System Performance

As it is always easier and in the end less costly to be accurate than inaccurate, the good engineer will always be accurate in all essentials, but he will not waste time in attempting unnecessary precision which does not add appreciably to the final value of his work.

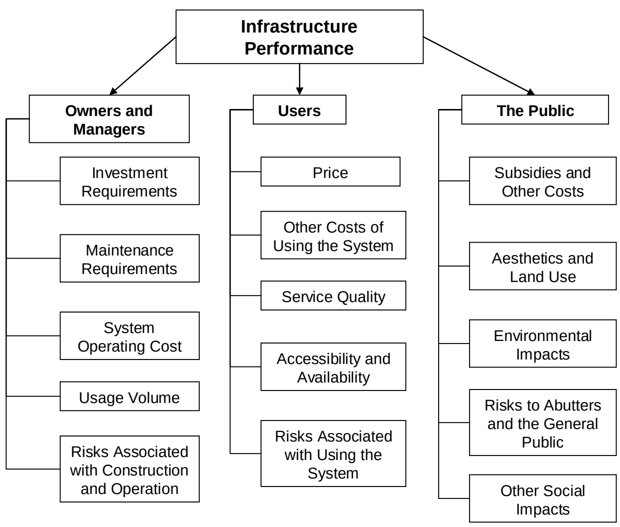
Arthur M. Wellington, **The Economic Location of Railways**, John Wiley & Sons, 1911

Performance of Infrastructure-Based Systems

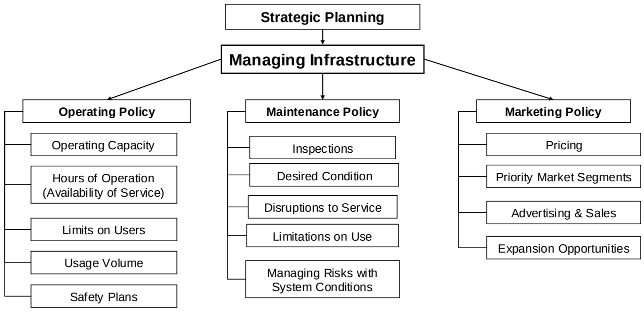
The performance of an infrastructure system cannot be captured by just one or two measures. Cost, quality of service, capacity, safety, environmental impacts and sustainability are all important, while the extent of coverage, accessibility, equity, and appearance can also be critical. Moreover, performance depends upon one's perspective, as owners and managers seek financial rewards, users seek good service at reasonable prices, and the public worries about such things as the need for subsidies, environmental impacts, safety, land use, economic development, and aesthetics (Figure 1). Designing, constructing, and managing these systems involves trade-offs among multiple factors, often with no obvious way of determining which factors are most important. Evaluating proposals for creating new systems or for expanding or modifying existing systems will always require judgment and will often require some sort of political process to determine which aspects of performance should be emphasized.

System performance can be documented, studied, modeled, predicted and managed. If system performance is well understood, and if there is a consensus about the relative importance of the various aspects of performance, then it is possible to provide a clear, objective basis for evaluating new projects and programs. If performance is poorly understood, then research and analysis may be able to clarify the potential trade-offs for various options. If there is no consensus as to which aspects of performance are most important, then research and analysis of performance can at least provide an objective framework for evaluating proposals.

Figure 1 Perspectives on Infrastructure Performance



Performance of infrastructure systems depends upon engineering and managerial issues, such as the nature, condition, and deterioration rates of structures and equipment. Infrastructure managers develop plans and policies that they use to guide operations and maintenance, usually with the hope of increasing profitability or other financial goals. The management structure will often have separate departments responsible for operations, maintenance and marketing, each of which will have their own concerns and ideas about what types of projects will be most beneficial in improving performance. Each department's objectives will ideally reflect strategic plans, which may include financial goals, goals for expanding or shrinking the system, plans for new services or markets, or goals related to service quality, risk management or interactions with the public (Figure 2).



Performance also depends upon many other factors, including the regulatory structure, operating capabilities, and business strategies pursued by the organizations and individuals that make use of the infrastructure and the existence and performance of competing infrastructure systems.

Most importantly, perhaps, performance reflects the demand for the service, especially when usage approaches system capacity. A highway that allows motorists to drive through the city at high speeds in the middle of the night can be more like a parking lot during rush hour. It is a very complex matter to predict the average speed on a highway during rush hour, for that will require some way to estimate demand taking into account the fact that the level of demand will affect performance. It is more straightforward to develop a **performance function** for the highway, i.e. to predict the performance of the highway for any particular level of demand.

Performance functions can be developed for any infrastructure system based upon the engineering characteristics of the infrastructure and the ways that it is used. Developing **engineering-based performance functions** enables planners and entrepreneurs to understand the options for developing or improving systems, to evaluate the impact of proposed projects, to understand the potential benefits of new technologies, and to select and design better projects. Performance functions may or may not be precise, depending upon the state of knowledge of the system and also upon the context for their use. The remaining sections of this essay focus on performance measures important for all infrastructure systems: cost, profitability, service, capacity, safety, and security.

System Cost

Owners and investors are concerned with the initial cost to construct infrastructure, as well as the continuing cost for operating and maintaining the infrastructure. Users are concerned with their costs of using the system, which would include the prices they are charged plus their own time and expense associated with using the system. Basic cost concepts include the following:

- Total cost: what is the total cost of the system for a given level of output?
- Average cost: what is the average cost per unit of output?
- Marginal cost: what is the cost for one additional unit of output?

- Incremental cost: what is the cost for an increment of output?

These concepts can be illustrated using an extremely simple equation expressing cost as a linear function of volume V:

(Eq. 1) Cost = a + b V

(Eq. 2) Average cost = (a + b V)/V = a/V + b

(Eq. 3) Marginal cost = (a + b (V + 1)) – (a + b V) = b

The fixed cost "a" is incurred whatever the volume, and "b" is the marginal cost per additional unit. The average cost per unit declines as the fixed cost is spread over additional volume. In any particular situation, these cost functions will be much more complex. Total cost will be a function of the resources that are used and their unit costs, and the resources that are used will depend upon the quality of service as well as the level of demand.

The fixed cost could well be the sum of many different types of costs that would not vary with changes in volume, such as the salaries of the senior administrative staff, the costs of establishing a maintenance facility, and the costs of acquiring basic equipment and machinery. Variable costs would likely include the labor and energy costs associated with operations. Which costs are fixed and which costs are variable depends upon the level of volume and the time frame under consideration. In the short run, many more costs are fixed, while in the long run, most costs can be affected by restructuring the system. Whatever the time frame, the simple cost function shown above as Equation 1 may (for a reasonable and meaningful range of volumes) actually be a useful approximation of a much more complex cost function. Hence, it is worth taking a closer look at this cost function.

Consider a situation in which following cost function is believed to be approximately valid for volumes V ranging from 10 to 100 units per day:

(Eq. 4) Total Cost = \$50 + \$1 (V)

The fixed cost is \$50, which plots as a straight horizontal line in Figure 3, while the variable cost plots as a straight line with a slope of 1. If this line were extended beyond the range of interest, it would intersect at the origin of the graph.

Figure 3 Plotting fixed, variable and total cost if C = 50 + V

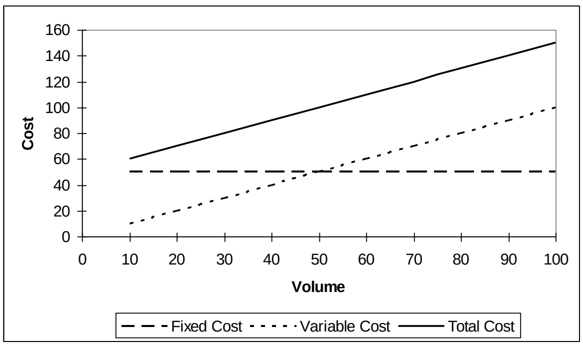
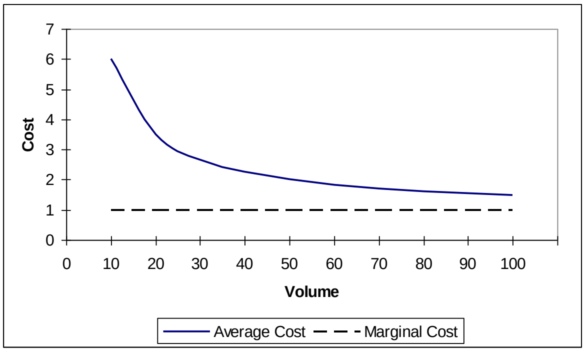


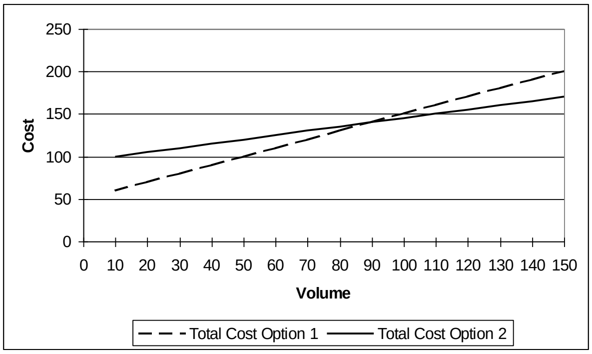
Figure 4 shows the average and variable cost for this cost function. The marginal cost is equal to \$1 over the entire range, while the average cost declines from \$6 when volume is 10 units per day to \$1.50 when demand is 100 units per day. The average cost function is non-linear, and the average cost approaches the marginal cost for high volumes.

Figure 4 Plotting average and marginal cost if C = 50 + V



It is very common to find that there is a trade-off among two options, one of which has higher fixed cost but lower marginal cost. Figure 5 shows a second cost curve in which the fixed cost has increased from \$50 to \$95 per day, but the variable cost has dropped from \$1 to \$0.50. The breakeven point is at 90 units per day: above this level, option 2 is preferred, while below this level option 1 is preferred.

Figure 5 The option with the higher fixed cost is preferred if the volume is greater than 90
(Cost of Option 1 = 50 + V; Cost of Option 2 = 95 + V/2)



For linear functions of the type TC = a + bV, the breakeven point V_b can readily be calculated:

If $\quad TC_1 = a_1 + b_1V$

and $\quad TC_2 = a_2 + b_2V$

then, at the point where the costs are equal for the two technologies, the following equation will hold:

(Eq. 5) $\quad a_1 + b_1V_b = a_2 + b_2V_b$

Solving for V_b yields:

(Eq. 6) $\quad V_b = (a_2 - a_1)/(b_1 - b_2)$

In words, the breakeven volume is calculated as the increase in fixed cost divided by the savings in variable cost per unit. Major CEE projects require investments that are aimed at reducing marginal cost. Larger projects typically provide an opportunity for greater reductions in marginal cost. A key question is whether there will be enough demand to cover the extra costs associated with the larger project. In general, smaller projects will be better until congestion and capacity concerns make a larger project desirable.

> L.F. Loree, Past-President of the Delaware & Hudson Railway and Chair of the Kansas City Southern, described how railroads worked hard to avoid investing unnecessarily in capacity:
>
> *We hang back and postpone as long as possible work to increase facilities the use of which may be increased by ingenuity and method.*
>
> L.F. Loree, **Railroad Freight Transportation**,
> D. Appleton & Co., 1922, p. 61

<u>Non-linear Cost Functions</u>

Linear cost functions are easy to draw, but more complex functions will often be necessary. The same logic applies to plotting total costs, average cost, or marginal costs and the same approach can be used to determine breakeven volumes. Figure 6 illustrates a situation where three technological options are available, each of which can be represented by a non-linear cost function:

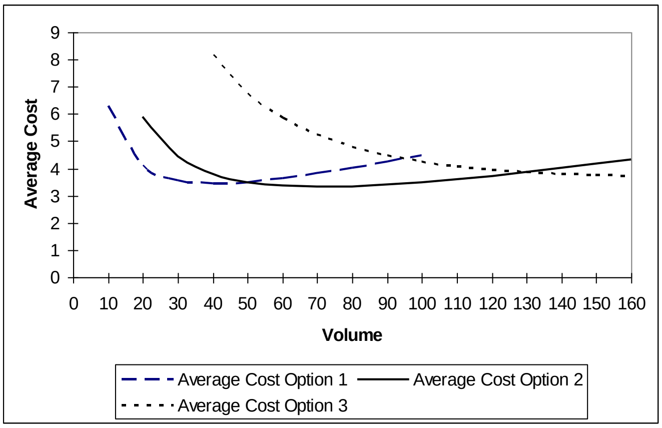
(Eq. 7) Total Cost Option 1 = $50 + V + 0.03V^2$

(Eq. 8) Total Cost Option 2 = $100 + 0.5V + 0.02V^2$

(Eq. 9) Total Cost Option 3 = $300 + 0.25V + 0.01V^2$

The third term in these equations means that the average costs eventually begin to increase, as the contribution of the squared term eventually offsets the savings from spreading the fixed costs over a larger volume. Figure 6 plots the average costs of the three technologies for the ranges of volumes for which they are feasible options. Option 1, the one with the lowest initial cost, has average costs greater than $6/unit for the minimum volume of 10 units; average costs drop below $4/unit for volumes of 30 to 50 units, then rise steadily as volume increases to 100 units, which is the maximum that can be handled by this technology. Option 3 is much too expensive for or ill-suited to low volumes, so the costs for that technology are only shown for volumes of at least 40 units. Figure 6 indicates that Option 1 is favored for volumes less than about 50, while Option 2 is cheapest for volumes between 50 and 130 and Option 3 is favored for greater volumes.

Figure 6 Long-Run Average Costs

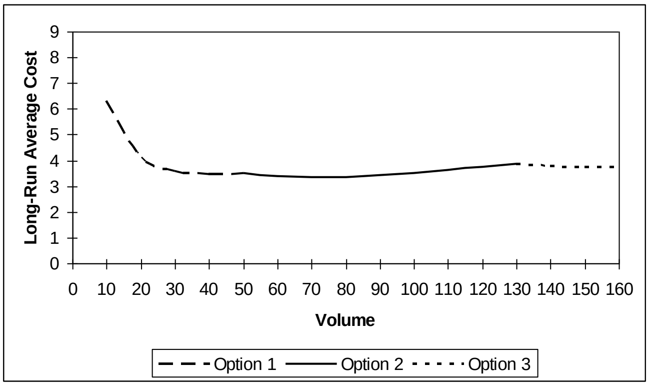


Short-Run vs. Long-Run Average Costs

The three cost curves shown above in Figure 6 each depict **short-run average costs** for a particular technology. Each curve shows the average costs that would result from using a particular technology for a specified range of volumes. They are called "short-run" average costs because they do not allow a shift to a more efficient technology.

It is often useful to understand how average cost of a particular type of process or production would change with volume, assuming that the best technology is used for each level of volume. This cost is called the **long-run average cost**, assuming that in the long-run it will be possible to adjust the production process to what is best for the actual volume served or produced. The long-run average cost curve can be constructed from the applicable short-run cost curves: for each volume, the long-run average cost is equal to the short-run cost of the technological option with the lowest short-run average cost. In other words, the long-run average cost will be the envelope of the minimum short-run costs, as depicted in Figure 7. This is a single cost function that is based upon the three cost functions shown in the previous figure. The long-run average cost in Figure 7 could be described as a little less than $4 for volumes ranging from 20 to 160.

Figure 7 Long-Run Average Costs for Technologies Shown in Figure 2-6



You (the developer, entrepreneur, or planner) know, or should know, your costs and technologies. You therefore should be able to develop an algebraic expression for your costs that accounts for the technological and design options that you may have. It the developer, entrepreneur, or planner doesn't know their costs or technologies, then you (the consultant, the researcher, or the smart young analyst) may be able to do some analysis and create relevant cost functions.

> *There is more than one way to skin a cat.*
>
> American Proverb

Resource Requirements

Costs ultimately are linked to resources: people, land, materials, energy, and capital. There will always be many different ways to use resources in constructing a particular project, and choices will probably be made by developers, engineers and planners seeking to minimize the cost of the project. From their perspective, minimizing cost is purely a financial exercise aimed at minimizing the cash required to complete the project.

The design of a project and the resources required for it will be influenced by both the current and the expected future availability and prices of people, land, materials, energy sources, and capital. If average salaries and wages are expected to rise relative to the cost of capital, then there will be a tendency to use fewer people and more machines. If the cost of energy is expected to rise relative to the cost of building materials, then there will be an incentive to construct buildings that require less energy and to use materials that are less energy intensive. If longer lasting materials are created, they will be used where their longevity provides benefits to a project. Standard methodologies of engineering economics make it possible to decide what system design is best for a project, which materials to use, and which construction techniques to follow, taking into account the initial investment cost and costs to the owners and users over the life of the project.

From an economic perspective, however, the price of a resource may not reflect the true economic cost of using that resource:

- The cost of using a resource is an **opportunity cost**; the use of resources on one project means that they are not being used on other projects or being reserved for future use.
- There may be **negative environmental impacts** associated with using a resource, such as the paving over of wetlands, emission of greenhouse gases in the manufacture of construction materials or the devastation of remote wild areas by mining activities.
- **Non-sustainable use** of renewable resources, such as ground water or timber, will lead to declines in the amount of such resources that can be used in the future.
- Extensive use of non-renewable resources can eventually lead to **resource depletion**.
- Exposure of workers to unsafe conditions or hazardous environments may lead to **risks of injury or serious illness**; unscrupulous companies may be able to avoid covering the costs associated with these problems, and governments may or may not regulate workplace conditions.
- There may be **government subsidies** for certain types of workers or government regulations that require the use of excessive numbers of workers.

Some of these costs may be reflected in the market prices for certain resources, such as the cost of capital and salaries and wages in developed countries. Other costs are notably absent from the prices that are charged, such as the effects of emissions on global warming or the environmental damages related to strip mining. If a company or a country already owns land that could be used for a project, they may not even consider the cost of using that land in the analysis, even though developing that land may forestall even better opportunities in the future.

If the prices for resources actually approximate the costs associated with using those resources, then financial analysis will perhaps produce a reasonable result. If the prices for some resources are markedly above or below their full economic costs, then the financial analysis could result in poor decisions from the perspective of society, even when those decisions do appear to increase the profitability of the project. Because of the discrepancy between financial and economic costs, public policy must enforce rules and regulations that promote or enforce consideration of the true economic costs of a project. Such regulations include:

- Restrictions on development of wetlands and other sensitive environments
- Regulations related to occupational safety and health
- Minimum wages and laws allowing the formation of unions
- Licensing of engineers and others involved in project design and implementation
- Regulations concerning the technologies and methods used in mining and other extractive activities
- Regulations requiring assessment of environmental and social impacts

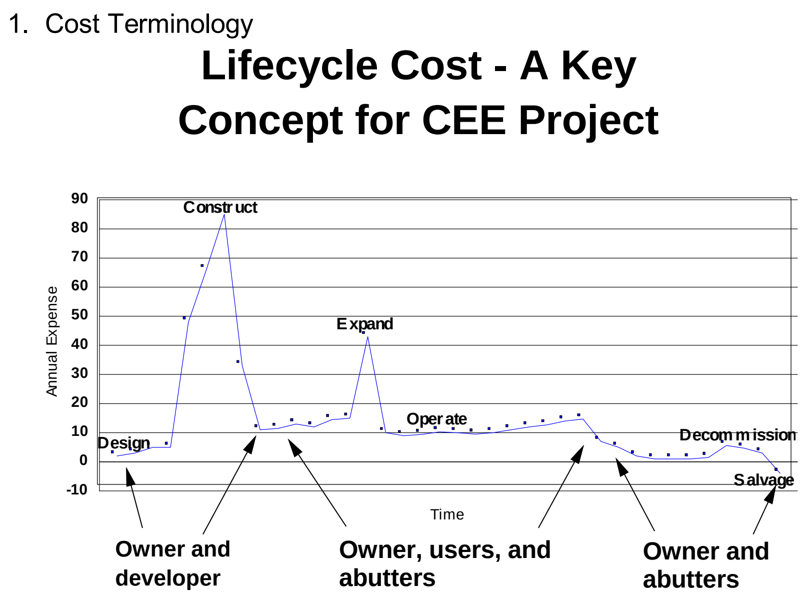
For most privately sponsored projects in developed countries, project evaluation will be strictly based upon a financial analysis. If government regulations are effective, then the private decisions will be reasonably good. If government regulations fail to capture significant aspects of economic costs – as has been the case with many environmental impacts – then private decisions could lead to increasing problems for society.

When evaluating projects, or when estimating costs, it will always be worthwhile a) to detail the actual resources that will be used in addition to how much money is needed and b) to document the methodologies and logic used in reaching decisions about design, choice of materials, and choice of construction methods. As prices of resources change relative to each other, as new technologies become available, and as more economic aspects of costs are recognized, different types of projects and different approaches to constructing those project will be needed. However, the original methodologies and logic may still be suitable for evaluating future projects.

Since many engineers and planners work in places with markedly different levels of development, it is important to recognize that the best solutions in one country can be quite different from the best solutions in another country. This is especially true when comparing systems in a developed country with systems in a developing country, where labor costs are apt to be much lower and capital costs much higher.

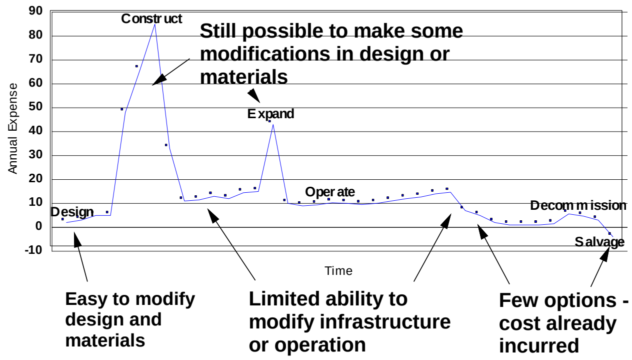
Lifecycle Cost

Since infrastructure lasts a long time, infrastructure design should consider costs over the entire life of the infrastructure. As shown in Figure 8, initial costs of design are likely to be modest compared to the construction costs that follow. Once the project is completed, there will costs of operation that may be borne by the owner, users or abutters. As th ... as a result of declining dema ... nd to salvage metal or whate ... while users and abutters ar



The design phase offers the greatest opportunity to affect the life cycle costs of a project (Figure 9). At this point, while all things will not be possible, many things will be. A major design consideration will be the extent to which the owner or developer considers the costs to users and abutters. Small changes in design conceivably produce more efficient operations or limit the negative effects on third parties, but only to the extent to which such costs are even considered in the design. Mistakes regarding the size of the project – too big, too small, too soon, too little flexibility, too difficult to rehabilitate, too much impact on neighbors – conceivably can be rectified with little or no additional time or expense related to construction. The opportunities for savings will be clearest when the owner will be responsible for operations for the indefinite future; the opportunities for misguided design (or fraud) will be greatest when the developer is interested only in minimizing the construction cost, as operations and maintenance will be the responsibility of others.

Lifecycle Cost - Greatest Potential For Lifecycle Savings is in Design!



Engineering-Based Cost Functions

Three approaches can be used to estimate cost functions: accounting, econometric, and engineering-based. Each approach has its usefulness, but for project evaluation, an engineering-based approach is essential.

Accounting approach: for an existing system, whether an apartment building, a transportation terminal or a wind farm, it should be possible to identify all of the expenses related to the system's construction and operation. If the accounting system is accurate and complete, then the costs associated with each phase of the design, construction and operation of the project should be evident. Complications will arise if there are many different owners and users or if it is difficult to allocate specific costs to specific purposes or users. Special studies may be required to support a realistic allocation of costs or to identify user costs. However, the companies or agencies that manage infrastructure are likely to have very good information concerning their own costs and good estimates of the costs borne by their users.

There are two main problems with using accounting costs for project evaluation. First, accounting costs will not readily relate to the costs of new projects, unless the new projects are very similar to existing projects. Second, accounting systems can provide very rich detail concerning existing operations, but they will not show how costs vary with demand, quality of service, capacity or technology.

Econometric approach: if cost data are available for a variety of completed projects, it may be possible to discern trends in cost as a function of the type or size or location of the project. Econometric analysis can be useful for public policy. For example, if data were available on the cost of constructing apartment buildings in a major city, it would be possible to determine the cost per apartment or the cost/square foot of living space for each project. This data could then be plotted to determine if the cost/apartment and the cost/square foot vary with the size of the project. If large projects involving a hundred or more apartments are markedly cheaper to construct than small projects, then public policy perhaps should be slanted toward facilitating larger projects.

Although the econometric approach can be very useful in determining some basic trends in cost, it is less useful in estimating costs for a particular project. It is very ill-suited toward estimating the costs of projects that use new designs or technologies – and these of course are among the characteristics of many large-scale infrastructure projects.

Engineering-based approach: the preferred method for estimating infrastructure costs is to break a project into well-defined pieces for which it is feasible to estimate unit costs. For each piece of the project, unit costs can be developed based upon past experience, expert judgment, or special studies. These unit costs can reflect new technologies or new designs, so that this approach is not restricted to past experience.

Profitability, Breakeven Volume, and Return on Investment

Profit

Entrepreneurs and owners will be concerned with the financial success of a project. There are three main questions that will be of interest. First, will the project be profitable? Second, will the profit be sufficient to justify the investment that is required? Third, once the project is completed, will it be worth more than it cost to build it? The project will be profitable if the revenues received from the project are sufficient to cover its costs. The revenue could include subsidies from government agencies as well as revenues from users of the project. Revenue from users will depend upon the price that is charged and the value of the project to potential users. For projects that add capacity within a competitive market, such as most real estate projects, the prices that can be charged will rise and fall with market forces. For projects where competition is difficult, such as new bridges or toll roads, the prices can to some extent be established by the owner. In the competitive situation, the question is whether the project can be constructed and operated so that it is possible to achieve a profit given expected market prices. In a monopolistic situation, the question is to choose the prices that will maximize profits, assuming that there is in fact a range of prices that could be profitable.

Let's begin with the simple situation we discussed in the previous section:

(Eq. 10) $\text{Total cost} = a + bV$

If this is a project that will add capacity to a competitive market, then the price P will be determined by the market and the total revenue can be expressed as:

(Eq. 11) $\text{Revenue} = PV$

Profit will be the difference between revenue and cost:

(Eq. 12)) $\text{Profit} = PV – (a + bV)$

Breakeven Volume

If a company has a linear cost function of the type shown above in Figures 1 and 2, then it must sell enough units so that the average cost of production is less than the average sale price. If the sale price is less than the variable cost, then the company will never make a profit. If the sale price is higher than the variable cost, then the company will receive enough cash from the sale to cover the variable cost and have something left over that could go toward covering fixed costs; once sales volume is sufficient to cover fixed costs, then each sale will add to profit. The difference between the sale price and variable cost is called the **contribution to fixed costs or profit** or the **contribution to overhead or profit**:

If $\text{Cost} = a + bV$

and $\text{Price} = P$

Then

(Eq. 13) $\text{Contribution} = P – b$

(Eq. 14) $\text{Profit} = PV – (a + bV) = (P\text{-}b)V – a$

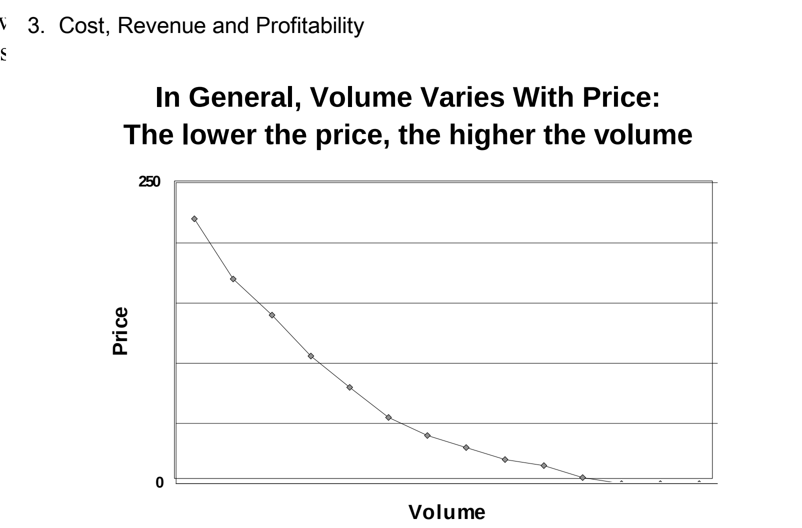
Sales will be enough to generate a profit if the total contribution is enough to cover the fixed costs. The point at which contribution equals fixed cost is known as the breakeven volume V_p relative to making a profit:

(Eq. 15) $V_p = a/(P - b)$

There are two conditions for profitable operation in this simple situation: price must be above variable cost and the volume must be above the breakeven volume. V_p is the volume where the incremental contribution to profit from each unit produced is sufficient to cover its share of the fixed cost.

Volume Varies with Price – the Demand Curve

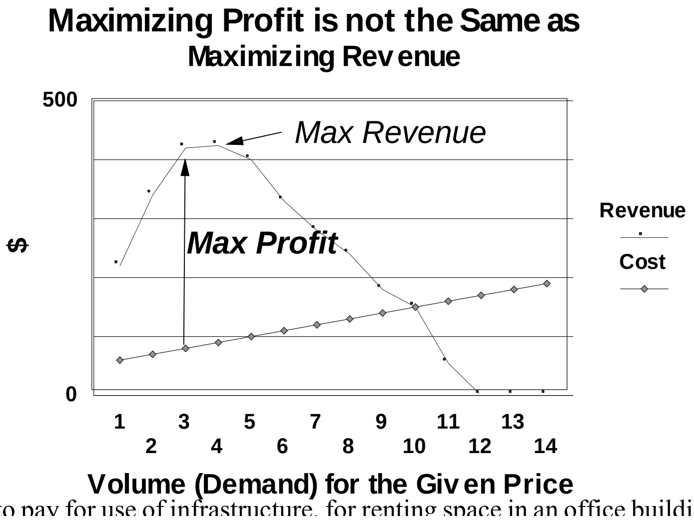
More generally, usage volume will vary with the price that is charged. The higher the price, the lower the volume. This is traditionally represented as a demand function, as shown in Figure 10. Note that price, presumably the independent variable, is shown on the y-axis while volume, presumably the dependent variable, is shown on the x-axis. Faced with a downward sloping demand function, someone trying to decide upon a price for their services is faced with a dilemm nd there won't be any profit, no matter hov n above (Eq. 14), and it will still be necess



The best price – at least from the owner's purely financial perspective - will be the price that maximizes profit. Figure 11 plots profit as a function of volume, where the price is defined by the demand curve shown above. The first point on each curve corresponds to a price of \$225 per unit and a volume of 1 unit per day. The third point corresponds to a price of \$140 and revenue of \$420 for sales of 3 units per day; this is the price that maximizes profit. If price is further lowered to \$110, the volume increases to 4 units per day and revenue rises to \$440 per day, but profit declines. With further price reductions, the total revenue declines. When price drops to \$15, the volume is much higher, but total revenue equals total cost. For even lower prices, the owner loses money. The project therefore would be profitable for any price between \$15 and \$225/unit. Note that the owner would rationally choose to keep prices high, even if there were additional benefits to society to be gained by attracting more users.

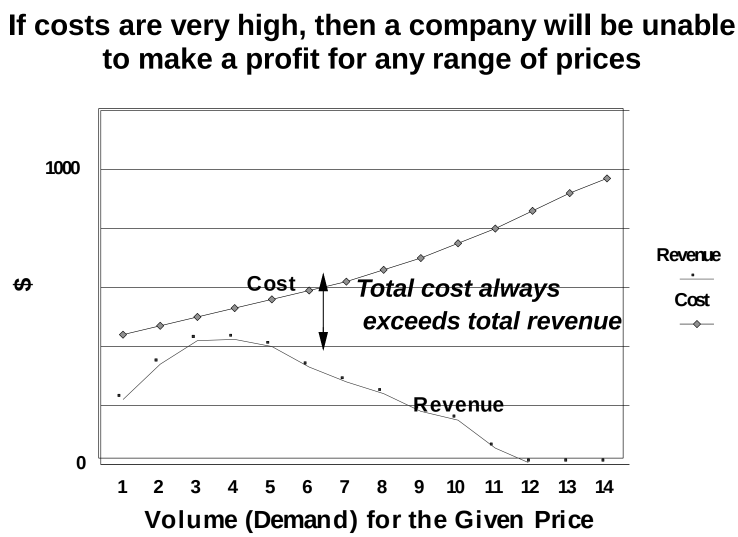
There will not necessarily be any price at which the owner could make a profit, as illustrated by Figure 12. If total cost always exceeds total revenue, then private companies will not provide the service. If the service is deemed to be desirable for society, then the public may decide either to provide the service via a public agency (e.g. public transit) or to subsidize the service so that the private sector will continue in the business (e.g. subsidized housing for low-income families).

3. Cost, Revenue and Profitability



Prices people are willing to pay for use of infrastructure, for renting space in an office building, or for any other service or product w … make a profit. General eco… …ges in building occupancy a…

3. Cost, Revenue and Profitability



Return on Investment

The financial success of a project is often measured by the return on investment, which is the annual profit divided by the initial investment in a project. If someone invests $100,000 in a project that provides a profit of $15,000 per year, then the return on investment is $15,000/ $100,000 or 15%. Usually the return on the investment will vary from year to year, because prices and usage volume will vary with the economy, competitive conditions, and other factors that will affect supply and demand. If return on investment is high, then investors will consider additional investments in similar projects; if returns are low or negative, then investors will turn to other types of projects, leave their money in the bank, or perhaps invest in other companies by buying their stocks and bonds.

Service

Potential customers for a project will be concerned about the quality of service as well as the cost. Moreover, there may be many different aspects of service quality that could be important to certain customers. For example, average trip time, trip time reliability, the probability of excessive delays, comfort, and convenience can all be important for passenger transportation systems. The relative importance of these measures will vary with the type of trip, the type of traveler, and the travel options that are available. For the movement of freight, key measures include the size of the shipment that can be carried and the probability that the shipment will be lost or damaged in addition to trip times and reliability. During the nineteenth century, the introduction of railroads revolutionized inter-city travel as it was so much faster than walking, riding a horse, or going down a canal in a barge. During the 20th century, the introduction of automobiles and air travel drastically reduced the role of railroads and led to massive investments in infrastructure for highways and airports. During the 21st century, high oil prices, congestion, and concerns about emissions may result in another major realignment of transportation systems and investments in transportation infrastructure.

Water resource systems have different measures of service quality. In developed countries, the main concern is likely to be the price, quality and reliability of the water supply. Potable water will be needed for every household, and sufficient quantities of water will need to be provided for all uses and all users at a reasonable price. In many regions of the world, where running water is a luxury, other measures will be useful, such as the percentage of homes with running water, the average distance to the nearest clean water in rural villages, and the probability of disruptions in supply. In some regions, the chemical content of the water is a major concern for public health. For example, arsenic has been found in rural water supplies in Nepal, resulting in serious health problems in many villages.

Historically, failure to keep drinking water separate from waste water and sewage resulted in outbreaks of cholera and other diseases that periodically would kill tens of thousands of people in major cities. Contamination of drinking water remains a critical problem today in situations where earthquakes or tsunamis endanger water supplies and in regions where wars or rumors of war cause thousands of people to move to refugee camps with inadequate infrastructure for either water or sewage.

Navigation, flood control and the use of water power to generate electricity are other matters that relate to the construction of dams and levees. Riverbanks and lakeshores are wonderful locations for parks – but also prime spots for industrial uses. The conflict between the natural and the manmade environment is very apparent as wetlands are filled in, whether to provide additional space for ports, seaside housing, or industrial uses. In many rural areas, vast expanses of wetlands have been drained (via extensive networks of drainage pipes, ditches, and channels) in order to provide fertile lands for agriculture.

Still other types of measures will be needed when considering energy projects. For example, the use of electricity is pervasive throughout the developed world. Normally, customers simply have access to electricity simply by flipping a switch, and their concerns about service relate to the time that the electric company requires to respond to infrequent outages caused by bad weather. Difficulties in finding an electrician to install a new appliance or disputes with the power company about billing are likely to cause more concerns than the availability of power. In the rest of the world, the availability of electricity can be a dominant concern, as many regions have either no access to the power grid or access that is limited by time of day or day of week because of limitations in capacity. Poorly maintained systems may have very frequent outages, and some systems are prone to disruption related to civil unrest.

Access to energy, limitations on the use of energy, likelihood of planned and unplanned disruptions, and the time required to restore service will be aspects of service related to other types of energy used directly by consumers, including gasoline, diesel fuel, natural gas, firewood.

Energy companies have options in terms of the kind of energy resources that they use and the origins of those resources. From their perspective, a critical measure of service quality is the reliability of the supply chain, which is the process by which coal, oil or other energy resource is mined, processed, and transported to their facility. In order to meet their customers' needs, energy companies prefer reliable supply chains for their fuel. If the supply chains are unreliable, companies have several options. They could maintain large inventories of fuel that will enable them to continue operations despite periodic breaks in deliveries, they could have multiple sources of fuel, or they could have redundant networks for providing their services, so that failure in one portion of the system does not shut down the entire system.

Engineering-Based Service Functions

As was the case with cost, engineering-based functions can be developed for each aspect of service. However, service is more complex than cost, because it is not possible to have a single measure (such as "dollars") that is readily accepted by owners and users and everyone else as a valid measure of performance. Thus there is a multi-step process in developing service functions. First, and probably most important, it is important to understand what aspects of service are most important to potential customers. Second, it is necessary to understand how design relates to the quality of service that can be offered.

Capacity

Capacity is another important aspect of the performance of infrastructure-based systems. Measurement of capacity is complicated by a number of factors, not the least of which is the difficulty in defining output. Infrastructure-based systems typically consist of networks of facilities that are managed or used by multiple organizations using labor, equipment, energy, and other resources to provide what may be a rather large array of multi-dimensioned services. To understand capacity, it is best to start with a small piece of the system, where it is possible to enumerate all of the key inputs, where volume is well-defined, and where it is straightforward to estimate what happens to performance as volume increases. As with both cost and service, it is very useful to develop engineering-based functions that relate capacity to the characteristics of the infrastructure and its users.

It is often useful to consider three types of capacity:

- **Maximum capacity** – the maximum flow through the system when everything is operating properly
- **Operating capacity** – the average flow through the system under normal operating conditions
- **Sustainable capacity** – the maximum flow through the system that allows sufficient time for maintenance and recovery from accidents or other incidents

Since demand may vary substantially on a daily, weekly or seasonal basis, it is also important to consider how well the system performs during periods of peak demand. There could well be delays to users or denials of service during peak periods, but the system may be able to recover soon after peak demands subside. Delays during peak periods do not mean that capacity has been reached; it is only when delays become serious impediments to users that capacity has been reached. The solution to peak period problems could be expansions to the infrastructure, changes in operations, changes in pricing, or other attempts to limit peak period demands.

Safety and Security

Safety and security are important aspects of performance for infrastructure systems. Both refer to the likelihood of injuries or fatalities to employees, customers, abutters, and others along with the probability of damages to the infrastructure or to the property of users or abutters. Security generally refers to the measures that might be taken to prevent deliberate attacks on people or property, whether those attacks involve pickpockets, thieves, or terrorists. Safety is generally used in reference to accidents or problems that occur in the course of normal operations of the system, although it may also be used with respect to the possibility of deliberate attacks. Safety records for a transportation company show such things as the number of accidents, the number of fatalities, and the number of fatalities per million miles traveled.

Risk is a broader concept that can be used to describe the potential for future accidents or incidents.[1] A methodology known as "probabilistic risk assessment" can be used to measure risk in a way that is very useful in evaluating system performance. Risk is the product of two factors:

- The probability of an accident or incident
- The expected consequences if an accident or incident occurs

Consequences could include fatalities, injuries, disruption of service, release of toxic chemicals, or inconvenience to people living in or moving through the neighborhood of the accident. A weighting scheme will be needed to compare the severity of the different types of accidents. Strategies for reducing risks could focus either on reducing the probability of accidents or on reducing the expected consequences if there is an accident.

The design of infrastructure systems always involves consideration of risks, including risks to those involved in construction of the facility as well as risks to users. Construction standards and choice of materials will depend in part upon the potential for natural disasters, such as hurricanes or earthquakes, and the need to provide a safe operating environment. It is never possible to eliminate all risks, and judgment will be necessary to determine which risks are worth worrying about.

Some adjustments in predicted risks may be necessary to reflect the perceived importance of certain types of accidents or consequences to various stakeholders. Quantifying perceived risks requires answers to questions such as "Who is at fault?", "Is it a catastrophic accident?", and "Is new technology involved". Public perceptions of risk will be greater for situations where that risk could result in catastrophic accidents with hundreds of fatalities or accidents with dreadful consequences, as could be the case with the release of toxic chemicals or radiation after an accident at a chemical factory or at a nuclear power plant. The public is more concerned with unknown risks that might be associated with new technologies such as drones or driverless cars than with well-known risks. People know that automobile accidents kill tens of thousands of people per year, but they still drive cars. Fears of accidents involving nuclear power plants hampered development of such projects in many countries, despite an excellent safety record.

A general approach to investigating the risks related to infrastructure encompasses the following basic concepts:

- Risk reflects both the probability of an accident and the consequences of the accident.
- Past experience can be a guide for estimating accident probabilities and expected consequences.
- Weights can be devised for comparing different types of consequences.
- Strategies for reducing risks can be evaluated by comparing the costs of those strategies to the expected reduction in risk.

Companies, operating agencies and agencies that create, use or regulate infrastructure are continually seeking cost-effective ways to reduce risks. Since resources are limited and risks can never be entirely eliminated, it is important to allocate those resources effectively. In practice, this means that the expected reduction in the consequences of accidents should be high enough to justify the costs.

Cost Effectiveness

When dealing with costs and finances, it is natural to measure everything in terms of money. When dealing with the other matters, it is more difficult to use money as a performance measure. Thus, if an investment is proposed to improve infrastructure performance, it will usually be difficult if not impossible to determine whether or not the non-

[1] The term "risk" is used in two different ways by safety engineers and financiers. To safety engineers, risk refers to accidents or incidents that result in property damage, injuries or fatalities, and that is the sense in which the term is used in this section. In finance, the term is used to include any of the many factors that may affect the success of a project, including such things as the risk of an economic downturn, the risk of new competition, and the risk of political upheaval as well as the risks related to safety and security. Both meanings are used in this text; which sense is being used should be clear from the context.

monetary benefits justify the financial cost of the investment. Ultimately, some kind of political process will be necessary to weigh all of the costs and benefits to determine whether or not a proposed project is justified.

The concept of **cost-effectiveness** is a valuable technique that can be used in evaluating potential investments aimed at achieving non-monetary benefits. Cost-effectiveness is the ratio of the benefit to the cost of achieving that benefit. Thus, as long as it is possible to measure a benefit, it is possible to calculate cost-effectiveness. If there are multiple ways of achieving the same type of benefit, then the most cost-effective approach is the one that costs the least per unit of improvement. Just because an option is the most cost-effective does not mean that it is worth pursuing – it may be better than any of the others, but it may still be judged to be too costly. Cost-effectiveness is therefore a concept that is most useful in eliminating projects from consideration and in identifying ones that deserve further consideration. A project will not be implemented simply because it is the most cost-effective.

Summary

Transportation, water resource and other infrastructure-based systems serve various needs of society. One aspect of the performance of such systems is therefore how well they actually serve those needs, as measured by the cost to the user and the quality of the service that they receive.

Total, average, marginal, and incremental costs are all important aspects of system performance. The average cost is the total cost divided by the total volume of usage; the units could be the cost/vehicle for a highway or the cost/gallon for a municipal water system. The marginal cost is the cost per additional user, which would be the cost for one more car on a highway or for an additional gallon of water. Pricing and operating decisions are often based upon marginal costs rather than average costs. The incremental cost is the added cost for a larger increment of users, e.g. a 10% increase in vehicles or water usage. Investment and longer-term operating decisions are often based upon consideration of incremental costs: what is the best way to handle the expected increase in usage over the next five years? It is possible to construct engineering-based cost functions that can be used to estimate costs based upon the engineering and operating characteristics of the system.

For many systems, it is useful to consider the distinction between fixed and variable costs. Fixed costs are those associated with making the system available, while variable costs are those that vary with the level of usage. For investments in infrastructure, there are often options that can provide better service or higher capacity, but that require higher fixed costs. A common question is therefore to decide whether or not there will be sufficient demand to justify the alternative with the higher fixed costs.

Since infrastructure lasts a long time, it is important to consider costs over the entire life of the infrastructure. Design costs are likely to be modest compared to the construction costs that follow. Once the project is completed, there will costs of operation that may be borne by the owner, users or abutters. As the facility ages, it may be desirable to expand or to rehabilitate the facilities. Eventually, as a result of declining demand or excessive costs of maintenance, it will be necessary to decommission the facilities and to salvage metal or whatever is left of value. Ideally, the design of infrastructure-based systems will address total lifecycle costs, rather than simply construction costs. Adding room for expansion, allowing more efficient operations, using designs that facilitate maintenance and ensuring safer operations may require additional investment, but result in lower costs over the life of the infrastructure.

The design phase offers the greatest opportunity to affect the life cycle costs of a project. At this point, while all things will not be possible, many things will be. A major design consideration will be the extent to which the owner or developer considers the costs to users and abutters. Small changes in design conceivably produce more efficient operations or limit the negative effects on third parties, but only to the extent to which such costs are even considered in the design. Mistakes regarding the size of the project – too big, too small, too soon, too little flexibility, too difficult to rehabilitate, too much impact on neighbors – conceivably can be rectified with little or no additional time or expense related to construction.

‘i.e. the profitability and the return on investment for the system. The return on investment is the annual profit divided by the total amount of the investment. While financial performance is not the only measure, or even the most important measure, if financial performance is deemed to be unacceptable then it will be difficult or impossible to find investors (or taxpayers) willing to invest more money in improving or expanding the system. The interaction between supply and demand will be an important factor determining the prices that can be charged for a service.

A third aspect of performance will be the ability of the system to handle growth in demand. The maximum capacity of a system is limited by the initial design, engineering factors, and operating constraints. The maximum capacity may be useful in design, but cannot be achieved except for brief periods. Operating capacity recognizes the importance of considering such things as the normal variations in volume, weather, and the need of routine maintenance. The operating capacity is what is achievable on most days when the system operates pretty much as planned. The limit to operating capacity is likely to be what users view as acceptable delays or restrictions on use during peak periods. The sustainable capacity is somewhat lower, because time must eventually be allowed for more maintenance and there will be periods when severe weather or accidents disrupt operations for a period of hours, days, or weeks. Sustainable capacity for most systems will be on the order of 70% of maximum capacity.

A fourth aspect of performance will be the safety and security of the system, i.e. the likelihood of accidents or disruptions to the system based upon system problems or attacks upon the system. Risk is a useful concept for considering safety and security issues, as risk is the product of two key factors: the likelihood of an accident or incident and the expected fatalities, injuries and other consequences if there is an accident or incident. Probabilistic risk assessment is a methodology that can be used to determine the cost-effectiveness of various options for reducing risk.

Figure 13 Union Pacific’s Rail Freight Terminal in North Platte, Nebraska
During the late 20th century, the rail industry struggled to increase capacity to keep pace with the demand for coal and the rapid growth in the shipment of containers by rail.