q) y_t = y_t \int_0^Q p(s) ds, \quad (3)$$

while the marginal gross consumer surplus with respect to Q writes as

$$\frac{\partial s(Q) y_t}{\partial Q} = \frac{\partial s(Q)}{\partial Q} y_t = \bar{p} y_t, \quad (4)$$

$\bar{p} y_t$ being the price of period t .

2.2 The firm's technology

As previously illustrated, the project requires an irreversible fixed cost I ; once the latter is borne, a total capacity \bar{Q} is made available for operation. The service is provided in quantity $Q \leq \bar{Q}$, which calls for the operational cost $c(Q) = cQ$. Consumers benefit from the new service immediately after the lump-sum investment is performed².

2.3 The socially optimal time of investment and quantity of service

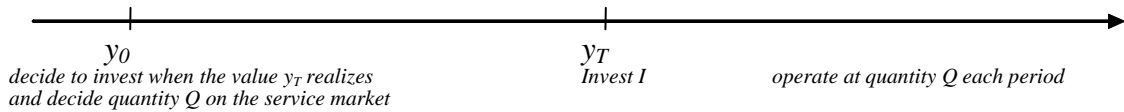


Figure 1: Evolution of the stochastic variable and performed decisions

The decision maker determines the optimal threshold y_T for performing the investment (at the stochastic date T) as well as the optimal scaling Q of operation on

²Strictly speaking, an operational cost function as in the text exhibits constant returns to scale. However, the proposed formulation is adopted because it well accomodates for the presence of capacity constraint. More precisely, the marginal cost of production is assumed to be constant for $Q \leq \bar{Q}$ and to become infinite at \bar{Q} . Hence, up to capacity, one has constant returns in operation, but overall increasing returns to scale due to the presence of I (see Tirole, 1988).

the service market. The timing of his decisions is represented in Figure 1 according to the evolution of the stochastic variable y_t . Notice that the service provision Q is assumed to be constant over time. In a more general case, Q would solely express the initial supply; during the operation interval, the offered quantity would optimally change, taking eventual adjustment costs into account. To accommodate for such a variation would make the analysis more complex, without adding any striking economic insight about the time-and -scale bundle; therefore, we rule out this scenario.

The expected payoff to be optimized is

$$E_{y_0} \left[(S(y_T, Q) - C(I, Q)) e^{-rT(y_T)} \right], \quad (5)$$

where E_y means that the expectation is taken over the current realization y_0 of the stochastic variable at the current date $t = 0$. Moreover

$$S(y_T, Q) = E_{y_T} \left[\int_0^\infty s(Q) y_{T+s} e^{-rs} ds \right] = s(Q) \frac{y_T}{r - \alpha} \quad (6)$$

is the gross social benefit of the project, evaluated at date T , when the value y_T of the stochastic variable is reached and the investment takes place. The total cost required for investment and operation, evaluated again at date T , is given by

$$C(I, Q) = \int_0^\infty c(Q) e^{-rs} ds + I = c(Q) \frac{1}{r} + I.$$

The expectation of the social return in (5) can be transferred from the space of time to the space of realizations of the stochastic variable, by using the closed form (see Dixit *et Al.*, 1999)

$$E_{y_0} [e^{-rT}] = \left(\frac{y_0}{y_T} \right)^\beta. \quad (7)$$

where β is the positive root of the quadratic equation

$$\frac{1}{2} \sigma^2 \beta (\beta - 1) + \alpha \beta - r = 0, \quad (8)$$

so that one has

$$\beta = \frac{1}{2} + \frac{\alpha}{\sigma^2} + \sqrt{\left[\frac{\alpha}{\sigma^2} - \frac{1}{2}\right]^2 + \frac{2r}{\sigma^2}} > 1. \quad (9)$$

From (7) and (5) the discounted net benefit writes as

$$\left(\frac{y_0}{y_T}\right)^\beta (S(y_T, Q) - C(I, Q)) \quad (10)$$

or, equivalently, as

$$\left(\frac{y_0}{y_T}\right)^\beta \left(S(Q) \frac{y_T}{r - \alpha} - c(Q) \frac{1}{r} - I\right). \quad (11)$$

All along our analysis the no bubbles condition is assumed to hold; it ensures that the social value of the project $W(y_T, Q)$ cannot go subject to speculative bubbles on the financial market (see Dixit and Pindyck, 1994).

Maximization the social welfare function with respect to the quantity and investment trigger provides the first-order conditions

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\varepsilon_{CQ}}{\varepsilon_{WQ}} \quad (12)$$

and

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\beta}{\beta - 1} \quad (13)$$

respectively, where $\varepsilon_{CQ} \equiv QC_Q/C$ and $\varepsilon_{SQ} \equiv QS'/S$ are the elasticities of cost function and social surplus with respect to quantity. Notice that ε_{SQ} is independent of the trigger y_T .

At this stage of the investigation, we make two additional assumptions to be relied upon in the subsequent discussion.

- *Assumption 3*: $\frac{\partial}{\partial Q} \left(\frac{S}{C}\right) < 0$. This condition holds if and only if $\varepsilon_{CQ} > \varepsilon_{SQ}$.
- *Assumption 4*: Define $R(y_T, Q) \equiv Qp(Q) \left(\frac{y_T}{r - \alpha}\right)$ the discounted revenues from the service provision at period T of investment. Moreover, denote the elasticity of R with respect to Q as $\varepsilon_{RQ} \equiv QR_Q/R$. Notice that $\varepsilon_{RQ} > 0$ if

$R_Q > 0$ and negative if $R_Q < 0$. We then assume that $\frac{\partial}{\partial Q} \left(\frac{S}{R} \right) \geq 0$. This inequality holds if and only if $\varepsilon_{SQ} \geq \varepsilon_{RQ}$. Similarly to ε_{SQ} , ε_{RQ} turns out to be independent of y_T .

As it will become clear, Assumptions 3 and 4 ensure that the scaling and an interior solution for Q in presence of production cost and dynamic change of demand, can be entailed ³. In particular, Assumption 3 is standard whereas Assumption 4 imposes quite mild restrictions. For instance, it holds at any level of Q when the demand function is linear; on the other hand, if the demand exhibits constant elasticity to price, the condition is satisfied with equality.

We next discuss the first-order conditions previously characterized. (12) suggests that, at the first-best value of y_T , the ratio between discounted gross social surplus $S(y_T, Q)$ and discounted cost of investment and operation $C(I, Q)$ has to equal the ratio between elasticity of cost and elasticity of gross surplus with respect to quantity. This is just equivalent to saying that marginal social revenue S_Q and marginal cost C_Q are equal.

(13) requires that the same ratio $S(y_T, Q)/C(I, Q)$ be equal to $\beta/(\beta - 1)$. In words, this means that the optimal period for performing the investment comes when the stochastic demand is such that the social revenue is larger enough than the total cost, so as to minimize the risk of subsequent decline. The solutions in (12) and (13) are equivalent to the ones in Sodal (2001) who had the intuition that it is easier to analyze the optimal joint decision of investment timing and operation scaling in terms of (absolute values of) gross surplus and overall cost. In what follows, we proceed analogously.

More precisely, Sodal (2001) studies the decision about investment period and scale of operation from the perspective of a monopolist facing stochastic demand; his solution has the same form as above, but S is replaced by the market revenues

³For a positive waiting period to be optimal, a sufficient condition is that for any infinitesimal quantity $\varepsilon > 0$, the initial value y_0 is such that $\frac{S(y_0, \varepsilon)}{C(I, \varepsilon)} < \frac{\beta}{\beta - 1}$. We take this condition as satisfied.

R . He assumes constant marginal cost⁴ and constant elasticity of demand with respect to the price, so that ε_{CQ} and ε_{RQ} are both constant. Since both $\varepsilon_{CQ}/\varepsilon_{RQ}$ and $\beta/(\beta - 1)$ are independent of the decision variables, different corner solutions emerge according to the relation between $\varepsilon_{CQ}/\varepsilon_{RQ}$ and $\beta/(\beta - 1)$.

In our context, the presence of the cost $I > 0$ involves increasing returns to scale, so that the elasticity ε_{CQ} is sub-unitary and increasing in Q . Moreover, under Assumption 4, ε_{SQ} (weakly) decreases in Q . Overall, it is possible to find a finite scale Q for which the ratio $\varepsilon_{CQ}/\varepsilon_{SQ}$ increases in Q . As a result, it adjusts to the level $\beta/(\beta - 1)$ in (12) and (13), so that an interior solution (y^{FB}, Q^{FB}) may exist. Indeed, at the optimal scale of operation Q^{FB} , one of the following two scenarios is realized

$$\begin{cases} \frac{\varepsilon_{CQ^{FB}}}{\varepsilon_{SQ^{FB}}} = \frac{\beta}{\beta - 1}, & \text{if } Q^{FB} < \bar{Q}, \\ \frac{\varepsilon_{CQ^{FB}}}{\varepsilon_{SQ^{FB}}} < \frac{\beta}{\beta - 1}, & \text{if } Q^{FB} = \bar{Q} \end{cases}, \quad (14)$$

while the first best trigger $y_T = y^{FB}$ is found in equation (13) evaluated at Q^{FB} .

We know from (9) that the uncertainty (as measured by the volatility σ in (13)) is directly proportional to $\beta/(\beta - 1)$. Indeed, (13) shows that given the quantity, higher uncertainty means longer waiting period. This is the standard solution we find in the Real Option theory for optimal delay. In the light of this, the equations in (14) prove to be particularly instructive, as we illustrate hereafter.

Firstly, the optimal scale of operation is also direct proportional to the level of uncertainty. At any $Q < \bar{Q}$ such that $\varepsilon_{CQ}/\varepsilon_{SQ} < \beta/(\beta - 1)$, the option to delay investment (the dynamic dimension of the model) embodies another option, that of upward scaling the operation (the static dimension), in case the investment is performed later, when the market conditions are more favorable. Once the equality in (14) is achieved, the benefit from immediate production overwhelms the benefit from waiting more, hence the investment takes place. In the presence of higher uncertainty, later investment and higher scale of operation are optimal. Furthermore,

⁴In Sodal (2001) the initial investment consists in buying a land to be used to build houses. The cost associated to such a purchase is not considered in the analysis. In our framework, on the opposite, the cost I is embodied in the overall cost function as relevant to the decision about timing.

(14) suggests that the first best scale is more sensitive to changes in uncertainty when the demand exhibits constant elasticity.

Secondly, the larger Q , the smaller the ratio $S(y_T, Q)/C(I, Q)$ appearing in (13) reduces. It follows that a supplementary increase in the waiting period (larger y_T) is required, as compared to the case where the scale is exogenous and fixed at the initial level. The additional delay is simply the consequence of the fact that the decision about the operational scale is embodied in that about the investment timing.

Nevertheless the option to augment output is limited by the capacity constraint⁵; hence, under higher uncertainty, it is more likely that such a constraint binds so that the second scenario in (14) is entailed. In the rest of our investigation we assume that capacity is not binding. Interestingly, it might be socially optimal to perform irreversible investments even in case the capacity is not entirely employed, as long as the uncertainty is not very high. The two cases are presented in Figure 2 in the Appendix.

3 The second-best environment: symmetric information and budget balance

The solution previously characterized was derived without being concerned by budget balance issues. If the social planner imposed the optimal threshold for investment and the optimal quantity of the service to the investor, the latter would not break-even and be forced to give up the project.

The operator breaks-even only in case the in-coming money flow is high enough to cover the overall cost $C(I, Q)$. We can easily notice that, as long as $I > 0$, at the first-best solution (y^{FB}, Q^{FB}) the market revenues are not large enough to ensure budget balance. On the other hand, public funds participation comes at marginal social cost λ . It means that an amount t from public funds for covering the initial

⁵We take the *size* of capacity as given, that is we do not allow it to be chosen. The decision is binary: either investment is performed in size \bar{Q} or it is not.

investment requires that an amount $(1 + \lambda)t$ be collected. We also suppose that consumer surplus and profits are attributed the same weight in social welfare, so that the programme to be solved writes as

$$\begin{cases} \left(\frac{y_0}{y_T}\right)^\beta [(S(y_T, Q) - R(y_T, Q) - (1 + \lambda)t) + (R(y_T, Q) + t - C(I, Q))] \\ \text{s.t. } R(y_T, Q) + t - C(I, Q) \geq 0 \end{cases} \quad (15)$$

The constraint says that the amount t of public participation has to be larger or equal to the difference between discounted cost and discounted revenues of the firm. The constraint is binding at optimum, so that the function to be maximized becomes

$$\left(\frac{y_0}{y_T}\right)^\beta [S(y_T, Q) - C(I, Q) + \lambda(R(y_T, Q) - C(I, Q))]. \quad (16)$$

The first-order condition with respect to Q writes as

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\varepsilon_{CQ}}{\varepsilon_{SQ}} + \frac{\lambda}{1 + \lambda} \left(\frac{S(y_T, Q)}{C(I, Q)} - \frac{\varepsilon_{RQ}}{\varepsilon_{SQ}} \frac{R(y_T, Q)}{C(I, Q)} \right), \quad (17)$$

whereas the optimality condition for y_T is given by

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\beta}{\beta - 1} + \frac{\lambda}{1 + \lambda} \left(\frac{S(y_T, Q)}{C(I, Q)} - \frac{R(y_T, Q)}{C(I, Q)} \right). \quad (18)$$

(17) is a different way to write the Ramsey type formula characterizing of optimal pricing, where $\frac{\partial S(y_T, Q)}{\partial Q} = p(Q) \frac{y_T}{r - \alpha}$ is the discounted expected price evaluated at the investment date. (17) and (18) show that, at the solution (y^{SB}, Q^{SB}) , the ratio S/C is greater than both $\varepsilon_{CQ^{SB}}/\varepsilon_{SQ^{SB}}$ and $\beta/(\beta - 1)$, so that (y^{SB}, Q^{SB}) differs from the first-best solution. Since S/C decreases in Q (by Assumption 3) and increases in y_T , the solution of the constrained problem is such that the quantity is downward distorted and the delay prolonged. Enlarging the waiting period involves postponing the public transfer t . Longer waiting period and/or smaller service provision increase the market revenues of the operator, so that the absolute value of t reduces.

In the previous Subsection we have shown that the objective function is optimized if the decision for any additional delay incorporates the option to scale the operation

upward. Whenever the investment is postponed with respect to the first best, the second best quantity results from two countervailing effects, namely the option to scale upward and quantity rationing. On the other hand, when less than the first best quantity is provided, the optimal delay that embodies the scale of operation is lower. In what follows, we discuss the quantity and timing solutions, taking into account the effects just mentioned.

Similarly to what we did in the previous Subsection, we combine (17) and (18) to characterize Q^{SB} ; we obtain

$$\frac{\beta}{\beta - 1} - \frac{\varepsilon_{CQ^{SB}}}{\varepsilon_{SQ^{SB}}} = \left(\frac{\lambda}{1 + \lambda} \right) \left(1 - \frac{\varepsilon_{RQ^{SB}}}{\varepsilon_{SQ^{SB}}} \right) \frac{R(y^{SB}, Q^{SB})}{C(I, Q^{SB})}. \quad (19)$$

An extreme and interesting case arises when demand exhibits constant price elasticity. Then we have $\varepsilon_{RQ} = \varepsilon_{SQ}$, which are also constant at any quantity level. The quantity distortion is perfectly compensated by the option to scale more through additional delay, so that the quantity solution, as characterized by (19) turns out to be equivalent to the first best one. Observe that the first best quantity is feasible by upward distorting the second best trigger in (18). This result is appealing in that the operator's budget constraint does not call for consumers rationing, as widely argued for this kind of problems, but only for postponing investment beyond the first best time. If Assumption 4 could be weakened to allow for $\frac{\partial \varepsilon_{SQ}}{\partial Q} > 0$ and, simultaneously, $\frac{\partial}{\partial Q} \left(\frac{\varepsilon_{CQ}}{\varepsilon_{SQ}} \right) > 0$, an interior solution for Q at first best would still be possible, whereas Q^{SB} would be upward distorted.

On the other hand, as soon as $\varepsilon_{RQ} < \varepsilon_{SQ}$, S/C in (17) should be higher than in the previous case, which means by Assumption 3 that the quantity rationing effect is more significant. Subsequently, the option to scale more does not entirely compensate the rationing effect so that $Q^{SB} < Q^{FB}$, as provided by (19). As this is the scenario where Assumption 4 is strictly satisfied, we look at it in what follows.

The optimal trigger $y_T = y^{SB}$ is found by replacing Q^{SB} in (18). Manipulating

(18) yields the following "Ramsey-type" equality

$$\frac{1}{y^{SB}} \left[y^{SB} - \frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q^{SB})/C(I, Q^{SB})} y^{FB} \right] = \left(\frac{\lambda}{1 + \lambda} \right) \left[1 - \frac{R(y^{SB}, Q^{SB})}{S(y^{SB}, Q^{SB})} \right].$$

We notice that the distortion away of y^{FB} is corrected for taking into account that changing the operation scale affects the optimal investment period as well (as expressed at previous Subsection). Under Assumption 3, the ratio $\frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q^{SB})/C(I, Q^{SB})}$ is sub-unitary. Once the correction induced by the different scale is considered, the cost of subvention requires the distortion of the trigger to increase proportionally to the difference $\left[1 - \frac{R(y^{SB}, Q^{SB})}{S(y^{SB}, Q^{SB})} \right]$. Uncertainty plays no role in the size of such a distortion; indeed, under the no bubbles condition, social and private value of the project at the date of investment are only expressed in expected terms.

At this point of the analysis it should be clear that also in the determination of optimal trigger two opposite effects show up. Due to this, the final result is ambiguous: the relative position of the second-best trigger with respect to the first-best one depends on which of the two effects dominates. Nevertheless, provided that both $\frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q)/C(I, Q)}$ and $\left[1 - \frac{R(y^{SB}, Q)}{S(y^{SB}, Q)} \right]$ increase in Q , we are able to establish that so does y^{SB} . The intuition is that, subsidies being socially costly, the pair (y^{SB}, Q^{SB}) solves the trade-off between the penalty induced by further postponing the investment and the one associated to more significant consumer rationing.

3.1 The trade-off between waiting and consumer rationing

In order to highlight the trade-off which arises in the second-best environment, we evaluate the unconstrained social welfare function at the second-best solution and obtain

$$F(y^{SB}, Q^{SB}) = \left(\frac{y_0}{y^{SB}} \right)^\beta [S(y^{SB}, Q^{SB}) - C(I, Q^{SB})]. \quad (20)$$

Using (20) allows to determine the marginal rate of substitution $MRS_{Q^{SB}, y^{SB}}$

between investment trigger and scale of operation, which writes as

$$\begin{aligned}
MRS_{Q^{SB}y^{SB}} &= \frac{E_{y_0} \left[\frac{\partial F(y^{SB}, Q^{SB})}{\partial y^{SB}} \right]}{E_{y_0} \left[\frac{\partial F(y^{SB}, Q^{SB})}{\partial Q^{SB}} \right]} \quad (21) \\
&= - \left(\frac{\beta - 1}{\frac{r - \alpha}{\alpha}} \right) \frac{1}{\alpha y^{SB}} \frac{s(Q^{SB}) \left[y^{SB} - y^{FB} \frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q^{SB})/C(I, Q^{SB})} \right]}{\frac{\partial s(Q^{SB})}{\partial Q^{SB}} \frac{y^{SB}}{r - \alpha}}.
\end{aligned}$$

The term $s(Q^{SB}) \left[y^{SB} - y^{FB} \frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q^{SB})/C(I, Q^{SB})} \right]$ measures the social benefit that is lost when the investment is postponed from $y^{FB} \frac{s(Q^{FB})/C(I, Q^{FB})}{s(Q^{SB})/C(I, Q^{SB})}$ up to y^{SB} . The term $-\left(\frac{\beta - 1}{\frac{r - \alpha}{\alpha}} \right) \frac{1}{\alpha y^{SB}}$, where $\frac{1}{\alpha y^{SB}} = E \left(\frac{1}{dy^{SB}} \right)$, expresses the rate of change of the discount factor over the expected change of hitting value y^{SB} . For the purpose of further interpreting, we rely on (8) and obtain

$$\frac{\frac{r - \alpha}{\alpha}}{\beta - 1} = \frac{1}{2} \sigma^2 \frac{\beta}{\alpha} + 1.$$

This equality suggests that, in case the process is deterministic, and so $\sigma = 0$, we have $(\beta - 1) / \left(\frac{r - \alpha}{\alpha} \right) = 1$. Otherwise, from (9) it follows that β reduces when σ increases. Then it turns out that $(\beta - 1) / \left(\frac{r - \alpha}{\alpha} \right) < 1$, involving that the expectation $E_{y_0} \left[\frac{\partial F(y^{SB}, Q^{SB})}{\partial y^{SB}} \right]$ reduces less when uncertainty increases. The cost of waiting is lower under uncertainty, as compared to the deterministic case, due to the fact that future investment will be performed only in case of favorable evolution of the stochastic variable. Finally, the term $\frac{\partial s(Q^{SB})}{\partial Q^{SB}} \left(\frac{y^{SB}}{r - \alpha} \right)$ expresses the marginal social gain associated to a unit increase in quantity.

Overall, (the negative of) MRS measures the amount by which Q should be augmented in order to compensate for a unit increase in y_T . This means that when the operator is let postpone the investment, it should be required to produce more

for social welfare not to decrease. By contrary, earlier investment, closer to social optimum, is compensated by higher quantity rationing.

In this sense, one can gain by bundling investment timing and scale of operation within the same problem, as it was the case in the benchmark of previous Subsection. The role of uncertainty is this time different. Higher uncertainty establishes the optimal trade-off at lower Q and higher hitting value y_T , with respect to the first best case.

4 The rent extraction efficiency trade-off

All along the previous analysis, we took the realizations of the stochastic demand as being observable by the social planner. However, it may be the case that the agent has private observation about it even though the two parties agree upon its law of motion. In the transportation field, it is generally known that operators have better information than public authorities about the dynamics of consumers willingness to pay.

In what follows, we abandon the previous environment and we assume that the dynamic variable y_t is entirely privately observed by operator. More precisely, the latter can base the demand forecast on its current observation, whereas the regulator cannot. Therefore, they have each period asymmetric expectations about future demand. If asked to truthfully tell the realizations of the stochastic variable, the operator would have an incentive to exploit the asymmetric information in demand prevision by reporting too low realizations and so artificially less optimistic forecasts. By doing so, the investment would take place at later stage, when consumers are willing to pay more for the same amount of provided output.

In an environment where uncertainty lasts two periods and the investment can be performed in only one of them by sinking a stochastic and privately observable cost, Antle *et Al.* (2004) demonstrate that, whenever the principal prefers the agent to invest during the first period, rather than the second, he needs to offer a rent that makes the agent indifferent between the two. The informational rent is composed by

paid at the date of investment is equal to

$$x = C(I, Q) - \left(\frac{\beta - 1}{\beta}\right) R(y_T, Q). \quad (24)$$

(24) shows that, given the investment trigger y_T , the subsidy x is to be sized to remunerate for the flexibility loss the agent bears by investing at the socially desirable period. Observe that, with $\left(\frac{\beta - 1}{\beta}\right)$ smaller than one, the transfer covers more than the excess cost with respect to revenues; hence, it is larger than t would be under complete information. Moreover, x decreases in β , that is the more important the uncertainty over future demand, the higher the agent's option to postpone investment.

The operator's discounted utility (expressing her option value) writes as

$$\left(\frac{y_0}{y_T}\right)^\beta (x + R(y_T, Q) - C(I, Q)) = \left(\frac{y_0}{y_T}\right)^\beta \frac{1}{\beta} R(y_T, Q) \quad (25)$$

Using the previous equalities, we are able to rewrite the problem of the benevolent public authority in a form which makes clear the trade-off between social efficiency and reduction of informational cost

$$\begin{aligned} \underset{y_T, Q}{Max} \left\{ \left(\frac{y_0}{y_T}\right)^\beta [S(y_T, Q) - C(I, Q) + \lambda(R(y_T, p) - C(I, Q))] \right. \\ \left. - \lambda \left(\frac{y_0}{y_T}\right)^\beta \frac{1}{\beta} R(y_T, Q) \right\}. \end{aligned} \quad (26)$$

The objective function in (26) is composed by two elements which we can easily identify. Firstly, the term $\left(\frac{y_0}{y_T}\right)^\beta [S(y_T, p) - C(I, Q) + \lambda(R(y_T, p) - C(I, Q))]$ is the same discounted social welfare function we had in the second-best environment. Secondly, the term $\left[\lambda \left(\frac{y_0}{y_T}\right)^\beta \frac{1}{\beta} R(y_T, Q)\right]$ expresses the agency cost of the discounted rent which is given up to the agent; indeed, under asymmetric information, the principal needs to increase the public participation for the purpose of inducing the agent to invest earlier than the latter would otherwise do.

Nevertheless, the transfer is costly for the regulator. Therefore, he necessitates to

trade-off the gain from earlier investment against the agency penalty, the instruments at his hands being investment time and operation scale. Optimizing (26) with respect to Q and y_T yields

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\varepsilon_{CQ}}{\varepsilon_{SQ}} + \frac{\lambda}{1 + \lambda} \left(\frac{S(y_T, Q)}{C(I, Q)} - \frac{\beta - 1}{\beta} \frac{\varepsilon_{RQ}}{\varepsilon_{SQ}} \frac{R(y_T, Q)}{C(I, Q)} \right) \quad (27)$$

and

$$\frac{S(y_T, Q)}{C(I, Q)} = \frac{\beta}{\beta - 1} + \frac{\lambda}{1 + \lambda} \left(\frac{S(y_T, Q)}{C(I, Q)} - \frac{\beta - 1}{\beta} \frac{R(y_T, Q)}{C(I, Q)} \right) \quad (28)$$

respectively. Denoting the solution as (y^{AI}, Q^{AI}) and proceeding as for the second best, we combine (27) and (28) to obtain the optimal scale of operation under asymmetric information. Indeed, at (y^{AI}, Q^{AI}) , we have

$$\frac{\beta}{\beta - 1} - \frac{\varepsilon_{CQ^{AI}}}{\varepsilon_{SQ^{AI}}} = \frac{\beta - 1}{\beta} \left(\frac{\lambda}{1 + \lambda} \right) \left(1 - \frac{\varepsilon_{RQ^{AI}}}{\varepsilon_{SQ^{AI}}} \right) \frac{R(y^{AI}, Q^{AI})}{C(I, Q^{AI})}. \quad (29)$$

In order to establish the relation between optimal quantities at the second-best level and under asymmetric information, we first fix y_T at its second-best value y^{SB} . Replacing into (29) shows that, as long as Assumption 4 holds, the difference $\left(\frac{\beta}{\beta - 1} - \frac{\varepsilon_{CQ^{AI}}}{\varepsilon_{SQ^{AI}}} \right)$ is smaller than if evaluated at Q^{SB} . Therefore, it must be the case that $\frac{\varepsilon_{CQ^{AI}}}{\varepsilon_{SQ^{AI}}} > \frac{\varepsilon_{CQ^{SB}}}{\varepsilon_{SQ^{SB}}}$ and so that $Q^{AI} > Q^{SB}$. We next determine the relation between triggers in the two environments by looking at (28); indeed, evaluating at y^{AI} , the following Ramsey-type equation holds

$$\frac{1}{y^{AI}} \left[y^{AI} - \frac{s(Q^{SB})/C(I, Q^{SB})}{s(Q^{AI})/C(I, Q^{AI})} y^{SB} \right] = \left(\frac{\lambda}{1 + \lambda} \right) \left[1 - \left(\frac{\beta - 1}{\beta} \right) \frac{R(y^{AI}, Q^{AI})}{S(y^{AI}, Q^{AI})} \right]. \quad (30)$$

Provided that, as previously established, we have $Q^{AI} > Q^{SB}$, it is as well true that $\frac{s(Q^{SB})/C(I, Q^{SB})}{s(Q^{AI})/C(I, Q^{AI})} > 1$, so that $y^{AI} > y^{SB}$. Observe that the difference between optimal triggers $(y^{AI} - y^{SB})$ gets larger as $\left(\frac{\beta - 1}{\beta} \right)$ decreases.

The result that $y^{AI} > y^{SB}$ is due to uncertainty. As at the previous Subsection, later investment means a reduction of direct subvention because expected market revenues are higher. However, because the uncertainty of future demand, waiting

for higher profit on the market instead of receiving a direct subvention reduces the discounted utility $\left(\frac{y_0}{y_T}\right)^\beta \frac{1}{\beta} R(y_T, Q)$ of the agent. The principal induces a larger waiting period than at the second best optimum, so that the expected rent it must leave for the revelation of true demand reduces.

The outcome $Q^{AI} > Q^{SB}$ can be interpreted by considering that, as output gets larger, the potential revenues on which the operator can cheat reduce. Interestingly, Q^{AI} is chosen closer to Q^{FB} , as compared to Q^{SB} , because this makes the informational rent which can be grasped less important. The advantage of bundling investment and operation within a single regulatory problem reappears. Under asymmetric information, disposing of Q as a second instrument helps shape the agent's incentive to invest. As at the previous Subsection, in the particular case of constant elasticity of the demand, Q is established at the first best level.

As a conclusion, under asymmetric information, the second-best solution can no longer be implemented. Both quantity and investment trigger are upward distorted due to the presence of the agency cost.

Finally, we would like to parallel the investigation we perform with the works of Lewis and Sappington (1988) on one hand, and Caillaud and Tirole (2001), on the other hand. The first two authors study the regulation of a monopolist under asymmetric information about demand in a static framework. In their context, the need of incentive regulation is determined by the lack of information about the demand curve and the non-observability of the quantity that is sold on the market. Caillaud and Tirole (2001) characterize the regulation of a privately informed monopoly in an inter-temporal context; more precisely, they consider a social planner who regulates the downstream market in such a way that essential facilities are optimally financed. In line with Lewis and Sappington (1988) we have explored the case of asymmetric information about demand; however, we concentrate on a dynamic setting, similarly to Caillaud and Tirole (2001). We mainly diverge from the latter in that, in our work, the market profitability changes dynamically and is stochastically drawn for all players, despite the monopolist has better expectations than the regulator. In our environment, with dynamic demand, the variable investment period becomes

essential for the optimal design of a monopoly franchise; furthermore, as we have shown, the variable scale of operation represents a valuable instrument at limiting the agency cost, provided that the relationship between principal and agent is beset by informational asymmetries.

5 Conclusion

We have analyzed the optimal choice about time of investment and scale of operation under both symmetric and asymmetric information about demand forecast. We have found that it is socially beneficial to bundle the two decisions at the evaluation stage of the project. *Ceteris paribus*, larger uncertainty calls for both longer waiting period and larger output provision. With exogenous size of capacity, the optimal bundle is such that the optimal scale of operation is directly proportional to the uncertainty level. Hence, despite previous literature, the optimal scale maybe sized at a level lower than the full capacity. When the operator is required to balance its budget, a trade-off appears between benefiting later from the service and increasing quantity rationing in each operational period. Under asymmetric information about demand prevision, the presence of the agency cost makes the trade-off more complex and it is necessary to upward distort both investment trigger and level of output with respect to the second-best environment.

The literature about public-private partnerships has highlighted that delegating both the task of building a facility and that of providing the associated service to the same operator is beneficial in the presence of positive externalities between the two activities. For instance, building the infrastructure well enough helps reduce the costs of maintenance which arise during the operational stage. Our result suggests that bundling becomes doubly advantageous as soon as the decision about the time of investment is also considered. Reasonably enough, the risk of renegotiation, which we ruled out under the full commitment assumption, might be reduced by embodying the volatility of forecasted benefits in the *ex ante* decisional process. Furthermore, the optimal downstream market structure through access regulation modifies when

the social policy embodies the benefit of investment timing and optimal scale within the same framework.

References

- Antle, R., Bogetoft P. and Stark A. W. (2004), Incentive problems and investment timing options, Working paper
- Caillaud B. and Tirole J. (2004), Essential facility financing and market structure, *Journal of Public Economics*, 88(3-4), 667-694
- Crampes, C. and Estache A., (1997), Regulatory trade-offs in designing concession contracts for infrastructure networks, World Bank, Policy Research Working Papers, No. 1857
- Dixit, A. and Pindyck R. S. (1994), Investment under uncertainty, Princeton University Press
- Dixit, A., R. S. Pindyck and S. Sodal (1999), A markup interpretation of optimal investment rules, *Economic Journal*, 109(455), 179-189
- Grenadier, S. and Wang N. (2004), Investment timing, agency, and information, forthcoming in *Journal of Financial Economics*
- Lewis, T. R. and Sappington D.E.M. (1988), Regulating a monopolist with unknown demand, *American Economic Review*, 78(5), 986-998
- McDonald, R. and Siegel D. (1986), The value of waiting to invest, *Quarterly Journal of Economics*, 101(4), 707-727
- Quinet, E. (2000) Evaluation methodologies of transportation projects in France, *Transport policy*, 7, 27-34
- Pindyck, R (1988) Irreversible investment, capacity choice and the value of the firm, *American Economic Review*, 78(5), 969-985

- Skamris, M. K. and Flyvbjerg B. (1997), Inaccuracy of traffic forecasts and cost estimates on large transport projects, *Transport Policy*, 4(3), 141-146
- Small, K. A. and Winston C. (1998), *The Demand for Transportation: Models and Applications*, Irvine Economics Papers, No.98-99-06
- Sodal, S. (2001), Monopoly with timing and scaling options, Working paper
- Szymanski, S. (1991), The optimal timing of infrastructure investment, *Transport Economics and Policy*, 25(3), 247-258
- Tirole, J. (1988), *The Theory of Industrial Organization*, The MIT Press, Cambridge
- Trujillo, L., Quinet E. and Estache A. (2002), Dealing with demand forecasting games in transport privatization, *Transport Policy*, 9, 325-334
- Trujillo, L., E. Quinet and A. Estache (2000), Forecasting the demand for privatized transport: what economic regulators should know and why, Preliminary draft

Appendix

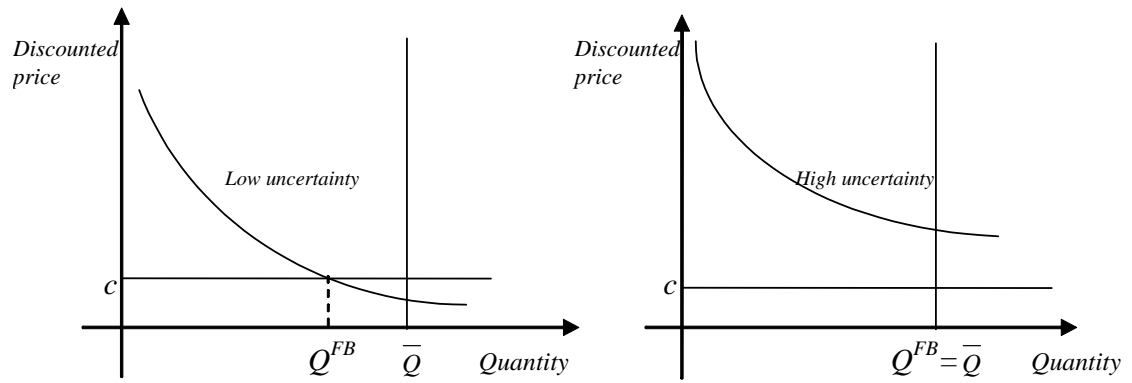


Figure 2: First best supply and discounted price
evaluated at investment date