This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 6, number 4

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm77-4

Publication Date: October 1977

Chapter Title: Models of Rail Road Passenger-Car Requirements in the Northeast Corridor: An Application of SeESAME

Chapter Author: Judith Gertler, Judith Gertler, Howard Simkowitz

Chapter URL: http://www.nber.org/chapters/c10523

Chapter pages in book: (p. 367 - 398)

Annals of Economic and Social Measurement, 6/4, 1977

MODELS OF RAILROAD PASSENGER-CAR REQUIREMENTS IN THE NORTHEAST CORRIDOR: AN APPLICATION OF SESAME

BY ROBERT FOURER,* JUDITH B. GERTLER,† AND HOWARD J. SIMKOWITZ†

We consider a general problem of determining optimal car allocations given a fixed schedule and predetermined demands. Requirements for car movements are modeled as a set of linear constraints having a transshipment structure, and alternative linear objectives are formulated. Various optimization techniques are developed for one or more objectives, and properties of the sets of optimal solutions are demonstrated. The model and optimization techniques are applied to projected rail service in the Northeast Corridor (Boston, New York, Philadelphia, Washington); derivation of a schedule and demands are explained, and results of a number of optimizations and analyses are displayed.

In 1973 Congress passed the Regional Railroad Reorganization Act, which became law on January 2, 1974. This complex piece of legislation called upon the U.S. Department of Transportation to improve passenger rail service in the Northeast Corridor, which extends from Boston, through New York and Philadelphia, to Washington, D.C. Subsequent planning for the improved service included engineering studies, financial analyses, and demand projections [1,2,6].

The research described herein began as an attempt to determine the minimum number of passenger cars required to serve the Northeast Corridor, given previously-determined schedules and estimates of demand. This is naturally viewed as a problem of constrained optimization. When the constraints imposed by demand and operating practices were expressed mathematically as equations and inequalities, the problem was seen to be an instance of a fairly general transshipment structure, as described in Section 1 of this paper. Such a structure is not specific to the Northeast Corridor, or to the movement of train cars (an application to locomotive requirements, for example, is given in §1.7). In addition, the constraints may be regarded as a fairly simple linear program, to which a feasible solution is easily found by standard methods.

Further analysis revealed that minimizing cars is but one of several

Models and computer routines described in this report were developed at the Computer Research Center of the National Bureau of Economic Research, under contract DOT-TSC-1179-1 from the Transportation Systems Center, U.S. Department of Transportation.

*National Bureau of Economic Research, Computer Research Center for Economics and Management Science, 575 Technology Square, Cambridge, MA 02139.

†Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142.

desirable objectives, and that each such objective may be viewed as a measure of cost of a particular kind: operating cost per mile, for example, or capital cost. Consequently, it was necessary to develop an approach to minimizing the "total cost" associated with two or more objectives, given a knowledge of the costs' relative magnitudes. This work is described in Section 2; much of it is applicable to linear programs generally. Moreover, the desired optimal solutions can be found by use of a standard "parametric" algorithm commonly employed in linear programming.

The remainder of this paper describes how the transshipment model was used to investigate rail service in the Northeast Corridor. For purposes of demonstration, a hypothetical case representing service on a busy day in 1982 was chosen as a basis for analysis. Base data for this case were estimated by the means described in Section 3. These data were incorporated in an appropriate instance of the model, which was solved and analyzed by use of NBER's SESAME interactive linear programming system [3,7,8] and supporting computer routines. Details of this base run, and some numerical results, are given in Section 4.

The base run was not intended as a thorough analysis of 1982 Corridor service, but as a test case to prepare the way for further analyses. Compilation of the base data, for example, led to development of techniques that are now available for more extensive studies. Output from the base run revealed some special properties of the Corridor network which in turn might be exploited in subsequent models (see, for example, §4.5 below).

In addition, application of the model requires an integrated set of interactive computer routines. These were developed and tested for the base run and are available to others via the NBERNET and TYMNET networks. Instructions for use of the computer routines are given in [5].

1. Formulation of the Model

It is desired to allocate "cars" of some sort in a transport network, subject to a fixed schedule and known demands for service. This section specifies the nature of such a network and the requirements that must be met by any feasible allocation of cars. To keep the discussion reasonably concrete, the model is described in terms of the railroad network that motivated it.

An informal statement of the problem occupies §1.1. The constraints are then formulated more precisely, first as a transshipment network (§§1.2–1.3), then as a linear program (§1.4) to which the simplex method may be applied.

The remainder of the section is concerned with extensions of the original problem to model corridor service with turnaround delays (§1.5), upper limits on train sizes (§1.6), and locomotive requirements (§1.7).

§1.1. *Statement of the Constraints*

A uniform fleet of passenger cars provides railroad service to a set of cities. Service is offered by means of a set of scheduled "trains", each comprising one or more cars and running between a given pair of cities. At any given time, each car in the fleet is either part of some currently running train, or is sitting in storage at one of the cities.

Two requirements constrain the size and deployment of the fleet: a fixed schedule, and known demands for scheduled trains.

Fixed schedule. The schedule lists all trains that depart in a chosen schedule-period (a day, for example). During the schedule-period, every scheduled train must be run, carrying one or more cars.

It is assumed that each schedule-period is followed immediately by another, identical schedule-period. Moreover, the same service is to be provided in every schedule-period: that is, the same schedule must be run, with the same allocation of cars to cities and trains.

Each entry in the schedule specifies a city of departure and a city of arrival, and corresponding departure and arrival times. In general, a train may arrive during the schedule-period (e.g., day) of departure, or during any subsequent period. For simplicity, however, it is assumed here that every train arrives either in the same period, or at an earlier time in the next period. (If the schedule-period is a day, this just says that a train arrives either the same day that it leaves, or the next day; and that every trip lasts less than 24 hours.)

A car that arrives at city c at time t is free to leave c in any scheduled train that departs at t or later. (Stopover delays at the arrival city — to discharge and board passengers, for example — are considered part of the preceding trip, and are reflected by adjusting the arrival time in the schedule accordingly.)

Demands. For each scheduled train there is a known demand which must be met; hence there is a minimum number of cars required in each train. A train may be larger than its minimum size, however, if circumstances require that extra (deadhead) cars be shifted from one city to another.

Table 1 shows a schedule and demands for a simple 2-city instance of this problem. Total demand from A to B requires 22 cars, while only 20 cars are required from B to A; consequently, in any feasible solution at least 2 extra cars will have to be deadheaded from B to A so that the stock of cars at A does not run out.

TABLE 1

SCHEDULE AND DEMANDS FOR A SIMPLE 2-CITY INSTANCE OF THE CAR-ALLOCATION PROBLEM. THE SCHEDULE-PERIOD IS A DAY. NOTE THAT THE LAST TRAIN FROM B LEAVES AT 23:00 EACH DAY AND ARRIVES AT 2:00 THE NEXT DAY.

CITY A to CITY B

Leave A	Arrive B	Demand	Cars required (75 pass./car)
10:00	13:00	398	6
12:00	15:00	177	3
16:00	19:00	259	4
18:00	21:00	557	7
21:00	24:00	121	2

CITY B to CITY A

Leave B	Arrive A	Demand	Cars required (75 pass./car)
9:00	12:00	209	3
11:00	14:00	280	4
15:00	18:00	373	5
19:00	22:00	421	6
23:00	2:00	90	2

§1.2. *Formulation As a Transshipment Network*

The train schedule is conveniently represented as a directed network whose unit of flow is one car. Nodes of the network correspond to the potential arrival or departure times at each city. Arcs represent the movement or storage of cars over time.

More specifically, partition the schedule-period into τ uniform intervals beginning at times $0, 1, \ldots, \tau - 1$. (If the schedule-period is a day, time t could be the beginning of the tth minute of the day.) Describe each train in the schedule by a departure city c, a departure time $t \in \{0, \ldots, \tau - 1\}$, an arrival city c', and an arrival time $t' \in \{0, \ldots, \tau - 1\}$. Clearly the schedule may be made as precise as desired by choosing τ sufficiently large.

Define one node in the network for each time in each city. If there are 4 cities and 1440 partitioning times, for example, the network has 4×1440 nodes.

Connect the nodes by arcs of two types, representing cars in storage and cars in trains, respectively:

Storage arcs. For each city, run an arc from the node for each time t to the node for the next time, $(t + 1) \bmod \tau$. The flow along such an arc represents cars held in storage at the city during the interval that begins at time t. (The last time, $\tau - 1$, is connected to the first time, 0, since the last interval of any schedule-period is followed immediately by the first interval of the next period.)

Train arcs. For each scheduled train, run an arc from the node representing the city and time of departure to the node for the city and time of arrival. Flow along this arc represents cars moving from one city to another in the scheduled train.

Flow around the network is constrained by the nature of the problem, in the following ways:

Conservation of flow. Since the fleet size is fixed, the number of cars in storage during interval t at a given city must equal the number in storage in the interval immediately before, plus the number that arrived at time t, less the number that departed at t. Equivalently, the net flow at every node must be zero: the network is built entirely of transshipment nodes.

Nonnegativity. All flows must be nonnegative. This amounts to requiring that trains cannot move backwards in time.

Integrality. Since cars are indivisible units, all flows must be integral.

Satisfaction of demand. The flow on each train arc must be greater than or equal to the number of cars needed to meet demand for the train. Demand thus places a lower limit on each arc. These lower limits are what force a positive flow around the network: they play the role of sources and sinks in more conventional transshipment-network formulations. (Indeed, an equivalent transshipment network without positive lower limits is easily constructed. One adds an appropriate sink for each departure at a node, and a source for each arrival.)

The network equivalent of Table 1's example is shown in Figure 1.

§1.3. *Reducing the Network*

If no trains arrive at or depart city c at time t, the node for c at t is connected to the rest of the network by only two storage arcs: an incoming arc from the previous time, and an outgoing arc to the following time. The flows on these two arcs must be the same in order to satisfy the conservation constraint. Consequently, one may remove the node and replace the two arcs with one. Other flows in the network are as before, and remain feasible if they were previously so; hence this transformation leaves the set of feasible solutions essentially unchanged.

When all such "inactive" nodes are removed, there remains a net-

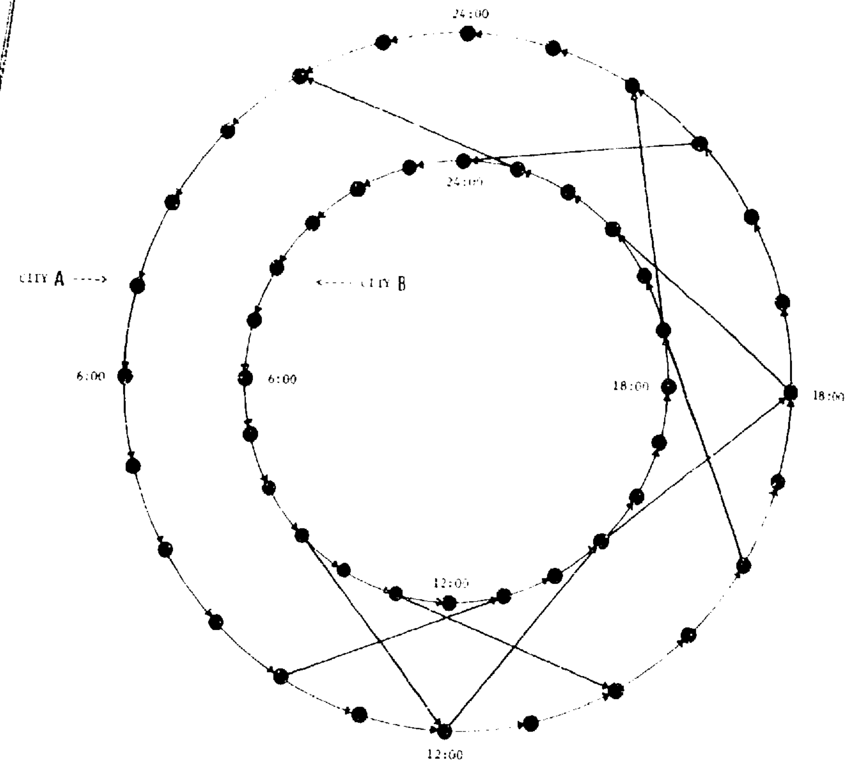

Figure 1 A network equivalent of the sample problem. The day is divided into $\tau = 24$ intervals, so that there is a node at each city at each hour.

work of minimum size for the problem. Figure 2 shows a reduced network of this sort, for the problem of Figure 1. When the number of intervals τ is quite large (the number of minutes in a day, for instance), reducing the network to active nodes is imperative if the network is to be kept to a manageable size. All cases run in the studies discussed later in this paper employed reduced networks.

It is possible to formulate the reduced problem directly, in terms of finite subsets of active times, one subset for each city, chosen from the interval $[0, \tau)$. To promote simplicity of notation, however, the results of the following sections are expressed in terms of unreduced networks.

§1.4. *Formulation as Linear Constraints*

Any network of the sort just outlined may be described by an equivalent linear-programming (LP) model. To each arc of the network there corresponds a structural variable, whose activity equals the arc's flow. Conservation constraints on flows become linear equalities in the vari-

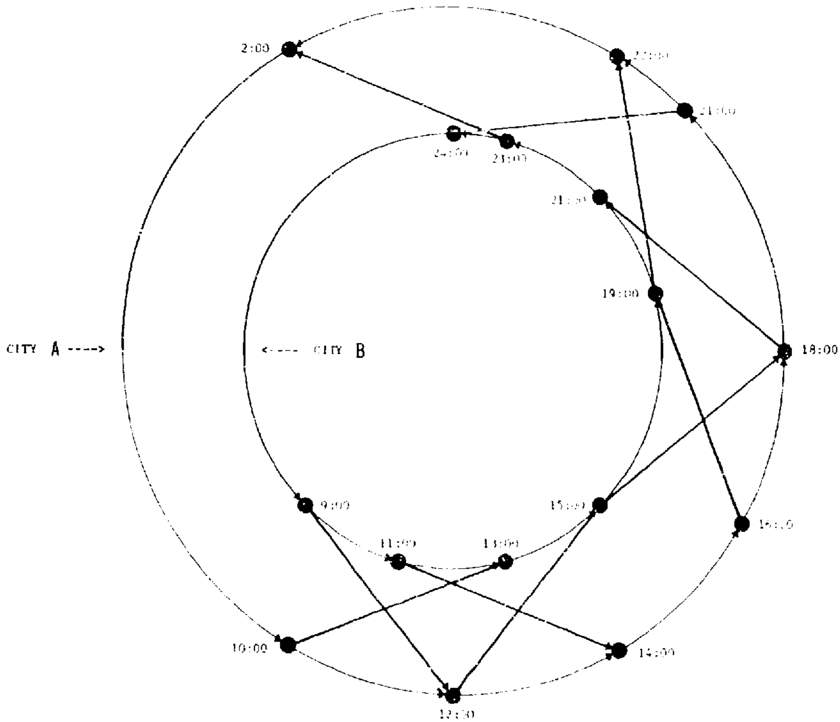

Figure 2 The reduced equivalent of the network in Figure 1.

ables, while common LP techniques can implicitly guarantee nonnegativity, integrality, and satisfaction of demand at every feasible basic solution.

To express the LP formally, define the following sets:

C — the set of cities

$T = \{0, \ldots, \tau - 1\}$ — the set of intervals into which the schedule-period is divided

$S \subset \{(c,t,c',t')\colon\ c \in C, c' \in C; t \in T, t' \in T; c \neq c'\}$

the schedule: each element represents a train that leaves city c at time t and arrives at city c' at t'

Represent the demands by

$d_{cc'}[t,t'] > 0$ — the smallest (integral) number of cars required to meet demand for train $(c,t,c',t') \in S$

Express the nodes of the network as:

$$\mathfrak{A}_c[t] \qquad \text{for all } c \in C, t \in T$$

The directed arcs representing storage of unused cars are then

$$\mathfrak{U}_c[t]: \mathfrak{A}_c[t] \to \mathfrak{A}_c[(t+1) \bmod \tau] \qquad \text{for all } c \in C, t \in T$$

The arcs representing movement of cars in trains are

$$\mathfrak{X}_{cc'}[t,t']: \mathfrak{A}_c[t] \to \mathfrak{A}_{c'}[t'] \qquad \text{for all } (c,t,c',t') \in S$$

Define an LP structural variable corresponding to each arc, and representing the flow over the arc:

$u_c[t]$	flow over $\mathfrak{U}_c[t]$, for all $c \in C, t \in T$
$x_{cc'}[t,t']$	flow over $\mathfrak{X}_{cc'}[t,t']$ for all $(c,t,c',t') \in S$

The constraints on network flow are expressed as follows:

Conservation of flow:

$$u_c[(t-1) \bmod \tau] + \sum_{(c_1,t_1,c,t) \in S} x_{c_1c}[t_1,t]$$
$$= u_c[t] + \sum_{(c,t,c_2,t_2) \in S} x_{cc_2}[t,t_2] \qquad \text{for all } c \in C, t \in T$$

Satisfaction of demand:

$$x_{cc'}[t,t'] \geq d_{cc'}[t,t'] \qquad \text{for all } (c,t,c',t') \in S$$

Nonnegativity:

$$u_c[t] \geq 0 \qquad \text{for all } c \in C, t \in T$$

Integrality:

$$u_c[t] \text{ integral} \qquad \text{for all } c \in C, t \in T$$

$$x_{cc'}[t,t'] \text{ integral} \qquad \text{for all } (c,t,c',t') \in S$$

Nonnegativity of the x variables is insured by satisfaction of demand.

Given that all $d_{cc'}[t,t']$ are integral, a fundamental property of transshipment problems guarantees that every basic solution to the above LP is an integral solution. Consequently, a feasible solution to the above problem—and hence a feasible allocation of cars to trains—may be determined directly by application of the (phase 1) simplex method. Given any linear objective function, the simplex method will also find an optimal feasible allocation.

Both satisfaction of demand and nonnegativity express simple lower bounds on the variables. Constraints of this sort are easily handled implicitly by the simplex method. Hence only the conservation-of-flow equations need appear explicitly as rows in the LP.

§1.5. *Corridor Service and Turnaround Delays*

A "corridor" is a set of cities related by a directional ordering that is complete, transitive, and irreflexive. In other words, the cities of a corridor may be indexed $c_1, c_2, \ldots, c_n$, such that c_i is in the given direction from c_j if and only if $i > j$. The Northeast Corridor is a corridor in this sense, ordered by the relation "north of".

Every train in a corridor must run in the ordering direction, or in the opposite direction. For convenience, these directions are here called north and south; they could just as well be east and west, or clockwise and counterclockwise. Trains are thus labeled northbound or southbound, accordingly.

In the initial formulation, stopover delay at the arrival city is implicit in the schedule and, therefore, it is the same for every car in a train. Within a corridor, however, it is reasonable to specify that the stopover delay for a car that changes direction is some number of intervals greater than the delay for a car that continues in the same direction along the corridor. Thus cars in a train from, say, Philadelphia to New York may continue to move north, after a minimal stop, in a train from New York to Boston; but cars in the Philadelphia-New York train that are to be taken off and sent back to Philadelphia are delayed in New York for a somewhat longer time. A similar "turnaround delay" is encountered in resuming service after one end of the corridor (say, Boston) is reached.

Turnaround delays cannot be modeled by simply adjusting the schedule because, in general, some cars in a train may continue in the same direction, while others are detached and turned around. A simple and feasible approach, however, is to duplicate the original network, creating two separate but similar parts: one for northbound trains, and one for southbound trains. Arcs connecting the two parts are added to represent cars being turned around.

Specifically, partition the schedule into two sets S^N and S^S of northbound and southbound trains, respectively. For the northbound trains, construct a full network as before:

$$\mathcal{A}_c^N[t] \qquad \text{for all } c \in C,\ t \in T$$

(nodes representing potential arrival and departure times of northbound trains)

$$\mathcal{U}_c^N[t]\colon\ \mathcal{A}_c^N[t] \to \mathcal{A}_c^N[(t+1) \bmod \tau] \qquad \text{for all } c \in C,\ t \in T$$

(arcs representing unused northbound cars in storage at each city and time)

$$\mathcal{X}_{cc'}[t,t']\colon\ \mathcal{A}_c^N[t] \to \mathcal{A}_{c'}^N[t'] \qquad \text{for all } (c,t,c',t') \in S^N$$

(arcs representing cars moving in northbound trains)

In the same way, define a separate network for southbound trains:

$$\mathfrak{A}_c^S[t] \qquad \text{for all } c \in C, t \in T$$

$$\mathfrak{U}_c^S[t]: \mathfrak{A}_c^S[t] \to \mathfrak{A}_c^S[(t+1) \bmod \tau] \qquad \text{for all } c \in C, t \in T$$

$$\mathfrak{X}_{cc'}[t,t']: \mathfrak{A}_c^S[t] \to \mathfrak{A}_{c'}^S[t'] \qquad \text{for all } (c,t,c',t') \in S^S$$

Represent the number of intervals required to change a car's direction by δ. Connect the northbound and southbound networks by two sets of arcs that represent unused cars in storage that are being turned around:

$$\mathfrak{U}_c^{NS}[t]: \mathfrak{A}_c^N[t] \to \mathfrak{A}_c^S[(t+\delta) \bmod \tau] \qquad \text{for all } c \in C, t \in T$$

(arcs representing formerly northbound cars, in storage at time t, that will be switched to run south δ intervals later)

$$\mathfrak{U}_c^{SN}[t]: \mathfrak{A}_c^S[t] \to \mathfrak{A}_c^N[(t+\delta) \bmod \tau] \qquad \text{for all } c \in C, t \in T$$

(arcs representing formerly southbound cars, in storage at time t, that will be switched to run north δ intervals later)

The construction of these connecting arcs guarantees that northbound cars reaching city c at time t must wait at least δ intervals before they can be incorporated in a southbound train.

The constraints on this expanded network are analogous in every respect to those on the original one: flow must be conserved at all nodes, all flows must be nonnegative and integral, and demand must be satisfied along the $\mathfrak{X}^N$ and $\mathfrak{X}^S$ arcs. As before, the network has a transshipment structure, and can be modeled by a linear program all of whose basic solution are integral.

For practical purposes, one can apply the methods of this section to the reduced network of §1.3, to produce separate reduced northbound and southbound networks having a reduced set of connecting arcs.

The corridor model is not fundamentally limited to the case of a single, fixed turnaround delay. One could easily incorporate a set of delays that vary with time, city, or direction, by making appropriate changes to the definitions of the $\mathfrak{U}^{NS}$ and $\mathfrak{U}^{SN}$ arcs. Extensions of these methods might also be applied to sets of cities that are not corridors.

§1.6. *Upper Limits on Train Sizes*

The model developed so far insures only that each train is allocated enough cars. One may also wish to specify that it is not allocated too many. For example, the number of cars in a train could be limited to twice the number needed to meet demand, to keep load factors at reasonable levels. Stations' platform lengths might also dictate some absolute bound on train sizes.

Upper limits are easily incorporated in the linear programs of §1.4 or §1.5. Define

$$h_{cc'}[t,t'] \geq d_{cc'}[t,t']$$

as the maximum feasible size of the train $(c,t,c',t') \in S$. Then the constraints on the x variables in the linear program are augmented to

$$d_{cc'}[t,t'] \leq x_{cc'}[t,t'] \leq h_{cc'}[t,t']$$

for all $(c,t,c',t') \in S$.

Upper limits of this sort do not destroy the model's transshipment structure. Hence all basic solutions are still integral, and the simplex method may be applied as before. Moreover, the augmented constraints on the x variables are still simple bounds that can be handled implicitly by the simplex method; the number of explicit rows in the LP is unchanged.

§1.7. *Modeling Locomotive Requirements*

In general, the number of locomotives required to haul a scheduled train depends on the number of cars assigned to the train. Since the number of cars may vary between feasible solutions, so may the number of locomotives.

By judicious choice of upper limits $h_{cc'}[t,t']$ (§1.6), however, one may be able to restrict the size of each train $(c,t,c',t') \in S$ so that its requirement for locomotives, $e_{cc'}[t,t']$, is fixed. Then the flow of locomotives may be modeled in exactly the same way as the flow of cars. One simply replaces car demands $d_{cc'}[t,t']$ in §§1.1–1.5 by the locomotive demands $e_{cc'}[t,t']$. Upper limits on the number of locomotives pulling each train may also be imposed, in the manner of §1.6.

Any of the optimization techniques described in section 2 may be applied to the locomotive-demand case. Many of the results expressed in terms of cars are also meaningful in terms of locomotives.

Application of these ideas to locomotive requirements in the Northeast Corridor is described in §4.7.

2. Objective Functions

A feasible set of car allocations for the problem formulated in the preceding section—if such a set exists—may be determined by application of the simplex algorithm, phase 1. Given that a feasible allocation exists, the next step is to seek an allocation that optimizes some functional in the

x and u variables. This paper is concerned with functionals of one particularly useful and tractable sort: linear objective functions related to costs.

Minimizing cost is a natural objective for any planning model. Since section I's network model, in particular, fixes the level of service and requires that all demands be met, cost is the principal criterion of difference between feasible allocations. In addition, certain classes of minimum-cost solutions may be characterized in particularly revealing ways.

Linear functionals have a purely practical justification: they may be minimized by straightforward application of the simplex method. Fortunately, several reasonable measures of cost are proportional to linear functionals, as shown in §2.1.

Approaches to minimizing more than one linear cost objective are discussed in §2.2. The case of two objectives is developed in §§2.3–2.4, and the results are applied in §2.5 to two objectives of particular interest.

For convenience of exposition, the schedule-period is hereafter taken to be a day. A set of solution activities of the x and u variables is written $(\mathbf{x}, \mathbf{u})$, and the value of a functional Z at the solution is $Z(\mathbf{x}, \mathbf{u})$.

§2.1. *Linear Functionals Representing Costs*

There is more than one sort of cost associated with railroad service, and consequently one may devise a number of linear forms that are proportional to cost of some sort. Three functionals of particular interest—associated with capital, operating, and switching costs, respectively—are formulated as follows:

Capital cost. The daily cost of amortizing the passenger-car fleet, here referred to as the "capital cost", may be considered proportional to the number of cars in the fleet. Hence minimizing fleet size serves to minimize capital cost.

The number of cars is easily represented by a linear form. Pick any time t^*, $0 \le t^* \le \tau - 1$, and sum (a) the number of cars in storage at each city in interval t^*, and (b) the number of cars in each train that is in transit during interval t^*. This sum is the total number of cars in the system at t^*. For a feasible solution, this sum must be the same at any t^* since cars may not enter or leave the system. For convenience, take $t^* = \tau - 1$; then the capital-cost objective is a linear combination

$$Z_{\mathrm{CAR}} = \sum_{c \in C} u_c[\tau - 1] + \sum_{\substack{(c,t,c',t') \in S \\ t' < t}} x_{cc'}[t, t']$$

The first sum covers all cars in storage during interval $\tau - 1$. The latter counts cars in only those trains which depart during one day and arrive the next: these are exactly the trains that are in transit during the last interval, $\tau - 1$, of the day.

Operating cost. Costs proportional to the number of car-miles run in a day, here called "operating costs", are another logical candidate for minimization. Letting the distance from c to c' be $m_{cc'}$, total car-miles per day is equal to the linear form

$$Z_{\text{MILE}} = \sum_{(c,t,c',t')\in S} m_{cc'} x_{cc'}[t,t']$$

Note that at any feasible solution Z_{MILE} is also a sum of integral multiples of the distances $m_{cc'}$. Moreover, when the cities form a corridor (§1.5), Z_{MILE} is a sum of integral multiples of the round-trip distances:

$$m_{c_i c_j} + m_{c_j c_i}, \quad i < j$$

since conservation of the flow of cars requires that the number of cars run north from c_i to c_j during a day is the same as the number run south from c_j to c_i.

Z_{MILE} is also closely related to load factor. Given fixed demands, it is reasonable to try to maximize system load factor in order to minimize the cost of providing service. By definition, system load factor is

$$Z_{\text{LF}} = \frac{\text{passenger-miles / day}}{\text{seat-miles / day}}$$

$$= \frac{(\text{passenger-miles / day}) / (\text{seats / car})}{\text{car-miles / day}}.$$

Since both passenger-miles/day and seats/car are fixed by the problem, Z_{LF} is inversely proportional to car-miles/day $= Z_{\text{MILE}}$. Hence minimizing operating cost is equivalent to maximizing the system load factor.

Switching cost. For the corridor model of §1.5, one may postulate an extra fixed "switching" cost incurred each time a car's direction is reversed. The number of car-reversals in a day is counted by the following linear form:

$$Z_{\text{TURN}} = \sum_{c\in C, t\in T} u_c^{\text{NS}}[t] + \sum_{c\in C, t\in T} u_c^{\text{SN}}[t]$$

The first term sums all northbound cars turned south, and the second all southbound cars turned north.

§2.2 *Combining Measures of Cost*

It was shown in §2.1 that there are several reasonable "costs" that are proportional to linear functionals in the u and x variables. As a consequence, no solution that merely minimizes one of these functionals is entirely satisfactory. For example, an allocation that minimizes the number of cars (capital cost) may nonetheless employ them inefficiently, running them more than the minimum car-miles/day (operating cost).

Some means is needed, therefore, of optimizing with respect to more than one cost objective. Two methods suggest themselves, combining objectives so that they are minimized simultaneously, and ordering objectives so that they may be minimized successively.

Combining objectives. Any n objective functions $Z_1, Z_2, \ldots, Z_n$ can be combined by choosing factors $p_1, p_2, \ldots, p_n > 0$, and minimizing the linear combination

$$Z = p_1 Z_1 + p_2 Z_2 + \cdots + p_n Z_n$$

Minimizing Z tends to minimize each of the Z_i. The value of Z_i at min Z is, however, generally greater than min Z_i; the extent of the discrepancy depends on the size of p_i with respect to the other factors.

Z has a natural interpretation when there is some cost proportional to each Z_i. Let p_i be the constant of proportionality, so that $p_i Z_i$ is the cost (in dollars, say) corresponding to any given level of Z_i. (If Z_1 is car-miles/day, for example, p_1 could be operating expense in dollars/car-mile.) Z is thus a "total variable cost" for the system, and minimizing Z can be seen as minimizing total cost.

The difficulty with this approach lies in determining true values for the constants p_i. Even small changes to the p_i can produce significant differences in the solution to min Z; yet, especially when a hypothetical system is being modeled, costs are often poorly known and the p_i can be determined only to within a wide tolerance. Hence it is necessary to treat the p_i as somewhat variable, and to find solutions for ranges of their values. (An efficient and exhaustive way of doing this when total cost is the sum of two costs is described in the following section.)

Ordering objectives. Another approach is to rank the objectives, minimizing Z_i subject to $Z_1, \ldots, Z_{i-1}$ being fixed at their previously attained values. One first computes min Z_1, the absolute minimum value of Z_1; then min $Z_2 \mid Z_1$, the minimum value of Z_2 given $Z_1 = \min Z_1$; then min $Z_3 \mid Z_2 \mid Z_1$, the minimum value of Z_3 given $Z_2 = \min Z_2 \mid Z_1$ and $Z_1 = \min Z_1$; and so forth. In general, min $Z_i \mid Z_{i-1} \mid \cdots \mid Z_1$ is greater than the absolute min Z_i, and the discrepancy tends to become greater as i does.

A solution to min $Z_2 \mid Z_1$ is found, in effect, by adding a new equality constraint ($Z_1 = \min Z_1$). The original problem's pure transshipment structure is thus violated. Nevertheless, an optimal integral solution is guaranteed by the following Proposition.

*Proposition 1.** For any linear forms $Z_1, Z_2, \ldots, Z_n$, there is an integral basic solution to min $Z_n \mid Z_{n-1} \mid \cdots \mid Z_1$.

*Proofs of all propositions in this section are given in [4].

Sequential optimization has the advantage of requiring only a preferential ordering of costs, rather than a full determination of their relative sizes. It is disadvantageous primarily in being less general than the "total cost" approach above. (The two approaches are closely related, however, as shown below in §2.4.)

§2.3. *The Case of Two Objective Functions*

When attention is restricted to two cost objectives, the set of all possible allocations can be described in a simple way. Moreover, the representative optima are easily found by use of an algorithm for parametric programming on the objective.

Denote the two objectives by Y and Z, and their respective expenses per unit by p_Y and p_Z. A total cost determined by Y and Z is thus $p_Y Y + p_Z Z$. The minimum total cost is:

$$\begin{aligned} \min\,[p_Y Y + p_Z Z] &= p_Y \min\,[Y + (p_Z/p_Y)Z] \\ &= p_Y \min\,[Y + \rho Z] \end{aligned}$$

where $\rho = p_Z/p_Y$ is the ratio of expenses per unit. Hence the minimum total cost is determined entirely by the choice of ρ.

The set of all solutions that can minimize total cost, given some choice of ρ, is characterized in the following Proposition:

Proposition 2. Let Y and Z be objectives for which min Y and min Z are finite. For any $(\mathbf{x}^*, \mathbf{u}^*)$, define:

$$R_{\mathbf{x}^*,\mathbf{u}^*} = \{k \geq 0 \mid (\mathbf{x}^*, \mathbf{u}^*) \text{ minimizes } Y + kZ\}$$

Then:

(a) There is a unique sequence

$$0 = \rho_0, \rho_1, \ldots, \rho_{n-1}, \rho_n = \infty, \qquad n \geq 1; \quad \rho_{i-1} < \rho_i, i = 1, \ldots, n$$

and there is a corresponding set of distinct basic solutions

$$(\mathbf{x}_i^*, \mathbf{u}_i^*) \qquad\qquad i = 1, \ldots, n$$

so that

$$R_{\mathbf{x}_i^*,\mathbf{u}_i^*} = [\rho_{i-1}, \rho_i] \qquad\qquad i = 1, \ldots, n$$

(b) for any solution $(\mathbf{x}^*, \mathbf{u}^*)$, exactly one of the following holds:

(i) $R_{\mathbf{x}^*,\mathbf{u}^*} = \phi$

(ii) $R_{\mathbf{x}^*,\mathbf{u}^*} = \{\rho_i\}$, for some $i \in \{0, \ldots, n-1\}$

(iii) $R_{\mathbf{x}^*,\mathbf{u}^*} = [\rho_{i-1}, \rho_i]$, for some $i \in \{1, \ldots, n\}$

(c) For every $i = 1, \ldots, n-1$,

$$Y(\mathbf{x}_i^*, \mathbf{u}_i^*) < Y(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

$$Z(\mathbf{x}_i^*, \mathbf{u}_i^*) > Z(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

What do the values ρ_i signify? They are the *critical ratios* p_Z/p_Y at which the allocation of cars must change to maintain optimal total cost. So long as p_Z/p_Y stays between some ρ_{i-1} and ρ_i, however, a single allocation $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ is guaranteed optimal.

Another way of looking at things is to note that, at critical point $p_Z/p_Y = \rho_i$,

$$Y(\mathbf{x}_i^*, \mathbf{u}_i^*) + \rho_i Z(\mathbf{x}_i^*, \mathbf{u}_i^*) = Y(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*) + \rho_i Z(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

which may be rewritten

$$p_Y[Y(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*) - Y(\mathbf{x}_i^*, \mathbf{u}_i^*)] = p_Z[Z(\mathbf{x}_i^*, \mathbf{u}_i^*) - Z(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)]$$

Proposition 2(c) says that changing from $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ to $(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$ involves a tradeoff: Z decreases while Y increases. At the critical point, the added cost from the increase in Y (left side of above equation) equals the cost saved by decreasing Z (right side). At $\rho < \rho_i$, the saved cost does not make up the added cost, and so $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ is preferable; at $\rho > \rho_i$, the saved cost more than makes up the added cost, and hence $(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$ is better.

The critical ratios ρ_i and solutions $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ are easily found by applying the standard parametric algorithm to the objective. In conventional terms, Y is the "base objective" and Z the "change objective". The algorithm starts with a solution for min Y, and "parameter" ρ at 0. Successive pivots either leave ρ unchanged, or step it to a new critical value that is generally one of the critical ratios ρ_i: the basis just before the step to ρ_i is $(\mathbf{x}_i^*, \mathbf{u}_i^*)$. The algorithm terminates when it finds a solution that is optimal for all parameter values greater than some critical value: this solution is $(\mathbf{x}_n^*, \mathbf{u}_n^*)$, and the critical value is ρ_{n-1}.

In some instances, the parametric algorithm identifies a supposed critical ratio ρ_i such that

$$Y(\mathbf{x}_i^*, \mathbf{u}_i^*) = Y(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

$$Z(\mathbf{x}_i^*, \mathbf{u}_i^*) = Z(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

This cannot be a true critical ratio, however, since the above equalities violate Proposition 2(c). Indeed, these equalities imply that both $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ and $(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$ minimize $Y + \rho Z$ for all ρ such that $\rho_{i-1} < \rho < \rho_{i+1}$, so that ρ_i is actually not critical at all. Spurious ratios of this sort are a side effect of degeneracy in the linear program.

§2.4. *Conditional Optima for the Case of Two Objectives*

The solutions $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ derived in Proposition 2 also have an interpretation in terms of min $Z \mid Y$, min $Y \mid Z$, and other conditional optima. This is shown in the following Proposition:

Proposition 3. The solutions $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ defined in Proposition 2 have the following properties:

(a) $(\mathbf{x}_1^*, \mathbf{u}_1^*)$ minimizes Y
 $(\mathbf{x}_1^*, \mathbf{u}_1^*)$ minimizes $Z \mid Y$

(b) $(\mathbf{x}_n^*, \mathbf{u}_n^*)$ minimizes Z
 $(\mathbf{x}_n^*, \mathbf{u}_n^*)$ minimizes $Y \mid Z$

(c) $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ minimizes $Z \mid (Y + \rho Z)$ when $\rho_{i-1} \leq \rho < \rho_i$ for $i = 1, \ldots, n$
 $(\mathbf{x}_i^*, \mathbf{u}_i^*)$ minimizes $Y \mid (Y + \rho Z)$ when $\rho_{i-1} < \rho \leq \rho_i$ for $i = 1, \ldots, n$

Proposition 3(a) says that minimizing $Y \mid Z$ yields the best solution when $\rho = p_Z / p_Y$ is small enough. In other words, when p_Y is sufficiently large relative to p_Z, Y dominates the total cost: the best solution is one that minimizes Y outright, then Z as much as possible. Proposition 3(b) makes the equivalent statement for the case where $\rho = p_Z / p_Y$ is sufficiently large that Z dominates total cost.

Note that if $n > 2$ there is at least one middle region of ρ where the best solution minimizes neither Y nor Z absolutely. When $n = 2$, the optimal solutions for total cost minimize either $Y \mid Z$ (for $\rho \leq \rho_1$) or $Z \mid Y$ (for $\rho \geq \rho_1$). When $n = 1$, there is a single solution that minimizes both Y and Z absolutely, and hence minimizes any $Y + \rho Z$.

§2.5. *Tradeoffs between Capital and Operating Costs*

Of special interest is application of the preceding section's results to functionals Z_{CAR} and Z_{MILE}, defined in §2.1 as proportional to notions of capital cost and operating cost, respectively. Total variable cost with respect to these two objectives is

$$p_{\mathrm{CAR}} Z_{\mathrm{CAR}} + p_{\mathrm{MILE}} Z_{\mathrm{MILE}}$$

where

$$\begin{aligned} p_{\mathrm{CAR}} &= \text{capital cost/car/day} \\ Z_{\mathrm{CAR}} &= \text{number of cars} \\ p_{\mathrm{MILE}} &= \text{operating cost/car-mile} \\ Z_{\mathrm{MILE}} &= \text{car-miles/day} \end{aligned}$$

The choice of a solution that minimizes total cost depends upon $p_{\mathrm{CAR}} / p_{\mathrm{MILE}}$, the ratio of capital cost/day to operating cost/mile.

The critical ratios for this problem have a special form related to the inter-city distances, as demonstrated by the following Proposition:

Proposition 4. (a) For objectives of the form

$$Z_{\text{MILE}} + (p_{\text{CAR}}/p_{\text{MILE}})Z_{\text{CAR}},$$

the critical ratios of $\rho = p_{\text{CAR}}/p_{\text{MILE}}$ (as defined in Proposition 2) have the form

$$\rho_i = \frac{1}{\kappa} \sum_{\substack{c,c' \in C \\ c \neq c'}} \alpha_{cc'} m_{cc'}$$

where $\alpha_{cc'}$ are integers, and κ is a positive interger satisfying:

$$\kappa \leq Z_{\text{CAR}}(\mathbf{x}_i^*, \mathbf{u}_i^*) - Z_{\text{CAR}}(\mathbf{x}_{i+1}^*, \mathbf{u}_{i+1}^*)$$

(b) If the cities constitute a corridor (§1.5) ordered $c_1, \ldots, c_k$ then under the assumptions in (a) the critical ratios have the form

$$\rho_i = \frac{1}{\kappa} \sum_{\substack{c_i, c_j \in C \\ i < j}} \alpha_{c_i c_j}(m_{c_i c_j} + m_{c_j c_i})$$

where $\alpha_{c_i c_j}$ are integers, and κ is a positive integer satisfying the inequality in (a).

Proposition 4 offers a characterization of the critical ratios for Z_{MILE} and Z_{CAR}. At ratios $\rho = p_{\text{CAR}}/p_{\text{MILE}}$ such that $\rho_i \leq \rho < \rho_{i+1}$, adding κ cars to the system makes possible a net saving of $\Sigma\alpha_{cc'}m_{cc'}$ car-miles/day. So long as $p_{\text{CAR}}/p_{\text{MILE}} > \rho_i$, however, the proposition implies that

$$\kappa p_{\text{CAR}} > (\Sigma\alpha_{cc'}m_{cc'})p_{\text{MILE}}$$

The cost of adding κ cars (left-hand side) is greater than the cost saved by the reduction in car-miles (right-hand side), and hence adding the cars is uneconomical. For $p_{\text{CAR}}/p_{\text{MILE}} < \rho_i$, the inequality is reversed, so that total cost is less when the cars are added. When $p_{\text{CAR}}/p_{\text{MILE}} = \rho_i$, however,

$$\kappa p_{\text{CAR}} = (\Sigma\alpha_{cc'}m_{cc'})p_{\text{MILE}}$$

Hence ρ_i is the ratio of capital to operating expense at which the capital cost of adding cars is exactly balanced by a resultant saving in operating cost.

For the Northeast Corridor data described later in this paper, all critical ratios had the especially simple form $\kappa = 1$, $\alpha_{c_i c_j} = 0$ or 1. At these ratios the capital cost of one added car equalled the operating cost over one round trip that was saved by adding the car: see further in §§4.4, 4.6. (It may be that under certain assumptions about the network and schedule, critical ratios must have simple forms like this; but such has not been shown formally.)

3. Base Data for the Northeast Corridor

As a demonstration case, the general transshipment structure was applied to anticipated Northeast Corridor service for 1982. This section describes how Corridor operations were modeled (§§3.1–3.2), and how base data for 1982 were derived (§§3.3–3.8).

The primary reference for data-gathering techniques is a pair of Corridor studies prepared by Peat, Marwick, Mitchell and Company [1, 6]. These are referred to in the sequel as the "PMM studies".

§3.1. *Characteristics of the Northeast Corridor*

The base-run Northeast Corridor comprises four *terminals:* Boston, New York, Philadelphia, and Washington. Scheduled trains connect these terminals on three north-south *segments* as follows:

Segment	*Length (miles)*
Boston–New York	232
New York–Philadelphia	90
Philadelphia–Washington	135

Cars arriving at a terminal may move on immediately in the same direction, or may be stored for use in later trains in either direction. A fixed minimum amount of time (in addition to the normal stopover time) is required to change the direction of a car.

Also in the Corridor are seven intermediate *stations*, as shown in Figure 3. Trains are scheduled to stop at these stations, but cars may not be stored or switched there. Including both terminals and intermediate stations, the corridor comprises 11 cities, connected by 10 north-south *links*.

For purposes of the 1982 base run, cars in Corridor service are assumed to have a uniform capacity of 75 passengers. Station size is taken to be 14 cars; trains requiring more than 14 cars are to be run in multiple sections of 14 cars or less each. Each section is assumed to require one locomotive.

§3.2. *Modeling Northeast Corridor Service*

The Northeast Corridor is modeled as a corridor network with turnaround delay, as defined in §1.5.

The set of C of cities in the model comprises the four Corridor terminals:

$$C = \{\text{Boston, New York, Philadelphia, Washington}\}$$

Intermediate stations can be omitted from this set, since they are not points at which cars may be stored or switched. (Demands to and from

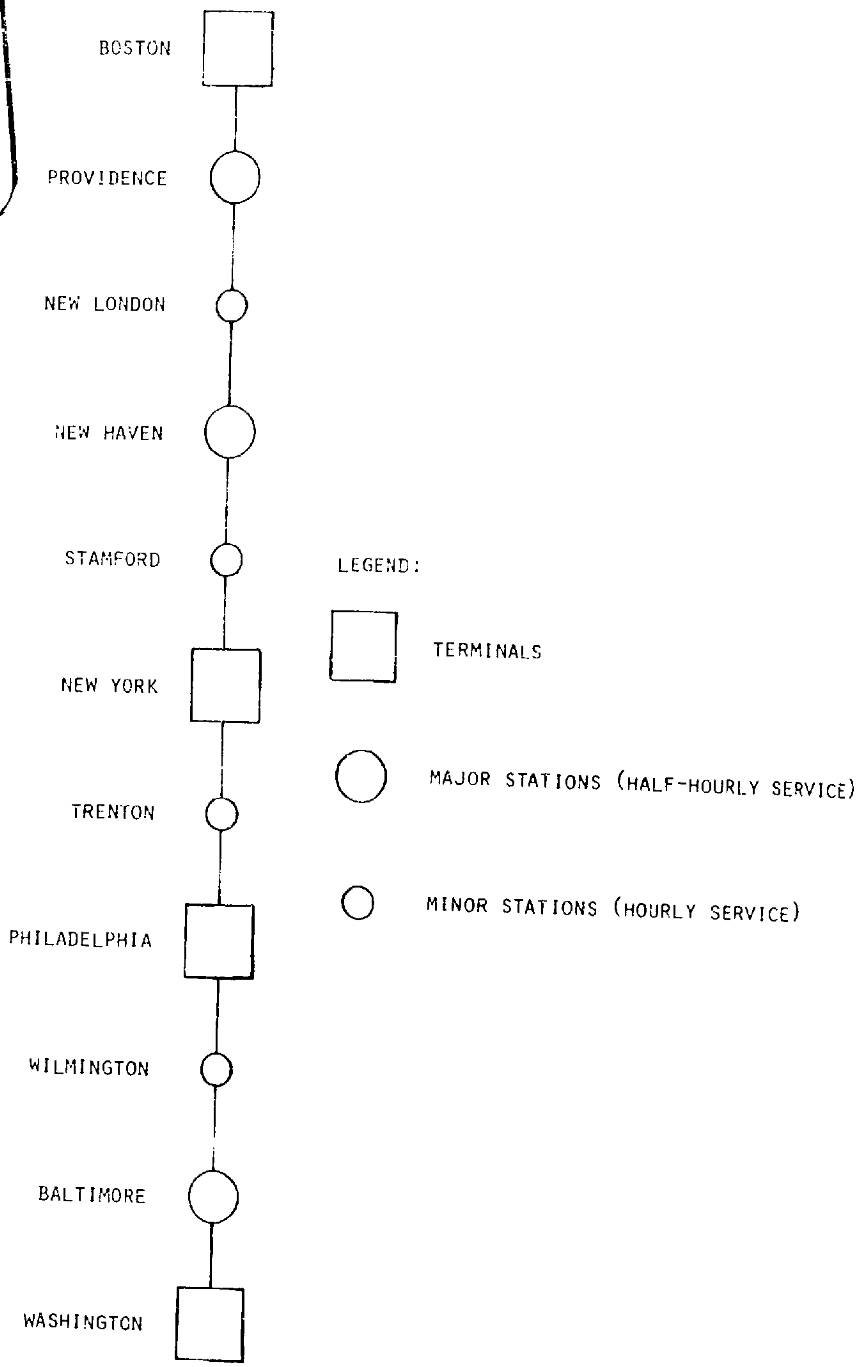

Figure 3 Terminals and intermediate stations in the Northeast Corridor as modeled by the base run.

intermediate stations are used to determine the minimum size of each train, however. See §§3.4–3.6.)

The model's schedule-period is one day, partitioned into a set of intervals T representing minutes of the day. Hence the number of partition intervals, τ, is 1440.

The schedule, S, lists the arrival and departure terminal of each train and the corresponding arrival and departure times to the nearest minute. Its construction is described in §3.3.

The demand for each train is calculated from annual patronage forecasts by the methods described in §§3.4–3.6. A lower limit $d_{cc'}[t,t']$ and upper limit $h_{cc'}[t,t']$ on each train's size is then derived from its demand, as explained by §3.7.

The turnaround delay δ is fixed at 20 minutes, for reasons set forth in §3.8.

§3.3. *The Schedule*

The 11-city base schedule is an updated version of that employed in the PMM studies [1,6]. It assumes generally half-hourly service to the terminals and major intermediate stations (Providence, New Haven, Baltimore), and hourly service at minor stations (New London, Stamford, Trenton, Wilmington). Appropriate reductions are made late at night and early in the morning, when demands are very low.

Segment trip times for 1982 are assumed to be approximately as follows:

Segment	*Trip time*
Boston–New York	3 hours 40 minutes
New York–Philadelphia	1 hour 1 minute
Philadelphia–Washington	1 hour 38 minutes

Trip times for individual links are calculated accordingly. Allowance is made in addition for stopover times of about 5 minutes at New York, and 1.25 minutes at other stations. It is assumed, however, that trains do not save any time when they skip stops at minor stations.

The full 11-city schedule is used in calculating demands, as described below (§§3.4–3.6). In forming the transshipment network, however, only the arrival and departure times at the terminal cities are employed. (The full base schedule is printed in [4].)

§3.4. *Design-day Patronage*

Annual patronage for 1982 was calculated by use of a computer-based model developed in one of the PMM studies [6]. The input data were those derived from PMM's "base assumptions" with the exception of trip times, which were increased to reflect the base schedule (§3.3).

The PMM model estimates annual two-way patronage for individual station-pairs in the Northeast Corridor. Annual one-way patronage is computed by halving the two-way figures. A few possible station-pairs are omitted, either because they could not be separated from other pairs, or because competitive commuter service is available for their travelers. All of these excluded pairs are short distance, and are deemed to be relatively insignificant to Corridor service.

The base run models patronage for a *design day*, calculated as 1/270 of the annual amount. This concept of design day, representing approximately the 10th busiest day of the year, has been employed in engineering studies of the Northeast Corridor [2, pp. 3 35] and in fleet-sizing experiments [1, Appendix C].

§3.5. *Demand Distributions*

The base run employs a set of cumulative demand functions to derive the patterns of demand between station-pairs over a day. Following a method of the PMM studies [1, pp. C.7 C.14], demand for service from a larger station to a smaller one is taken to be departure-based (that is, dependent upon the time of departure), while demand for service from a smaller to a larger station is arrival-based (dependent upon time of arrival). Demand between cities of comparable size is determined by averaging arrival-based and departure-based distribution functions.

The demand distributions employed in the base run are derived from bimodal gaussian-like probability distributions* fit to actual arrival and departure counts for Tuesday, May 21, 1974. This date was chosen because it afforded actual ticketing data, and was uninfluenced by special weekend or holiday patterns. Counts could be made, however, for only a small number of station-pairs, especially as no information was available for trips passing through New York. In consequence, actual distributions were fit for ten particular pairs only, and these are used to approximate the distributions for other station-pairs (see [4] for further details).

§3.6. *Effective Demands Over Segments*

For every scheduled train over a segment, there is an *effective demand*: the number of passengers that the train must accommodate to guarantee everyone a seat at all points on the route. Effective demands are determined for the base run in the following steps:

Station-pair demands. Given one-way patronage data (§3.4) and cumulative demand functions (§3.5), a design-day demand is calculated

*These distributions were derived and estimated by Walter Messcher and Alan Wellington at the Transportation Systems Center, U.S. Department of Transportation. See further in the Appendix to [4].

for every scheduled trip between a pair of stations in the 11-city schedule (excluding certain station-pairs as explained in §3.3).

Link demands. Total demand for any train over a given link is computed as a sum of all station-pair demands that involve travel over that link.

For example, total demand for a typical scheduled train over the Wilmington-Philadelphia link does not include only passengers who get on at Wilmington and disembark at Philadelphia. Some passengers who get off at Philadelphia began their trip in Washington or Baltimore; some who start at Wilmington will stay on to Trenton, New York, or a station further north; and some passengers both start south of Wilmington and terminate north of Philadelphia. Demand for the train for all such station-pair trips must be added to determine total demand for the train over the Wilmington-Philadelphia link.

Maximal-link demands. For every train over a particular segment, there emerges from the link demands a *maximal link* over which demand is highest. A train accommodates all passengers over a segment only if it meets demand over the maximal link, since cars cannot be added within the segment. Hence the effective demand for each train is equal to the train's maximal link demand.

For instance, say demand for a Washington-Philadelphia train is 197 passengers over the Washington-Baltimore link, 237 over the Baltimore-Wilmington link, and 225 over the Wilmington-Philadelphia link. The maximal link for that train is then Baltimore-Wilmington, and effective demand for the train is 237.

§3.7. *Minimum and Maximum Train Sizes*

For the base run, cars are assumed to hold 75 passengers. Hence if d is the effective demand for a train, its minimum size is:

$$\langle d/75 \rangle$$

(Here angle brackets denote the least integer greater than the enclosed value.)

The maximum size of a train for the base run is the lesser of two limits, one related to load factor, the other to station length:

Load-factor limit. Due to imbalances in demand throughout the day, some trains will have to be run with more than the absolute minimum number of cars. It is reasonable, however, to limit the number of these deadhead cars to some proportion of the train. Specifically, in the base run no train is allowed to have more cars than required to meet twice its effective demand, with the proviso that every train may have at least 2 cars.

In terms of d, this limit is:

$$\max(\langle 2d/75 \rangle, 2)$$

Its effect is to require load factor over the maximal link to be at least reasonably near 50%, the requirement becoming stricter at larger demands.

Station-length limit. Plans for 1982 assume that stations will hold at most 14-car trains (§3.1). When more than 14 cars are assigned to a train, one or more extra sections must be put on, employing an equal number of extra locomotives. To prevent unnecessary extra sections, the base run requires that the number of sections actually run be no greater than $\langle (d/75)/14 \rangle$, the number of sections required to meet effective demand. This translates to an upper limit on cars of:

$$14 \langle (d/75)/14 \rangle$$

If $d/75$ is 12.6, for instance, this upper limit is 14; but if $d/75$ is 15.2, two sections are needed in any event, and the limit is 28.

The load-factor upper limit is the lesser one for demands under 525 passengers (7 cars). At larger demands, the station-length limit predominates.

For the base run, only 5 trains require as many as two sections: three from Philadelphia to New York in the morning, and two from New York to Philadelphia in the afternoon. Most other trains of 7 or more cars are also on the New York-Philadelphia segment.

§3.8. *Turnaround Delay*

For the base run a delay of 20 minutes (in addition to the regular stopping time built into the schedule) is postulated whenever the direction of a car is changed. This time is deemed sufficient to cover switching under 1982 conditions plus any lags in the arrival of extra sections.

It happens that for the base schedule any turnaround delay from 9 to 20 minutes has the same effect. A delay of more than 20 minutes requires additional cars at Philadelphia.

4. Base Runs With the Northeast Corridor Data

Computer processing and its results for the base run are discussed in this section. The principal computing tool was the SESAME linear programming system developed at the National Bureau of Economic Research [3,7,8].

Generation of the base data (§4.1) and the LP model (§4.2) were necessarily performed first. Optimal solutions were then found for a variety of objectives: minimum cars and car-miles, and maximum load factor (§4.3); minimum total operating and capital cost (§4.4); and minimum turnaround switching (§4.5). Further analyses included sensitivity to demands (§4.6) and requirements for locomotives (§4.7).

§4.1. *Computing the Base Data*

Estimates of 1982 rail patronage (§3.4) were produced by running a computer simulation program adapted specially for the PMM demand study [6]. This program projects business and non-business use of four modes of travel: rail, bus, air, and car. A subroutine was added to file total rail patronage only, in a format suitable for subsequent processing.

The patronage data file, plus a file representing the full schedule, then served as input to a demand-calculating program. This program employs cumulative demand functions for station-pairs to compute effective demands, and consequent upper and lower limits, for each train (as described in §§3.5–3.7).

Principal output from the demand program is a set of tables, representing the schedule and other information, that can be read by an LP matrix generator (§4.2). In addition, sets of alternative train-size limits are filed in a form that allows any one set to be read into the matrix.

§4.2. *Generating the Model*

An LP equivalent of the network model was generated in a form suitable for computer processing by DATAMAT, a subsystem of SESAME [3]. A program in the DATAMAT macro language was written for this purpose.

Upper and lower limits on train-size variables are not generated as explicit constraint rows: they are incorporated in a "bound set" that is enforced implicitly by SESAME's simplex algorithm. Actual limit values are also absent at this stage: they are read in from a separate file just before the model is solved or analyzed. This arrangement facilitates working with several sets of limits, as was done, for example, in the sensitivity analysis described in §4.6.

The LP generated by the DATAMAT program represents a reduced network, duplicated to distinguish northbound and southbound cars in the corridor (§§1.3, 1.5). For the base schedule, the LP representation required 1275 structural variables and 528 constraint rows.

§4.3. *Minimizing Cars and Car-Miles*

The base-run LP was solved by use of SESAME's standard primal simplex algorithm. A feasible solution was obtained (starting from an all-slack basis) in 665 iterations, and an optimal solution for the minimum-cars objective, Z_{CAR}, in an additional 28 iterations. An optimal solution

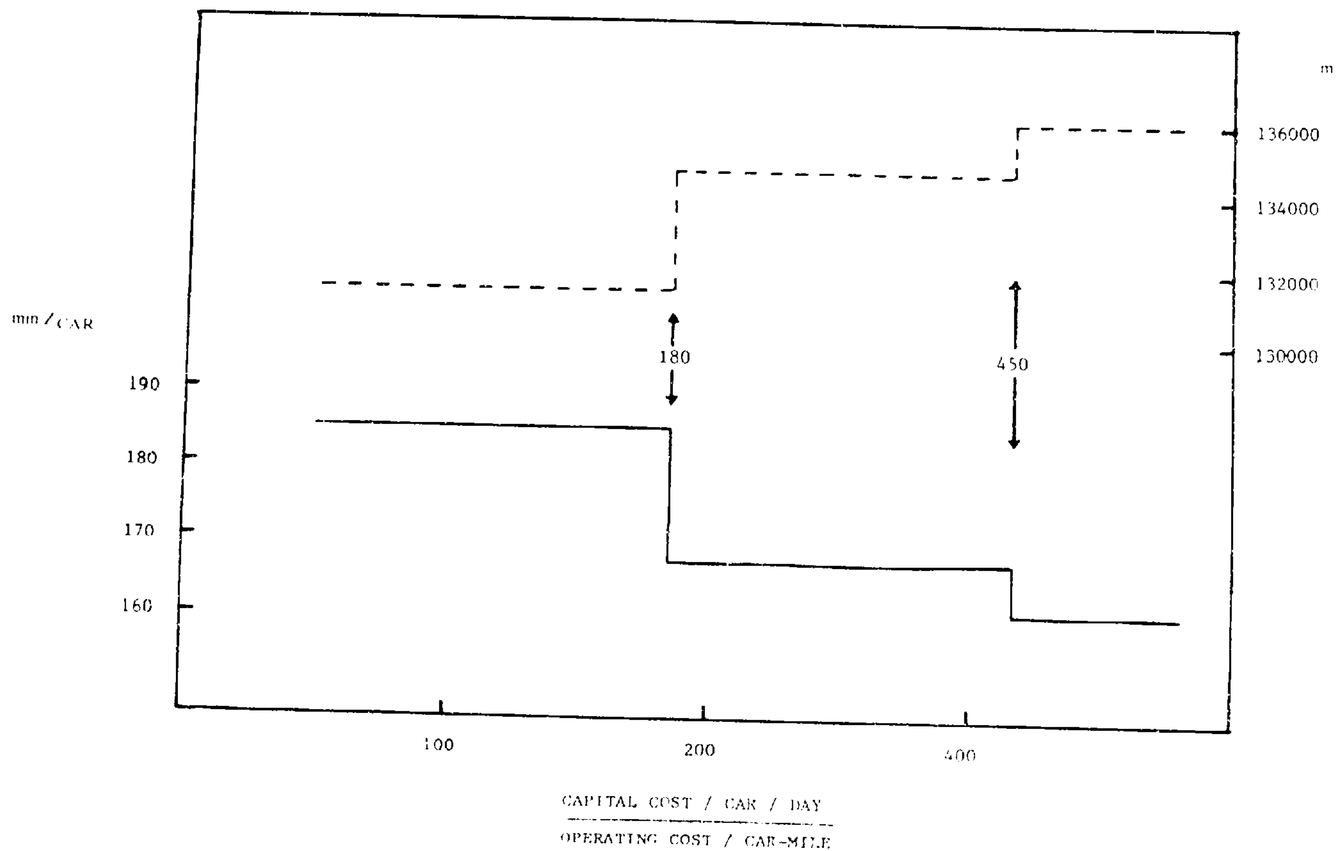

Figure 4. Cars and car-miles when total cost is minimized, as a function of the ratio of capital cost/day [illegible]

for the minimum-car-miles objective, Z_{MILE}, was also found. A maximum system load factor, Z_{LF}, was determined from Z_{MILE}, as these two objectives are inversely proportional (§2.1).

The values of the objectives at their optima for the base data are:

$$\min Z_{CAR} = 164 \text{ cars}$$
$$\min Z_{MILE} = 131388 \text{ car-miles}$$
$$\max Z_{LF} = 74.15\%$$

§4.4. *Minimizing Operating Plus Capital Cost*

Following the analysis set forth in §§2.3–2.5, the next step was to minimize total "operating" and "capital" cost of the base model. SESAME's algorithm for parametric analysis of the objective function was employed for this purpose. Part of the process was automated by use of small programs written in the SESAME command language.

The properties of an optimal solution depend upon the value of p_{CAR}/p_{MILE}, the ratio of capital cost/day to operating cost/mile. For the base data, there are three significantly different regions into which this ratio may fall:

(1) *Capital cost/day ≥ 450 × operating cost/mile.* Here capital cost dominates: in any optimal solution the number of cars is at its absolute minimum, 164. The minimum number of car-miles per day, given 164 cars, is 135978; and the system load factor (which is inversely proportional to total car-miles) is 71.65%.

(2) *450 × operating cost/mile ≥ capital cost/day ≥ 180 × operating cost/mile.* At this level the influence of capital cost declines somewhat. The number of cars in an optimal solution increases to 167; car-miles per day decline to 134628 (system load factor 72.37%).

(3) *Capital cost/day ≤ 180 × operating cost/mile.* Here operating cost dominates. In an optimal solution car-miles/day is at its absolute minimum, 131388 (system load factor 74.15%), while the number of cars in the system increases to 185.

The results are shown graphically in Figure 4. Clearly the biggest jump is at critical ratio $p_{CAR}/p_{MILE} = 180$, the round-trip distance between New York and Philadelphia. At ratios below this point, buying an extra car is economical even if it saves just one New York-Philadelphia run. At higher ratios it pays to buy a smaller fleet, running each car (on the average) more miles every day. The size of the jump—about a 10% difference in fleet size—is not surprising. Demand is heaviest along the New York-Philadelphia segment, and is highly unbalanced: northbound travel peaks in the morning, while southbound demand is highest in the

Figure 4. Cars and car-miles when total cost is minimized, as a function of the ratio of capital cost/day to operating cost/mile. The critical ratios are 180 and 450.

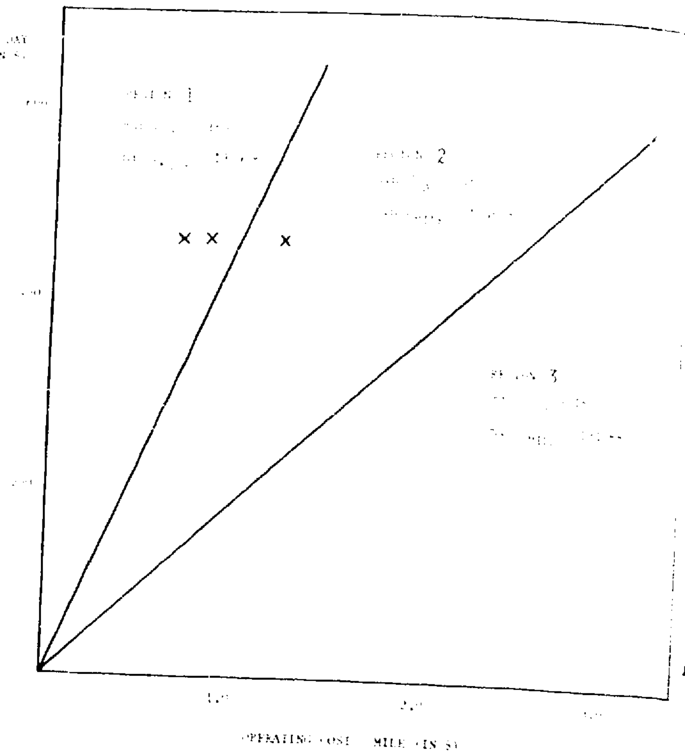

Figure 5 The three solution regions for the base run, plotted as a function of capital cost/day and operating cost/mile. Alternative projections of the actual ratio of these amounts in 1982 are indicated by X's.

afternoon. Consequently, a fair amount of deadheading can be avoided if a larger fleet is available.

The other jump, at $p_{\mathrm{CAR}}/p_{\mathrm{MILE}} = 450$, represents a point at which the cost of a car equals the cost of running it from New York to Washington and back. This is a fairly insignificant critical ratio, however, as the optimum at ratios below 450 requires only three cars more than the optimum above 450.

Several estimates of the actual $p_{\mathrm{CAR}}/p_{\mathrm{MILE}}$, derived from a PMM financial analysis [1], are plotted against the critical ratios in Figure 5. The estimates suggest that $p_{\mathrm{CAR}}/p_{\mathrm{MILE}}$ probably falls into region (1), and hence that capital cost is probably predominant. (Moreover, if the ratio is not in region (1) it appears very likely to be in region (2), where the optimal solution is not much different.)

§4.5. *Minimizing Turnaround*

Z_{TURN}, the number of times per day that the direction of a car is changed (§2.1), was also considered. Since capital cost seemed likely to predominate, a solution was found to:

$$\min Z_{TURN} \mid Z_{MILE} \mid Z_{CAR}$$

The optimal value of Z_{TURN} is not particularly revealing; but the flow of cars being turned around at New York and Philadelphia is of interest. No northbound car is ever turned around at Philadelphia, and no southbound car is ever turned at New York. Cars running north from Philadelphia are held in storage at New York mostly in the morning, when northbound travel on the Philadelphia-New York segment predominates. Cars running south from New York are held at Philadelphia mostly in the afternoon, when southbound traffic is dominant on the segment.

In effect, many cars are needed only for the Philadelphia-New York segment, to satisfy peak demand northbound in the morning and southbound in the afternoon. This suggests a revised schedule in which New York-Philadelphia shuttle trains are run at peak hours, in addition to the usual through trains.

§4.6. *Sensitivity to Demand*

Demand projections are inherently uncertain. They are based on approximate data, and their postulations are open to question. A PMM study of Northeast Corridor demands [6], for example, estimates 1982 patronage at anywhere from 11 to 23 million passengers, depending upon assumptions about costs and travel times.

It is thus essential that the model be solved for a range of demands. Fortunately, this can be done by SESAME in an especially efficient way, by taking advantage of two characteristics of the model.

First, a change in demands does not change the model's row and column structure: demands affect only the lower and upper limits on the train-size variables. Consequently, the LP matrix need be generated only once for each combination of schedule and turnaround delay. Sets of limit values are filed separately; just before the model is to be solved or analyzed, SESAME is instructed which set of limits to use with the previously-created matrix. Any different set of limits is easily substituted whenever desired.

Second, different sets of demand limits for the same model tend to be similar, and hence their optimal solutions are generally close together. As a result, it is not necessary to solve from scratch for each set of demands. An optimal basis for one demand set is a very good starting basis for iterating to an optimum for any similar set. SESAME's dual simplex algorithm is especially useful for this purpose, since changing upper and lower limits does not violate dual feasibility.

For the base run, alternative estimates of effective demands were first derived through scaling the base patronage estimates by a constant

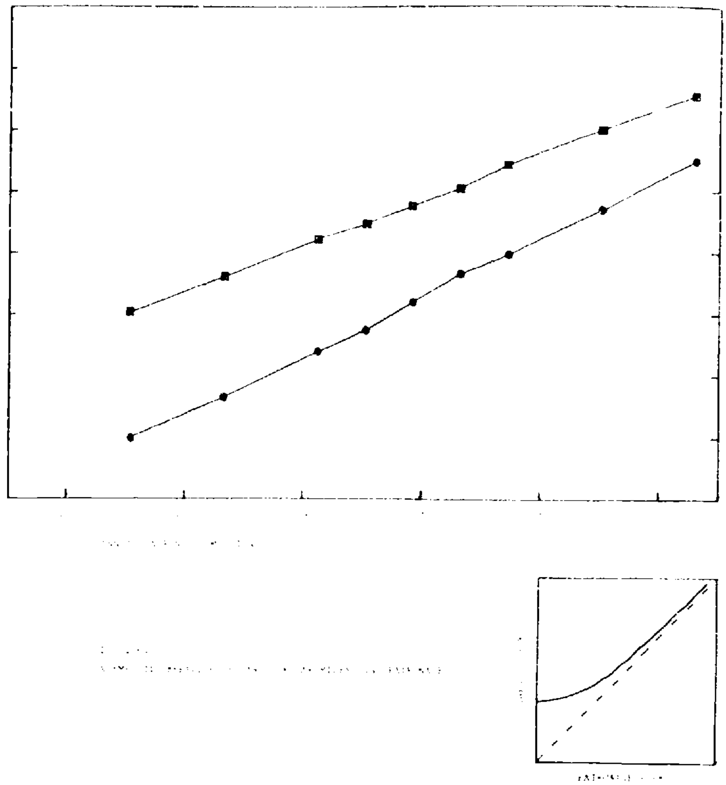

Figure 6 Min Z_{CAR} (right scale) and min Z_{MILE}/Z_{CAR} (left scale) as a function of annual patronage. The small graph shows the general forms of these functions as patronage approaches zero and infinity.

factor; then upper and lower limits were determined as before. Nine factors, ranging from .7 to 1.3, were chosen. For each, a separate set of upper and lower limits was filed by the demand program (§4.1).

An analysis of total capital and operating cost was performed, in the manner of §4.4, for each set of scaled demands. The overall pattern is the same as that for the base demands: the only large jump is at p_{CAR}/p_{MILE} = 180, where the capital cost of a car equals its operating cost from New York to Philadelphia and back. There is some variation in the minor jumps, the one at 450 (New York-Washington) sometimes omitted, and one at 462 (New York-Boston) occasionally appearing; but none of these jumps is associated with a significant change in the solution.

Figure 6 shows cars and car-miles plotted against total annual patronage for the case in which capital cost predominates. These slightly convex curves are fairly close to lines through the origin, especially within a limited range (say, 20% around the base data). Hence as a rule of thumb one may say that both the minimum fleet size, and the minimum number of car-miles that must be run with minimal fleet, are roughly proportional to total patronage:

$$\min Z_{\mathrm{CAR}} \approx 000103 \cdot (\text{total annual patronage})$$
$$\min Z_{\mathrm{MILE}} \mid Z_{\mathrm{CAR}} \approx .0086 \cdot (\text{total annual patronage})$$

(In fact, both cars and car-miles do approach proportionality to patronage as the latter goes to infinity. This is because at very high demands the problem is virtually continuous, so that any increase in total demand can be met by just increasing the size of each train in the same proportion, with rounding in negligible amounts. At fairly small demand, on the other hand, the integrality of the problem comes into play. A relatively large amount of excess capacity is run simply because demands are rounded up to the next integer, and hence the actual curve for cars or car-miles runs somewhat above the line of proportionality see small graph in Figure 6.)

Many more sophisticated sensitivity analyses are conceivable if one allows patronage between different station-pairs to vary at different rates. For example, one might use annual patronages computed under different assumptions; or one might apply different cumulative probability distributions to one set of annual patronages.

§4.7. *Locomotive Requirements*

The upper-limits rules for the base run (§3.7) insure that the number of 14-car sections that must be run to meet each train's demand is fixed: if demand is 14 cars or less, one section is run; if demand is greater than 14 but not more than 28, two sections are run; and so forth. Hence, assuming one locomotive per section, one can tell exactly how many locomotives will be required for each train in the schedule, in any feasible solution. The analysis of §1.7 is thus applicable: locomotive demands can be substituted for car demands to determine the number of locomotives required and how far they must travel.

Only 5 trains in the base run required two sections (and hence two locomotives); the remainder all required one. One-section trains were given an upper limit of two locomotives, and two-section trains an upper limit of three (for up to 21 cars) or four (for 22 28 cars). Sets of limits were computed and filed by the same demand program used for modeling cars (§4.1).

Solving the model with the techniques applied previously to car demands, it was determined that a single solution minimized both the number of locomotives required (31) and the number of locomotive-miles run (34074). Only 4 sections, all southbound, had to be run with an extra locomotive.

Sensitivity analyses analogous to those run for car demands (§4.6) were also applied to locomotives. The case at 70% of base demand requires only one locomotive for every scheduled train; hence 29 locomotives is an absolute minimum for the base schedule.

National Bureau of Economic Research
U.S. Department of Transportation

Submitted August 1976
Revised January 1977

REFERENCES

[1] Baker, H. S. and M. O. Laughlin. "Financial Analysis of the Northeast Corridor Development Project." Report no. DOT-TSC-FRA-NCD-76-3. Transportation Systems Center, U.S. Department of Transportation, June 1976.

[2] Bechtel, Inc., "Northeast Corridor High-Speed Rail Service Improvement Project Task 1. Demand Analysis." Report No. FRA-ONECD-75-1, Office of Northeast Corridor Development, Federal Railroad Administration, April 1975.

[3] *DATAMAT Reference Manual* (Third Edition). Order no. D0078. Computer Research Center for Economics and Management Science, National Bureau of Economic Research, July 1975.

[4] Fourer, Robert. "Models of Railroad Passenger-Car Requirements in the Northeast Corridor. Volume I: Formulation and Results." Report no. FRA/NECPO-76/21. Transportation Systems Center, U.S. Department of Transportation, September 1976. (Available through the National Technical Information Service, Springfield, VA 22161; or from the authors of this paper.)

[5] Fourer, Robert. "Models of Railroad Passenger-Car Requirements in the Northeast Corridor, Volume II: User's Guide." Report no. FRA/NECPO-76/22. II. Transportation Systems Center, U.S. Department of Transportation, September 1976. (Available through the National Technical Information Service, Springfield, VA 22161; or from the authors of this paper.)

[6] Prokopy, J. C. and D. E. Rinna. "Demand Projections for the Northeast Corridor Financial Analysis." Report no. DOT-TSC-FRA-NCD-76-4. Transportation Systems Center, U.S. Department of Transportation, June 1976.

[7] *SESAME Primer* (Second Edition). Order no. D0076. Computer Research Center for Economics and Management Science, National Bureau of Economic Research, January 1976.

[8] *SESAME Reference Manual*. Order no. D0077. Computer Research Center for Economics and Management Science, National Bureau of Economic Research, July 1975.