Chapter 5

Transportation Economics
5.1 HISTORY, DEFINITION AND SCOPE OF TRANSPORTATION ECONOMICS

5.1.1 Historical Context

Up to the 18th century, the most important commercial cities in the world were maritime cities due to the relatively low costs of water transportation. However, the invention of the steam engine in the eighteenth century marked a watershed in the history of transportation by allowing for greater economy in transportation of goods and passengers, and therefore shifted the balance in favor of land transportation. After World War II, advancements in road construction technology and mass production of the automobile led to increasing use of highways for land transportation. Currently, highway transportation accounts for approximately 30% of all ton-miles (representing 90% of all overall value) of freight transportation, and over 95% of passenger transportation in terms of person-miles (BTS, 2001). Such modal shifts that have been observed over the years arose from improvements in transportation technology, and resulted in the reductions in transportation costs and time. Reductions in transportation cost and time have in turn led to increased availability of goods, lower prices of goods and services, price stabilization and equalization, changes in land values, urbanization, and equity (Locklin, D.P., 1960).

In recent years, certain developments have greatly influenced the economics of the various modes of transportation. These include the deregulation of the transportation industry (1977-1980), which enabled shippers and carriers to negotiate the best mutually beneficial rates and service packages, and Just-in-Time logistics systems, which reduced the need for inventory and therefore lowered the holding costs of goods. The other developments are increasing demands of customers for improved quality of service (which includes ensuring that a product is transported to a destination when it is needed, in the right quantities and in undamaged condition), and globalization of business, as companies are increasingly seeking to purchase their production inputs or market their products regardless of global location (Wood and Johnson, 1996).

5.1.2 Definition

Economics has been defined as “the study of how people and society end up choosing, with or without the use of money, to employ scarce productive resources that could have alternative uses to produce various commodities and distribute them for consumption, now and in the future, among various persons and groups in society. It analyzes the costs and benefits of improving patterns of resource allocation” (Samuelson, 1976). Transportation refers to the movement of persons, goods and services from one point to another, with the aid of fixed facilities and/or vehicles such as bridges, highway pavements, pipelines, aircraft, etc. Transportation, from
the economist’s viewpoint, may therefore be considered in terms of either supply (the available quantity and quality of the fixed facilities and vehicles) or demand (the “desire” of persons or goods to be transported and ability to pay for it). Transportation economics can be described as the study of how scarce productive resources are used to produce and distribute various transportation services for consumption by the society.

5.1.3 Scope of Transportation Economics

The study of economics is divided into macro-economics and micro-economics. Micro-economics is associated with the wealth of society on a regional scale, and deals with the behavior of aggregate concepts. On the other hand, micro-economics involves the behavior of relatively smaller entities such as firms and individuals. Transportation economics, while considered a branch of applied micro-economics, is associated with certain unique issues (Khisty and Lall, 2002) such as:

- the demand for transportation is not direct, but is derived
- the consumption of each transportation facility (i.e., each trip) is unique in time and space
- technological differences among different modes and economies of scale
- governmental interventionist policies and regulations in transportation

Transportation economics specifically addresses demand of transportation services, supply of transportation facilities, elasticities of demand and supply, price mechanisms, and transportation cost analysis.

5.2 Importance of Transportation Economics

Constituting the largest government-owned asset in the United States, transportation facilities such as highways and bridges are associated with annual investment levels exceeding 1 trillion dollars (FHWA, 1996; AASHTO, 1996). Such investments are in the form of new construction, rehabilitation and maintenance, and operations. Transportation agencies at all levels of government have the responsibility of effectively managing the performance and usage of their physical assets so that such assets can be kept in acceptable condition to provide desirable levels of service with available resources. Given the ever increasing commercial and personal travel demands vis-à-vis limited resources, this task is more critical than ever before. Managers of transportation facilities are now being perceived as stewards of a vast public asset, and are expected to provide operational and financial accountability of any investment decision. The management of transportation assets, defined as a systematic process of maintaining, upgrading, and operating physical assets cost-effectively (FHWA, 1999) that combines engineering principles with sound business practices and economic theory, has been touted as a means of achieving more organized, logical and integrated approaches to decision making involving transportation systems. The recent issuance of Governmental Accounting Standards Board Statement 34 (GASB34) that established new financial reporting requirements for state and local governments to ensure safekeeping and appropriate use of public resources and operational accountability (GASB, 1999), brought a new dimension to the importance of economics in transportation. Such trends, coupled with increasing public expectation and extraordinary advances in technology, have ushered in a new era of the economics of transportation systems.
Furthermore, such new perspectives in the transportation environment underscore the need for transportation policy makers, engineers, managers and administrators to be well trained in formal economics and finance.

5.3 TRANSPORTATION DEMAND AND SUPPLY

5.3.1 Analysis of Transportation Demand

Like all other goods and services, the demand for any specific transportation facility demands on factors pertaining to the consumer such as income, and characteristics of the facility such as the cost associated with its use (in terms of time and price) relative to rival facilities. A typical example of such demand is that for auto travel: lower incomes of consumers, coupled with lower costs and travel times associated with transit are expected to lead to reduced demand for auto travel. Transportation demand analysis involves demand functions (which represents the willingness of consumers to purchase the transportation “product” at various alternative prices, i.e., the demand-price curve, and demand models (which estimate the probability that an individual (or fraction of a set of individuals) will choose a particular product over the other. This section focuses on demand functions, while the concept of demand modeling is discussed briefly in a later section of this chapter.

A hypothetical example of an aggregate transportation demand function is provided as Figure 5-1. This represents the amount of travel people are willing to make by transit at various transit fare (price) levels. Transportation demand functions, either in the form of a graph or an equation, are useful in transportation planning because they enable the determination of expected demand at any price. A specific demand curve represents the demand-price relationship given a set of conditions specific to the transportation product in question (referred to as alternative-specific attributes, such as travel time, comfort, convenience), and also specific to the users (income levels and other socio-economic characteristics). Changes in such conditions often result in changes in the levels of transportation demand, even at fixed price of that product. For example, increased unemployment would likely lead to reduced demand for travel. Also, an increase in costs associated with auto use is likely to result in increased transit demand, even if transit fares remain the same. When such changes in conditions (other than price) occur, they are represented as a shift in the demand curve shown as Figure 5-2 (upward shift for increased demand, $D_1 \rightarrow D_2$; and downward shift for decreased demand, $D_1 \rightarrow D_3$).
Transportation Demand Functions

A basic feature of transportation systems analysis is the prediction of transportation demand or changes thereof. The level of transportation demand \( D \) may be expressed as a function of activity system attributes \( A \), and the prevailing level of service \( S \), as shown as Equation 5.1.

\[
D = f(A, S) \tag{5.1}
\]

Transportation demand models, such as that shown as Equation 5.1, are used to determine the volume of travel demanded, at various levels of service and have been described as “a representation of human behavior which can be used to predict how individuals or firms [or groups thereof] will change transportation choices in response to changes in future conditions” (Manheim, 1979). Within the context of transportation economic, a trip maker is defined as a consumer, in the economics meaning of the word, as the trip maker, by planning a trip, seeks to consume the service offered by transportation facilities. There are two types of demand functions:

1. **Disaggregate demand functions**: these predict the behavior of a single consumer in response to changes in future conditions,

2. **Aggregate demand functions**: these predict the behavior of a group of consumers such as a household, in response to changes in future conditions.

Compared to the latter, disaggregate demand functions are relatively young, and recent work in this area include those by McFadden, 1985, and Ben-Akiva, 1988. For aggregate demand functions, aggregation transportation demand into market segments is typically done on the basis of geographic region called traffic analysis zones (TAZs). Manheim (1979) provides examples of transportation demand models as shown below:
The Gravity Model: This is described as a classic transportation demand model which is analogous to Newton’s law of universal gravitation. The general equation of a gravity model is as follows (Manheim, 1979):

\[ V_{12} = Y_1 \times Z_2 \times L_{12} \]  

(5.2)

Where

- \( Y_1 \) = some measure of intensity of activity at zone 1, such as population or employment.
- \( Z_2 \) = some measure of intensity of activity at zone 2, such as population or employment.
- \( L_{12} \) = effect of transportation service attributes on demand for travel between zones 1 and 2.

It has commonly been assumed that \( L_{12} \) is directly related to the travel time between zones 1 and 2.

The Kraft-SARC Model

\[ V_{kdm} = \phi_0 (P_k P_d)^{\phi_{1}} (I_k I_d)^{\phi_{2}} (t_{kdm} t_{kdm})^{\phi_{3}} (t_{kdm} t_{kdm})^{\phi_{4}} \]  

(5.3)

c) The McLynn Model

\[ V_{kdm} = \phi_0 (P_k P_d)^{\phi_{1}} (I_k I_d)^{\phi_{2}} \left( \sum_q c_{kdm}^{\phi_{3}} (t_{kdm} t_{kdm})^{\phi_{4}} + \sum_q c_{kdm}^{\phi_{5}} (t_{kdm} t_{kdm})^{\phi_{6}} \right) \]  

(5.4)

c) Baumol-Quandt Model

\[ V_{kdm} = \phi_0 (P_k P_d)^{\phi_{1}} (I_k I_d)^{\phi_{2}} \left( c_{kdm}^{\phi_{3}} (t_{kdm} t_{kdm})^{\phi_{4}} + c_{kdm}^{\phi_{5}} (t_{kdm} t_{kdm})^{\phi_{6}} \right) \]  

(5.5)

where

- \( V_{kdm} \) = volume of travel between zones \( k \) and \( d \) by mode \( m \)
- \( P_k \) = population in zone \( k \)
- \( I_k \) = median income in zone \( k \)
- \( t_{kdm} \) = travel time between zones \( k \) and \( d \)
- \( c_{kdm} \) = travel fare between zone
- \( t_{kdm} \) = travel time by fastest mode
- \( c_{kdm} \) = fare by cheapest mode
- \( \phi, \theta, \delta \) = parameters of the model

Note: the subscripts indicate whether the parameters are mode-dependent (\( \theta_{mj} \)) or mode-independent (\( \theta_1 \)).
Table 5-1: Classification of Demand Models

<table>
<thead>
<tr>
<th></th>
<th>Stochastic</th>
<th>Deterministic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>Primarily theoretical</td>
<td>This has been the main area of development, historically for travel demand analysis</td>
</tr>
<tr>
<td></td>
<td>Used for expository purposes</td>
<td></td>
</tr>
<tr>
<td>Disaggregate</td>
<td>There has been extensive development in these models (for both practical and theoretical applications) in recent years</td>
<td>No theoretical or practical developments</td>
</tr>
</tbody>
</table>

Source: Manheim (1979)

5.3.1 Analysis of Transportation Supply

The supply of a transportation product represents the quantity of that product a producer is willing to offer at a given price. However, transportation supply may also be associated with the quality of the product. At a given time, transportation facility supply depends on price: higher prices are an incentive for producers to make more profits who therefore increase supply levels. A hypothetical example of a transportation supply function is provided as Figure 5-3. This represents the amount of transportation products that suppliers are willing to make available at various prices. Transportation supply functions are useful in transportation planning because they enable the determination of expected supply at any future price. A specific supply curve represents the supply-price relationship given a set of conditions specific to the transportation product in question (referred to as alternative-specific attributes, such as travel time, comfort, convenience), and also specific to the producers (such as technology, policy and governmental intervention through policies and regulation. Changes in such conditions often result in changes in the levels of transportation supply, even at fixed price of that product. For example, improved increased unemployment would likely lead to reduced demand for travel. Also, an increase in costs associated with auto use is likely to result in increased transit demand, even if transit fares remain the same. When such changes in conditions (other than price) occur, they are represented as a shift in the supply curve shown as Figure 5-4 (upward shift for increased demand, $S_1 \rightarrow S_2$; and downward shift for decreased demand, $S_1 \rightarrow S_3$) (Manheim, 1979).

Increases in transportation supply may be traditionally thought of in terms of increasing the fleet size of a transit company or building new roads or increasing the number of lanes for existing roads. However, it is possible to increase supply without such physical capital-intensive investments (Manheim, 1979). For instance, the use of intelligent transportation systems could lead to increased supply without any physical enlargements of the road network.
Equilibrium of Transportation Demand and Supply

Like the supply function, the transportation demand function is a relation between quantities of goods and price. The general forms of the demand and supply functions (Figures 5-3 and 5-4) suggest that there could exist a point at which the demand of a transportation commodity is equal to its supply. Such a state is indicative of equilibrium.

**Example 1** (Khisty and Lall, 2002)

A company determines the price of a seat on a particular rote to be \( p = 200 + 0.02n \). The demand for this route by air has been found to be \( n = 5000 - 20p \), where \( p \) is the price in dollars, and \( n \) is the number of seats sold per day. Determine the equilibrium price charged and the number of seats sold per day.

**Solution:**

\[
\begin{align*}
p &= 200 + 0.02n \\
n &= 5000 - 20p
\end{align*}
\]

Solving these two equations simultaneously, we obtain \( p = 214.28 \) and \( n = 714 \) seats.
Example 2 (Khisty and Lall, 2002)

The travel time on a stretch of highway lane connecting two activity centers has been observed to follow the equation representing the service function:

\[ t = 15 + 0.02v \]

where \( t \) and \( v \) are measured in minutes and vehicle per hour respectively. The demand function for travel connecting the two activity centers is \( v = 4000 - 120t \).

a) Sketch these two equations and determine the equilibrium time and speed of travel.

b) If the length of the highway lane is 20 miles, what is the average speed of vehicles traversing this length?

Solution:

\[ t = 15 + 0.02v \]
\[ v = 4000 - 120t \]

Solving these two equations simultaneously yields:

\( v = 647 \) vehicle/hr
\( t = 27.94 \) minutes

Therefore speed = \((20 \times 60)/27.94 = 42.95 \) mph

Figure 5-5: Static Equilibrium of Demand and Supply

Example 3 (Salvatore, 1982)

The following structural equations represent a simple demand-supply model:

\[
\begin{align*}
\text{Demand: } Q_t &= a_0 + a_1P_t + a_2Y_t + u_{1t} \\
\text{Supply: } Q_t &= b_0 + b_1P_t + u_{2t}
\end{align*}
\]

\( a_1 < 0 \) and \( a_2 > 0 \)

\( b_1 > 0 \)
Where $Q_t$ is the quantity demand, $P$ is the price, and $Y$ is consumer’s income. It is assumed that the market is cleared in every year so that $Q_t$ represents both quantity bought and sold in year $t$.

(a) Why is this a simultaneous-equations model?

(b) Why would the estimation of demand and supply function by OLS give biased and inconsistent parameter estimates?

Solution:

(a) The given demand-supply model represents a simple simultaneous-equations market system, as $Q$ and $P$ are mutually or jointly determined. If the prevailing price is less than equilibrium price, the quantity demanded would exceed the quantity supplied, and vice versa. At equilibrium, the demand curve meets the supply curve, thereby jointly (or simultaneously) determining the equilibrium conditions $Q$ and $P$.

(b) The endogenous variables of the model are $Q$ and $P$. These are the variables determined within the model. $Y$ is the only exogenous variable of the model because it is the only variable determined outside the model.

(c) The endogenous variable $P$ is also an explanatory variable in both the demand and supply equations. Therefore $P$ is correlated with $u_{1t}$ in the demand equation and also with $u_{2t}$ in the supply equation. This violates a key OLS assumption that the explanatory variable should not be correlated with the error term. As a result, estimating the demand and supply functions by OLS would yield estimates that are not only biased but also inconsistent.

Sensitivity of Travel Demand

In the planning and evaluation of transportation systems and associated investments, it is often useful to have knowledge of the changes in transportation demand caused by changes in attributes of the transportation system or its environment. A particular instance is the change in demand for a given mode in response to changes in price of that mode. Given the functional form of the travel demand function, it is possible to derive a marginal effects model that estimates any one of the following:

- Change in demand in response to unit change in attribute
- Change in demand in response to unit percent change in attribute
- Percent change in demand in response to unit percent change in attribute

The following table presents some standard functional form of demand functions, as well as their marginal effects models (specifically, their point elasticities).

Transportation demand elasticity may be defined as the degree of responsiveness of transportation demand in response to a unit change in demand-related or attributes such as price or income. This is typically expressed as follows:

$$ e_p = \frac{\partial q}{\partial p} \frac{p}{q} = \frac{\partial q}{\partial p} \times \frac{p}{q} $$  \hspace{1cm} (5.6)
Arc elasticity can be calculated as follows:

\[
\text{Arc Elasticity} = \frac{\partial q / q}{\partial p / P} = \frac{\partial q}{\partial p} \times \frac{p}{q} = \frac{Q_1 - Q_0 (P_1 + P_0)}{P_1 - P_0 (Q_1 + Q_0)}/2
\]  

(5.7)

**Interpretation of Price Elasticities**

Values of price elasticities with respect to demand make it possible for transportation operators to predict the impact of changing transportation prices on total revenue.

<table>
<thead>
<tr>
<th>Nature of Demand</th>
<th>Nature of Relation between Price and Revenue</th>
<th>Impact of Price Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>e &gt; 1 elastic</td>
<td>negatively related</td>
<td>reduce revenue</td>
</tr>
<tr>
<td>e &lt; 1 inelastic</td>
<td>positively related</td>
<td>increase revenue</td>
</tr>
<tr>
<td>e = 1 Unit elasticity</td>
<td>none</td>
<td>remain the same</td>
</tr>
</tbody>
</table>

**Example 4: Point Elasticity** (Khisty and Lall, 2002)

An aggregate demand function is represented by the equation

\[ q = 200 - 10p \]

where \( q \) is the quantity of a good, and \( p \) is the price per unit. Find the price elasticity of demand when

\( q = 0, q = 50, q = 100, q = 150, q = 200 \) units

corresponding to

\( p = 20, p = 15, p = 10, p = 5, p = 0 \) cents

**Solution:**

\[ e_0 = 1 - \frac{200}{200} = 0 \]

![Demand function showing elasticities at various quantities.](image-url)
When the elasticity is less than –1 (i.e., more negative than –1), the demand is described as being elastic, meaning that the resulting percentage change in quantity will be larger than the percentage change in price. In this case, demand is relatively sensitive to price change. However, when elasticity is between 0 and –1, the demand is described as being inelastic or relatively insensitive. These ranges are shown in Figure 5-6(b).

\[ e_p = 1 - \frac{\alpha}{q} \text{ where } \alpha = 200 \]

\[ e_5 = 1 - \frac{200}{150} = -0.333 \]

\[ e_{10} = 1 - \frac{200}{100} = -1 \]

\[ e_{15} = 1 - \frac{200}{50} = -3 \]

\[ e_{20} = 1 - \frac{200}{0} = -\infty \]

Discussion:

From Figure 5-6(a), it is obvious that when the price per unit is 20 cents, no units are bought. Also, when nothing is charged per unit, 200 units are bought. Notice that the price elasticity for this system varies from 0 to -\infty, with unit elasticity when \( p = 10 \).
Example 5: Arc Elasticity (Khisty and Lall, 2002)

When admission rate to an amusement park was $5 per visit, the average number of visits per person was 20 per year. Since the rate has risen to $6, the demand has fallen to 16 per year. What is the elasticity of demand over this range of prices?

Solution:

Arc price elasticity, \( e_p = \frac{\Delta Q (P_1 + P_2)/2}{\Delta P (Q_1 + Q_2)/2} = \frac{4 \cdot (5 - 5)}{(1)(18)} = -1.22 \)

Therefore, elastic.

Discussion:

Note that there are problems connected with arc price elasticity because it will differ from point elasticity, the difference increasing as \( \Delta P \) or \( \Delta Q \) increase. Also note that elasticity is a unit-free measure of the percent change in quantity demanded (or supplied) for a 1 percent change in price.

5.4 FACTORS AFFECTING ELASTICITY

As stated earlier, elasticity is the change in demand in response to a unit change in levels of attributes of the transportation system or its environment. Such attributes include characteristics of the transportation system such as price and level of service associated with a given mode, price and level of service of competing modes, and characteristics of the socio-economic system such as income, level of employment, household size, car ownership, etc. Of these factors, of particular interest are price and income. The elasticities of demand in response to price and income are known as price elasticity and income elasticity, respectively.

5.4.1 Income Elasticities

The elasticity of demand with respect to income, or income elasticity, is the change in demand for a good in response to a unit change in income of the consumer of that good. Income elasticities have a special significance in travel demand modeling. Often, the transportation planner seeks to evaluate the impact of changing socio-economic trends on the demand for or share of various modes of transportation. A major indicator of economic trends is income. In disaggregate demand modeling, it is sought to determine the sensitivity of changing income on the demand for a particular mode. Income elasticity is generally defined as the change in demand in response to a unit change in income. In transportation economics, a good service is considered normal if there is a direct relationship between the demand for that commodity and the income of the consumer. Besides increased demand, if the share of the demand for that good (in the consumer’s total income) also increases, then the good is described as a superior good. If the demand of a good decreases with increasing income then the good is described as inferior. In developed countries, automobile travel is considered superior while mass transit is considered an inferior good.
5.4.2 Price Elasticities

The elasticity of demand with respect to price, or price elasticity, is the change in demand for a good in response to a unit change in the price of the good. A study of price elasticities is important because it is often sought to assess the impacts of changing prices of a good or rival goods (due to past supply and demand conditions) on the demand of that good. The level of price elasticity depends on factors such as price of rival goods, income-share of the good, the scope of definition of the good, and whether the good is considered a luxury or a necessity.

Price of rival goods: A consumer who spends a substantial percentage of income on a particular good is more likely to seek a substitute good when the price of the good increases.

Scope of definition: Goods that have narrow definitions are more likely to have more substitutes, and are therefore expected to have a more elastic demand.

Price an availability of rival goods: The lower the price and greater availability of substitutes, the greater the elasticity of demand of the good with respect to price.

Luxury vs. substitute goods: Goods that are considered necessities typically have price inelasticities, while luxury goods are relatively elastic.

Example 6 (Khisty and Lall, 2002)

A bus company’s linear demand curve is $P = 10 - 0.05Q$, where $P$ is the price of a one-way ticket, and $Q$ is the number of tickets sold per hour. Determine the total revenue along the curve.

$$R = 10Q - 0.05Q^2$$

Figure 5-7: Total Revenue Curve.
Solution:

\[ P = 10 - 0.05Q \]
\[ R = Q(10 - 0.05Q) \text{ where } R = \text{total revenue} \]
\[ R = 10Q - 0.05Q^2 \]
\[ \frac{dR}{dQ} = 10 - (0.05 \times 2)Q \]

and this is equal to zero when \( R \) is maximum

Therefore, \( Q = 100 \) when \( R = 500 \) (maximum).

Discussion:
Starting from a price of $10 at near zero tickets sold and decreasing the price eventually to half ($5), the revenue steadily increases to a maximum of 4500/hour (over the elastic portion). After that, the revenue decreases as the price further decreases and finally approaches near zero, when the demand approaches 200 (over the inelastic portion).

Direct and Cross Elasticities

Direct elasticity is the effect of a change in the price of a good on the demand for the same good, while cross elasticity refers to the degree of responsiveness of the demand for a good in response to a unit change in price of another good.

Substitute goods: When consumers buy more of good A when good B’s price increases, good A is typically described as a good substitute of good A. An example of such goods is auto use and transit: When the price of auto-use goes up, the demand for transit increases. In this case, cross elasticity is positive.

Complementary goods: When an increased price of B results in increased demand for B, goods A and B are typically described as complementary. In this case, cross elasticity is negative.

Example 7 (Khisty and Lall, 2002)
A 15% increase in gasoline costs has resulted in a 7% increase in bus patronage and a 9% decrease in gasoline consumption in a mid-sized city. Calculate the implied direct and cross elasticities of demand.

Solution:

Let

\[ P_0 = \text{initial price of gas} \]
\[ P_1 = \text{price of gas after the hike} \]
\[ Q_0 = \text{quantity of gas consumed before} \]
\[ Q_1 = \text{quantity of gas consumed after} \]
For direct elasticity:
\[ Q_0(\text{gas}) \times 0.91 = Q_1(\text{gas}) \]
\[ P_0(\text{gas}) \times 1.15 = P_1(\text{gas}) \]

From the elasticity formula,
\[ e = \frac{\Delta Q / Q}{\Delta P / P} = \frac{-0.09 / (1 + 0.91)}{0.15 / (1 + 1.15)} = -0.675 \]

And let \( B_0 = \) initial bus patronage
\( B_1 = \) bus patronage after the hike

For cross elasticity:
\[ B_0(\text{bus}) \times 1.17 = B_1(\text{gas}) \]
\[ B_0(\text{bus}) \times 1.15 = P_1(\text{gas}) \]

From the elasticity formula,
\[ e = \frac{\Delta B / B}{\Delta P / P} = \frac{0.07 / (1 + 1.07)}{0.15 / (1 + 1.15)} = +0.48 \]

Discussion:
In the case of direct elasticity we are calculating the percent change in gasoline consumption due to a 1 percent change in the price of gasoline, while in the case of cross elasticity we are calculating the percent change in bus patronage due to a 1 percent change in gasoline price.

5.5 APPLICATIONS OF PRICE ELASTICITIES: CONSUMER SURPLUS AND LATENT DEMAND

Consumer surplus is a measure of the monetary value made available to consumers by the existence of a facility. It is defined as the difference between what the consumers might be willing to pay for a service and what they actually pay. A patron of a bus service pays a fare of, say, 50 cents per trip but would be willing to pay up to as much as 75 cents per trip. In this case, her consumer surplus is 25 cents.

Consumer surplus may be determined from a survey of customers to determine how much they are willing to pay for a good or service. Recent studies have evaluated the willingness of road users to pay for improvements that would enhance safety on their highways (Islam and Sinha, 2002). Such information can be combined with data on actual payments to determine the amount of consumer surplus per person or for a group of persons. Alternatively, consumer surplus can also be determined from the demand-supply curve, as the
demand curve can be considered as an indicator of the utility of the good or service in terms of price, as illustrated in Figure 5-8. The area ABC represents the total consumer surplus. Maximization of consumer surplus is the maximization of the economic utility of the consumer (Khisty and Lall, 2002). The use of the consumer surplus concept is common in the area of the evaluation of transit systems.

![Figure 5-8: The Concept of Consumer Surplus.](image1)

![Figure 5-9: Change in Consumer Surplus.](image2)

Generally, the economic impacts of improvements in a transportation system can be evaluated in terms of consumer surplus, and can be represented as the area under a demand curve and a shift in the supply curve. Figure 5-9 indicates the case of an urban arterial street with a traffic supply $S_1$, intersecting a demand curve at $E_1$. An additional lane is added, thus shifting the supply curve to $S_2$ and therefore intersecting the demand curve at $E_2$. The change in consumer surplus can be quantified as the trapezoidal area $P_1, P_2, E_1, E_2$. In other words,

$$E = (P_1 - P_2)(Q_1 + Q_2)/2.$$  

The consumer surplus can be measured as the difference between the maximum amount which consumers are willing to pay for a specified quantity of a good or service rather than going without it. In general, the area $AQB$ in Figure 5-9 represents the total community benefit, $BCOQ$ is equal to the market value, and $ACB$ is equal to the consumer surplus or net community benefit. For a constant supply curve, a higher elasticity is associated with a smaller consumer surplus, and vice versa.

Figure 5-8 also illustrates an additional concept that is useful to transportation engineers: latent demand. It can be observed that travelers between $Q$ and $D$ do not make trips at the current time, but would do so if the price per trip were lower than the equilibrium price. The number of such potential travelers is referred to as latent demand (Khisty and Lall, 2002). An application of this concept is the investigation of latent demand associated with introduction of incentives for non-peak travel, such as reducing transit fares during non-peak hours. From Figure 5-8, it can be seen that if transit travel were free (zero trip price), the travel demand (quantity of trips demanded) would be $Q_L - Q$. 

---

**Figure 5-8:** The Concept of Consumer Surplus.

**Figure 5-9:** Change in Consumer Surplus.
Example 8 (Khisty and Lall, 2002)

A bus company with an existing fleet of one hundred 40-seater buses increases its fleet size by 20% and reduces its $1.00 fare to 90 cents per ride. Determine the impact of the price change on the company’s profitability, change in consumer surplus, and the price elasticity of demand. It is assumed that the current load factor of 90% would increase to 95% after the price change. Assume that all buses in the fleet are being used during the peak hours.

Solution:

Existing situation:

\[
\text{Demand} = 100 \text{ buses} \times 40 \text{ seats} \times 0.90 \text{ (load factor)} = 3600 \text{ persons per hour}
\]

\[
\text{Revenue} = \text{Demand} \times \text{Price} = 3600 \times 1.00 = $3600 \text{ per hour}
\]

After the price drop and supply increase:

\[
\text{Demand} = 120 \text{ buses} \times 40 \text{ seats} \times 0.95 = 4560 \text{ persons/hour}
\]

\[
\text{Revenue} = 4560 \times 0.90 = $4104 \text{ per hour}
\]

After the price change, the company gains $4104 - $3600 = $504 per hour.

Change in consumer surplus = \((1.00 - 0.90) \times (3600 + 4560)/2 = $408/hr\)

Price elasticity of demand

\[
\text{Price elasticity of demand} = \frac{(Q_i - Q_0) \cdot (P_i + P_0)/2}{(P_i - P_0)/2 \cdot (Q_i + Q_0)/2}
\]

\[
= \frac{-960}{0.10 \times 4080} = -2.235
\]

It is interesting to observe that even if the supply were not increased, the decrease in price alone would have increased overall revenue by $200. It is therefore possible to evaluate the economic impacts of a change in price, a change in supply, or both.

5.6 COST ANALYSIS IN THE EVALUATION OF TRANSPORTATION SYSTEMS

A complete and balanced evaluation of alternative investment options is possible only by giving due consideration to both benefits and costs associated with each alternative. For this reason, it is essential to have knowledge of the costs of each aspect of the provision of a transportation good or service. This way, the future costs of such aspects can be determined using average cost values or better still, cost models that estimate cost as a function of investment and facility attributes.

In economic theory, three types of costs are encountered: fixed costs (which are independent of the volume of goods produced), variable costs (which depend on the volume of good produced), and total costs (the sum of fixed and variable costs). For each of these costs, it is possible to find the average cost (dividing the total
production cost by the number of goods produced) and marginal cost (the incremental cost of producing an additional unit). It is therefore possible to determine average fixed costs, marginal fixed costs, average total costs, etc.

Economic Laws Related to Costs

Two cost-related concepts are encountered in costs analysis of transportation systems evaluation:

Law of Diminishing Returns: States that an increase in input of one unit of a factor of production generally causes an increase in output, but only up to a point, after which increasing inputs of that factor will result in progressively less increase in output.

Law of Increasing Returns to Scale: States that in practice, the production of units is often likely to increase at a faster rate than the increase in the factors of production. This may be due to technological features, specialization.

Average Cost

The total cost of a product can be expressed mathematically as follows:

\[ C = c q = \alpha + \beta(q) \]  

(5.8)

Where

- \( C \) = total cost of a product
- \( c \) = unit cost
- \( q \) = magnitude of the output
- \( \alpha \) = fixed cost of production
- \( \beta \) = variable cost of production

The average cost of each item produced is equal to:

\[ \bar{c} = \frac{C}{q} = \frac{cq}{q} = \frac{\alpha + \beta(q)}{q} = \frac{\alpha}{q} + \frac{\beta(q)}{q} \]  

(5.9)

The relationships of the total and average cost functions are shown in Figure 5-10. It can be seen that as output \( q \) increases, the average cost of production decreases and then increases at higher levels of production. When the production level reaches \( q' \), the average cost is a minimum (\( \bar{c} \)). The decrease in average cost with increasing output is referred to as economies of scale. In the figure, there is obviously there is economy of scale.
for production levels between 0 and $q'$. However, there is no economy of scale beyond $q'$ because the average cost increases. This concept is useful to engineers in deciding whether additional capacity or growth would yield higher profits, and is important in the economic evaluation of transportation system improvements.

![Figure 5-10: Total and Average Costs.](image)

**Marginal Costs**

The marginal costs of a transportation good or service is the additional cost associated with the production of an additional unit of output. The following example (Khisty and Lall, 2002) illustrates the concepts of average and marginal costs.

Table 5-3 presents the cost of running a train system with variable number of wagons. For each system size, the fixed and variable costs are provided in the first three columns. The total, average and marginal costs are then computed and presented in the next three columns.
The various costs are computed as follows:

\[ \text{Total Cost} \quad = \quad TC(x) = FC = VC(x) \]  \hspace{1cm} (5.10)

\[ \text{Average Total Cost} \quad = \quad AC(x) = \frac{TC(x)}{x} = \frac{FC}{x} + \frac{VC(x)}{x} \]  \hspace{1cm} (5.11)

\[ \text{Marginal Total Cost} \quad = \quad MC(x) = TC(x) - TC(x-1) \]  \hspace{1cm} (5.12)

Table 5-3: Costs associated with Train System Operation

<table>
<thead>
<tr>
<th>Number of wagon/train</th>
<th>Fixed cost, FC</th>
<th>Variable cost, VC</th>
<th>Total cost, TC</th>
<th>Average cost, AC</th>
<th>Marginal cost/unit, MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55</td>
<td>30</td>
<td>85</td>
<td>85.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>55</td>
<td>110</td>
<td>55.0</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>75</td>
<td>130</td>
<td>43.3</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>105</td>
<td>160</td>
<td>40.0</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>155</td>
<td>210</td>
<td>42.0</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>225</td>
<td>280</td>
<td>46.7</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td>55</td>
<td>315</td>
<td>370</td>
<td>52.9</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>425</td>
<td>480</td>
<td>60.0</td>
<td>110</td>
</tr>
<tr>
<td>9</td>
<td>55</td>
<td>555</td>
<td>610</td>
<td>67.8</td>
<td>130</td>
</tr>
<tr>
<td>10</td>
<td>55</td>
<td>705</td>
<td>760</td>
<td>76.0</td>
<td>150</td>
</tr>
</tbody>
</table>

Figure 5-11b presents the curves corresponding to average cost, marginal cost and total costs for the train system. It can be seen that the point of minimum cost ($40) occurs at the intersection of the average cost (AC) and marginal cost (MC) curves. It is also observed that the projection of this point to Figure 5-11a corresponds to the point where the gradient of the tangent drawn from the origin has the minimum slope.
When the output is a continuous function, the differential form of the marginal cost is used, in which the marginal cost is the rate of change of total cost with respect to a change in output. In this form the equation is:

\[ MC(x) = \frac{dTC(x)}{dx} = \frac{dVC(x)}{dx} \]  

(5.13)

From the geometry of the AC and MC curves, it is also seen that the average costs is proportional to the slope of a line connecting the origin of the total cost curve with a point on that curve corresponding to the total output. In our example, the slope of such a line begins at infinity at zero output and then decreases to its lowest point, when \( x = 4 \). Beyond this point, the slope increases again. On the other hand, the marginal cost curve is the slope of the tangent drawn at any point on the total cost curve.

**Costs and Production**

In general, a private company will continue to produce and market a good or service as long as it is returning a profit. The net profit (\( P \)) is equal to the total revenue (\( R \)) minus the total costs (\( C \)).

\[ P = R - C = pq - cq \]

Where \( p \) is the selling price of one unit of product \( q \), and \( c \) is the production cost of one unit. The necessary condition for profit maximization is:

\[ \frac{dP}{dq} = \frac{dR}{dq} - \frac{dC}{dq} = 0 \]
or

\[
\frac{dP}{dq} = \frac{d(pq)}{dq} - \frac{d(cq)}{dq} = 0
\]

Thus,

\[
\frac{d(pq)}{dq} = \frac{d(cq)}{dq}
\]

Let

\( MR = \text{marginal revenue} = \frac{dR}{dq} = \frac{d(pq)}{dq} \)

\( MC = \text{marginal cost} = \frac{dC}{dq} = \frac{d(cq)}{dq} \)

Therefore,

\( MR = MC \)

This equation implies that the firm seeking to maximize profits should produce at the point where marginal revenue equals marginal costs.

**Example 9** (Khisty and Lall, 2002)

A transport company hauling goods by truck has cost function \( C = 1.5q^{1.25} \), where \( C \) is the total costs of supply, \( q \).

(a) Determine the average cost and the marginal cost of production

(b) Prove that the cost elasticity is 1.25

(c) Is there an economy of scale?

**Solution:**

(a) \( \bar{C} = \frac{C}{q} = \frac{15q^{1.25}}{q} = 15q^{0.25} \) which is the average cost

\( MC = \frac{dC}{dq} = (15 \times 1.25)q^{0.25} = 18.75q^{0.25} \)

(b) \( e = \frac{MC}{AC} = \frac{18.75q^{0.25}}{15q^{0.25}} = 1.25 \) (see Eq.5.14 below)

(c) Economy of scale does not exist because the average cost increases with increases \( q \).
Cost Elasticity

The cost elasticity of a good or service is defined as the ratio of percentage change in cost \( C \) to a unit percentage change in supply \( q \). The difference between cost elasticity and price elasticity is obvious.

\[
e_c = \frac{\% \text{ in cost}}{\% \text{ in supply}} = \frac{(\Delta C / C) \cdot 100}{(\Delta q / q) \cdot 100} = \frac{q \cdot \Delta C}{C \cdot q}
\]

In the limit when \( \Delta q = 0 \), \( e = (q/C)(dC/dq) \). Rearranging terms yields:

\[
e = \frac{dC}{dq} \cdot \frac{C}{q} = \frac{MC}{AC}
\]

(5.14)

5.7 CONGESTION PRICING

Urban congestion is currently a serious problem in most countries, causing over $70 billion in wasted fuel, travel time and air pollution in the US alone. Besides applying an array of traffic mitigation measures such as physical capacity increases, intelligent transportation systems, and travel demand management, transportation agencies in many countries are considering the imposition of penalties as a disincentive to travel in congestion-prone areas such as central business districts. In such a scheme, road users pay a higher price to travel within an area at a time when congestion is expected to be higher. It is envisaged that introducing such an element of marginal cost pricing may help address the problem of urban congestion.

It is generally reasonable to characterize the short-run travel price paid by motorists by an average cost function as indicated by curve AC in Figure 5-12. The marginal costs curve MC is also shown. The demand curve is represented by D-D. After traffic flows up to OL per hour, the cost per trip is OA, comprising time and operating expenses. Beyond this point, the speed of vehicles falls, and therefore the average user cost per trip rises. In the absence of any further interference, the flow will equilibrate at ON vehicles per hour at a cost of OB.

As traffic flows above OL vehicles per hour, each additional vehicle is slowing down the entire flow of traffic and raising the operating costs of all other vehicles in the stream flow. In accord with the pricing principle, the price should equal the marginal cost, to give a flow of OM vehicles. This should be achieved by imposing a tax of GF, thus raising the average cost curve to achieve the optimal traffic flow. The benefit from this action is the reduction in the operating cost of all the remaining vehicles, and the loss is the loss in benefit from the MN trips than cannot be undertaken. This marginal cost pricing policy may result in most efficient use of the system (Khisty and Lall, 2002).
Figure 5-12: Congestion pricing
REFERENCES

1. BTS, 2001
2. Locklin, D.P., 1960
3. Wood and Johnson, 1996
4. Samuelson, 1976
5. Khisty and Lall, 2002
6. FHWA, 1996; AASHTO, 1996
7. FHWA, 1999
8. GASB, 1999
9. Manheim, 1979
10. McFadden, 1985, and Ben-Akiva, 1988
11. Salvatore, 1982
12. Islam and Sinha, 2002

(Full detail of incomplete references will be provided to the students in due course.)