TRANSPORTATION RESEARCH

RECORD

No. 1298

Aviation

Public-Sector Aviation Issues

Graduate Research Award Papers 1989–1990

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1991

Transportation Research Record 1298
Price: $13.00

Subscriber Category
V aviation

TRB Publications Staff
Director of Publications: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Kathleen Solomon, Norman Solomon
Graphics Coordinator: Diane L. Ross
Production Coordinator: Karen S. Waugh
Office Manager: Phyllis D. Barber
Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Public sector aviation issues : graduate research award papers, 1989–1990.
p. cm.—(Transportation research record ; no. 1298)
"The papers were presented at the 70th TRB Annual Meeting in January 1991"—Foreword.
"A peer-reviewed publication of the Transportation Research Board."
ISBN 0-309-05107-X
1. Airports. 2. Aeronautics, Commercial. 3. Aeronautics–Accidents. I. National Research Council (U.S.). Transportation Research Board. II. Series: Transportation research record ; 1298.
TE7.H5 no. 1298
[TL725]
388 s—dc20
[387.7]
91-24111
CIP

Sponsorship of Transportation Research Record 1298

TRB Selection Panel for Graduate Research Award Program on Public-Sector Aviation Issues
The panel evaluates applications from graduate students and recommends candidates for awards on the basis of academic and other qualifications.
Chairman: Francis X. McKelvey, Michigan State University
John W. Fischer, Vicki L. Golich, Greig W. Harvey, Alice Herman, Sally D. Liff, Dorn C. McGrath, Jr., Dennis W. Mewshaw, Ronald W. Pulling, Ted Tedesco, Karl R. Zaeske.

FAA Liaison: James Veatch

TRB Staff
Aviation Specialist: Larry L. Jenney
Program Coordinator: E. Thomas Burnard
Senior Secretary: Marcela O. Deolalikar

The organizational units, officers, and members are as of December 31, 1990.

Transportation Research Record 1298

Contents

Foreword v

Peak-Load-Congestion Pricing of Hub Airport Operations with Endogenous Scheduling and Traffic-Flow Adjustments at Minneapolis–St. Paul Airport 1
Joseph Daniel

U.S. Cabotage Policy 14
David P. Agnew

Human Orientation and Wayfinding in Airport Passenger Terminals 25
Anthony D. Andre

Understanding the Role of Human Error in Aircraft Accidents 33
Daniel J. Berninger

Analysis of Relationship Between Financial Health and Maintenance Cost Structure of U.S. Airline Industry 43
Edward J. Rodowicz

Foreword

The papers in this Record are reports on research topics chosen by graduate students selected for awards from a nationwide competition under the fourth (1989–1990) Graduate Research Award Program on Public-Sector Aviation Issues. This program is sponsored by the Federal Aviation Administration and administered by the Transportation Research Board. Its purpose is to stimulate thought, discussion, and research by those who may become the future managers and policy makers in aviation. The papers were presented at the 70th TRB Annual Meeting in January 1991. The authors, their university affiliations, their faculty research advisors, and their TRB monitors are as follows.

Joseph Daniel, a doctoral candidate in economics at the University of Minnesota, examined price and scheduling competition among hub airlines. His faculty research advisor was Herbert Mohring, Professor of Economics at the University of Minnesota. TRB monitors were Greig W. Harvey, Eakin, Harvey, Skabardonis, Inc., and Francis X. McKelvey, Professor of Civil Engineering at Michigan State University.

David P. Agnew, a master's degree candidate in public policy at the John F. Kennedy School of Government at Harvard University, examined the current U.S. policy on aviation cabotage and potential alternatives. His Harvard faculty research advisor was Jose A. Gomez-Ibañez. TRB monitors were John W. Fischer, Congressional Research Service, and Vicki L. Golich, Pennsylvania State University.

Anthony D. Andre, a doctoral candidate in engineering psychology at the University of Illinois, researched human factors in the design of airport passenger terminal wayfinding systems. His faculty research advisor was Jefferson M. Kounce at the Institute of Aviation, Aviation Research Laboratory, University of Illinois. TRB monitors were Sally Liff, Bellomo-McGee, Inc.; Andrew C. Lemer, National Research Council Building Research Board; and Dennis W. Mewshaw, National Association of State Aviation Officials.

Daniel J. Berninger, a doctoral candidate in systems engineering at the University of Pennsylvania, sought an understanding of the role of human error in aircraft accidents. His faculty research advisor was Kenneth A. Fegley, Chair, Systems Engineering Department. TRB monitors were Richard F. Pain, TRB Transportation Safety Coordinator, and Harold C. Van Cott, National Research Council Commission on Behavioral and Social Sciences and Education.

Edward J. Rodowicz, a master's degree candidate in aviation business at Embry-Riddle Aeronautical University, analyzed the relationship between the financial and mechanical health of the major U.S. airlines. The faculty research advisor was Linda J. Block, Associate Professor of Aviation Busness at Embry-Riddle. TRB monitors were Ronald W. Pulling, Ronald W. Pulling Associates, and Stephen Godwin, TRB Special Projects Division.

Peak-Load-Congestion Pricing of Hub Airport Operations with Endogenous Scheduling and Traffic-Flow Adjustments at Minneapolis–St. Paul Airport

JOSEPH DANIEL

Hub airlines schedule banks of flights that create periodic demand peaks. During these peaks, arrival and departure rates approach or exceed airport capacity and queues develop. Under the current weight-based landing-fee structure, excessive delays result because airlines ignore the delays their operations impose on other airlines and their passengers. Congestion-based airport pricing would encourage airlines to use larger aircraft with lower service frequency and to shift their operations away from the peak, thereby reducing congestion. A model is presented of the adjustment of flight schedules and traffic flows in response to weight-based and congestion-based fee structures. With data from the Minneapolis–St. Paul airport, the model is applied to calculate equilibrium congestion fees, schedule frequencies, traffic patterns, landing and takeoff costs, airport revenues, and resource savings from peak-load-congestion fees.

Most major airports in the United States assess landing fees that are proportional to aircraft weight and independent of time of operation. The social cost of a landing or takeoff, however, consists primarily of the additional delay the operation imposes on all the aircraft and travelers using the airport at approximately the same time. These costs are essentially independent of aircraft weight and vary considerably with time of operation. Weight-based fees encourage frequent service by small aircraft during peak periods and fail to appropriately manage demand. Use of weight-based fees has led to unnecessarily high levels of congestion and delay.

FAA estimates that airport congestion delay costs the airlines and their passengers $5 billion annually in increased operating costs and travel time. A Transportation Research Board report (*1*) indicates that 21 large hub airports each experience more than 20,000 plane-hours of flight delay annually and predicts that within the next decade 39 airports will exceed that level. Delays at Chicago, Atlanta, and Denver could approach 100,000 plane-hours annually. Air traffic is expected to double by early in the next century. In the past, the congestion problem has been addressed mainly by increasing capacity. Little effort has been made to manage demand. Expansion of an existing airport or construction of a new airport, however, can cost several billion dollars. Given the high cost of increasing capacity, it would seem wise to make efficient use of airports before resorting to expansions or new airport construction.

Department of Economics, University of Minnesota, 1035 Management and Economics, 271 19th Avenue South, Minneapolis, Minn. 55455.

Economists have a standard solution to the congestion problem—use the price system to allocate scarce airport capacity and bring demand into line with short-term supply. The optimal fee equals the marginal external delay costs that an airplane imposes on other airplanes and travelers using the airport at approximately the same time. Such fees would cause the airlines to internalize the congestion externality they create, thereby encouraging cost-minimizing scheduling decisions. It has been shown that if airport capacity exhibits constant returns to scale, congestion fees will yield revenues exactly sufficient to pay for optimal long-run capacity (*2*).

The role of atomistic hub-and-spoke route networks (ones in which each route is served by a different airline) is examined as the cause of traffic peaks that exacerbate the congestion problem. Existing models of traffic flows in the transportation economics literature generally assume that the traffic is atomistically operated—each car on the highway is owned and operated independently of other cars. Previous models of airport congestion also implicitly make this assumption (*3–6*). This paper continues in that tradition by modeling an airport serving an atomistic hub network. The model has historical relevance; the Civil Aeronautics Board prevented single airlines from dominating prederegulation hub airports. In the postderegulation environment, however, many airports are dominated by a single airline that accounts for more than half the airport's operations. In scheduling airport arrivals and departures, a profit-maximizing dominant airline would internalize the delay that one of its aircraft imposes on another. An atomistic airline, on the other hand, would take delay as parametric. Dominant airline operations, therefore, impose less external congestion than atomistic airline operations. If a constrained-optimal fee structure is not to discriminate between dominant and atomistic airlines, it must be a compromise between a fee equal to the external congestion imposed by the dominant airline and a fee equal to the external congestion imposed by atomistic airlines. The model presented here is a preliminary step toward modeling the more difficult and realistic case of an airport with a dominant hub airline and an atomistic fringe.

The model of atomistic hub scheduling [an adaptation of Mohring's (*7*) model of direct-service bus scheduling to an air-service network] is integrated with a model of traffic flows through an airport bottleneck. [The bottleneck model draws on the work of Vickrey (*8*).] In the scheduling model, the

airlines choose plane size and service frequency to minimize the sum of their own and their passengers' costs. In the bottleneck model, each airline chooses its flight's arrival time at the airport to minimize the sum of its queuing costs and layover costs. Traffic patterns at the bottleneck are determined endogenously. Previous congestion-based pricing models either treat traffic patterns as exogenous or assume zero intertemporal cross-elasticity of demand (i.e., demand is a function of the current period's price alone, and no shifting of demand to other periods occurs). The scheduling model captures the effects of intertemporal demand shifts and predicts changes in aircraft size and schedule frequency in response to congestion fees. The bottleneck model captures the effects of peak-spreading in response to the fees. Together, the models enable calculation of equilibrium fees and traffic patterns. Finally, the models introduce layover delay into the analysis of airport congestion and show that the hub airlines face a trade-off between congestion and layover delays. Gains from reduced congestion are partially offset by increased layover delays for the hub airline and its passengers.

The model is applied to traffic data from the tower logs of Minneapolis–St. Paul (MSP) airport. The data set gives the time of every operation at the airport during the first week of May 1990. Because no reliable data on actual queue lengths exist, a queuing model is used to infer the expected queue lengths given the actual arrival rates. A discrete time version of a bottleneck model is then fit to the arrival and queuing data. The model estimates the equilibrium congestion fees, traffic patterns, queuing delays, layover costs, airport's revenues, real resource savings from reduced congestion, and changes in the airlines' schedule frequency and aircraft size.

ECONOMICS OF HUB-AND-SPOKE NETWORKS

Airlines face significant economies of scale and scope in providing their services. The fundamental scale economy arises at the conveyance level. The cost of an airplane and its crew increases less than proportionately with increases in plane size, so cost per seat-mile decreases as the number of seats on a plane increases. Economies of scope result from joint production of trips between many different city pairs. Hub-and-spoke route networks enable airlines to fly passengers with the same origin but different destinations on a single flight to the hub. Similarly, passengers with the same destination but different origins can be combined on a single flight from the hub. Because a given number of passengers can be served more frequently with fewer flights on larger planes, the cost to an airline of conveying each passenger in a hub-and-spoke network is less than that by direct service. This reduction in cost is partially offset, however, by the increased circuity of routes and by the introduction of layover delay.

Airline passengers play both a consuming and a producing role in air travel. They combine airline services with their own time inputs to produce trips. The passengers' time inputs consist of time in transit and schedule, congestion, and layover delay. (Time in transit is the time passengers actually spend on the plane, exclusive of the time spent because of congestion delay. Schedule delay is the difference in time between the passengers' preferred departure time and the closest scheduled departure time. Congestion delay is the time spent sitting at the gate or on the taxiway waiting for clearance to take off plus the time spent circling in the air waiting for clearance to land. Layover delay is the time that connecting passengers spend at a hub airport between their first flight's arrival and their second flight's departure.)

The passengers' time inputs required to complete a trip depend on the rate at which the airline provides its services. If the airline increases its schedule frequency, its passengers' schedule delays decrease, on average, but its cost per seat increases on the smaller airplanes. Similarly, if the airlines schedule more frequent landings and takeoffs at the hub, the average layover delay will decrease but the average congestion delay will increase. The full cost of a flight is the sum of the passengers' time costs and the airline's operating costs.

Given equal elasticities of demand across groups of passengers with different time values, a profit-maximizing airline will choose the socially optimal service frequency—that which minimizes the sum of its own and its passengers' costs. An intuitive explanation of this fact is as follows. Suppose that, by increasing its service frequency, the airline could reduce its passengers' costs by more than the increase in its own costs. Then a simultaneous increase in frequency and price by an amount equal to the reduction in its passengers' costs would not change the full cost of service to passengers and, hence, the rate at which they travel. The simultaneous change in frequency and price, therefore, would increase its profits. The airline's profit-maximizing scheduling departs from socially optimal scheduling, however, insofar as uncompensated congestion externalities are imposed on other airlines and their passengers.

If the airport had unlimited capacity and flights were always on schedule, airlines would schedule all their planes to arrive at the hub airport at the same instant. There would be a brief interchange period while passengers transferred to their connecting flights. All planes would then depart at the same instant. Unfortunately, flights deviate randomly from their expected arrival times, and airports can accommodate only a limited number of landings or takeoffs in a given period of time. As the volume of traffic increases, landing and takeoff queues develop. If the airlines scheduled all flights to arrive precisely at the beginning of the interchange period, the flights that arrived close to their scheduled times would face long delays because of the high traffic volume near the beginning of the interchange. Some of the flights would arrive early and experience small queues but have to wait longer until the interchange period began. Others would arrive after the interchange began and these passengers would either delay or miss their connections. Each airline has an incentive to adjust its flight's arrival schedule to minimize the sum of its expected costs of arriving early, waiting in the queue, and arriving late. An equilibrium traffic flow results when the arrival schedules and expected queue lengths are such that no airline can reduce its flight's expected costs by changing its schedule. Assuming that all aircraft have identical costs, the equilibrium expected cost of scheduling an arrival at all periods must be identical—otherwise, flights scheduled at high-cost periods could reduce costs by moving to low-cost periods. These observations about airline scheduling will be formalized in a model in the next two sections.

MODEL OF ATOMISTIC HUB-AND-SPOKE ROUTE NETWORK SCHEDULING

Atomistic airlines operate a hub-and-spoke route network that serves a large hub city located at the center of a circular market and n rim cities located at equal intervals around the circle's circumference. It takes h hours to fly from hub to rim. The cost per hour of flying an airplane with P passengers is aP^b, where a and b are cost parameters. A bank of $2*n$ flights can convey travelers between all cities in the route network. The cost to the network of flying all the airplanes in one bank from the rim cities to the hub and back again is $\$2naP^bh$.

Travelers in each direction on hub-to-rim (or rim-to-hub) and rim-to-rim routes have desired departure times that are uniformly distributed at the rate of D_h and D_r per hour, respectively. Random variations in demand are ignored. On each airplane, the hub combines passengers on a hub-to-rim (or rim-to-hub) route with passengers on $(n - 1)$ rim-to-rim routes. The effective demand density for each flight in a bank is therefore $D = D_h + D_r(n - 1)$ travelers per hour. If the airplanes all have P passengers, there will be D/P banks per hour and the interval between banks will be P/D hours. On average, a passenger will experience a schedule delay equal to $1/\alpha$ of the interval between banks, where $1/\alpha$ is in the interval $1/4 \leq 1/\alpha \leq 1/2$. If all passengers choose the flight closest to their desired departure time, then $1/\alpha = 1/4$, and if all passengers pick the flight that will arrive at a destination before a given time, then $1/\alpha = 1/2$. Let $r = D_r(n - 1)/D$ be the fraction of indirect passengers on each flight. In each bank there are nP passengers originating at the rim cities and $(1 - r)nP$ passengers originating at the hub. If the average passengers pays $\$v_0$ per hour to avoid schedule delay, then the total cost of all passengers' schedule delay is $[(2 - r)nPv_0]P/(\alpha D)$ per bank.

There are rnP indirect passengers in each bank who have to spend $2h$ hours in transit and $2(1 - r)nP$ direct passengers in each bank who spend only h hours in transit. If the average passenger pays $\$v_1$ per hour to shorten the length of a trip, then the total cost of all passengers' time in transit is $2nPhv_1$ per bank.

The airlines and their passengers also experience costs caused by congestion and layovers at the hub airport. Let C_a and C_d denote the sum of congestion and layover costs experienced by an airplane and its passengers on landing at and taking off from the hub airport. Let C_i denote the cost of time spent during the interchange period. C_a and C_d are determined below in the bottleneck model and C_i is a parameter. The total cost of congestion and layover is $n(C_a + C_d + C_i)$.

The sum of the passengers' and airlines' cost of operating the hub-and-spoke network for an hour is

$$\frac{D}{P}\{2naP^bh + [(2 - r)nPv_0]\frac{P}{\alpha D} + 2nPhv_1 + n(C_a + C_d + C_i)\} \qquad (1)$$

A profit-maximizing or full-cost-minimizing hub network would choose P to minimize expression 1.

To simplify the previous scheduling problem, all the rim cities in the hub-and-spoke network are assumed to be equidistant from the hub, and all have identical demand densities. As a result, the optimal plane sizes and service frequencies are identical for all the hub's routes. This greatly simplifies the mathematics of the model. In reality, however, the airline faces different demand densities and different flight distances, so its bank interval cannot simultaneously optimize service on each route. Some cities with low demand density may not be served in every bank. Increasing airport fees may cause some cities to be served in fewer banks, resulting in lower congestion and higher schedule delay. The model does not capture this effect and therefore tends to underestimate the reduction in congestion costs that would result from congestion pricing.

DISCRETE TIME BOTTLENECK MODEL WITH STOCHASTIC ARRIVALS

Let S_t denote the number of flights scheduled to arrive during period t and p_j the probability that a flight will arrive j periods from its scheduled time. If A_t denotes the expected number of arrivals during period t, then

$$A_t = \sum_j p_j S_{t-j}$$

Let $Q_t = Q(Q_{t-1}, A_{t-1})$ denote the expected length of the queue at the beginning of period t, a function of the expected queue at the beginning of the previous period and the expected number of arrivals during the previous period. Let k be the number of airport operations that can be performed in one period. The expected delay, D_t, experienced by planes arriving in period t is $D_t = (Q_t + Q_{t+1})/(2k)$. Let C_q be the amount that an airline and its passengers would be willing to pay to avoid one period of queuing delay. Similarly, let C_e and C_l be the amount that they would be willing to pay to avoid a period of earliness or lateness. The beginning of the interchange period occurs at period T_0. The expected cost of a landing scheduled for period t is

$$C_t = C_q \sum_j p_j D_{t+j} + C_e \sum_{j \leq T_0 - t - D_{t+j}} p_j[T_0 - (t + j + D_{t+j})] + C_l \sum_{j > T_0 - t - D_{t+j}} p_j[(t + j + D_{t+j}) - T_0]$$

A no-fee atomistic bottleneck equilibrium is a sequence $\{S_t, A_t, Q_t, D_t, C_t\}$, $T = (-\infty. . .\infty)$, that, given p, C_q, C_e, C_l, k, and a queuing process, $Q(\cdot)$, satisfies

$$A_t = \sum_j p_j S_{t-j} \qquad (2a)$$

$$Q_t = Q(Q_{t-1}, A_{t-1}) \qquad (2b)$$

$$D_t = (Q_t + Q_{t+1})/(2k) \qquad (2c)$$

$$C_t = C_q \sum_j p_j D_{t+j} + C_e \sum_{j \leq T_0 - t - D_{t+j}} p_j [T_0 - (t + j + D_{t+j})] + C_l \sum_{j > T_0 - t - D_{t+j}} p_j [(t + j + D_{t+j}) - T_0] \qquad (2d)$$

and

$$(S_t > 0) \rightarrow (C_t \leq C_{t'}) \quad \text{for all } t \text{ and } t' \tag{2e}$$

The optimal congestion fee, F_t, is equal to the marginal external congestion imposed by a landing in period t. Define the sequence A' such that $A'_i = A_i$ for $i<>t$ and $A'_i = A_i + \varepsilon$ for $i = t$. Define the sequence Q' such that $Q'_i = Q(Q'_{i-1}, A'_{i-1})$ and the sequence D' such that $D'_i = (Q'_i + Q'_{i+1})/(2k)$. The delay cost incurred by a flight that actually arrives at period t is

$$C_t = C_q D_{t+j} + C_e\,[T_0 - (t + D_{t+j})] \quad \text{if } (t \leq T_0 - D_t)$$

and

$$C_t = C_q D_{t+j} + C_l[(t + D_{t+j}) - T_0] \quad \text{if } (t > T_0 - D_t)$$

Define the sequence C' such that

$$C'_t = C_q D'_{t+j} + C_e\,[T_0 - (t + D'_{t+j})] \quad \text{if } (t \leq T_0 - D'_t)$$

and

$$C'_t = C_q D'_{t+j} + C_l[(t + D'_{t+j}) - T_0] \quad \text{if } (t > T_0 - D'_t)$$

Now the optimal congestion fee can be written as

$$F_t = \sum_{t \leq i} (A_i C'_i - A_i C_i)/\varepsilon$$

A congestion-fee atomistic bottleneck equilibrium is a sequence $\{S_t, A_t, Q_t, D_t, C_t\}$, $t = (-\infty. . .\infty)$, that, given p, C_q, C_e, C_l, k, and $Q(\cdot)$, satisfies Equations 2a, 2b, 2c, 2e, and

$$C_t = C_q \sum_j p_j D_{t+j} + C_e \sum_{j \leq T_0 - t - D_{t+j}} p_j[T_0 - (t + j + D_{t+j})] + C_l \sum_{j > T_0 - t - D_{t+j}} p_j[(t + j + D_{t+j}) - T_0] + \sum_j p_j F_{t+j} \tag{2d'}$$

Identical equilibriums can be defined for takeoff schedules, except that p, C_q, C_e, C_l, k, and $Q(\cdot)$ may take different values. There is much less randomness in the departure than the arrival process, so p has a different distribution. Similarly, C_q changes because the cost of being in the takeoff queue is lower than that of being in the landing queue. The relative values of C_e and C_l in the takeoff schedule problem are the reverse of those in the landing problem. As stated previously, the equilibrium values of the expected cost of a scheduled landing (takeoff) are identical for all t in which arrivals (departures) are scheduled. The equilibrium values of C_t for arrivals and departures are the C_a and C_d that appear in the scheduling model.

Two assumptions made previously greatly simplify the bottleneck model—that all flights have identical cost parameters, C_q, C_e, and C_l, and that all flights are part of the hub route network with the same desired arrival and departure times. As a consequence of these assumptions, the model ignores cost savings resulting from a relatively greater incentive for small flights and nonhub flights to shift their arrival and departure times away from the peaks. More will be said about this point subsequently. Much of a typical airport's traffic, however, is part of its hub network. Code-sharing airlines cooperate with the dominant hub airline in scheduling and marketing their flights—they clearly are part of the network. Some unaffiliated regional-carrier flights and general-aviation flights transfer passengers to the hub flights: they, too, prefer to operate close to the hub's interchange period. Although nonhub commercial carriers do not carry many passengers who want to make interline transfers, these flights are often scheduled for times close to the hub's flights to match the hub's service times and because the hub's banks occur at popular times of the day for travel. The nonhub carriers must have a substantial preference for operating at the same time as the hub's bank, because they are willing to incur the substantial delays associated with it. Assuming that they prefer to land at the interchange times exaggerates the costs of shifting these flights off the peak and tends to underestimate the benefits of congestion pricing but is probably not too unrealistic.

DATA AND EMPIRICAL SPECIFICATION OF THE MODEL

To implement the model requires estimation of $Q(\cdot)$, p, C_q, C_e, C_l, and k. Landing and takeoff data were gathered from the tower logs of MSP airport for all operations on May 1, 2, 3, 4, and 8, 1990. The data show the flight number, aircraft type, destination (if a departure), and time that the aircraft contacted the tower to join the landing or takeoff queues. Figure 1 shows the average number of arrivals and departures that occurred during each 10-min interval on those days. The different shading of the bars indicates how many of the operations were attributable to Northwest Airlines (NWA); its code-affiliates, Express (NWX) and Mesaba (MES); other national carriers (OTH); general-aviation (GA); or air freight, military, and independent regionals (MISC). The importance of Northwest Airlines' (i.e., the hub's) banks in creating the peak demand periods is evident from the graphs. It is also clear, as discussed earlier, that large numbers of other aircraft are willing to incur delays and operate during the bank periods. This implies that their preferred schedule time is during the bank. The double vertical lines indicate the scheduled interchange periods, which are generally 30 min long. An interesting feature is that many departures overlap the subsequent bank's arrivals, even when minor adjustments would seem to avoid the overlap. This suggests that a mix of landings and takeoffs may require little more time than would be required for the landings alone. Note also that arrival peaks are lower and less steep than departure peaks—a fact attributable to the greater randomness in the arrival process.

Table 1 summarizes the peak demand periods at MSP. Northwest operates nine banks of arrivals and departures in about 17 hr each day. The average interval between banks is approximately 1.9 hr. Northwest's banks alternate service between east-to-west routes and west-to-east routes, so that a given route in a given direction is generally served on every other bank. The highest-density routes, however, may be served in each direction on virtually every bank, whereas the lowest-density routes may be served only once a day. The largest banks interchange passengers between nearly every city in the

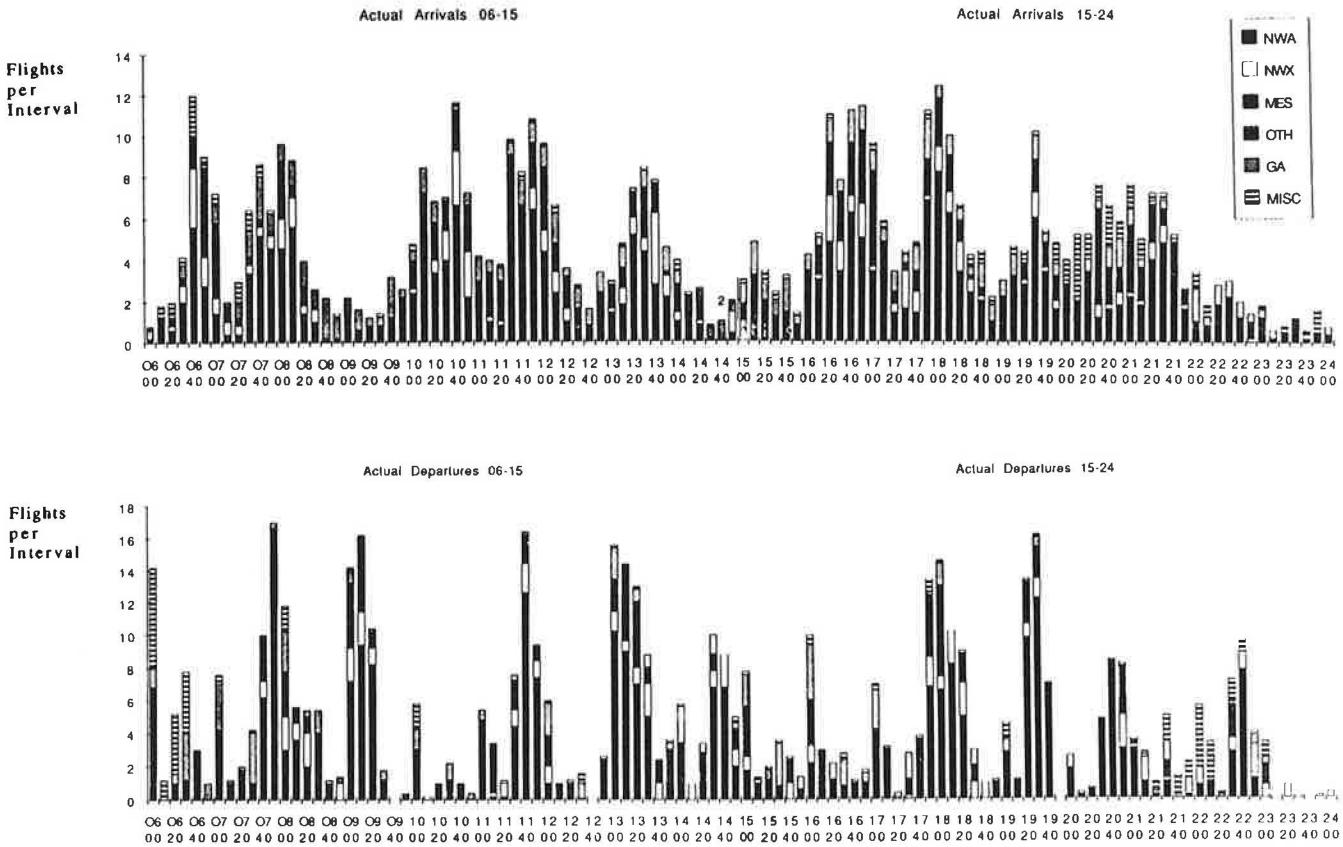


FIGURE 1 Arrivals and departures by 10-min intervals.

hub-and-spoke network, whereas the small banks serve only the denser routes.

Unfortunately, no reliable data exist about the lengths of the landing and takeoff queues. It is necessary, therefore, to use a queuing model to infer the lengths of the queues given airport capacity, the arrival and departure process, and several simplifying assumptions about the queuing discipline. Previous studies of airport queues have assumed a single queue operated on a first-come, first-served basis with no distinction between landings and takeoffs. It is more convenient here to assume a polar-opposite queuing discipline—two independent queues, each operated on a first-come, first-served basis with no interaction between landings and takeoffs. The truth probably lies somewhere between these poles. Using hourly

TABLE 1 TIMING AND SIZE OF FLIGHT BANKS AT MSP

Time	Type	NWA	NWX	MES	OTH	GA	MISC	TOTAL
06:00-07:00	Arrivals	15	7	8	4	7	0	41
07:40-08:30	Departures	28	7	6	11	9	1	62
07:30-08:40	Arrivals	27	6	2	4	10	1	50
09:00-09:20	Departures	26	5	6	4	7	0	48
09:40-11:00	Arrivals	28	8	5	11	7	3	62
11:30-12:00	Departures	26	6	2	7	4	0	45
11:40-12:20	Arrivals	29	4	6	9	9	0	57
12:50-13:30	Departures	33	6	9	6	3	0	57
13:00-14:00	Arrivals	20	7	2	12	4	0	45
14:30-15:00	Departures	19	6	2	7	2	0	36
16:00-17:20	Arrivals	34	7	9	15	16	2	83
17:40-18:20	Departures	30	9	7	9	4	1	60
17:40-18:40	Arrivals	32	6	5	4	9	3	59
19:20-19:40	Departures	30	2	5	3	1	0	41
19:00-20:10	Arrivals	17	2	2	9	10	8	48
20:30-21:10	Departures	19	4	3	3	3	1	33
20:20-22:00	Arrivals	28	4	5	10	6	9	62
22:30-22:50	Departures	12	5	3	1	2	4	27

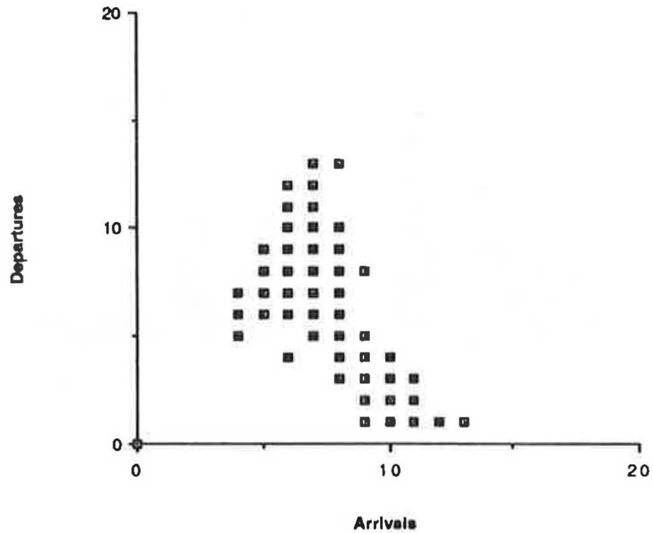


FIGURE 2 Landing and takeoff capacity for 10-min intervals.

arrival, departure, and delay data, Morrison and Winston (*9*) found that landings have about twice the effect of takeoffs on delays experienced by an arriving airplane. Similarly, takeoffs have about twice the effect of landings on delays experienced by a departing airplane. Figure 2 shows the number of landings and takeoffs performed at MSP during 10-min intervals on the afternoon of April 24, 1990. The graph of the landing and takeoff possibilities frontier indicates some interdependence between landing and takeoffs, but it is not linear as implied by the single-queue model. In the absence of adequate data to calibrate a more complicated queuing model, the assumption of two independent queues is adopted.

Arrivals and departures are assumed to be Poisson distributed, with time-varying Poisson parameters equal to the average arrival and departure rates for each 10-min interval as given in Figure 1. The expected queue lengths are calculated for the given arrival and departure distributions using Omosigho and Worthington's (*10*) discrete time queuing model for single-server queues with inhomogeneous arrival rates and discrete service time distributions. The model uses the probability distribution on queue lengths as a state vector and uses recurrence relations to calculate the state vector for each service interval given the time-varying arrival rate distributions. Because constant capacity is assumed, the model is similar to a Markov transition model, with the probability distribution on queue sizes as the state and transition matrices for each service interval, which change every 10 min to reflect the current distribution of arrivals. The calculations are based on an assumed airport capacity, k, of nine landings and nine takeoffs per 10-min interval.

Figure 3 shows the arrival and departure rates and the resulting expected queue lengths at the beginning of each 10-min interval. The arrival banks tend to peak and plateau at between 10 and 12 arrivals per 10-min interval. Arrival queues develop only during the hub banks and return to nearly zero as the bank ends. Their peaks do not plateau, and range in height from about 4 to 10 aircraft, depending on the size of the bank. The arrival rates and queue sizes increase and decrease at roughly the same rates across banks with different numbers of flights. The departure rates and queue sizes are relatively more highly peaked, and increase and decrease more quickly. The rates of departure peak at between 10 and 18 flights per interval, and the queue sizes achieve similar magnitude. Departure queues are significant only during the hub banks.

Although the queuing model is necessary for initial estimation of expected queue lengths, it is too computationally cumbersome to use in the bottleneck model. A regression model, however, can be fit to the queue estimates and the arrival and departure data. The regression equation is

$$Q_t = \alpha + \beta Q_{t-1} + \gamma A_{t-1} + \delta Q_{t-1}A_{t-1} + \varepsilon Q_{t-1}^2 + \zeta A_{t-1}^2$$

The regression estimates (with t-values in parentheses) are

$$\begin{aligned} Q_t = {} & -0.046 + 0.07Q_{t-1} + 0.021A_{t-1} + 0.045Q_{t-1}A_{t-1} \\ & (-1.012) \quad (4.093) \quad (1.363) \quad (33.739) \\ & + 0.027Q_{t-1}^2 + 0.028A_{t-1}^2 \qquad R^2 = .995 \\ & (26.384) \quad (23.676) \end{aligned}$$

Figure 4 compares the evolution of the queuing systems as estimated by the queuing model and the regression model. The simple regression model appears to describe the evolution of the queues remarkably well.

To estimate the probabilities that the actual arrival and departure times deviate from the scheduled time by j periods (i.e., p_j), the actual arrival and departure times reported in the tower log were matched with the flights' scheduled times reported in the *Official Airline Guide* (*11*). The p_j histogram for arrivals is shown in Figure 5. The histogram indicates the probability that a flight scheduled to arrive at the gate at period $t = 0$ actually arrives at the queue during periods $t = -5$ to 2, where each period is 10 min long. For example, 5 percent of all flights arrive at the landing queue during the 10-min interval centered on the time they are scheduled to arrive at the gate. On average, flights arrive at the queue 18 min before their scheduled arrival time at the gate, to allow ample time to land and taxi to the gate. For departing flights, 97 percent of the times reported in the tower logs were within the same 10-min interval as their scheduled departure times. The tower log reports a few deviations from scheduled departure times, but no other time interval accounted for as much as 1 percent of the deviations. Frequent fliers will be excused for suspecting that the logs do not tell the whole story, but in the absence of better information, all flights are assumed to depart at their scheduled time (i.e., $p_0 = 1$).

The cost parameters C_q, C_e, and C_l can be estimated from observations of expected queue lengths, expected early-time deviation, and expected late-time deviation. Manipulating the equilibrium cost equation yields

$$\sum_j p_j D_{t+j} = \frac{C^*}{C_q} - \frac{C_e}{C_q} \sum_{j \leq T_0 - t - D_{t+j}} p_j[T_0 - (t + j + D_{t+j})] + \frac{C_l}{C_q} \sum_{j > T_0 - t - D_{t+j}} p_j[(t + j + D_{t+j}) - T_0]$$

It follows that least squares estimates of C_e/C_q and C_l/C_q can be obtained by regressing observations of expected delay

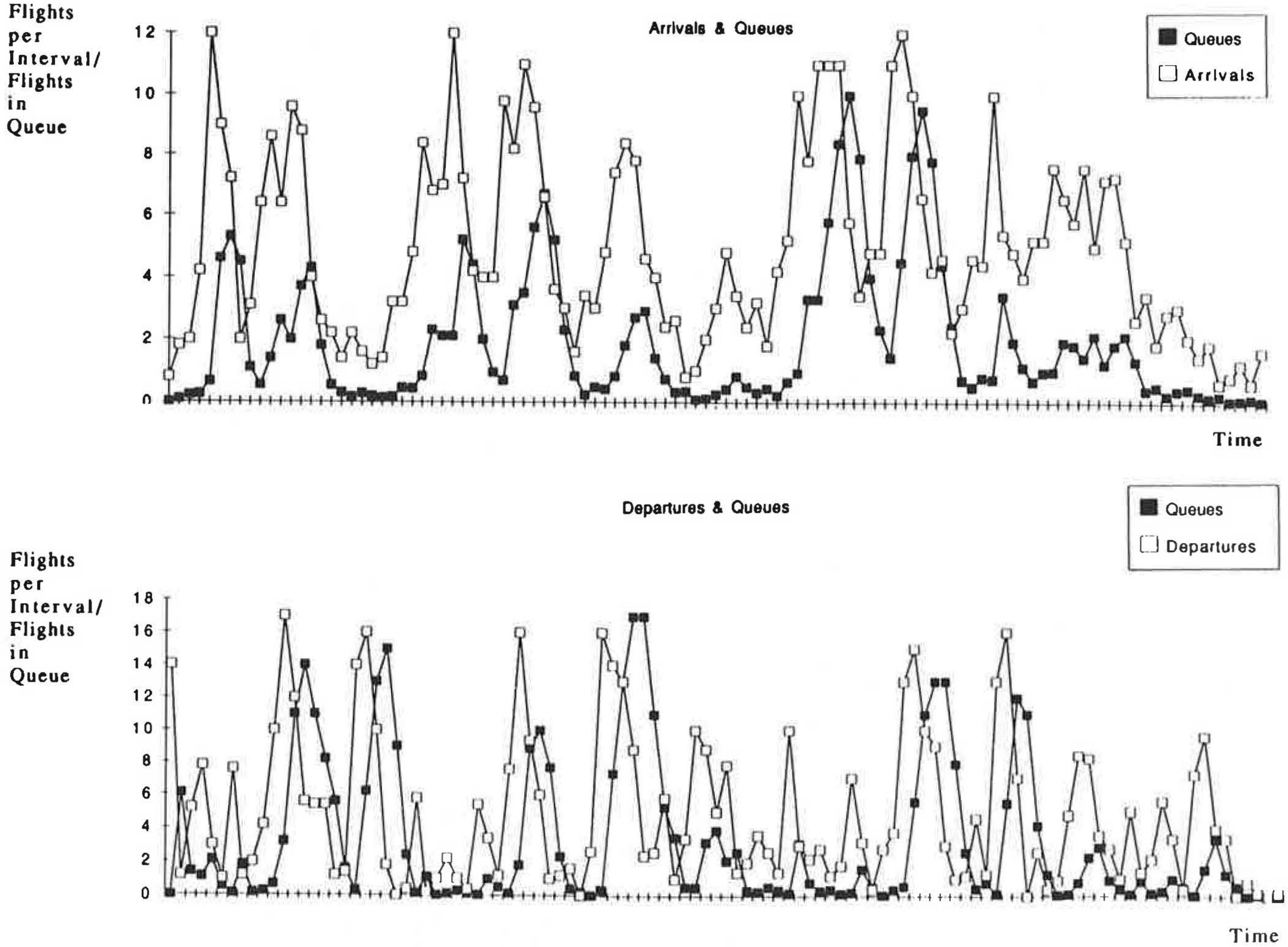


FIGURE 3 Arrivals, departures, and queue lengths.

$\Sigma_j p_j D_{t+j}$ on observations of expected early time

$$\sum_{j \le T_0 - t - D_{t+j}} p_j \left[T_0 - (t + j + D_{t+j})\right]$$

and expected late time

$$\sum_{j > T_0 - t - D_{t+j}} p_j \left[(t + j + D_{t+j}) - T_0\right]$$

Only time periods during which arrivals or departures were actually scheduled can be used in the regression. Because C^* may vary for each bank, a dummy variable must be entered in the regression for each bank (except the first) from which data are used. The regression constant plus the relevant dummy coefficient can be interpreted as an estimate of C^*/C_q for that bank. Using data from the six largest arrival banks, the estimates for the arrival costs and their t-values are

$$\begin{aligned}\text{Exp delay} = {} & \underset{(6.317)}{0.673} + \underset{(0.872)}{0.037B_2} + \underset{(0.81)}{0.034B_3} + \underset{(6.416)}{0.282B_4} + \underset{(9.239)}{0.392B_5} + \underset{(10.831)}{0.513B_6} \\ & + \underset{(5.405)}{0.134\,(\text{exp early time})} + \underset{(1.112)}{0.932\,(\text{exp late time})} \qquad R^2 = .961\end{aligned}$$

Because actual departure periods deviate practically not at all from scheduled departure periods and there are virtually no departures during the interchange periods, expected early time is omitted in the departure cost regression, and the early-time cost is assumed to be sufficiently large that no flights are scheduled to depart during the interchange. Using data from the seven largest departure banks, the estimates for the departure costs are

$$\begin{aligned}\text{Exp delay} = {} & \underset{(5.425)}{0.826} + \underset{(0.695)}{0.114B_2} - \underset{(-1.731)}{0.284B_3} + \underset{(1.874)}{0.308B_4} - \underset{(-4.311)}{0.724B_5} + \underset{(0.051)}{0.008B_6} \\ & - \underset{(-0.8)}{0.131B_7} + \underset{(3.737)}{0.169\,(\text{exp late time})} \qquad R^2 = .868\end{aligned}$$

The high R^2s in the regressions suggest that the airlines do trade off expected queuing delays against expected early and late times in accordance with the model.

Although the bottleneck model depends only on the ratios C_e/C_q and C_l/C_q, the scheduling model requires assigning some monetary values to the parameters. Suppose that 10 min of queuing delay on landing costs the airlines and their passengers \$350 per flight. Because early time on arrival is identical

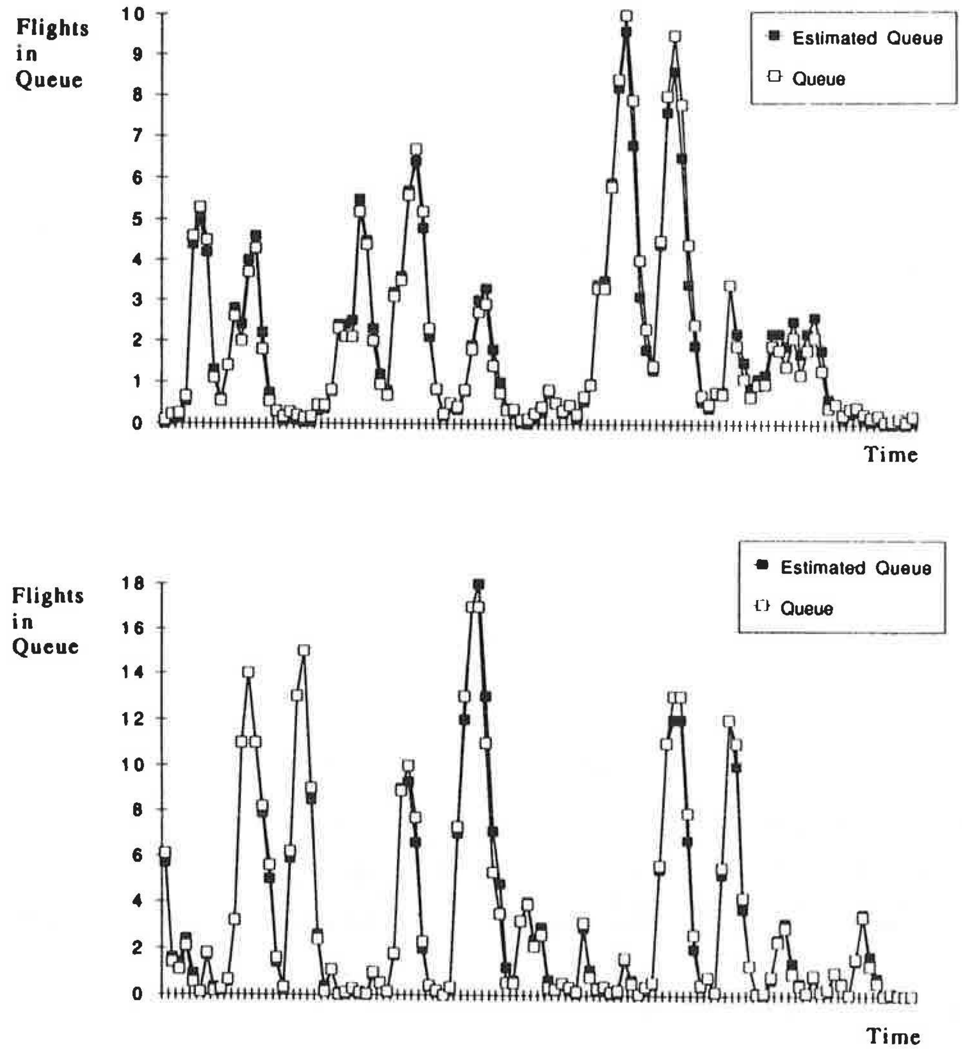


FIGURE 4 Comparison of queues in regression model and queuing model.

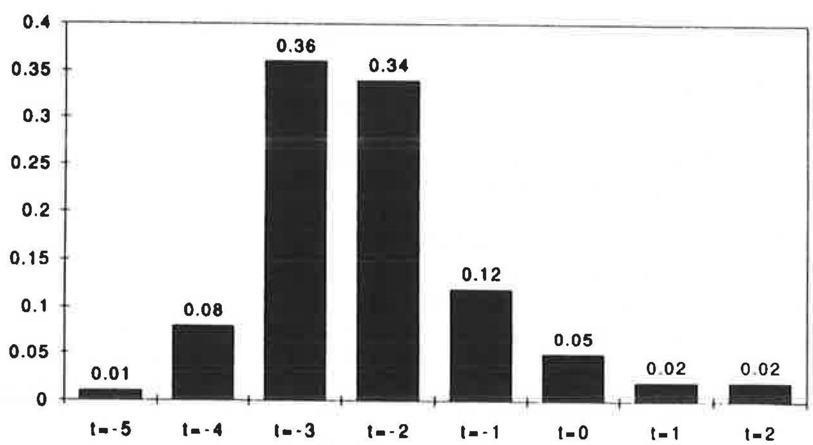


FIGURE 5 Deviations of actual arrival times from scheduled arrival times.

to late time on departure, it follows that $C_q = 350$, $C_e = 46.9$, and $C_l = 326.2$ on arrival and that $C_q = 277.5$ and $C_l = 46.9$ on departure.

Having estimated a queuing process, $Q(\cdot)$, the probabilities, p_j, and the cost parameters, C_q, C_e, and C_l, the equilibrium S, A, Q, D, C, and F can be determined. The solution technique takes an initial sequence, S, and iteratively moves incrementally smaller numbers of scheduled arrivals from the highest-cost period to the lowest-cost period. After each iteration's change in S, the sequences A, Q, D, F, and C are recalculated. The algorithm quits when the expected cost of a landing (takeoff) in any period during which flights are scheduled converges to within 0.1 percent of all other such periods' costs. The system converges to the same equilibrium from widely differing initial sequences.

RESULTS

Figure 6 compares the equilibrium arrival and departure rates in the no-fee and congestion-fee equilibriums for hypothetical banks of 40, 45, 50, 55, and 60 flights. The hypothetical no-fee arrival banks look quite similar to their similar-sized counterparts in Figure 3. In both, the banks begin to plateau at about 10 operations per 10 min, with larger banks spreading operations away from the interchange periods. As expected, the congestion-fee equilibrium arrival banks are more spread out than the no-fee banks. Their arrival rates peak more slowly and reach only 80 percent as high.

As with the departure banks of Figure 3, the hypothetical no-fee departure banks of Figure 6 have much steeper and higher peaks than their corresponding arrival banks. In contrast to Figure 3, however, they are much higher and oscillate between high departure rates and no departures. In the absence of a fee, many flights attempt to leave immediately after the interchange period. A large queue develops, and no new flights join the departure queue until it diminishes. When the queue is short enough, there is another rush to leave and the process is repeated. These oscillations appear to dampen as the number of flights in a bank increases. The hypothetical congestion-fee banks, on the other hand, have quite modest

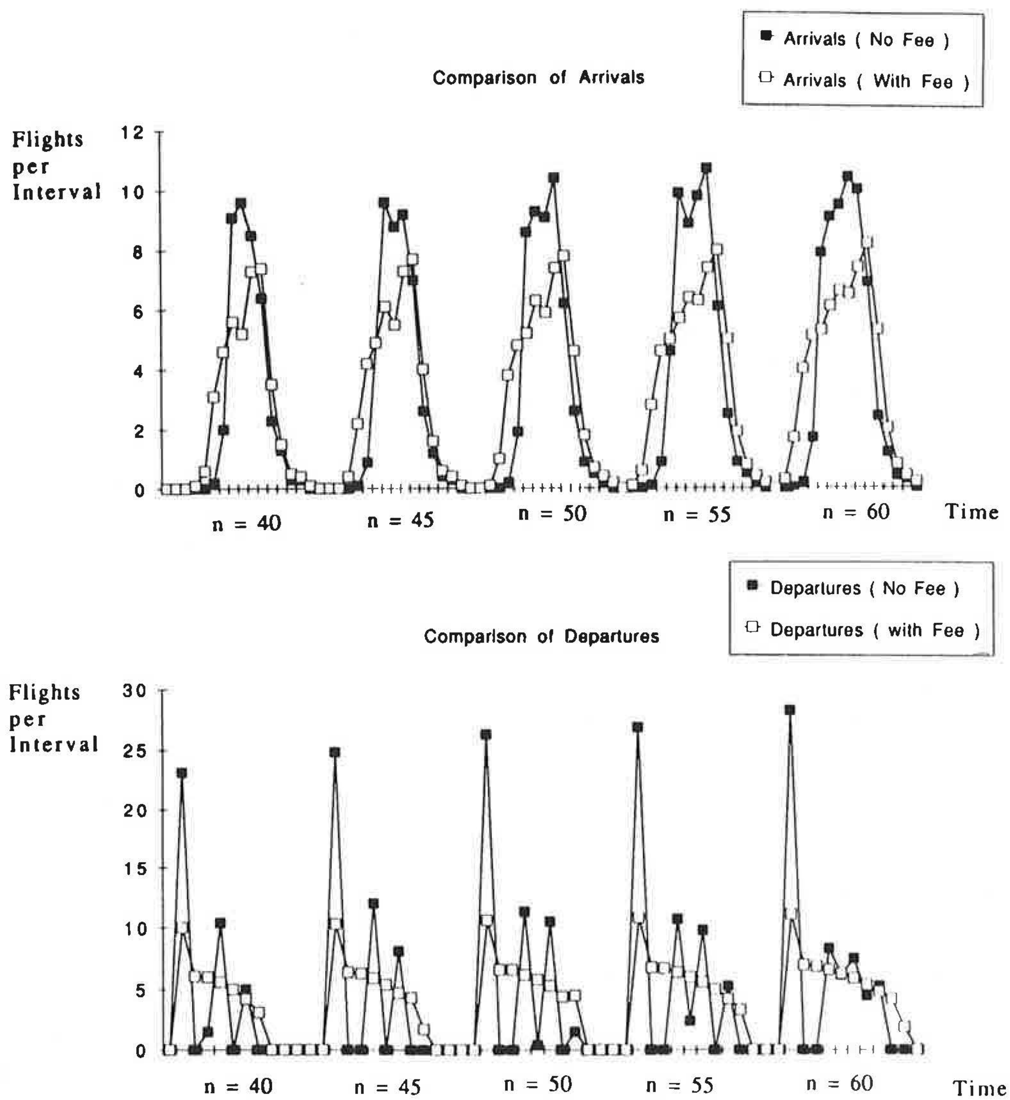


FIGURE 6 Effect of congestion fees on arrival and departure rates.

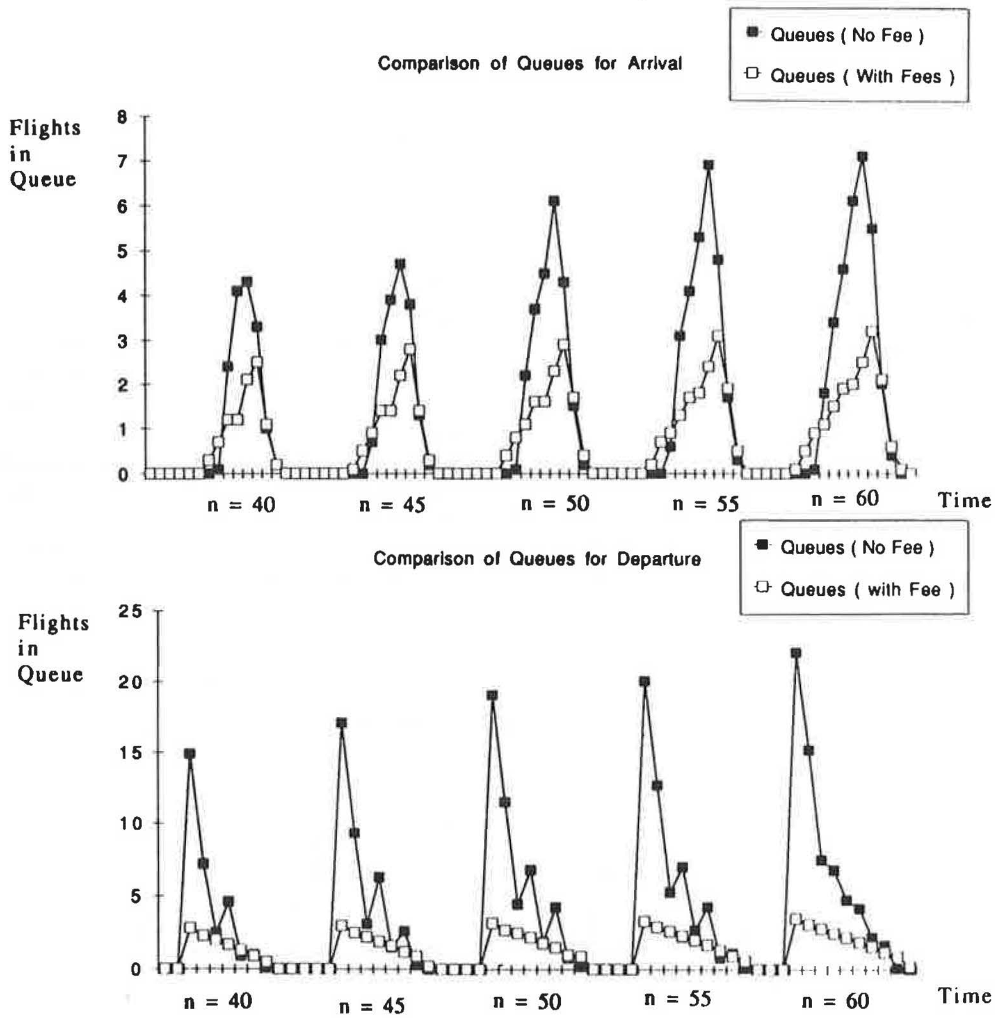


FIGURE 7 Effect of congestion fees on arrival and departure queue.

peak-departure rates immediately following the interchange and very even and gradual tapering off of departures.

Figure 7 shows the changes in queues that result from imposing equilibrium congestion fees. The no-fee hypothetical arrival queues look very similar in slope and magnitude to those of Figure 3. They tend to peak with similar slopes regardless of bank size; larger banks have progressively higher queue peaks. Congestion-fee arrival queues are more spread out and achieve only about 50 percent of the peak size of no-fee queues. The reduction in departure queues is much more dramatic. Because of more even spreading of the congestion-fee departure rates, the departure queues are very small. The peak congestion-fee departure queue is about 17 percent as long as the corresponding no-fee queue. On average, flights experience much less queuing delay in the congestion-fee equilibrium but experience higher early- and late-time costs because of the peak-spreading. Those familiar with standard formulations of bottleneck models may wonder why any queues exist in the congestion-fee equilibrium. The standard formulation assumes no randomness in the arrival or departure process. Queues only develop when traffic exceeds capacity. The queuing process used here is estimated from a queuing model with Poisson arrivals and departures that generate positive queues with arrival or departure rates below capacity. Eliminating the queues would require very low arrival or departure rates and very long layover delays.

Figure 8 shows the change in marginal external congestion between the no-fee and congestion-fee equilibriums. The marginal external congestion schedules in the congestion-fee cases are, of course, the equilibrium congestion-fee schedules. The peak external congestion costs caused by arrivals decrease by about 50 percent in response to the congestion fees. The peak external congestion costs caused by departures decrease by between 66 and 75 percent. The equilibrium arrival fee schedule increases almost linearly with time as flights approach the interchange period. It peaks just before the interchange and then decreases very quickly, almost linearly. The equilibrium departure fees peak in the period following the interchange and decrease gradually with time, almost linearly. Thus, the optimal fees can be approximated using a simple piecewise linear fee schedule.

Two additional observations should be made regarding Figures 6 through 8. Randomness in the arrival process appears to mitigate the congestion externality problem. Peak external congestion levels are 50 to 66 percent less for no-fee arrival banks than for no-fee departure banks. Bottleneck models

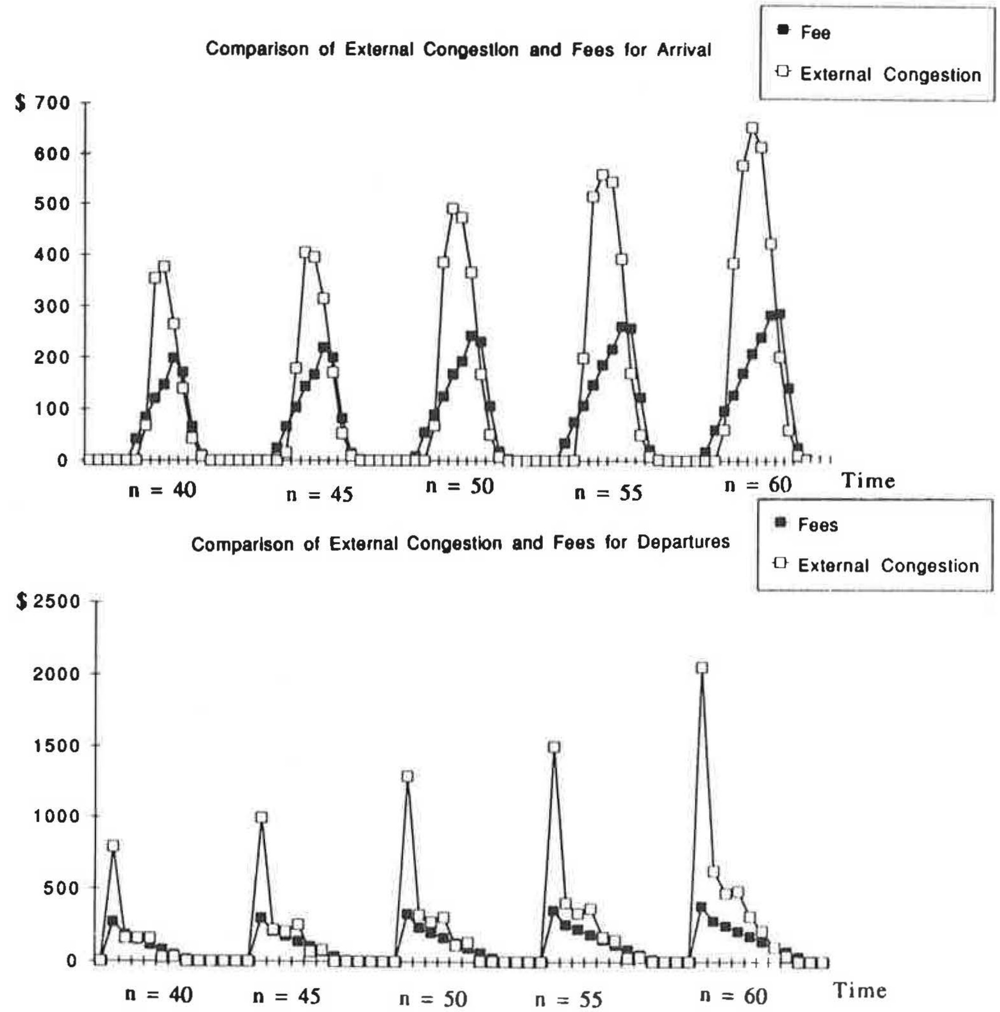


FIGURE 8 Comparison of external congestion and equilibrium fees.

that ignore randomness in the arrival or departure processes may seriously overstate the levels of congestion and the benefits of congestion pricing. That departures may be more random than indicated by the tower log data may explain why the departure rates and queues for the hypothetical no-fee banks are much higher and steeper than those in Figure 3. A second explanation may be that internalization of delay by the hub airline accounts for the difference. The actual rates of departure shown in Figure 3 fall between the hypothetical no-fee and the congestion-fee departure rates, as might be expected if the hub airline were partially internalizing delays.

Table 2 shows the changes in landing and takeoff costs per flight resulting from imposition of congestion fees. Columns 2 through 5 are the equilibrium values of C_a and C_d for the no-fee and congestion-fee equilibriums, respectively. Column 5 is the sum of C_a, C_d, and the weight-based fee for an average-sized aircraft (130,000 lb). In 1990, the weight-based fee at MSP was $0.70/1,000 lb, so the average fee was $91. Column 6 is the sum of C_a and C_d for the congestion-fee equilibrium. Column 7 shows the increase in cost per flight resulting from the congestion fees. Depending on bank size, the increase in cost is between 60 and 95 percent of the weight-based fee and about 10 percent of total landing and takeoff costs with a weight-based fee.

Table 3 shows the changes in airport revenues and resource savings per flight bank resulting from switching to congestion pricing. Airport revenues from the fees would increase three- to four-fold. The average increase in airport revenue per flight more than offsets the additional landing and takeoff cost per flight, resulting in a resource savings per flight of $125 to $205, depending on the number of flights in the bank. These resource saving are 25 to 30 percent of the total landing and takeoff costs experienced under the weight-based fee structure.

In addition to spreading out the airport's arrival and departure peaks, congestion fees may also change the frequency of arrival and departure banks. The airlines choose aircraft size and flight frequency to minimize the full cost of operating the hub network as given in expression 1. To minimize expression 1 requires knowledge of how C_a and C_d vary with changes in aircraft size, P, which in turn requires knowledge of how C_q, C_e, and C_l vary with P. The cost parameters, C_q, C_e, and C_l, are composed of passenger time costs, which are essentially proportional to P, and aircraft costs, which increase at approximately the 0.75 power of P. So, for example, $C_q = \alpha P + \beta P^{0.75}$. Unfortunately, the data do not enable estimation of the coefficients, α and β, which apportion these costs between passengers and airlines. Using what it is hoped are reasonable assumptions about these coefficients, the following

TABLE 2 COMPARISON OF LANDING AND TAKEOFF COSTS

Flights in Bank	Landing Costs per Flight ($)		Takeoff Costs per Flight ($)		Landing and Takeoff Cost per Flight ($)		Increase in Cost per Flight[a] ($)
	Weight Fee	Congestion Fee	Weight Fee	Congestion Fee	Weight Fee	Congestion Fee	
40	282	332	314	320	596	652	56
45	295	360	354	353	649	713	64
50	322	391	391	389	713	780	67
55	344	418	408	416	752	834	82
60	363	446	444	448	807	894	87

[a]Including fees.

estimates of the relationship between landing and takeoff costs and aircraft size fit simulated data from the bottleneck model extremely well:

$$C_a + C_d = 13.558P^{0.834} \quad \text{for the no-fee equilibrium}$$

$$C_a + C_d = 13.749P^{0.877} \quad \text{for the congestion-fee equilibrium}$$

The weight-based fees at MSP can also be estimated as a function of P:

$$F = 0.24P^{1.23}$$

By substituting reasonable values for the parameters in expression 1, it can be rewritten as

$$\frac{61}{P}\{2*60*34*P^{0.75}*2 + [(2 - 0.5)*60*P*10.66]*\frac{P}{3/8*61}$$

$$+ 2*60*P*2*25 + 60\,[f*13.749*P^{0.877} +$$

$$(1 - f)*(13.558*P^{0.834} + 0.24*P^{1.23}) + (8*P + 4*P^{0.75})]\}$$

where $f = 1$ for congestion fees and $f = 0$ for weight-based fees.

These parameters are chosen so that the solution for the weight-based-fee case is $P = 117$ and the interval between banks is 1.9 hr—the average values for MSP. The solution for the congestion-fee case is $P = 117.625$, which indicates that, given this parameterization of the model, the congestion fees have virtually no effect on service intervals. An interesting question is whether this result would hold for networks serving markets with different demand densities. Table 2 shows that significant reductions in cost per plane would result if there were fewer planes in a bank, but serving smaller markets with lower frequency than every bank would increase schedule delay in the small market. The current model does not answer whether the cost savings from smaller banks would exceed the increase in schedule-delay costs.

FURTHER RESEARCH AND CONCLUSIONS

The model suggests a number of issues for further research. Foremost among these is the need to model the internalization of congestion by a dominant hub airline. Imposing atomistic fees on it would cause it to overinternalize the delay its airplanes impose on one another, thereby spreading its arrival and departure banks out too much. Because it would be unacceptable to have fees that favor the hub airline, the constrained-optimal single-fee structure must balance overinternalization by the hub against underinternalization by the nonhub airlines. This issue has been overlooked in the past, but it must be significant at airports where a single airline accounts for more than half of the traffic.

The model's realism would be improved by relaxing the assumptions that all flights serve identical markets, are the same size, and have the same desired arrival and departure times at the hub. A primary political objection to implementing congestion pricing is its effect on service frequency in lower demand-density markets. The hub scheduling model could be extended to model networks in which some routes are served with less frequency than every bank, thereby providing some insight about the effect of congestion pricing on service frequency in marginal markets.

Allowing market density to vary would also require changing the bottleneck model to accommodate flights of different sizes. Different-sized flights have different queuing and layover-delay costs. In a bottleneck model with flights of different sizes, the smaller planes with lower aggregate passenger and aircraft-delay costs would shift further away from the peak. In equilibrium, similar flights would still have identical costs regardless of arrival and departure time, but flights of different sizes would have different costs. (Smaller flights generally have smaller congestion fees than larger flights because they operate at less desirable times and because they impose delay mostly on other lower-cost flights.)

Finally, the model's implications for optimal airport capacity should be studied. Previous models of airport capacity have assumed that demand peaks are independent of capacity.

TABLE 3 AIRPORT REVENUES AND RESOURCE SAVINGS

Flights in Bank	Total Airport Revenues ($)		Increase in Airport Revenues ($)	Avg Increase in Revenues per Plane ($)	Total Resource Gain ($)	Avg Resource Gain per Flight ($)	Resource Gain as Percent of No-Fee Costs
	Weight Fee	Congestion Fee					
40	3,640	10,867	7,227	181	4,987	125	24.7
45	4,095	13,460	9,365	208	6,485	144	25.8
50	4,550	16,647	12,097	242	8,747	175	28.1
55	5,005	19,508	14,503	264	9,993	182	27.5
60	5,460	22,951	17,491	292	12,271	205	28.6

The bottleneck model implies that additional capacity would increase the height and steepness of the demand peaks as airlines attempt to shorten the average layover period their passengers experience. Properly accounting for the benefits of additional capacity requires a model that includes layover costs and endogenous demand peaking.

Several innovations in the airport congestion-pricing literature have been introduced here. An atomistic hub-and-spoke network's scheduling problem is modeled, thereby explaining the underlying causes of the periodic demand peaking experienced at hub airports. The model explains how airlines choose aircraft size and service frequency to minimize full costs of service and it captures changes in service frequency in response to congestion pricing. The hub-network model also motivates the use of a bottleneck model of the timing of arrivals and departures within the network's flight banks. The bottleneck model captures peak-spreading and the important trade-off between layover and congestion delay. The standard bottleneck model is extended to allow flights to deviate randomly from their schedules and is given a discrete-time specification that facilitates empirical application of the model. Implementing the model for MSP describes the airport's arrival peaks quite well but overestimates the peak rates of departure and the resulting queues and congestion. The optimal fees have a simple structure that leads to modest increases in airline and passenger costs and significant increases in airport revenues and net social welfare.

ACKNOWLEDGMENTS

The author is deeply grateful to Herbert Mohring of the University of Minnesota for suggesting this topic and providing guidance and advice at every stage of its development. E. Thomas Burnard and Larry L. Jenney of TRB, Greig W. Harvey of Stanford University, and Francis X. McKelvey of Michigan State University provided useful comments, suggestions, and encouragement. Special thanks are due Steve Thompson of the University of Minnesota, who was instrumental in procuring the tower log data and entering the data into computer files.

The author also wishes to thank FAA for its financial support, TRB for administering the Graduate Research Award Program, and the Center for Transportation Studies at the University of Minnesota for additional financial support.

REFERENCES

1. *Special Report 226: Airport System Capacity: Strategic Choices.* TRB, National Research Council, Washington, D.C., 1990, 134 pp.
2. H. Mohring and M. Harwitz. *Highway Benefits: An Analytical Framework.* Northwestern University Press, Evanston, Ill., 1962.
3. B. Koopman. Air-Terminal Queues Under Time Dependent Conditions. *Operations Research*, Vol. 20, 1972, pp. 1089–1114.
4. A. Carlin and R. Park. Marginal Cost Pricing of Airport Runway Capacity. *American Economic Review*, Vol. 60, June 1970, pp. 310–319.
5. A. Carlin and R. Park. A Model of Long Delays at Busy Airports. *Journal of Transport Economics and Policy*, Vol. 4, 1970, pp. 37–52.
6. S. Morrison and C. Winston. *The Economic Effects of Airline Deregulation*, Brookings Institution, Washington, D.C., 1986.
7. H. Mohring. *Transportation Economics.* Ballinger Publishing Company, New York, N.Y., 1976.
8. W. Vickrey. Congestion Theory and Transport Investment. *A.E.R. Proceedings*, Vol. 59, 1969, pp. 251–260.
9. S. Morrison and C. Winston. Enhancing the Performance of the Deregulated Air Transportation System. *Brookings Papers on Microeconomics*, 1989, pp. 61–123.
10. S. E. Omosigho and D. J. Worthington. "The Single Server Queue with Inhomogeneous Arrival Rate and Discrete Service Time Distribution. *European Journal of Operations Research*, Vol. 22, 1985, pp. 379–407.
11. *Official Airline Guide.* Richard A. Nelson, Oak Brook, Ill., May 1, 1990.

U.S. Cabotage Policy

David P. Agnew

The continuing liberalization of the international aviation regime has forced the reexamination of many long-standing policies and accepted conventions. Cabotage, the right of a foreign carrier to transport domestic traffic within another country, has been almost universally banned throughout aviation history. Recently, however, cabotage has surfaced as a major issue in various international aviation forums. The regionalization of aviation, as evidenced by the recent moves toward unity in Europe, has given special force to calls for cabotage. As the most robust aviation market in the world, the U.S. market is an obvious target for foreign carriers and governments that want to share the enormous amount of U.S. air traffic. Calls from some European governments and carriers have been particularly strong and may, if the trend toward European integration continues, force an alteration of U.S. cabotage policy. There are significant legal, political, and practical barriers to such an alteration. Despite these barriers, however, U.S. policymakers should continue their recently revealed willingness to use cabotage at the bargaining table. The potential advantages of allowing foreign carriers to carry cabotage traffic in the United States include increasing competition in U.S. markets, providing international service to secondary U.S. gateways, and, perhaps most important, improving the negotiating position of the United States vis-à-vis other countries or regions. Potential costs of allowing cabotage, such as those to U.S. carriers, must also be a factor in the decision. An analysis of the potential impact of limited cabotage (allowing foreign carriers to operate only a limited number of domestic routes) suggests that allowing a limited form of cabotage would have neither an extremely negative nor an extremely positive impact on U.S. and foreign carriers, respectively. Because of the formidable level of U.S. carrier opposition and other factors, the foreign carrier market share would probably be minimal. The Quality of Service Index (QSI) model, when used to analyze potential European carrier limited-cabotage routes, clearly points to this outcome. In the routes most likely to be targeted by European carriers, the QSI model predicts market shares of less than 5 percent. U.S. policymakers should consider using cabotage as a negotiating tool with European countries. As the European Community becomes a more organized and effective economic region, this willingness may become a necessity. In the meantime, the potential benefits and costs of allowing limited cabotage within the U.S. market should be analyzed carefully. The winners and losers of the various altered scenarios should be identified. Scenarios should be tested with respect to the impact on U.S. carriers and consumers. Finally, the United States should take a proactive role in the worldwide discussion of cabotage and actively identify and pursue those ends that are of the greatest value to U.S. interests.

The institutions and policies of international aviation have undergone a dramatic restructuring in recent years. Worldwide, in domestic markets and in the international marketplace, deregulation and globalization have altered the assumptions that governed aviation for the past half-century. Deregulation in the United States has produced a level of competition and industrywide restructuring that would have been unthinkable 15 years ago. The forces of the European market are reducing the importance of national boundaries in the region and may force changes in the negotiation strategies of non-European Community (EC) countries. As some of these changes occurred, new aviation issues emerged in the international arena. Cabotage is one of these issues.

John F. Kennedy School of Government, Harvard University, Cambridge, Mass. 02138. Current affiliation: Price Waterhouse, Office of Government Services, 1801 K Street, N.W., Washington, D.C. 20006.

Cabotage, the carriage of domestic traffic within another country by a foreign carrier, has been banned almost universally throughout modern aviation history. Exceptions to the cabotage ban have been few. As regions band together for negotiating purposes and liberalization reduces the need to protect individual carriers, however, the logic behind a strict ban on cabotage begins to unravel. The EC, for instance, has started to realize that its market power would dramatically increase if its members were able to develop a coherent regional aviation community. The purpose of this paper is to consider U.S. cabotage policy, particularly in light of the important market events in Europe. After an examination of the current U.S. cabotage policy, the calls for change, the obstacles to change, and the implications of various alternative scenarios will be discussed.

First, the terms of discussion must be identified. As a general term, cabotage includes any form of domestic transport by a foreign carrier. Fifth-freedom rights are those that allow a carrier to pick up passengers from a state other than its own and transport them to a third state, also not its own. As explained subsequently, fifth-freedom rights are important because they are sometimes equated with cabotage rights by foreign carriers and governments. Beyond rights are those that allow a flight to operate an extension of an international route within the United States without being able to transport passengers on the domestic leg. Limited cabotage allows foreign carriers to transport domestic passengers on the beyond-right routes. Long-haul cabotage is limited cabotage that occurs on longer domestic flights, those that are often cited as potential targets.

U.S. CABOTAGE POLICY

The term *cabotage*, sometimes known as the eighth freedom of the air, has its origins in coastal shipping, also traditionally banned to foreign ships (*1*). (The term *cabotage* may derive from *cabot* or *chabot*, French terms for a small vessel. Alternatively, according to *Black's Law Dictionary*, it may be derived from the Spanish word *cabo*, meaning "cape," which was used to describe navigation proceeding from cape to cape along the coast without going into the open seas.) The current U.S. cabotage law has its roots in the proceedings of the Convention on International Civil Aviation (the Chicago Convention of 1944) and in several U.S. aviation statutes.

The Chicago Convention was called by Roosevelt and Churchill to "establish a workable and efficient international aviation system through the establishment of multilateral agreements for the exchange of commercial air rights" (*2*). Article 7 of the Chicago Convention has served as the international law on cabotage since 1944 (*3*):

> Each contracting State shall have the right to refuse permission to the aircraft of other contracting States to take on in its territory passengers, mail and cargo carried for remuneration or hire and destined for another point within its territory. Each contracting State undertakes not to enter into any arrangements which specifically grant any such privilege on an exclusive basis to any other State or an airline of any other State, and not to obtain any such privilege from any other State.

Signatory states decided to restrict cabotage for several reasons: (a) fully aware of aviation's military applicability and their territorial vulnerability, states sought to ensure national security; (b) cabotage prohibitions were employed to protect each nation's fledgling airlines; and (c) air transportation, unlike most sea transportation, involves deep penetraton into another state's sovereign territory, adding to the discomfort of those concerned with foreign access to interior markets and geography (*1*, p. 1061).

Several U.S. statutes also prohibit cabotage. Section 6(c) of the Air Commerce Act of 1926 prohibited cabotage by stating: "No foreign aircraft shall engage in interstate or intrastate air commerce" (*1*, p. 1065). The Civil Aeronautics Act of 1938 amended Section 6(c) to read that "no foreign aircraft shall engage in air commerce otherwise than between State, Territory, or possession of the United States, or the District of Columbia and a foreign country" (*1*, p.1066). In 1953 the cabotage section of the Air Commerce Act of 1926 was altered to reflect the provisions of Article 7 of the Chicago Convention. Finally, the current Federal Aviation Act (P.L. 85-726), passed in 1958, incorporated Section 6 with only minor modifications. Section 402(a) of the Federal Aviation Act is also recognized as a further restriction on cabotage within the United States.

Despite the historical tenacity and virtually universal acceptance of the ban on cabotage, recent liberalization efforts have triggered new thinking on the issue. Although most countries remain concerned about protecting the interests of their own carriers and are unwilling to open up their markets to foreign competition, some have started to call for the use of cabotage within the markets of their bilateral partners. With the largest aviation market in the world, the United States is an obvious target for foreign governments and airlines interested in cabotage. Intriguingly, these calls for an altered U.S. cabotage policy have come from within the United States as well as from foreign governments and carriers.

CALLS FOR CHANGE

Within the United States

To some domestic observers of American aviation, cabotage appears to be an ideal way of injecting competition into post-deregulation markets. Proponents of the concept argue that on many routes dominated by the hub-and-spoke operations of a particular carrier, fares are higher and service quality is down. Cabotage, they argue, would increase competition, and thus benefit the consumer, without a commitment to reregulation of the industry. A 1988 Congressional Budget Office report called for "allowing foreign carriers to provide domestic service" to increase competition (*4*). Some also view cabotage as a way of serving local economies with more connections, a benefit that groups such as USA BIAS (a group of U.S. airports seeking international service) would welcome heartily. Others cite the potential benefits to the U.S. economy from cabotage, because foreign carriers "would need to hire U.S. employees to man their local operations work, pay local landing and gate rental fees, and buy most of their fuel here" (*5*). The infusion of foreign capital into the aviation system might be substantial. Finally, supporters of an altered cabotage policy maintain that additional benefits would accrue to U.S. carriers and consumers by way of an improved negotiating position. U.S. carriers, they argue, would fare well in a deregulated world marketplace. As the argument goes, even if foreign carriers were allowed access to certain domestic U.S. markets, U.S. carriers would handle that challenge well and prosper in the markets around the globe that would open up as a result of the altered policy. Some see the relaxations as a way of getting Europeans to allow U.S. carriers the right of price leadership on fifth-freedom routes that they already operate (*5*, p. 36). The murmurings from within the ranks of U.S. carriers in support of an altered cabotage policy may be based on this type of reasoning. As a negative motivation, some international routes might be taken away from U.S. carriers in the event that bilateral or multilateral partners become angry enough over this issue to act.

In Foreign Countries

Foreign airline and government officials have direct incentives to push for a change in the U.S. cabotage policy. By all measures, the U.S. commercial aviation market is the largest and most robust in the world. Access to that market for a competitive carrier represents tremendous opportunity. Many foreign carriers complain that U.S. carriers are able to use their vast domestic route networks to feed their international flights and thus secure a substantial competitive advantage (*5*). Many foreign carriers would, of course, prefer unlimited access to all U.S. domestic markets. Barring that possibility, however, these carriers claim that even limited cabotage rights would enable them to justify some services that are now economically infeasible because of beyond rights. Some foreign carriers claim that they would be able to justify more routes of this sort if allowed to carry limited cabotage traffic.

The value of cabotage routes, however, is not universally recognized. According to those who doubt the significance of the issue, cabotage can be requested repeatedly with little chance of its actually being realized. These skeptics view the potential economic value of cabotage traffic as secondary to the value of perceptions of an intransigent United States. Thus, the claims of unfair treatment may be used in negotiations to exact concessions from the United States. Despite these misgivings about the motivations of foreign negotiators and the value of cabotage rights, the fact remains that cabotage is an issue that U.S. and foreign governments must deal with at some point in the future. As long as cabotage

remains on the table, U.S. policymakers must develop a sound response either altering or reaffirming the current U.S. ban. The adopted or affirmed policy must balance the potential benefits and costs that will be experienced by both U.S. carriers and consumers.

The pro-liberalization entities within the European aviation community have been particularly persistent in their requests for a change in U.S. cabotage policy. As the prospect of a unified European aviation community has become more realistic, European officials have started to equate the fifth-freedom rights of U.S. carriers within Europe with cabotage rights for themselves in the U.S. market. If Europe comes to be viewed as a single entity, these officials ask, then why should European carriers not enjoy the same rights in the United States that U.S. carriers enjoy (i.e., present fifth-freedom rights) within Europe?

To fathom the impact of these changes in the European market, an analysis of the current dynamics, grounded in the history of the market, is essential. A brief summary of the recent history of the European aviation community will be followed by a discussion of the implications of the present situation on U.S. cabotage policy.

EUROPEAN MARKET: HISTORICAL PERSPECTIVE

The European aviation market is changing. After years of operating in a heavily regulated market, European government officials and commercial carriers are facing changes that will force them to rethink their operating assumptions. No longer will national flag carriers be able to depend on a government-guaranteed market share on particular routes. No longer will discount fares be ruled out when smaller carriers are trying to challenge the majors. Perhaps most important, no longer will the aviation community as a whole be able to depend on exemption from the rules of the European Economic Community as laid down by the Treaty of Rome. No one denies that change is on the horizon. The extent to which the market will develop and the pace of change, however, are still subjects of debate both inside and outside Europe.

In 1957 the Treaty of Rome established the European Economic Community (*6*). The treaty dealt briefly with aviation, and contained one article that pointed directly "towards the adoption of a common transport policy, another establishing a target of eradicating state boundaries for business, and six (articles 85–90) prohibiting cartels and doing away with anticompetitive practices" (*6*). These articles laid out an ideal that was unworkable in the air transport world. In 1961 air transport was removed from the Treaty of Rome with respect to competition rules. The ultimate decision on European Economic Community aviation policies was, in effect, postponed by Article 84, which dictated that "the Council would develop a policy on air transport eventually" (*7*). Since 1961, European aviation has continued to be governed by bilateral relationships negotiated between the governments of the member states. These "bilaterals nominated the airlines that could fly, the fares that could be charged, provided for pooling agreements under which departure timings were agreed, and revenue split under an agreed formula" (*6*). Airlines were therefore exempt from any procompetitive policies of the European Economic Community and could depend on a predictable regime that allowed them to exist without a great deal of competitive flexibility. The "interests" of the carriers were most often the prime considerations of the state transport officials. Pricing and flight availability were set primarily according to airline planning needs.

This regime was not challenged until recently. A European Court ruling in 1974 hinted that civil aviation was subject to the general rules of the treaty (including the competition provisions). Despite this ruling, real change did not become imminent until the early 1980s, when it became apparent to many that the European aviation structure simply had to change (*7*). The liberalization process was hastened by U.S. initiatives such as the extension of U.S. antitrust laws to the International Air Transport Association and the renegotiation of bilaterals with the United Kingdom and the Netherlands. The deregulation experience in the United States also spurred some proponents of European liberalization to argue that consumers had suffered under an aviation regime that sacrificed their interests in the name of carrier profitability and planning. These proponents received some valuable support in 1982 from the Competition in Air Services (COMPAS) report, which was issued by the European Civil Aviation Conference. The COMPAS report urged liberalization in several important areas, including route entry, capacity, and pricing.

The calls for liberalization fell on sympathetic ears within the EC as well. In 1984 the Commission issued its Second Memorandum on Air Transport, which proposed to maintain the bilateral regimes between member states but sought to loosen the restrictions on the industry to encourage efficiency and innovation on the part of Europe's airlines (*8*). Though this was certainly a modest step, it signaled an evolving attitude. As noted by Reed (*6*):

> By 1986 EC opinion had hardened to the extent that airlines that failed to rid themselves of cartel-like fares, pooling agreements and all the other trappings of the "bad old days" were being threatened with references to the European Court on the grounds that they were in violation of the competition clauses of the Treaty of Rome.

This was a radical departure from the days of regulation. These changes, however, represented a gradual shift in policy and attitude rather than a U.S.-style deregulation of the industry.

Proponents of a U.S.-style deregulation usually cited potential consumer gains, such as lower fares and improved flight availability, in a deregulated environment. The opponents cited a long list of reasons why this type of total deregulation would never work in Europe. The U.S. market, said the opponents, was larger and was unique in its uniformity of laws, consumers, and government institutions. The U.S. market was also free from the wide diversity of languages and cultures that would hamper European deregulation. The complexities of Europe were alluded to by Karl-Heinz Neumeister, Secretary General of the Association of European Airlines, in a speech given in May 1989 to the Airport Operators Council International: ". . .above all we are part of a political, economic, and social process of European integration" that ". . .goes much deeper and its scope far beyond the narrow aspect of just finding new rules for the airline business." The market also had a robust demand and a geographic size that

made it special. Finally, and perhaps most important, the United States was not encumbered by the political considerations of flag carriers and the national pride of 12 independent states. This resistance by European officials and carriers to U.S.-style deregulation was based on a belief that reform should proceed by slowly introducing more freedoms into the existing system without shaking its foundations. These evolutionists carried the day. Thus, the European aviation market was never destined to undergo a one-shot, radical transformation.

Nevertheless, European reformers could learn a great deal from the U.S. experience. McGowan and Seabright (*7*) argue that deregulation can produce major gains for the European consumer as long as European officials deal effectively with air traffic control and airport congestion problems and maintain a vigorous competition policy, areas commonly perceived as failures of the U.S. goverment. They see the scarcity of landing slots at European airports and potential merger activity as tools with which the more established carriers would be able to keep out new entrants and abuse their dominant position. Their views represent some of the latest thinking on how to apply the U.S. deregulation experience to European liberalization.

In December 1987 the EC Council of Ministers, concerned with the state of competition in European skies, clarified the rules of procedure for European Economic Community competition rules and their application to the air transport sector. This package started the formal process of change but was viewed by many as a poor first step in the liberalization process. A representative of the Federation of Air Transport Users said that the plan was (*9*) "the very minimum step forward in the process of liberalization" and that "any action to limit the liberalization plan would be resisted by consumer groups." In spite of these criticisms, the package did, at least, begin the process of liberalization, not an insignificant event in a market dominated by heavy regulation for nearly half a century.

Specifically, the 1987 reform package contained elements pertaining to tariffs, capacity, multiple designation, and fifth-freedom rights, such as:

- Allowing for "discount and deep discount zones within which fares would be automatically approved" (*10*) (fares outside these zones would be approved as long as it could be proved that the proposed changes were cost-related);
- Providing for a gradual movement away from the strict 50/50 percent capacity sharing agreements that have dominated for so long, lowering from 45 to 40 the percentage of a particular market share that a state can demand for its own carriers (*10*);
- Allowing for multiple designation of carriers on city pairs with more than 250,000 passengers annually; and
- Granting fifth-freedom rights between hubs and other airports as long as they were an extension of existing third- and fourth-freedom services and did not exceed 30 percent of passenger seats.

Thus, the 1987 Brussels package started the process of liberalization but stopped short of really reforming the system.

The EC Council of Ministers extended the liberalization of the European market during a June 1990 meeting held in Brussels. The new measures replace the 1987 package and became effective November 1, 1990. Though not likely to transform the face of European aviation in a radical fashion, the 1990 package represents significant progress for those who envision a European transport market with no barriers. The package contains specific provisions in the areas of market access and capacity sharing, passenger fares, and exemptions from EC competition rules.

The 1990 package reduces the traffic threshold at which member states must accept multiple designation to 140,000 in 1991 and 100,000 in 1992 and allows the capacity share to be increased periodically by 7.5 percent, with all restrictions on capacity share to be abolished in 1993. In addition, fifth-freedom traffic is now allowed on all routes with a limit of 50 percent of all seats. The restrictions on discount fares are also relaxed, allowing for additional zones of automatic approval and removing certain restrictions. Though not formally in the documents, true intra-EC cabotage is also introduced as a goal for 1993.

From the perspective of the United States, other interesting questions concern the likely positions to be taken by the EC in its aviation relationships with non-EC countries. The internal market changes are certain to bring about fundamental shifts in attitudes and institutions in this area. But what effect will these changes have on the bilateral regimes now in place between the EC member states and non-EC countries? Will these existing bilaterals be replaced altogether? Will the member states of the EC ever surrender their negotiating powers to an EC institution? If any of these hypothetical situations do come about, at what pace can they be expected? These are the questions that should be considered in a discussion of U.S. cabotage policy.

At the very least, continued liberalization of the European market must overcome significant legal, structural, political, and economic barriers. To name a few of the most striking:

- Each member state has its own interests in terms of the division route rights and other benefits of negotiation and member states are still tied to the concept of flag carriers.
- Until the EC irons out its own internal aviation policy, hopes for a united external policy are diminished.
- It is still not clear which institution will carry the torch of European aviation policy.

Before the internal aviation market proceeds with reforms, the answers to these questions must crystalize.

Despite the uncertainty surrounding the previous questions, the fact remains that if the EC unites, the region could demand treatment as a single aviation entity. This declaration could transform existing U.S. carrier fifth-freedom rights into European cabotage. As of January 1, 1990, U.S. carriers had rights to 84 intra-EC routes (Table 1). European officials would have to decide whether to renew or extend these routes. Though they are bound by international treaties to allow current routes to continue operation, it is possible that they would be willing to grant additional cabotage rights to U.S. carriers either within the EC or individual European countries. The United States must therefore be prepared to pay for its fifth-freedom traffic in Europe (both existing and future) or to consider a liberalization of U.S. cabotage policy.

TABLE 1 U.S. CARRIER TRAFFIC IN EUROPE

U.S. Airline	Weekly Flights Scheduled to or via EC States	Flights Using Intra-EC 5th-Freedom Rights
American	105	14
Continental	29	0
Delta	53	0
Northwest	26	0
Pan Am	122	42
TWA	133	28
USAir	7	0

SOURCE: International Civil Aviation Organization (Based on scheduled traffic January 1, 1990).

CHANGE IN U.S. CABOTAGE POLICY

Because of the likelihood of European change, the United States should be prepared for a new negotiating scenario with the European aviation community. If Europe does reach the point of a unified negotiating position, the United States should have at least considered potential scenarios ". . .so that whatever policy it does create will still provide benefits to its citizens, no matter how large the negotiating partner" (*11*).

The United States has been deflecting demands from the Europeans for cabotage rights for several years. The basis for these refusals has been outlined at length by U.S. officials, who doubt both the legitimacy and value of these claims. The ability of Europe to suddenly declare itself a cabotage area is questioned, as is the value of cabotage rights for foreign carriers.

Despite these misgivings, the United States should explore cabotage as a possibility, perhaps by experimenting with limited cabotage on certain routes. The outright ban, though well founded historically, does not maximize potential value for the United States. U.S. interests would be best served if cabotage were at least prepared for as an option. The costs and benefits of an altered policy for U.S. carriers and consumers should be weighed against the likely impact on negotiations with the EC. Even if the EC aviation community does not organize to the point at which it can demand cabotage rights for several years, the United States should be prepared for the occasion. As this reality becomes more established in the minds of Europeans, the necessary changes and assignment of powers are likely to be subjugated to the realization of these potential benefits.

Barriers to Change

Before the type of cabotage policy the United States might adopt can be addressed, the barriers to change are worth discussing. The potential hurdles to cabotage are significant:

- The political realities of striking unions and angry carriers would not be a welcome sight for most aviation officials or politicians.
- Legal barriers still exist, both in U.S. domestic law and in international law.
- European carriers might find that even with the route rights they would be unable to compete with the powerfully based and more efficient U.S. carriers.
- Potential logistical difficulties exist, which include slot limitations on the requested routes and European airframe and pilot shortages.

Political realities being what they are, any change in the current U.S. cabotage policy will be met with fierce resistance. The Air Line Pilots Association (ALPA), for example, threatens, ". . .if cabotage is ever adopted, we would shut down the U.S. air system in protest" (*12*). Of these types of threats by unions such as ALPA, a member of the TRB Aviation Economics and Forecasting Committee states, "The threat of striking unions and angry carriers has often been put on the table, but has seldom really achieved anything. It is a tactic which will be used no doubt, but it will probably not be effective" (V. Golich, unpublished data). (Indeed, cabotage will soon be on the table in the U.S.-Canadian talks, which represents a positive sign.) There are ways that the U.S. negotiators might reduce the likelihood of such all-out opposition. Phasing in the changes or trying out an altered policy for a trial period represent two possibilities. Also, cabotage as an issue might lose some of its emotional charge as it becomes discussed more frequently.

As to the legal barriers, domestically, a change in the U.S. cabotage policy would require congressional and presidential approval. Given the opposition of the unions and most of the carriers at the present time, this is not likely to be an easy task. Though congressional opposition to a change has traditionally been intense, this opposition might subside after the debate on the issue became commonplace. In addition, changes in the regulatory framework must comply with existing U.S. antitrust legislation.

In terms of international law, questions still exist about the legality of cabotage with respect to the rules of the 1944 Chicago Convention. The language in Article 7 of the Convention on International Civil Aviation suggests that any cabotage offering must immediately be extended to all other signatory states, thus removing any potential for using cabotage as a bargaining tool. Such an interpretation would be compatible with the prevailing philosophical guidelines that informed international agreements following World War II, namely, multilateral governance and commitment. These principles are evident in the General Agreement on Tariffs and Trade mandate and operating rules (V. Golich, unpublished data).

The proponents of cabotage, however, argue strenuously against this interpretation. Several alternative approaches to dealing with the wording of Article 7 have been stressed. The first opposing view hinges on the interpretation of the words "specifically" and "exclusive" and holds that a country should be able to grant cabotage rights as long as the possibility is held open for grants to other countries at a later date. In other words, a country would not be able to grant exclusive cabotage to any one country. A country granting cabotage must, in good faith, hold open the possibility that cabotage might be granted to other countries at a later date. This interpretation would allow a given country to grant cabotage to another country without opening its market to every country in the world, thus preserving the value of cabotage as a negotiating tool.

The second approach has been used effectively by the Scandinavian countries in structuring their cabotage agreements: "The Scandinavian agreements remain in force as long as no hostile bid to carry cabotage traffic is received. Any non-approved seeker of cabotage can only play 'spoiler' by ending all cabotage; in no event can it obtain cabotage for itself, which removes the main reason for attempting to intervene" (*13*, p. 714). This second approach might face more difficulty in the United States, however, given the fierce competition to enter the world's largest market. Some excluded countries might not mind playing spoiler.

European carriers are likely to find that, even with the right to provide service, they would face stiff competition from U.S. carriers, which are, after the consolidation that followed deregulation, "in an excellent position to compete with any carriers that enter the market, foreign or domestic" (*13*, p. 716). There is little doubt that foreign carriers would be hard pressed to compete effectively, because, according to G. James in an address given at the 15th Annual FAA Aviation Forecast Conference in March 1990, they "must still face severe competition from U.S. carriers who have the strength of their domestic hub-and-spoke systems to counter this new, foreign competition." Foreign carriers would, at the least, be faced with the competitive disadvantages of relatively few offerings and lack of name recognition and customer loyalty. This is realized by many Europeans as well, including H. A. Wassenbergh, who stated in 1988, "It should be noted that U.S. cabotage rights have limited value for European carriers as the competition from U.S. domestic carriers would be very strong" (*14*).

Likely logistical difficulties include slot limitations on the requested routes and European airframe and pilot shortages. Many destinations in the U.S. market would be hard pressed to accommodate an influx of additional carriers because of the lack of airport capacity. Because the slots have already been divided among U.S. carriers, it appears that the foreign carriers might face some difficulties obtaining the necessary slots.

Possible Changes

If these barriers are overcome, the question of possible changes must be addressed. The United States could alter its cabotage policy in a variety of ways. First, the United States could maintain the status quo. Second, the United States could open its skies completely to foreign carriers. Third, limited variations of cabotage could be allowed. Finally, U.S. policies that forbid foreign ownership and control could be changed, thus diminishing the relevance of foreign calls for cabotage. How specific changes in foreign ownership and participation regulations would affect the cabotage policy debate merits additional discussion in the appropriate policy forums.

Status Quo

The United States may choose to maintain the status quo. By not changing policy, the United States would risk losing rights that it now possesses and would definitely never gain the additional route rights that an altered policy would make available. In the 1977 Bermuda II Agreement, U.S. carriers lost fifth-freedom rights from the United Kingdom to many European cities. A failure to yield to demands for cabotage could theoretically result in similar losses, though the value of these routes appears to be relatively small because the revenue from true fifth-freedom traffic within Europe accounts for a relatively small portion of the entire route revenue. Perhaps more important, however, the United States would be missing out on certain route opportunities that might be offered by other countries or groups of countries if the status quo were maintained. Finally, by maintaining the status quo, the United States would be passing up the opportunity to create some benefits for U.S. consumers. On the other hand, by maintaining the present laws, the United States would not be upsetting the unions and some U.S. carriers opposed to any type of relaxation on cabotage.

Open Skies

At the other end of the spectrum, the United States could allow foreign carriers unlimited access to U.S. skies. The obvious drawbacks are the substantial political difficulties and uncertainty. After the many unforeseen developments that accompanied U.S. deregulation, U.S. and foreign carriers and government officials are wary of all-out change. No one can predict precisely what might happen in the U.S. market if foreign carriers were allowed to function as domestic carriers. For this reason, many U.S. carriers would prefer not to face a round of imported competition.

On the other hand, open skies would be likely to have some positive impact for the U.S. consumer in the form of additional domestic competition and more international route availability. If pursued multilaterally, open skies might benefit U.S. carriers in the form of more international route availability. Consumers would benefit if such a deregulated environment allowed foreign carriers to inject competition into hub-dominated routes. Proponents of open skies decry the consolidation of the industry and argue that foreign competition within the United States is a logical extension of U.S. deregulation. The addition of a foreign carrier on particular routes may indeed benefit U.S. consumers marginally. Though predictions about what might happen have limited value, the important point is the feasibility of such policy changes at this time. Barring major upheavals, piecemeal changes in U.S. cabotage policy appear much more realistic than an all-out reversal.

Limited Cabotage

A third possibility is to alter the U.S. policy slightly and allow limited cabotage in the form of fill-up rights on the U.S. leg of flights entering the United States from Europe in one U.S. city and continuing to another. For example, three European carriers currently operate eight flights that have a leg entirely within the United States but do not pick up domestic traffic on the U.S. segment. Presumably, the additional fill-up rights would make these types of routes more economically feasible for European carriers to operate and might open up service to additional gateways. This policy would have benefits similar

to those of opening the market to foreign carriers but would be much more workable. Specifically, this type of change would have several implications:

- The negotiating position of the United States would probably be improved vis-à-vis the European countries (assuming that U.S. negotiators pushed for concessions in return), thus benefiting some U.S. carriers.
- U.S. consumers would benefit from the additional competition on domestic routes, however small, and would have the additional international flights on their list of possibilities.
- The incremental nature of this change would not carry with it the uncertainty of all-out change and would be more politically feasible than a drastic change in the status quo.
- Such a change would continue the process of opening up the international aviation order and could thus represent a step toward longer-term gains for U.S. carriers and consumers.
- By making an incremental change, the United States would force the hand of European carriers on the issue of cabotage without risking too much in return.

Given these advantages, it appears that this is the type of cabotage that U.S. policymakers should, and will be likely to, consider. On the assumption that the limited fill-up rights policy is the most likely type of U.S. response, it is necessary to determine which foreign carriers are interested in cabotage and decide what they expect to gain.

LIMITED CABOTAGE: IMPACT ON THE MARKET

To begin the analysis of an altered cabotage policy, a look at existing gateways, carriers, and route operations is instructive. To determine which routes might be candidates for limited fill-up rights by European carriers, several different possibilities should be examined. If allowed limited fill-up rights, European carriers are likely to target two types of routes:

- Routes that currently exist as beyond rights for European carriers (Scenario 1).
- Routes that are logistically convenient extensions of existing international services (Scenario 2).

To assess the potential impact of allowing limited fill-up rights to European carriers, both of these scenarios should be examined by using the Quality of Service Index (QSI) methodology.

QSI Model

The QSI model was developed in pre-deregulation days to predict the market share that would be captured by a carrier that introduced an additional flight in a particular city-pair market. Although no model can account for all of the factors that influence a consumer's decision to take the flight of a particular carrier, the QSI model correctly identifies many of the most significant.

The QSI model employed in this paper predicts the probable market share of a carrier that introduces a single flight to an existing market on the basis of the aircraft size, the number of stops, the number of connections necessary, and the weekly frequency. First, a value is assigned to each variable for every flight that currently operates on the given route. The flights are assigned values according to the following inputs:

Variable	*Value*
Aircraft size	Average capacity/100
Number of stops	
Nonstop	1.0
One	0.5
Two	0.25
Three	0.125
Number of connections	
Direct	1.0
one change	0.5
two changes	0.25
Weekly frequency	
Seven days	7.0
Six days	6.0
Five days	5.0
Four days	4.0
Three days	3.0
Two days	2.0
One day	1.0

The totals for each flight and the entire market are then computed. Finally, the total value for the additional flight is divided by the total value of flights currently operating in the market, giving the predicted market share of the added flight.

For example, say the route in question was Paris-Washington, D.C. Air France currently operates a flight from Paris that stops in Boston, unloads passengers, and continues to Washington, D.C. What market share would the Air France flight capture if allowed to transport additional passengers from Washington to Boston? According to the previous variables, the total value of the current operating schedule is 544.98, with the Air France flight operating a 747 aircraft 7 days a week with no stops and no connections. The parameters give the Air France flight a value of 12.5. The QSI methodology predicts that the Air France cabotage flight would capture 2.29 percent (12.5/544.98) of the Washington-Boston market.

Limitations

The limitations of the QSI methodology merit discussion. The QSI methodology was developed during the days of regulation and therefore does not take into account the effect of price on the market share. Also, the market share predictions might be biased upward for European carriers, because the model does not take into account the factors of convenience and customer loyalty. The offerings of the U.S. carriers would be much more convenient than the once-a-day offerings of the European carriers. Also, assuming that the U.S. carriers have been able to develop some sort of customer loyalty through frequent-flier programs, the market share of an additional carrier, particularly a foreign one, on a domestic U.S. route would not be as large as that predicted by the QSI model. Finally, the market share of the foreign carriers on these routes would be limited because of the inherent disadvantages associated with international flights in a domestic market setting. These foreign-carrier-operated domestic flights are only

offered once a day, suffer from worse on-time records, have different staffing and airport needs, and arrive and depart at times that are not convenient for many domestic U.S. passengers.

Thus, although the QSI model predicts the general effects of allowing limited fill-up rights on particular routes, its output is somewhat limited in predictive value, given the important variables omitted. A more useful model would include price, convenience of flight times, and customer preference of domestic carriers. A model developed specifically for the assessment of the effects of cabotage would be extremely beneficial in evaluating potential changes.

The limited ability to predict what carriers might do if allowed limited fill-up rights also complicates prediction. Though Alitalia, KLM, and Scandinavian Air Systems (SAS) do not currently operate beyond-right routes, they might very well begin if limited cabotage were made available. Also, these and other carriers might attempt to develop some sort of international feeder network in the United States. This change might make a difference in the economic viability of certain routes. The QSI methodology, despite these limitations, is useful because it provides a general first cut at predicting what might happen in the event of an altered cabotage policy.

Application to Scenarios

Scenario 1, in which European carriers target their existing route networks, provides a useful starting point for analysis. A review of the international flight schedules of seven major European carriers (British Airways, Lufthansa, Air France, Alitalia, KLM, SAS, and Swissair) reveals that these carriers currently operate 14 beyond-right flights during the course of the year. British Airways operates five such trans-Atlantic flights serving four gateways: London-Philadelphia-Pittsburgh, London-Washington, D.C.-Miami (Concorde flight), London-Los Angeles-San Diego, London-Newark-Philadelphia, and London-Washington, D.C.-Pittsburgh. Air France operates four such flights: Paris-New York-Washington, D.C., Paris-New York-Philadelphia, Paris-Washington, D.C.-Boston, and Paris-San Francisco-Los Angeles. Lufthansa operates one such flight: Frankfurt-Boston-Philadelphia, and Swissair also operates one such flight: Zurich-Boston-Philadelphia.

Alitalia, KLM, and SAS do not offer any flights that have stops in two North American gateways. All of their service in the United States is point-to-point from Europe. The QSI methodology predicts uniformly low market shares for a foreign carrier operating on the above routes (Table 2).

As a secondary approach, a review of the U.S. gateways of European carriers might also reveal the routes that would be targeted by European carriers if they were allowed limited cabotage (Scenario 2). Once limited fill-up rights were made available, it is quite possible that European carriers would simply extend some of the flights that are currently operating, thus adding another leg to the existing service. The ability to transport the cabotage traffic might make an extension economically viable. Assuming that these extensions would originate in existing gateways allows certain routes to be identified as potential targets and analyzed using the QSI methodology. Two gateways were selected for seven European carriers (British Airways, Lufthansa, Air France, Alitalia, KLM, SAS, and Swissair), for a total of 14 additional gateways. A logistically convenient extension was then selected for each of the gateways. The QSI methodology was applied to each of these routes to arrive at a predicted market share for the additional European carrier operation. The results of the exercise also point to extremely low market share for European carriers operating a domestic U.S. route. Again, the QSI methodology predicts uniformly low market shares for the carriers (Table 3).

TABLE 2 PREDICTED MARKET SHARE: SCENARIO 1

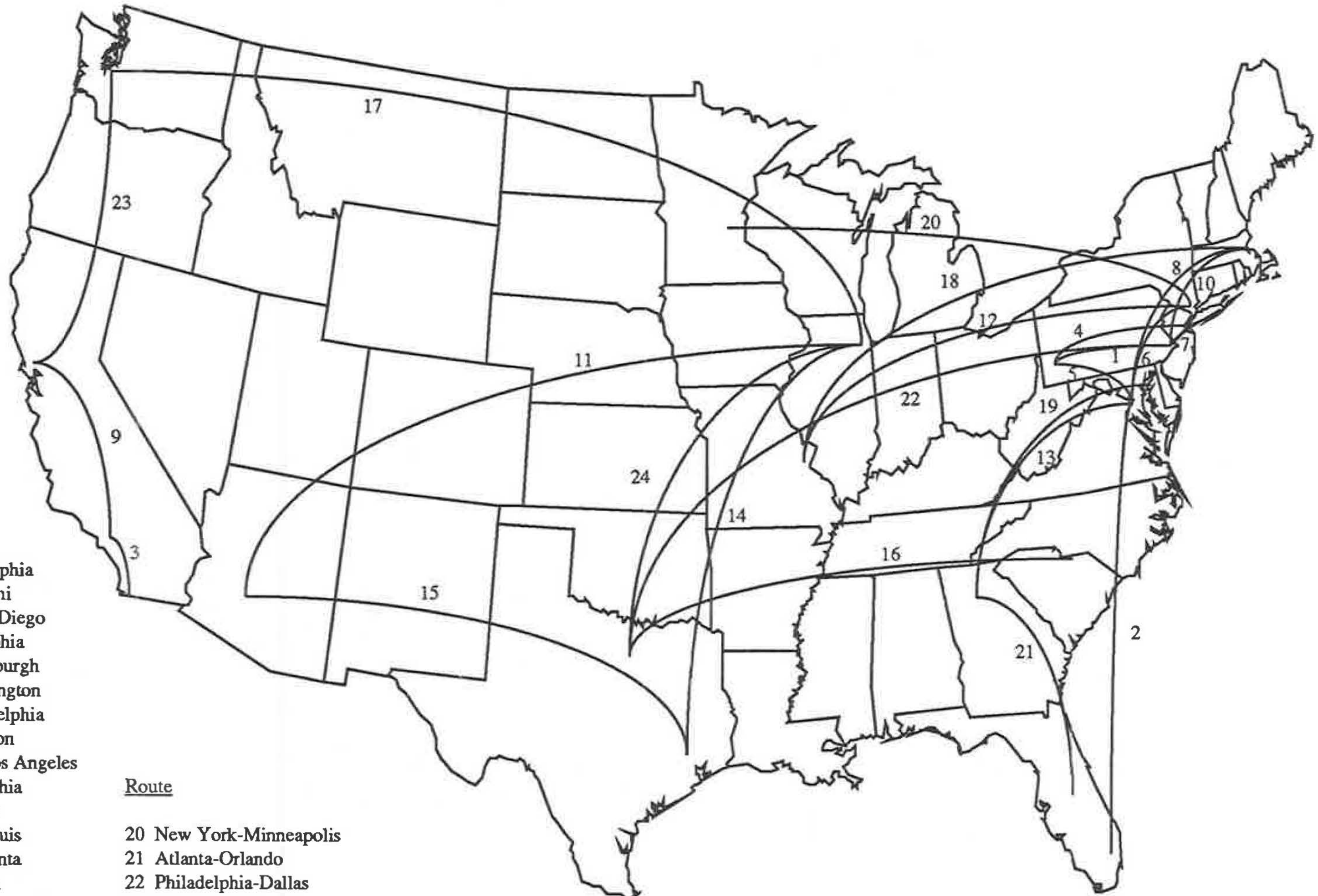
Airline and Route	Predicted Market Share (%)
British Airways	
London-Philadelphia-Pittsburgh	8.99
London-Washington-Miami	5.40
London-Los Angeles-San Diego	1.45
London-Newark-Philadelphia	3.52
London-Washington-Pittsburgh	5.18
Air France	
Paris-New York-Washington	0.88
London-New York-Philadelphia	3.59
Paris-Washington-Boston	2.29
Paris-San Francisco-Los Angeles	0.62
Lufthansa	
Frankfurt-Boston-Philadelphia	3.05
Swissair	
Zurich-Boston-Philadelphia	3.05

Summary of Conclusions

The QSI model, when applied in both of the preceding scenarios, predicts that allowing a very limited form of cabotage to selected routes would result in small market shares for foreign carriers and, of course, slight losses for U.S. carriers. The analyzed were diverse (Figure 1). Of the 24 routes analyzed, the average predicted market share for a single European

TABLE 3 PREDICTED MARKET SHARE: SCENARIO 2

Route	Predicted Market Share (%)
British Airways	
London-Chicago-Phoenix	2.61
London-New York-St. Louis	2.14
Air France	
Paris-Washington-Atlanta	2.71
Paris-Chicago-Houston	2.36
Lufthansa	
Frankfurt-Houston-Phoenix	2.88
Frankfurt-Charlotte-Dallas	4.81
Alitalia	
Milan-Chicago-Seattle	3.37
Milan-Boston-St. Louis	4.00
KLM	
Amsterdam-Baltimore-Atlanta	9.50
Amsterdam-New York-Minneapolis	3.01
Swissair	
Zurich-Atlanta-Orlando	4.89
Zurich-Philadelphia-Dallas	2.76
SAS	
Copenhagen-Seattle-San Francisco	3.35
Copenhagen-Chicago-Dallas	3.05



Route

1 Pittsburg-Philadelphia
2 Washington-Miami
3 Los Angeles-San Diego
4 Newark-Philadelphia
5 Washington-Pittsburgh
6 New York-Washington
7 New York-Philadelphia
8 Washington-Boston
9 San Francisco-Los Angeles
10 Boston-Philadelphia
11 Chicago-Phoenix
12 New York-St. Louis
13 Washington-Atlanta
14 Chicago-Houston
15 Houston-Phoenix
16 Charlotte-Dallas
17 Chicago-Seattle
18 Boston-St. Louis
19 Baltimore-Atlanta

Route

20 New York-Minneapolis
21 Atlanta-Orlando
22 Philadelphia-Dallas
23 Seattle-San Francisco
24 Chicago-Dallas

FIGURE 1 Routes analyzed.

carrier flight was 3.5 percent, hardly the earth-shattering result predicted by many proponents and opponents of cabotage. This highlights the intense competition that a foreign carrier would face on any of the long-haul cabotage routes. Foreign carriers would be likely to pick up enough passengers to fill their one daily service between the two cities, an apparently small impact in most cases. The implications of this level of impact are discussed below.

Costs and Benefits of Long-Haul Cabotage

The value of an altered U.S. cabotage policy must be viewed in light of the potential costs and benefits to both U.S. carriers and consumers. U.S. negotiators have always been faced with the difficult task of forging agreements that recognized both of these competing claims. The issue of cabotage poses significant problems for negotiators, because an altered policy would entail a set of costs and benefits for U.S. carriers different from that for U.S. consumers. An appropriate U.S. policy must get beyond the rhetoric and attempt to realize each set of benefits to the fullest extent possible, realizing that gains in one area might be accompanied by losses in others.

From the standpoint of the U.S. consumer, limited cabotage appears to have minor potential benefits. If long-haul cabotage were allowed, U.S. consumers would benefit from the increased competition through the occasional super-bargain fare, somewhat lower prices offered by domestic carriers on the routes (probably a marginal difference), and increased availability of flights. Though these benefits to the consumer are not likely to be overwhelming, they must be factored in as an advantage of allowing long-haul cabotage. The quantification of these benefits would require a route-by-route analysis of the specific carriers to be affected.

From the perspective of U.S. carriers, the costs and benefits of allowing long-haul cabotage are more difficult to predict. On the cost side, the primary losses to U.S. carriers would occur on the domestic segment. As demonstrated previously in the QSI analysis, the impact on any single carrier is likely to be extremely limited because of the competition that any foreign carrier would face from U.S. carriers. Provided that the grants of cabotage were made strategically, U.S. carriers would have little to fear from the addition of a single flight with extremely limited capacity.

On the benefit side, U.S. carriers might receive additional international routes in exchange for the cabotage rights within the United States. Presumably, by altering its cabotage policy, the United States would open up more fifth-freedom and other international routes to its carriers. These gains would be achieved as along as U.S. negotiators used cabotage as a tool for opening foreign markets to U.S. carriers. The potential for this type of bargaining does exist. This stress on the value of international rights comes at a time in which international routes hold special significance for U.S. carriers. The U.S. Department of Transportation Secretary's Task Force on Competition noted in its report *International Air Service* (*15*) that international air services contributed a disproportionate percentage of the operating profits to U.S. carrier operations, more than 25 percent of the profits but never more than 25 percent in revenue passenger miles. The report noted that higher profit margins of the Atlantic market were particularly important. As noted by Golich, "With airline deregulation forcing razor-thin profit margins domestically, several U.S. airlines have targeted successful international expansion as critical to providing the passenger feed they need to survive" (*16*).

This observation leads to an important caveat: U.S. carriers have differing attitudes toward cabotage. These differences stem from differing corporate strategies, different existing route rights, and variance in the expected returns in the international marketplace. Some U.S. carriers, as noted by James in the address referred to earlier, "are rushing to gain access to Western European gateways before 1992." Although some larger U.S. airlines may support cabotage as a mechanism to expand their access to foreign markets, some smaller, less internationally focused carriers are inclined to believe that the removal of cabotage restrictions would siphon some of their domestic traffic with no commensurate benefits. Also, many opponents of cabotage will probably continue to oppose any form of cabotage because of the "foot in the door" syndrome, the fear that once the barrier has been broken, the U.S. market would eventually be open to foreign competition. Thus, it is not likely that U.S. carriers will reach consensus on the issue. Rather, there will be a gradual shift in the attitude of U.S. carriers who expect to gain internationally and continued resistance by those who plan to remain strictly domestic or those who are already satisfied with their international routes.

CONCLUSIONS AND RECOMMENDATIONS

U.S. aviation officials should view long-haul cabotage as a tool to achieve three basic goals:

1. Realizing limited gains for U.S. consumers;
2. Forcing the hand, in a fairly risk-free manner, of the European carriers and governments that have been demanding cabotage; and
3. Continuing the process of international aviation liberalization, a process that might hold significant rewards for both U.S. carriers and consumers in terms of additional competition and route networks.

To this end, U.S. officials should develop a clear picture of how particular routes might be affected by long-haul cabotage traffic. A better understanding of these potential markets would enable U.S. officials to grant cabotage on routes that are most in need of the additional competition or where U.S. carriers would not be damaged significantly by the additional competition. The analysis in this paper, although limited, does highlight the conclusion that limited cabotage holds neither benefits nor costs of great proportions. This is the realization that should inform future debate and analysis on the issue of cabotage. To exaggerate the implications of the issue is to diminish the constructive role that it might play in future negotiations. Policy discussions of the cabotage issue must develop around the costs and benefits that it realistically implies rather than the exaggerated claims of extreme potential costs and benefits.

To overcome traditional opposition to a change in cabotage policy, U.S. officials might consider an incremental approach in which limited grants with severe restrictions would be followed by more generous offers of cabotage if they proved advantageous to U.S. interests. An agreement with a single country might be viewed as an experiment and, if successful, be applied to additional countries later. Some in the industry believe that limited cabotage might be acceptable if accompanied by restrictions on the nature of the flights and the number of domestic passengers carried on the cabotage legs. If U.S. officials were to make a proposal spelling out these restrictions and stressing the limited nature of the cabotage grant, the path to change might be smoothed.

This strategy would also, presumably, be combined with aggressive attempts to open up the international aviation markets to U.S. carriers. One reason that carriers may be voicing their opposition so loudly is to ensure that U.S. officials do not give away cabotage rights without extracting European route rights in exchange. Long-haul cabotage rights in the United States should be traded on an incremental basis for rights of comparable value around the world. These long-haul cabotage rights might also be used as a way of encouraging the liberalization of international air transport to realize possible longer-term gains for U.S. carriers and consumers. The rights might be granted, for example, exclusively to countries with procompetitive aviation agreements with the United States. In the case of Europe, U.S. negotiators might be able to tie these rights to the continued liberalization of the European aviation market (i.e., until certain changes occurred with regard to subsidizing flag carriers, the U.S. would not grant long-haul cabotage rights to European carriers). These efforts by U.S. officials to open international skies should serve to benefit U.S. carriers and consumers and should not serve as a protective device for selected U.S. carriers.

Finally, the implications of an altered U.S. cabotage policy vis-à-vis other bilateral partners must be considered. Any grant of cabotage rights is important to all U.S. bilateral relationships. Assuming that questions of international law were resolved and a grant of cabotage affected only the U.S. negotiating position, it is quite possible, and even likely, that pressure on the U.S. government to continue to relax its cabotage policy would increase. That fact alone, however, does not diminish the potential value of limited grants of cabotage. If limited experiments with European carriers are successful, other useful experiments might include Pacific Rim countries, particularly on such routes as Honolulu-Los Angeles.

Cabotage appears to be an issue that will be of abiding interest in international aviation. To take full advantage of this interest and of its size and position in global aviation, U.S. officials should treat cabotage as a potential reality. If the grants of cabotage rights are made carefully, tied to equally valuable concessions from other countries, and used to encourage change in other, related areas, long-haul cabotage might very well prove to be a winner for both U.S. carriers and U.S. consumers.

ACKNOWLEDGMENTS

The author expresses gratitude to FAA for sponsoring the Graduate Research Award Program. The author also appreciates the generous assistance provided by John Fischer of the Library of Congress, Vicki Golich of Pennsylvania State University, Mark Diamond of Kurth & Company Inc., and Pablo Mendes de Leon of Leiden University. A special thanks is extended to Tony Gomez-Ibañez of Harvard University for his invaluable support during initial stages of this project. Finally, the patience and assistance of Larry L. Jenney and E. Thomas Burnard of the TRB staff were extremely helpful. Any errors that remain in the manuscript are solely the responsibility of the author.

REFERENCES

1. D. Lewis. Air Cabotage: Historical and Modern-Day Perspectives. *Journal of Air Law and Commerce*, Vol. 45, 1980, p. 1059.
2. *Cabotage in International Air Transport, Historical and Present Day Aspects*. 7-E. Institut du Transport Aerien, Paris, France, 1969, p. 10.
3. Convention on International Civil Aviation, *opened for signature* Dec. 7, 1944, 61 Stat. 1180, Treaties and Other International Acts Series No. 1591.
4. U.S. Congress. *Policies for the Deregulated Airline Industry*. Congressional Budget Office, July 1988, p. 63.
5. J. Koten. Foreign and U.S. Airlines Clash Over Access to Domestic Routes. *Wall Street Journal*, Dec. 7, 1987, p. 27.
6. A. Reed. Europeans Warily Eying 1992 Open Skies Deadline. *Air Transport World*, Dec. 1987, p. 20.
7. F. McGowan and P. Seabright. Deregulating European Airlines. *Journal of Economic Policy*, Oct. 1989, p. 294.
8. EEC Commission. *Civil Aviation Memorandum No. 2*. Doc. COM(84) 72 final, March 15, 1984.
9. M. Feazel. European Court Blocks Bid to Halt Fare Regulation. *Aviation Week & Space Technology*, May 5, 1986, p. 36.
10. When Is 1992? *Avmark Aviation Economist*, June 1989, p. 2.
11. J. M. Feldman. U.S. Could Be Forced to Change Negotiating Stance. *Air Transport World*, May 1988, p. 19.
12. R. Hall. Cabotage. The Grand Illusion. *Air Line Pilot*, Feb. 1989, p. 49.
13. G. Davidson. Don't Reregulate Airlines: Open American Skies to Foreign Competition. *Journal of Policy Analysis and Management*, Vol. 7, No. 4, 1988, p. 714.
14. H. A. Wassenbergh. EEC-Cabotage After 1992!? *Air Law*, Vol. 13, No. 6, Nov. 6, 1988, page 282.
15. Secretary's Task Force on Competition in the U.S. Domestic Airline Industry. *International Air Service*. U.S. Department of Transportation, Washington, D.C., Feb. 1990.
16. P. Proctor. U.S. International Carriers Maneuver to Exploit Transport's Range. *Aviation Week & Space Technology*, March 14, 1988, p. 197.

Human Orientation and Wayfinding in Airport Passenger Terminals

ANTHONY D. ANDRE

Passengers and visitors to major U.S airport terminals frequently complain that they have problems finding their way through the vast and complex structures. Such problems are related to the extent to which it is possible to reach various destinations within reasonable limits of time and effort, which is an important aspect of a terminal building's environmental quality. This paper provides an evaluation of current airport terminal wayfinding systems, focusing specifically on visual information elements, such as signs, maps, and directions. A multiple-method strategy of respondent self-reporting and behavior tracing was used to draw a composite picture of passenger wayfinding information needs and to identify particular problems that lead to poor spatial orientation or wayfinding performance. Significant problems were reported and observed with respect to the design of you-are-here maps and, to a lesser extent, corridor directional signs. These problems are addressed in the context of a theoretical framework for understanding how humans acquire and represent navigational information. This framework, in turn, is used both to evaluate existing terminal designs and to suggest specific guidelines for effective redesign.

Passengers and visitors to major U.S. airports frequently encounter problems of spatial orientation and wayfinding. This may be caused, in part, by the rapid growth of air travel in the last decade and resistance to the construction of new major airports, both of which have greatly contributed to the overcrowding of airport terminals (*1*). Furthermore, the vast structure of the passenger terminal creates a complexity that most airport planners and architects are not equipped to simplify; that is, they lack a formal, theoretical framework for understanding human spatial cognition and for relating its implications to the design of the terminal or its wayfinding system (e.g., signs, maps, directions).

The present research addresses the problem of passenger wayfinding using a multiple-method strategy of respondent self-reporting and behavior tracing both to draw a composite picture of passengers' wayfinding information needs and to identify particular problems that lead to poor wayfinding performance. A psychological, or information-processing, model of how humans acquire and represent spatial and navigational information is then proposed to identify potential design shortcomings of present wayfinding systems and to suggest pragmatic, feasible design guidelines for effective redesign.

PASSENGER ORIENTATION AND WAYFINDING

> The very word "lost" in our language means more than simple geographical uncertainty; it carries overtones of utter disaster (*2*).

Institute of Aviation, Aviation Research Laboratory, University of Illinois, Urbana-Champaign, Ill. 61820.

The term *wayfinding* describes a person's abilities to reach spatial destinations and refers to one's static relation to space as well as the dynamics involved in his or her purposeful mobility (*3*). Accordingly, two variations of lostness may occur; lostness in orientation—knowing where you are, but not where you are going—and lostness in location—knowing where you are going, but not where you are. Simply stated, the combined effects of both variations represent the degree to which a person deviates from the most direct route to a desired location (*4*).

The susceptibility to problems of spatial orientation and wayfinding reflects the extent to which it is possible for passengers to reach various destinations within reasonable limits of time and effort, which is an important aspect of a terminal building's environmental quality (*5*). Indeed, the longer the time that passengers spend waiting, walking, or trying to find a particular facility, the lower the perceived level of service (*6, 7*). Moreover, poor spatial orientation often causes considerable annoyance and psychological stress (*8*).

As the burden of airport finance shifts from federal government to local government, airport operators will become, or want to become, more competitive in attracting travelers through their terminals. Logically, a common way to attract travelers is to upgrade the level of service at a terminal by increasing or improving the services and facilities that are offered (e.g., shops, restaurants, and business services). Yet, overriding these aspects of terminal design, providing another "level" of service, is the appropriate location of facilities in the terminal area and their quick identification by the passenger (*9*). Indeed, this aspect of the terminal design may also have a major impact on capital and operating costs, because passengers can only spend money at facilities they can locate.

INFLUENCE OF PHYSICAL ENVIRONMENT

> Environments . . . do not really function properly unless users are able to find their way around (*5*).

A total wayfinding system comprises many elements (*10*), but wayfinding ability is most affected by the influence of the physical environment (*11*). Therefore, in order to forecast wayfinding problems, physical variables that are likely to affect the ease with which spatial orientation and wayfinding are accomplished must be specified (*5*). Garling et al. (*5*) have proposed three such variables: the degree of differentiation, the degree of visual access, and the complexity of spatial layout.

The degree of differentiation refers to the degree to which different related parts of the environment look the same and unrelated parts are distinguishable. A high degree of differentiation should enhance spatial orientation and wayfinding by making parts of the environment more easily recognizable.

The degree of visual access refers to the extent to which different parts of the environment are visible from many vantage points. High visual access will increase both the recognition and localization of distant destinations, thereby improving spatial orientation and wayfinding (*12*). Seidel (*11*) reports some evidence that passengers arriving at a gate with direct visual access to the baggage claim area have an easier wayfinding experience.

The complexity of the spatial layout will also affect spatial orientation and wayfinding (*13*). The spatial layouts of environments such as the airport passenger terminal are naturally complex because of their sheer size and the large number of possible destinations and routes.

Much research has approached the issue of passenger orientation and wayfinding under the assumption that its efficiency is a factor of space and time constraints (*7*). In other words, it is assumed that the more space passengers have, and the less distance they need to walk, the greater the ease of mobility. There is growing reason to believe, however, that these parameters fall short of explaining passengers' abilities and attitudes toward circulation in the terminal. In fact, a survey of 30,000 passengers conducted by Condom (*15*) found that availability of space is not a variable of fundamental importance.

More recently, Martel and Seneviratne (*1*) have suggested that several variables besides availability of space and waiting time should be considered when passenger terminal performance is evaluated in terms of passenger wayfinding needs. In their study, 53 percent of the respondents to a personal interview survey perceived information as the most significant variable affecting their ability to circulate in the terminal effectively, whereas 38 percent chose walking distance as the most critical variable. Fewer than 10 percent of the respondents, though, chose space availability and level changes as most important. In a study of passenger information needs in subway systems, Beck (*16*) also notes the critical role that guidance information plays in facilitating the efficient movement of passengers through transit facilities.

IMPORTANCE OF INFORMATION

Of course, improvements in the three physical characteristics may be achieved by varying the size, form, or architectural style of the passenger terminal environment (*13*). But perhaps the most logical and feasible alternative under the present circumstances would be to increase efficiency through better management of existing facilities and resources for aiding passenger orientation and wayfinding (Seneviratne and Martel, unpublished data). Thus, the focus shifts from physical form to information. Guidance information sources, in the form of signs and maps, must then also be taken into account, especially when one attempts to predict the wayfinding difficulties that newcomers will have in any specific environment (*5, 14*).

Information, however, is a rather broad term, and thus could reflect any number of guidance sources (*1*). For example, one can minimize walking distance and level changes if the appropriate information is available to direct the passenger to the correct destination (*1*). It is seemingly important, then, to define and categorize the various types of information available to passengers. Here, two general forms of guidance information are discussed: visual guidance signs and verbal route directions.

Visual Guidance Information Sources

The visual guidance system can be broken down into two main components: corridor guidance signs and airport map displays.

Corridor Guidance Signs

Corridor guidance signs usually take the form of directional arrows. One problem with these signs, however, is that they are trying to portray three-dimensional information on a two-dimensional or flat surface. This often causes ambiguity as to whether a sign is pointing up or straight ahead, leaving the passenger guessing as to the terminal layout ahead.

Airport Maps

Airport maps are publicly displayed pictorial or schematic layouts of the terminal and may be stationary or portable. Stationary maps, often in the form of the you-are-here design shown in Figure 1, are placed in large display cases so that passengers and visitors can view them easily. Portable maps, so called because passengers can carry them as they navigate through the terminal, are available in most airline magazines or at the airport information and customer-service booths.

Verbal Directions

Many passengers, either confused by the visual guidance information or simply lacking the time needed to interpret their relative location accurately from a map, consult airport personnel for directions to their desired designations. In most large airports, information booths or customer-service centers are operated by both airport managers and airlines. In addition, customer-service representatives are placed throughout the terminals, and often at arriving gates, in an effort to assist passengers.

EXPERIMENTAL EVALUATION OF AIRPORT WAYFINDING SYSTEM

Clearly, the design of any terminal should not proceed without knowledge of the mix of passenger traffic envisaged, because this knowledge can be used by airport owners to tailor the design of the terminal to the needs of the passengers (*17*). Moreover, any effort aimed at improving the quality of the wayfinding systems in passenger terminal buildings should be based on an understanding of the critical factors that underlie

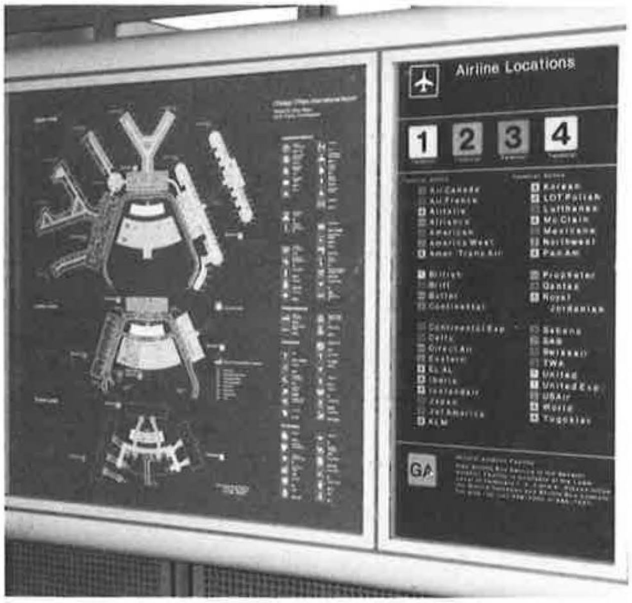


FIGURE 1 Typical you-are-here map display.

passenger performance and satisfaction. Accordingly, in the present study an evaluation was undertaken to (a) identify which guidance information sources are most important to passengers while they are navigating the terminal and which are perceived as needing the most improvement in their design or implementation; (b) determine the passengers' ability to use these sources; and (c) provide feasible guidelines for the optimal design of guidance information.

The experimental data were collected in three stages through:

1. an analysis of requested information at airport information booths;
2. a personal interview survey, which was used to determine passenger attitudes toward the importance of various types of directional information and the need for improvements in their design; and
3. interviews with, and observations of, "lost" passengers as they consulted a you-are-here map display.

All data were collected at Chicago's O'Hare International Airport, Terminals 1 and 2, over a three-month period, April through June 1990.

Experiment 1: Verbal Guidance Analysis

An analysis of the nature of information requested at the information booths located in Chicago's O'Hare International Airport was conducted on data obtained for 1989, using the SAS statistical package. These booths are intended to provide more in-depth information than that offered by the visual information systems (*18*). Figure 2 shows the relative proportions and the nature of the information requested during 1989 from the four information booths located in the airport. As shown, of the 1,046,957 persons who requested information during this year-long period, the majority (74 percent) asked questions involving directions to various terminal facilities ($p < .05$). Not surprisingly, a smaller but significant

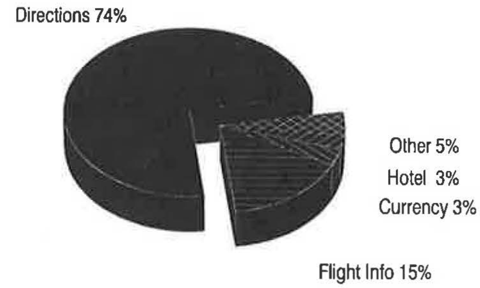


FIGURE 2 Analysis of passenger requests at O'Hare International Airport information booths.

proportion (15 percent) requested information pertaining to the scheduled arrival or departure of various flights ($p < .05$). In addition, a survey of 10 airport and airline customer-service representatives revealed that more than 90 percent of their requests are for directional information.

Given that few of these passengers missed their flights, these findings suggest that information booths, and the verbal route directions they provide, are an important navigational aid to many passengers. Of course, it is difficult to tell whether this finding demonstrates that the terminal architecture and the visual guidance system are not adequately designed to address the information needs of passengers navigating the terminal. It may, instead, merely demonstrate that passengers do not devote time to the study of maps and signs, and opt instead to ask for verbal route directions. In either case, though, one can argue that a more efficient visual guidance system would not only assist those passengers who rely on maps and signs but also alleviate the need for passengers to consult the information booths. Indeed, such an improvement would potentially provide a substantial cost reduction for airport operators and airline owners, who carry the burden of providing customer-service representatives at information booths and throughout the airport terminals.

Experiment 2: Personal Survey

Various terminal passengers were administered a personal survey that consisted of three parts. The survey sought to obtain

1. Demographic information (age, sex, nationality, residence) and information relating to the purpose of the trip (business or leisure);
2. Information on the frequency of travel from O'Hare (infrequent or frequent), the mode of arrival at the airport (local or transfer), and the time of departure; and
3. Passengers' ratings of the relative importance of four directional information aids (corridor signs, airport or magazine carry maps, you-are-here map displays, and verbal directions) and of the need for any corresponding improvements.

The personal survey was completed by 118 passengers in various segments of the terminals.

Procedure

After providing demographic and type-of-travel information, participants were instructed to rank order (1 to 4) the variables

by their relative importance for use in navigating the airport, with a rank of 1 signifying the variable as the most important. Subjects were also asked to rank order (1 to 4) the variables by the need for improvement in their design or implementation, with a rank of 1 now signifying a variable as needing the most improvement.

Results

The passenger responses to the survey were coded and analyzed using the SAS package. The variables sex, age, nationality, and trip purpose were not included in the analysis, which focused primarily on the frequency of travel out of, and, hence, familiarity with, the airport. Of the 118 respondents, 56 (47 percent) identified themselves as infrequent passengers and 62 (53 percent) identified themselves as frequent passengers. An analysis of variance (ANOVA) procedure was performed to determine the significance of the difference between the mean ranks of the variables.

Figure 3 plots the percentage of passengers, by frequency of travel through O'Hare, who chose each of the four variables as the most important. It is apparent from these data that the majority of infrequent travelers perceived either the corridor signs or the you-are-here maps as the most important guidance-information sources but that frequent travelers chose only the corridor signs as most important. This suggests that with increasing exposure to the airport, passengers rely more on the corridor signs and less on the you-are-here maps. An ANOVA performed on the mean ranking showed these trends to be significant ($p < .05$).

Figure 4 plots the percentage of passengers, by frequency of travel through O'Hare, who chose each of the four variables as needing the most improvement. The results suggest that the you-are-here maps provide the most significant source of confusion for infrequent passengers, and hence are most in need of improvement ($p < .05$). Corridor signs and verbal directions are also noted, although to a lesser degree, as in need of improvement. Frequent passengers rate the corridor signs as most in need of improvement ($p < .05$). The you-are-here maps and verbal directions are also noted, although to a lesser degree, as in need of improvement.

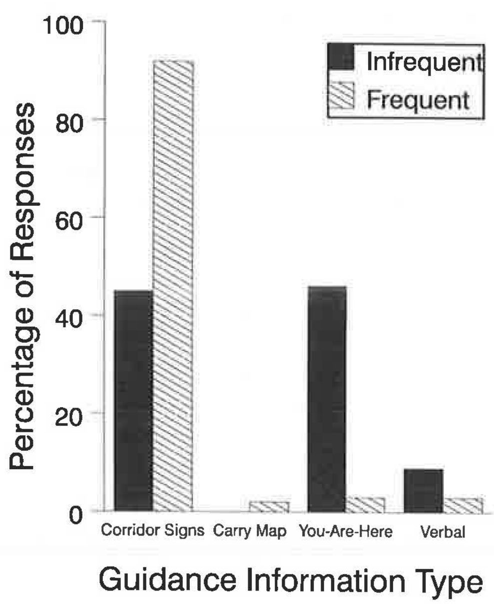


FIGURE 3 Effect of airport exposure (frequency) on passengers' perception of the most important directional information sources.

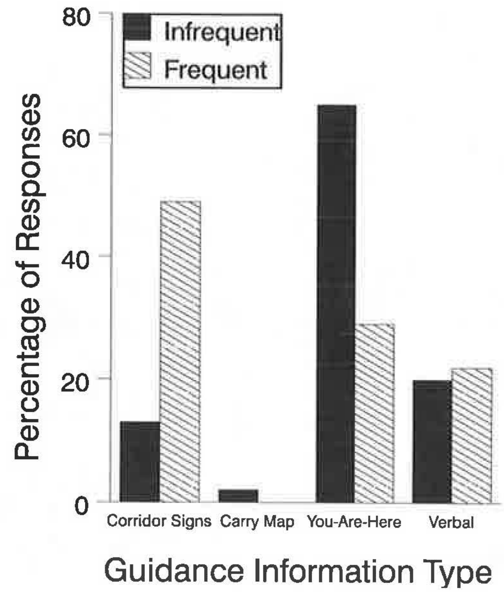


FIGURE 4 Effect of airport exposure (frequency) on passengers' perception of the directional information sources most in need of improvement.

Discussion

A comparison of Figures 3 and 4 provides a rather informative and consistent picture. Those passengers unfamiliar with the airport rated both corridor signs and you-are-here maps as important sources of guidance information, but also suggested that these sources, in addition to the verbal directions provided by the airport personnel, are in need of substantial improvement. Those passengers more familiar with the airport rated only the corridor signs as important sources of guidance information, yet they cited the same set of information sources (i.e., corridor signs, you-are-here maps, and verbal directions) as in need of improvement. Perhaps this latter finding reflects the business travelers' previous experiences with you-are-here maps and verbal directions at a time when they were less familiar with the airport layout.

Experiment 3: Passenger Performance with the You-Are-Here Map Display

The results of Experiment 2 suggest that newcomers rely extensively on you-are-here maps, yet they also suggest that such passengers, along with more experienced ones, believe that you-are-here maps are most in need of improvement. Previous research, however, has shown that subjective ratings sometimes differ from objective performance (*19*). The objective of Experiment 3, therefore, was to determine what

proportion of those passengers trying to use the you-are-here maps could do so effectively.

To meet this objective, a final set of data was collected as passengers viewed one of the airport's you-are-here map displays (see Figure 1). These maps represent a static view of the airport layout, with Terminal 2 always located at the top of the map display. More than 90 of these displays are located through O'Hare International Airport.

Over a 3-hr period, data were collected from all of those passengers (n = 19) who consulted the you-are-here map. Respondents were categorized in one of two ways:

1. Successful, if they were able to determine the appropriate heading and direction to the desired facility, or
2. Unsuccessful, if they could not.

All 19 passengers identified themselves as being unfamiliar (i.e., newcomers) with O'Hare Airport.

Results

Figure 5 plots the passengers' ability to determine the proper navigational course successfully, as well as to identify the specific terminal facility in question. Clearly, the results are discouraging, because only 2 of the 19 passengers were able to determine the correct course of action from their study of the you-are-here map display. In fact, many of the passengers had trouble just determining where they were located on the map in relation to the terminal. Frequent comments included, "I'm not even sure where I am on this map" and "Now that I know where it is, how do I get there?"

Discussion

The results of Experiment 3 suggest that the majority of passengers who attempted to use the you-are-here maps could not do so. Thus, the passengers' subjective rating of the poor

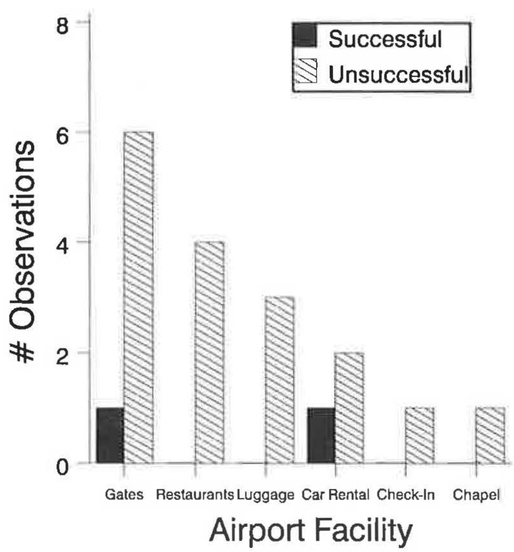


FIGURE 5 Analysis of passenger performance with you-are-here maps.

design quality of these maps corresponded rather well with their inability to use them effectively. It should be noted, however, that although suggestive, the reliability of these findings is limited because of the small sample size obtained.

THEORY AND GUIDELINES

Collectively, the experimental results obtained suggest that an overwhelming majority of passengers perceive the corridor directional signs and the you-are-here map displays to be the most important sources of directional information; yet, at the same time, they note substantial flaws in their design or implementation. Furthermore, the results show that inexperienced passengers, unfamiliar with the airport terminal, rely less on the corridor signs and more on the you-are-here maps than do experienced passengers.

These results are consistent with those of Martel and Seneviratne (*1*), who found that for efficient circulation in the airport terminal, business travelers perceive walking distance (i.e., time) as the most important variable, whereas leisure travelers feel that information (for guidance) is the most important variable. The authors conclude that "these differences seem to reflect the variance in the value of time for the two groups, as well as the variance in familiarity with the airport" (*1*).

Clearly, the efficient movement of passengers through transit facilities should be the primary concern of terminal planners (*16*). That is, the internal guidance-information system of the passenger terminal should be simple to follow and easy to negotiate. But this rather intuitive human factors aspect of design does not seem to have been given proper consideration in the preparation of most passenger terminals (*9*), including the one evaluated here.

Why is the design process seemingly devoid of human factors? Perhaps bringing in another group of experts would undermine the architects' autonomy (*20*). A more likely reason, however, is the planners' and architects' concerted lack of understanding of the way humans acquire and represent navigational information and the optimal formats for presenting this information. For without this knowledge, planners and architects have only their common sense to assist them with the difficult task of minimizing passenger disorientation and confusion while they circulate in the airport terminal.

Theoretical Approach to Design of Guidance Information Systems

Although it is undoubtedly important to highlight systematic problems experienced by passengers as they attempt to navigate the terminal, it is equally important to illustrate how these problems may be overcome through redesign. Such an endeavor naturally falls within a human factors approach, whereby the information needs of the passengers are assessed and subsequently serve as the basis for proposed designs. A good theory, relating these factors, can be used to identify the consequences of proposed designs early in the design process, when various alternatives are still being generated.

Thorndyke (*21*) proposed that, as people become increasingly familiar with a geographical environment, the nature of

their knowledge of that environment undergoes qualitative as well as quantitative changes (*22*). The qualitative changes are characterized by a progression through three levels of knowledge.

Initially the representation is characterized by landmark knowledge. Here, orientation is guided exclusively by highly salient visual landmarks (e.g., statues, buildings, restaurants) that provide little more than a crude representation of the environment.

One soon progresses from landmark knowledge to route knowledge (i.e., one acquires the ability to navigate between points). This understanding is expressed from an ego-centered frame of reference whereby landmarks or other visual features seen while navigating the environment generate the decision to turn left or right or continue straight ahead at a given intersection (*22*).

Sufficient navigational experience eventually provides one with survey knowledge. Here, the knowledge resides in the form of an internalized "cognitive map" (*23*), which is analogous to the true physical map of the environment. This representation is expressed from a world-centered frame of reference, so one navigates with a top-down perspective of the layout of the environment. Hence, at this level one is able to describe the relative location of two landmarks even though one may never have traveled a route that connects them.

Thus, beyond describing two phases of navigational learning, route and survey knowledge forms may be contrasted in their "canonical" or preferred frame of reference. During early acquisition, one navigates using route knowledge from an ego-centered frame of reference, and thus one's representation of the environment corresponds directly to what one sees as he or she follows a route. With survey knowledge, on the other hand, one navigates from a world-centered frame of reference, and thus the internal map is independent of the particular view one has of the environment.

As Thorndyke (*21*) has proposed, a logical consequence of this difference is that possession of route knowledge is optimal for judgments made from one's own frame of reference. These would include such tasks as pointing to a given part of the airport that is not visible (orientation) or judging the actual walking distance that must be traveled between two points and actually navigating that route. For example, the instruction "turn right" (a command based on route knowledge) will lead to different actions, depending upon whether one is facing north or south. In contrast, "turn westward" (a command based on survey knowledge) will lead to the same ultimate action, independent of the initial orientation.

The progression from route to survey knowledge with training suggests that the internal model slowly progresses from an ego-centered, context-dependent representation to a world-centered, context-free representation. Evidence for this progression is provided by the observation that experts in map reading tend to orient maps in a "north-up" or fixed direction, whereas novices tend to rotate the map in the direction ("track-up") they are heading (*24*).

Design Guidelines

Maps

Several experiments have shown that the mental representation of a paper map is analogous to the physical map itself (*25, 26*), and as previously discussed, that inexperienced travelers navigate according to their own frame of reference (i.e., "up" in the direction they are heading). However, if the environment is not aligned with the map, the traveler must then perform some time-consuming and error-prone mental operation (i.e., mental rotation) to bring the map and the environment into correspondence. An optimal map design, therefore, would consider the orientation of the passenger as he or she views the display, and would provide a map whose orientation (i.e., frame of reference) is congruent with the passenger's current orientation in the terminal.

The implication of the previous discussion is that newcomers would benefit when the map was aligned in the same direction they were heading (i.e., track-up), whereas experienced passengers would prefer fixed maps. Clearly, however, this guideline stands in contrast to the airport map display evaluated in this study. As shown in Figure 1, a fixed-map orientation was used, with Terminal 2 always at the top of the map. It is therefore not surprising that passengers had difficulty locating their relative position on the map and ascertaining the appropriate navigational route to the terminal, facility, or gate in question. Moreover, because the data suggest that only inexperienced travelers rely extensively on you-are-here map displays, these travelers should be provided with the map format that is most compatible for them. Track-up maps not only would improve their orientation and wayfinding but also would facilitate their progression from route knowledge to survey knowledge as they became more familiar with the terminal environment.

In a detailed account of map-design issues, Levine (*26*) suggests the following additional guidelines:

- The map should be placed in some asymmetric location, preferably close to some prominent landmark, to facilitate the observer's locating himself or herself on the map. The map orientation could be indicated symbolically as aligned properly with respect to the environment by means of a horizontal line at its appropriate location on the map, and the observer in front of the map symbolized with an upward-pointing you-are-here symbol.
- Facilities and services (restaurants, restrooms, information booths, etc.) should be indicated on the map either by visual shape or by some type of symbol. A written description would be less salient and more difficult for foreign travelers to understand.
- Color should not be used in way that would contradict normal stereotypes or meanings with which certain colors are known to be associated (e.g., red—emergency/help; blue—sky/up).

Corridor Directional Signs

The previous discussion on map design shows that it is essential that signing be designed carefully and that careful consideration be given to guidelines that oppose each other. On the one hand, the passenger must be given sufficient signage to find the facility or the direction sought. On the other, there must not be such a proliferation of signs that there is confusion.

More specifically, Beck (*16*) suggests three guidelines for the design of directional signs in transit terminals.

Directional Association. Directional information should be placed in a way that will be easily and unambiguously associated with the pathway choice the user must make. Thus, designers must be aware of both where signs are located and the location from which they can be seen.

Message Content. The message content of a directional sign should be in a form that is easily understood and useful to the user. Designers must ensure that only relevant information is displayed and that combinations of displays are avoided, when possible. Combinations of displays introduce the possibility that display elements will interact in ways not intended for the task by the system designer. As an example, the display shown in Figure 6 was found in a U.S. airport and reported by Kantowitz et al. (*27*). One can see how a passenger might be confused about the correct association of gate numbers to directional arrows. For this example, the difficulty is easily solved by the introduction of an appropriate line (either vertical or horizontal, depending on whether gates 1 to 5 are to the left or the right) between the display elements (*27*).

Redundancy. Because of the built-in limitations in human short-term memory, a certain amount of redundancy of information is necessary. This may be in the form of a visual icon or symbol complementing a written description or confirmation signs that let the passenger know that he or she has chosen the correct path.

Terminal Architecture

Finally, although not the focus of the present paper, it is acknowledged that planning an airport terminal in such a manner that its internal layout minimizes possible disorientation significantly contributes to passenger satisfaction (*9*, *28*). Moreover, passengers will not be likely to be able to use the terminal's facilities and appreciate its architectural features if they cannot easily navigate it. Hence, the architectural design of passenger terminals should, ideally, be structured so that the recognition and localization of its inner facilities are continually apparent to its users.

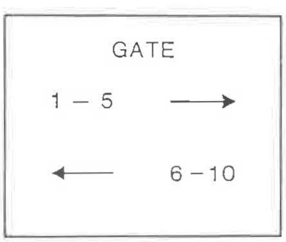


FIGURE 6 In which direction should one turn to reach Gate 4? (*27*)

CONCLUSION

This research has attempted to determine the significant sources of information that passengers use while orienting themselves to various facilities within and between airport terminals, how successful those sources are, and what design qualities facilitate or inhibit the passenger's wayfinding abilities. Perhaps not surprisingly, the current results suggest the need for vast improvements in the design and implementation of guidance information in airport passenger terminals. Specifically, it was found that inexperienced passengers rely more on airport maps than do experienced passengers, yet they both find these same maps difficult to interpret or use for navigation.

A theoretical review of human navigation and spatial cognition indicated that people unfamiliar with an environment navigate from an ego-centered (i.e., self) reference. By providing geographical information congruent with this vantage, wayfinding performance can be improved substantially. An evaluation of the you-are-here maps at O'Hare International Airport indicates that all of these maps show an identical, fixed (Terminal 2-up) orientation, regardless of their placement in the terminal. Previous research has revealed severe wayfinding decrements for misaligned maps (*26*, *29*), thus explaining the wayfinding problems observed by inexperienced passengers in the present study. It is, therefore, imperative that you-are-here maps, such as those studied here, are designed with previous knowledge of their location within the environment. To carry out this process in the reverse—first designing the map and then looking for its ideal placement—would leave the ultimate alignment of the map to chance (*29*).

Time spent in a passenger terminal is an increasingly important component of the overall travel experience for the passenger. This time, however, is often spent in a state of anxiety, confusion, and disorientation because of the complex architecture and dysfunctional guidance-information systems inherent in many airport passenger terminals. It is therefore important that terminal planners and designers make every possible effort to enact a total guidance system, including architect, signs, maps, and personnel, that facilitates rather than impedes the passengers' orientation process. Furthermore, the design of the system should ensure that these elements are mutually reinforcing (*10*).

Navigating an unfamiliar environment is a formidable task for many. Improving the visual wayfinding system by providing signage at choice points and placing you-are-here maps so that they are aligned congruent with one's forward view would significantly improve the wayfinding abilities of passengers unfamiliar with the airport terminal.

In conclusion, this paper has attempted to bridge the gap between theoretical and applied issues of human orientation and wayfinding in airport terminals. In doing so, these findings attempt to illustrate why the terminal-design process in general, and the design of terminal guidance-information displays in particular, should be guided by the needs of passengers and should therefore reflect an awareness of their attitudes and behavior. For information to be of use to architects, social scientists must have useful and relevant knowledge to contribute a form that is meaningful and timely for designers. It is hoped that the present research shows that many applied-design problems can be answered, or at least guided, through

the process of generalizing and applying results of theoretically based research.

ACKNOWLEDGMENTS

This research was supported through the Graduate Research Award Program on Public-Sector Aviation Issues administered by TRB and sponsored by FAA. The author would like to thank Jefferson M. Koonce for his valuable support and knowledge of and enthusiasm for this project. Appreciation and thanks also go to Thomas Burnard, Larry Jenney, Andrew Lemer, Dennis Mewshaw, and Sally Liff for volunteering their time and expertise in overseeing the nature and course of the project and for thoughtfully reviewing the manuscripts. Finally, the author would like to thank Susan Smolinski and the multilingual staff at O'Hare International Airport for providing data and valuable insight, and Karen Andre for her help, love, and support.

REFERENCES

1. N. Martel and P. N. Seneviratne. Analysis of Factors Influencing Quality of Service in Passenger Terminal Buildings. In *Transportation Research Record 1273*, TRB, National Research Council, Washington, D.C., 1990, pp. 1–10.
2. K. Lynch. *The Image of the City*. MIT Press, Cambridge, Mass., 1960.
3. R. Passini. Spatial Representations: A Wayfinding Perspective. *Journal of Environmental Psychology*, Vol. 4, 1984, pp. 153–164.
4. G. Best. Direction Finding in Large Buildings. In *Architectural Psychology* (D. Canter, ed.), Royal Institute of British Architects, London, 1970.
5. T. Garling, A. Book, and E. Lindberg. Spatial Orientation and Wayfinding in the Designed Environment. *Journal of Architectural Planning Research*, Vol. 3, Jan. 1986, pp. 55–64.
6. N. Ashford. Level of Service Design Concept for Airport Passenger Terminals—A European View. *Transportation Planning and Technology*, Vol. 12, 1988, pp. 5–21.
7. J. J. Fruin. *Pedestrian Planning and Design*. Metropolitan Association of Urban Designer and Environmental Planners, Inc., New York, N.Y., 1971.
8. C. M. Zimring. Stress and the Designed Environment. *Journal of Social Issues*, Vol. 37, No. 1, 1981, pp. 145–171.
9. S. K. Modak and V. N. Patkar. Transport Terminal Design and Passenger Orientation. *Transportation Planning and Technology*, Vol. 9, 1984, pp. 115–123.
10. J. R. Carpman, M. A. Grant, and D. A. Simmons. Hospital Design and Wayfinding. *Environment and Behavior*, Vol. 17, No. 3, May 1985, pp. 296–314.
11. R. S. Seidel. Way-finding in Public Spaces: The Dallas/Fort Worth, USA Airport. *Environmental Design Research Association*, Vol. 14, 1983, pp. 129–138.
12. T. Garling, E. Lindberg, and T. Mantyla. Orientation in Buildings: Effects of Familiarity, Visual Access, and Orientation Aids. *Journal of Applied Psychology*, Vol. 68, 1983, pp. 177–186.
13. G. W. Evans, C. Smith, and K. Pezdek. Cognitive Maps and Urban Form. *Journal of the American Planning Association*, Vol. 48, 1982, pp. 232–244.
14. J. Weisman. Evaluating Architectural Legibility. *Environment and Behavior*, Vol. 13, 1981, pp. 189–204.
15. P. Condom. Thirty Thousand Passengers State Their Preferences. *Interavia*, Vol. 11, 1987, pp. 1177–1179.
16. R. Beck. Designing for Passenger Information Needs in Subway Systems. *ITE Journal*, Jan. 1986, pp. 17–24.
17. N. Ashford and P. Wright. *Airport Engineering*, 2nd ed. John Wiley and Sons, Inc., New York, N.Y., 1984.
18. N. Ashford, H. P. M. Stanton, and C. A. Moore. *Airport Operations*. John Wiley and Sons, Inc., New York, N.Y., 1984.
19. Y. Y. Yeh and C. D. Wickens. Dissociation of Performance and Subjective Measures of Workload. *Human Factors*, Vol. 30, 1988, pp. 111–120.
20. R. Sommer. *Social Design. Creating Buildings with People in Mind*. Prentice-Hall, Englewood Cliffs, N.J., 1983.
21. P. W. Thorndyke. *Performance Models for Spatial and Locational Cognition*. Technical Report R-2676-ONR. Rand Corporation, Washington, D.C., 1980.
22. C. D. Wickens. *Engineering Psychology and Human Performance*. Scott, Foresman & Co., Glenview, Ill., 1984.
23. E. C. Tolman. Cognitive Maps in Rats and Men. *Psychological Review*, Vol. 55, 1948, pp. 189–208.
24. J. R. Anderson. *Cognitive Psychology*. Academic Press, New York, N.Y., 1979.
25. A. J. Aretz. Spatial Cognition and Navigation. In *Proceedings of the 33rd Annual Meeting of the Human Factors Society*, Human Factors Society, Santa Monica, Calif., 1989.
26. M. Levine. YOU-ARE-HERE Maps: Psychological Considerations. *Environment and Behavior*, Vol. 14, 1982, pp. 221–237.
27. B. H. Kantowitz, T. J. Triggs, and V. E. Barnes. Stimulus-Response Compatibility and Human Factors. In *Stimulus-Response Compatibility* (R. W. Proctor and T. G. Reeve, eds.), North Holland Publishing, New York, N.Y., 1990, pp. 365–388.
28. J. P. Braaksma and W. J. Cook. Human Orientation in Transportation Terminals. *Transportation Engineering Journal of ASCE*, Vol. 106, 1980, pp. 189–203.
29. M. Levine, I. Marchon, and G. Hanley. The Placement and Misplacement of You-Are-Here Maps. *Environment and Behavior*, Vol. 16, No. 2, March 1984, pp. 139–157.

Understanding the Role of Human Error in Aircraft Accidents

DANIEL J. BERNINGER

The commercial aviation industry has achieved an enviable record of safety, but accidents still occur. In roughly two-thirds of aircraft accidents, aviation's human link receives the blame, and the proportion of accidents attributed to human error has not changed appreciably in 20 years. Most human error that leads to accidents surfaces in the performance of flight crews and air traffic controllers. The strategies used to address human error can be placed in two categories: introduction of technology that reduces the role of humans in the system and changes to the system and training suggested by human factors considerations. The pursuit of these approaches has largely become distinct, but they are both characterized by several basic assumptions. Both technologists and human factors specialists attribute human error to human fallibility and accept in varying degrees the inevitability of human error. Both accept the notion that humans are the most unreliable element in aviation. Both place emphasis on flight crews and air traffic controllers. Supporters of both approaches hold doubts as to the value of the other; in particular, the technologists view human factors as being too untidy to be the basis of design. The system that fails in an aircraft accident can be divided into animate (human) and inanimate components. If assumptions are reconsidered, there are mechanisms by which the inanimate system can contribute to causing the human error that leads to accidents. There is a spectrum of possible accident causes between the extremes of entirely human error or entirely inanimate system malfunction. Current interventions are heavily weighted toward the human error end of the spectrum, but this paper suggests an additional approach to interventions that alleviates system problems that cause human errors.

The commercial aviation industry has achieved an enviable record of safety, but accidents still occur. The distribution of accident causes for the world jet air carrier fleet from 1960 through 1981 is as follows (*1*):

Causal Factor	*Percent*
Cockpit crew	70–75
Airframe, power plants, systems	13
Maintenance	3
Weather	5
Airport, air traffic control	4

The proportion of accidents attributed to cockpit crew error has not changed appreciably since these figures were assembled (*2*). The greater than 70 percent of accidents attributed to human error seems to be a strong indictment of human performance and a powerful motivator for research that would reduce the frequency of human error.

Within the field of human factors there are numerous efforts to prevent the commission of errors in aviation, as recently documented or proposed in *The National Plan for Aviation Human Factors*, jointly developed by FAA and other agencies (*3*). For example, cockpit resource management (CRM) is a philosophy that seeks to enhance the effectiveness of the cockpit team through increased cooperation. CRM is already integrated into the pilot training of most major airlines, and, for example, United Airlines acknowledged the contribution of CRM in minimizing the loss of life in the July 1989 accident in Sioux City in which the flight crew was forced to improvise procedures to cope with loss of directional control.

By far the largest effort to prevent human error (measured by funding, personnel, or industry support) is through automation that seeks to reduce the role of humans. These efforts are motivated by the availability of powerful new information-processing technologies and the premise that reducing the human element will reduce error. The most recent generation of large air transports incorporates automation for flight control, augmentation, and management (ascent, cruise, flare, and landing), as well as automatic throttle control and fuel load management. The National Airspace System Plan proposes to automate many air traffic control functions (*4*). For example, the automated en route air traffic control system will automatically detect and advise controllers on potential airspace conflicts; the controller would no longer be required to detect conflicts through mental projections of aircraft trajectories (*5*).

Human factors is receiving increasing attention in the design of systems, especially automation systems, but there remains a lack of communication between the engineering and human factors disciplines. System development (automation or otherwise) requires the integration of a number of specialties. Most systems interact with humans on some level (design, manufacturing, operation, or maintenance). Thus, human factors should be a consideration, but it is not now a major one. If human factors is addressed at all, it is usually through specialists who act as advisers to the design team. As outsiders, these specialists have difficulty influencing the final product. Often they are called in after the fact to approve the design, when changes are impractical. In contrast, the concerns of electrical, mechanical, and aeronautical engineers, among others, are not neglected. This arises from a (seemingly obvious) distinction between the importance of attaining precise electrical or mechanical functionality versus the imprecision of the "best" human system interface.

Human error has persisted. This paper proposes an alternative paradigm for understanding human error and an additional avenue (called soft deficiencies) to address it. The paper is motivated by questions that do not seem to be well addressed by the current approaches:

Department of Systems, University of Pennsylvania, Philadelphia, Pa. 19104-6315. Current address: 234 Chelsea Court, Cherry Hill, N.J. 08003.

- Why are the most highly trained, paid, and professional humans in the system blamed for a majority of accidents?
- Why are pilots blamed for 70 percent of aircraft accidents when they represent less than 5 percent of the aviation community?
- Why has the percentage of accidents caused by human error been so consistent over the past 20 years?
- Why has the design community been resistant to changes suggested by the human factors community?

The ideas presented in this paper are advanced as motivation for an alternative paradigm in addressing human error that addresses the previous questions. Some distinctions are semantic, but the end result is additional interventions to those already being attempted. The problem is a complex one, and research efforts are under way on many fronts. This paper is offered as but a first step in enhancing the current approach to human error. The presentation might be somewhat novel, but the definition of the problem will be familiar to human factors specialists. The paper is directed at those in the aviation community who are not already enlightened about the importance of human factors considerations in design. The central discussion of the paper, headed Soft Deficiency and System Design, provides a basis for intervention that can be carried out jointly by human factors specialists and engineers. Focus on soft deficiencies would provide these groups with a common language and goals that are now lacking.

As indicated by the title, the focus in this paper is on understanding human error. The discussion is on the system-level "macro" issues as opposed to a specific definition of the perfect human-machine interface. There is always trial and error, but identifying actions that will prevent human error requires an understanding of human error.

Human error has been the subject of considerable research over the years, and numerous studies and publications address the issue. A recent book by Reason, *Human Error*, (*6*) provides a comprehensive overview of what is and is not known about human error. Science has yielded numerous theories that classify and explain vagaries of human behavior. Although these theories offered some inspiration, they were not the basis for the findings in this paper. The problem is that they are all very complicated, and a complicated explanation leads to a complicated intervention. The understanding of human error discussed here is the intuitive understanding that drives decision making. Fortunately, this theory of human error is simple and provides simple interventions. The present effort makes no claim to furthering the science of human error. The intent is to identify approaches that would reduce the number of aircraft accidents caused by human error. For example, no distinction is made between error and mistake. The author considers human error to be the type of action identified as human error in accident statistics. The examples used center on cockpit crews and pilots, because pilot error is believed to be the principal cause of accidents, but other examples, such as air traffic control or maintenance, are possible.

A few other clarifications of terminology are needed. Managers, engineers, and others who advocate or are involved in the implementation of automation as a solution are referred to as technologists. A system is a composite of skilled people, procedures, materials, tools, equipment, facilities, and software that provides an operational capability. Human interaction with the rest of the system will be referred to as human system interaction. The discussion of systems is in terms of the animate (human) and inanimate components. For the sake of simplicity, the inanimate system is sometimes referred to as the machine component of the system, but it includes all of the inanimate elements identified previously. The intention of design is referred to several times when the capabilities of humans are compared with those of the rest of the system. Although machines, tools, and procedures might be designed with certain intentions in mind, it is not clear that there is a corollary intent in human design, which is the common limitations and capabilities of humans. Several other distinctions in the use of words are made as they appear in the text.

HUMAN ERROR AS HUMAN MALFUNCTION

The root of the perception that humans are unreliable lies in the intuitive understanding of human error. Humans often learn new concepts through analogies with what they already know. For example, humans can make intuitive assessments about atomic structure equipped only with knowledge of the solar system and the fact that the two are similar. The use of analogy is a powerful learning mechanism, but it is not without problems when the analogy does not adequately represent reality. Observation demonstrates that the analogy used to understand human error by managers, engineers, and perhaps even scientists who study human error is that human error is like malfunction. Like malfunction, human error is the result of faulty design or other weakness. Like malfunction, human error is itself a bad thing. As in malfunction, intervention for human error requires fixing or replacing the human. As in malfunction, the one who commits the least human error is the best. At the very least, this idea of human error is hard on one's self-esteem. This section will address why equating human error with malfunction is probably not appropriate and certainly not productive.

Machines malfunction but, strictly speaking, humans do not. Malfunction not only indicates that something went wrong, but it also implies that the outcome was not consistent with the intention of design. Malfunction is often descriptive of machine failure, because machines often do not perform within the intention of design. It would be presumptuous, however, to assume that humans do not perform within the intentions of design. Indeed, although human performance can be maladaptive and have adverse effects, the performance itself is always the result of combined innate and learned capabilities. What might a human do that is not consistent with the intention of design? Placing a hand in a fire and enjoying it would be a malfunction. Desiring a life without food and shelter would be a malfunction. However, humans do not do these things. No humans would keep their hands in a fire or go indefinitely without food unless there were an accumulation of experience that made those events desirable. In the latter case, there is no malfunction, because the human is responding appropriately to experience or immediate circumstances. Errors that result from excessive workload, overtaxing complexity, or long periods with low stimulus are also not malfunctions. Humans have limitations of endurance, cognitive

abilities, and attention span that cannot be overcome. Similarly, when a jet engine's performance decreases after being in an "over-speed" condition, no one claims that the engine has malfunctioned. Malfunction is not a good analogy for human error.

The fact that human error is not characterized well by malfunction means that assuming so is problematic and unproductive. This is illustrated when the malfunction/human error analogy is taken further. Machine malfunctions can be the result of worn parts, faulty design, or misuse. Intervention might involve replacing parts, correcting the design, or retraining the operator. What are the analogies of worn parts, faulty design, or misuse, in the case of humans? "Worn parts" might be like old people, but replacing humans is more involved than replacing machines. For example, humans cannot be replaced piece by piece. Some may accept that human error is the result of faulty design, but few would claim knowledge of a better design. The idea of human error being the result of human "misuse" shows promise, but retraining the operator (the inanimate system) may not be so easy. The point is that the interventions suggested by the malfunction analogy are all dead ends, or nearly so.

The problem with the malfunction analogy comes down to one thing: humans cannot be "fixed." The medical field is in the business of "fixing" humans, but addressing human error does not usually involve medicine. Training and retraining is normally considered the "fix" for humans, but even this is not exactly correct. Training is not a "fix" because it is not repair of something that is not working. Training involves knowledge acquisition through the human ability to learn. Even though human factors specialists may not view themselves as fixing humans, this is an underlying reason why decision makers are often reluctant to fund human factors projects. Few are willing to predicate plans and expend dollars on the undefinable process of fixing humans.

HUMAN ERROR AS SYSTEM MALFUNCTION

An assumption is integral to the blaming of 70 percent of aircraft accidents on human error. The system that fails in an accident has both animate and inanimate components. Humans are highly dependent on the rest of the system. Even if the cockpit crew did not follow procedures or was not adequately vigilant, it is an assumption that the inanimate system did not cause the crew to act in this way. It is not satisfying to blame inanimate objects, but human performance is too dependent on the environment to automatically be held accountable independent of the environment. The evidence accumulated to support the conclusion of human error only proves that the human could have prevented the accident (i.e., through use of proper procedures). Proving that the pilot could have prevented the accident is not the same as proving that the pilot caused it.

This is more than an academic issue. Although there are always legal implications, the principal reason for determining cause is to identify actions that would prevent future accidents (7). Blaming 70 percent of aircraft accidents on humans has led to the current drive for automation, but if the system is causing the human errors, additional interventions might be warranted. Although the theory of automation (no human, no human error) sounds plausible, the reality is not. In reality, replacing humans is a very complex, time-consuming, and expensive process. In the end, something less than total automation is always the result, and humans are still left to fill in the gaps. These gaps are usually not the result of careful planning, but are those parts of the system that the designers found too difficult to automate. Humans may be left with a role for which they are even less suited, and may not be able to maintain the situation awareness that is necessary to take over when the automation fails.

Attempts to evaluate the suitability of automation are usually inconclusive. Technologists can always point to an emerging technology that will overcome any limitations identified. Artificial intelligence (AI) is now expected to further extend the role of automation and surmount any shortcomings in the current generation. The author does not intend to pass judgment on automation, merely to suggest the need for some "reality testing." Five years of designing automation systems and 6 years of studying AI makes the author suspect that automation may not live up to expectations and will always be waiting for a technology that is "just around the corner." It appears that humans are at a disadvantage because they are held more accountable (no one expects that an enhanced version will be available next year). On the other hand, the version that is already available could be used more effectively and is quite powerful.

Is it any more plausible that human error is caused by the system? Consider an accident whose cause seems to be a clear case of human error. The following excerpt is from a National Transportation Safety Board (NTSB) report (7):

> About 0734 e.d.t., on September 11, 1974, Eastern Air Lines, Inc., Flight 212, crashed 3.3 statute miles short of runway 36 at Douglas Municipal Airport, Charlotte, North Carolina. the flight was conducting a VOR DME nonprecision approach in visibility restricted by patchy dense ground fog. Of the 82 persons aboard the aircraft, 11 survived the accident. One survivor died of injuries 29 days after the accident. The aircraft was destroyed by impact and fire.
> The National Transportation Safety Board determines that the probable cause of the accident was the flight crew's lack of altitude awareness at critical points during the approach due to poor cockpit discipline in that the crew did not follow prescribed procedures.

The type of accident described is known as "controlled flight into terrain." It is still one of the most common types of accidents, in which, inexplicably, cockpit crews fly aircraft into the ground even though all instruments and warning systems function properly. These accidents are the apparent result of the cockpit crew's lack of situation awareness. Perhaps the crew did not notice a disconnected autopilot (even given flashing lights and bells) or misjudged altitude because they were distracted (even though they are supposed to read instruments and not judge altitude). In any case, how could the (inanimate) system have caused the human error that led to the accident? Specifically, how might the system have contributed to the diminished situation awareness of the pilots? The landing procedures required may be complicated and inefficient, leading the pilots to seek shortcuts. The ground proximity warning system may be activated so frequently (every landing) that they desensitize the pilots. The automated systems may lay the foundation for boredom. Deferred-maintenance items

(identified by fluorescent tags on instruments) might distract the pilots' attention. Difficulty in reading instruments may cause the pilots to favor visual landings, causing overdependence on visual cues. The list could go on, but the one factor that prevents all of these from causing an accident is pilots' ability to remain vigilant.

Pilots are very capable individuals who are expected to remain vigilant, but consider the factors that contribute to vigilance. Personal traits may not be dominant. All of the elements identified previously work against pilots' vigilance. Positive factors are also conceivable. For example, awareness of a recent accident "due to lack of vigilance" will certainly increase the cockpit crew's efforts to remain vigilant. Taken individually, each of these might be ignored as insignificant "noise." The problem is that they are additive: each one requires additional cognitive processing by the crew. Pilots can discipline themselves to carry out procedures properly (even if they seem to be inefficient); they can remember what each aural warning means (even if approximately 15 of the more than 100 possible alarms sound every flight); they can entertain themselves when the autopilot is in control (for 98 percent of the flight); they can remind themselves that those instruments tagged by maintenance are probably not needed; and they can remember to trust their instruments during landings. They usually manage to do all of these things, but even if it were possible for some pilots to overcome all of these factors every day for their entire career, the probability that all pilots will do so is remote. Not only does vigilance vary on the basis of external factors, but also the amount of vigilance needed (to avoid an accident) varies, depending on the phase of flight or flying conditions, such as weather and congestion.

Figure 1 shows a block diagram of a mechanism in which accidents are a function of vigilance and flight conditions—specifically, pilot effectiveness and flight conditions. Pilot effectiveness is the combination of competence and vigilance. Pilot effectiveness and flight conditions can be viewed as stochastic (random) processes. This is not to say that they are totally random, because there will be typical distributions based on the aircraft and routes flown. Nonetheless, their status at any given time is random, because the sum of the factors affecting vigilance and the flight conditions at any given moment will be random. Figure 2 shows a hypothetical distribution of flying conditions and pilot effectiveness for a single flight. A relative scale is used. The graph shows deviation of conditions and effectiveness above and below normal. It illustrates that it is not the absolute level of effectiveness or flight conditions that are important but their relative levels. If the level of effectiveness drops below the level needed by current flying conditions, an accident (or incident) will result. In the case shown, the levels coincide during landing. Note that the crash did not occur during the most demanding

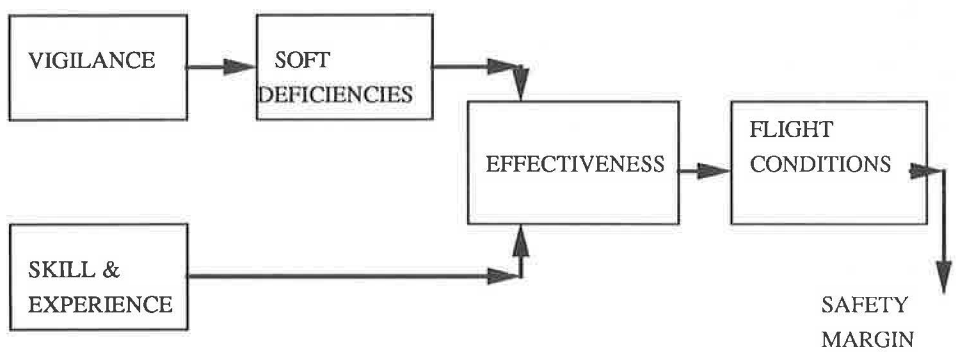


FIGURE 1 Mechanism for system design causing aircraft accidents.

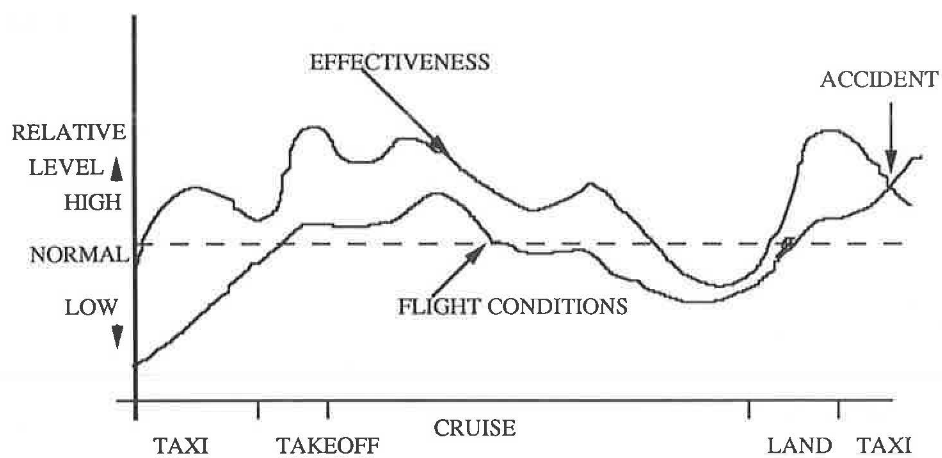


FIGURE 2 Hypothetical distribution of aircraft conditions and pilot effectiveness for a single flight.

flight conditions or while vigilance was lowest. It should be further noted that the time when two stochastic processes will coincide is also random. This model suggests a surprising conclusion: in effect, aircraft accidents are random events and are not necessarily a function of pilot performance. This at least is not contradicted by the findings of aircraft accident investigations. No one has ever shown a clear correlation between aircraft accidents and pilot performance. This will be discussed further.

Some additional terminology is needed. The factors affecting vigilance identified previously are labeled *soft deficiencies* by the author. Soft deficiencies are a broad class of system characteristics that work against human performance. Hard-system deficiencies (i.e., insufficient durability) cause the hardware to fail, and soft-system deficiencies (i.e., inefficient procedures) cause the human to fail.

All of the factors listed previously have contributed to the crash of Flight 212. The captain and the first officer had very good performance records and were experienced in flying this type of aircraft. The flight was the first of the day, and it is safe to assume that it was routine. The NTSB reference to "lack of altitude awareness" refers to the conversation between the pilots recorded during the last minutes of the flight. It seems that the pilots were looking for an amusement park tower they normally used as a visual cue to assess their position relative to the airport. The problem was that there was patchy fog and they were never sure whether they saw the tower. In their preoccupation with identifying the tower, they ignored the ground proximity warnings and did not adequately follow procedures. The reference to "poor cockpit discipline" refers to the crew's failure to follow procedures and to the fact that some of the conversation was on nonoperational issues.

The NTSB's pilot error cause and the author's soft-deficiency cause represent two ends of a spectrum. The crash of Flight 212 represents one failed landing out of many thousands, or perhaps even millions, of successful landings. Accepting that pilot error caused the accident requires accepting the fact that the crew of Flight 212 was an exception. It presumes that there was something unique about their lack of discipline. The mechanism of soft deficiency is largely independent of the crew. In this view, accidents are the result of external factors and need not be linked to pilot performance. Reality probably lies somewhere in between. The question is, where? It is not possible to know directly, but interventions based on the exception side of the spectrum are already being applied. It might be time to consider additional interventions suggested by the alternative.

HUMAN FACTORS

The field of human factors seeks to apply the knowledge of human limitations and capabilities (design intention) to the design of systems (*8*). The goal is systems that are compatible with humans. The field of human factors has been in existence for more than 40 years, and a great deal is now known about the limitations and capabilities of humans. Yet aviation systems (or systems in general) that are compatible with humans are the exception (*3*). One reason may be that there are still a few elements missing before consideration of human factors can become an integral component of system development. The missing elements might be the following:

- Conviction that human performance can be reliable.
- Causative linkage between aviation accidents and human system incompatibility.
- Clear interventions for ensuring human compatibility.

Naturally there is a diversity of opinions on these points, but the dominant consensus is the following:

- Humans are unreliable, and design for compatibility is secondary to automation, because human behavior is unpredictable no matter what the design.
- Aircraft accidents are caused by human errors that are the result of a lack of discipline. Aircraft systems sometimes do not help the situation but cannot be blamed for human fallibility.
- Human factors are too "fuzzy" to be used as the basis of design. It is impossible to say anything conclusively about interventions, given individual human differences.

Although definitive proof is beyond the scope of this paper, the arguments presented in the previous two sections take aim at the first two elements;

- Humans do not malfunction; thus, if system design is compatible with human capabilities and limitations, humans are completely reliable.
- Aircraft accidents are caused by the accumulation of soft deficiencies. Soft deficiencies are the elements of the environment that work against the pilot's vigilance. Soft deficiencies are the result of system design that inadequately considers human capabilities and limitations.

In the remainder of this paper, the author addresses the third point.

First, though, a few additional words on human factors are necessary. The author is told that the primary obstacle to interventions based on human factors is the complexity of human behavior and individual differences. Actually, complexity and individual differences present an obstacle to predicting the precise nature of individual human behavior but not to the design of systems that are compatible with humans. Observation of humans in most situations rapidly demonstrates the complexity and unpredictability of human performance (e.g., observing cockpit crew interaction, even when the high-fidelity simulator scenario is repeated). Yet from other viewpoints, human performance is very uniform. For example, humans share many common motivations, such as the pursuit of nutrition, shelter, peer approval, respect, and self-preservation. Further, although there are individual differences, humans share the same basic cognitive and physical abilities. For example, unlike computers or machines, humans process information symbolically and have limitations of endurance. Humans are more alike than different. The characteristics that humans share are those that are innate (possessed at birth). But if all humans are similar, what is the source of the complexity? The answer is individual experience. Individual experiences are as unique as fingerprints. Behavior is a function of innate and learned (from experience)

components. Complexity of behavior arises from the learned component. This can be directly observed. Complexity surfaces when there is more than one way or no clear way to satisfy innate motivations with innate capabilities. Life rarely presents situations in which innate motivations can be satisfied through innate capabilities; thus we rarely see uniform behavior. Consider a simple example. If 10 people are asked to get a kite out of a tree, they will probably apply 10 different techniques. On the other hand, if a ladder is put in place and the same assignment is made, chances are that all 10 will use the ladder. Variability of human behavior is the result of the application of experience to situations in which humans are forced to adapt. Everyone has similar inherent capabilities and limitations and the same innate motivations. Rather than trying to characterize all of the varieties of human behavior, then, one should simply recognize that if the system is completely compatible with human capabilities and limitations, behavior will be consistent. To take this a step further, the existence of variability in human performance in any given situation is an indication that the humans involved are being forced to adapt. In other words, inconsistent performance may be the result of faulty system design (insufficient consideration of human factors) and not human fallibility.

This discussion does not seek to prove that humans never make mistakes. Clearly, training will remain essential to increase human performance. The distinction is that training addresses the learned contribution of human performance and not innate factors. Humans will always lose vigilance in certain conditions. It is not sufficient for machines merely to perform within functional requirements; it is essential that those requirements be compatible with innate human capability. In any case, this focus on the system does not reduce the responsibility of the individual. Systems will remain dependent on individuals performing to the best of their abilities. Soft deficiency is recognition that "to the best of their abilities" is not always enough. The final two sections of this paper address the concern that there are no clear design interventions suggested by consideration of human factors.

THEORY OF HUMAN FACTORS

Clear and concise guidelines for intervention require a theoretical foundation. The field of human factors has accumulated a considerable body of knowledge about human capabilities and limitations, but these do not yet form a theoretical basis. A theory is a system of assumptions, accepted principles, and rules of procedure devised to analyze, predict, or otherwise explain the nature of behavior of a specified set of phenomena. Electrical engineering is based on theories about electricity. Aeronautical engineering is based on the theory of aerodynamics. But the human factors system of assumptions, accepted principles, and rules of procedure used to analyze and predict does not adequately explain the behavior of humans. Until it does, human factors will not be incorporated in system development with priority equal to that of the disciplines that have a theoretical basis.

Fortunately, the addition of a few more assumptions and consideration of soft deficiencies is all that is needed to round out a theory of human factors. The assumptions needed have already been alluded to: (a) humans are reliable (i.e., performance is completely consistent with design intention) and (b) innate human behavior is uniform (i.e., variability in human behavior is the result of applying individual experience in situations in which humans are forced to adapt). These are, perhaps, impossible to prove and are thus assumptions, but the current consensus identified in the previous section is also an assumption.

The inanimate system is currently given the benefit of the doubt in cases of human error. As it stands, humans are expected to adapt to the limitations of the rest of the system. It seems more appropriate that systems should be designed to adapt to humans. Actually, this is academic, because the history of aircraft accidents demonstrates that humans cannot adapt sufficiently to the inanimate system, no matter how hard they try. The continued struggle of highly trained and professional commercial pilots should leave no doubt about this. Existing capabilities of system development hold great promise for the design of systems compatible with humans, but it will require a new way of perceiving the human component of the system.

The addition of these assumptions gives human factors a true theory capable of explaining the nature of the behavior of a specified set of phenomena. The behavior of interest is innate human behavior, and the specified set of phenomena is human system interaction. The claim can be made that human error can always be traced to some element in the system. This is not unlike the operative claim now used by technologists that all human error can be prevented by automation. Neither statement has value in a literal sense, but both are hypothetically plausible and ensure that perseverance will eventually lead to a solution. The only question is in the level of perseverance that will be cost-effective. As it stands, the human factors perspective allows attribution of performance irregularities to the complexity and unpredictability of humans. This says that the tools of human factors are not sufficient to understand all aspects of human performance. In effect, the outcome will still be a matter of chance. This may be reasonable and true based on the current assumptions, but it is not what a decision maker wants to hear. The implication is that funding of human factors projects is a gamble, because there is a possibility that no matter what the duration of the project is or its success, error may still persist. This may start to explain the disparity between the funding of human factors and automation projects.

The challenge does not just involve developing quantitative criteria. For example, the bulk of the information used by electrical engineers to design systems is qualitative. Design is based on guidelines for elements such as circuit function, layout, grounding, cooling, packaging. Engineers become skilled in design only through experience, because much of what is required cannot easily be conveyed in a text. Nonetheless, the impact of this qualitative aspect of engineering design on the performance of the final product is very clear in terms of cost of fabrication and reliability. This knowledge is what separates new engineers from veterans. Soft-system design will probably also involve a combination of quantitative and qualitative knowledge.

SOFT DEFICIENCY AND SYSTEM DESIGN

Relative to human performance, system design currently comes into question most often where there is a direct link (e.g., a

case in which an instrument's location facilitates accidental engagement or disengagement) (*9*). The pursuit of soft deficiencies will be much broader. A candidate for soft deficiency may be any aspect of the system that is incompatible with innate human motivations, capabilities, and limitations (physical and cognitive). Previous examples suggest some changes in the cockpit that might make it easier for the pilot to remain vigilant. The solutions are not sophisticated; the difference is a matter of emphasis or priority. Requiring humans to adapt to the "minor" inconveniences of the system will no longer be standard operating procedure; system design should adapt to humans, or at least explicitly recognize where adaptation is not feasible.

The concept of soft deficiency is designed to facilitate the process by placing under one heading a broad range of human system incompatibility issues. Soft deficiencies must be pursued jointly by human factors specialists, engineers, and others. Thus, disparate disciplines are provided with a joint language and goals. Pursuing soft deficiencies will be fundamentally different from focusing on human error. Human error is no longer the problem; human error is a symptom of the problem. Soft deficiency provides motivation for "human centered design" beyond the desire of not being second to a machine. Automation should be implemented as a tool to make the system more compatible with humans, not as a replacement for humans. The existence of human error in system operation has implications for system design first and training second.

The search for soft deficiencies can start in those elements of the system that are vulnerable to lapses in pilot discipline. A number of soft-deficiency examples and potential interventions are listed below. The examples are already the subject of in-depth investigations, and the brief discussion here does not seek to provide definitive interventions. A complicated balance exists in the aviation system, and changes require thorough analysis and testing. The following examples are designed to highlight the alternative perspective and potential interventions suggested by consideration of soft deficiencies. A central theme of the examples is that improving pilot performance should be the intervention of last resort. All of the examples focus on errors committed by pilots (most have been identified as the cause of one or more accidents), but all the examples have interventions independent of pilots.

Example 1

Human Error

Pilots occasionally ignore aural warnings and flashing lights that indicate important conditions.

Soft Deficiency

The warnings are issued whether or not the pilot needs to be notified. For example, the ground proximity warning sounds every time the aircraft passes through the elevation of 1,000 ft. Pilots are constantly turning off alerts that are superfluous (they are already aware of the condition). Thus, turning off alerts becomes relatively routine. It becomes automatic. Occasionally, pilots will turn off alerts without giving sufficient thought to the meaning of the alert.

Intervention

This dynamic has led to accidents that were blamed on pilot error (*7*). Recommendations made after accident investigations were designed to ensure that pilots paid closer attention to the alerts in the future. The soft-deficiency perspective suggests something different. It is only natural for humans to become insensitive to repetitive stimuli. A change in the alert system is warranted: the alert should activate only if the pilot has demonstrated a lack of knowledge about the condition. There are numerous ways to achieve this. One approach might be to modify the alert circuit so that the pilot could turn off the alert during a window before activation. Then, for example, if the ground proximity alert ever were to activate, it would explicitly represent the pilot's lack of awareness.

Difference from Current Practice

This intervention recognizes that alerting the pilot is not simply a matter of sounding an alarm. Current alarms are incompatible with humans in two ways. First, the alarms activate too frequently. Activation when the pilot is already aware of the condition is a false alarm. Too many false alarms lead the pilot to ignore the warning. Second, cognitive processing is necessary to identify the meaning of the alarm. Pilots may lose their motivation to do the processing (too many false alarms), or there may not be enough time to do the processing.

Example 2

Human Error

Pilots occasionally fail to maintain situation awareness.

Soft Deficiency

Boredom is a known problem during long automated flights (*4*). While actively flying the aircraft, the pilot necessarily does whatever is necessary to maintain situation awareness. When the autopilot is in control, the pilot is less likely to work so hard to maintain situation awareness, and there may be lapses.

Intervention

Although it may not be possible to maintain a high level of stimulation in long automated flights, it should at least be recognized that automated flight is an adverse environment for humans. The system should facilitate mental activities that enable the pilot to remain aware. Maintaining situation awareness requires integrating information from a number of sources, and this involves considerable and continuous mental effort. The trick is to make it fun or at least interesting. Perhaps pilots could periodically test themselves. Perhaps the effort to maintain situation awareness could be incorporated into some sort of training. Perhaps small competitions could be set up between the pilots or between pilots and the automatic

pilot. The point is to use innate human motivations (i.e., competitive spirit) or capabilities to offset other human limitations (i.e., attention span). No matter what is done, it should be possible for the crew to have nonoperational discussions. The long-run intervention may be to change the representation of flight information, so the pilot would not have to integrate information mentally from a dozen instruments.

Difference from Current Practice

Some work is under way to address this issue, but the principal intervention involves expecting pilots to stay more alert. Soft deficiency emphasizes the importance of facilitating the pilot's effort to stay alert.

Example 3

Human Error

Pilots occasionally fail to follow landing procedures (e.g., to call out certain altitudes during the landing cycle).

Soft Deficiency

The procedures are fixed and do not accommodate the possibility that the crew may not be able, for one reason or another, to carry out every aspect of the procedures. The procedures are rigid by design to elicit uniform performance. It is unrealistic, however, to believe that such uniform performance is possible. There are no guidelines for modifying the procedures, so it is left to the pilot to decide what should go and what should stay. It is inevitable that situations will arise in which the crew has to take shortcuts in following procedures. Policies and procedures should account for this.

Intervention

The intervention would be to leave intact the current requirements (which remain satisfactory for 99 percent of flights) but provide guidelines for adapting the procedures, when necessary. Pilots should learn the priority and motivation for each step in the procedure. It should become ingrained that missing an altitude call-out increases the risk of crashing short of the runway.

Difference from Current Practice

Current effort focuses on getting all pilots to carry out procedures perfectly during every flight. The soft-deficiency perspective suggests that this is unrealistic.

Example 4

Human Error

A pilot occasionally decides to seek visual cues in conditions that warrant using instruments.

Soft Deficiency

The information provided by instruments falls short of information available to pilots under Visual Flight Rules (VFR). Humans learn to fly under visual conditions and become accustomed to integrating the tangible and intangible stimulus of that environment. Instrument flying does not provide sufficient stimulus. Thus, humans inevitably favor visual cues.

Intervention

Flying an aircraft requires more than knowing heading, altitude, and velocity. It requires an integrated mental picture of the relationship between the aircraft, ground, and other aircraft. It requires skill and experience to develop this picture from existing instruments. Instruments should be designed to present information to the pilot that is closer to the information available during VFR.

Difference from Current Practice

Some effort is under way to improve pilot displays, but the principal intervention currently involves demanding increases in pilot discipline. The soft-deficiency perspective suggests reversing the priority of these interventions.

Example 5

Human Error

Pilots occasionally fail to maintain cockpit discipline or commit other lapses in professional conduct and standards.

Soft Deficiency

Cockpit crew conversations that have been recorded in the final minutes before accidents often are not focused on the immediate task of flying. Given the eventual outcome, it is disconcerting to see the pilots apparently disconnected from their duties. However, it is natural for humans to use casual conversation to reduce the monotony of day-to-day flying. Insisting that casual conversations be avoided will not achieve the goal of having pilots be more attentive to their duties. Pilots can monitor themselves on this account. Excessive dependence on discipline is symptomatic of other problems. Discipline is needed only when humans are expected to do things that they are uncomfortable doing.

Intervention

The cockpit environment, procedures, training, and policies should be revisited to determine what makes the process uncomfortable for pilots. Approaches that are more compatible with humans should be adopted. When improvements are not feasible, it should be recognized explicitly that a particular aspect of the process is incompatible with natural human behavior.

Difference from Current Practice

There is considerable focus on increasing cockpit discipline. The soft-deficiency perspective suggests that cockpit discipline is as good as it is going to get. Alternative interventions are needed. Alternative interventions are suggested by scrutiny of what is vulnerable to lapses in pilot discipline. Discipline is needed to make humans do things that make them uncomfortable, so these elements are thus prime candidates for enhancements suggested by human factors.

Summary

Expecting systems to take the blame for human error may seem to be an excessively burdensome requirement. It is a change of emphasis, but it probably will not result in a net increase of system complexity or cost. The additional design complexity caused by incorporation of human factors is balanced by a decrease in system objectives (it need not replace the human). Ambitious efforts to automate pilot duties contributed to costs for aircraft that have risen far faster than inflation. In any case, the idea is to integrate consideration of human factors and not to add steps to the design process. The army program MANPRINT (manpower and personnel integration) has integrated human factors into the acquisition process (proposal, selection, design, test, and evaluation). Although contractors were not initially enthusiastic about the idea, all found that the process did not lead to additional costs (*10*). It is important, also, to remember the reason for incorporating human factors in the first place (i.e., the payoff of enhanced system performance).

Systems are judged not so much by whether they meet functional requirements but by their reliability in meeting functional requirements. It has already been established that no system is 100 percent reliable. The proposal is to consider soft deficiencies in the assessment of system reliability. In other words, end the practice of distinguishing between system reliability and human reliability. The two cannot be separated in a meaningful manner. Although it will become more difficult for the system (human and machine) to achieve high levels of reliability, the term "reliability" gains more meaning. A machine might now claim to be highly reliable when its real-world performance, as a part of a system that includes humans, is poor. Reliability that includes human performance is more representative of reality than is hardware reliability alone. This version of reliability emphasizes that the human factors specialist is as integral to system design as the electrical engineer. Once this fact is accepted by all involved, it will promote cooperation, because neither can reach his or her goals without the other.

CONCLUSION

This paper presented alternatives to existing assumptions and suggested additional interventions to prevent the human errors that lead to aircraft accidents. The understanding of human error presented makes the case that human error is not human malfunction. Human error is not fundamentally different from human behavior that is not considered to be error. In both cases, human behavior is driven by attempts to satisfy innate motivations with innate capabilities. Unpredictable individual differences arise when humans apply experience in situations that require adaptation. Preventing human error is a matter of designing systems that do not force humans to adapt. The elements of the system that force humans to adapt are labeled soft deficiencies. Human factors effort should focus on the common innate elements of human behavior.

There are already examples of programs that successfully integrate human factors and system design. The army has one of the largest in its MANPRINT program, which links several aspects of the acquisition process and makes consideration of human factors a major evaluation issue. For example, system failure cannot be blamed on the skills of the soldier during test and evaluation, because system designers are aware of soldier skills during the entire design process. The initial apprehension of contractors is usually diminished by the end of the process, and the results have been very good. For example, the tools required to maintain one type of engine were reduced from 140 specialized tools and fixtures to a little more than a dozen that can be found in most homes. Yet there seems to be little indication that without government intervention companies will pick up the process on their own. Given the ideas in this paper and the success demonstrated by MANPRINT, there is hope that a process of education might turn the situation around.

Progress in this area should not pause for a debate on whether human error or the system causes aircraft accidents. Given that both require assumptions, the debate could be sustained indefinitely. Aviation does not have forever. Not that aviation is unsafe, but the shear number of aircraft expected to be in service means an increase in accidents. The increase might be sufficient to further alarm a public that already is not particularly comfortable with flying. There is room for both viewpoints. Considerable effort is under way based on interventions suggested by human error, and additional interventions suggested by system design are warranted.

There is a need to further solidify a theory of human factors, but the first step is one of awareness. The more people who can be educated about the importance of human factors, the more resources that will be available to develop a theory. The human factors community understands the problem and can evaluate the ideas in this paper. However, the Human Factors Society has a membership of around 5,000, and the Institute of Electrical and Electronic Engineers has more than 300,000 members. Other design organizations boast similar membership numbers. The greatest education challenge is in the design community. The author will begin the process by presenting the ideas initiated by this paper to design-oriented forums. If there are readers who are interested in volunteering ideas and knowledge or combining efforts in support of this education process, please contact the author.

One issue not addressed in this paper is perhaps the most difficult obstacle to the incorporation of human factors in system design. It is the difficulty of assessing the value of changes suggested. Soft deficiencies do not become readily apparent until the system is operational, and even then debate is likely. How does one know which soft deficiencies are tolerable and which must be addressed? In any case, once a system is operational it is too late for the changes to be cost-

effective. The objectives of this paper are ambitious, but not so ambitious as to expect to solve this issue. The concepts presented are designed to broaden the acceptance of human factors in design and to establish a framework for addressing this issue. Once the design community is willing, the resources necessary to address this issue can be brought to bear. This paper represents one step in the journey to the higher system performance that can be attained when systems are more compatible with humans. Further progress will require the cooperative effort of the entire community. One of the greatest strengths of humans is the ability to solve problems. Now that we are agreed on the problem, it surely will be solved.

ACKNOWLEDGMENTS

This paper was written for the Graduate Research Award Program sponsored by FAA and administered by TRB. The author would like to thank E. Thomas Burnard, Program Officer, and Larry L. Jenney, Aviation Specialist, TRB, faculty adviser Kenneth Fegley, Systems Engineering Chair, University of Pennsylvania; and TRB research advisers Harold van Cott, National Research Council, and Richard F. Pain, TRB. In addition, the author wishes to acknowledge the helpful comments of the numerous individuals who reviewed drafts of this paper including: Joseph Keegan, Harold Booher, Fred Crenshaw, Judith Mahaffy, Carl Berninger, Elizabeth Rahdert, and James Taylor.

REFERENCES

1. E. Wood and G. Bates. *Two Decades of Air Carrier Jet Operation.* Flight Mechanics and System Design Lessons from Operational Experience, Advisory Group for Aerospace Research and Development, Oct. 1983.
2. *National Plan to Enhance Aviation Safety Through Human Factors Improvements.* Air Transport Association, April 1989.
3. *The National Plan for Aviation Human Factors.* FAA, U.S. Department of Transportation, Nov. 1990. NTIS PB91-100339.
4. D. Hughes. Glass Cockpit Study Reveals Human Factors Problems. *Aviation Week & Space Technology*, Aug. 7, 1989.
5. *Controller Perspective of AERA 2.* FAA, Report MP-88W00015. MITRE Corporation, Feb. 1990.
6. J. Reason, *Human Error.* Cambridge University Press, Cambridge, Mass., 1988.
7. *Aircraft Accident Report—Eastern Air Lines, Inc., Douglas DC-9-31, N8984E, Charlotte, North Carolina, September 11, 1974.* NTSB-AAR-75-9. National Transportation Safety Board, Washington, D.C., May 1975.
8. E. McCormick. *Human Factors in Engineering and Design,* 4th ed. McGraw-Hill Inc., New York, N.Y., 1976.
9. E. Wiener. Fallible Humans and Vulnerable Systems: Lessons Learned from Aviation. *Information Systems: Failure Analysis.* NATO ASI Series, No. F32, 1987.
10. H. R. Booher, ed. *MANPRINT, An Approach to Systems Integration.* Van Nostrand Reinhold, New York, N.Y., 1990.

Analysis of Relationship Between Financial Health and Maintenance Cost Structure of U.S. Airline Industry

EDWARD J. RODOWICZ

The financial structure of the airline industry is examined and compared with the cost of maintenance and fleet age. The U.S. airlines studied are the nine that were classified as major airlines as of January 1, 1990. The Spearman rank correlation test is used to show that there is a correlation between the financial structure and maintenance cost structure of the airline industry. A causal forecasting model is then proposed to predict future maintenance expenditures based on the debt ratio. However, because of inadequacies with the measurements chosen, as well as the complex nature of the airline industry, the results of the study must be considered inconclusive. Although the airlines may appear superficially similar to each other, the characteristics of their composition, such as management, debt policy, route structure, fleet mix, unionization, and marketing policies, make them unique, and it may be inappropriate to attempt to extrapolate results of the individual airlines to the industry as a whole.

The airline industry is expected to grow through the turn of the century, yet the average age of the aircraft fleet continues to increase. Expansion, combined with efforts to replace older aircraft, has resulted in large backlogs in aircraft orders with all the major manufacturers. A study by Rose (*1*) examined the financial health of the industry compared with accidents and incident rates for 35 airlines over a 20-year period. The results of those findings indicated a link between financial conditions and safety, especially among smaller carriers. This study is not about safety, but rather an investigation of the financial structure of the U.S. airline industry to determine if there is a relationship between financial status and fleet age and maintenance costs of the industry.

For purposes of this study, airlines will consist of those carriers considered as Section 401 major airline carriers by FAA as of January 1, 1990. In order to receive this classification, a domestic carrier must have yearly revenues exceeding $1 billion. The nine airlines that fall into this class are American, United, Delta, Northwest, USAir, TWA, Pan Am, Continental, and Eastern. The paper will first examine the financial structure of the major airlines and then explore the relationship, if any, to maintenance costs and fleet age.

To provide the most accurate picture possible of the financial condition of the airlines would require at minimum a detailed examination of the balance sheet and income statements of each of the airlines. Even then, the basis for comparison could be debatable. Most economists or financial analysts would probably not agree on which key measurements should be used. However, considering the capital-intensive nature of the airline industry, three indicators are proposed for this study. They are the capital structure, the cost of capital, and interest expense as a percentage of operating profits. Theoretically, each is an important indicator of financial condition and will be discussed subsequently. These are then compared with a "measure of maintenance cost structure" that will also be discussed later.

The hypothesis is that there is a relationship between the financial structure and maintenance costs and fleet age in the airline industry. The Spearman rank correlation test is used to test the null hypothesis that there is no significant correlation between the rankings of the financial and mechanical costs of the airlines. A regression analysis to predict maintenance expenses from financial indicators is also analyzed.

It is expected that those airlines with poorer financial ratios will have a high cost of capital and that the high cost of capital will affect their profits. Although most airlines may be able to survive this during growth periods, the real financial difficulties will occur during periods of stagnation or recession. Even in good economic periods, if an airline's financial position is poor, it may not have sufficient access to funds (internal or external) for the purchase of new aircraft as replacements for older aircraft. The increasing age of the fleet may, in turn, cause an increase in maintenance costs.

BACKGROUND

The United States is now in the era of mass consumption flying. Since deregulation took place in the airline industry, passenger miles have nearly doubled. The aviation industry is still feeling the effects of the Aviation Deregulation Act 13 years after its enactment in October 1978. The FAA review, as published by the U.S. Department of Transportation (*2*), indicates that the industry has gone through three distinct phases during this period:

1. There was a period of expansion from 1978 to 1985 that saw an increase in the number of large air carriers from 30 to 105. Many airlines also increased their route structures during this period.
2. The period from 1986 to 1988 saw a considerable amount of consolidation, with the number of active carriers dropping from 105 to 61. Several large mergers changed the scope of the major airlines. United acquired Pan Am's Pacific routes

Embry-Riddle Aeronautical University, 808 N. Halifax Drive, Ormond Beach, Fla. 32176.

in early 1986. TWA acquired Ozark Airlines in August 1986. Republic Airlines and Northwest were merged in August 1986. American purchased AirCal in November 1986. Western Airlines was purchased by Delta in December 1986. People's Express and Frontier Airlines were acquired by Texas Air to be combined with Continental Airlines and New York Air, and Eastern Air Lines was acquired to operate as a separate airline. USAir purchased Pacific Southwest Airlines in June 1987 and Piedmont Aviation in November 1987. This consolidation process led to the next phase, concentration.

3. The concentration phase covers the period from 1988 to the present. According to FAA statistics published by the U.S. Department of Transportation (*2*), the four largest carriers accounted for 60.4 percent of traffic in 1988, up 7.9 percent from 52.5 percent in 1978. The composition of this "big four" has also changed. In the decades before deregulation, the industry was dominated by United, TWA, American and Eastern. Today, as the industry enters the 1990s, a new "big four" has emerged, with Delta and USAir joining American and United. Many regional and commuter airlines have become increasingly integrated with the large scheduled air carriers through code-sharing agreements or through acquisition in part or totally by their larger partners.

The next phase in this continuing evolution may well be globalization. The proposed deregulation of the European Common Market by December 1992, along with other "free market" movements around the world, opens the possibility for the creation of megacarriers throughout the world. Competition among the world's air carriers will focus on which airline can put together the most effective global system through marketing agreements, code sharing, and equity stakes in other carriers.

It was expected that both airline revenues and net profit would grow following deregulation. This did not happen. The economic recession in the early 1980s partially accounts for these results. Growth in passenger traffic was seen in the mid 1980s. However, profits for the industry as a whole were still below 5 percent. It was not until the late 1980s that the airlines saw the rewards of growth. Record profits for the industry were seen in 1987 and 1988. According to Henderson (*3*), airline executives and industry analysts were virtually unanimous in their predictions of a third straight year of record profits in 1989. Unexpected events such as the Eastern Airlines strike and subsequent bankruptcy filing, ill-advised promotions, Hurricane Hugo, and the San Francisco earthquake negatively affected industry profitability. FAA statistics published by the U.S. Department of Transportation (*2*) indicate that operating profits fell from $3.2 billion in 1988 to $2.7 billion in 1989. Interest expense on the $11 billion in outstanding long-term debt reduced the industry's 1988 operating profit of $2.7 billion to a net income of only $1.2 billion.

There is considerable disparity among the financial performance of the individual carriers. American and Delta reported combined operating profits of almost $1.6 billion in 1989, up from $1.2 billion the previous year. At the other end of the scale, Eastern and Braniff entered into Chapter 11 bankruptcy proceedings. Although the combined net income for American and Delta totaled almost $1 billion in 1989, Eastern and Pan Am posted a combined net loss of more than $1 billion. Over the last 10 years, the net income of five carriers—American, Delta, Northwest, USAir, and Piedmont—has accounted for 85.9 percent of the entire industry's profits. This brief overview gives some idea of the volatility of the industry. With that cautionary note, the objectives of this study are now presented.

METHODOLOGY

Capital Structure Analysis

The items on the right-hand side of a firm's balance sheet are its capital components. Theoretically, any increase in total assets must be financed by an increase in one or more capital components. These components are various types of debt, preferred stock, and common equity. Because no major airline has any outstanding preferred stock, the discussion here will be limited to the various types of debt and common equity. Included in debt are all capital leases. No distinction is necessary between capital leases and other forms of debt, because capital leases are a form of long-term debt. The capital structure of the firm is important because it indicates the proportion of funds invested by stockholders versus debtors. Theoretically, this relationship influences a firm's risk, which affects the price of the stock and the cost of capital.

Debt itself is not necessarily a bad thing. It allows a firm to acquire assets or resources that can be used to generate income. What is important is how a firm manages debt in relationship to its ability to generate earnings. There is "risk" associated with taking on debt. If the firm is unable to meet the financial obligations of the debt from operating earnings, it may be forced to liquidate assets. Contraction of business operations through liquidation has had a spiraling impact already on several airlines, eventually leading them into bankruptcy.

The debt/equity ratio measures a firm's degree of indebtedness. The degree of indebtedness refers to the proportion of a firm's assets financed by debt relative to the proportion financed by equity. Total liabilities include current liabilities, noncurrent liabilities, and deferred credits. Capital leases are included in total liabilities. It is recognized that operating leases are a form of "off-balance sheet" financing and do not appear in this ratio. The subject of leases is addressed later. Stockholders' equity includes paid-in capital and retained earnings. The source for the data used in calculations for Tables 1 and 2 is the Office of Aviation Information Management (*4*). The data presented in this paper are as reported to the Department of Transportation on RSPA Form 41 Schedule, required by all large certified air carriers holding a certificate under Section 401 of the Federal Aviation Act of 1958. All dollars included in this study are reported as current dollars.

The debt ratios calculated in Table 1 are for the 3-year period from 1987 through 1989. In an attempt to eliminate any temporary fluctuations as a result of significant short-term changes, a 3-year average was also calculated and is used to rank the airlines within the table. Although the top five airlines are closely grouped, there is a significant drop to the sixth-ranking airline. The two lowest-ranking airlines have a negative debt ratio, because their total stockholders' equity was negative. When this condition occurs, their debt ratio is said to approach infinity and is indicated by an INF in Table 1.

TABLE 1 AIRLINE DEBT RATIOS (*4*)

RANKING	AIRLINE	DEBT RATIO 1987	1988	1989	AVG 1987-9
1st	Northwest	1.2	1.2	1.3	1.3
2nd	USAir	1.5	1.6	1.1	1.4
3rd	Delta	1.6	1.5	2.3	1.8
4th	American	2.8	2.4	2.3	2.5
5th	United	3.8	4.1	3.5	3.8
6th	Eastern	8.1	49.3	INF	16.7
7th	Continental	8.3	39.7	24.5	24.2
8th	Pan Am	INF	INF	INF	INF (-4.7)
9th	TWA	7.2	INF	INF	INF (-51.7)
			Industry		3.5

Formula: TOTAL LIABILITIES / TOTAL STOCKHOLDERS EQUITY

The time series analysis of the industry debt ratio for the 10-year period since deregulation indicates that the debt ratio has climbed from 2.5 in 1980 to 3.5 in 1989, with a peak of 3.7 in 1988, as shown in Table 2. This increased use of debt is exposing the airlines to greater financial risk. A temporary decrease in operating revenues could hurt those carriers with heavy debt. Fotos (*5*) indicates that FAA believes that the industry's commitment to debt is its single largest financial problem. The growth rate in the debt ratio, as shown in Table 2, represents a 40 percent increase over the 10-year period, or an average annual increase of 4.4 percent. Projecting this annual average growth rate over the next 10 years (1989 through 1998) would result in a debt ratio of 4.9 in 1998.

The previous projection is based on the average annual increase since deregulation to 1989. Selection of an alternative starting date (1982/1983) would result in a smaller projected increase in the future. The magnitude of the increase may be debatable, but the upward trend is not.

Not reflected in the previous analysis is an alternative means of acquiring equipment through the use of "off-balance sheet" financing (i.e., operating leases). Operating leases allow airlines to expand or replace their fleets without the capital outlays associated with purchasing.

One of the primary attractions of operating leases is the ability to acquire aircraft without effect on the balance sheet. The cost of the operating lease is an expense, and affects the income statement only during the periods the aircraft is rented. Recent corporate tax law changes have made operating leases more attractive. Premo (*6*) indicates that the elimination of the investment tax credits, coupled with lower tax rates and the introduction of corporate alternative minimum tax, has resulted in a dramatic increase in operating leases. From the operational aspect, operating leases give airlines flexibility in fleet planning by allowing for the short-term acquisition of needed aircraft.

Table 3, generated from statistics compiled by Flint (*7*), shows that the major U.S. airline fleet grew from 2,851 aircraft in 1987 to 2,972 aircraft in 1988, an increase of 4.2 percent. The number of aircraft owned (including those on capital leases) by the airlines fell 3.0 percent, from 2,004 in 1987 to 1,943 in 1988. Operating leases increased 30.5 percent, from 558 to 728. Operating leases now account for 24.5 percent of all aircraft, up from 19.6 percent in 1987. These figures do not include Continental and Eastern, because their parent corporation, Texas Air, does not release this information.

Pan Am has the highest percentage of its total fleet on operating leases, at 70 percent, followed by Delta, at 38.8 percent, USAir, at 38.3 percent, and TWA, at 37.4 percent. The two lowest are United, at 14.5 percent, and Northwest, at 19.3 percent. Most of the airlines increased use of operating leases between 1987 and 1988. Delta remained constant, at 38 percent, and TWA had a slight reduction from 39.0 percent to 37.4 percent. Northwest increased from 9.9 percent to 19.3 percent and United increased from 6.5 percent to 14.5 percent. Flint (*7*) points out that many industry analysts expect leasing to be the dominant means of acquiring new aircraft in the future. Delta has indicated that it will continue to purchase aircraft. Other airlines with adequate liquidity, American and United, may see purchasing as an effective use of assets.

Cost of Capital

The term *cost of capital* contains two key words—*capital* and *cost*. The word *capital* refers to a firm's capital structure, which is its component mix of long-term financing. Each component of the capital structure has a cost attached to it. The cost of capital is important to airlines in analyzing whether new projects, such as aircraft purchases or other capital assets, should be undertaken. The expected return on the project must exceed the cost of capital in order for the project to be profitable. In theory, companies should use the weighted average cost of capital (WACC), which includes long-term debt,

TABLE 2 DEBT RATIO TIME SERIES ANALYSIS (*4*)

YEAR	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
RATIO	2.5	2.7	3.4	3.4	2.9	3.0	3.0	3.4	3.7	3.5

Formula: TOTAL LIABILITIES / TOTAL STOCKHOLDERS EOUITY

TABLE 3 ANALYSIS OF AIRCRAFT OWNERSHIP

	AIRCRAFT OWNED*	OPERATING LEASES	TOTAL FLEET	%OPERATING OF TOTAL
AMERICAN				
1987	335	75	410	18.3%
1988	352	116	468	24.8%
DELTA				
1987	230	144	374	38.5%
1988	240	152	392	38.8%
NWA				
1987	273	30	303	9.9%
1988	251	60	311	19.3%
PAN AM				
1987	44	82	126	65.1%
1988	42	98	140	70.0%
TEXAS AIR (SEE NOTE)				
1987	347	N/A	636	N/A
1988	312	N/A	613	N/A
TWA				
1987	130	83	213	39.0%
1988	134	80	214	37.4%
UNITED				
1987	361	25	386	6.5%
1988	349	59	408	14.5%
USAIR				
1987	284	119	403	29.5%
1988	263	163	426	38.3%
INDUSTRY				
1987	2004	558	2851	19.6%
1988	1943	728	2972	24.5%
CHANGE	-3.0%	30.5%	4.2%	

NOTE: TEXAS AIR IS PARENT CORPORATION OF CONTINENTAL AND EASTERN

*OWNED AIRCRAFT INCLUDE THOSE ON CAPITAL LEASES (EXCEPT FOR TEXAS AIR)

Formula: %OPERATING OF TOTAL = OPERATING LEASES / TOTAL AIRCRAFT

Source: Numbers, Flint (7), percentages calculated by author.

preferred stock, and the cost of equity. Because no major airline has issued preferred stock in the past 10 years, the following equation will be used for WACC:

$$\text{WACC} = (\%\ \text{debt})\ (\text{interest rate})\ (1 - \text{tax rate}) + (\%\ \text{equity})\ (\text{cost of equity})$$

The percentages of debt and equity are derived from the underlying data in Table 1. The interest rate is determined from the weighted average of debt issued by the airline as indicated in *Moody's Bond Record* (*8*) and *Moody's Transportation Manual* (*9*). Interest is a tax-deductible expense. It produces tax savings, which reduce the net cost of debt. The interest rate is adjusted downward to account for the tax treatment of debt. The tax rate used is the actual rate for 1989 (Table 4). For those firms with losses, no adjustment is required.

For purposes of this study, the dividend yield plus growth rate model [i.e., discounted cash flow (DCF)] is used to calculate the cost of equity (COE). According to Brigham (*10*), financial analysts estimating the cost of equity tend to rely most heavily on this method. It uses the following formula:

$$\text{COE} = (\text{expected dividend/price}) + \text{growth}$$

The expected dividend is calculated by multiplying the last dividend payment by the expected growth rate. The price is the stock price selected from the *Wall Street Journal* as of the close of business on July 3, 1990. Growth is the expected growth rate as estimated by the security analysts at Value Line, Inc. (*11*).

The problem with using the DCF approach is that only two of the nine major airlines paid dividends in 1989 (Table 5). As an alternative, the capital asset pricing model (CAPM) is used. This methodology was used by McMullen (*12*) in a prederegulation study on the cost of capital.

The CAPM offers a theoretical explanation of the relationship between risk and return. The model implies that rational investors will accept additional risk if compensated by an expected higher return. They require a risk premium above the risk-free rate. In the CAPM, the expected risk premium varies in direct proportion to β. Beta measures the volatility or variability of an individual stock compared with

TABLE 4 TAX RATE CALCULATIONS IN THOUSANDS OF DOLLARS FOR 1989

AIRLINE	EBT	TAXES	TAX RATE
AMERICAN	783,610	281,895	35.97%
CONTINENTAL	(31,489)	8,618	-27.37%
DELTA	764,120	270,011	35.34%
EASTERN	(671,395)	0	0.00%
NORTHWEST	424,820	142,305	33.50%
PAN AM	(337,067)	1,147	-0.34%
TWA	20,839	(305)	-1.46%
UNITED	592,013	236,579	39.96%
USAIR	(60,054)	(19,385)	32.28%

Formula: TAXES / EBT

Source: EBT and Taxes are from Office of Aviation Information Management data (*4*), while Tax Rate was calculated by author.

the total market. Although systematic risk cannot be eliminated by diversification, the company-specific or nonsystematic risk can largely be eliminated in a fully diversified portfolio. Thus, investors are only compensated for the systematic risk they bear, and beta is a measurement of that risk. The CAPM formula is as follows:

$$\text{COE} = \text{RF rate} + \beta(\text{market return} - \text{RF rate})$$

RF stands for "risk free" and can be thought of as the U.S. Treasury bill (T-bill) rate and has a β of 0. According to the *Wall Street Journal*, the current T-bill rate was 7.73 percent as of July 5, 1990.

The difference between the market return and the risk-free rate times the β is referred to as the risk premium. According to the Consolidated Capital Communications Group (*13*), the average risk premium over the past 50 years has been 8.8 percent. Because the T-bill rate is known and the β of the market portfolio is 1, the market return is 16.53 percent [7.73 percent + 1(8.8 percent)]. The β values for the individual airlines have been calculated by Value Line (*11*).

The COE, using the CAPM model, is shown in Table 6. The industry average for COE was 19.46 percent, with Delta and Pan Am having the lowest figure at 16.97 percent, whereas Texas Air (Eastern and Continental) had the highest, at 22.25 percent. The surprise in this table is the COE for Pan Am. The stock price for Pan Am has been very low and has not fluctuated significantly. Because β is a measure of the variability of returns, Pan Am's low β results in a low COE. All the data elements are now available to calculate the WACC.

The WACC is presented in Table 7 and includes all the elements of the equation as well as the source of the data. The airlines are ranked in Table 8 according to their WACC, with the lowest figured being the most desirable. The five airlines with a WACC below the industry average of 12.74 percent are also the five airlines that had the lowest debt ratios (Table 1). The next section will look at how the cost of capital affects the airlines' profits.

Profitability

Historically, the airlines have not been highly profitable. The airline industry showed an operating profit of $2.7 billion in 1989, according to FAA statistics published by the U.S. Department of Transportation (*2*), yet the industry's net profit totaled only $1.2 billion. A substantial portion of this difference is the interest expense on the $11 billion in outstanding long-term debt. This highlights FAA's concern that the industry's commitment to debt is its single largest financial problem. Table 9, calculated from Office of Aviation Information Management data (*4*), shows the percentage of operating profits

TABLE 5 COST OF EQUITY AS OF JULY 3, 1990

AIRLINE	DIVIDEND	PRICE	GROWTH	COE
AMERICAN	0.00	63.000	-42.0%	-42.0%
CONTINENTAL	0.00	6.125	-60.4%	-60.4%
DELTA	1.20	72.125	23.1%	25.2%
EASTERN	0.00	6.125	-60.4%	-60.4%
NORTHWEST				
PAN AM	0.00	2.250	-83.9%	-83.9%
TWA	0.00	11.125		0.0%
UNITED	0.00	149.125	-26.5%	-26.5%
USAIR	0.12	25.375	-76.9%	-76.8%

Formula: (DIVIDEND/PRICE) + GROWTH

Source: Dividend and growth, Value Line (*11*); Price, <u>Wall Street Journal</u>; COE calculated by author.

TABLE 6 CAPM COST OF EQUITY CALCULATIONS AS OF JULY 5, 1990

AIRLINE	BETA	COE
AMERICAN	1.40	20.05%
CONTINENTAL	1.65	22.25%
DELTA	1.05	16.97%
EASTERN	1.65	22.25%
NORTHWEST	1.30	19.17%
PAN AM	1.05	16.97%
TWA	1.50	20.93%
UNITED	1.15	17.85%
USAIR	1.25	18.73%
AVERAGE		19.46%

Formula: RF RATE + BETA(MARKET RETURN - RF RATE)

Source: Beta, Value Line *(11)*; COE calculated by author.

TABLE 7 WEIGHTED AVERAGE COST OF CAPITAL (WACC) AS OF JULY 1990

AIRLINE	%DEBT	INT. RATE	TAX RATE	%EQUITY	COE	WACC
AMERICAN	70.7%	9.52%	35.97%	29.3%	20.05%	10.18%
CONTINENTAL	93.4%	15.04%	0.00%	6.6%	22.25%	15.52%
DELTA	63.0%	9.79%	35.34%	37.0%	16.97%	10.27%
EASTERN	99.9%	15.04%	0.00%	0.1%	22.25%	15.05%
NORTHWEST	55.2%	9.07%	33.50%	44.8%	19.17%	11.92%
PAN AM	100.0%	14.40%	0.00%	0.0%	16.97%	14.40%
TWA	96.5%	15.16%	0.00%	3.5%	20.93%	15.36%
UNITED	78.6%	12.85%	39.96%	21.4%	17.85%	9.88%
USAIR	57.2%	10.47%	32.28%	42.8%	18.73%	12.07%
					Average	12.74%

Formula: (%DEBT)(INT. RATE)(1 - TAX RATE) + (%EQUITY)(COE)

Source: %DEBT & %EQUITY calculated from data used to generate Table 1
INTEREST RATE represents the weighted average of the cost of outstanding debt *(8)*.
TAX RATE calculated in Table 4
COE using CAPM from Table 6

expended for interest during the period 1987 to 1989. In an attempt to eliminate any temporary fluctuations as a result of significant short-term change in interest or operating profits, a 3-year time span was calculated. Total interest expense for the three years was divided by total operating profit for the same period.

TABLE 8 COST OF CAPITAL AS OF JULY 1990

RANKING	AIRLINE	WACC
1st	United	9.88%
2nd	American	10.18%
3rd	Delta	10.27%
4th	Northwest	11.92%
5th	USAir	12.07%
6th	Pan Am	14.40%
7th	Eastern	15.05%
8th	TWA	15.36%
9th	Continental	15.52%

The data in Table 9 show that a significant portion of operating profit for all nine carriers is expensed on interest. Figures exceeding 100 percent (Continental and TWA) indicate that interest expense exceeded operating profits. Negative figures indicate that there was an operating loss. Five airlines—American, Delta, Northwest, United, and USAir—earned a profit after interest payments. Table 10 ranks the airlines based on lowest percentage of operating profits expensed on interest. It also indicates their ranking based on debt/equity ratios and WACC.

The five airlines earning a net profit expensed an average of 29.72 percent of operating profits on interest. TWA's and Continental's interest expense exceed operating earnings by 53.3 percent and 187.1 percent. Eastern and Pan Am did not generate any operating profit, and their net losses for the period were increased by the cost of debt.

With the understanding that there may be special opportunities for a firm to have significantly different financial ratios, Table 10 includes rankings of the airlines with respect to all three financial indicators selected. In general, the higher

TABLE 9 PERCENTAGE OF OPERATING PROFIT EXPENDED ON INTEREST (*4*)

AIRLINE	1987	1988	1989	AVG 1987-9
AMERICAN	46.2%	40.8%	20.8%	35.3%
CONTINENTAL	842.7%	322.5%	166.9%	287.1%
DELTA	21.1%	15.3%	9.3%	14.2%
EASTERN	477.3%	-132.7%	-38.3%	-96.3%
NORTHWEST	38.8%	29.8%	15.8%	25.7%
PAN AM	-56.5%	-104.1%	-37.8%	-54.9%
TWA	99.8%	127.9%	302.1%	153.2%
UNITED	78.1%	29.2%	29.3%	34.5%
USAIR	14.2%	38.7%	231.5%	38.9%

Formula: INTEREST EXPENSE / OPERATING PROFIT

TABLE 10 RANKING BY PERCENTAGE OF PROFIT EXPENDED ON INTEREST, 1989

RANKING	AIRLINE	INT/O.P.	DEBT RATIO	WACC
1st	Delta	14.2%	3rd	3rd
2nd	Northwest	25.7%	1st	4th
3rd	United	34.5%	5th	1st
4th	American	35.3%	4th	2nd
5th	USAir	38.9%	2nd	5th
6th	TWA	153.3%	9th	8th
7th	Continental	287.1%	7th	9th
8th	Pan Am	-54.9%	8th	6th
9th	Eastern	-96.3%	6th	7th

the ranking, the better the financial situation of the carrier. It might be noted that the top five airlines—American, Delta, Northwest, United, and USAir—remained in the top five in all three categories. The lower group of four consists of Continental, Eastern, Pan Am, and TWA. This could imply a bimodal financial distribution in the industry. The maintenance cost structure of the industry is now presented in order to compare it with the financial measures just discussed.

Fleet Age

Aircraft when designed have an estimated economic design life objective, measured in cycles (a cycle being a takeoff and landing), hours of operation, and years. Table 11 presents the current status of the Boeing fleet, indicating the number of aircraft exceeding 75 percent and 100 percent of the manufacturer's specified design objectives. Table 12 indicates the

TABLE 11 BOEING JET AIRCRAFT FLEET STATUS, 1989

Model	Fleet Size	Economic Design Life Objective	Number Exceeding Percent Of Design Objective 75%	100%
707	213	20,000 cycles	170	59
		60,000 hours	152	47
		20 years	192	123
720	14	30,000 cycles	14	7
		60,000 hours	14	6
		20 years	14	14
727	1,649	60,000 cycles	170	2
		50,000 hours	785	358
		20 years	784	435
737	1,051	75,000 cycles	52	3
		51,000 hours	226	38
		20 years	297	9
747	614	20,000 cycles	84	5
		60,000 hours	222	68
		20 years	180	0

Source: Boeing.

TABLE 12 MCDONNELL DOUGLAS JET AIRCRAFT FLEET STATUS, 1989

Model	Fleet Size	Economic Design Life Objective	Number of Airplanes Exceeding
DC-8	350	25,000 cycles	49
		50,000 hours	203
		20 years	350
DC-9	924	40,000 cycles	504
		30,000 hours	635
		20 years	184
MD-80	496	50,000 cycles	0
		50,000 hours	0
		20 years	9
DC-10	425	42,000 cycles	0
		60,000 hours	0
		20 years	0

Source: Aircraft Marketletter.

status of the McDonnell Douglas fleet, with the number of aircraft exceeding the design objectives.

A study of aging aircraft by the U.S. General Accounting Office (*14*) indicates that experts agree that a properly maintained and inspected aircraft can fly indefinitely. The aircraft age issue, for most airlines, becomes an issue of economic trade-offs. Airlines weigh the price, operating costs, route structure, fleet mix, and availability of new aircraft against the cost of operating and maintaining older, less efficient aircraft. The current backlog of orders for new aircraft may also affect an airline's decision.

Table 13, generated from data compiled by Avmark, Inc. (*15*), breaks down the commercial fleets, indicating the average age for each of the major airlines' fleets. The youngest fleet belongs to Delta, with an average of 9.4 years, whereas Northwest has the oldest fleet, with an average of 15.4 years. The question to be explored in the next section is whether there is any relationship between fleet age and maintenance costs.

Maintenance Costs

All aircraft, regardless of age, require routine maintenance. Older aircraft may require additional maintenance expenditures. According to *Aviation Week & Space Technology* (*16*),

TABLE 13 AIRCRAFT FLEET AGES AS OF JULY 1989 (*15*)

RANKING	AIRLINE	AVG. FLEET AGE (YEARS)
1st	Delta	9.4
2nd	USAir	10.0
3rd	American	11.0
4th	Continental	11.5
5th	Pan Am	13.1
6th	TWA	14.3
7th	United	14.6
8th	Eastern	14.8
9th	Northwest	15.4

TABLE 14 MAINTENANCE DOLLARS PER BLOCK HOUR (*15*)

AIRLINE	FLEET SIZE	Q4 - 88	Q1 - 89	Q2 - 89	Q3 - 89	AVG.
AMERICAN	482	409.71	427.14	449.19	496.49	445.63
CONTINENTAL	329	437.15	489.84	475.82	452.36	463.79
DELTA	404	386.75	371.90	384.24	397.77	385.17
EASTERN	95	731.51	746.36	1485.02	674.89	909.45
NORTHWEST	310	534.10	556.30	556.29	533.26	544.99
PAN AM	153	964.55	915.67	817.77	897.00	898.75
TWA	217	680.53	739.72	688.86	790.55	724.92
UNITED	414	610.55	597.57	638.39	700.89	636.85
USAIR	425	384.03	387.34	406.85	469.65	411.97
ALL MAJORS	314	570.99	581.32	655.83	601.43	602.39

TABLE 15 RANKING BY MAINTENANCE EXPENDITURES AND FLEET AGE, 1989

RANKING	AIRLINE	DOLLARS/BLOCK	FLEET AGE
1st	Delta	$385.17	1st
2nd	USAir	$411.97	2nd
3rd	American	$445.63	3rd
4th	Continental	$463.79	4th
5th	Northwest	$544.99	9th
6th	United	$636.85	7th
7th	TWA	$724.75	6th
8th	Pan Am	$898.75	5th
9th	Eastern	$909.45	8th

the Airworthiness Assurance Task Force has estimated that it will cost $800 million in inspections and repairs to keep today's fleet of Boeing 727s, 737s, and 747s; McDonnell Douglas DC-8s, DC-9s, and DC-10s, and Lockheed L-1011s flying. A major overhaul of an American Airlines Boeing 727 costs more than $546,000 in labor and $286,000 in materials, for a total of $832,000. The estimated resale value of the aircraft is $6 million.

Table 14 indicates the expenditures by the major airlines on maintenance per block-hour for a year. A block-hour is measured from the time the blocks are removed from the wheels under the aircraft departing the gate to when they are replaced at the next airport. Maintenance expenditures are for both scheduled (preventive) and unscheduled (corrective) maintenance, including the airframe and engine as well as a maintenance burden (overhead). The source of the data is Avmark, Inc. (*15*).

Table 15 ranks the airlines based on the amount of maintenance expenditures per block-hour compared with the ranking based on fleet age. This table shows a high correlation between fleet age and maintenance costs among the top-four-ranking airlines. The lower-five-ranking airlines in both categories show some inconsistency among their respective rankings. Northwest ranked fifth in maintenance expenditures versus ninth in fleet age. Possible explanations include variations in fleet composition and maintenance efficiencies. Flight lengths may also affect maintenance costs measured per block-hour, because an airline with many long international routes, such as Northwest, might have a lower cost per block-hour than an airline with predominantly short domestic routes. There is a wide discrepancy in maintenance dollars spent between first and last. Eastern spent more than twice the amount of Delta. Contributing to Eastern's situation, in addition to their eighth-ranking fleet age, might be high union labor costs.

RESULTS

The financial structure of the major airlines has been examined in terms of debt/equity ratios, WACC, and interest expense as a percentage of operating profits. All three indicators are theoretically important, so there is no justification for weighing one factor more than another. Table 16 uses an average of the three to determine overall financial ranking.

The maintenance cost structure of the major airlines has been measured in terms of fleet age and maintenance costs. Based on these findings, Table 17 ranks the overall maintenance cost structure. Table 18 combines the previously discussed rankings. Using a Spearman rank correlation test and a T-test, correlation values are calculated. The procedure is explained in Appendix A. Based on the results of both the Spearman rank correlation test and the T-test, the null hypothesis that there is no relationship between the financial health and the maintenance cost structure of the airline industry can be rejected at the 5 percent level of significance.

Although the results of the Spearman rank correlation test were for the industry as a whole, there are differences among the individual airlines. Referring to Table 18, both Delta and Pan Am have the lowest value, at 0.2. However, they have the same value for different reasons. Delta scored well in both the financial and maintenance cost structure categories, at 1.5 and 1.0, whereas Pan Am was near the other end of the

TABLE 16 FINANCIAL RANKING OF MAJOR AIRLINES, 1989

RANKING	AIRLINE	DEBT/EQUITY	WACC	INT/OP PROFIT	AVERAGE
1st*	Delta	3	3	1	2.3
1st*	Northwest	1	4	2	2.3
3rd	United	5	1	3	3.0
4th	American	4	2	4	3.3
5th	USAir	2	5	5	4.0
6th*	Eastern	6	7	9	7.3
6th*	Pan Am	8	6	8	7.3
8th*	Continental	7	9	7	7.7
8th*	TWA	9	8	6	7.7

NOTE: * indicates tie

TABLE 17 MAINTENANCE COST STRUCTURE RANKING OF MAJOR AIRLINES, 1989

RANKING	AIRLINE	FLEET AGE	MAINTENANCE	AVERAGE
1st	Delta	1	1	1.0
2nd	USAir	2	2	2.0
3rd	American	3	3	3.0
4th	Continental	4	4	4.0
5th	Pan Am	5	8	6.5
6th*	TWA	6	7	6.5
6th*	United	7	6	6.5
8th	Northwest	9	5	7.0
9th	Eastern	8	9	8.5

spectrum, at 6.5 and 6.0. Continental and Northwest represent 66 percent of the total variance. These two airlines deviate the most from the industry findings. Continental's financial ranking was very low, at 8.5, whereas the maintenance cost structure ranking was near the middle, at 4.0. Expansion may have left Continental too highly leveraged, or the airline may not have recovered from the 1983 bankruptcy proceeding. Northwest's high variance results from a good financial ranking, at 1.5, versus a high maintenance cost structure ranking, at 8.0. Part of Northwest's expansion was achieved through acquisition of other airlines with older fleets, and these may have affected the maintenance cost structure measures. These examples show that the general findings for a group cannot be applied wholesale to each member in the group.

The Spearman rank correlation test indicates whether there is any correlation among rankings and does not imply causation; therefore, a causal forecasting model is proposed. A regression analysis is used to predict maintenance expenses based on the financial variables. The dependent variable is maintenance dollars per hour, and the independent variables are the debt ratio and the percentage of operating profits expended on interest payments. The regression equation takes the following form:

$$Y = a + b_1X_1 + b_2X_2$$

Refer to Appendix B for an explanation of the equation.

The data used in the analysis are those of the nine airlines for the 10-year period from 1980 through 1989, a total of 90 observations. The 10-year period was chosen to provide more observations, and coincides with the period of deregulation. The maintenance dollars per hour were calculated by dividing the total maintenance expenditure by the aircraft revenue-hours for each year. The source for the maintenance expense data is the Office of Aviation Information Management (*4*). Aircraft revenue-hours are defined as the airborne hours in revenue service computed from the moment an aircraft leaves the ground until it touches the ground again. The source of these data is the Civil Aeronautics Board (*17*) for 1980 through 1983 and the U.S. Department of Transportation (*2*) for 1984 through 1989. The responsibility of maintaining air carrier statistics fell to the Department of Transportation, following the demise of the Civil Aeronautics Board. The debt ratio is a 10-year extension of the data used for Table 1, and the percentage of profits lost is a 10-year extension of the data used for Table 9.

If the independent variables have a positive impact on the dependent variable, then this could indicate that the airlines are not replacing their fleets because of increased debt (already defined). This, in turn, would lead to higher maintenance costs because of older aircraft. A negative result might suggest that airlines are cutting corners in maintenance expenditures because of increased debt; this result is considered highly unlikely because of government regulations.

The multiple regression resulted in the following equation:

$$Y = 580.0 + 3.2X_1 + 0.3X_2$$

With the interest as a percentage of operating profit held constant, maintenance costs per hour will increase by $3.2 for every unit change in the debt ratio, and with the debt ratio held constant, maintenance costs will increase by $0.3 for every unit change in the percent profit lost.

The debt t value (a measurement of the reliability of a coefficient) of 5.13 tells us that we can be 99.95 percent sure that the debt ratio does have an effect on maintenance costs

TABLE 18 COMBINED RANKING, 1989

AIRLINE	FINANCIAL (X)	MAINTENANCE (Y)	X - Y	$(X - Y)^2$
American	4.0	3.0	1.0	1.0
Continental	8.5	4.0	4.5	20.3
Delta	1.5	1.0	0.5	0.2
Eastern	6.5	9.0	-2.5	6.3
Northwest	1.5	8.0	-6.5	42.2
Pan Am	6.5	6.0	0.5	0.2
TWA	8.5	6.0	2.5	6.3
United	3.0	6.0	-3.0	9.0
USAir	5.0	2.0	3.0	9.0
				94.5

per hour. The profits lost t value of 0.63 is quite small, indicating little or no relationship with maintenance costs per hour. Based on these results, it appears that the debt ratio is related to maintenance costs, but the percent profit loss is not. The R^2 (a measurement of the overall effect of the regression) for this multiple regression is .24. This means that 24 percent of the variance is explained by the regression, whereas 76 percent of the variance is not.

Co-linearity between the two independent variables was tested with a separate regression, with debt being the independent variable and profit lost the dependent variable. The results were:

$$Y = 62.2 + 0.1X$$

The t value was 0.69 and the R^2 was .01, indicating no relationship. Thus, we can be reasonably sure that the insignificance of profit loss is not caused by a colinear relationship with debt ratio. Based on these results, profit lost adds little or no explanatory power to the regression and is dropped in further analysis.

An important assumption of the classical linear regression model is that the variances in the error term are equal. Because the data were cross sectional, heteroscedasticity was tested for using the Park test and the Goldfeld-Quandt test. (Refer to Appendix B for an explanation of these two tests.) The results of both tests seem to indicate some presence of heteroscedasticity. Although it might be possible to compensate for this condition, no changes were made to the model.

To investigate individual differences, separate regressions were run for each airline. The results are tabulated in Table 19. Attempts were made to improve the model by grouping specific airlines or leaving specific airlines out of the matrix. Grouping the top five airlines as determined by the financial analysis in this study (American, Delta, Northwest, United, and USAir) resulted in a t value of 3.12 and an R^2 of 0.17, both lower than the original model. A grouping of the bottom four airlines (Continental, Eastern, Pan Am, and TWA) resulted in a t value of 2.66 and an R^2 of 0.16, both of which are lower than the original model and the top-five model.

The original model appears to provide the most reliable information. As an example of what the regression might predict, we can use the average debt ratio for 1989 (the most recent year for data in the model) and calculate the maintenance costs per revenue-hour for the industry as a whole. The results of this are

$$Y = 601.5 + 3.3(3.5) = \$613.05$$

The time series analysis of the debt ratio predicted that the debt ratio for the industry would grow to 4.9 by the year 1999. Substituting this value into the equation gives the following:

$$Y = 601.5 + 3.3(4.9) = \$617.67$$

If the results of the regression analysis and the debt ratio time series analysis can be accepted, maintenance costs over the next decade will increase by $4.62 per revenue-hour, or 1 percent. This alone should not significantly affect future financial performance.

ANALYSIS

Although the results of the Spearman rank correlation and the regression indicate a possible relationship between the financial and maintenance cost structure aspects of the industry, there are several factors affecting the data that should be discussed. A balance sheet shows the book value of the assets, debt, and equity accounts, and this may not represent fair market values. Had market values been used, results may have been different. The debt equity ratio as well as the WACC would probably change. Airlines are increasingly using operating leases, and this may bias the debt/equity ratio used in this study. More work and analysis is needed in this complex area.

Although one of the highest correlations within the Spearman rank correlation test was between age of fleets and maintenance expenditures, age may not be the best measurement of the condition of an aircraft. Other factors, such as cycles, the environment in which the aircraft operates, and the efficiency of inspection and maintenance procedures, also play a major role. A more in-depth study should consider all of the factors.

Maintenance expenditures between airlines may also vary as a function of fleet variations, geographic locations, and union versus non-union shops. As an alternative to maintenance costs per block-hour, a future study might use main-

TABLE 19 RESULTS OF MULTIPLE REGRESSION FOR EACH AIRLINE, 1980–1989

AIRLINE	X	t	R^2
AMERICAN	-51.4	1.48	0.22
CONTINENTAL	-0.6	0.67	0.05
DELTA	17.3	0.54	0.04
EASTERN	5.4	5.36	0.78
NORTHWEST	149.4	1.68	0.26
PAN AM	5.5	6.20	0.82
TWA	0.6	0.92	0.10
UNITED	4.5	0.09	0.00
USAIR	-25.6	0.23	0.01

tenance costs per available seat-mile. This would take into account aircraft size and flight length, but not model variance or mix efficiency.

Although the regression analysis projected a 1 percent increase in maintenance costs, the result is predicted on a "projected" debt increase. Additionally, maintenance expenses are just one element in the overall cost structure.

CONCLUSIONS

The findings of this study are inconclusive and suggest that more research needs to be done in the area of financial health, maintenance costs, and fleet age. The measures selected had numerous drawbacks. It is unclear from this analysis what direction the future capital structure will take. What effect will future liquidations, bankruptcies, and mergers have on the overall industry? Eastern has declared bankruptcy for the last time. Its assets (aircraft, gates, and routes) are being acquired by some of the remaining airlines. Will the Eastern liquidation strengthen the balance sheets and income statements of the other carriers? It is difficult to say if there will even be any short-turn gains, considering the current economic and political environment.

Recent developments in the economy at home and the political situation in the Middle East may affect the airline industry. A recession will most likely result in a reduction in forecasted growth rates, and could result in an actual decrease in passenger traffic. The Middle East crisis has resulted in increased fuel costs in the fourth quarter of 1990. Normally, increased fuel cost are passed on to the traveler in the form of fare increases. The airlines were unable to do this because of the recession. The slump comes on top of the announcement that the initial analysis of 1990 revenues for the industry shows record losses of $2 billion. Although operating revenues may be decreasing and some expenses increasing, the highly leveraged airlines are still subject to the same high levels. This could be a serious problem.

The airlines may appear superficially similar, yet their management, debt policy, route structure, fleet mix, unionization, and marketing policies can make them unique. The analysis of the Spearman rank correlation test and regression analysis results of the individual airlines compared with the industry results (Tables 18 and 19) point toward the uniqueness of the airlines. The complex nature of an airline makes it difficult to evaluate the financial structure or maintenance cost structure of the airline industry and then relate the two. None of the measures, especially the maintenance cost structure indicators, appear comprehensive enough to be considered totally adequate. It may be inappropriate to extrapolate results of the individual airlines to the industry as a whole. The results and conclusions of this study must be considered in that context.

ACKNOWLEDGMENTS

The author would first like to thank FAA for sponsoring the Graduate Research Award program and TRB for administering the program, and, in particular, Larry L. Jenney and E. Thomas Burnard for their roles in making the program a success. The author also wishes to thank the faculty members and library staff at Embry-Riddle Aeronautical University for their suggestions, help, and encouragement.

APPENDIX A Spearman Rank Correlation Test Explanation

The Spearman rank correlation is a nonparametric statistical test. It has the following advantages: (a) It uses rankings, (b) it makes no assumptions about the distribution of the population, and (c) it does not assume normality. The Spearman rank correlation coefficient formula is

$$R = 1 - 6 * \mathrm{SUM}(X - Y)^2/N(N^2 - 1)$$

R is the correlation coefficient that is calculated in the equation. The $\mathrm{SUM}(X - Y)^2$ represents the sum of the squares of the difference in ranks, with X representing the financial status and Y the mechanical status. The difference or variance is squared for two reasons. First, the test is concerned with the absolute value of all deviations. Second, the test exaggerates large variances because they are considered critical to the results. One variance of 9 is considered to have more of an impact on the results than three variance of 3, which would add up to 9. Thus, by squaring the variances, the one variance of 9 contributes 81 to the total variance, whereas the three variances of 3 would contribute 27 (3×3) to the total variance. N stands for the number of airlines that have been ranked.

Substituting the data from Table 18,

$$R = 1 - 6 * 94.5 / 9(81 - 1) = .2125$$

The T-test formula is

$$t = R / [(1 - R^2)/(N - 2)]^{\frac{1}{2}}$$

R is the correlation coefficient calculated using the Spearman rank formula as indicated above. $N - 2$ represents the degrees of freedom (df), and N is the population.

Substituting the values for this study,

$$t = .2125 / [(1 - (.2125)2/(9 - 2)]^{\frac{1}{2}} = .5754$$

APPENDIX B Regression Analysis Explanation

The regression equation takes the following form:

$$Y = a + b_1X_1 + b_2X_2$$

Y represents the dependent variable maintenance dollars per hour, and a represents the y-axis intercept. X_1 represents the first independent variable, the debt ratio. The coefficient b_1 represents the partial effect on Y caused by a change in the debt ratio. X_2 represents the second independent variable,

the percentage of operating profits expended on interest payments. The coefficient b_2 represents the partial effect on Y caused by a change in the percentage of operating profits expended on interest payments.

The Park test is run using the following regression:

$$\ln e^2 = a + b \ln X$$

The natural log is represented by ln. If the coefficient of the independent variable, b, turns out to be statistically significant, it would suggest that heteroscedasticity is present in the data. The park test is a two-stage procedure. First, the normal regression equation is run to obtain e. Then the preceding equation is run using the natural log of e^2 and the natural log of the independent variable.

To perform this the Goldfeld Quandt test, the data are divided into two samples, one with the low values and the other with the high values. A separate regression is then run for each set of data. *ESS* is then calculated. An F value is then determined:

$$F = ESS_1 / ESS_2$$

An F value greater than the critical value as determined in the F table would indicate the presence of heteroscedasticity.

REFERENCES

1. N. Rose. Profitability and Product Quality: Economic Determinants of Airline Safety Performance. *Journal of Political Economy*, Vol. 98, No. 5, 1990, pp. 944–964.
2. U.S. Department of Transportation. *FAA Aviation Forecasts—FY 1990–2001*. Report FAA-APO-90-1. Office of Aviation Policy and Plans, Washington, D.C., 1990.
3. D. Henderson. Thorny Issues Cloud Profit Picture for 1990. *Air Transport World*, January 1990, pp. 32–55.
4. Office of Aviation Information Management. *Air Carrier Financial Statistics Quarterly*. Office of Aviation, Washington, D.C., 1990.
5. C. Fotos. Heavy Debt Could Put Airlines at Risk If Traffic Growth Slows. *Aviation Week & Space Technology*, March 20, 1989, p. 206.
6. D. Premo. Fleet Financing. *Airline Executive*, February 1988, pp. 22–31.
7. P. Flint. The Joy of Leasing. *Air Transport World*, June 1989, pp. 24–6.
8. *Moody's Bond Record*. Moody's Investment Service, New York, N.Y., 1989.
9. *Moody's Transportation Manual*. Moody's Investment Service, New York, N.Y., 1989.
10. E. Brigham. *Fundamentals of Financial Management*. Dryden Press, Chicago, Ill., 1989.
11. *The Value Line Investment Survey*. Value Line, Inc., New York, N.Y., March 1990.
12. B. S. McMullen. *Profits and the Cost of Capital to the United States Trunk Airline Industry*. Unpublished doctoral dissertation, University of California, Berkeley, 1979.
13. *The Financial Desk Book*. Consolidated Capital Communications Group, Inc., Emeryville, Calif., 1985.
14. U.S. General Accounting Office. *Aging aircraft*. GAO/RCED-90-75. General Accounting Office, Washington, D.C., 1990.
15. *Quarterly Aircraft Operating Costs and Statistics*. Avmark, Inc., Arlington, Va., 1989.
16. Economics, Not Safety Is Key Issue in Replacement of Older Transports. *Aviation Week & Space Technology*, July 24, 1989, p. 69.
17. Civil Aeronautics Board. *Air Carrier Traffic Statistics*. Government Printing Office, Washington, D.C., 1984.