Transport infrastructure investment and demand uncertainty

by

Stef PROOST
Saskia VAN DER LOO

Energy, Transport and Environment

February 2008

DISCUSSION PAPER

Transport Infrastructure Investment and Demand Uncertainty

Stef Proost*   Saskia van der Loo[+]

February 22, 2008

Abstract

In transportation planning there can be long lead times to adapt capacity. This paper addresses two questions. First, in a one mode world (say rail or road), what is the optimal capacity choice when faced with uncertain demand, long lead times and congestion. Using a simple analytical model it is shown that when demand is inelastic, it is socially optimal to invest more than if only the expected level of demand is taken into account. In this case it may be beneficial to overinvest in capacity because congestion costs are a convex function of relative use. This result holds with or without optimal tolling. The second question deals with two competing modes and where only one mode has long lead times for capacity while the other has flexible capacity. This is typical for the competition between High Speed Rail and air for the medium distance trips (500 to 1000 km), or for the competition between inland waterways and trucks for freight. We find that overinvestment is less justified because the substitute mode can more easily absorb the high demand outcomes.

Acknowledgements and author information

* Proost: Department of Economics, Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven, Belgium
(stef.proost@econ.kuleuven.ac.be),
[+] Van der Loo: Department of Economics, Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven, (saskia.vanderloo@econ.kuleuven.ac.be).

1   Introduction

Transport infrastructure is known as a lumpy investment with long lead times. The construction of a new motorway, a new high speed rail line or a new canal may take 10 years or more. Whether to take on a new project or not and in order to choose the right capacity, one needs demand forecasts for the next 10 to 30 years. Studies of past large transport infrastructure projects have shown that demand has been systematically overestimated and that costs have often been underestimated (Flyberg, Bruzelius and Rothengatter [6]).

There are many sources of uncertainty in infrastructure projects. In this paper we concentrate on only one possible source of uncertainty: the level of future demand. Given that future demand is uncertain and that one needs to decide on the capacity level now, is there a justification to overinvest rather than to underinvest in infrastructure capacity? To discuss this question we use two analytical models: a one mode model and a two mode model.

In the one mode model we are able to show analytically that "overinvestment" is under certain conditions a better strategy than to chose the capacity in function of the expected level of demand. These conditions are surprisingly simple: the demand elasticity has to be smaller than one. This one mode model makes more sense when there are no easy transport alternatives for a given trip.

In the second model we deal with two modes that can easily be substituted and where one mode has long lead times in capacity additions, but the other mode can easily adjust its capacity. Examples are the competition between high speed rail (long lead times) and air transport for passengers and the competition between inland waterways (long lead times) and trucks for freight. In this case it is no longer possible to show that "overinvestment" is systematically optimal because the other mode serves as an escape route for high demands.

In section 2 we introduce the topic and illustrate the main results graphically. Section 3 links our paper to the literature. In section 4 we use the one mode model, section 5 analyses the two mode model and section 6 concludes.

2   Transport investments under uncertainty: a graphical illustration

In this section we introduce our problem setting using graphical illustrations. We do this primarily for the case where no tolls exist. In section 4 we will show that most results carry over to the case with optimal or non optimal tolling. We first illustrate the one mode case and then extend our illustration to the two mode case.

There is a transport planner who faces an uncertain level of future demand for trips between two points. There is only one mode of transport, demand levels, once realized, are constant over time and the capacity of the transport infrastructure has to be decided before the level of demand is known. Typical examples are new motorways, high speed rail connections, canals and ports whose construction or extension may take 10 or more years. The level of capacity that is chosen determines the users' costs. In Figure 1 we represent the simplest case where demand (for a representative period) is price inelastic and is either equal to a level N^- or equal to N^+, both with equal probability. The expected demand level is shown as $\bar{N}$. The Y axis represents the generalized cost of a trip.

To simplify the analysis we put non time costs equal to zero. We furthermore restrict ourselves to a linear average time cost function $(ATC(N, \bar{K}))$. The average time cost of a trip is an increasing function of the level of use and a decreasing function of the level of capacity. In Figure 1, we show the average time cost function that corresponds to the capacity level $\bar{K}$. This capacity level is the level of capacity that maximizes welfare (consumer surplus minus costs of capacity) if the decision maker only takes into account the expected demand level $\bar{N}$.
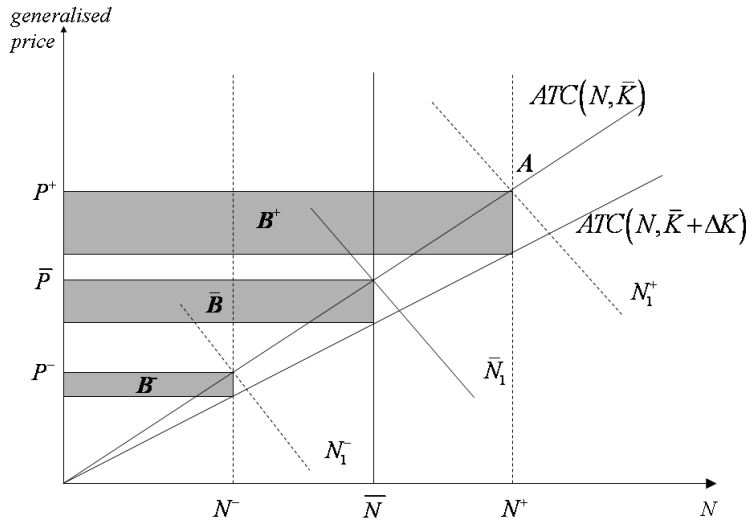

Figure 1: Uncertain demand levels call for larger capacities if demand is price inelastic

We will now show that in this case it is beneficial to invest more than $\bar{K}$ when, instead of the expected demand level $\bar{N}$, one takes into account the explicit distribution of demand levels N^- and N^+. We illustrate this by demonstrating that the expected savings in users' costs of a small increase in capacity ΔK are higher when explicit account is taken of the two possible demand levels N^+ and N^-. The equilibrium time cost when the demand level would be $\bar{N}$ equals $\bar{P}$, the corresponding equilibrium user prices equal P^- and P^+ if demand turns out to be low or high. Consider now the savings in users' costs when capacity is increased by ΔK. The savings in users' costs now equal B^+ if demand is high and B^- when demand is low. The weighted sum of these areas $[0.5B^+ + 0.5B^-]$ is clearly larger than the gross benefits $\bar{B}$ associated to the use of the expected demand level only. Hence marginal benefits of capacity extension are clearly larger when the uncertainty of demand is explicitly taken into account. As can be seen, the benefits of a capacity extension are a quadratic function of the level of realized demand and this explains that it is beneficial to "overinvest" compared to the expected demand level.

In this paper we show that this result holds more generally whenever the price elasticity of demand is smaller than one. Why this is the case can be

3

easily shown in Figure 1 by rotating the demand function in point A to the price elastic function N_1^+. A more price sensitive demand level implies that the realized demand increases when capacity is extended as demand reacts to the lower time costs. At the same time the savings in user benefits become smaller: demand is larger but the savings per user become smaller. When the elasticity of demand is larger than one, the reduction in expected user cost savings becomes smaller than the user cost savings associated to the expected demand level. It then becomes beneficial to underinvest compared to the expected demand level.

Let us now turn to the two mode case. The typical problem we have in mind is the investment in a new rail line for passengers between two cities. Creating a new rail line requires investments long before the service is operational. If both cities have an airport it could be easier to set up a direct air connection between the two cities. So in this example, air would be the second (competing) mode.

The simplest case is where both modes are perfect substitutes (only the average user cost matters), where total demand for both modes is price inelastic but uncertain and where all modes are priced at marginal social cost (MSC). The mode with the long time lag has a fixed infrastructure cost. Its variable infrastructure cost is proportional to the level of congestion. The flexible mode is characterized by an average time cost that is constant.

We can now start with a graphical illustration that is identical to Figure 1. In Figure 2 we represent the average time cost ($ATC(N, \bar{K})$) and the social marginal time cost ($MSC(N, \bar{K})$) and concentrate our attention first on the expected total demand level $\bar{N}$. In contrast to Figure 1, we assume perfect pricing of the rail mode: the user price consists of the average time cost plus a charge that equals the extra time losses imposed on others. Let $\bar{K}$ be the capacity level that minimizes the user costs associated to the expected demand level $\bar{N}$.
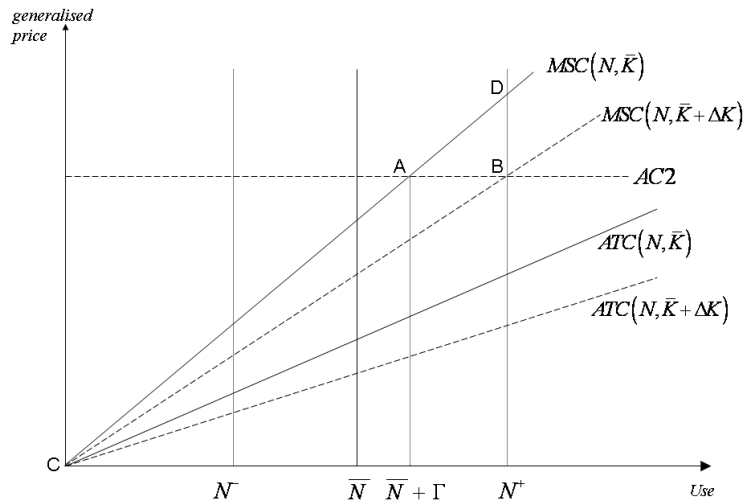

Figure 2: Uncertain demand levels in the two mode case.

4

We now bring in the second mode by drawing an average cost function $AC2$. We have drawn an $AC2$ curve such that the social marginal cost of the rail mode $\left(MSC(\bar{N}, K)\right)$ is lower at a demand level $\bar{N}$. In this case there are two possible solutions: either the fixed cost of rail is relatively high and it is better to rely completely on the second mode (air), or it makes sense to invest in rail and then rail takes the whole market. This is a first simple result that is useful to address the case of uncertain demand.

We now introduce a demand level that can be either low (N^-) or high (N^+) and return to our question whether "overinvestment" in capacity can make sense. Contrary to the one mode model with uncertain and inelastic demand, one can now show that it is no longer always optimal to "overinvest" in rail capacity. Start with capacity $\bar{K}$, which is the optimal capacity when only the expected demand $\bar{N}$ is taken into account. In that case we assume it is interesting that rail serves the whole market $\left(MSC\left(\bar{N}, \bar{K}\right) < AC2\right)$. When demand is high (N^+) it may be interesting not to extend capacity to $K + \Delta K$ and let from the extra demand $\left(N^+ - \bar{N}\right)$ only a part be satisfied by rail (Γ) and the rest $\left(N^+ - \bar{N} - \Gamma\right)$ by using the second flexible mode. Consider now the benefits of adding extra rail capacity; this will lower the trip costs in the cases where N^- and N^+ materialize. Consider the benefits if N^+ materializes. Then the benefits are limited to the area ABC instead of $ABCD$. The presence of the other mode has taken away part of the congestion reduction benefits we observed in the one mode case.

3 A brief review of the literature

There is a large body of literature on investment decisions under uncertainty. There is the option theory that attaches an extra value to any project that can be adjusted or postponed until more is known about demand. This theory is now the basis for investments in firms (Dixit and Pindyck [4]) and is also used in cost benefit assessments (Graham [7]). Another related debate is the use of a higher discount rate in function of the riskiness of the project. Here we take a social cost benefit approach and in this case the risks of a transport project are small and diversifiable, so that the risk free discount rate makes sense (Arrow and Lind [2]).

When it comes to investments in transport, there are two strands of literature. We classify them in function of the type of demand uncertainty. In the first strand demand is "stochastic", in the sense that users, making their decision to make a trip or not, do not know the aggregate level of demand or the available capacity and therefore face an uncertain user cost. Kraus [8] has a simulation model with uncertainty in individual demand, risk averse individuals and ex ante optimal tolling, and finds that highway capacity could be 3 to 12% higher. In [5], D'Ouville and McDonald use an analytical model to treat the same question. They have optimal ex ante tolls, no risk aversion and also find that optimal capacity is always larger relative to the mean level of demand than in the case of no uncertainty.

Arnott, de Palma and Lindsey [1] use a model with uncertain demand and capacity levels, a constant elasticity of demand function and no tolls. They have one result which is of interest to us. With perfect information for the users, the optimal capacity level is higher in the uncertainty case than in the certainty

case if the demand elasticity is lower than one.

The second strand of the literature deals with "macro-economic uncertainty" in the level of demand. Ex ante, demand levels are unknown to the planner but once the investment has been made, the level of demand and user's costs are known to all users. This is the problem we want to address.

Saphores and Boarnet [9] deal with the optimal investment time for a congestion relief investment in a city where the population growth is uncertain. They find that, when the uncertainty is high and the lead times are long, it may be better to advance the investment time compared to a traditional cost benefit analysis. Our model uses a simpler representation of uncertainty than theirs; they use a geometric Brownian motion formulation, while we use a simple demand uncertainty representation with a high and a low value. Other differences are that we explore the role of congestion pricing for the capacity decision and that we only deal with the choice of the capacity level and not with the timing.

4 One mode model

4.1 Model structure

We have opted for a simple model that can be solved analytically. The (generalized) average user cost before toll τ, capacity cost and demand functions are assumed to be:

$$C\left(N,K\right) = \theta\left(\frac{N}{K}\right)^{s} \tag{1}$$

$$F\left(K\right) = kK^{\varepsilon} \tag{2}$$

$$N\left(P\right) = q_{0}P^{-\eta} \tag{3}$$

where N is the volume, P the generalized price, K the capacity of the mode, θ value of time of the (homogenous) users, the parameter $s \geq 0$ denotes the elasticity of user cost with respect to the volume-capacity ratio, $\varepsilon \geq 0$ is the elasticity of construction costs and $\eta \geq 0$ the demand elasticity. The model defined in (1), (2) and (3) has been used previously by de Palma and Lindsey [3] to study cost recovery properties of tolls. The average user cost function (1)[1] becomes independent of usage and capacity when $s = 0$, for high values of s we have a user cost function that increases strongly once one hits the capacity K. The average cost of capacity (2) can be decreasing in K ($\varepsilon < 1$), constant ($\varepsilon = 1$) or increasing in K ($\varepsilon > 1$). In order to make capacity costs comparable to the user benefits we make two assumptions. First, the cost of capacity has been corrected with a discount factor to correct for the lag between investment and the realization of demand. Second, we measure capacity costs as a rental cost per unit period using an annuity of the capital cost. The demand function given in (3) is a constant elasticity function where η can vary between 0 and ∞.

---

[1] One could generalise the user cost function by adding a constant term. This would not affect Proposition 1.

4.2 No Uncertainty

4.2.1 Equilibrium and optimal use

It is well known that, in the presence of congestion, it is optimal to charge a toll that equals the marginal external congestion cost. Since optimal congestion tolling may not be a realistic option because of implementation costs or political opposition we will consider the case with and without optimal tolls. We can in general write for the toll

$$\tau = f \cdot C_N N, \tag{4}$$

where subscripts denote a derivative e.g. $\frac{\partial C}{\partial N} \equiv C_N$. When $f = 0$, we have no tolling and when $f = 1$ the toll is set optimally, any non optimal toll can be considered by using $f \neq 1$.

Equilibrium usage is determined by the condition that the willingness to pay of the last user equals the users cost, that equals the time cost plus the toll:

$$P(N) = C(N, K) + \tau. \tag{5}$$

Given (3), (1) and (4), this solves for N :

$$N^*(K) = \left[(q_0)^{\frac{1}{\eta}} A K^s \right]^{\frac{\eta}{s\eta+1}}, \tag{6}$$

where $A \equiv \frac{1}{\theta(1+fs)}$. Usage is a positive function of capacity and a negative function of the toll (f).

4.2.2 Optimal capacity

In our simple partial equilibrium model, welfare is equal to the area below the demand curve minus the total social cost:

$$\Omega = \int_{n=0}^{N} P(n) \, dn - C(N, K) N - F(K). \tag{7}$$

The first order condition for optimal capacity is:

$$\frac{\partial \Omega}{\partial K} = -C_K N - F_K + [P(N) - C(N, K) - C_N N] \frac{\partial N}{\partial K} = 0 \tag{8}$$

Substituting (5) and (4) the first order condition can be rewritten as

$$\underbrace{-C_K N - F_K}_{(1)} + \underbrace{(f-1) C_N N \frac{\partial N}{\partial K}}_{(2)} = 0 \tag{9}$$

Term (1) in (9) represents the savings in user costs of an extra unit of capacity minus the marginal capacity cost. With optimal pricing ($f = 1$), the second term (2) dissapears. Term (2) in (9) incorporates the effects of induced demand. Suppose now that usage is under-priced ($f < 1$). Given $\frac{\partial N}{\partial K} > 0$, term (2) is negative which might suggest that the marginal benefits of capacity expansion are reduced (since $-C_K N > 0$), as a result the second-best capacity would be less than first-best capacity. However, with a suboptimal toll usage, N and C_K

in term (1) are greater than in the first best. The net effect of these opposing effects is ambiguous a priori.

Solving the first order condition with respect to the capacity yields:

$$K^* (N) = B^{\frac{1}{s+\varepsilon}} N^{\frac{s+1}{s+\varepsilon}}, \tag{10}$$

where $B \equiv \frac{\theta s}{\varepsilon k} \frac{1+\eta f s}{1+\eta s}$.

Solving (6) and (10) simultaneously, we obtain the following expressions for the optimal usage and capacity in function of the parameters $f, \theta, \varepsilon, s$ and η that are included in A, B and C.

$$N^* = \left[B \left[A (q_0)^{\frac{1}{\eta}} \right]^{\frac{(s+\varepsilon)\eta}{s\eta}} \right]^{\frac{s\eta}{C}} \text{ and} \tag{11}$$

$$K^* = \left[B \left[A (q_0)^{\frac{1}{\eta}} \right]^{\frac{(s+1)\eta}{s\eta+1}} \right]^{\frac{s\eta+1}{C}}, \tag{12}$$

where $C \equiv (s + \varepsilon) + s\eta (\varepsilon - 1)$. We see that the optimal usage and capacity are decreasing in capacity cost k. Depending on the values of s, ε and η, the optimal capacity for non-optimal tolling will be either larger or smaller than the first-best capacity.

4.3 Uncertainty on demand

In our model we deal with an uncertainty of demand that is long term or macro-economic in nature. Economic growth and location of firms are not perfectly known at a horizon of 5 to 15 years and this means that the aggregate demand level is unknown. We assume, however, that when the infrastructure has been constructed, the level of demand, congestion and the (perfect or imperfect) toll are known by everybody and there is no remaining uncertainty for the users. This is the justification for using utility functions without risk aversion. This contrasts with the models of Kraus [8], D'ouville and McDonald [5] that deal mainly with "stochastic" uncertainty at the level of the individual user. Their users decide while not knowing what is the available capacity. Arnott, et al.[1] deals with the two cases: what they call "perfect information case" is useful for our problem, their imperfect information case deals more with the "stochastic" demand.

The simplest way to model demand uncertainty is to assume that demand can take only two values (low or high)[2][3]. With probability $0 \leq \mathcal{P} \leq 1$ the demand function is given by:

$$N (P) = q_0^- P^{-\eta} \tag{13}$$

with probability $1 - \mathcal{P}$ it is given by:

$$N (P) = q_0^+ P^{-\eta} \tag{14}$$

---

[2] We could generalise our results by introducing continuous probability distributions. See [1] for an illustration.

[3] We use a model formulation where the demand level for a given price is uncertain. An alternative formulation would start from uncertain willingness to pay levels for a given price. Would one obtain the same basic insights? The answer is yes but one has to take care of the effect of the model calibration.

where $q_0^+ > q_0^-$, so that, for equal user cost, demand will be higher. The "expected" demand function is

$$N(P) = \bar{q}_0 P^{-\eta} \tag{15}$$

with

$$\bar{q}_0 \equiv \mathcal{P} q_0^- + (1 - \mathcal{P}) q_0^+ \tag{16}$$

We are interested in comparing two solutions. One where no account is taken of the uncertainty range and where an average ("expected") demand function is used. In the second approach (the "expected welfare" approach) we explicitly take into account the range of demand functions that can materialize. In the following sections we compare the capacity levels that are optimal when only the expected demand functions are taken into account, with capacity levels that are optimal when explicit account is taken of the range of potential demand realizations.

4.3.1 Expected demand function approach

The optimal usage $\bar{N}^*$ and optimal capacity $\bar{K}^*$ when the expected demand function is used, are given by (11) and (12), with $\bar{q}_0$ instead of q_0 :

$$\bar{N}^* = \left[B \left[A \left(\bar{q}_0 \right)^{\frac{1}{\eta}} \right]^{\frac{(s+\varepsilon)\eta}{s\eta}} \right]^{\frac{s\eta}{C}} \quad \text{and} \tag{17}$$

$$\bar{K}^* = \left[B \left[A \left(\bar{q}_0 \right)^{\frac{1}{\eta}} \right]^{\frac{(s+1)\eta}{s\eta+1}} \right]^{\frac{s\eta+1}{C}}. \tag{18}$$

4.3.2 Expected welfare

If instead of using the expected demand function, one can maximize the expected welfare with respect to N and K, taking explicitly the two realizations into account. The expected welfare is given by:

$$
\begin{aligned}
\Omega = \ & \mathcal{P} \left[\int_{n=0}^{N} P(n) \, dn - C(N, K) N - F(K) \right]_{q_0 = q_o^-} \\
& + (1 - \mathcal{P}) \left[\int_{n=0}^{N} P(n) \, dn - C(N, K) N - F(K) \right]_{q_0 = q_0^+}
\end{aligned} \tag{19}
$$

Solving the first order condition with respect to K yields

$$\hat{K}^* = \left[B \left[\mathcal{P} \left[A \left(q_0^- \right)^{\frac{1}{\eta}} \right]^{\frac{(s+1)\eta}{s\eta+1}} + (1 - \mathcal{P}) \left[A \left(q_0^+ \right)^{\frac{1}{\eta}} \right]^{\frac{(s+1)\eta}{s\eta+1}} \right] \right]^{\frac{s\eta+1}{C}} \tag{20}$$

Proposition 1 *If there is uncertainty in demand , then the use of an expected demand function approach leads to too low capacity levels if the demand elasticity is smaller than one ($|\eta| < 1$).*

Proof. The statement that the use of an expected demand function approach leads to too low capacity levels corresponds to stating that

$$\bar{K}^* < \hat{K}^*$$

rewriting $\bar{K}^*$ given in (18) as

$$\bar{K}^* = \left[BA^{\frac{(s+1)\eta}{s\eta+1}} \left(\mathcal{P}q_0^- + (1-\mathcal{P}) q_0^+ \right)^{\frac{s+1}{s\eta+1}} \right]^{\frac{s\eta+1}{C}},$$

substituting $\alpha \equiv \frac{(s+1)\eta}{s\eta+1}, x \equiv q_0^-$ and $y \equiv q_0^+$, this becomes

$$\bar{K}^* = \left[BA^\alpha \left(\mathcal{P}x + (1-\mathcal{P}) y \right)^{\frac{a}{\eta}} \right]^{\frac{s\eta+1}{C}}.$$

Making the same substitutions in eq(20) gives

$$\hat{K}^* = \left[BA^\alpha \left[\mathcal{P}(x)^{\frac{a}{\eta}} + (1-\mathcal{P}) (y)^{\frac{a}{\eta}} \right] \right]^{\frac{s\eta+1}{C}}$$

Using the above two expressions, the inequality $\bar{K}^* < \hat{K}^*$ is equivalent with

$$[\mathcal{P}x + (1-\mathcal{P}) y]^{\frac{a}{\eta}} < \mathcal{P}x^{\frac{a}{\eta}} + (1-\mathcal{P}) y^{\frac{a}{\eta}}$$

Since $0 \leq \mathcal{P} \leq 1$ this is the same as stating that the function $f(x) = x^{\frac{a}{\eta}}$ is convex. Since x is always positive, the function f is convex whenever its second derivative is positive, which is true if and only if $\frac{a}{\eta} > 1$, or $\eta < 1$. Thus, we can conclude that if $\eta < 1$ then $\bar{K}^* < \hat{K}^*$. ∎

It is useful to analyze a few special cases in order to understand the intuition behind this result.

If $s = 1$ (linear congestion function) and $\eta \to 0$ (inelastic demand) then we see that the marginal benefits of an increase in capacity will increase more than proportionally with the realized demand level. In this case it is clear that it will be optimal to invest more than the optimal capacity for the expected demand level $\bar{N}$. This was illustrated in Figure 1. When demand elasticity is high ($|\eta| > 1$), there is a need to invest less than the expected demand function would tell us to do. The reason is simple; a high (generalized) price elasticity means that it is easy to reduce the number of trips (substituting to other modes etc.). This implies that a capacity shortage, when demands happens to be high, has only a small welfare cost.

It is important to note that our Proposition 1 holds, regardless of the pricing regime since f can take any value. Of course the value of f will affect usage and optimal capacity, but it will do this in a consistent way. One can, however, conjecture that with optimal pricing ($f = 1$) the welfare loss of non-optimal capacity levels will be lower because tolls are a more efficient rationing device than congestion costs.

Our Proposition 1 is in line with Theorem 1 of Arnott et al [1]. They also find that optimal capacity is larger when demand uncertainty is explicitly taken into account. Their result is more general in that they use a more general form of demand uncertainty. Our result is more general in that it holds for any tolling policy and also for any iso-elastic capacity cost function.

5 Analyzing a two-mode problem with demand uncertainty

One often faces investment problems where two modes compete. A typical example is the rail-air competition where both rail and air compete for the same travellers (N) going from one city to another. While air travel is supplied by airlines that can easily adapt capacity by hiring extra planes, the supply of rail is characterized by an important fixed cost and requires long lead times.

The fact that for rail there is a large fixed cost and a long investment lag while air can adjust its capacity fairly easily (provided that there is already some airport infrastructure) adds an additional feature; it implies that rail has to decide first whether it wants to invest or not and if it does, how much. A similar problem exists in freight transport where the construction or improvement of inland waterways has a long lead time while the capacity of road freight can be more easily adjusted at least when the road network is dense. Throughout the rest of this text we will use the rail-air interpretation.

We assume that the cost function for rail is given by

$$CC_r = K_F + l_r n_r + kK \tag{21}$$

where K_F are the fixed costs, l_r the constant variable cost per passenger, n_r the number of passengers, k is the constant average cost of capacity and K the capacity (tracks, trains etc.). The combination of a high fixed cost and a constant marginal capacity cost allows us to approximate cost functions with increasing returns to scale[4].

For air, we assume that the fixed costs can be neglected and that the variable costs per passenger (l_a) are constant. The cost function is given by:

$$CC_a = l_a n_a \tag{22}$$

We use the simplest demand specification. Demand is fixed (unlike the one mode case) at level N and users select the mode with the lowest generalized price (Wardrop principle).

We are interested in how demand uncertainty will influence investment decisions in rail. To obtain a better insight into the problem we will start with the case where there is no uncertainty.

5.1 No uncertainty case

5.1.1 User equilibrium

We assume the same functional forms for the user cost of rail as in the previous section but take the special case that $s = 1$ (linear in volume capacity ratio). The generalized price for rail use is then,

$$UC_r = \alpha + \frac{\theta n_r}{K} + \tau_r \tag{23}$$

---

[4]The cost function for rail is highly simplified here. Compared to other modes like air and road it has important economies of scale in density. The simpliest representation is a linear formulation with fixed costs.

where α is a mode specific user cost. We assume that rail is optimally priced, i.e. toll equals the resource cost plus the social marginal external congestion cost, i.e.

$$\tau_r = \frac{\theta n_r}{K} + l_r \tag{24}$$

By assumption, in air transport there is no congestion and the users' costs for air are simply equal to the fee charged by the airline (τ_a). Because we assume that average costs for air transport supply are constant, we can assume that there is perfectly competitive pricing[5] as well:

$$UC_a = \tau_a = l_a \tag{25}$$

We assume perfect substitutability between the two modes and assume that in equilibrium, users' costs are equalized across modes. The market allocation between modes will depend on the difference between the fixed user cost of both modes. Let $L \equiv l_a - l_r - \alpha$, if the average cost for air transport is lower than the fixed users costs for rail ($L < 0$) everyone will travel by air ($n_r = 0$). For high average costs for air transport $\left(L > \frac{2\theta N}{K} \right)$, rail will serve the whole market ($n_r = N$). For values in between $\left(0 \leq L \leq \frac{2\theta N}{K} \right)$, both modes will attract travellers: $n_r = \frac{LK}{2\theta}$ and $n_a = N - n_r$.

5.1.2 Optimal rail capacity

There is one interesting result in our certainty case: if it makes sense to invest in rail it is always beneficial to serve the whole market.

Proposition 2 *Assuming full substitutability between modes, fixed total demand, and given the cost functions for the two modes defined in (21) and (22), if it is beneficial to invest in the high fixed cost mode (rail), it is optimal to serve the whole market.*

The proof of this proposition is given in Appendix A.

The intuition can be understood as follows. Forget about the fixed cost for the moment. The best rail capacity is the capacity which minimizes the average time cost plus the variable capacity cost. This total optimised variable cost is constant for the capacity function we selected. Consider now total demand N, for this demand either the total optimised variable cost is sufficiently lower than the average variable air transport cost so that the rail advantage can cover the fixed rail cost or this is not the case. This implies that we have either rail or air covering the whole market.

The following proposition tells us under which circumstances it makes sense to invest in rail and if this is the case, what the optimal capacity level is.

Proposition 3 *Under the same assumptions as in Proposition 2 it makes sense to invest in rail if and only if*

$$k < \frac{L^2}{4\theta} \tag{26}$$

---

[5] One could introduce monopolistic pricing by airlines but this would distract our attention away from the central issue: the optimal level of capacity under uncertainty.

and

$$K_F < \left(L - 2\sqrt{\theta k}\right) N \tag{27}$$

If both conditions are satisfied, then the optimal level of capacity is:

$$K^* = \sqrt{\frac{\theta}{k}} N \tag{28}$$
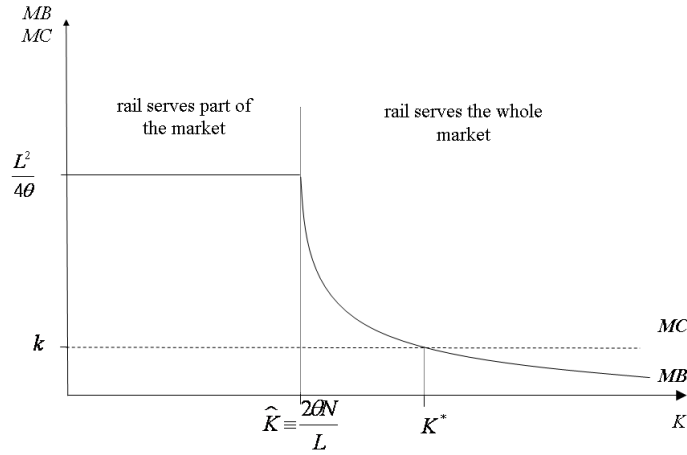
For a proof, see Appendix B.

Figure 3: Marginal benefit (solid line) and marginal cost (dashed line) of a capacity extension.

In Figure 3 we represent the marginal cost and marginal benefits of a capacity extension for rail. We need $k < \frac{L^2}{4\theta}$, before any investment in rail makes sense as air will otherwise always be cheaper. For any k that is smaller than $\frac{L^2}{4\theta}$, the optimal level of capacity increases with a lower k. For an optimal choice of K (covering the whole market with rail) the savings in total variable cost for rail and air has to be sufficiently high so that the fixed rail cost can be compensated.

5.2 With Uncertainty

Suppose that there are two possible realizations of N : with probability $\mathcal{P}$ we have $N = N^-$, and with probability $(1-\mathcal{P})$ we have $N = N^+$ where $N^+ > N^-$. The expected number of users is equal to: $\bar{N} \equiv \mathcal{P}N^- + (1-\mathcal{P})N^+$.

Investment decisions will be made before knowing the realization of N. Once the investments are made, demand is observed and prices are set with perfect knowledge of demand levels. We are interested in finding the welfare gains a planner can achieve when he explicitly uses the information that demand is

uncertain rather than to use the expected demand level only. We will show that, although we work with an inelastic demand assumption, the result of Proposition 1, that overinvestment is beneficial, does not carry over to our two mode setting. The main reason is that in the unexpectedly high demand case, calling upon the second mode may be more economical than investing in extra capacity.

5.2.1 Expected demand approach

If the social planner takes the expected demand as only information, it is clear that the results will be the same as before without uncertainty and with $N = \bar{N}$ we have that: if and only if

$$k < \frac{L^2}{4\theta} \tag{29}$$

and

$$K_F < \left(L - 2\sqrt{\theta k}\right)\bar{N} \tag{30}$$

then rail will invest and will do so until

$$\bar{K}^* = \sqrt{\frac{\theta}{k}}\bar{N} \tag{31}$$

5.2.2 Expected welfare approach

The expected welfare is given by

$$EW\left(K\right) = \mathcal{P}W\left(K\right)|_{N=N^-} + \left(1-\mathcal{P}\right)W\left(K\right)|_{N=N^+} \tag{32}$$

We can distinguish three cases (or regions) which are depicted in Figure 4, where the solid line represents the user cost for rail and the dotted (red) line represents the user cost for air (note that both functions are shifted downwards by a constant factor $\alpha + l_r$).
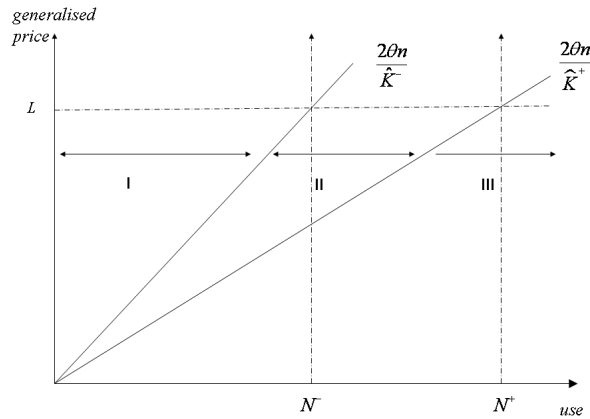

Figure 4: The three possible regions.

14

In the first case (region I in Figure 4), we have that $K < \frac{2\theta N^-}{L} \equiv \hat{K}^-$. In this case, rail will only serve part of the market whether or not demand is low. Uncertainty of demand would thus have no influence on the results. We know from Proposition 2 that this can never be optimal. If it is useful to build a new rail line it will take the whole market (so at least N^-) and capacity will be at least $\hat{K}^-$.

In the other two cases, it may be beneficial for rail to invest. In region III of Figure 4 $\left(\text{where } K > \frac{2\theta N^+}{L} \equiv \hat{K}^+\right)$, rail is certain to serve the whole market even in the high demand case, while in region II of Figure 4 $\left(\text{where } \hat{K}^- < K < \hat{K}^+\right)$ rail serves all users if $N = N^-$, but if $N = N^+$, then rail shares the market with air.

The resulting expected welfares are different in both cases:

$$EW^{III}(K) = \mathcal{P}W^H(K)|_{N=N^-} + (1 - \mathcal{P})W^H(K)|_{N=N^+} \quad (33)$$
$$EW^{II}(K) = \mathcal{P}W^H(K)|_{N=N^-} + (1 - \mathcal{P})W^P(K)|_{N=N^+} \quad (34)$$

where $W^H(K)$ (given in (42)) is the welfare when rail serves all users and $W^P(K)$ (given in (39)) is the welfare when rail shares the market with air. Solving the first order conditions for an optimal choice of K gives:

$$K^*_{III} = \sqrt{\frac{\theta}{k}}\sqrt{\left[\mathcal{P}(N^-)^2 + (1 - \mathcal{P})(N^+)^2\right]} \quad (35)$$

as optimal capacity if $K > \hat{K}^+$. Under the constraint that $\hat{K}^- < K < \hat{K}^+$, optimal capacity is:

$$K^*_{II} = 2\theta N^- \sqrt{\frac{\mathcal{P}}{4\theta k - (1 - \mathcal{P})L^2}}. \quad (36)$$

Whether the social planner chooses a capacity equal to K^*_{II} or K^*_{III} will depend on the parameter values of the problem. In Appendix C, we derive the conditions that determine the regime.

Will the social planner invest more or less when he uses the expected welfare approach rather than using the expected demand approach? Using Proposition 1 we see that $K^*_{III} > \bar{K}^*$, and in this case the social planner will overinvest. In this regime the social planner is sure to be able to serve the whole market independently of the final realization of the demand. The intuition behind the "overinvestment" is the same as in the one mode situation. In the second regime the social planner is not sure whether it will have to share the market with air or not. We can show that the planner overinvests (invests more than $\bar{K}^*$) only if

$$k < \frac{L^2}{4\theta}\left[\frac{(1 - \mathcal{P})(\bar{N})^2}{(\bar{N})^2 - \mathcal{P}(N^-)^2}\right]. \quad (37)$$

The higher the probability of having a low demand, the less likely it will be that any overinvestment occurs. If the difference between high and low demand is large, the social planner will be more cautious and overinvestment will again be less likely.
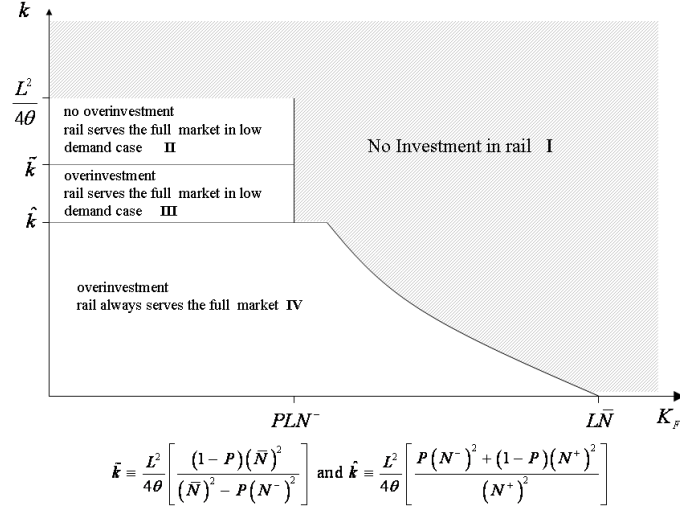
15

$$\bar{k} \equiv \frac{L^2}{4\theta} \left[\frac{(1-P)(\bar{N})^2}{(\bar{N})^2 - P(N^-)^2} \right] \text{ and } \hat{k} \equiv \frac{L^2}{4\theta} \left[\frac{P(N^-)^2 + (1-P)(N^+)^2}{(N^+)^2} \right]$$

Figure 5: Capacity choices in the two mode problem.

5.2.3 Summary

The results of the two-mode model are summarized in Figure 5. The two critical parameters are the variable investment cost k (vertical axis) of the rail mode and the fixed investment cost of the rail mode K_F (horizontal axis). We have 4 types of solutions. First, if the variable and the fixed costs are too high it is better not to invest in rail at all: these are the North, East and North-East regions (zone I) in Figure 5. In this case all traffic is served by the other mode (say air). When it makes sense to invest in rail, one will always invest such as to take the full market in the low demand case. In the second type of solution (zone II in Figure 5) one invests less in the expected welfare approach than in the expected demand approach ($K^* < \bar{K}^*$). This will be the case when the variable cost of capacity is relatively high or when the probability of low demand outcome is high. In the third regime (when the variable cost of capacity is lower, zone III in Figure 5) it makes sense to provide a capacity K^* in excess of what would be optimal for expected demand ($\bar{K}^* < K^* < \hat{K}^+$). In the fourth case (zone IV in Figure 5) it is optimal to provide a level of capacity that always takes the whole market even if the demand level turns out to be high ($\bar{K}^* < \hat{K}^+ < K^*$).

Overall, "overinvesting" in capacity ($K^* > \bar{K}^*$) is not in general better in this model. When the variable cost of capacity is relatively high it is better to serve the "unexpected" part of the market with the other mode that is flexible. As in the one mode model, "overinvestment" can be optimal when the cost of capacity is relatively high because congestion costs remain a convex function of the demand over capacity ratio.

16

5.3 Conclusions

In this paper we have analyzed whether overinvestment in capacity, in the sense of investing more than for the expected demand level, can be beneficial. In both the one mode and the two mode case we found that this can be the case. The main reason is the presence of congestion costs that tend to be convex in the demand over capacity ratio. Overinvestment is not optimal when demand is relatively elastic or when a more flexible substitute mode is available. Even if the full range of possible demand realizations are taken into account, demand uncertainty remains costly when capacity decisions have long lead times. The "overinvestment strategy", studied in this paper, should be compared with other strategies such as investment in improved information on future demand and more flexible capacity extension options.

A   Proof of Proposition 2

Assume that rail invests so that its capacity is insufficient to serve the whole market, in other words $K < \frac{2\theta N}{L} \equiv \hat{K}$, where $\hat{K}$ is by definition the minimum capacity level that decreases the users cost of rail sufficiently to take the whole market. The market share of rail will be equal to $n_r = \frac{LK}{2\theta}$.

Welfare of providing rail is equal to the sum of toll revenues (TR), minus the total cost of providing rail (TC):

$$W^P = CS + TR - TC = \frac{\theta n_r^2}{K} - kK - K_F \tag{38}$$

Substituting n_r;

$$W^P = \frac{L^2 K}{4\theta} - kK - K_F \tag{39}$$

The first order condition for optimal capacity is

$$\frac{\partial W}{\partial K} = \frac{L^2}{4\theta} - k = 0 \tag{40}$$

We see that, as usual, the first order condition tells us that the social planner will invest up to the point where marginal benefits equal marginal costs. Consequently the social planner will choose not to invest when

$$k > \frac{L^2}{4\theta}, \tag{41}$$

since then welfare will be negative.
If, however, k is smaller than $\frac{L^2}{4\theta}$ and since the welfare function W^P is linear in capacity, the social planner will want the capacity to be as large as possible. At a certain point K will become larger then the boundary value $\hat{K}$ and rail will attract all the users.

B   Proof of Proposition 3

If capacity is such that the user costs for rail are always inferior to that of air, or if capacity satisfies $K \geq \hat{K} = \frac{2\theta N}{L}$, then rail will serve the whole market and

$n_r = N$. Welfare is in this case equal to

$$W^H = LN - \frac{\theta N^2}{K} - kK - K_F \tag{42}$$

The first order condition for an optimal choice of capacity is:

$$\frac{\partial W^2}{\partial K} = \frac{\theta N^2}{K^2} - k = 0 \tag{43}$$

We now see that the marginal benefit of an increase in capacity $\left(= \frac{\theta N^2}{K^2}\right)$ decreases with capacity while the marginal cost (k) is constant.
Optimal capacity is

$$K^* = \sqrt{\frac{\theta}{k}} N \tag{44}$$

The condition that it is beneficial to continue to invest beyond the minimum level to have the whole market $(K^* > \hat{K})$ is:

$$K^* = \sqrt{\frac{\theta}{k}} N > \hat{K} = \frac{2\theta N}{L} \Leftrightarrow k < \frac{L^2}{4\theta} \tag{45}$$

which is fulfilled whenever the social planner chooses to invest. Substituting the expression for K^* in the welfare function, we see that it will only be positive if the fixed cost is not too high:

$$K_F < \left(L - 2\sqrt{\theta k}\right) N \tag{46}$$

C   Conditions on capacity cost and fixed cost

The social planner will invest in rail only if two conditions are fulfilled. The first condition requires that the marginal benefits of a capacity expansion should exceed the marginal costs $(MB > MC)$. The second condition implies that the planner will only invest if the welfare is positive $(W > 0)$. While the first condition can be translated into a condition on the capacity cost k, the second condition gives us a condition on the fixed cost K_F.
The marginal benefits of a capacity expansion are different depending on whether or not rail serves the whole market, and can be derived from the welfares given in (39) and (42):

$$MB^P = \frac{L^2}{4\theta}, \quad MB^W = \frac{\theta N^2}{K^2}.$$

Using these we can derive the marginal benefits in the three different regimes:

$$MB^I = \frac{L^2}{4\theta},$$

$$MB^{II} = \mathcal{P}\frac{\theta (N^+)^2}{K^2} + (1 - \mathcal{P})\frac{L^2}{4\theta},$$

$$MB^{III} = \frac{\theta}{K^2}\left[\mathcal{P}(N^-)^2 + (1 - \mathcal{P})(N^+)^2\right].$$

The marginal costs of a capacity expansion are the same for all three regimes: $MC = k$.

Knowing that in order to invest we need the marginal benefits to be larger than the marginal costs. We see that if

$$k > \frac{L^2}{4\theta},$$

the marginal benefits will always be inferior to the marginal costs and it will never be interesting to invest in rail. If

$$MB^{III}\left(\hat{K}^+\right) < k < \frac{L^2}{4\theta},$$

then it will never be beneficial to invest in a capacity that exceeds $\hat{K}^+$, in other words the social planner will not invest as to ensure full coverage of the market in the case of high demand. Finally, if

$$k > MB^{III}\left(\hat{K}^+\right) = \frac{L^2}{4\theta}\frac{\left[\mathcal{P}\left(N^-\right)^2 + (1-\mathcal{P})\left(N^+\right)^2\right]}{\left(N^+\right)^2},$$

then it is so cheap to invest that the social planner will have an interest to invest in rail so that it will always serve the whole market. These conditions are in so far that the welfare is positive, for this to be true the fixed costs may not be too large. The conditions on the fixed cost K_F are:

$$EW^{III}\left(K_{III}^*\right) > 0 \Leftrightarrow K_F < L\bar{N} - 2\sqrt{\theta k \left[\mathcal{P}\left(N^-\right)^2 + (1-\mathcal{P})\left(N^+\right)^2\right]} \quad (47)$$

$$EW^{II}\left(K_{II}^*\right) > 0 \Leftrightarrow K_F < \mathcal{P}LN^-. \quad (48)$$

References

[1] Arnott only , R., de Palma A. and Lindsey L. (1996). Information and Usage of Free-access Congestible Facilities with Stochastic Capacity and Demand. *International Economic Review* 37 (1), 191-203.

[2] Arrow K. J. and Lind R. C. (1970). Uncertainty and the evaluation of public investment decisions, *The American Economic Review* 60 (3), 364-378.

[3] de Palma A. and Lindsey R. (2007). Financing needs and cost structure. Chapter 2 in A. de Palma, Lindsey R., Proost S. (eds.), Investment and the use of Tax and Toll Revenues in the Transport Sector, Elsevier publ.

[4] Dixit A. and Pindyck S. (1999). A Markup Interpretation of Optimal Investment Rules. *The Economic Journal* 109 (455), 179-189.

[5] D'Ouville, E. and McDonald J.F. (1990). Effects of Demand Uncertainty on Optimal Capacity and Congestion Tolls for Urban Highways. *Journal of Urban Economics* 28 (1), 63-70.

[6] Flyberg B., Bruzelius N.and Rothengatter W. (2003). Megaprojects and Risk: an anatomy of Ambition. Cambridge University Press.

[7] Graham (1981). Cost-Benefit Analysis under Uncertainty. *American Economic Review* 71 (4), 715-725.

[8] Kraus M. (1982). Highway Pricing and Capacity Choice under Uncertain Demand. *Journal of Urban Economics* 12 (1), 122-128.

[9] Saphores J-D M & Boarnet M. (2006). Uncertainty and the timing of an urban congestion relief investment. The no-land case. *Journal of Urban Economics* 59 (2), 189-208.