

EVALUATING TRANSIT SUBSIDIES IN CHICAGO

by

IAN SAVAGE and AUGUST SCHUPP

Department of Economics
Northwestern University

Correspondence Address

Professor Ian Savage
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, Illinois 60208

Ph: (847) 491-8241

Fax: (847) 491-7001

e-mail: ipsavage@northwestern.edu

November 1996

This is a working paper version of a paper published in the Winter 1997 *Journal of Public Transportation* (volume 1(2), pages 93-117. We would like to thank Ken Dallmeyer of the Chicago Transit Authority, Mark Fialkowski of the Chicago Area Transportation Study, Duana Love and Richard Starr of the Illinois Department of Transportation, Cynthia Nethercut of the Regional Transportation Authority, and Robert Seyfried of Northwestern University's Traffic Institute for providing data and professional insight. However, the analytical use of that data and the conclusions drawn are solely those of the authors. This work was funded by the United States Department of Transportation's University Transportation Centers Program at the Great Lakes Center for Truck and Transit Research, University of Michigan.

ABSTRACT

This paper presents a model that calculates the social welfare benefits of using additional subsidy to reduce fare levels or improve service levels of public transit in Chicago. The model differentiates between the effects in peak and off-peak periods for both bus and rapid rail service.

The potential welfare benefits of transit improvements accrue to several different groups of people: (1) existing transit riders; (2) automobile users who decide to switch to public transit; (3) road users who benefit from reduced road congestion; and (4) people who undertake new trips. These benefits are compared with the welfare cost of the additional sales taxes that are used to fund transit subsidies in Chicago. Results of the analysis are that:

- Bus fares should be reduced, especially during off-peak and weekend periods;
- Rail fares are "acceptable" in that the marginal benefit of using subsidy to reduce fares is close to the excess burden of raising the subsidy;
- Bus service levels are broadly acceptable, except for the peak period where they are too high; and
- Rail service levels are too high at all times of the week, but especially in the peaks and on Sundays.

In general, it is more advantageous to use subsidy monies to reduce fares than improve service levels. Even if overall subsidy levels were not increased, society would be better off if service levels were reduced, and the money saved channeled into reductions in fares.

Keywords: transit - Chicago - subsidies - fares - service levels

1. INTRODUCTION

The Chicago Transit Authority (CTA) provides comprehensive bus and rapid rail service in the City of Chicago with a peak vehicle requirement of 1,700 motor buses and 800 railcars. Currently all of the capital costs and 50% of the operating costs are funded by public subsidies. Operating subsidies are primarily raised from a local sales tax levy. Federal operating support is less than 5% of costs. This paper investigates CTA operations in 1994 to see whether this level of subsidy can be justified.

2. ECONOMIC THEORY

Passengers face a generalized cost of riding transit which is composed of the fare that they pay (P) and the time costs associated with waiting at stops, and riding on the vehicle. Each user faces a generalized cost (GT) of:

$$GT = P + w(M) + t(M, Q_T) \quad (1)$$

where M is the number of vehicle miles operated, and Q_T the number of transit riders. The $w(M)$ function represents the monetary equivalent of the time taken waiting at stops. As the level of service increases average waiting times at stops should fall. The $t(M, Q_T)$ function is the monetary equivalent of the time taken on the vehicle ("in-transit time"). This will vary with the average number of people on the vehicle. As the vehicle gets more crowded it will have to stop more often and for longer periods for people to board and alight. This function will be increasing in Q_T but decreasing in M . The transit agency will face a demand function:

$$Q_T = d(GT) \quad (2)$$

The transit agency's costs (C) can be thought of as a combination of fixed costs (F) and the marginal cost (α) of providing each vehicle mile:

$$C = F + \alpha M \quad (3)$$

The transit agency can choose the levels of P and M , and the market will determine GT and Q_T . If the transit agency was a social welfare maximizing monopolist without any budget constraint, welfare would be maximized when:

$$\max W = \int_{gt(P,M)}^{\infty} d(GT) \partial GT + P Q_T - F - \alpha M \quad (4)$$

This gives the following first order conditions:

$$\frac{\partial W}{\partial P} = -d(GT) \frac{\partial GT}{\partial P} + Q_T + P \frac{\partial Q_T}{\partial P} = 0 \quad (5)$$

$$\frac{\partial W}{\partial M} = -d(GT) \frac{\partial GT}{\partial M} + P \frac{\partial Q_T}{\partial M} - \alpha = 0 \quad (6)$$

These equations can be simplified by removing the partial derivatives of Q and GT with respect to P and M . This can be done by differentiating equations (1) and (2) with respect to P and M and then solving using Cramer's Rule. This will result in the simplified first order conditions:

$$P = Q_T \frac{\partial t(\cdot)}{\partial Q_T} \quad (7)$$

$$\alpha = Q \frac{\frac{\partial w(\cdot)}{\partial M} + \frac{\partial t(\cdot)}{\partial M}}{\frac{\partial d(\cdot) \partial t(\cdot)}{\partial GT \partial Q_T} - 1} - P \frac{\partial d(\cdot)}{\partial GT} \frac{\frac{\partial w(\cdot)}{\partial M} + \frac{\partial t(\cdot)}{\partial M}}{\frac{\partial d(\cdot) \partial t(\cdot)}{\partial GT \partial Q_T} - 1} \quad (8)$$

The first condition is that fare should be set equal to the delay caused to all existing riders due to the boarding and alighting of the marginal rider. The second condition is that the operating cost of a marginal vehicle mile should be set equal to the benefit to riders of the marginal vehicle mile on their waiting and in-transit time, less the revenue gained from the new passengers attracted to the improved service. The latter condition also recognizes that while additional service will generally lower generalized cost, there is a countervailing effect in that the additional passenger trips generated may slow service because of their boarding and alighting time.

There are three implications that are important to this work. The first is that fares and service levels are both policy variables for the transit agency, so any additional subsidy can be used to either reduce fares or augment service levels. The second is that passengers impose externalities on each other. On one hand, additional passengers increase the travel times of existing riders because of additional time spent at stops. On the other hand, any exogenous increases in demand will result in the provision of additional service, and this will generate more ridership because generalized cost will fall as waiting times and the number of people on each vehicle falls. This latter economy of scale in the number of passengers on user costs is commonly referred to as the "Mohring (1972) effect." The third implication is that the first order conditions define an "optimal" mix of fares and service levels, which will be referred to in this paper as the point at which fares and service levels are "balanced."

Driving is a substitute for transit. The demand for driving (Q_A), measured in vehicle miles, will be determined by the attractiveness of public transportation (P and M) and positively related to the average speed that traffic moves on the roads (S). If congestion leads to the lower average speeds then roads become less attractive to potential users:

$$Q_A = a(P, M, S) \quad (9)$$

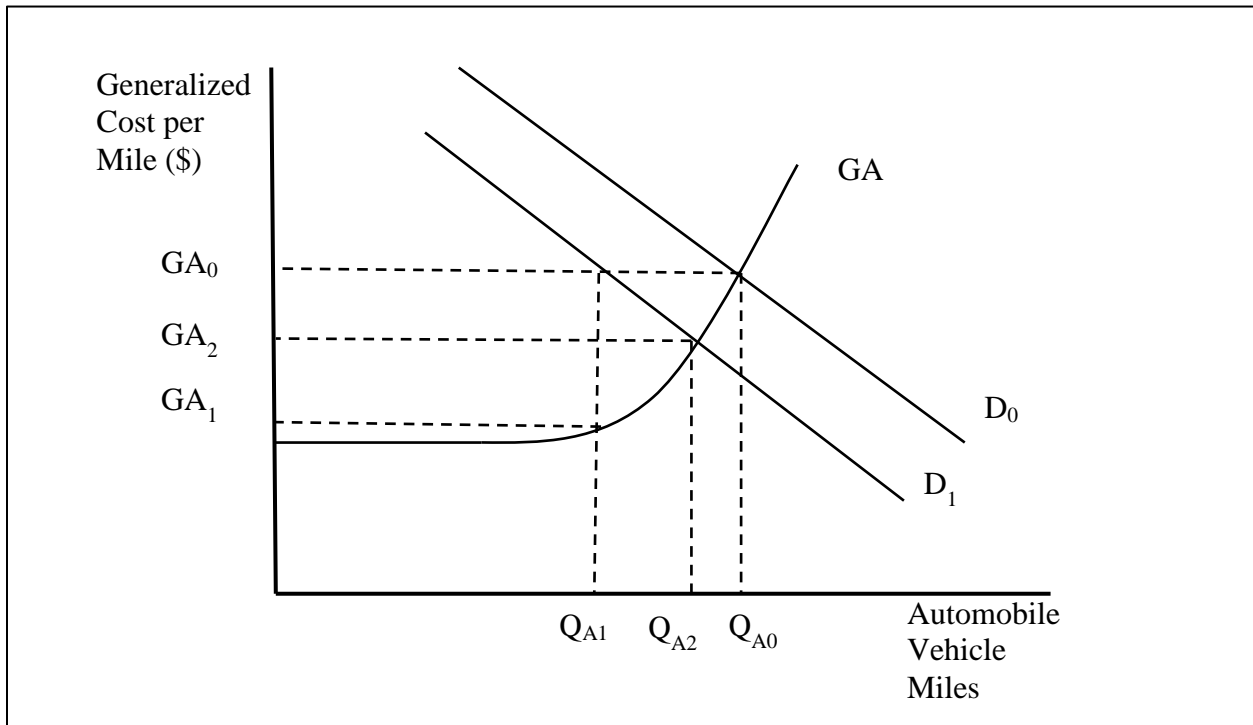
The cost of road travel to each user (GA) comprises the taxation payment to support the cost of building and maintaining the roads (R), the private operating costs of vehicles (O), and user time costs which will be an inverse function of S , which is itself a function of the total number of people wishing to use the highway:

$$GA = R + O + \Phi S(Q_A) \quad (10)$$

For welfare maximization, the first order condition would require that a "congestion toll" be charged equal to the time penalty imposed by the marginal user on all of the existing users who now have to travel slower because the road is more congested. However, congestion pricing of roads is not used in Chicago, and is not likely to be in the foreseeable future. A second best alternative is to use subsidies to make transit more attractive and thereby encourage some road users to switch modes. This will reduce road congestion, and improve travel times for those people who continue to use the roads.

Figure 1 shows the market for road travel. The average user cost curve (GA) is flat at low levels of demand because additional vehicles do not impede existing traffic. However, beyond a certain point, congestion develops and travel times start to increase. If transit was priced commercially there would be a demand curve D_0 for road travel, and in equilibrium Q_{A0} miles are driven. If transit is subsidized, the demand curve moves inward to D_1 , and Q_{A2} vehicle miles are driven. The Q_{A1} users who continue to drive each gain $(GA_0 - GA_2)$ because the road is less congested. In addition $Q_{A1} - Q_{A2}$ miles are generated because the roads are more attractive. The *additional* surplus that is generated in the road sector from transit subsidies would justify lower fares and greater service levels than would be suggested from the model in equations (4)-(8). Glaister and Lewis (1978) estimated such a model for London.

Figure 1: Market for Road Travel



The provision of subsidy funds is not costless. In the case of Chicago, additional transit subsidies are provided by a sales tax levy. Raising tax rates will increase the cost of goods and services and produce a deadweight loss θ for every dollar of subsidy monies raised. According to Jorgenson and Yun (1991), the marginal excess cost per dollar of tax revenue generated from

a sales tax is 26.2 cents. This "shadow value (or excess burden) of public funds" should not be confused with the costs of solely administering raising tax dollars.

Optimal transit subsidies, fares and service levels would therefore be based on a modified version of equation (4) which includes both the benefits of reduced road congestion and the excess burden of transit subsidies, which would be, in its most simplistic version, given by the term:

$$- \theta (F + \alpha M - PQ_T)$$

Dodgson and Topham (1987a) provide a more sophisticated version of this model allowing for income distribution issues to be considered.

3. PREVIOUS LITERATURE

In Britain in the early 1980s there was considerable concern by the national government that socialist-leaning local authorities in the major cities were providing "too much" subsidy to the local publicly owned bus companies. The British Transport Act of 1983 required local authorities to measure the benefits of subsidy as part of their planning process. The national Department of Transport sponsored the development of a computer model, the Method for Evaluating Transport Subsidies (METS) that is reported in Glaister (1987). This model contained a set of simultaneous equations expressing the demand for public transportation and road transportation, the congestion conditions of the roads, and the user costs for both transit and driving. Later, another British transport economist John Dodgson (1987) estimated a more stripped down version of the model for the major cities in Australia.

The results of Glaister's and Dodgson's work are reported in table 1. The second and third columns report the benefit-cost ratios of a marginal pound or dollar of subsidy spent on reducing fares or improving service levels. The cost is simply the nominal amount of subsidy and does not include any shadow value of raising the public funds. In nearly all cases the benefits of subsidizing fares exceed those of subsidizing expanded service levels. This indicates that none of the systems are near the "balance" described in the previous section. Social welfare could be improved considerably, even if overall subsidy levels are kept constant, by reducing service levels and using the money saved to reduce fares. Glaister's British work calculates the changes in fare and service levels necessary to bring them into balance. These are shown in the fourth and fifth columns. In general service levels should be reduced by 15% and the savings used to reduce fares by 20%. When fares and service levels have been balanced, one can investigate whether additional subsidies are justified. These benefit-cost ratios are shown in the final column and should be compared with the shadow value of public funds. Dodgson and Topham (1987b) report a shadow value of about 11% for the property tax used to fund transit in Britain. Therefore additional subsidies could be justified in all cities except for Sheffield, a city where the price of transit had been held at very low levels.

Table 1: Results of Previous Literature in the UK and Australia

City (Bus systems unless otherwise indicated)	Benefit-Cost Ratio For		Changes to Balance Fares and Service Levels		Benefit- Cost Ratio at Balance Point
	Reduced Fares	Increased Service Levels	Fares	Service Levels	
Birmingham	1.21	1.41	+5%	+4%	1.24
Leeds	1.29	0.81	-24%	-13%	1.18
Manchester	1.33	0.71	-23%	-17%	1.19
Liverpool	1.31	1.15	-6%	-3%	1.26
Sheffield	1.03	1.03	0	0	1.03
London - bus	2.12	0.37	-28%	-31%	1.28
London -subway	1.26	1.79	-11%	+19%	
Sydney	1.43	0.37			
Newcastle		0.46			
Melbourne		0.32			
Brisbane		0.53			
Adelaide		0.43			
Perth		0.47			
Hobart		0.48			
Canberra		0.48			
Sydney - Rail		1.25			
Melbourne - rail	1.25	0.27			
Brisbane - rail	1.33	0.53			
Adelaide -rail	1.55	0.34			
Perth - rail	1.55	0.26			

A more stunning result is that cost-benefit ratios are less than unity for improving service levels in Leeds, Manchester, London buses, and all services in Australia. This suggests that unambiguously *too much* service is provided in these cities. There would seem to be a natural tendency for transit agencies to maintain a level of service far in excess of that justified, and as a result charge higher fares to remain within their budgets.

4. INNOVATIONS IN THIS WORK

In general this work follows that of Dodgson in that a comprehensive interactive demand model is not used. However, Dodgson calculates welfare at the margin, whereas we follow Glaister in calculating infra-marginal welfare changes. Our model therefore permits calculation of the balance point of fares and service levels.

The innovation in our work is the introduction of different time periods within the week. Both Dodgson and Glaister simply calculate an overall daily figure. In this work the week is divided into four time periods: weekday peaks (6am-9am, 3pm-6pm), weekday off-peak, Saturdays, and Sundays. This should permit identification of where the imbalance of fares and frequencies is occurring.

5. DETAILS OF THE MODEL

5.1 The Transit Demand Model

The generalized cost (GT) for each rider is:

$$GT = P + v_t T + v_w W \quad (11)$$

where P is the fare paid, T is in-transit time, W is the time taken waiting at the stop, and v_t and v_w are the values of time for in-transit and waiting respectively. Demand is taken as a linear function of GT . The relationship between demand and GT will be expressed as the elasticity of demand with respect to generalized cost of travel (ε_{GT}). The literature does not generally report empirical values of ε_{GT} , but there is abundant literature on price elasticities of demand (ε_P). These two elasticities are related to each other in that:

$$\frac{\varepsilon_{GT}}{\varepsilon_P} = \frac{GT}{P} \quad (12)$$

A study was conducted for the CTA in the late 1980s which produced some very specific price elasticities for time of day and mode (LTI Consultants, 1988). These are shown in the numbers on the edges of table 2. Knowledge of the number of riders in each time and mode category allowed us to infer the elasticities in the middle of the table. One will note that the demand for rapid transit rail service is very inelastic even in off-peak periods. In contrast bus service is less-inelastic especially in the off-peak. Much of the difference is explained by the longer journey lengths on the rail service compared with bus service (6 miles compared with 2), and the radial nature of the rail system which is oriented to trips to downtown, whereas the bus

service serves the neighborhoods. Therefore both walking and the automobile are easier substitutes for bus trips than rail trips.

Table 2: Price Elasticities

	Bus	Rail	Total
Peak	-0.25	-0.12	-0.19
Off-Peak	-0.50	-0.16	-0.44
Total	-0.40	-0.14	-0.34

Information is available from the annual "Section 15" report submitted to the Federal Transit Administration (CTA, 1995) on the average fare paid on the bus and rail modes. This is calculated by dividing total revenue by total trips on the mode. However, average fare paid is likely to vary across different time periods because the type of passengers will vary. School children and the elderly pay reduced fares, and commuters with monthly passes and prepurchased tokens have a discounted fare. Fortunately the electronic fareboxes used on the buses store such information, and we were able to use a management report to be able to determine the average fare in each time period. We were not able to obtain similar information for the rail system as fare collection is not automated. We therefore assumed that the average fare paid on the rail system would vary about the weekly mean in the same ratio as on the bus system. However, inter-time-period differences in average fare are not very large, and are at most 5¢.

Information is available in the Section 15 reports to calculate in-transit time for each mode and time period. The average journey length is calculated by dividing passenger miles by passenger trips. This is then converted into minutes by dividing by the average speed of operation, found by dividing vehicle miles by vehicle hours.

The model of bus service is constructed in such a way to allow the in-transit time to vary according to average load factor. As fares and service levels are changed, the average number of people on each bus will vary, and hence in-transit time will vary because the bus will have to stop more frequently and/or for longer periods to allow the extra people to board or alight. The change in average travel time is the number of extra people on each bus multiplied by the average boarding and alighting time (*BAT*). For the configuration of vehicle and type of fare collection used by the CTA, this is 2½ seconds a person (Transportation Research Board (TRB), 1985, p 12:13). The change in in-transit time can therefore be expressed as:

$$\Delta T = \left[\frac{Q_{T1}}{M_1} - \frac{Q_{T0}}{M_0} \right] BAT \quad (13)$$

where the 0 subscript is the *status quo* and the 1 subscript represents the situation after a fare and/or service level change. One will immediately note that the Q_{TI} on the right hand side of this

equation is itself a function of ΔT . Therefore the equation will have to be rationalized to collect all of the terms in ΔT on the left hand side. In general this will produce:

$$\Delta T = \frac{\Delta M.GT_0 + \frac{\varepsilon_{GT}(v_w \Delta W + \Delta P).BAT.Q_{T_0}}{M_0 + \Delta M}}{GT_0 - \frac{\varepsilon_{GT}v_T Q_{T_0} BAT}{M_0 + \Delta M}} \quad (14)$$

This equation simplifies considerably when only fares change because ΔM and ΔW will be zero. The model for rail service does not include this effect because the number of stops is predetermined and station dwell time less sensitive to changes in load factor.

Waiting time is taken as a function of the headway (H) between bus or train arrivals at a stop. Prior research has indicated that for headways of up to twelve minutes, passenger arrivals at stops are random and W is one-half of H ; for longer headways transit users attempt to arrive at stops close to the time of departure, and W becomes less than one-half H . Our research uses the seminal relationship found by Seddon and Day (1974). This quadratic relationship relates headway and waiting time in seconds:

$$W = 11.39 + 0.49H - 0.00009H^2 \quad (15)$$

Information on average headways was obtained from a CTA management document that summarized the published schedules for each bus and rail route. An average for each time period for each mode was obtained by weighting the headway on each route by the vehicle hours operated on that route.

A standard approach was taken to valuing in-transit and waiting time (Bein et al., 1994). In-transit time is valued at half of the average wage rate, and waiting time at the average wage rate. Bureau of Labor Statistics (1996) figures were obtained on the average wage rate in Chicago in November 1994, which produced a value of in-transit time of about 10¢ a minute, and a value of waiting time of about 20¢ a minute. A summary of the data used for the transit demand analysis is shown in table 3.

Table 3: Base Data for the Transit Demand Model

Bus	Weekdays		Weekends	
	Peak	Off-Peak	Saturday	Sunday
Annual Passenger trips (million)	136.4	136.4	33.8	24.8
Annual Bus Miles (million)	23.7	33.8	7.9	7.1
Average Fare (\$)	0.73	0.75	0.79	0.79
Average Headway (min)	6.1	10.9	10.7	13.1
Average Trip length (miles)	2.4	2.3	2.3	2.2
Average Trip Time (min)	14.0	13.4	13.1	11.8
Average Load Factor (passengers/bus)	13.8	9.3	9.8	7.5

Rail	Weekdays		Weekends	
	Peak	Off-Peak	Saturday	Sunday
Annual Passenger trips (million)	74.6	52.0	9.8	7.6
Annual Train Miles (million)	2.6	4.8	1.1	1.2
Average Fare (\$)	0.74	0.75	0.80	0.80
Average Headway (min)	5.5	9.2	9.0	8.3
Average Trip length (miles)	6.4	5.5	5.5	6.0
Average Trip Time (min)	16.2	14.0	13.3	13.7
Average Load Factor (passengers/train)	181.9	59.5	47.6	37.7

5.2 The Transit Cost Model

The model assumes that the marginal cost for an additional passenger on a predetermined level of service is zero. Therefore, if subsidy is used to change fares there will be no change in the CTA's cost of operations to the CTA. However, if subsidy is used to change service levels there are cost implications of running additional vehicle miles, and changing the size of the fleet. Section 15 financial reports on operating expenses (CTA, 1995) and information on capital expenditures were used to separate costs into three types: (1) cost that vary with vehicle hours operated, (2) costs that vary with the number of vehicles owned, and (3) other costs which are

taken to be invariant with service. Table 4 indicates how the various Section 15 cost categories are classified in our model.

Table 4: Cost Allocation System

Section 15 Cost Categories	Vary with Vehicle Hours	Vary with Vehicles Owned	Invariant with Service
Operating Expenses	Operators' wages & fringe benefits, Fuel (buses), Electricity (rail), Liability		Administrative & Support Staff
Vehicle Maintenance	Tires, Lubricants, Oil	Labor, Materials	Utilities, Administration
Non-Vehicle Maintenance			All
General Administration			All
Capital Costs		Annualized Vehicle Purchase	All Other Capital Expenses

While Section 15 reports do include information on capital expenditures, it is not particularly useful to use data from only one year in that capital expenditures on buses and rail cars are lumpy expenses. An alternative approach was adopted of calculating an annual capital expense if the CTA continually replaced its bus fleet on a 12 year cycle and its railcars on a 35 year cycle with a refurbishment after 25 years. The Regional Transportation Authority provided figures on the purchase price of the most recent series of buses (\$218,000 each), railcars (\$855,000) and the midlife railcar refurbishment cost (\$400,000).

Based on these data, the variable cost of running an extra bus hour is \$47, an extra train hour \$210, and the annual cost of vehicle ownership is \$67,600 for a bus, and \$78,000 for a railcar. Our model assumes that if service levels are increased in the off-peak the CTA only bears the additional variable cost of running the extra bus or train hours. However, if service is increased in the peak, the CTA bears the cost of a proportionate increase in the number of vehicle owned in addition to the cost of additional bus or train hours. The peak to midday base vehicle requirement ratio for the CTA is 1.72 for buses, and 2.54 for railcars. Off-peak service can be expanded considerably without the need for additional vehicle ownership.

5.3 Transit Welfare Calculations

The demand curve for transit service is assumed to be linear, and therefore any benefits to newly generated transit trips can be evaluated in the usual way by the "rule of a half." Changes in transit consumer surplus (ΔCS_T) will therefore be:

$$\Delta CS_T = \frac{1}{2}[GT_0 - GT_1][Q_{T0} + Q_{T1}] \quad (16)$$

and the change in producer surplus will be the change in total revenue less change in total cost. This will, of course, be equal to the total amount of extra subsidy that will have to be granted.

5.4 Highway Model

Of the generated transit travel ($Q_{T1} - Q_{T0}$) some portion will be people who switch modes from the automobile, and some portion will be entirely new trips. Some people in the former category may have been auto passenger or users of van-pools and although they changed mode there will not be a reduction in automobile traffic.

Empirical evidence on the proportion of new-to-transit trips that were previously auto drivers or taxi users can be found from ridership surveys after new transit line construction in Chicago: the Blue Line Extension to the northwest side (CATS, 1986), and the opening of the Orange Line to the southwest side (LaBelle and Stuart, 1995). The proportion was 20.1% and 51.7% respectively. While both extensions serve airports, airport traffic is a relatively minor proportion of generated ridership. This analysis will assume that 50% of generated transit trips were mode shifters from the automobile. Given knowledge of average transit trip lengths for both bus and rail, it is possible to calculate the number of vehicle miles removed from the roads.

The 50% assumption implies a cross-elasticity in the peak period between vehicle miles and transit fares of 0.0011, and vehicle miles and transit waiting time of 0.009 based on auto travel on congested roads in the whole of Cook County which covers the City of Chicago and most of the inner suburbs. These elasticities are low compared with the fare cross-elasticity of 0.14 calculated by Webster and Bly (1980) and transit waiting time cross-elasticities in the range of 0.02-0.14 estimated by Peat, Marwick (1972). However, one should remember the base number of vehicle miles includes trips for which the CTA is not a substitute, and the miles of freight vehicles. Unlike in Europe, the CTA is a marginal player in urban transportation with a market share of only 21% of work trips made to destinations within the City of Chicago.

The basic highway model was shown in figure 1. If transit becomes more attractive and the road is congested, the movement inward of the demand curve from D_0 to D_1 will produce increases in welfare to other road users. However, if the demand curve intersects the average cost of travel curve on its flat portion, which is to say that the road is uncongested, then even if automobile users switch to transit there will not be a change in the average speed on the road, and there will therefore be no additional benefit to the remaining road users. Therefore, benefits will only accrue to the road sector where the roads are already congested. Our model will assume that there is no road congestion in the off-peak and on weekends. In addition, during the peak there will be some roads that operate under free-flow conditions.

The current demand for peak-period road travel in Cook County, measured in vehicle miles, is classified into nine categories depending on the type of highway and the level of congestion. These are shown in the table 5. The data represent demand on a typical weekday, and are derived from the Chicago Area Transportation Study's (CATS) regional demand model.

The boundaries between the three levels of congestions were based on CATS calculations of the ratio of traffic volume to theoretical road capacity on each road link. Deriving the information in the table was not an easy task, and is described in more detail in the appendix.

Table 5: Weekday Peak Vehicle Miles

Road Type	Congestion Level		
	Free-Flow	Moderate	Severe
Freeway	4,550,000	5,050,000	3,350,000
Arterial	7,500,000	10,600,000	4,950,000
Collector	450,000	312,000	90,000

The vehicle miles that are removed from the highway by the mode switch to transit are assumed to come from roads in the moderately and severely congested categories. This is because free-flow conditions generally exist in parts of suburban Cook County where the CTA does not provide service. The mileage removed from the roads is subtracted from the six categories of road / congestion level in proportion to current demand.

When subsidies are used to improve service levels, the number of bus miles operated increases and this will add to the number of vehicles on the road. These additional miles are assumed to occur on moderately and severely congested arterials roads, and in proportion to current road demand. The increase in vehicle miles on these roads is calculated based upon the passenger car equivalent (PCE) that a bus represents. In Chicago, buses stop in the roadway rather than bus bays, and according the Highway Capacity Manual (TRB, 1985, p 12:41) a bus is equivalent to 4.37 cars.

A model, similar to that shown in figure 1, is estimated for each of the six types of moderately and severely congestion roads, with daily peak-period vehicle miles on the horizontal axis, and generalized cost per mile on the vertical axis. The calculation of the original level of demand (Q_{A0}), and the amount of traffic that switches to transit, less any increase in the number of buses operated, ($Q_{A0}-Q_{A1}$) have been described in the previous paragraphs.

The next step is to derive an algebraic expression for the average user cost (GA) curve. While there are standard formulations used by traffic engineers, this model calibrates a curve based on actual conditions in the Chicago area. The calculations were made for two types of roads: freeways, and arterial roads. The relationship for collector roads was taken to be equivalent to that for the arterial roads. Average speeds for both expressways and arterial roads for the downtown area, the rest of the City of Chicago, and suburban Cook County are reported in CATS (1996). These speeds were taken to be equivalent to severe congestion, moderate congestion and free-flow respectively. The traffic volume to capacity ratios for severe ("level of service E") and moderate ("level of service C") congestion were obtained from the Highway Capacity Manual for 50 mph designed speed freeways and class III arterial roads.

This gives two data points on volume-to-capacity ratio and average speed for each of the road types. A linear estimate was made of the relationship between these two variables. This relationship allowed calculations to be made of the relationship between traffic volume and user cost. The ratio of Q_{A1} to Q_{A0} vehicle miles directly indicates the change in volume to capacity ratio, and hence to the change in speed. There is a direct relationship between changes in speed and the change in the time taken to drive one mile. The change in time is valued using the work of Bein et al. (1994) who find that people value their time at 65% of the average hourly wage rate on moderately congested roads, and 78% of the hourly wage rate on severely congested roads. Therefore, one can calculate $(GA_0 - GA_1)$, if one assumes that vehicle operating costs per mile do not change. The latter assumption is not unreasonable given that under the most extreme changes we look at, those necessary to balance fares and frequencies, average speeds on freeways change by about 3 mph, and those on arterial roads by less than 1/2 mph.

Knowledge of Q_{A0} , Q_{A1} and $(GA_0 - GA_1)$ permits calculation for the ultimate equilibrium (GA_2, Q_{A2}) providing the slope of the demand curve is known. Chan and Ou (1978) find that the elasticity of vehicle miles to travel time is -0.8. The ultimate change in consumer welfare for each category of road is:

$$\Delta CS_A = \frac{1}{2}[Q_{A1} + Q_{A2}][GA_0 - GA_2] \quad (17)$$

For the case where transit fares are reduced, each vehicle mile removed from the road produces a benefit to other road users of 22½¢. This varies from 8¢ a mile on a moderately congested freeway to 55¢ a mile on a severely congested arterial road.

Unfortunately for the case where service levels are improved, the addition of buses on already congested arterial roads makes road users *worse off*, not better off. This is because load factors on the buses are very low, and therefore the additional bus occupies more passenger-car units of road space than the number of riders it attracts from driving.

6. RESULTS

6.1 Changes in Current Fares and Service Levels

Table 6 shows the gross social welfare benefit per dollar of subsidy based upon either a 10% decrease in fare or a 10% increase in service levels for both rail and bus service during each of the four periods. The peak periods for both bus and rail include their respective contributions to the welfare changes on the roads. The numerator of the benefit-cost ratio only includes the changes in consumer surplus. Producer surplus change, which is equivalent to the subsidy requirement, is the denominator.

Table 6: Results of Benefit-Cost Analysis

Benefit per \$1 of subsidy	Weekdays		Weekends	
	Peak	Off-Peak	Saturday	Sunday
FARES DECREASED BY 10%				
Bus	1.39	1.77	1.77	1.80
Rail	1.26	1.18	1.18	1.18
SERVICE LEVELS INCREASED BY 10%				
Bus	0.21	1.11	1.24	1.16
Rail	0.34	0.97	0.77	0.54

The baseline comparison for each of these figures is 1.262, which is the cost of the dollar of subsidy plus the excess burden of raising that dollar. Therefore one may draw the following conclusions about appropriate uses of subsidy:

- Bus fares could be reduced, especially during off-peak and weekend periods;
- Rail fares are "acceptable" in that the marginal benefit of using subsidy to reduce fares is close to the excess burden of raising the subsidy;
- Bus service levels are broadly acceptable, given the tolerance of the model, except for the peak period where they are too high; and
- Rail service levels are too high at all times of the week, but especially in the peaks and on Sundays.

These conclusions are consistent with the results of Dodgson (1987), indicating that Chicago, and most likely American cities in general, are more like Australia with respect to public transit than they are like Britain.

Bus fares have a high return to subsidy because of the current CTA fares policy. The CTA charges a flat fare that is the same on both modes. This is despite the fact that bus journeys are both more elastic and have considerably shorter average journey lengths. In addition people wishing to transfer between buses have to purchase a transfer, while transfer between rapid transit lines is free. Clearly, there is considerable evidence to suggest that the CTA should charge differential fares between bus and rail, especially in the off peak.

A striking feature is the over-supply of capacity in the peak. At the margin the cost of providing peak service is very high. The peak period sets the standard for the number of vehicles required. Recent attempts by the CTA to stem its budget deficits have focused on trimming off-peak service, yet peak service is the area where service decreases can lead to major cost reductions. Some people may argue that it is impossible to reduce peak service without leaving people behind at stops. It is true that the CTA does operate at "crush loads" for short periods at certain parts of its system. It is likely that people will not be able to board the first bus or train that arrives if service is reduced. However, that situation does not occur on all parts of the CTA

system. Even if one assumes that there are no riders travelling in the reverse direction to the peak flow, which is clearly not true, there is currently an average of 28 people on each bus and 70 people per train car during the peak periods.

In all time periods for both bus and rail service, there is clearly a greater benefit to subsidizing lower fares as opposed to increasing service levels. The current mix of fares to service levels is not at the social welfare optimum described earlier in this paper. Even if subsidies are held constant, service levels should be cut across the board to finance lower fares. The causes for this imbalance can be found by looking at the history of the CTA in the ten year period 1984-1994. Bus ridership fell by 31% and rail ridership by 6% yet vehicle miles and train miles did not change. Fares, on the other hand, increased by 14% in real terms. It is clear that the CTA has tried to maintain output in the face of falling demand, and has increased fares to correct any resulting budget deficit. Service cuts provoke very vocal opposition from staff and specific groups of riders. The opposition to fare changes is a lot more defuse, and hence less politically effective. By ducking service cuts, the CTA has actually made the citizens of Chicago worse off rather than better off.

Even knowledgeable observers of the industry frequently argue that transit should maintain service levels, even at the expense of higher fares, because transit demand is more responsive to frequency than it is to fares. That is certainly true, but demand is still inelastic with respect to frequency, and therefore expanding service levels will lead to declining average load factors. This would not be a serious problem except that load factors in Chicago are already so low that the number of riders on the marginal vehicle is not sufficient to justify the cost of running it. In other words, while passengers are relatively sensitive to the level of service, the cost of providing any additional capacity is high.

6.2 "Balancing" of Fares and Service Levels

A decision on whether the current overall levels of subsidy to the CTA are justified can only be made after fares and service levels have been adjusted to efficiently use the existing subsidy. If one had the data to do so, one would proceed by balancing fares and service levels on each mode and in each time period independently. To do so one would need to know the cross-elasticity effects between periods and modes. If bus fares are reduced by more than rail fares, what mode shift would occur? If peak service is reduced, what would be the effect on off-peak ridership?

Unfortunately such information is not readily available. We therefore proceeded to look at the policy option of changing fares and service levels by the same percentage for both modes in all time periods, so as not to change the relative value of the generalized cost of transit travel. Service levels were reduced, and then fares were reduced so as to keep producer surplus, i.e., subsidy, constant. This was done until the combined consumer surplus from both the transit and road modes, was maximized. At this point, fares and service levels are balanced. The balance point required service levels to be reduced by 31%, and the savings allow fares to be reduced by 59%. The change in the mix of fares and service levels to the balance point produced a gain in surplus of \$76 million a year or about \$15 per person in Cook County. On the average, peak load factors on both bus and rail come nowhere near capacity. However, it is likely that some traffic

may be "choked of" at certain points in the system. Indeed, it may be socially optimal for the CTA to not serve some peak demand when the costs of peak operation are considered.

When fares and service levels were balanced, the benefit-cost ratio of increasing the overall level of subsidy was calculated. This was found to be \$1.16 per \$1 of subsidy. The return on transit subsidies is slightly less than the shadow value of sales taxes. The approximate nature of the model makes it difficult to conclude with certainty that subsidies are currently "too high." Certainly one could not argue that subsidies are grossly wasteful. If fares and service levels were balanced in each time period and mode individually, it is likely that additional transit subsidies would be justified particularly for bus service in the off-peak.

6.3 Sensitivity to Cost Reductions

In many parts in the world competitive contracting, or outright private competition, has been introduced into urban transit provision (Cox et al., 1995). A major objective has been to reduce unit cost levels that were believed to be too high under the existing monopoly public provision. The CTA is a publicly-owned monopoly. Typically cost reductions of 20% or more have been experienced from competitive contracting. There is currently some discussion of introducing a limited experiment of competitive contracting in Chicago. The model was re-estimated using unit cost reductions of 10%, 20% and 30%. Such reduction will only affect the returns to subsidizing level of service, as it is assumed that fare reductions are costless. The benefit-cost ratios for various levels of cost reduction are shown in table 7.

Table 7: Effect of Cost Reduction on Benefit per \$1 of Subsidy for Increased Service Levels

	BUS				RAIL			
	Peak	Weekday Off-Peak	Saturday	Sunday	Peak	Weekday Off-Peak	Saturday	Sunday
10%	0.23	1.32	1.47	1.37	0.38	1.09	0.87	0.60
20%	0.27	1.61	1.82	1.68	0.43	1.26	1.00	0.69
30%	0.31	2.07	2.39	2.17	0.49	1.48	1.17	0.80

It is clear, especially for off-peak bus service, that cost reduction would not only bring fares and service levels much more into balance but also make the case for additional subsidies much more clear cut.

7. POLICY IMPLICATIONS

There are three major policy recommendations for the CTA. The first is to thoroughly investigate the cost effectiveness of the provision of additional peak capacity. The second is to consider a discounted fare on the buses during off-peak hours. The third is to pursue policies to

achieve unit cost reduction. If the CTA can make strides towards these three goals then it should not only be able to justify current subsidy levels, but also make a case for increased subsidies.

APPENDIX: CALCULATION OF HIGHWAY PEAK AUTOMOBILE VEHICLE MILES

Data do not exist that show the amount of traffic during the peak periods in Cook County. CATS have a traffic demand model which can calculate vehicle miles on different links of the network for a 24-hour period on a summer weekday (CATS, 1996). CATS conducted a special run of their model to produce a matrix of vehicle miles for Cook County broken down into three categories of roads: expressways, arterials, and collectors. Within each of those three categories, the data are separated into free-flow, moderate, and severe levels of congestion, based on the volume-to-capacity ratio of the link. While the model was run for the year 2007, CATS scaling factors were used to produce mileage data for 1996.

The problem was to determine the proportion of daily vehicle miles travelled during the peak. This was relatively straightforward for the freeway system because the Illinois Department of Transportation (IDOT) counts hourly on all Cook County freeways using counting loops placed in the lanes. Data was obtained for six weekdays in March 1996 for 11 freeway locations, and the proportion of traffic travelling in the peaks estimated.

Table A1: Ratio of Peak to Daily Traffic

Road Type	Congestion Level		
	Free-Flow	Moderate	Severe
Freeway	43.7% (3)	40.7% (4)	35.5% (4)
Arterial	44.1% (31)	42.2% (54)	43.5% (29)
Collector	44.5% (50)	45.8% (21)	44.6% (15)

Hourly data for non-freeway roads in Cook County is limited to vehicle counts by IDOT at selected intersections. The intersections are selected primarily because IDOT is considering highway improvements at these locations. Counts are conducted over two mid-week days between 6am to 6pm. Data are available for the hourly counts, and an estimate of the 24-hour volume (average annualized daily traffic - AADT). Peak percentage of daily traffic was calculated by dividing the sum of vehicles counted during the defined peak periods by the AADT. We then assigned each intersection a road congestion designation of severe, moderate, or free-flow, according to its peak volume-to-capacity ratio. The capacity value is based upon the number of lanes, distribution of "green time" at an intersection, and parking and other traffic restrictions. A typical intersection has a capacity of 700 vehicles per lane per hour. The Highway Capacity Manual (TRB, 1985) considers Class III Urban Streets, the type found in cities, to be severely congested when the ratio is 1.00 or higher and moderately congested when the ratio

is between 0.63 and 1.00. Based on our classification of freeway locations and intersections by congestion level, we were able to determine the percentage of 24-hour traffic that occurs during the peak. This is shown in the table A1 along with the number of sites in our sample in parentheses.

REFERENCES

Bein, Peter, Ted Miller, and William G. Waters II. (1994). "British Columbia road user unit costs." *Proceedings of the Canadian Transportation Research Forum*, 29th annual meeting, Victoria, British Columbia, pp 714-727.

Bureau of Labor Statistics. (1996). *Employment and Earnings*, vol 43(1). Washington, D.C.: U.S. Department of Labor.

Chan, Yupo, and Fong-Lieh Ou. (1978). "Tabulating demand elasticities for urban travel forecasting." *Transportation Research Record*, vol 673 pp 40-46.

Chicago Area Transportation Study. (1986). *The O'Hare Rapid Transit Extension: Before-After Analysis*. Chicago: Chicago Area Transportation Study.

Chicago Area Transportation Study. (1996). *TIP: Transportation Improvement Program for Northeastern Illinois*. Chicago: Chicago Area Transportation Study.

Chicago Transit Authority. (1995). *Section 15 Report to the Federal Transit Administration for 1994*. Chicago: Chicago Transit Authority.

Cox, Wendell, Jean Love, and Nick Newton. (1995). "A summary of international urban transport competition with case studies; Copenhagen, London and San Diego." *Proceeding of the Fourth International Conference on Competition and Ownership in Land Passenger Transport*, Rotorua, New Zealand, pp 28-42.

Dodgson, John S. (1987). "Benefits of changes in urban public transport subsidies in the major Australian cities." In Stephen Glaister (ed.), *Transport Subsidy*. Newbury, UK: Policy Journals, pp 52-62.

Dodgson, John S., and Neville Topham. (1987a). "Benefit-cost rules for urban transit subsidies: An integration of allocational, distributional and public finance issues." *Journal of Transport Economics and Policy*, vol 21(1) pp 57-71.

Dodgson, John S., and Neville Topham. (1987b). "Shadow value of public funds: A survey." In Stephen Glaister (ed.), *Transport Subsidy*. Newbury, UK: Policy Journals, pp 114-119.

Glaister, Stephen. (1987). "Allocation of urban public transport subsidy." In Stephen Glaister (ed.), *Transport Subsidy*. Newbury, UK: Policy Journals, pp 27-39.

Glaister, Stephen, and David Lewis. (1978). "An integrated fares policy for transport in London." *Journal of Public Economics*, vol 9(3) pp 341-355.

Jorgenson, Dale W., and Kun-Young Yun. (1991). "The excess burden of taxation in the United States." *Journal of Accounting, Auditing & Finance*, vol 6(4) pp 487-508.

LaBelle, Sarah, and Darwin Stuart. (1995). "The CTA's Orange Line: Market share gains after sixteen months." *Proceedings of the Tenth Annual Metropolitan Conference on Public Transportation Research*. Chicago: Metropolitan Conference on Public Transportation Research.

LTI Consultants, Inc. (1988). *Fares Structure Analysis: Demonstration Study*. Report to the Chicago Transit Authority. London: LTI Consultants.

Mohring, Herbert. (1972). "Optimization and scale economies in urban bus transportation." *American Economic Review*, vol 62(4) pp 591-604.

Peat, Marwick, Mitchell & Co. (1972). *Implementation of the N-Logit Model*. San Diego: Comprehensive Planning Organization.

Seddon, P.A., and M.P. Day. (1974). "Bus passenger waiting times in Greater Manchester." *Traffic Engineering and Control*, vol 15(9) pp 442-445.

Transportation Research Board. (1985). *Highway Capacity Manual*. Special Report 209. Washington, D.C.: National Research Council.

Webster, F. Vernon, and Phillip H. Bly (eds.). (1980). *The Demand for Public Transport*. Crowthorne, UK: Transport & Road Research Laboratory.