UC Berkeley
Working Papers

Title
An Economic Analysis Of Network Deployment And Application To Road Pricing

Permalink
https://escholarship.org/uc/item/8hx7190m

Author
Levinson, David

Publication Date
1998

eScholarship.org Powered by the California Digital Library
University of California

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

An Economic Analysis of Network Deployment and Application to Road Pricing

David M. Levinson

California PATH Working Paper
UCB-ITS-PWP-98-1

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Report for MOU 275

January 1998

ISSN 1055-1417

An Economic Analysis of Network Deployment and Application to Road Pricing

funded as part of MOU 275

by David M. Levinson,

Institute of Transportation Studies

McLaughlin Hall, Rm. 109

University of California at Berkeley

Berkeley, CA 94720

home (address mail here):

David M. Levinson

1822 Francisco St. #11

Berkeley, CA 94703

dmlevins@uclink2.berkeley.edu

http://www.ce.berkeley.edu/~dmlevins/

(510) 848-3057

November 24, 1997

ABSTRACT

This paper develops an economic framework for developing strategies necessary to deploy networks, and applies the framework to the deployment of road pricing. The cost structure of highways are discussed. A graphical method for measuring welfare with road pricing is presented. The relationship of space and financing mechanism is reviewed. A network model of the economy is presented. This is followed by a discussion of network externalities, and how those relate to both the deployment and emergence of technologies. Finally, the deployment of three main elements relating to road pricing: use of electronic toll collection on existing toll roads, construction of new toll roads, and conversion of existing untolled road to toll roads are discussed.

Key Words: Road Pricing - Deployment, Congestion Pricing, Toll Roads, Turnpikes, Network Financing

TABLE OF CONTENTS

1. INTRODUCTION 7

2. COST STRUCTURE 9

3. CONGESTION PRICING 13

4. JURISDICTION SIZE AND LOCAL WELFARE 22

5. MONOPOLY AND COMPETITION 25

6. A NETWORK ANALYSIS OF THE ECONOMY 29

7. POSITIVE FEEDBACK AND NETWORK EXTERNALITIES 33

8. THE DEPLOYMENT OF ROAD PRICING 43

9. REFERENCES 46

TABLE OF FIGURES

FIGURE 2-1: MARGINAL AND AVERAGE COST VS. DEMAND 10
FIGURE 3-1: OPTIMAL CONGESTION TOLL AND WELFARE LOSS WITHOUT TOLL 15
FIGURE 3-2: (A) TIME VS. FLOW; (B) IMPLICIT AND REVEALED DEMAND VS. PRICE 17
FIGURE 3-3: WELFARE ANALYSIS, WITH AND WITHOUT PRICING 19
FIGURE 5-1: COMPETITION ON NETWORKS 25
FIGURE 6-1: A NETWORK MODEL OF THE ECONOMY 31
FIGURE 7-1: POSITIVE FEEDBACK BETWEEN NUMBER OF VEHICLES AND SIZE OF NETWORK 33
FIGURE 7-2: S-CURVES, A SCHEMATIC 34
FIGURE 7-3: GROWTH PATH OF CO-EVOLVING VEHICLE AND INFRASTRUCTURE SYSTEMS: A SCHEMATIC 34
FIGURE 7-4: CONSTRUCTION OF REVEALED DEMAND (FULFILLED EXPECTATIONS) CURVE WITH POSITIVE NETWORK EXTERNALITIES 37

Executive Summary

This research identifies critical technological, economic, and political factors associated with the deployment of technological networks in general and road pricing in particular. The development of an undeployable technology is of little use. In contrast to the large thread of research which focuses on optimal congestion prices, this research analyzes the political and economic constraints on, and opportunities for, deployment.

The deployment of road pricing is really comprised of three related but distinct elements: the deployment of electronic toll collection on existing turnpikes, which is already proceeding apace; the tolling of presently unpriced roads, presumably with electronic toll collection; and the construction of new toll roads. The elements are similar in their analysis, though the particular factors involved make some easier than others.

Section two of this paper considers the cost structure facing highway transportation today, reviewing the conventional economic theory and results from an empirical study of the costs of intercity highways. The cost structure, and whether marginal costs are higher or lower than average costs, will have a great influence upon the choice of the most appropriate financing mechanism.

In section three the theory of congestion pricing is discussed as is welfare measurement under tolling. Congestion reduction and more efficient allocation of resources are cited as some of the main benefits of road pricing, particularly peak period pricing. A new graphical approach is suggested which disentangles revealed demand at the given level of service (recognizing that level of service and demand are jointly determined) and underlying demand for travel at a given level of service. It is this underlying demand which should be used for welfare calculations, and which can suggest new approaches to differentiating the road network by level of service.

The relationship between jurisdiction size and road pricing is summarized in section four. Smaller jurisdictions are more likely to have non-local trips, a local welfare

maximization criteria suggests they will have greater incentives to toll. Decentralized control is thus a critical issue in the efficacy of road pricing for deployment.

Section five of this paper discusses how monopoly, competition, and hierarchy influence how road pricing might be deployed. Different actors: a government attempting to maximize welfare, regulated monopolies, or private unregulated firms, will have different desired revenue mechanisms and thus different pricing consequences. Deployment of road pricing needs to consider the regulatory regime and competitive situation of the network. Hierarchy is a related issue relating both to monopoly and competitiveness as well as cost structure and demand levels on the road network.

Section six of this paper develops a network analysis of the economy. Production, exchange, and transportation are described in network terms. This network analysis of the economy enables an economic analysis of the transportation network. The economic analysis is required to understand and speculate about the deployment of advanced technologies, such as road pricing.

The ideas of positive feedback in the economy through network externalities, learning curves, path dependence and cumulative causation, and self-fulfilling prophesies are discussed in section seven. The success of network deployment depends on these conditions. S-Curves can be used to model how a given technology is deployed over time, showing the gestation period, take-off, and saturation of a technology. A model of co-evolution shows technological change over time, and the interdependence of complementary technologies.

Finally, these seemingly disparate parts are brought together to develop a scenario and strategy for the deployment of electronic toll collection, new priced roads, and converting existing untolled roads to toll roads. While deploying electronic toll collection on existing toll roads should be neither politically nor technologically difficult, it is important that consistant standards be employed so that users can easily switch from one toll network to another without needless inconvenience. New limited access roads have a reasonable likelihood of being toll roads when they are locally funded and serve both local and non-local traffic, especially as electronic toll collection reduces transaction costs.

Converting existing untolled roads to toll roads poses the most difficulty - though a scenario involving first establish toll cordons, rings around metropolitan areas which require payment to cross, has some potential to be the critical first step for tolling to become widespread. New networks, such as automated highway systems, will also require a financing mechanism targeted to their users, and tolls are promising there too.

1. INTRODUCTION

Policy makers face the problems of roadway congestion, air pollution from cars, and the dearth of resources to finance new infrastructure. Transport economists often suggest road pricing, charging users a monetary toll for the use of a specific part of the roadway network, as a solution to these problems. While tolls are common for certain expensive facilities such as tunnels and bridges, they are less common on streets and highways, which are more typically funded from user taxes or general revenue. However new electronic toll collection technologies are changing the costs associated with toll financing, while new technologies such as automated highways will require financing. This research, a companion piece to the Case Study *Road Pricing In Practice* (Levinson 1997), identifies critical technological, economic, and political factors associated with the deployment of technological networks in general and road pricing in particular. The development of an undeployable technology is of little use. In contrast to the large thread of research which focuses on optimal congestion prices, this research analyzes the political and economic constraints on, and opportunities for, deployment.

The deployment of road pricing is really comprised of three related but distinct elements: the deployment of electronic toll collection on existing turnpikes, which is already proceeding apace; the tolling of presently unpriced roads, presumably with electronic toll collection; and the construction of new toll roads, again with electronic toll collection. The elements are similar in their analysis, though the particular factors involved make some easier than others.

This report is organized in a star-shaped, rather than linear, fashion. Each of the sections is in some ways independent of the others, but they are all brought together in the concluding section on road pricing deployment strategy. Each section provides some insight into a salient factor bearing on road pricing, network deployment, or both. The report is intended to inform the implementers of road pricing of some of the relevant issues. In particular it focuses on how the nature of the network, its internal organization, cost structure, location, market situation, and historical situation influences the success and

value of deployment. It is hoped that the separate ideas presented here will be useful in the consideration and development of a road pricing deployment strategy.

Section two of this paper considers the cost structure facing highway transportation today, reviewing the conventional economic theory and results from an empirical study of the costs of intercity highways. The cost structure, and whether marginal costs are higher or lower than average costs, will have a great influence upon the choice of the most appropriate financing mechanism.

In section three the theory of congestion pricing is discussed as is welfare measurement under tolling. Congestion reduction and more efficient allocation of resources are cited as some of the main benefits of road pricing, particularly peak period pricing. A new graphical approach is suggested which disentangles revealed demand at the given level of service (recognizing that level of service and demand are jointly determined) and underlying demand for travel at a given level of service. It is this underlying demand which should be used for welfare calculations, and which can suggest new approaches to differentiating the road network by level of service.

The relationship between jurisdiction size and road pricing is summarized in section four. Smaller jurisdictions are more likely to have non-local trips, a local welfare maximization criteria suggests they will have greater incentives to toll. Decentralized control is thus a critical issue in the efficacy of road pricing for deployment.

Section five of this paper discusses how monopoly, competition, and hierarchy influence how road pricing might be deployed. Different actors: a government attempting to maximize welfare, regulated monopolies, or private unregulated firms, will have different desired revenue mechanisms and thus different pricing consequences. Deployment of road pricing needs to consider the regulatory regime and competitive situation of the network. Hierarchy is a related issue relating both to monopoly and competitiveness as well as cost structure and demand levels on the road network.

Section six of this paper develops a network analysis of the economy. Production, exchange, and transportation are described in network terms. This network analysis of the economy enables an economic analysis of the transportation network. The economic

analysis is required to understand and speculate about the deployment of advanced technologies, such as road pricing.

The ideas of positive feedback in the economy through network externalities, learning curves, path dependence and cumulative causation, and self-fulfilling prophesies are discussed in section seven. The success of network deployment depends on these conditions. S-Curves can be used to model how a given technology is deployed over time, showing the gestation period, take-off, and saturation of a technology. A model of co-evolution shows technological change over time, and the interdependence of complementary technologies.

Finally, these seemingly disparate parts are brought together to develop a scenario and strategy for the deployment of electronic toll collection, new priced roads, and converting existing untolled roads to toll roads.

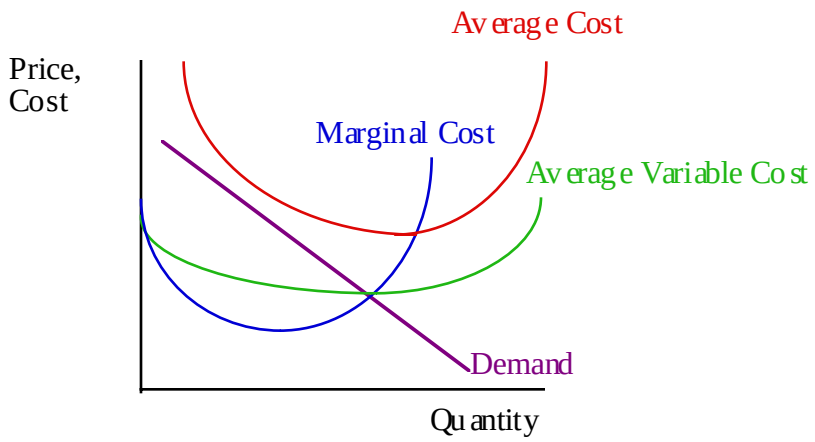
2. COST STRUCTURE

Network operators are distinguished from many businesses by their cost structure. The cost structure will have a great influence upon the choice of the most appropriate financing mechanism. For instance, congestion pricing won't recover the full costs of highway infrastructure if average costs are higher than the marginal costs used to price the roadway. This section reviews the theory and empirical evidence underlying the cost structure of highways.

Economists often speak of a U-shaped average cost curve which combines fixed costs and variable costs. At low levels of demand, fixed costs dominate, while at high demand, the variable costs dominate as the average fixed cost steadily declines. Average costs decrease initially as the fixed costs are spread over more and more units, and then rise as the marginal costs increase due to using more and more of a fixed factor of production. A manufacturer of widgets must expend some fixed sum of money to build a factory, plus some cost per widget manufactured. The cost per widget may depend on the number of widgets manufactured. In a thin market, a market which our widget manufacturer

dominates as a monopsonistic or oligopsonistic consumer of raw materials, at high volumes, costs may rise due to scarcity in raw materials.

Figure 2-1: Marginal and Average Cost vs. Demand



In an efficient and competitive market, the good is supplied along the marginal cost curve, the optimal point of production is where marginal cost intersects the point of lowest average variable cost and the demand curve. Implicit in this analysis are smooth continuous curves, and increasing marginal costs. Moving from left to right, first average cost is high, due to the fixed cost of investment which is spread over few goods, but the share of fixed cost per unit output declines as output increases. There is a point however where rising variable costs per unit overtake the declining fixed costs.

To what extent do physical networks in general, and highway networks in particular follow this structure? Clearly there are high fixed costs: right-of-way, grading, construction. However variable costs are in large part borne by users rather than the network operator in terms of time wasted in delay. Levinson and Gillen (1997) studied the cost structure of state highway departments empirically, conducting an econometric analysis of their costs as a function of input prices (labor, capital, materials) and use (vehicle kilometers traveled by car, single unit trucks, and combination trucks). Short run costs included the costs of operating, maintaining, and administering highways. Long run costs included the short run costs plus the cost of constructing infrastructure, which was measured using capital stock. The following table (2-1) comparing the short and long run costs of intercity highway travel by cost category is reproduced here:

Table 2-1: Long and Short Run Marginal and Average Incremental Infrastructure Costs and Scale Economies ($/vehicle-kilometer)

	Auto	Single Truck	Combination Truck
Long Run			
Marginal	0.0188	0.0431	0.0514
	(0.0072-0.0331)	(0.0205-1.33)	(0.0193-0.1349)
Average Incremental	0.017	0.063	0.101
S = IC/MC	0.92	1.45	1.96
Economies of Scale	Decreasing	Increasing	Increasing
Short Run			
Marginal	0.0055	0.0075	0.0003
Average Incremental	0.00075	0.0298	0.0032
S = AIC/MC	0.14	3.97	10.67
Economies of Scale	Decreasing	Increasing	Increasing

Note: Parentheses refer to range of state level highway agency costs.
ref.: Levinson and Gillen, 1997, Table 7

For automobiles, there are slight decreasing economies of scale in the long term, so that each additional vehicle kilometer traveled increases costs more than the preceding vehicle kilometer traveled. The opposite is true of trucks. This suggests the system is operating on the right side of the U-shaped cost curve for automobiles, and to the left side for trucks. This is an interesting, and perhaps counter-intuitive result. To understand it, consider that one main reason for rising marginal costs is the presence of congestion in the peak period. These are not congestion costs per se, but the costs of solving the congestion problem, namely building wider roads in developed areas where land costs are highest. When we recognize that peak period travel is disproportionately by automobile, while trucks tend to use roads more heavily before and after the peak, it is not as counterintuitive to see rising marginal costs on the auto and not the truck side. However the range of costs by state is fairly broad, indicating that there is not a definitive answer to the question. Short run costs are much smaller than long run costs, indicating that the fixed costs are the dominant consideration, but that the cost of capital facilities rises the more travel there is. Reasons include the increasing costs of the inputs to capital facilities such as land and labor increase with the amount of travel, as both reflect factors like urban density and size which bid up prices.

Table 2-2 compares infrastructure costs to other costs of travel. Time costs (congestion and freeflow time) have been converted to monetary costs using a $10/hour value of time. Accident costs are calculated with both direct monetary costs and assuming

a $2.9 million value of life. A fuller description is available in Levinson and Gillen (1997). The analysis shows that infrastructure is a relatively small component in total costs. This table measures the costs in dollars per vehicle kilometer traveled. Other elements, such as congestion, which clearly do have rising marginal costs, are far more significant.

Table 2-2: Average and Marginal Long and Short Run Costs by Category ($/vehicle-kilometer)

Cost Category	Short Run Marginal Cost	Short Run Average Cost	Long Run Marginal Cost	Long Run Average Cost
User Fixed + Var.	$0.049	$0.130	$0.049	$0.130
Infrastructure	$0.0055	$0.00075	$0.019	$0.0174
Freeflow Time	$0.15	$0.15	$0.15	$0.15
Congestion	$0.049	$0.0045	$0.049	$0.0045
Accidents	$0.035	$0.031	$0.035	$0.031
Noise	$0.009	$0.006	$0.009	$0.006
Air Pollution	$0.0056	$0.0056	$0.0056	$0.0056
Total	$0.30	$0.33	$0.32	$0.34

ref.: Levinson and Gillen, 1997, Table 10

The cost structure of a highway department or a trip is the aggregation of the costs of numerous individual links which show greater variation. The vast majority of links are uncongested, though a high percentage of travel occurs, as would be expected, on the relatively congested links. Adding a lane-kilometer to a large network will involve a small increase in cost, which can be modeled as a continuous, rather than lumpy or scalloped, function. However, from the point of view of the individual link, one sees a much "lumpier" investment pattern, lanes can only be added in integer values, so that each expansion of a link involves a significant percentage increase in costs.

Just as networks may have an economy of consumption, where more users spread fixed costs results in saving to each users, there are diseconomies of consumption. More users result in either congestion (the congestion externality) or higher prices in an attempt to allocate a scarce good (a pecuniary externality). Demand in excess of some fixed capacity over a period of time leads to delay - the all too familiar being stuck in traffic due

to queueing at a bottleneck or greater inter-vehicle interactions. Congestion is not restricted to transportation systems. For phone service there are missed connections and busy signals, in electricity transmission there may be brownouts when circuits overload. However these problems are not the result of peak pricing, but the result of incomplete and imperfect peak pricing, and are largely reduced by the presence of congestion pricing in those networks. Compare the failures of phone or electric networks with the amount of delay on the unpriced or underpriced road system. To the user/consumer and for the system overall, there may be some advantage to congestion pricing to ration scarce resources.

3. CONGESTION PRICING

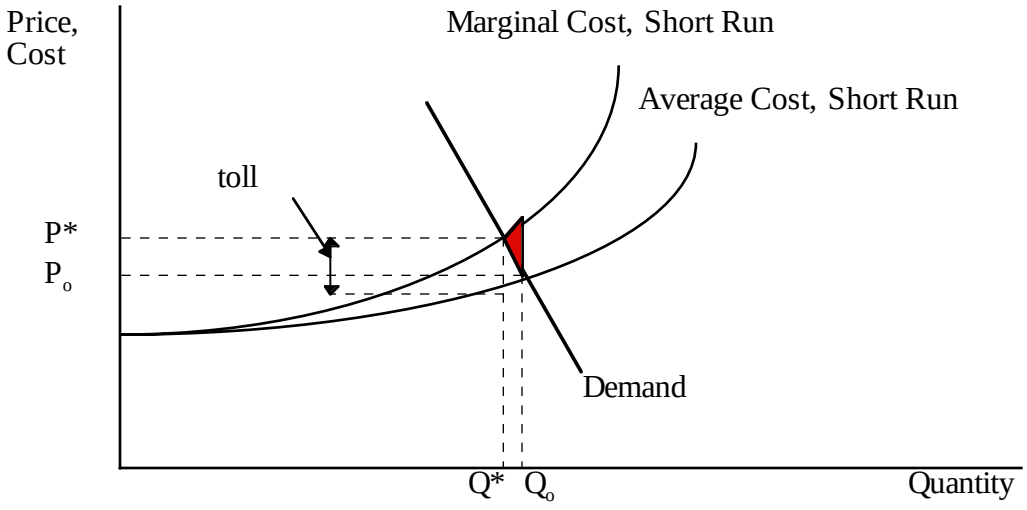
In this section, the theory of congestion pricing is discussed as is welfare measurement under tolling. Congestion reduction and more efficient allocation of resources are cited as some of the main benefits of road pricing, particularly peak period pricing. Qualitatively, the idea behind congestion pricing is this: person R has a high value of time, person P a low value of time. Without pricing, persons R and P both travel at a slow speed. But if roads are priced, person R will be able to pay money and travel faster, while person P will not pay the money and not travel at that time (or travel on more congested and slower free roads). To work, the money collected needs to be redistributed to persons R and P in some fashion, either through lowering other taxes, through direct payments, or by reinvesting it in transportation. If person R's value of time saved plus the amount returned is greater than the amount paid, R is better off. If the amount of money returned to person P is greater than the cost of deferring the trip (or traveling at a slower speed), then P is better off. Road pricing will inevitably create both winners and losers (and usually losers) without redistribution of the toll which was collected. However, under the right redistribution policy, most people can be made better off.

In this section, a new graphical approach is suggested which disentangles revealed demand at the given level of service (recognizing that level of service and demand are jointly determined) and underlying demand for travel at a given level of service. It is this underlying (or implicit) demand which should be used for welfare calculations, and which can suggest new approaches to differentiating the road network by level of service.

Explaining the advantages of congestion pricing to a non-technical, or even non-economist audience is difficult. The task is made more difficult by the choice of graphs and assumptions used in the explanation. Often, the graphs do not permit the use of standard economic tools like consumer's surplus. The difficulty lies with the use of generalized cost and a revealed demand curve, rather than the use of a money cost and the multiple underlying demand curves reflecting different demands for road use at different levels of service (travel times). This section seeks a more straight-forward development of the justification of congestion pricing from a graphical perspective.

The conventional explanation of road pricing, found in various sources, uses a variation of Figure 3-1. On the y-axis is a measure of generalized cost (e.g. price plus monetized time), on the x-axis is flow in vehicles per hour. In the absence of any toll, equilibrium occurs at (Q_o, P_o), where demand intersects the short run average cost curve. Any traveler who values a trip more than P_o will travel, anyone who doesn't won't travel. The shaded (red) area on the graph is considered the welfare loss, the benefit which is lost when tolls are not imposed. The loss is due to the difference between the cost a driver imposes on society (the short run marginal cost) by making everyone else's trip take a little bit longer, and the cost that driver bears personally which is spent in traffic congestion due to all the other cars on the road (short run average cost). The imposition of a marginal cost toll moves the equilibrium to (Q*,P*) and eliminates the welfare loss due to the congestion externality.

Figure 3-1: Optimal Congestion Toll and Welfare Loss Without Toll



Legend:

P* = Optimal Price with Tolls Q* = Amount of Travel with Tolls

P_o = Price without tolls Q_o = Amount of Travel without Tolls

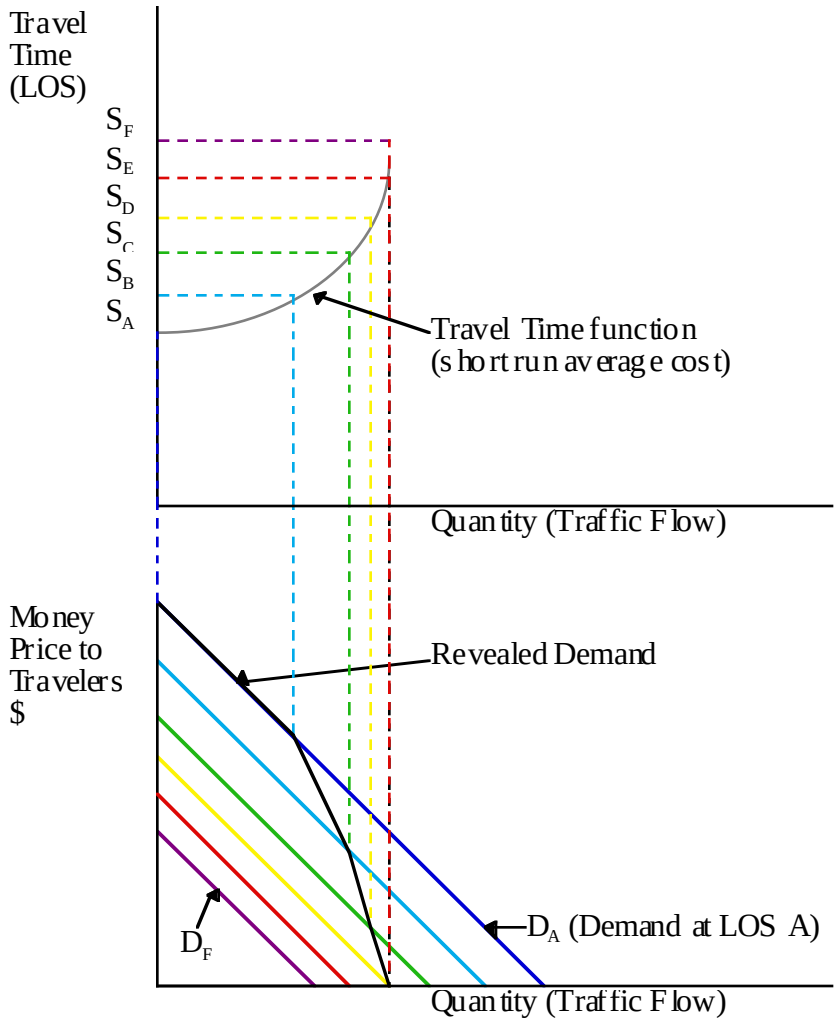
Welfare Loss without Tolls

However, the use of a single demand curve on the graph confuses the issue, since clearly moving from the short run average cost to short run marginal cost has a welfare implication, raising the price lowers the demand, and thus the area that is conventionally thought of as consumer surplus gets smaller. But whether consumer's surplus gets larger or smaller depends on how individuals value the time savings. The conundrum results from the fact that individual drivers would pay more for a better level of service (LOS), and thus in actuality, the movement from short run average cost to short run marginal cost implies a movement from a demand curve with poor LOS (D_F) to a demand curve with a better LOS (D_A).

The demand for a graded commodity at a given price depends on the grade of the commodity. In the case where the commodity in question is road use, the grade is the level of service, the time it takes to traverse the road. At better levels of service (lower travel times) the demand will be higher at the same money price than at lower levels of service. We will describe LOS as ranging from S_A to S_F, with S_A being best.

Suppose that there is some money price (a toll) charged by the agent managing the road, such that, even if the travel time is zero, the quantity demanded will be very small or zero. At a zero price, even if the travel time is small or zero, the quantity of travel will be limited. Similarly, there is a travel time at which demand will be small or zero, even at zero price. This can be represented by the following graphic (Figure 3-2).

Figure 3-2: (A) Time vs. Flow; (B) Implicit and Revealed Demand vs. Price



A. (Top) Travel Time as Function of Traffic Flow

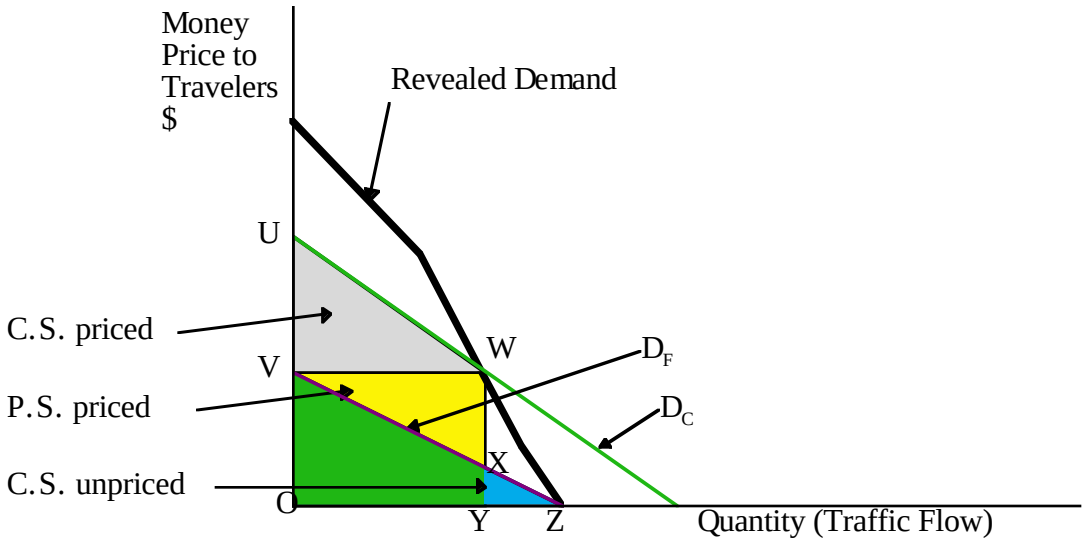
B. (Bottom) Traffic Flow (Implicit and Revealed Demand) as a function of Money Price to Travelers

The top part of Figure 3-2(A) shows schematically the travel time to a driver (short run average cost) at a bottleneck or on a capacitated link resulting from various levels of approach flow. The travel time function relates travel time (or delay) and approach traffic flow. The greater the approach flow, the higher the travel time. At flows below capacity (level of service A (S_A) or B (S_B)), traffic flows smoothly, while at high approach flows, those above capacity, traffic is stop and start and is classified as level of service E (S_E) or F (S_F).

The bottom part of Figure 3-2(B) shows schematically the implicit demand for travel on a link as a function of the travel time. All else equal (for instance the price charged to users), demand to travel on a link at level of service A (D_A) is higher than demand at level of service F (D_F). However the demand and the travel time on a link are not independent, as shown in Figure 3-2(A). So the implicit demand and revealed demand are not identical, rather the revealed demand is formed by projecting the travel time at a given flow onto the implicit demand curves. So for instance, when the price charged users is high, the revealed demand coincides with the implicit demand at level of service A (D_A). As the prices are lowered, the revealed demand crosses the implicit demand curve at level of service B (D_B), then D_C, D_D, D_E and finally at a zero money price it crosses D_F. While the actual prices that generate specific demand levels vary from place to place with local circumstances, demand preferences, and market conditions, the general trend (higher prices gives lower approach flow gives better level of service) is simply an application of the law of demand from economics along with traffic flow theory.

In other words, the change in welfare with and without congestion pricing depends not only on both the change in price and quantity, but also on the change in reservation price, the price travelers would be willing to pay at a given level of service. And at better levels of service, travelers (and potential travelers) have a higher reservation price.

Figure 3-3: Welfare Analysis, With and Without Pricing



Legend

Consumer's Surplus Priced = UVW	Producer's Surplus Priced = OVWY
Consumer's Surplus Unpriced = VOZ	Producer's Surplus Unpriced = 0

The movement along the revealed demand curve follows the shape of the curve shown above because of the relationship between traffic flow (quantity demanded) and travel time. Assume for instance that each level of service category represents a one minute increase in travel time from the immediately better travel time. So in the graph, let the level of service for a one minute trip be denoted S_A , and for a six minute trip, S_F. The amount of traffic necessary to move from 1 minute to 2 minutes exceeds the amount to move from 2 to 3 minutes. In other words there is a rising average (and thus marginal) cost in terms of time.

The concepts in Figure 3-2 can be used to develop the welfare analysis shown in Figure 3-3. There are several areas of interest in Figure 3-3. The first is defined by the lower left triangle (the blue + green) (triangle VOZ) which is the consumer surplus when the road is unpriced. The second is the producer surplus (profit) to the road authority when the road is priced, illustrated by the rectangle formed in the lower left (yellow + green) (rectangle OVWY). The third is the consumer surplus when the road is priced, shown in

gray (triangle UVW). This consumer surplus represents a higher reservation price than the other because the level of service is better when flow is lower. That first area needs to be compared to the sum of the second and third areas. If the sum of the second and third areas (OUWY) is larger than the first (OVZ), then pricing has higher welfare than remaining unpriced. Similarly, two price levels can be compared. In other words, the welfare gain from pricing is equal to the yellow + gray area (VUWX) minus the blue area (XYZ). In this particular figure, consumer's surplus is maximized when the good is free, but overall welfare (including producer's surplus) is not. Whether consumer's surplus is in fact higher in a given situation depends on the slopes of the various demand curves.

Welfare is maximized by maximizing the sum of the producer's surplus rectangle and the consumer's surplus "triangle" (it may not be a true triangle), recognizing that the consumers surplus triangle's hypotenuse must follow an underlying demand curve, not the revealed demand curve. Differentiating the level of service (for instance, providing two different levels of service at two different prices) may result in higher overall welfare (though not necessarily higher welfare for each individual).

How welfare is measured and how it is perceived are two different things. If the producer's surplus is not returned to the users of the system somehow (through rebates of other taxes or reinvestment in transportation), the users will perceive an overall welfare gain as a personal loss because it would be acting as an additional tax. It should be noted that the entire argument can be made in reverse, where consumer and producers surplus are measured in time rather than money, and the level of service is the monetary cost of travel. This however has less practical application.

In low volume situations, those which are uncongested, it is unlikely that the revenue from marginal cost congestion pricing will recover long term fixed costs. This is because the marginal impacts of an additional car when volume is low is almost zero, so that additional revenue which can be raised with marginal cost pricing is also zero. Imagine a road with one car - the car's marginal impact is zero, a marginal cost price would also be zero, its revenue would thus be zero, which is less than the fixed costs. Add a second car, and marginal impacts are still nearly zero - a phenomenon which remains true until capacity is approached.

4. JURISDICTION SIZE AND LOCAL WELFARE

In this section (adapted from Levinson 1998) some issues relating to roadway network financing over space are developed. The relationship between jurisdiction size and road pricing is summarized. Smaller jurisdictions are more likely to have non-local trips, a local welfare maximization criteria suggests they will have greater incentives to toll. Decentralized control is thus a critical issue in the efficacy of road pricing for deployment.

The idea of decentralized, local control and multiple jurisdictions distinguishes this approach from one where a central authority maximizes global welfare. The main idea in the analysis described below is that jurisdictions responsible for network financing have the objective of local welfare maximization. Local welfare reflects the consumer's surplus of residents of the jurisdiction and the profits accruing to the locally controlled network authority that the jurisdiction owns and manages.

The network operator has several actions which can be taken to maximize local welfare. The set of actions of interest to us is the selection of a revenue instrument (such as taxes or tolls, with various payment schedules) and the setting of a price level. Collectively the revenue instrument and the price level are called the revenue mechanism. The main complication is the joint production and consumption of the key good (network services) by the jurisdiction and its residents. Jurisdiction residents use both local and non-local network, and the network is used by both local and non-local residents. The proportion of trips using the network which originate in the jurisdiction of control directly shapes the local welfare resulting from a particular revenue mechanism, and itself is a function of jurisdiction size. The choice of financing instrument must trade-off the number of spatial free riders, system users who don't pay their cost because of the location, and the costs of collection. The price charged with a given instrument is limited by the elasticity of demand for use of the network on those who are charged.

While development of the model and its application are beyond the scope of this report (see Levinson 1998), the results are fairly straight-forward. The central question facing jurisdictions is "Should I Tax or Should I Toll?". The answer is that it depends. It

depends on the various empirical constants defined here which relate to the unit cost of various cost components and demand elasticity. It also depends on jurisdiction size.

Levinson (1998) identified the conditions where tax, toll, or mixed tax/toll policies were preferred. Cordon tolls (tolls placed on jurisdiction boundaries) by themselves are economically unsustainable as jurisdictions get large. Large jurisdictions are more likely to impose taxes or a mixed financing policy than only cordon tolls because cordon tolls raise insufficient revenues to cover costs, as revenue levels off above a certain point. In uncongested conditions with low variable costs, use of interior (non-cordon) tolls does not enhance local welfare as any additional revenue raised compared with cordon tolls comes from local residents, except to the extent that the tolls reduce over-use and social loss (see section 3).

Similarly, the higher the cost of collecting tolls, the less likely tolls will be the preferred revenue mechanism. The welfare maximizing toll may not fully recover costs, and thus still require subsidy (thus toll-only financing may be unsustainable). The maximum welfare from taxes may exceed those of tolls under certain circumstances, depending on model parameters, however if jurisdictions are sufficiently small, demand sufficiently high, and collection costs relatively low, then tolls will be preferred. Hence collection costs need to be fairly high before no-tolls is a better solution than some tolls.

The gains to a jurisdiction of imposing tolls exceeds the gains from taxes under certain circumstances. The gains come foremost from residents of other jurisdictions. This problem, a finance externality, is well known in certain cases, for instance the reliance by local governments on some mix of sales, income, and property taxes, each of which are borne by a different set of people, not all of whom are local.

Congestion pricing has long been a goal of transport economists, who argue that it will results in more efficient use of resources. The path for implementing such a system has been strewn with political potholes, pricing inevitably creates winners and losers. An alternative approach, one which would create the local winners necessary to implement road pricing, is required before congestion pricing can be expected to become widespread. Levinson (1998) suggests one approach, one that would decentralize the decision about

whether to tax or toll before attempting to impose road pricing. Road pricing is a necessary prerequisite to congestion pricing. Once tolls are in place, peak period pricing differentiation is not a difficult problem, but placing tolls on untolled roads in the first place is difficult. And tolls are a rational financing mechanism for a sufficiently small jurisdiction, particularly with the advent of electronic toll collection systems. In short, the prospects for future success of toll roads depend on several factors, including the relative centralization of control of the highway sector, and the transaction costs of collecting revenue. Factors that would be conducive to a return to turnpikes are a reduction in collection costs and a decentralization of authority. Should the governance become more decentralized, and collection costs continue to drop, tolls could return to prominence as the preferred means of financing roads for both local and intercity travel.

5. MONOPOLY AND COMPETITION

On an unpriced transportation network owned by the government, the issue of the monopoly power of links is seldom raised. It only matters to the extent that those links form bottlenecks and are expensive to expand, or are vulnerable to catastrophe. However when roads are priced, and possibly not owned by the government - the monopoly power held by networks needs to be considered. This section discusses monopoly and competition on networks.

Four sources of monopoly power are identified here: (1) spatial location, (2) economies of scale and scope, (3) patents, and (4) other regulatory. Types (1), (2), and (4) are of particular concern in road pricing. So long as roads are dispersed and transportation is not costless, each road segment has some semblance of spatial monopoly. Objective economies of scale and scope may dictate that only one producer is economically efficient. Patents for unique processes provide protection in law or in fact to prevent other companies from performing the same technological transformation. The regulatory regime may create monopolies even in the absence of economies of scale, just as patents protect the producer of certain items or certain production processes.

5.1 Spatial Monopoly

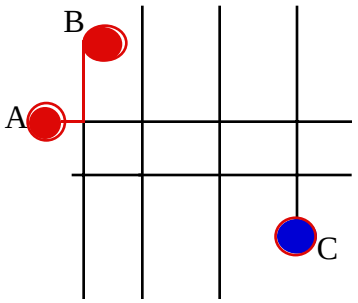
Especially for the access rather than movement function of transportation networks, strong spatial monopoly power of individual segments exists. While there is always some alternative (walking, if not transit), the costs of the alternative are often significantly higher than the costs of using the monopoly provider, even when the monopoly provider is priced and unregulated. The degree to which the costs of the alternative is higher suggests the strength of the monopoly.

The monopoly property can be illustrated using a simple network. Between adjacent points there may only be one path, between distant points there are multiple paths, though all the paths may require using certain critical links.

In Figure 5-1, travel from A to B entails only one path (other paths use the required path plus additional links, which means they cannot be the shortest), in fact anyone attempting departing A must use one particular link. However, the entire path from A to C can include a variety of different links (in addition to the link on which A is found. In this case all of the dead-end (or cul-de-sac) links are monopolies, while other links are potentially competitive.

The model can be extended from strictly defined cul-de-sacs, to "functional cul-de-sacs" which are links which, while nominally interconnected remain under one management to achieve economies or operate as cul-de-sacs in that long distance (non-local) traffic won't use them because of their poor service level/low speeds.

Figure 5-1: Competition on Networks



The ideas developed here are in the context of transportation, though they can be applied to other networks. Purely physical networks such as a steel frame structure or a storm sewer system have similar force (flow) distribution properties as a road network without involving economics. A load in a building will definitely fall on certain members, but will be transmitted in some equilibrium fashion (depending on other loads) down the building to the ground.

The hierarchical model (different functions for different links) can apply not only to roads, but to local telecommunications and energy distribution networks. As noted above, competition is often found between modes (networks of alternative technologies broadly serving the same function) rather than multiple providers of the same mode. In long distance transportation, there is competition between rail, road, and air. In communication, there is competition between wired and wireless telephony. In energy there is competition

between electricity, home heating oil, and natural gas for certain uses (e.g. home heating and cooling, certain appliances). Electricity can be privately produced, there are opportunities for co-generation, and large users can bypass the local grid. Monopoly network operators are often enmeshed with potentially competitive economic units. Local electricity distribution (from high-voltage transmission lines to the end-user) are monopolies owned by electric utility. However energy generation (power plants) are potentially competitive enterprises which need not be. The utility can be treated as a common carrier, which sells carriage of electricity between competitive producer and sovereign consumer. This would be analogous to the current environment for long distance telephony.

5.2 Economies of Scale and Scope

Section 2 of this report investigated the cost structure of highway networks and found no economies of scale in the provision of highways at the scale of the state. However, at a more local level, such economies almost surely exist. It must be cheaper to manager all of the streets in a town with one organization than with multiple organizations, very basic road maintenance and traffic operations issues have certain scale economies, which while not necessarily applicable at very large levels, do exist at smaller jurisdictions.

Economies of scale and scope may be internal or external to the organization The network externality, described in section 7, discusses more fully the ideas behind economies which are external to the organization These economies exist when multiple individuals join a network and adopt the same standard to connect with each other, making that network more valuable than if they were all free-standing and unconnected. A given network may have key components controlled by a single organization , which thus enjoys monopoly power not because of internal scale economies, but because everyone else is tied into that organization's network.

5.3 Regulatory Regime

Different actors: a government attempting to maximize welfare, regulated monopolies, or private unregulated firms, will have different desired revenue mechanisms

and thus different pricing consequences. A deployment strategy for road pricing needs to consider the regulatory regime under which priced roads will operate.

The examples of electric power distribution and the "local loop" in telephony may suggest lessons for how priced roads (either in the private sector or as an independent public authority) might be regulated. Electricity distribution and local telephone tend to have monopoly power, and are generally treated as regulated monopolies, under the presumption that regulation will improve welfare. A difficulty is that regulation in general, and regulation of monopolies in particular leads to price structures determined not by an efficient market but by lobbying, bargaining, or negotiation. The motives of the players in the regulation game (the regulators, the regulated producers, the public or consumers) are distinct, the regulated have strong motives and are concentrated, while the public is diffuse, and the regulators, seeking to mediate between the two feel more pressure from one side than the other, leading to what has been termed "regulatory capture" of the regulating agency by the regulated.

6. A NETWORK ANALYSIS OF THE ECONOMY

This section develops a network analysis of the economy. Production, exchange, and transportation are described in network terms. This network analysis of the economy enables an economic analysis of the transportation network. The economic analysis is required to understand and speculate about the deployment of advanced technologies, such as road pricing, and how those technologies interact and depend on each other.

Several economic properties of networks are relevant for understanding the effectiveness of deployment strategies. However, we should begin by describing what we mean by a network. The term network comes from "net", and its definition is "1. an openwork fabric or structure in which rope, thread, or wires cross at regular intervals. 2. Something resembling a net" (Websters II, 1984). The central idea is thus of connections between links which reinforce each other. These links can be physical (threads, wires, beams, highways, rails, pipes) or socio-economic (kinship, social, or exchange relationships). The market on the other hand is a place (real or virtual) where exchange takes place. An economic network may be comprised of multiple markets. A market may sell the right to use, or the ownership of, physical networks.

Figure 6-1 illustrates conceptually the idea of an economic network. There are three main elements: the site of production/consumption (material transformation), the site of exchange (ownership transformation), and the connection between the two (spatio-temporal transformation). While each of these elements is modeled as a link or node, it should be remembered that each can be expanded to form a subnetwork of itself if there is a desire to increase the detail or resolution of the analysis. A production/consumption agent in an economic network has both suppliers and customers, and can be modeled as an "agent node" on a network. On Figure 6-1, the open or hatched circles indicate agent nodes. Because production and consumption are two sides of the same coin, they are referred to together, any process consumes inputs to produce outputs. The "exchange nodes" are defined by the convergence of "connection links," and are analogous to markets. On Figure 6-1, exchange nodes are represented by filled circles. The agent nodes are connected to exchange nodes by special "connection links" (shown on Figure 6-

1 as dashed lines). Connection links account for transportation or communication costs in the production system. The flows in one direction are goods that are input into the production process, transformed and output as a refined good(s). The flows in the other direction represent money (or a monetary equivalent) that is paid for the goods.

In the model represented by Figure 6-1, an agent (firm or individual) purchases goods in an input market (Stage 1), and may be supplied by any (or all) firms in that input market. The goods are brought to the "factory," (the term is used loosely) transformed (Stage 2), and sold in the output market to any or all customers (Stage 3). The firm is complementary to any firm in the input market and to its customers, while it is competitive with parallel and unconnected nodes.

Clearly this situation is idealized. Some firms may have different degrees of vertical integration, that is they may internalize what is represented here as an input market or the output market. However, this figure does reflect that a production process may have economies of scope, so that a single firm produces for more than one output market, as is shown in Figure 6-1 between Stage 2 and Stage 3. In the illustration, there are three stages (1,2,3 from left to right) several markets in each stage (for instance a market for capital and a market for labor) and multiple firms in each market. Extending the chain far enough to the left and to the right, and incorporating enough of the economy, the markets connect with each other again, as the ultimate final consuming "firm" is the individual consuming goods and an ultimate input "firm" is an individual producing labor.

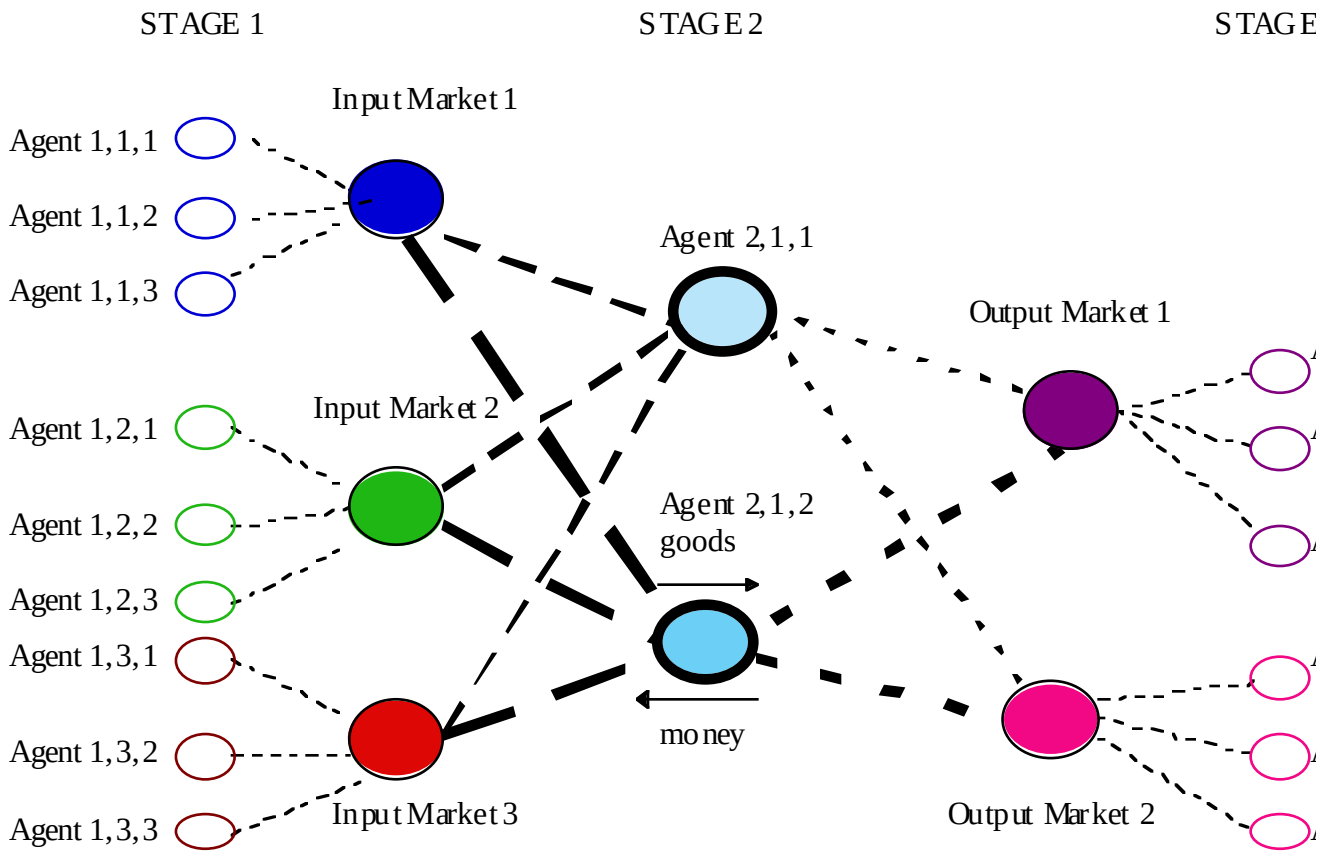
To compare with a conventional transportation network, a roadway link is a composite of the "agent node" and the "connection link". For each link on a highway, there is only one input market and one output market, each identified with a single node (an intersection), which makes the graphic representation and analysis simpler as the agent nodes are unnecessary because the transformation is only spatial, not material. While there is "conservation of flow" in the network, flows can be one way, the link moves traffic in one direction with nothing in return. As part of a larger system, the link (more precisely, an agent: Department of Transportation, Turnpike Authority, private firm acting on behalf of the link) receives revenue from government or users, which is used to maintain the link.

In one sense, the link is selling the right to be traveled on and is paid by users or government for this right. If it is not paid, it deteriorates over time (the payment comes from the link's own capital stock which is dissipated). The more generalized version of a graphed economy subsumes the transportation network as a special case. The use of this framework serves to incorporate, at least conceptually, financing in the standard highway network analysis, and thereby allows us to identify some pertinent issues.

In particular if we identify links with firms, the issue of payment becomes clear. In order to operate, the link must be subsidized by government, be paid for directly by users, or allow its capital stock to deteriorate. Direct payment from users equal to the marginal cost is clearly more efficient, it does not entail the social loss described in section 2 due to overuse and subsidy, and does not impose deadweight losses inherent in certain taxing structures. Imposing road pricing is a natural conclusion to these problems.

Figure 6-1 is a snapshot, it describes the processes and relationships at a given point or window of time. Over a long period of time, links and nodes are added and deleted as the economy grows and contracts, markets change, and innovation occurs in response to entrepreneurship and invention. The purpose of this analysis is to provide a tool to examine how networks and relationships in general do happen. We might extend the standard network flow idea of the least cost path to the process. Then "final" customers on the right side purchase a bundle of goods which provides the highest utility or lowest cost, profit seeking production/consumption agents in the middle will act as efficient customers for the initial producers on the left, and efficient producers/transformers in their own right. The network will generate welfare maximizing flows under the usual strong assumptions from microeconomics: well defined property rights and the absence of externalities (or when there is internalization of externalities), the presence of competitive links throughout, convex cost functions, etc. The interesting cases are in the absence of one or more of those conditions. Furthermore, the degree to which the network itself is efficient is another, much more complex (and important) question.

Figure 6-1: A Network Model of the Economy



Nomenclature: Agent: stage s, market m, firm number n

Open or hatched circles indicate production/consumption agent nodes

Filled circles indicate market or exchange nodes

Lines indicate links connection markets and agents

7. POSITIVE FEEDBACK AND NETWORK EXTERNALITIES

Networks entail economies of scope, joint and common costs, spill-overs, externalities, and cross-subsidies. Though these processes are difficult to disentangle link by link or route by route, they should be recognized. Networks (including both the economic network and its physical network components) are complex systems. There are multiple demand curves corresponding to multiple customer classes with differing price elasticities. As described in the previous section, producers create multiple outputs (different services) for different users.

Positive feedback is one term encapsulating the idea that more begets more. It is in contrast to the idea of negative feedback: more begets less. Several sources of positive feedback exist in networks. The first is in the law of the network: a network becomes more valuable the more members (users, destinations, etc.) it has. Second, standards and compatibility are another application of the idea of networks. A third is the process of cumulative causation and historical path-dependence - the longer a particular technological path is followed, the harder it is to switch as more and more new technologies, business decisions, etc. have been made with a certain environment as the default assumption. Finally endogenous growth suggests that new opportunites and market niches are created as the network expands, thereby expanding the network and creating new niches. The ideas of positive feedback in the economy and its sources: network externalities, learning curves, path dependence and cumulative causation, and self-fulfilling prophesies are discussed in this section. The success of network deployment depends on these conditions.

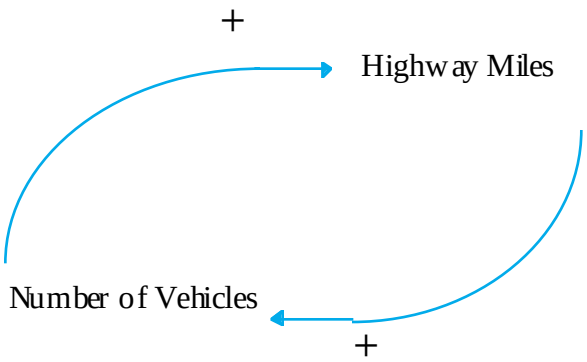
7.1 Positive Feedback, S-Curves, and Co-evolution

Positive feedback properties, from whatever source, make it hard to unseat one network with a new one in essentially the same niche. The net present value of the new network, including its complete construction costs, must be greater than the existing network for users to switch. With a paucity of members, a disadvantage all new networks face, that may be difficult to accomplish. But the expansion of the network will become easier the more users it has. This property applies to many kinds of networks, not just

transportation, from communications: telephony, fax, and email come readily to mind as examples.

In the case of the vehicle-highway system, the number of vehicles and miles of highway act as a positive feedback loop (Figure 7-1), particularly in the early years. The increase in the number of vehicles and increased ease of travel increased demand for highways - the more highways, the better the market for cars; the more cars, the larger the market for highways, which would be built better, spaced closer together.

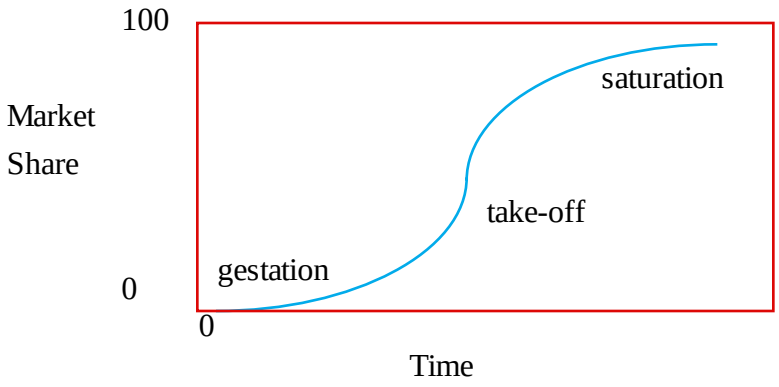
Figure 7-1: Positive Feedback Between Number of Vehicles and Size of Network



However, positive feedback curves are not generally inexhaustable, dimishing marginal returns tend to set in after a point. S-Curves, (as shown in Figure 7-2) can be used to describe how a given technology is deployed over time, showing the gestation period, take-off, and saturation of a technology. The S-Curve shows the cumulative amount of a technology as a share of its total potential market penetration, and can be viewed as the cumulative version of a normal distribution. The theory underlying the S-curve is straight-forward, and can be seen as an application of network externalities (described below). As knowledge of a technology and realization of its benefits spreads, the rate of adoption increases. Each project acts as a demonstration to potential new users. Furthermore, the advantages to adoption may increase with the number of users if there are network or inter-firm scale, scope, or sequence economies. As the technology diffuses, those who expect to attain the most benefit adopt it first. After a point, diminishing marginal returns set in. It is expected that, after complete exposure, technology is adopted by those who gain the most, and then by those gain less and less from it, until it is fully deployed. The life of technology may be cut short by competing technologies (such as

canals and railroads in the case of turnpikes) or because a technological problem is discovered (as in the case of plank roads).

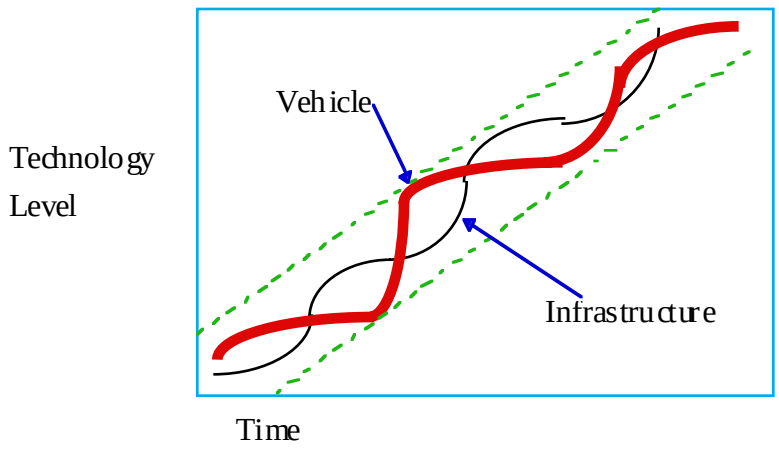
Figure 7-2: S-Curves, A Schematic



Over-time, the idea of co-evolution links interdependent S-curves, that is interdependent complementary technologies. Understanding this interdependence is critical to understanding the pitfalls of deploying a new technology or redeploying an old technology.

One underlying constraint behind technological advancement in complex systems is the requirement of "co-evolution" (Figure 7-3). Co-evolution is another example of the network externality phenomenon. Complex elements require the proper environment (network of related technologies) in which to work, and so cannot emerge in isolation. The environment here is defined broadly, to include the entire socio-technical system outside of the technology element in question.

Figure 7-3: Growth Path of Co-Evolving Vehicle and Infrastructure Systems: A Schematic



To illustrate, consider the automobile - highway system available today, and how it arrived. In the 1890's the first automobiles were being tested, but there were no hard, smooth surfaced roadways on which to drive them, cars had to be sufficiently durable to be able to drive on the dirt and macadam roads of the day. But even on those roads, they provided some benefits over the horse and wagon vehicles they replaced. Those benefits, whether mobility benefits, the pleasures of those who enjoy gadgets or simply conspicuous consumption, the desire to show one's economic status to the neighbors, is unimportant, they were sufficient to create a self-sustaining market for autos.

As auto deployment expanded, it became feasible to start constructing good roads, roads designed for motorized vehicles. Furthermore, the increasing number of auto users put pressure on government to provide better roads. These roads were more expensive than previous efforts, and would provide little or no advantage to animal powered vehicles, but were quite useful for cars, bicycles, and other wheeled mechanical vehicles.

The better roads enabled cars to further evolve. Rubber wheeled vehicles, problematic on dirt roads, worked well on hard roads. Further advances such as radial tires were made necessary by limits to the quality of surfaces (hard roads still had pot-holes), but could not emerged without advances in vulcanization, steel, and tires made over the previous century, and would not have emerged without the presence of a large market for vehicles riding on highways.

A vehicle designed to the technological level of the upper right of the co-evolution figure may not even operate on a roadbed at the lower left, and would not have emerged had roadbeds remained at the lower left. The "reverse salient" terminology of Hughes (1983) may prove useful. There is always one technology at any given time which is constraining progress in any set of related technologies, once this bottleneck technology (a reverse salient) is improved, progress can be made until the next bottleneck technology is reached. A vehicle can only progress so long as the infrastructure to support it progresses as well.

This basic concept can be applied at multiple scales, where the vehicle is any component in question of a system, and the infrastructure represents the rest of the system. These two elements interact to shape the complete system.

In economic terms, the environment needed for a technology to be viable can be considered as a fixed cost of that technology. Usually it is a hidden fixed cost. Moving from the lower left to the upper right in one step requires a large increase in fixed costs, while gradual changes only entail smaller marginal increases. Furthermore, the fixed costs in a multiple technology co-evolving system are borne externally at least in part. The cost of employing one advanced technology depends on the presence of earlier technologies. The cost of those earlier technologies is not directly present in the advanced technology, yet if those earlier technologies had not existed, moving directly to the advanced technology would require inventing them.

7.2 Law of the Network

The law of the network asserts that a network becomes more valuable the more users (destinations) that it has. This is referred to as a network externality. "Thus, the demand slopes downward but shifts upward with increases in the number of units expected to be sold. ... The key reason for the appearance of network externalities is the complementarity between the components of a network." (Economides 1996) An example is the telephone system, a telephone which was hard-wired between two users (a, b) has some value (V) (say 2: ab, ba) if those users are in frequent contact, but add a third user to the network (c), the value increases to six: (ab, ba, ac, ca, bc, cb), add a fourth user (d) and now 12 different connections can be made . The equation more precisely is:

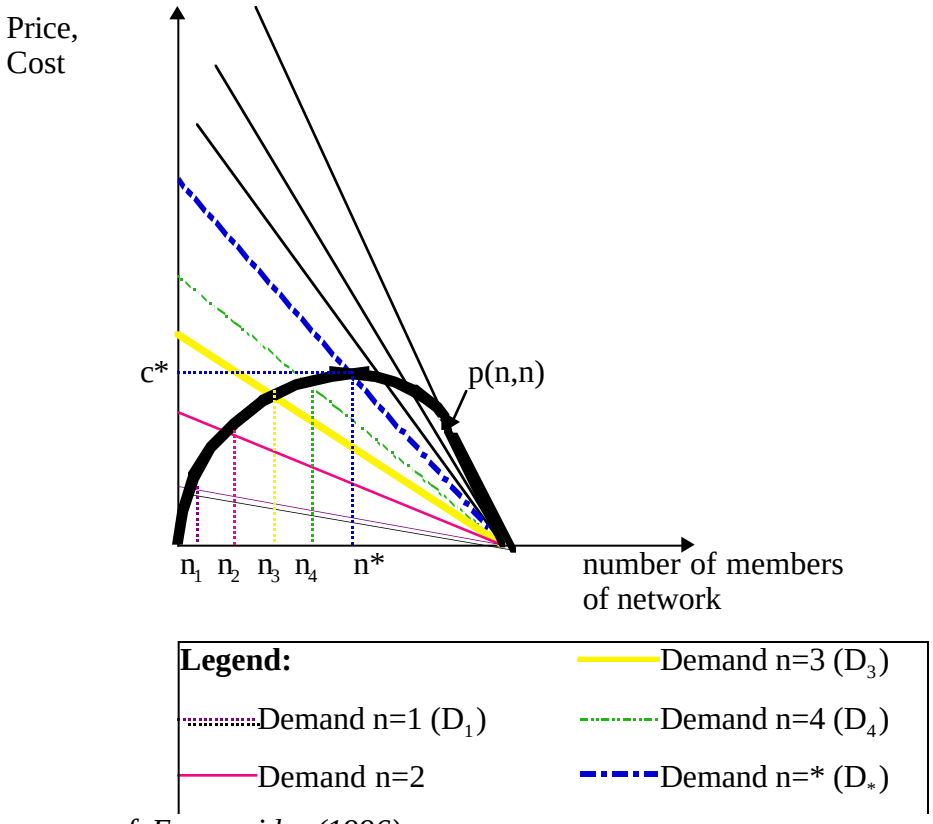
$$V \propto N(N-1)$$

where: N is the number of users

Figure 7-4 constructs the revealed demand curves for positive network externalities in a similar fashion to that undertaken in section 3 when the issues was congestion and road pricing. Let $P(n; n_e)$ be the willingness to pay for the nth unit of the good when n_e units are expected to be sold (assume each consumer purchases only one unit of the good). The network is more valuable the more units are sold. With only one consumer, (n=1), the network is not particularly valuable, so the implicit demand at n=1 (D_1) is low, lower than at D_2, which is lower than D_3, etc. Drawing a line between the number of consumers (n) and the implicit demand curve at that number (D_n) traces out an approximately parabolic shape, P(n, n). P(n, n) is the equilibrium price where the demand curve for a network of size n (D_e) intersects the vertical projection of the network size when the number of

consumers (network size) is e. P(n, n) is thus the fulfilled expectations (or revealed demand) curve, the set of prices that the nth consumer would actually pay to join the network which would sustain n-consumers. Economides (1996) argues that the fulfilled expectations demand is increasing for small n if any one of three conditions hold: "(i) the utility of every consumer in a network of zero size is zero, or (ii) there are immediate and large external benefits to network expansion for very small networks, or (iii) there is a significant density of high-willingness-to-pay consumers who are just indifferent on joining a network of approximately zero size." While demand rises with the number of members, thereby exhibiting positive critical mass under perfect competition, there is a saturation point, such that increasing the number of members does not add value. Such a system exhibits multiple equilibria (the largest of which is stable), and under perfect competition, the amount of network may be undersupplied because the positive externalities cannot be internalized to the producing firms.

Figure 7-4: Construction of Revealed Demand (Fulfilled Expectations) Curve with Positive Network Externalities



ref. Economides (1996)

A network externality on the supply side concerns the number of suppliers (which is a function of the number of consumers). The more providers there are, the "thicker" the market is, and the less dependent a consumer is on any one provider, the network becomes more valuable. This has advantages by making the network more robust, more secure from failure. Just as four one-lane bridges separated in space are less likely to catastrophically fail all at once than one four-lane bridge, the same issues apply to networks. Four suppliers are less likely to go bankrupt at once than one supplier. The competition from multiple suppliers affects not only system reliability but also price, competitors generally drive down costs (in the absence of major scale economies) and eliminate monopoly pricing, in the absence of cartels.

7.3 Historical Path Dependence and Cumulative Causation

Path dependence and cumulative causation are two other terms which are related to this phenomenon of increasing returns and network externalities, applying the ideas over time. Over time, being big creates advantages which makes you bigger. It often makes more sense to connect a new link to an existing network than to leave it in isolation in hopes that a new network will form around it. Of course this is not always true, new networks do form, but usually of dissimilar or greatly superior goods. When roads were first formed, they connected with towns and ports, and thus the existing transportation system. This process is path dependent, what happens now depends on what happened in the past.

In the case study "Road Pricing in Practice" the process of cumulative causation was described in road formation, where an existing path attracted new users who helped keep the path clear of brush, thereby attracting more. This process is present throughout networks as they become denser and denser. However, at some point, the net benefits of each additional improvement become proportionately (and perhaps absolutely) smaller with each iteration of the cycle. More current examples include the downward spiral facing transit systems. Poor conditions on transit drive away users, forcing service cutbacks, making the service worse, driving away more users.

7.4 Standards and Compatibility

A similar property happens with standards and compatibility. A classic example is the competition between VHS and Betamax to become the standard in video cassettes which resulted in the emergence of a VHS-only consumer market (Arthur 1990). The hypothesis for this rests on the idea of path dependence, because VHS had a slight edge in the number of users at a critical point, it thus had more films released and stores renting them, which made VHS models more valuable (and cheaper because of the spread of fixed costs), which further entrenched the technology. The reason for its slight edge at the critical point is unknown and probably as much due to chance as to better product or marketing. The adoption of a technology leads to an historic path dependency where future technology must be compatible with or far superior to previous technology, even if the old technology is suboptimal in some or many respects. Other examples include typewriter keyboards, gas tank and pump sizes, screws, and much computer and communications technology.

Application to transportation can be seen as well with railroads adopting standard gauges and interface protocols so that cars from one railroad could be shipped on the tracks of another. A rail line from a port to the mountains may have some value (as did the initial railroads on the east coast from Charleston, Baltimore, and Boston to the west), but if that same line runs from the port to everywhere else with a rail connection on the network, it is much more valuable, it will carry more traffic and generate more traffic, thereby providing incentives to expand the network. Once all points are efficiently connected, network additions may somewhat reduce the time or cost of travel, but will not add new places to the network, and will thus be less valuable.

7.5 Hubbing

While the law of the network looks simply at how many points are on the network, when we introduce space and time to the analysis, we need to make the analysis more sophisticated. The value of the connection between two points depends on the costs to go between them, the lower the cost the higher the value. Hubbing or any other means for aggregating small flows into larger ones which can be served more efficiently is one mechanism for lowering the internal costs and thereby making the network more valuble. The costs include both travel time as well as waiting time. Hubbing in transit services

groups passengers from multiple origins or destinations onto a single flight. Hubbing allows users to save schedule delay on a trip by increasing the frequency of service between destinations and allowing relatively easy connections. Similarly, the road hierarchy groups traffic from multiple origins and destinations onto common links which are built to higher design speeds. Higher demand reduces the spacing between major highways, thereby decreasing backtrack costs.

A key factor is thus that service is both spatially and temporally differentiated in nature, and that users contribute their own resources which depend on the location and schedule of transportation service. If the number of routes are cut back, travelers must travel farther (to a less convenient hub while in the air travel system, to a less convenient airport to access the air travel system), increasing their cost, if the frequency of flights is reduced, travelers must wait longer or travel earlier than they want to.

Network synergies may create monopolies to everyone's advantage. For instance, the concentration of flights at a single airport in a region enables a great deal of connecting traffic and thus improves local service. Splitting that traffic to multiple sites reduces network connectivity. On the other hand, concentration has costs including the additional congestion and travel time compared with dispersion.

Airlines have tried to further tie users to their systems over time by creating a long-term temporal network, their frequent flyer plans. These plans have even been extended across sectors, now being coupled with long distance and credit card use.

7.6 Knowledge, Belief and Prophecy

Network externalities are long term phenomena, and occur both in reality as described above, and in the perception of users. As knowledge about the network increases, it becomes a fixed cost among network users, which a new network has to overcome. Knowledge reduces the information and transactions costs in using a network. For instance, if I know who and where my suppliers are, I don't need to research them before purchase. Knowledge creates confidence in a certain level of performance, which an unknown quantify, a new network, does not possess.

Decisions have to be made based on some assumptions about the future. A form of knowledge, belief about the future, comes from many sources, and many implicit or

explicit predictions. Predictions which are believed change actions, potentially leading to the fulfillment of those predictions - the idea of self-fulfilling prophecy. Associated with self-fulfilling prophecy is self-negating prophecy, where a prediction about a (presumably bad) outcome leads to it being avoided. So belief in the success of a new network can lead to decisions which make its success more likely. Similarly, disbelief can kill it.

7.7 Endogenous Growth

Endogenous growth suggests that the size of the system itself cannot be analyzed in isolation of the system. While the clearest application of network economics is to a physical network, the network can be conceived as something broader, for instance, including land development in the transportation network. New development increases demand on the transportation network both by generating new trips and attracting them from existing sites. It also makes the network more valuable, a factor which will help the network grow (both the transportation subsystem and the land use subsystem).

8. THE DEPLOYMENT OF ROAD PRICING

Physical networks require long lead times to construct, and last for a long time. However, they may be used in very different ways than intended. Road networks laid out over 2000 years ago, with some refurbishment, still function in formerly Roman Europe. New York's water system is over 150 years old. Any analysis of network deployment should consider their longevity.

Perhaps the ultimate application of road pricing technology will require advanced highway systems. In an era of smart cars and smart roads, one can conceive of quasi-competitive highways altering their toll dynamically in response to demand (and travel time) and drivers (or pre-programmed cars) selecting routes in response to the price structure and their individual value of time. However, in the interim, a path to deploying road pricing and electronic toll collection as it is currently technologically available is needed. In the introduction, three different road pricing deployment problems were identified: deployment of ETC on existing roads, constructing new toll roads, and converting existing roads to toll roads. These are discussed in turn.

8.1 Electronic Toll Collection on Existing Toll Roads

The first issue concerns the deployment of electronic toll collection on existing turnpikes, toll roads, and toll bridges and tunnels. The conversion from human toll operators and automatic coin deposit boxes to electronic toll collection (ETC) systems is presently being undertaken. However, different systems use different technologies. While this may suffice for the vast majority of local trips, and may be a necessary interim step to winnow out technological winners and losers, over the long term some standardization is necessary. With the provision of compatibility between regions, users can use more than one toll facility under different management while only having one electronic toll collection device in their vehicle. As many individuals use different toll facilities from time to time, the presence of an electronic toll collection unit in the vehicle will become common, at least in certain parts of the country. Thus, many of the transaction costs (delay due to stopping, labor costs, construction of toll booths) associated with implementing tolls on a new

facility can be minimized as new facilities can be tolled using the standard ETC system. Costs of ETC should decline as the fixed costs of development and initial deployment are spread over a wide number of users. Network externalities should be exploited. Special facilities which are already tolled, including tunnels, bridges, turnpikes, selected new highways, and high occupancy/toll (HOT) lanes need to adopt a standard electronic toll collection mechanism.

8.2 New Toll Roads

The second issue was the construction of new toll roads. Since the completion of the interstate highway system, new road construction has been relatively sparse, in some sense, applying the idea of the S-curve, the limited access highway network has reached saturation. Much of the new, albeit limited, freeway construction is being toll financed because of scarce resources. First, finding financing from a higher level of government for local projects is difficult in the absence of a national road-building program. Second, roads paid for by one jurisdiction serve both local and non-local residents - generating revenue from non-local residents requires a mechanism like tolling.

Furthermore, the dearth of new construction also implies that no new travel speed increases from highway travel will come about, as current roads get saturated with traffic which continues to grow. Proposals to construct automated highway systems (AHS), which promise higher speeds and flows, will require financing to be implemented. As these systems will, at least initially, only support a subset of the vehicle fleet, it seems reasonable to suppose that a special financing mechanism (tolls) will be preferred to a more general mechanism (funding out of general revenue). If AHS becomes a dominant technology, it can be expected to bring tolling back as the primary revenue mechanism. The toll collection mechanism for these roads should be consistent with the technology used elsewhere, so that users need only support one in-vehicle toll communication device.

8.3 Tolling Existing Unpriced Roads

The third and hardest case is the tolling (or re-tolling) of presently unpriced roads, presumably utilizing electronic toll collection. The idea of local welfare maximization with

decentralized decision making should be employed. The following is a scenario of how that might happen. Central cities need to establish cordon rings in lieu of or in addition to other financing mechanisms so long as the dollars collected remain within the transportation sector. This is akin to a commuter tax which several cities already assess on individuals working but not living within their geography. The idea of cordon rings is not as unlikely as it seems. Already several cities (Singapore, Oslo, Bergen, and Trondheim) have imposed explicit cordon rings. Other cities, like Manhattan in New York have implicit cordon rings, one cannot enter that island without paying tolls from most directions. Similarly San Francisco has a partial cordon ring from the north and east, though the revenue collected remains with the bridge authorities rather than being used for city streets. Cordons can be established at convenient locations, most often natural barriers like bodies of water, but also artificial barriers like beltways.

In response, suburbs would likely establish cordons to toll city residents at convenient boundaries. It would be perceived as unfair that suburban residents pay tolls to enter the city, but city residents can drive on suburban roads without a similar price tag. Whether the suburban cordons would require separate facilities than the city cordon, or simply share the revenue from those crossing the cordon is a secondary issue to its presence.

Once they are initially constructed, cordon rings can be made more efficient when they get drawn tighter and tighter over time. The smaller the area enclosed within a cordon, the more direct the pricing of the network, and the closer use and revenue coincide. At a very tight cordon, this method approaches link specific tolls, particularly on excludable facilities. The traditional downside of "perfect" tolls on excludable facilities is that spacing between exits is increased, so backtracking and slowtracking costs are increased as well, however, with low transaction costs associated with ETC, this problem need not arrive.

It should be emphasized that these tolls are primarily a substitute for existing road financing systems (gas tax and general revenue), a substitute which is more efficient because it directly collects revenue from users on a specific facility and thereby can be used to provide incentives to reduce the welfare loss associated with excess use (where marginal cost exceeds marginal benefit). Of course, the burden associated with tolling will shift, but

if decision-making is sufficiently decentralized, the shift will be politically palatable because it shifts from local to non-local residents.

Pricing should initially be applied to the appropriate level of the hierarchy of roads. Limited access links dedicated to movement are the first candidate for pricing. These roads have a cost structure where users face increasing costs with additional users. Cordons around subareas, networks used mainly for access and short movement, are the second candidate for road pricing - although this would be mainly to recover fixed costs and some maintenance costs rather than to increase the size of the local network or reduce congestion. Local streets are more likely to be operating on the left side of the U-shaped cost curve, the area of declining average costs.

8.4 Final Words

The economic goal of congestion pricing is achievable after the implementation of road pricing in general. Over time, direct road pricing can be structured to provide off-peak discounts, and can thus be converted to time-of-day pricing, which is more efficient than "one size fits all" pricing. Congestion pricing requires the peak users to pay for the additional capacity that the peak requires, while not requiring off-peak users to pay for the excess capacity they don't need. However it is clear that the acceptance of toll roads is required before time-of-day differentiation, much less dynamic pricing, can be deployed.

Finally, differences in demand for different levels of service need to be recognized. Some users would pay more to have a better level of service. This can be exploited to raise revenue for new infrastructure. Some new private roads (such as SR91 in Southern California) charge a premium for high level of service travel in parallel with existing congested roads. However, when devoting capacity to a high level of service increases congestion on the rest of the links (e.g. queue jumping at a bottleneck), the social equity impacts must be carefully considered.

9. REFERENCES

Alleman, James and Emerson, Richard ed. (1989) Perspectives on the Telephone Industry Harper & Row, NY

Arthur, Brian (1990) "Positive Feedbacks in the Economy". Scientific American, Feb. 1990, pp. 92-99

Economides, Nicholas (1996) "The Economics of Networks". Journal of Industrial Organization

Hughes, Thomas (1983) Networks of Power, Johns Hopkins University Press, Baltimore

Kahn, Edward (1988) Electric Utility Planning and Regulation American Council for an Energy-Efficient Economy with Universitywide Energy Research Group - University of California

Krugman, Paul (1996) The Self-Organizing Economy Blackwell Publishers, Cambridge MA

Levinson, David (1997) Case Study: Road Pricing in Practice. California PATH Research Report. UCB-ITS-PRR-97-38. Berkeley CA

Levinson, David (1998) On Whom The Toll Falls: A Theory of Network Financing, Ph.D. Dissertation, University of California at Berkeley