**Title**

On the Design of Public Infrastructure Systems with ElasticDemand



**Author**
Daganzo, Carlos F.

**Publication Date**
2012-02-01

INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

**On the Design of Public Infrastructure Systems with Elastic Demand**

**Carlos F. Daganzo**

WORKING PAPER
UCB-ITS-VWP-2012-2

UC Berkeley Center for Future Urban Transport
A VOLVO Center of Excellence

**February 2012**

# On the design of public infrastructure systems with elastic demand

WORKING PAPER

Carlos F. Daganzo

UC Berkeley Center for Future Urban Transport,
A Volvo Center of Excellence

February 23, 2012

## Abstract

This paper considers the optimization of public infrastructure systems, recognizing that these systems serve multiple user classes. Example application domains include: public transportation systems, electricity distribution grids, urban water distribution systems, and maintenance of pavement and bridge systems. Under the guidance of a policy-making body, the analyst chooses both the system design, including its layout and control, and the prices to be charged for the service. The goal of the optimization is to maximize society's welfare recognizing that the system's performance will in general depend on the system's demand, and vice versa.

The optimization problem is first formulated in its full complexity, where the prices, the demand and the system design are to be determined. It is then shown that if the user classes recognized in the analysis can be priced independently, and if the policy setting body specifies either the demand levels or the generalized prices experienced by each user group, then the problem decomposes into three sub-problems that can be solved sequentially: demand estimation, system design and pricing.

It also turns out that the optimum design can always be obtained by minimizing the generalized cost of the system to society for a known fixed demand, as is conventionally done in practice. The resulting design is independent of how net user benefits are measured.

If the policy-making body does not specify cost or demand targets, and instead assesses benefits by means of consumer surplus then the optimum design is still the solution of a conventional design problem with fixed demand. In this case, however, the demand has to be obtained iteratively using a marginal cost pricing rule.


## 1 The problem

Public infrastructure systems (e.g., a light rail network or an electricity system for a new suburb) are often designed for a fixed target demand. Although this assumption simplifies the resulting optimization problem, it is often questioned: would the optimum system design change if the effects of elastic demand were included as part of the optimization? Could the induced demand and added externalities of a good design negate some of its benefits? These questions are reasonable because the effects of induced demand can be quite significant; see e.g. Downs (1962), Owens (1985), Hansen and Cervero (2002). Therefore it makes sense to include them explicitly in the design framework.

This paper addresses these questions. It unveils conditions under which the the optimum (welfare-maximizing) design for elastic demand is the optimum design for the fixed "optimum" demand. These conditions allow the complex, joint demand estimation/pricing/design optimization problem to be simplified by being broken into parts.

The paper is organized as follows. Section 2, below, formulates the welfare maximization problem in its full complexity, with endogenous demand and a generic net user benefit function. Section 3, shows that if a policy-making body specifies either the generalized prices that should be experienced by the various user groups, or the demand that each group should generate, then the full problem decomposes into parts. Furthermore, the design problem can be treated as if the demand was fixed, by minimizing a generalized cost function for a known fixed demand. Section 4 examines the special case where the policy-setting body does not specify any price or demand targets, and instead stipulates that net user benefits are to be measured by the consumer surplus metric. Finally, Section 5 summarizes the main findings.

# 2 The modeling framework

The general optimization problem is now introduced. Section 2.1 introduces the notation and some basic ideas, and Section 2.2 expresses the mathematical program to be solved.

## 2.1 Definitions

It is assumed that the system's design is encapsulated by a vector $\boldsymbol{x}$ of decision variables. For a transit system, this vector would include the set of routes, stops, schedules, vehicle runs and driver tasks. This list could be time-dependent, for a planning horizon that could span days, months or years depending on the application.

The design is to be evaluated for a set of user classes $\{i = 1, 2, ...I\}$. A user class consists of people from a demographic group that use the system in a similar way. For a transportation system, a user class could consist of all the users that travel by bus from a specific origin region to a destination region during the rush hour of a particular day or a particular year. The number of users in each class is called the demand, and will be expressed as a vector: $\boldsymbol{q} = \{q_i\}$.

To provide and operate the design, the agency (or agencies) will endure monetary costs that depend on $\boldsymbol{x}$ and $\boldsymbol{q}$. It also assumed that the monetary costs of the system's externalities can be expressed as a function of $\boldsymbol{x}$ and $\boldsymbol{q}$. The sum of the external and agency costs will be denoted $S = S(\boldsymbol{x}, \boldsymbol{q})$.

It is also assumed that associated with each user class there is a non-monetary user cost $L_i$ that is expressed in monetary terms. This "level-of-service" cost is also a function of $\boldsymbol{x}$ and $\boldsymbol{q}$. The dependence on demand expresses the fact that level of service may drop if too many users use the system; e.g., roads get crowded during the rush hour, brown-outs may arise during hot weather, etc. Note that the level of service for a specific class is allowed to depend on all the design and demand variables, so the formulation is quite general. The non-monetary user costs for all classes combined will also be expressed in vector form: $\boldsymbol{L} = \boldsymbol{L}(\boldsymbol{x}, \boldsymbol{q})$. It is also assumed that a separate monetary price can be assigned to each user class, and these prices are also expressed as a vector: $\boldsymbol{p} = \{p_i\}$. Finally, the sum of the non-monetary and monetary prices for a user class is the generalized price for the class, $c_i = L_i + p_i$; and these too will be expressed in vector form, $\boldsymbol{c}$.

The net benefit for users is assumed to be a function of the generalized prices and the demand: $U = U(\boldsymbol{c}, \boldsymbol{q})$. This function is notoriously hard to define. Thus, it will be assumed to exist, but no particular form will be attributed to it. The special case where $U$ is the consumer surplus metric will also be discussed.

Finally, it is assumed that the demand that materializes is a function of the design and the system attributes through the generalized prices only; i.e., there is a vector-valued demand function such that $\boldsymbol{q} = \boldsymbol{D}(\boldsymbol{c})$. Again note, that the demand for one user group is allowed to depend on the generalized prices of all the user classes, not just its own. Thus, this demand framework is quite general. It includes as special cases most of the models used in the evaluation of public service systems; e.g., all random utility model forms, and combinations of them.

## 2.2 The optimization problem

The objective is maximizing the net benefit to society, i.e., the welfare, $W$, which is the sum of the user net benefits, $U$, and the net benefit to the rest of society. The latter is the difference between the revenue extracted from the users, $\sum_i q_i p_i$, and $S$. To be determined are: $\boldsymbol{c}, \boldsymbol{q}, \boldsymbol{p}$ and $\boldsymbol{x}$. The welfare maximization problem is therefore:

$$\max_{\boldsymbol{x},\boldsymbol{p},\boldsymbol{c},\boldsymbol{q}} \left\{ W = U(\boldsymbol{c},\boldsymbol{q}) + \sum_i q_i p_i - S(\boldsymbol{x},\boldsymbol{q}) \right\} \tag{1a}$$

$$\text{s.t.} \quad \boldsymbol{q} = \boldsymbol{D}(\boldsymbol{c}) \quad \text{and} \tag{1b}$$

$$\boldsymbol{c} = \boldsymbol{L}(\boldsymbol{x},\boldsymbol{q}) + \boldsymbol{p}. \tag{1c}$$

Note, this formulation is quite general because demand is endogenous, the effects of decreased user experience with crowdedness are captured, and benefits and costs are captured by arbitrary functions. The demand model is quite general too. The main limitation is that there is no constraint on the price vector, which can take positive or negative values, expressing subsidies. This means that user classes can be differentially priced in any way. This level of pricing freedom may be reasonable if the demographic classes are defined by location, time and age, but not reasonable or feasible for finer gradations, e.g., using gender or race. For most design, problems, however, the level of disaggregation cannot be great, for otherwise the optimization problem becomes too complex. So we proceed, but with with this caveat in mind.

# 3 The solution

This section presents two different ways in which optimization problem (1) can be solved, depending on the targets set by the policy-setting body.

## 3.1 The forecasting approach

It is assumed here that the policy-setting body has specified a set of generalized prices that should be achieved, and the goal is to maximize welfare subject to this constraint. Note from (1b) that if $\boldsymbol{c}$ is given ($\boldsymbol{c} = \boldsymbol{c}^o$) then, $\boldsymbol{q}$ is also given ($\boldsymbol{q} = \boldsymbol{q^o} = \boldsymbol{D}(\boldsymbol{c^o})$). Therefore the only remaining variables after this forecasting step are $\boldsymbol{x}$ and $\boldsymbol{p}$, and the first term of the objective function becomes a constant that can be dropped. Furthermore, $\boldsymbol{p}$ can be eliminated from the optimization by inserting (1c) into (1a). This is possible because we can find a posteriori one (or more) sets of prices that will achieve the original generalized prices regardless of the non-monetary prices that arise from the design optimization. The result of the insertion is the following unconstrained objective function for maximization with $\boldsymbol{x}$ as the decision vector: $\sum_i q_i^o(c_i^o - L_i(\boldsymbol{x},\boldsymbol{q^o})) - S(\boldsymbol{x},\boldsymbol{q^o})$. Ignoring the constant term, we can rewrite this as the following minimization problem:

$$\min_{\boldsymbol{x}} \left\{ G(\boldsymbol{x}) = \sum_i q_i^o L_i(\boldsymbol{x},\boldsymbol{q^o}) + S(\boldsymbol{x},\boldsymbol{q^o}) \right\} \tag{2}$$

Note, this is the conventional way in which design problems are often formulated–by minimizing a generalized cost to society for a fixed demand.[1] Clearly, if the generalized cost function $G(\boldsymbol{x})$ is chosen as recommended here, then the conventional way achieves the same result as the original welfare-maximization problem with endogenous demand.

To complete the analysis we need to identify the optimum prices. Let $\boldsymbol{x^o}$ denote the solution of (2) and $\boldsymbol{L^o} \doteq \boldsymbol{L}(\boldsymbol{x^o},\boldsymbol{q^o})$ the optimum non monetary price arising from (2). Then, the optimum

[1] Prices do not appear in the conventional approach because they are just a transfer between members of society.

prices $\boldsymbol{p^o}$ are given by:

$$\boldsymbol{p^o} = \boldsymbol{c^o} - \boldsymbol{L^o}. \tag{3}$$

One nice thing about the described approach is that the user net benefit function does not influence the result and therefore does not have to be known. The generalized price targets are a proxy for this function. In fact, when the policy-setting body specifies the generalized prices it is indirectly stipulating how welfare is to be evaluated. The procedure requires a demand model, $\boldsymbol{D}(\boldsymbol{c})$, however.

### 3.2 The backcasting approach

The sequence in which the analysis is carried out can be changed because most demand models $\boldsymbol{D}$ are surjective mappings; i.e., a generalized price vector always exists for any demand vector. In these cases no generality is lost if we assume that the policy-setting body specified $\boldsymbol{q^o}$ instead of $\boldsymbol{c^o}$, because we could always obtain a suitable $\boldsymbol{c^o}$ by solving an inverse demand problem.

Note, if the demand is specified one can start directly with the design problem (2) to obtain $\boldsymbol{x^o}$; then find a $\boldsymbol{c^o}$ consistent with $\boldsymbol{q^o}$ by solving the inverse demand problem; and finally solve (3) to find $\boldsymbol{p^o}$. The result is the same as if we had used the forecasting approach, starting with the $\boldsymbol{c^o}$ obtained in step two. So the two approaches are equivalent if their two initial data vectors are mutually consistent with the demand model.

If, as before, we use $\boldsymbol{L^o}$ to denote the optimum non-monetary prices obtained in the design step, the last two steps can be reduced to solving for $\boldsymbol{p}$ in:

$$\boldsymbol{q^o} = \boldsymbol{D}(\boldsymbol{L^o} + \boldsymbol{p}). \tag{4}$$

This equation merely states that the system should be priced to achieve the target demand. This is always possible because the demand function is invertible and prices/subsidies can be freely chosen.

What is different with this approach is that it is less dependent on demand modeling. Note that the optimum design is identified first without any modeling to precede it because a target demand rather than a set of generalized prices were specified. This specification is not just analytically simpler, it may also be more likely to happen in practice since policy-making bodies may find it easier to specify a target demand vector than a target vector of generalized prices. This proclivity can perhaps be seen in the behavior of long term transportation planning agencies, which often stipulate demand targets, such as desired modal splits.

The demand model may also be unnecessary to determine the prices in the second step. If the demand in a time period only depends on the prices in the same period, and if these prices can be set in real-time, period by period, then it may be possible to tweak the prices in the real world until the desired demand materializes, ensuring that (4) is satisfied. Of course, a demand model can be used to accelerate convergence of the tweaking process, but lack of a good model model is not an impediment for success. The name "backcasting" alludes to this feature of the approach: one chooses the target demands to optimize the system and then manages it in real time to make sure that they arise in retrospect.

## 4 The special case of consumer surplus as the net user benefit

In some cases, e.g., when the population is not disaggregated into diverse socioeconomic subgroups, it may make theoretical sense to use consumer surplus, $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{U}}(\boldsymbol{c})$, as the metric for net user benefit.

This metric does not exist for all demand functions, however, so the results in this section are a special case.

Consumer surplus only exists for $\boldsymbol{D}(\boldsymbol{c})$ that can be expressed as the gradient of a scalar function, and then it is defined as the negative of this function;[2] i.e.:

$$D_i(\boldsymbol{c}) = -\partial\tilde{U}(\boldsymbol{c})/\partial c_i \ . \tag{5}$$

We assume now that consumer surplus exists and that the decision-making body wishes to maximize welfare using $\tilde{U}$ as the measure of net user benefit, and without stipulating any demand or generalized price targets. Thus, the problem to be solved is (1) with $\tilde{U}$ substituted for $U$.

To do this, insert (1b) and (1c) into (1a) to eliminate $\boldsymbol{q}$ and $\boldsymbol{p}$ from the formulation. The result is the following unconstrained minimization problem in $\boldsymbol{x}$ and $\boldsymbol{c}$:

$$\max_{\boldsymbol{x},\boldsymbol{c}} \left\{ W = \tilde{U} + \sum_i D_i(c_i - L_i(\boldsymbol{x}, \boldsymbol{D})) - S(\boldsymbol{x}, \boldsymbol{D}) \right\}, \tag{6}$$

where it should be remembered that $\tilde{U}$, $D_i$ and $\boldsymbol{D}$ are functions of $\boldsymbol{c}$ alone.

Consider now the first order optimality conditions for (6):

$$\frac{\partial W}{\partial x_n} = -\sum_i D_i \frac{\partial L_i}{\partial x_n} - \frac{\partial S}{\partial x_n} = 0 \ , \quad \forall n, \tag{7a}$$

$$\frac{\partial W}{\partial c_j} = -D_j + \sum_k \frac{\partial D_k}{\partial c_j}(c_k - L_k) - \sum_{ik} D_i \frac{\partial L_i}{\partial D_k}\frac{\partial D_k}{\partial c_j} + D_j - \sum_k \frac{\partial S}{\partial D_k}\frac{\partial D_k}{\partial c_j} = 0 \ , \quad \forall j. \tag{7b}$$

Note condition (7a) is the first order condition of (2) so it is satisfied by the solution of the forecasting approach. Note too that condition (7b) can be rewritten as:

$$\frac{\partial W}{\partial c_j} = \sum_k \frac{\partial D_k}{\partial c_j} \left\{ (c_k - L_k) - \sum_i D_i \frac{\partial L_i}{\partial D_k} - \frac{\partial S}{\partial D_k} \right\} = 0 \ , \quad \forall j. \tag{8}$$

If, as is usually the case, the Jacobian matrix of the demand is non-singular[3] then (8) is satisfied if and only if the quantity in braces vanishes for all $k$. Thus, an equivalent condition to (8) is:

$$(c_k - L_k) - \sum_i D_i \frac{\partial L_i}{\partial D_k} - \frac{\partial S}{\partial D_k} = 0 \ , \quad \forall k. \tag{9}$$

[2]Recall from microeconomics that consumer surplus exists only if the line integral of the demand between two price points is path-independent. In our case, since the demand and price vectors have the same dimension, we know from vector calculus that for the the integral to be path independent the demand function $\boldsymbol{D}(\boldsymbol{c})$ must be the gradient of a scalar function of the prices. Furthermore, if the demand function is differentiable, vector calculus also implies that the Jacobian matrix of $\boldsymbol{D}(\boldsymbol{c})$ must be symmetric. Although this is somewhat restrictive, the property is satisfied by many commonly used demand models. It is satisfied for example if $q_i = D_i(c_i)$, as in this case the Jacobian matrix is diagonal. The property is also satisfied by all demand models derived from random utility, since consumer surplus is in this case the "satisfaction function" expressing the sum of the expected values of the experienced utilities across all users; see e.g., Daganzo (1979), Sec. 4.4.

[3]This is true of demand models derived from random utility choice theory if one assigns zero cost to a no-travel alternative that is not included as part of the demand function.

Define now the following generalized cost function:

$$G(\boldsymbol{x}, \boldsymbol{q}|\boldsymbol{q^o}) = \sum_i q_i^o L_i(\boldsymbol{x}, \boldsymbol{q}) + S(\boldsymbol{x}, \boldsymbol{q}), \tag{10}$$

which weighs the non-monetary user costs functions with the fixed demand, $\boldsymbol{q^o}$. Remembering from (1c) that $c_k - L_k = p_k$, one can rewrite necessary condition (9) as follows:

$$p_k = \frac{\partial G(\boldsymbol{x}, \boldsymbol{q}|\boldsymbol{q^o})}{\partial q_k} \ , \quad \forall k. \tag{11}$$

This simply restates a classical result of microeconomic theory: at optimality the prices for each user group should equal the marginal generalized cost with respect to the demand of the group, taking the weights of each user group as constants.

This suggests that to solve the problem one can use the forecasting method with some initial guess $(\boldsymbol{c^o}, \boldsymbol{q^o})$ to identify a test solution: $(\boldsymbol{x^o}, \boldsymbol{p^o})$. If this solution satisfies (11) approximately, one is done. Otherwise, one can use the the new prices predicted by (11), $\boldsymbol{p^n}$, to compute a new set of generalized prices: $\boldsymbol{c^n} = \boldsymbol{L^o} + \boldsymbol{p^n}$, and use $\boldsymbol{c^n} - \boldsymbol{c^o}$ as a search direction in which to move from $\boldsymbol{c^o}$ and approach a solution.

It is characteristic of design problems of type (2) in the logistics and public transportation fields that both the optimum designs and the cost they produce are rather insensitive to the input data; see e.g., Daganzo (1991). This suggests that in at least some applications, the iterative procedure can be stopped after the first iteration so that the design problem is only solved once.

## 5 Conclusion

The main finding of this note is that if a system designer uses as inputs demand levels given to him or her by a policy-making body, and then uses these inputs to obtain an optimum design by minimizing the system's generalized cost for the given demand, the solution so obtained maximizes welfare. The user net-benefit function does not have to be known to obtain the design: the target demands act as proxy for it. The only caveat to this result is that society must be able and willing to intervene after the design is implemented to ensure that the assumed demand materializes by pricing the various user groups appropriately. If prices cannot be independently charged to each user class then the result does not hold.

In applications where consumer surplus is used as the metric for user benefit, the same procedure can be adapted within an iterative process that converges when the predicted prices match the marginal costs of increased demand to society.

## Conclusion

Research supported by the Volvo Research and Educational Foundations and by NSF Grant CMMI-0856193.

## References

1. Cervero, R. and Hansen, M. (2002) "Induced travel demand and induced road investment: a simultaneous equation analysis" *J. Transportation Economics and Policy* 36(3), 469-490.

2. Daganzo, C.F. (1979) *Multinomial probit: The theory and its application to demand forecasting*, Academic Press, N.Y.

3. Daganzo, C.F. (1991) *Logistics systems analysis*, Springer-Verlag, Heidelberg, Germany.

4. Downs, A. (1962) "The law of peak hour expressway congestion" *Traffic Quarterly* 16, 393-209.

5. Owens, W. (1985) "Transportation and world development" *Transportation Quarterly* 39, 365-375.