Volume 8 Number 2, 2005
ISSN 1094-8848

JOURNAL OF TRANSPORTATION AND STATISTICS

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

JOURNAL OF TRANSPORTATION AND STATISTICS

A PEER-REVIEWED JOURNAL

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 8 Number 2, 2005
ISSN 1094-8848

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

The *Journal of Transportation and Statistics* releases three numbered issues a year and is published by the

Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 7th Street SW, Room 4117
Washington, DC 20590
USA
journal@bts.gov

Subscription information

To receive a complimentary subscription:
mail Product Orders
Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 7th Street SW, Room 4117
Washington, DC 20590
USA
phone 202.366.DATA
internet www.bts.dot.gov

Information Service
email answers@bts.gov
phone 800.853.1351

Cover and text design Susan JZ Hoffmeyer
Cover photo Marsha Fenn

The Secretary of Transportation has determined that the publication of this periodical is necessary in the transaction of the public business required by law of this Department.

JOURNAL OF TRANSPORTATION AND STATISTICS
Volume 8 Number 2
2005

Contents

Papers in This Issue

Precision of Geocoded Locations and Network Distance Estimates

V.S. CHALASANI[1]
J.M. DENSTADLI[2]
Ø. ENGEBRETSEN[2]
K.W. AXHAUSEN[1,*]

[1] Institute for Transport Planning (IVT)
ETH Zürich
8093 Zürich, Switzerland

[2] Institute of Transport Economics (TØI)
P.O. Box 6110
Etterstad
0480 Oslo, Norway

ABSTRACT

This paper addresses the accuracy of the geocoding of travel diaries, the relationships between different network-based distance estimates, and how exact estimates are when distances are self-reported. Three large-scale surveys in Norway and Switzerland demonstrate that very high precision is possible when survey protocol emphasizes the capture of addresses. The study uses the relevant and available databases and networks. Crow-fly, shortest distance path, shortest time path, and mean user equilibrium path distances are systematically related to each other, the pattern of their relationships is matched to theoretical expectations, and the impact of network resolution is reported. In the examples studied, medians of self-reported distances by distance band provide reasonable estimates of crow-fly and shortest distance path distances.

HOW MUCH PRECISION IS POSSIBLE?

Measuring distances traveled is a central task of transport statistics, as these data are not only key descriptors of travel behavior, but also essential for the calculation of derived statistics, such as exposure to risks (accidents, pollution), volume of externalities (emissions, congestion), speeds, incidence of

Email addresses:
* Corresponding author: axhausen@ivt.baug.ethz.ch
V.S. Chalasani—chalasani@ivt.baug.ethz.ch
J.M. Denstadli—jmd@toi.no
Ø. Engebretsen—oen@toi.no

KEYWORDS: Geocoding, travel diary, precision, network distances, detour factors.

taxation, and so forth. It is also central, directly or indirectly, to all choice models estimated from travel behavior data. Thus, it is not surprising that recent technological innovations, such as geographic information systems and the vast expansion of spatially referenced databases and networks have been adopted quickly by transport statisticians and modelers. This adoption process is ongoing, and professional standards for appropriate use must be formulated. This paper contributes to the current discussion: first, by highlighting various questions about the availability of these new resources and second, by reporting results from our work with these systems in Norway and Switzerland.

The gold standard of distance measurement is an uninterrupted trace of Global Positioning System (GPS) points matched to a complete and geometrically correct network model. The currently available GPS datasets are neither uninterrupted nor matched to complete and geometrically correct network models (see, for a recent example, Hackney et al. 2004; Marchal et al. 2004), but they are much closer to this standard than the alternatives discussed below. Some studies come quite close (see, e.g., Wolf et al. 2003). Lacking data of this quality, the researcher has various second-best alternatives to locate (geocode) origins and destinations of stages or trips observed[1] (Axhausen 2003) and to estimate distances between them. Data sources assumed available for further discussion are the following: travel diary surveys (Richardson et al. 1995; Axhausen et al. 2003; Resource Systems Group 1999), address databases, and network models suitable for shortest path calculations.

The quality of geocoding will depend on the details reported by travelers, as well as the details of the address databases to which these reports are matched. Travelers' difficulties with reporting addresses are well known: full street addresses may not be known for shops and other locations; correct postal codes are forgotten, even when the street address is known; or no unique names exist for common meeting points in parks or other public spaces. Address databases have similar problems: no entries for points in public spaces; arbitrary allocation of reference points for large complexes, such as train stations, airports, or shopping centers; and some missing street addresses.

Using zones for modeling convenience or privacy protection increases both complexity and the possibility for error. The definition of a reference point for a zone is an additional problem in its own right. Should one use the geographical mean of the zone, the built-up area, the center of gravity of the population, the city hall, or the post office for zones defined by a postal code?

Currently available detailed network models for vehicle navigation will be almost perfect from a topological perspective, as they include (nearly) all street addresses and all nodes. However, minor delays in the updating of such databases can cause minor errors. The larger issue is the coding of link types and associated mean speeds for link types. The same problems (with larger impacts on accuracy) occur with planning networks, that is, networks used in planning applications for assignment or other transport flow algorithms (Ortuzar and Willumsen 2001; Sheffi 1985). These contain far fewer links and nodes, causing inconsistencies between shortest paths calculated using them in comparison with using navigation networks. An added complication is their use of zones to represent space with all the related definition problems discussed above. Further, network models employ special types of links to connect zones with networks. One such connector is required to produce a complete description of the area, but many users employ two or more, which again will impact shortest path calculations.

Road geometry in network models only approximates the true geometry of real road alignments. As long as the true length of links is known, locating a street address along a link will add only minor errors.

Network models can be used to calculate path distances between origins and destinations for different criteria that might or might not have the same values—for example:

- shortest distance path,
- shortest time path,
- paths included in the set of paths traveled at user equilibrium,

[1] A stage is the movement with one mode; a trip is the sequence of stages between two activities; a journey is a sequence of trips starting and ending at the current residence of the traveler, generally the home (Axhausen 2003).

- paths included in the set of paths traveled at stochastic user equilibrium, and
- paths included in the set of paths traveled at system optimum.

For the last three criteria, one would need to define summaries of returned path distances, for example, mean, median, or minimum. The complexities involved in estimating origin-destination matrices required for these calculations are not included here (see Ortuzar and Willumsen 2001 for details).

Calculation of the shortest distance path distance is unambiguous, which is not the case for shortest time path distance, which requires the modeler to make assumptions about traveling speeds on the various links. One obvious assumption is the free-flow speed, normally the posted speed limit, available in all assignment networks. Most networks set up for navigation purposes assume a mean speed for each link type. These are substantially lower than free-flow speeds. Other a priori choices are possible. We can also calculate the straight line (crow-fly) distance between two points, either as Euclidian distance or as Great Circle distance (Hubert 2003), that takes the Earth's spherical shape into account.

When we consider the number of possible combinations and choices in network distance calculation, traveler-reported distances are at least unambiguous. Travelers choose a path based on specific preferences and situations. We can expect their self-reported distances will deviate from any modeled distance because of their tendency to estimate distances imprecisely (Bovy and Stern 1990; Rietveld et al. 1999; Raghubir and Krishna 1996). In many cases, though, this is the only information available. Thus, patterns of deviations between reported and modeled distances are of interest.

Although not yet undertaken, a study of the interactions between all these elements would be helpful. This paper focuses on many of the relevant issues that provide some missing background and allow other results to be assessed:

- What degree of accuracy is possible in the geocoding of addresses obtained from travel diaries? The results of three studies, the Swiss national travel diary survey (Mikrozensus 2000), the 2003 Thurgau six-week diary (Thurgau 2003), and the 2001 Norwegian national passenger travel survey (NPTS 2001) are compared.
- How large are the differences between various distance estimates? Using a current national assignment model for Switzerland (Vritc et al. 2003; Vritc and Axhausen 2004), shortest distance path distances, shortest time path distances, and mean user equilibrium path distances will be calculated and compared.
- What are the differences between reported distances and calculated distances? The three datasets will be used to answer this question.

DATASETS

2001 Norwegian National Passenger Travel Survey

The 2001 NPTS is the latest in a series of Norwegian travel surveys, which are undertaken on a four-year cycle (Denstadli et al. 2003). The respondents, all of whom are at least 13 years old, reported both their trips for one day and all trips over 100 kilometers made during the last month in a computer-aided telephone interview (CATI). They were asked to fill in a "memory jogger" before the interviews. Respondents were drawn from the national person register, which allows pre-geocoding of home and work place addresses.

The published dataset gives addresses at the level of the approximately 14,000 statistical wards, which is how the census office divides Norway. These vary in population from 0 to 3,500, with a mean of 320. The geocoding of the 64,240 daily trips and 27,507 long-distance journeys involved two automatic matches and two manual correction phases against a set of address databases, including one with the names of firms and organizations (Denstadli and Hjorthol 2003).

Swiss National Travel Survey

The Swiss Federal Office of Statistics (BFS) and the Federal Office of Spatial Planning (ARE) conducted the Mikrozensus 2000, the sixth in a series dating back to 1974 (BFS 2001 and 2002). A number of cantons provided further support by financing additional respondents at marginal costs. The CATI-interview covered the stages of one entire day and

long-distance and air travel for longer periods. The feasibility of geocoding the stage data was still uncertain during the survey's design phase, so exact street addresses or their equivalents were obtained only for trips to, within, and from the 10 largest cities in Switzerland (40,000 to 340,000 inhabitants). The names of stations and public transport stops were carefully recorded as part of the stage-based interview, as well as home addresses. However, the quality of the address information was not a prime concern for the survey.

The geocoding (Jermann 2003) of the 144,000 stages (about 100,000 trips[2]) was performed some time after the field phase of the survey, as part of a different project. Using geocoded address databases of the BFS, canton Zürich, and the Swiss Federal Railways stations and stops, we implemented a semi-automatic matching process after normalizing and correcting street addresses in the Mikrozensus 2000 records (spelling, punctuation, removal of diacritical marks, etc.). The remaining addresses were matched by hand, as far as possible, using maps, telephone books, and information on the internet, especially for place names and leisure facilities. (The address-matching tools in ArcInfo and MapInfo were unsuitable, because they embed too many assumptions valid only in the context of the United States).

2003 Thurgau Six-Week Diary

This survey replicates and improves on the six-week Mobidrive survey (Axhausen et al. 2002). A total of 99 households with 230 members were recruited in the rural and small town canton of Thurgau; they reported their travel for a continuous six-week period, using six one-week trip diaries (about 36,000 trips). The data were then coded and the field worker called respondents to clarify any omissions, particularly omitted or unclear addresses. (Address information quality was a priority for everyone involved in the survey.)

The geocoding was undertaken (Machguth and Löchl 2004) after the completion of the field work using the same type of databases employed for the geocoding of the Mikrozensus 2000 and adopting the same process. In contrast to the Mikrozensus, destinations abroad were coded to street block level in Germany and to municipality level elsewhere.

[2] Microzensus deliberately omitted many stages, in particular those under 100 meters; these omissions were exacerbated by interviewer error.

QUALITY OF GEOCODED LOCATIONS

In the preceding section, we asked what level of quality could be achieved for such large-scale exercises when they rely primarily on automatic matching steps. The quality of geocodes can be evaluated by how precisely addresses can be pinpointed. In the Norwegian study, quality was rated by quantifying the number of wards to which an address could belong. Table 1 gives details of the criteria for quality rankings. In nearly 90% of the cases, it was possible to locate the address within one ward. However, address locations for both ends of the trip were possible in only about 80% of the cases, raising problems later with distance calculations (table 2). Trip purpose, mode, and area were investigated for impacts on accuracy. The first two were not significant, but the type of area, predictably, had an impact. Better databases for larger urban areas substantially improved quality, particularly when the wards considered are smaller in these areas.

The matching quality of data on location in Mikrozensus 2000 needed to be examined individually for each stage, as these were the basic units of the data collection. Varying quality of underlying databases produces differences. Because some addresses were available only with street names, and in most cases only as municipalities, the collection of addresses differed for various areas during the survey. Table 3 details the quality ratings and table 4 shows the qualities available at the origins and destinations of the stages.

Matching was very precise for stages with stations on either end, relatively good for both bus and tram stops. When street addresses were available, coding was simple. However, in one-third of the cases, respondents could only recall the street, or only a street could be identified for the location. The municipalities were matched precisely. Note that cases rated C2, which refers to locations for available street addresses, were so incomplete that matching could only be achieved at the municipal level. Slightly more than 70% of the stages could be matched at both ends to level 1 (including 14%

TABLE 1 2001 NPTS: Geo-Information and Accuracy Level

Type of information	Accuracy level
1. Pre-geocoding of home address (verified by respondent) 2. Pre-geocoding of work place address (verified by respondent) 3. Street address, postal number, and municipality; location using GIS and address databases 4. As in 3, but with some inaccuracies—manually controlled and verified 5. As in 3, but using a manual method for location 6. Insufficient information (e.g., name of store, postal code, etc.), but GIS or manual checks made possible exact location 7. Location to city center in small urban settlements (few cases)	Exact location of statistical ward
8. As in 6, but 2 possible wards	Approximate location (2 possible wards)
9. As in 6, but 3 possible wards	Approximate location (3 possible wards)
10. As in 6, but 4 or more possible wards	Inexact location (4 or more possible wards)
11. Insufficient information—only possible to locate municipality	No location
12. Geocoding impossible or destination abroad	No location

TABLE 2 2001 NPTS: Accuracy of the Geocoded Trip Origins and Destinations by Area and Location

Accuracy of geocoding	Exact location of ward (%)	Approximate location—2 or 3 possible wards (%)	Inexact location—4 or more possible wards (%)	Municipality only (%)	Sample size
Metropolitan areas with cities with over 100,000 inhabitants	81	4	10	5	18,204
Cities/towns of 40,000–100,000 inhabitants	82	5	8	5	12,690
Smaller towns/villages	78	5	11	6	13,868
Sparsely populated areas	74	4	16	6	19,478
Trip origin	89	2	6	3	64,240
Trip destination	89	2	6	3	64,240
Origin and destination	78	4	11	6	64,240

municipality to municipality stages) and 85% to level 1 or 2, which is roughly comparable to the Norwegian results. Considering that the average Swiss municipality has only about 2,500 inhabitants, and given that the Mikrozensus was mostly conducted without considering geocoding of locations, this result is quite good.

The geocoding quality for the 2003 Thurgau followed the Mikrozensus example, but was supplemented by a new type of coding that translated the previous codes into a more comprehensible metric (table 5). The code "<100m" understates the accuracy, because it covers mainly exactly coded street addresses. The quality of the geocoding is very high, reflecting the attention given to it during the survey process. With 60% of trips captured within 100 m of their true origins and destinations, this brings us very close to ideal conditions for the distance estimation.

DIFFERENCES BETWEEN DISTANCE ESTIMATES

Swiss and Norwegian data allow comparison of network estimates against reported distances, as well as against each other. This section focuses on the comparison between the various network estimates discussed above.

TABLE 3 Mikrozensus 2000: Rating of the Matching Quality by Type of Location

Rating	Description	Quality
Building address available		
A1	Precise match	Precise
A2	Varying address spelling, certain match	Certain
A3	Strongly varying spelling, uncertain match	Uncertain
Street name available		
B2	No house number available; employed lowest known number in the street	Certain
B3	As above, but uncertain match	Uncertain
Municipality known		
C1	No street address	Precise
C2	Street address given, but not identifiable locally	Certain
C3	Dubious information in the Mikrozensus	Uncertain
Bus or tram stop		
D1	Precise match	Precise
D2	Varying address spellings, certain match	Certain
D3	Strongly varying spellings, uncertain match	Uncertain
Station		
E1	Precise match	Precise
E3	Strongly varying spellings, uncertain match	Uncertain
F	Not identifiable; abroad	No match

TABLE 4 Mikrozensus 2000: Matching Quality by Stage End
Percentage of 144,329 stages

	From														
To	**A1**	**A2**	**A3**	**B2**	**B3**	**C1**	**C2**	**C3**	**D1**	**D2**	**D3**	**E1**	**E3**	**F**	**Sum**
A1	4.0	0.4	0.0	2.6	0.1	3.4	0.1	0.0	1.5	0.3	0.1	6.0	0.0	0.7	19.3
A2	0.4	0.2	0.0	0.2	0.0	0.8	0.0	0.0	0.1	0.0	0.0	1.0	0.0	0.1	3.0
A3	0.0	0.0	0.0	0.0	0.0	0.0	—	0.0	0.0	0.0	—	0.0	—	0.0	0.1
B2	2.4	0.2	0.0	1.9	0.0	1.0	0.0	0.0	0.5	0.1	0.0	1.3	0.0	0.2	7.9
B3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	—	0.0	0.0	0.0	0.0	—	0.0	0.2
C1	3.2	0.7	0.0	0.9	0.0	12.2	0.3	0.0	0.2	0.1	0.0	4.0	0.0	0.4	22.1
C2	0.1	0.0	0.0	0.1	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.0
C3	0.0	0.0	—	0.0	—	0.0	0.0	0.0	0.0	0.0	0.0	0.0	—	0.0	0.1
D1	1.5	0.1	0.0	0.5	0.0	0.2	0.0	0.0	1.7	0.2	0.1	0.8	0.0	0.1	5.3
D2	0.3	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.2	0.1	0.0	0.2	0.0	0.0	1.1
D3	0.1	0.0	—	0.1	0.0	0.0	0.0	—	0.1	0.0	0.0	0.1	0.0	0.0	0.5
E1	5.7	1.0	0.0	1.2	0.0	4.1	0.4	0.0	0.9	0.2	0.1	21.4	0.1	0.4	35.5
E3	0.0	0.0	0.0	0.0	—	0.0	0.0	—	0.0	0.0	0.0	0.1	0.1	0.0	0.3
F	0.7	0.0	0.0	0.2	0.0	0.4	0.0	0.0	0.1	0.0	0.0	0.6	0.0	1.5	3.7
Sum	18.5	2.8	0.1	7.9	0.2	22.6	0.9	0.1	5.3	1.1	0.5	36.0	0.3	3.6	100.0

Key: — = combinations with no observation.
Note: Columns and rows may not equal their marginal sum because of rounding.

In a first step for Mikrozensus 2000, the stage-based information was used to geocode the trips. The best available geocode was attached to the start of the first stage and the destination of the last stage (table 6). The main mode of the trip was determined, as is usual in this situation, by an a priori ranking of the modes involved, in which the various public transport modes have priority before private motorized vehicles and slow modes. Further analysis in this section is restricted to car driver and pas-

TABLE 5 2003 Thurgau: Matching Quality
Percentage of 36,824 trips

	Quality at destination					
Quality at origin	**< 100 m**	**100–500 m**	**500–1,000 m**	**Municipality**	**Unknown**	**Sum**
< 100 m	60.3	13.4	0.1	2.7	0.6	77.1
100–500 m	13.4	3.2	0.0	0.9	0.1	17.6
500–1,000 m	0.1	0.0	0.0	0.0	—	0.3
Municipality	2.6	1.0	0.0	0.6	0.0	4.2
Unknown	0.6	0.1	—	0.0	0.0	0.7
Sum	77.0	17.7	0.2	4.2	0.7	100.0

Note: Columns and rows may not equal sum because of rounding.

TABLE 6 Mikrozensus 2000: Quality of the Geocoding of Trip Origins and Destinations
Percentage of 104,215 trips; all modes

	Trip destination			
Trip origin	**Postal code, street name, and house number**	**Postal code and street name**	**Only postal code**	**Total**
Postal code, street name, and house number	16.8	0.0	6.2	23.0
Postal code and street name	0.0	0.0	6.3	6.3
Only postal code	0.0	0.0	70.7	70.7
Total	**16.8**	**0.0**	**83.2**	**100.0**

senger trips, as no detailed walking and cycling network information was available.

Network distance calculations were performed using a national assignment model available at the Institute for Transport Planning and Systems (Vritc et al. 2003; Vritc and Axhausen 2004), which divides Switzerland into 3,066 zones, 14,798 nodes, and 19,664 links. The associated origin-destination matrix of average annual weekday flows was calibrated for the year 2001. The geocode for a postal code is the geocode of the associated post office's address. As a municipality is normally the same as a postal code area and a zone in the national network model, this address was also used to describe the center of gravity of the zones. The distance between the network and the center of gravity, that is, the length of the centroid connector, was set to the Euclidian distance between the relevant node and the centroid.

Crow-fly distances were calculated as Euclidian distances between the origin and destination of the trip, at the precision available. For network-based calculations, each trip end was associated with the relevant zone and, therefore, its zonal centroid. Distances were calculated using VISUM 8.0 (PTV AG 2002) for about 3,000 zones with an average of 2,500 residents. Shortest distance path distances included lengths of centroid connectors at either end of trips. Shortest time path distances were calculated assuming free-flow speeds for links. User-equilibrium (UE) assignment distances were calculated as weighted average distances of paths used at equilibrium between any two locations. The matrix of average weekday traffic flows was assigned with the assumption that daily link capacities are 12 times hourly link capacities. We excluded all trips inside a zone from further analysis, as they have, by definition, a distance of zero in network models, better interpreted as a missing value.

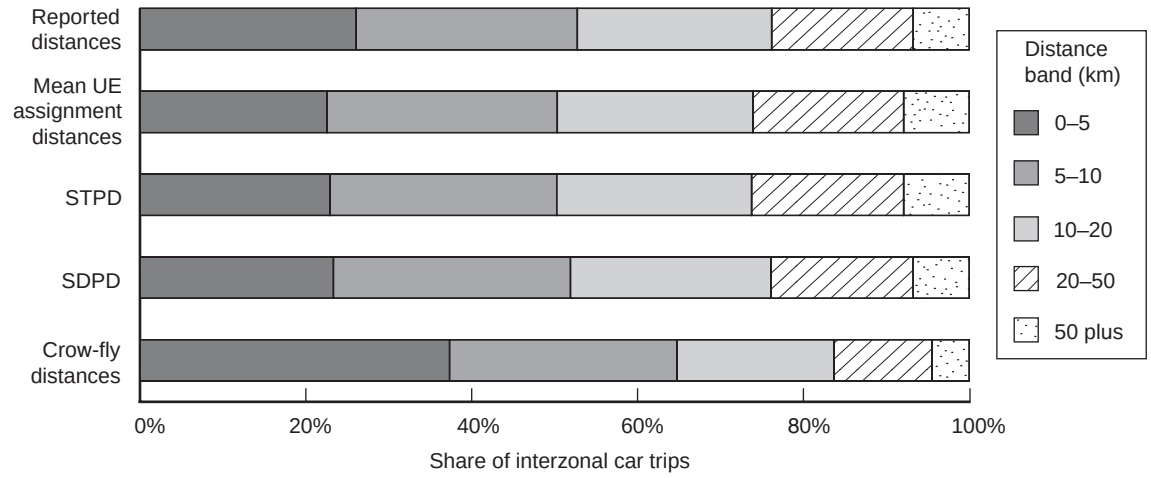
A comparison of distance distributions (table 7 and figure 1) highlights the differences between the three sources of information. The largest share of crow-fly distance trips lies in the one to five kilometer band. The mean crow-fly distance in this band is substantially smaller than the mean distances in all other bands. Network distance distributions are similar, but, as one would expect, shortest time path and mean UE assignment path distances are slightly longer. This effect is pronounced for longer distances, where routings via roads with higher speeds

TABLE 7 Mikrozensus 2000: Distribution of the Reported and Calculated Distances
34,195 car passenger and driver interzonal trips

	Crow-fly		Shortest distance path		Shortest time path		Mean UE path		Reported	
Distance band (km)	Share (%)	Class mean (km)	Share (%)	Class mean (km)	Share (%)	Class mean (km)	Share (%)	Class mean (km)	Share (%)	Class mean (km)
0–5	37.34	3.1	23.32	3.5	22.93	3.5	22.59	3.5	26.10	3.6
5–10	27.45	7.1	28.60	7.3	27.35	7.3	27.76	7.3	26.65	7.9
10–15	12.32	12.3	15.77	12.3	15.18	12.3	15.19	12.4	14.67	13.1
15–20	6.55	17.3	8.38	17.3	8.34	17.2	8.36	17.2	8.77	18.3
20–25	4.26	22.4	5.51	22.3	5.54	22.2	5.51	22.2	5.72	23.3
25–50	7.63	34.2	11.69	34.1	12.75	34.2	12.66	34.2	11.40	35.5
50–75	2.20	60.4	3.17	61.0	3.66	61.1	3.64	61.0	3.05	62.0
75–100	1.07	85.7	1.37	86.2	1.60	86.9	1.63	86.6	1.71	88.6
>100	1.18	135.2	2.20	148.2	2.67	161.0	2.65	161.5	1.94	158.7
Total	100.00	13.1	100.00	17.9	100.00	19.6	100.00	19.6	100.00	18.4

Key: UE = user equilibrium.
Note: Percentages may not add to 100 because of rounding.

FIGURE 1 Mikrozensus 2000: Comparison of the Distance Distributions
34,195 car passenger and driver interzonal trips



Key: SDPD = shortest distance path distances
STPD = shortest time path distances
UE = user equilibrum

start to pay off. Alpine topography, including the many large lakes in the foothills of the Alps, explains the large differences in the shares of trips over 100 kilometers distance vs. crow-fly distances. Mean reported distance lies between the shortest distance path and shortest time path estimate. Given that neither of the two network-based estimates reflects actual behavior fully, this mean value is a credible estimate for all trips.

In many cases, it is useful to convert one distance estimate to another. Such conversions, using the mean ratios of the relevant estimates, often called detour factors, have been reported previously but only for certain pairs of distance estimates (e.g., by Qureshi et al. 2002). Table 8 provides six comparisons for Mikrozensus 2000 based on the estimates described above. A clear difference can be observed in detour factor change patterns. Calculations are based on all observations in the sample, even if crow-fly distances were longer than model-based estimates. This can happen, especially for shorter trips, when the distance between zonal centroids is

TABLE 8 Mikrozensus 2000: Detour Factors Between Different Distance Estimates
34,195 car passenger and driver interzonal trips

	Average detour factor with respect to					
	Crow-fly distance			Shortest distance path		Shortest time path
Distance band (km)	SDPD	STPD	Mean UE distance	STPD	Mean UE distance	Mean UE distance
0–5	1.83	1.87	1.88	1.01	1.02	1.01
5–10	1.39	1.46	1.46	1.04	1.05	1.00
10–25	1.35	1.47	1.47	1.09	1.09	1.00
25–50	1.31	1.46	1.46	1.11	1.11	1.00
50–75	1.31	1.47	1.47	1.12	1.12	1.00
75–100	1.32	1.49	1.49	1.13	1.13	1.00
>100	1.26	1.48	1.48	1.16	1.16	1.00
Total	1.54	1.62	1.62	1.05	1.05	1.00

Key: SDPD = shortest distance path distance; STPD = shortest time path distance; UE = user equilibrium.

TABLE 9 2003 Thurgau: Detour Factors Between Different Distance Estimates

	Average detour factor with								
	Public transport			Car driver and passenger			Slow modes		
	Crow-fly distances		SDPD	Crow-fly distances		SDPD	Crow-fly distances		SDPD
Distance band (km)	SDPD	STPD	STPD	STPD	SDPD	STPD	SDPD	STPD	STPD
0–5	1.33	1.38	1.05	1.46	1.50	1.04	1.44	1.49	1.04
5–10	1.46	1.51	1.02	1.35	1.40	1.02	1.67	1.73	1.01
10–25	1.26	1.32	1.05	1.25	1.32	1.05	1.81	1.85	1.03
25–50	1.20	1.32	1.10	1.21	1.32	1.09	1.26	1.36	1.08
50–75	1.25	1.40	1.09	1.26	1.39	1.08			
75–100	1.30	1.43	1.12	1.30	1.46	1.12			
>100	1.28	1.34	1.07	1.19	1.29	1.11			
Total	1.28	1.36	1.06	1.38	1.43	1.04	1.45	1.50	1.04

Key: SPDP = shortest distance path distance; STPD = shortest time path distance.
Note: All values shown are based on 30 or more observations.

smaller than the actual distance traveled (see above). Detour factors fall as crow-fly distances become longer. While they are well above the square root of two (a factor of the Manhattan metric for short distances), they are also much smaller for longer distances. Factors for the three network distances are, for practical purposes, identical for the shortest distance band, but diverge after this, reflecting different objective functions behind their calculation.

The pattern is reversed for shortest distance paths detour factors, where the factors grow as shortest path distances increase. This is predictable, as the chance to use a faster, but longer route via the less-crowded high-capacity network increases with trip length.

In the 2003 Thurgau survey, the distances (shortest distance path and shortest time path) were calculated using high resolution Vektor 25, a network of the Swiss ordinance survey, employing the gecodes described above. This allowed the inclusion of all trips, except for cases where respondents returned to the same address after a walk or drive. The patterns revealed in table 9 are similar to those discussed for the Mikrozensus 2000, but their levels

TABLE 10 2001 NPTS: Mean and Median Detour Factors Between STPD and Crow-Fly Distance by Distance Band
20,700 car passenger and driver trips below 100 km

	Detour factor	
Distance band (km)	**Mean**	**Median**
0–9	1.56	1.48
10–19	1.42	1.34
20–29	1.40	1.33
30–39	1.37	1.32
40–49	1.40	1.36
≥50	1.43	1.35
Total	1.51	1.42

Key: STPD = shortest time path distance.

are markedly lower for crow-flow distance ratios, reflecting the finer network employed and the absence of centroid connectors.

Distance estimate comparisons for the Norwegian data are possible only for shortest time path distance at this time. However, results confirm the pattern revealed by the Mikrozensus data; the detour factor is significantly larger in the shortest distance band (table 10). The national-level planning network data were provided by the Norwegian highway authority and the path calculation included travel times, distances, and various bridge and ferry tolls.

Figure 2 illustrates the results for the shortest time path distances. The ratio level seems to depend on resolution of the networks used. The national-level planning networks used for the Mikrozensus 2000 and 2001 NPTS produced larger ratios than the finer network used for the 2003 Thurgau survey. This is especially obvious for the shorter distance bands, while differences start to disappear over long distances.

REPORTED AND ESTIMATED DISTANCES

Unknown errors in the differences between the true length of a trip and the reported length have led modelers to avoid the use of travelers' reported distance estimates whenever possible. Expressly, when estimating choice models, the consistent errors of network models are preferable to travelers' unknown, idiosyncratic errors. But, in many cases, neither full traces nor geocodes nor network models are available. Thus, the quality of reported distances is important, especially if the differences were to cancel out for averages or other sample summaries.

One partial way to assess reported distance quality is to compare it with the shortest distance path distance derived from a network model. Such a comparison must be partial, as one cannot know if the traveler deviated from the predicted path. If the distance estimates for the model are zone-based, their measurement uncertainties due to the differences between interzonal distances can be assessed and compared with the distances between addresses.

In the 2001 NPTS, geocodes refer to statistical wards of differing size. To determine measurement uncertainty, mean distances between all ward addresses and their respective centroid were calculated for each ward (for details, see Denstadli and Engebretsen 2004). To avoid large measuring uncertainties, in the later calculations we eliminated trips to and from wards with a mean distance of more than 1.0 kilometers between addresses and the centroid. In addition, trips with obvious geocoding errors and trips where the measurement uncertainty for either statistical ward was larger than one-quarter of the network distance estimate were removed. Finally, we omitted trips that started and ended in the same ward.

Table 11 shows the resulting relative deviations by distance band for all car driver and passenger trips below 100 kilometers, which applied to the vast majority of all such trips. The measurement uncertainty is nearly independent of trip distance and fairly small, with a mean of about 0.6 kilometers. The overall deviation decreased with distance. The shares of trips in the various deviation bands were redistributed. The large share of distance estimates within the measuring uncertainty was greatest for the lowest distance band. This share went down as distance rose with a nearly matching increase in the below 5% deviation band. About 45% of trips were estimated within 10% of the shortest time path distance. Additional analysis revealed minor differences between various trip purposes, young and middle-aged people, sexes, and urban and rural areas.

Deviations in reported distances are due not only to respondent errors, but may also be caused by interviewer misinterpretation, recording errors, or

FIGURE 2 Ratios of Shortest Time Paths with Crow-Fly Distances
By distance band

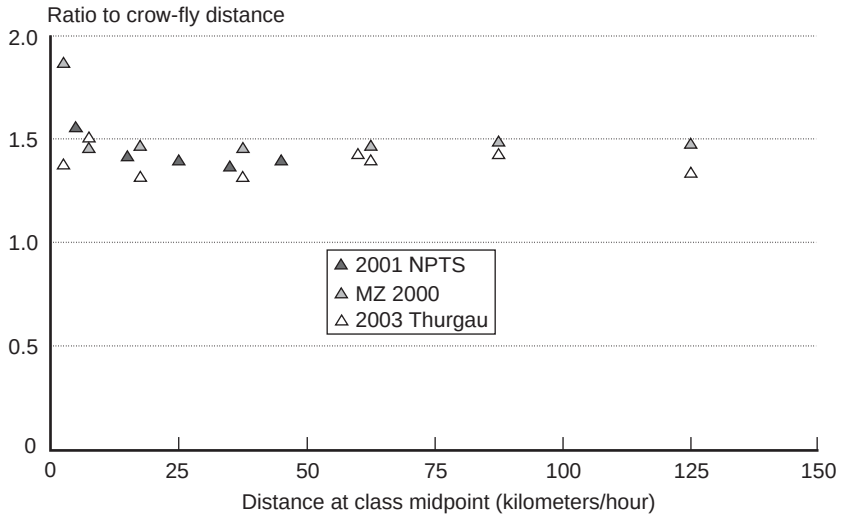


TABLE 11 2001 NPTS: Distribution of the Relative Deviations of Reported to STPD Estimates by Distance
20,700 car passenger and driver trips below 100 km

	Share of trips with relative deviations of reported to SDPD estimates (%)						
STPD (km)	Within the measuring uncertainty	<5	5–10	10–25	25–50	≥50	Total
0–9	28.2	8.7	8.2	19.5	19.1	16.3	100.0
10–19	17.1	15.9	13.2	23.8	15.9	14.1	100.0
20–29	12.6	18.0	18.1	26.4	14.1	10.7	100.0
30–39	13.4	24.3	14.8	22.6	12.1	12.8	100.0
40–49	9.5	23.9	27.6	22.1	5.5	11.3	100.0
≥50	7.1	25.5	18.4	24.1	9.4	15.5	100.0
Total	23.6	12.0	10.7	21.1	17.4	15.3	100.0

Key: SDPD = shortest distance path distance; STPD = shortest time path distance.
Note: Rows may not add to 100.0 because of rounding.

routes with freely chosen detours. We expect deviations of this kind to be more random. Note that a consistent share of deviations are in excess of 50%. Plots of reported distances against distances from the network model show that, except for some outliers, distance estimates correlate highly. Omitting the outliers, we can conclude that deviations seem randomly and asymptotically normally distributed (for details, see Denstadli and Engebretsen 2004), with the result that the mean detour factor is close to 1.0 across all distance bands (table 12).

Repeating this analysis for the 2000 Mikrozensus and 2003 Thurgau data (tables 13 and 14) also reveals a similar pattern for public transport trips. Mean detour factors are dominated by outliers over short distances. Over longer distances, the median converges quickly to 1.0 for car trips and to 1.1 for longer public transport trips. The factor drops below 1.0 for longer car trips and to about 1.2 for public transport trips. To obtain a credible estimate of distance traveled, this pattern requires adjustment of reported distances by distance band. The poorer estimates for public transport reflect the longer routing of public transport services, a lack of active navigation by the traveler, and slow access and egress to the station or stop.

The pattern is also visible in Thurgau 2003, but not as clearly. It is obvious that the very large detour

TABLE 12 2001 NPTS: Detour Factors Between Reported and Shortest Time Path Distance
20,700 car passenger and driver trips below 100 km

Shortest time path distance (km)	Detour factors	
	Mean	Median
0–9	1.11	0.96
10–19	0.99	0.99
20–29	1.00	1.03
30–39	0.96	1.02
40–49	0.99	1.02
≥50	0.91	1.01
Total	1.07	0.99

TABLE 13 2000 Mikrozensus: Detour Factors Between Reported and Shortest Distance Path Distance
Car passenger and driver and public transport interzonal trips

Distance band (km)	Average detour factor with			
	Public transport		Car driver and passenger	
	Mean	Median	Mean	Median
0–5	4.12	3.34	1.58	1.20
5–10	1.59	1.55	1.16	1.02
10–25	1.44	1.28	1.07	1.00
25–50	1.18	1.04	1.05	1.00
50–75	1.17	1.07	0.99	1.01
75–100	1.11	1.15	0.94	1.00
>100	1.16	1.18	0.83	0.99
Total	1.48	1.23	1.21	1.03

factors for short distances in Mikrozensus 2000 data are a product of omitted intrazonal trips. The very low reported distances in the longer distance band are due to the omission of hiking and cycling paths in the network model used; these can be crucial in hilly terrain. It should be noted that the speed assumptions chosen for shortest time paths were overly optimistic resulting in reported travel time underestimates of about one-third. This is far too much, even allowing for biases inherent in reported travel times. One would assume that this would lead to longer-than-realistic distances for longer trips.

The pattern of change suggests a relationship with trip speed and mode. Based on the distance bands used above, this pattern is visible in figure 3. The same pattern, but without the outlier for the short interzonal distances, can be seen in the 2003 Thurgau data.

For the Mikrozensus 2000 data, which represent a more typical situation, the dependence of the detour factor on the reported speed was modeled using aggregate values for distance bands of 2 kilometers up to 50 kilometers and of 5 kilometers beyond that. Table 15 presents the best fitting model. (For an alternative approach, see Zmud and Wolf 2003.)

CONCLUSIONS AND FURTHER RESEARCH

The three questions raised at the beginning of this paper were:

- What level of accuracy of geocoding of addresses can be obtained from travel diaries?
- How big are the differences between various distance estimates?
- What are the differences between reported distances and calculated distances?

The experiences reported here show that, in urban areas, it is possible to geocode almost all locations to within 100 meters of their true geocode, if the survey process emphasizes this aspect of the work. With even lower accuracy requirements, higher rates are possible. This carries forward to the joint accuracy of the trip length estimate, as the probability increases that both trip ends are well coded. It should be noted, though, that these rates require very good address databases, especially for firms, commercial outlets, common locations without street addresses, and public transport stations and stops. The last two categories require particular attention, as these addresses are often not available from either the relevant Census office or commercial providers. (In the case of Norway and Switzerland, it was possible to obtain relevant databases from public transport operators or the national government.) National public transport timetables include some geocoding information, but their station and stop names sometimes differ from local nomenclature.

A lower location rate for trips undertaken outside urban areas (noticeable in the 2001 NPTS, as

TABLE 14 2003 Thurgau: Detour Factors Between Reported and Shortest Distance Path Distance

	Average detour factor with					
	Public transport		Car driver and passenger		Slow modes	
Distance band (km)	Mean	Median	Mean	Median	Mean	Median
0–2.5	1.32	1.16	1.16	1.07	1.17	1.04
2.5–5	0.97	1.01	1.03	1.02	0.81	0.92
5–10	1.20	1.20	1.12	1.07	0.90	1.11
10–25	1.15	1.15	1.10	1.13	0.65	0.10
25–50	1.01	1.11	1.02	1.09		
50–75	1.10	1.17	1.14	1.13	0.33	0.06
75–100	1.12	1.16	1.04	1.08		
>100	1.13	1.14	1.10	1.06		
Total	1.32	1.16	1.16	1.07	1.17	1.04

FIGURE 3 Statistics for the Distributions of Reported Distance Deviations with Respect to Calculated Distances

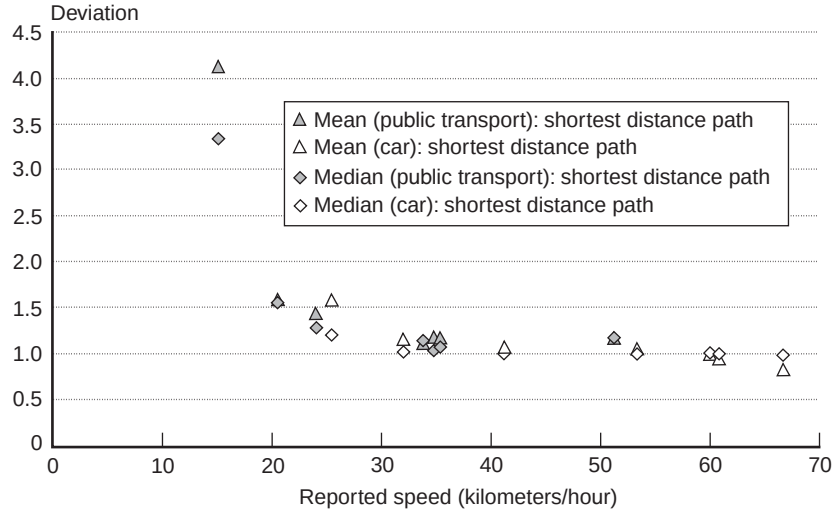


TABLE 15 Mikrozensus 2000: Linear Regression of Detour Factors Between Reported and Shortest Distance Path Distance on Reported Speed

Variable	Parameter	*t*-value
Constant	–1.94	–3.129
Reported speed	0.771	6.234
Inverse reported speed * 100	0.028	3.707
N	118	
Adjusted R^2	0.491	

well as other surveys) raises some concern. The low location rate is due to a lack of street names and identifiable landmarks like shops, churches, etc. It is important that the interviewer keep this in mind. If the respondent cannot provide an address or a landmark close by, the interviewer must make him or her describe the place in alternative ways, that is, by asking for distance and direction to the nearest lake or urban settlement, or any other marker that can help locate the trip.

We found large and systematic differences in network distance estimates, as expected. It is crucial that the modeler report the assumptions behind the estimates used. The 2003 Thurgau data show that speed assumptions behind the shortest time path distances can be crucial; detour factors provided here give a first impression of their size and pattern. However, they cannot be corroborated until the lit-

erature provides further estimates of their value. Still, the impact of network resolution is already visible in the results reported here.

Differences between reported and estimated distances can be very large for an individual trip. These errors do not cancel out for large samples. A systematic difference remains, but its pattern is predictable and depends on the trip distance. For longer trips, the medians of reported distances match the shortest distance path distances. Correcting for reported speed, there are no differences in detour factors between modes. The strong dependence on reported speed suggests a reasonable way to correct estimates.

Although we do not recommend using self-reported information as the only data for travel distances, self-reported distances are useful when assessing the quality of geocoding. Large deviations between two distance measures may indicate that the error lies in an incorrectly located start or end point and not the respondent's stated travel distance. There may also be errors in digital road data or logical defects in models determining the route (and consequently the distance). In addition, as long as objective measurements relate only to distances between zones (e.g., statistical wards), self-reported distances represent valuable additional information on short trips and intra-zone trips.

Three surveys do not allow wide generalizations. Replication of this work is required to establish the robustness of the results presented here. Discrepancies due to different formulations of network models are especially important, as substantial variance in professional practice exists, which should be reduced to improve accuracy and consistency of the model results. This zeros in on the most important element missing for further research: a high-quality GPS dataset matched to an equally high-quality network model as the basis for detailed studies.

ACKNOWLEDGMENTS

The authors are grateful for the support of H. Machgut and J. Jermann during the geocoding of the Swiss data and for the support of Mr. M. Vrtic and Mr. T. Hamre, who provided the network distance estimates for the 2000 Mikrozenus and the 2001 NPTS, respectively. The results are our own and do not reflect the assessment of the owners of the datasets used.

REFERENCES

Axhausen, K.W. 2003. Definitions and Measurement Problems. *Capturing Long Distance Travel.* Edited by K.W. Axhausen, J.L. Madre, J.W. Polak, and P. Toint. Baldock, Hertfordshire, England: Research Science Press.

Axhausen, K.W., A. Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt. 2002. Observing the Rhythms of Daily Life: A Six-Week Travel Diary. *Transportation* 29(2):95–124.

Axhausen, K.W., J.L. Madre, J.W. Polak, and P. Toint, eds. 2003. *Capturing Long Distance Travel.* Baldock, Hertfordshire, England: Research Science Press.

Bovy, P.H.L. and E. Stern. 1990. *Route Choice: Wayfinding in Transport Networks.* Dordrecht, Netherlands: Kluwer.

Bundesamt für Raumentwicklung, Bundesamt für Statistik (BFS). 2001. Mobilität in der Schweiz, Ergebnisse des Mikrozensus 2000 zum Verkehrsverhalten, Bern und Neuenburg.

____. 2002. Mikrozensus Verkehrsverhalten 2000, Hintergrundbericht zu, Mobilität in der Schweiz, Bern und Neuenburg.

Denstadli, J.M. and Ø. Engebretsen. 2004. Testing the Accuracy of Self-Reported Geoinformation Travel Surveys, paper submitted to the Conference on Progress in Activity-Based Analysis, Maastricht, Netherlands, 28–31 May.

Denstadli, J.M. and R.J. Hjorthol. 2003. Testing the Accuracy of Collected Geoinformation in the Norwegian Personal Travel Survey: Experiences from a Pilot Study. *Journal of Transport Geography* 11(1):47–54.

Denstadli, J.M., R. Hjorthol, A. Rideng, and J.I. Lian. 2003. *Travel Behaviour in Norway,* TØI report, 637/2003. Oslo, Norway: Institute of Transport Economics.

Hackney, J., F. Marchal, and K.W. Axhausen. 2004. Monitoring a Road System's Level of Service: The Canton Zürich Floating Car Study 2003, paper presented at the 84th Annual Meetings of the Transportation Research Board, Washington, DC, January 2005.

Hubert, J.P. 2003. GIS-Based Enrichment. *Capturing Long Distance Travel.* Edited by K.W. Axhausen, J.L. Madre, J.W. Polak, and P. Toint. Baldock, Hertfordshire, England: Research Science Press.

Jermann, J. 2003. Geokodierung Mikrozensus 2000. *Arbeitsbericht Verkehrs- und Raumplanung,* 177. Zürich, Switzerland: Institute for Transport Planning.

Machguth, H. und M. Löchl. 2004. Geokodierung 6-Wochenbefragung Thurgau 2003. *Arbeitsbericht Verkehrs- und Raumplanung,* 219. Zürich, Switzerland: Institute for Transport Planning.

Marchal, F., J.K. Hackney, and K.W. Axhausen. Forthcoming. Efficient Map-Matching of Large GPS Data Sets: Tests on a

Speed Monitoring Experiment in Zurich. *Transportation Research Record*.

Ortuzar, J. de D. and L.G. Willumsen. 2001. *Modelling Transport*. Chichester, England: Wiley.

Planung Transport Verkehr AG (PTV AG). 2002. U*ser Manual VISUM 8.0*. Karlsruhe, Germany.

Qureshi, M.A., H. Hwang, and S. Chin. 2002. Comparison of Distance Estimates for the Commodity Flow Survey Based on the Great Circle Distance Versus Network Based Distances. *Transportation Research Record* 1804:212–216.

Raghubir, P. and A. Krishna. 1996. As the Crow Flies: Bias in Consumers' Map-Based Distance Judgments. *Journal of Consumer Research* 23(1):26–39.

Resource Systems Group. 1999. Computer-Based Intelligent Travel Survey System: CASI/Internet Travel Diaries with Interactive Geo-Coding, report to the U.S. Department of Transportation.

Richardson, A.J., E.S. Ampt, and A.H. Meyburg. 1995. *Survey Methods for Transport Planning.* Melbourne, Australia: Eucalyptus Press.

Rietveld P., B. Zwart, B. Van Wee, and T. van den Hoorn. 1999. On the Relationship Between Travel Time and Travel Distance of Commuters: Reported Versus Network Travel Data in the Netherlands. *The Annals of Regional Science* 33(3):269–287.

Sheffi, Y. 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Englewood Cliffs, NJ: Prentice-Hall.

Vrtic, M. and K.W. Axhausen. 2004. Forecast Based on Different Data Types: A Before and After Study, paper presented at the 10th World Conference on Transport Research, Istanbul, Turkey, July.

Vrtic, M., P. Fröhlich, and K.W. Axhausen. 2003. Schweizerische Netzmodelle für Strassen- und Schienenverkehr. *Jahrbuch 2002/2003 Schweizerische Verkehrswirtschaft*. Edited by T. Bieger, C. Laesser, and R. Maggi. St. Gallen, Switzerland.

Wolf, J., M. Oliveira, and M. Thompson. 2003. The Impact of Trip Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study, paper presented at the 83rd Annual Meetings of the Transportation Research Board, Washington, DC, January.

Zmud, J. and J. Wolf. 2003. Identifying the Correlates of Trip Misreporting: Results from the California Statewide Household Travel Survey GPS Study, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, August.

Respondent Behavior in Discrete Choice Modeling with a Focus on the Valuation of Travel Time Savings

DAVID A. HENSHER*
JOHN M. ROSE

Institute of Transport and Logistics Studies
Faculty of Economics and Business
University of Sydney
NSW 2006 Australia

ABSTRACT

For models of discrete choice and their parameter estimates we examine the impact of assuming that all attributes are deemed relevant to some degree in stated choice experiments, compared with a situation where some attributes are excluded (i.e., not attended to) by some individuals. Using information collected exogenous of the choice experiment on whether respondents either ignored or considered each attribute of the choice task, we conditioned the estimation of each parameter associated with each attribute and compare, in the context of tolled vs. free routes for noncommuting car trips, the valuation of travel time savings under the assumption that all attributes are considered and the alternative assumption of relevancy. We show empirically that accounting for the relevance of attributes will have a notable influence on the valuation of travel time savings.

INTRODUCTION

> What lies ahead for discrete choice analysis? . . . The potentially important roles of information processing, perception formation and cognitive illusions are just beginning to be explored, and behavioral and experimental economics are still in their adolescence. (McFadden 2001)

Email addresses:
* Corresponding author—davidh@itls.usyd.edu.au
J. Rose—johnr@itls.usyd.edu.au

KEYWORDS: Stated choice experiment, willingness to pay, attribute relevance.

Stated choice (SC) experiments have become a popular method to model choice behavior in transportation contexts (see Louviere et al. 2000 for an overview). The outputs of SC models (e.g., willingness-to-pay estimates), have been used extensively to understand and model choice behavior (e.g., Jovicic and Hansen 2003; Jou 2001; Lam and Xie 2002), including the determination of the viability of new infrastructure projects such as proposed toll roads (e.g., Ortúzar et al. 2000; Hensher 2001). Given the risks often associated with these projects and the potential for large financial losses if they fail, it has become increasingly important that the outputs of SC models, such as the value of travel time savings (VTTS), be both reliable and unbiased estimates of the true population behavioral parameters that they purport to represent.

Realism in SC experiments can be captured by asking respondents to make "choices" between a finite but universal set of *available* alternatives, similar to those actions they would take in real markets. However, for any individual respondent, realism may be lost if the alternatives, attributes, and/or attribute levels used to describe the alternatives do not realistically portray that respondent's experiences or, in terms of "new" or "innovative" alternatives, are deemed not to be credible (e.g., Green and Srinivasan 1978, 1990; Cattin and Wittink 1982; Wittink and Cattin 1989; Lazari and Anderson 1994). An example in which individuals sometimes make decisions that deviate strikingly and systematically from the predictions of the standard SC model is the phenomena called *availability* effects, where responses rely heavily on readily retrieved information and too little on background information (e.g., rules adopted to process information) and the relevancy of such information. Information processing is distorted by what are called *regression* and *superstition* effects, in which we are too quick to attribute elaborate causal patterns to coincidences and attach too much permanence to fluctuations, failing to anticipate regression to the mean (McFadden 2001).

Regarding the attributes and attribute levels used within an SC experiment, significant prior preparation on behalf of the analyst may reduce the possible inclusion of irrelevant or improbable product descriptors within the choice sets shown to respondents (Hensher et al. 2005). Additionally, for quantitative variables, pivoting the attribute levels of the SC task from a respondent's current or recent experience is likely to produce levels within the experiment that are consistent with those experiences and, hence, produce a more credible or realistic survey task for the respondent (e.g., Hensher In press (a)).

Researchers have expended significant effort on the design of statistically efficient choice experiments (e.g., Bunch et al. 1996; Huber and Zwerina 1996; Kanninen 2002; Kuhfeld et al. 1994; Sandor and Wedel 2001) that minimize the amount of thought required of respondents (e.g., Louviere and Timmermans 1990; Oppewal et al. 1994; Wang et al. 2001; Richardson 2002; Swait and Adamowicz 2001a and b; Arentze et al. 2003). These efforts, however, appear to have been developed without adequate recognition that respondents may process SC tasks differently. That is, there may exist heterogeneity in the information processing strategies employed by respondents. SC surveys should, therefore, be tailored to be as realistic as possible at the level of the individual respondent.

Advances in econometric modeling of discrete choices, in the form of latent class and mixed logit models, may help in uncovering preference heterogeneity for attributes. However, experience suggests that, depending on the random parameter distribution, these models will likely assign non-zero parameter estimates to individual decisionmakers, even though their marginal utility for an attribute may be zero.[1] This may apply to only a small number of decisionmakers, but a bias in the population parameter estimates is still likely to exist. Therefore, the econometric models used to estimate SC outputs need to be conditioned to assign to those individuals, who either ignore an attribute or do not have that attribute present, a zero parameter estimate.

This paper examines how we can use exogenous information on the attribute processing strategies (APS) employed by individual respondents under-

[1] This will be the case if the constrained triangular or log-normal distributions are used. While these distributions force the parameter estimates to be of the same sign, they also ensure that few, if any, individual-specific parameter estimates will be zero.

taking SC tasks, and how such information can aid in conditioning the parameter estimates derived from the econometric models fitted. Additional nondesign information that may be captured in SC surveys and assist in revealing the APS include the inclusion/exclusion plan for each attribute as well as an aggregation plan (e.g., adding up attributes such as components of travel time). In this paper, we concentrate only on the attribute inclusion/exclusion strategy employed by individual respondents.

Experimental evidence and self-reported decision protocols support the view that heuristic rules are the proximate drivers of most human behavior (McFadden 2001). The question remains as to whether rules themselves develop in patterns that are broadly consistent with random utility maximization postulates. If there are preferences behind rules, then it is possible to define them and correctly evaluate policies in terms of these underlying preferences. It not, economics will have to seek a new foundation for this task. While many psychologists argue that behavior is far too sensitive to context and effect to be useful in relating to stable preferences, this is a somewhat pessimistic view. A number of authors have challenged this position (e.g., Hensher In press (a); McFadden 2001; Swait and Adamowicz 2001a). Many behavioral deviations from the economist's standard model can be attributed to perceptual illusions, particularly in the way in which we process information, rather than a more fundamental breakdown in the pursuit of self-interest. Many of the rules we do use are essentially defensive, protecting us from mistakes that perceptual illusions may induce.

There is a link between the topic here and the debate about self-explicated methods in conjoint analysis. This is especially true in light of the use of this method in Sawtooth's ACA software, which has an option to ask respondents prior to the conjoint tasks to indicate which attribute levels they would find unacceptable. ACA then deletes these declared unacceptable levels from the experimental design for the particular individual. The debate about this method has focused on whether respondents can reliably indicate which levels are unacceptable. Evidence shows that respondents often do consider or accept levels that they initially rejected. The method used for this paper is less affected by this issue, because the attribute screening task is presented after respondents have seen all the profiles or choice sets. So, when indicating which attributes they use or consider, the respondents know the complete attribute space.[2]

Adaptive choice-based conjoint (e.g., see Toubia et al. 2004) such as ACA also customizes the attribute levels of an SC experiment shown to a respondent using the previous choices made. This, however, is not the same as customizing the actual alternatives or attributes in order to make the choice task more realistic or believable to the individual respondent. Rose and Hensher (2004) addressed the mapping of alternatives in terms of their presence or absence in reality to choice experiments at the individual respondent level; however, presence or absence of attributes at the individual level is lacking in the literature. This is somewhat surprising given that, in real markets, there will likely exist heterogeneity in the information respondents have about the attributes and attribute levels of alternatives, as well heterogeneity in terms of the salience of and preference for specific attributes. For example, one respondent may have perfect information on the safety of using a tolled route compared with a free route and possess a positive marginal utility for the attribute, while a second respondent may have no understanding of the attributes' applicability to specific routes or the attributes in general and hence possess no marginal utility for the attribute at all. SC experiments assume that all respondents have perfect information (at least on the attributes included within the experiment) and that all respondents process these attributes in the same way.

MODEL DEVELOPMENT

Consider a situation in which $q = 1,2,\ldots,Q$ individuals evaluate a finite number of alternatives. Let subscripts j and t refer to alternative $j = 1,2, \ldots,J$ and choice situation $t = 1,2, \ldots,T$. Random utility theory posits that the utility for alternative j present in choice situation t may be expressed as

$$U_{jtq} = \theta_q' x_{jtq} + \varepsilon_{jtq} \quad (1)$$

[2] We thank a referee for highlighting the point of distinction.

where

U_{jtq} is the utility associated with alternative j in choice situation t held by individual q,

x_{jtq} is a vector of values representing attributes belonging to alternative j, characteristics associated with sampled decisionmakers q, and/or variables associated with context of choice situation t,

ε_{jtq} represents unobserved influences on utility, and

θ'_q is a vector of parameters such that $\theta = \theta_1, \theta_2, ..., \theta_K$ where K is the number of parameters corresponding to the vector x_{jtq}.

In the most popular choice model, multinomial logit, the probability that alternative i will be chosen is given as

$$P(i|j) = \frac{e^{V_{itq}}}{\sum_{j=1} e^{V_{jtq}}}, \ \forall j = 1,...,i,...,J, \ \forall s = 1,...,T, \quad (2)$$

where

$$V_{jtq} = \theta'_q x_{jtq}. \quad (3)$$

Assuming a sample of choice situations, $t = 1,2,...,T$, has been observed with corresponding values x_{jtq}, and letting i designate the alternative choice situation t, the likelihood function for the sample is given as

$$L\theta = \prod_{t=1}^{T} P(i|J) \quad (4)$$

and the log-likelihood function of the sample as

$$L^*(\theta) = ln[L(\theta)] = \sum_{t=1}^{T} ln(P(i|j)). \quad (5)$$

Equation (5) may be rewritten to identify the chosen alternative i

$$L^*(\theta) = \sum_{t=1}^{T} \left[V_{itq} - ln\left(\sum_j e^{V_{jtq}} \right) \right]. \quad (6)$$

Given that θ is unknown, it must be estimated from the sample data. To do this, we used the maximum likelihood estimator of θ, which is the value of $\hat{\theta}$ at which $L(\theta)$ is maximized. In maximizing equation (6), it is usual to use the entire set of data for V_{jtq}. That is, it is assumed that across all t, all V_{jtq} and hence x_{jtq} are considered, and, as such, the levels assumed by each x in the x_{jtq} matrix are used in determining the value at which $\hat{\theta}$ maximizes the likelihood estimator of θ.

Assuming that over a sample of choice situations t, not all k variables within the x_{jtq} vector are considered in the decision process, the value of $\hat{\theta}$, which is conditioned on the assumption that all x_{jtq} are considered, will likely be biased. For those choice situations in which attribute k is excluded from consideration in the choice process, $\hat{\theta}_k$ should be equal to zero. Note that this is not the same as saying that the attribute itself should be treated as being equal to zero.[3]

In cases where attribute k is indicated as being excluded from the decision process, rather than set the value for the kth element in the x_{jtq} vector to zero and maximizing equation (6), the algorithm that searches for the maximum of equation (5) excludes that x from the estimation procedure and automatically assigns it a parameter value of zero. The parameter estimate $\hat{\theta}_k$ is then estimated solely on the sample population for which the variable was not excluded. In this sense, the process is analogous to selectivity models (which censor the distribution, as distinct from truncation). To demonstrate, consider a simple example in which there are only two variables, x_1 and x_2, associated with each of j alternatives. Denote N as the number of attribute processing strategies such that $n = 1$ represents those decisionmakers who consider only x_1 in choosing between the j alternatives, $n = 2$ represents those decisionmakers who consider only x_2, and $n = 3$ represents those decisionmakers who consider both x_1 and x_2. The likelihood is defined by the partitioning of observations based on the subset membership defined above. The likelihood function is therefore given as

[3] To demonstrate, consider the situation where attribute x_{jtq} is the price for alternative j in choice situation t. For all but Giffen goods, setting the price to equal zero will likely make that alternative much more attractive relative to other alternatives in which the price is not equal to zero. Further, the procedure for maximizing $L^*(\theta)$ will be ignorant of the fact that setting $x_{jtq} = 0$ represents the exclusion of that attribute in the choice process and will estimate a value of $\hat{\theta}_k$ assuming that the value observed by the decisionmaker in choice situation t was zero for that attribute when indeed it was not. As such, setting $x_{kjt} = 0$ will not guarantee that the parameter for that attribute will be equal to zero for that choice situation. It is, therefore, $\hat{\theta}_k$ that should be set to zero in the estimation process, not x_{kjt}.

$$L^{*}(\theta) = \sum_{t=1}^{T} \sum_{n=1}^{N} ln(P(i|j)). \quad (7)$$

The derivatives of the log likelihood for groups n_1 and n_2 have zeros in the position of zero coefficients and the Hessians have corresponding rows and columns of zeros. This partitioning of the log-likelihood function may be extended to any of the logit class of models, including the nested logit and mixed logit family of models. We used a mixed logit specification in the empirical study, in which we accounted for preference heterogeneity in the specification of random parameters where their mean and standard deviation are a function of contextual influences.

$$\boldsymbol{\theta}_{qk} = \boldsymbol{\theta}_k + \boldsymbol{\delta}_k' \boldsymbol{z}_q + \boldsymbol{\sigma}_k \exp(\boldsymbol{\theta}_k' \boldsymbol{h}_q) \boldsymbol{v}_q. \quad (8)$$

The distribution of $\boldsymbol{\theta}_{qk}$ over individuals depends in general on underlying structural parameters $(\theta_k, \delta_k, \sigma_k)$, the observed data $\boldsymbol{z}_q$, a vector $\boldsymbol{h}_q$ of M variables such as demographic characteristics that enter the variances (and possibly the means as well), and the unobserved vector of K random components in the set of utility functions

$$\boldsymbol{\eta}_q = \boldsymbol{\Gamma} \boldsymbol{\Sigma}^{1/2} \boldsymbol{v}_q.$$

The random vector $\boldsymbol{\eta}_q$ endows the random parameter with its stochastic properties. In isolating the model components, we defined v_q to be a vector of uncorrelated random variables with known variances. In the empirical study, we adopted a Rayleigh distribution (defined below) as the analytical representation of v_q. The heteroskedastic mixed logit model is detailed in Greene et al. (2006). In the next section, we discuss the empirical application in which we estimate models of the form described above.

EMPIRICAL APPLICATION

The data used to contrast models that do and do not account for the attention paid to each attribute are drawn from a study undertaken in Sydney in 2004, in the context of car-driving noncommuters making choices from a range of level-of-service packages defined in terms of travel times and costs, including a toll where applicable. The sample of 223 effective interviews, each responding to 16 choice sets, resulted in 3,568 observations for model estimation.

To ensure that we captured a large number of travel circumstances, which will enable us to see how individuals trade off different levels of travel times with various levels of tolls, we sampled individuals who had recently taken trips of various travel times (called trip length segmentation) in locations with tollroads.[4] To ensure some variety in trip length, three segments were investigated: no more than 30 minutes, 31 to 60 minutes, and more than 61 minutes (capped at two hours).

A telephone call was used to establish eligible participants from households stratified geographically, and a time and location agreed on for a face-to-face computer-aided personal interview. An SC experiment offers the opportunity to establish the preferences of travelers for existing and new route offerings under varying packages of trip attributes. The statistical state of the art of designing SC experiments has moved away from orthogonal designs to D-optimal designs (see below and Rose and Bliemer 2004; Huber and Zwerina 1996; Kanninen 2002; Kuhfeld et al. 1994; Sandor and Wedel 2001). The behavioral state of the art has moved to promoting designs that are centered around the knowledge base of travelers, in recognition of a number of supporting theories in behavioral and cognitive psychology and economics, such as prospect theory, case-based decision theory and minimum-regret theory.[5] Starmer (2000, p. 353) makes a very strong plea in support of the use of reference points (i.e., a current trip):

> While some economists might be tempted to think that questions about how reference points are determined sound more like psychological than economic issues, recent research is showing that understanding the role of reference points may be an important step in explaining real economic behaviour in the field.

The two SC alternatives are unlabeled routes. The trip attributes associated with each route are summarized in table 1. These were identified from reviews of the literature and through the effectiveness of previous VTTS studies undertaken by Hensher (2001).

[4] Sydney has a number of operating tollroads; hence, drivers have a lot of exposure to paying tolls. Indeed, Sydney has the greatest amount on urban kilometers under tolls than any other metropolitan area.

[5] See Starmer 2000; Hensher 2004; Kahnemann and Tversky 1979; Gilboa et al. 2002.

TABLE 1 Trip Attributes in Stated Choice Design

Routes A and B
Free flow travel time
Slowed down travel time
Trip travel time variability
Running cost
Toll cost

All attributes of the SC alternatives are based on the values of the current trip. Variability in travel time for the current alternative was calculated as the difference between the longest and shortest trip time provided in non-SC questions. The SC alternative values for this attribute are variations around the total trip time. For all other attributes, the values for the SC alternatives are variations around the values for the current trip. The variations used for each attribute are given in table 2.

The experimental design has 1 version of 16 choice sets (games), with no dominance given the assumption that less of all attributes is better. The distinction between free flow and slowed down time is designed to promote the differences in the quality of travel time between various routes—especially a tolled route vs. a nontolled route—and is separate from the influence of total time. Free flow time is interpreted with reference to a trip at 3 a.m., when there are no traffic delays.[6] Figure 1 illustrates an example of an SC screen, and figure 2 shows a screen with elicitation questions associated with attribute inclusion and exclusion.

In choosing the most statistically efficient design, the literature has tended toward designs that maximize the determinant of the variance-covariance matrix, otherwise known as the Fisher information matrix, of the model to be estimated. These so-called D-optimal designs require explicit incorporation of prior information about the respondents' preferences.[7] In determining the D-optimal design, it is usual to use the inversely related measure to calculate the level of D-efficiency, that is, minimize the determinant of the inverse of the variance-covariance matrix. The determinant of the inverse of the variance-covariance matrix is known as D-error and will yield the same results maximizing the determinant of the variance-covariance matrix.

The log-likelihood function of the multinomial logit model is shown as

$$L = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{njs} ln(P_{njs}) + c \qquad (9)$$

where y_{njs} is a column matrix with 1 indicating that an alternative j was chosen by respondent n in choice situation s and 0 otherwise, P_{njs} represents the choice probability from the choice model, and c is a constant. Maximizing equation (9) yields the maximum likelihood estimator, $\hat{\theta}$, of the specified choice model given a particular set of choice data. McFadden (1974) showed that the distribution of $\hat{\theta}$ is asymptotically normal with a mean, θ, and covariance matrix

[6] This distinction does not imply that there is a specific minute of a trip that is free flow per se, but it does tell respondents that there is a certain amount of the total time that is slowed down due to traffic, for instance, and hence a balance is not slowed down (i.e., the trip is free flow like that observed typically at 3 a.m.).

[7] Orthogonal designs also require prior information in order to choose the attribute levels in such a way that dominating and inferior attributes are avoided. Optimal designs will be statistically efficient but will likely have correlations; orthogonal fractional factorial designs will have no correlations but may not be the most statistically efficient design available. Hence, the type of design generated reflects the belief of analysts as to what is the most important property of the constructed design. Carlsson and Martinsson (2003) used Monte Carlo simulation to show that D-optimal designs, like orthogonal designs, produce unbiased parameter estimates but that the former have lower mean.

TABLE 2 Profile of the Attribute Range in the SC Design

	Free flow time	Slowed down time	Variability	Running costs	Toll costs
Level 1	–50%	–50%	5%	–50%	–100%
Level 2	–20%	–20%	10%	–20%	20%
Level 3	10%	10%	15%	10%	40%
Level 4	40%	40%	20%	40%	60%

FIGURE 1 Example of a Stated Choice Screen

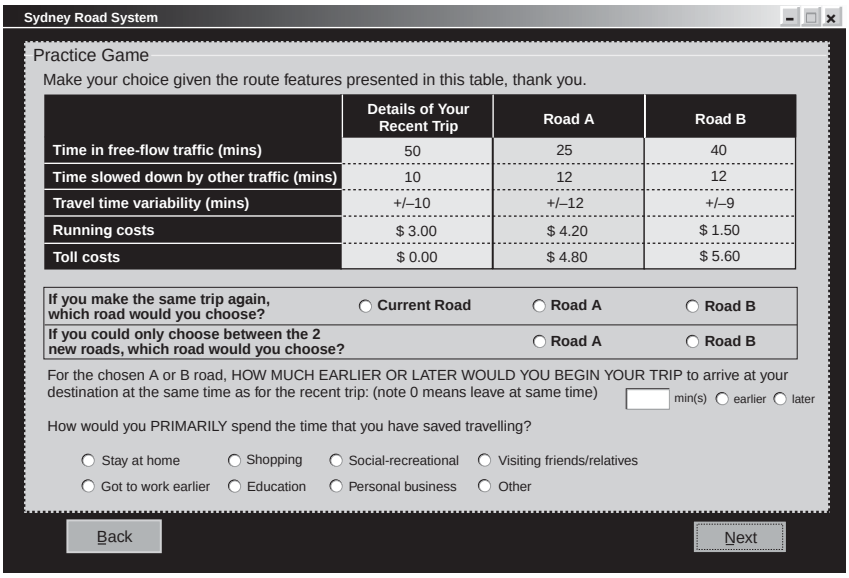
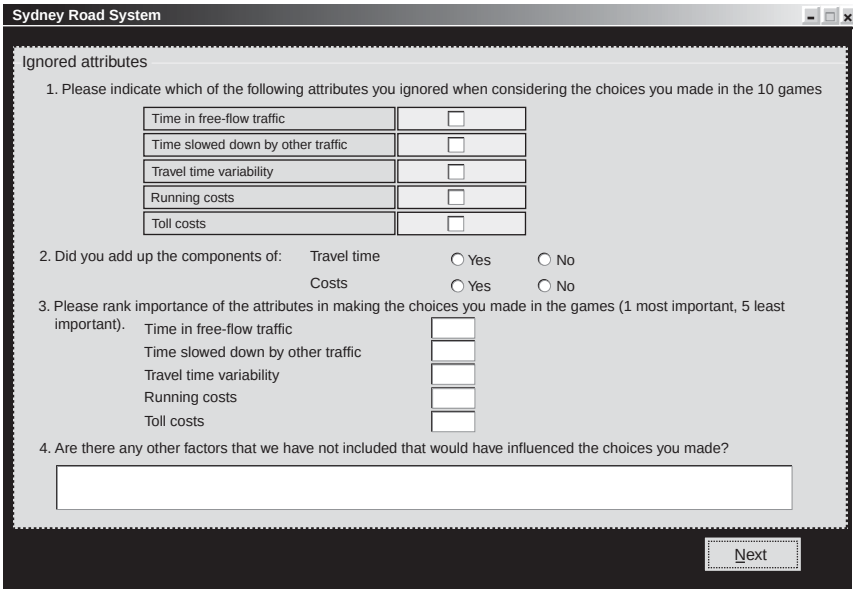


FIGURE 2 Computer-Aided Personal Interview Questions on Attribute Relevance



$$\Omega = (X'PX) = \left[\sum_{m=1}^{M} \sum_{j=1}^{J} x'_{njs} P_{njs} x_{njs} \right] \quad (10)$$

and inverse,

$$\Omega^{-1} = (X'PX)^{-1} = \left[\sum_{m=1}^{M} \sum_{j=1}^{J} x'_{njs} P_{njs} x_{njs} \right]^{-1} \quad (11)$$

where P is a $JS \times JS$ diagonal matrix with elements equal to the choice probabilities of the alternatives, j, over choice sets, s. For Ω, several established summary measures of error have been shown to be useful when contrasting designs. The most popular summary measure is known as *D-error,* inversely related to D-efficiency.

$$D\text{-}error = (det\Omega^{-1})^{1/K} \quad (12)$$

where K is the total number of *generic* parameters to be estimated from the design. Minimization of equation (12) will produce the design with the smallest possible errors around the estimated parameters.

MODEL RESULTS

Table 3 presents the model results for the experiment. Model 1 uses all data irrespective of whether a sampled individual indicated they had ignored an attribute throughout the experiment or not. Model 2 took into account the exogenous information on attribute relevance.

A profile of attribute inclusion and exclusion is shown in table 4. This is the attribute processing choice set for the sample. Just over half (52%) of the sample attended to every attribute and not one respondent attended to none of the attributes. Running cost was the attribute most likely to be ignored (17.9% of the sample); in contrast, the toll cost was attended to by 96% of the sample. Free flow time was not attended to by 13% of the sample, with 8.5 percentage points of this being when both components of travel time were ignored and the focus was totally on cost. The key message is that 78% of the sample attended to the components of travel time and 69% attended to the components of cost.

For both models, all parameters associated with the design attributes were specified as generic random parameter estimates. These parameters, with the exception of travel time variability, are statistically significant and of the expected sign. In specifying the mixed logit models, we drew the parameters associated with the design attributes from an unconstrained Rayleigh distribution. Hensher (In press (b)) showed that the Rayleigh distribution in its unconstrained and constrained forms has attractive properties. In particular, it does not have the long tail that the log normal exhibits and appears to deliver a relatively small proportion of negative VTTS when the function is not globally signed to be positive. The Rayleigh distribution probability function is given in equation (13).

$$P(r) = \frac{re^{-r^2/2s^2}}{s^2} \qquad (13)$$

for $r \in [0, \infty)$. The moments about 0 are given by

$$\mu'_m \equiv \int_0^\infty r^m P(r)dr = s^{-2}\int_0^\infty r^{m+1}e^{-r^2/2s^2}dr$$

$$= s^{-2}I_{m+1}\left(\frac{1}{2s^2}\right),$$

where $I(x)$ is a Gaussian integral. The Rayleigh variable[8] is a special case of the Weibull density,[9] with parameters 2 and $s/2$ where s is the desired scale parameter in the Rayleigh distribution. The mean is centered as $s^*\sqrt{\pi/2}$ and the standard deviation is $\sqrt{(4-\pi)s^2/2}$. This distribution has a long tail, but empirically appears much less extreme than the log normal. We obtained the random parameter estimates of the mixed logit models using 500 Halton draws.

A comparison of models 1 and 2 reveals significant differences in their parameter estimates. Caution in interpretation, however, is required, because we have estimated complex nonlinear attribute functions as per equation (8), and so individual parameter estimates for the random parameters are not meaningful in isolation. The VTTS comparison, our behavioral output of interest for toll road patronage forecasting studies, provides a valid contrast and accounts for any scale differences.

The results in table 3 show the importance of accounting for heterogeneity in the mean of random parameters and heteroskedasticity in these parameters via decomposition of the standard deviation parameter estimate. The attribute inclusion rule influences the contributing effect. For example, all three random parameters are conditioned on the trip length in kilometers through decomposition of the standard deviation with strong statistical significance, yet the sign changes with respect to slowed down time. All other effects being held constant, when combined with the standard deviation of the random parameter (all being positive as required), we found that as trip length increased the standard deviation decreased, resulting in reduced heterogeneity in preferences over longer trips. The exception was when all data were considered relevant for slowed down time, with preference heterogeneity increasing as trip length increased.

Seven variables had a statistically significant influence on the mean of the three random parameters when all attributes were included; but when we

[8] In the current paper, we use a conditional (on choice made) distribution, but for an unconditional distribution, the empirical specification used is
Rayleigh = 2*(abs(log(*rnu*(0,1))))$^{0.70710678}$ where *rnu* is the uniform distribution.

[9] The Weibull(b,c) is: $w = b*(-\log U)\wedge(1/c)$.

TABLE 3 Summary of Empirical Results
500 Halton draws; 3,568 observations

	Model 1: Full data		Model 2: Account for relevance	
	Estimated parameter	**t-ratio**	**Estimated parameter**	**t-ratio**
Mean random parameters				
Free flow time	0.0893	3.95	0.0451	2.14
Slowed down time	–0.2449	–6.67	–0.0164	–0.64
Toll route constant	2.8599	5.47	2.2173	4.75
Standard deviation of random parameters				
Free flow time	0.1565	7.90	0.1303	6.99
Slowed down time	0.0360	2.81	0.1872	5.89
Toll route constant	5.0973	6.10	5.6430	6.90
Fixed parameters				
Running cost	–0.4519	–11.21	–0.3781	–10.84
Toll cost	–0.7733	–12.72	–0.6224	–11.95
Heterogeneity in the mean of random parameters: free flow time				
Lead to improved pedestrian safety	0.0016	2.84	0.0015	3.05
Heterogeneity in the mean of random parameters: slowed down time				
Provide better access to and from the city	–0.0026	–2.64	–0.0031	–3.78
Lead to improved pedestrian safety	–0.0017	–2.41	–0.0006	–0.94
Avoid traffic lights	0.0032	2.53	0.0034	2.72
Heterogeneity in the mean of random parameters: toll cost				
Provide better access to and from the city	0.0404	2.62	0.0133	0.96
Lead to improved pedestrian safety	–0.0290	–2.73	–0.0095	–0.99
Avoid traffic lights	0.0648	4.02	0.0699	4.54
Heteroskedasticity in random parameters				
Free flow by trip kms	–0.0056	–3.22	–0.0091	–3.66
Slowed down time by trip kms	0.0149	4.34	–0.0293	–3.84
Toll route constant by trip kms	–0.0229	–8.03	–0.0270	–8.80
LL(B)	–2,609.99		–2,637.38	

TABLE 4 Incidence of Mixtures of Attributes Processed

Attribute processing profile	Sample no. of observations = 3,568
All attributes attended to (v1)	1,856
Attributes not attended to:	
Running cost (v2)	640
Running and toll cost (v3)	192
Toll cost (v4)	96
Slowed down time (v5)	192
Free flow and slowed down time (v6)	304
Free flow time (v7)	112
Slowed down time and running cost (v8)	64
Free flow and slowed down time and toll cost (v9)	48

allowed for attribution exclusion for the same set of influences, three became statistically insignificant. These influences on heterogeneity around the mean are opinion variables, derived from a weighting of a response on a seven-point Likert scale of the importance of such factors associated with toll roads in general and a seven-point *"likely to deliver"* Likert scale for specific tolled routes that respondents use. A positive parameter indicates, all other influences remaining fixed, that the opinion reflects something of greater importance and/or greater likelihood of it being delivered. For example, given that the mean estimate of the random parameter for slowed down time was negative and "avoid traffic lights" had a positive parameter estimate, the presence of a strong positive effect reduces the marginal (dis)utility of slowed down time. Again, we remind readers that, strictly, the signs cannot be interpreted independently of the full effect of all contributing sources aligned with the mean, the standard deviation and the sources of decomposition around the mean, and standard deviation parameter estimates. For example, the full marginal (dis)utility effect of free flow time for model 1 is:

$$\theta_q = \{0.0893 + 0.0016 \times lead\ to\ improved\ pedestrian\ safety + 0.1565 \times \exp[-0.0056 \times trip\ kms]r\}_q \quad (14)$$

In interpreting the parameter estimates for model 2, it is important to note that the estimates are specific only to sample population segments that consider an attribute while undertaking the choice experiment. For those in the population who do not consider an attribute, the parameter estimate expression in equation (14) for that individual is zero. That is, the parameter estimates are specific to each attribute inclusion/exclusion strategy. In terms of segmentation and benefit studies, this is an important development. In traditional models, these benefit segments may be lost if the segment is small relative to the total population size.

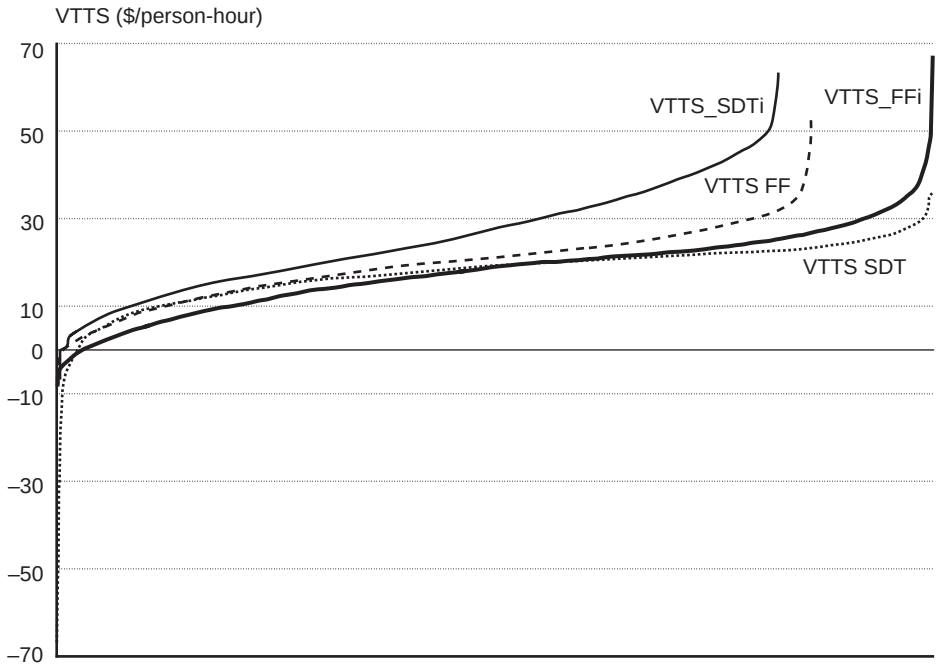
Willingness to pay (WTP) distributions for travel time savings can be derived from the conditional "individual specific" parameter estimates obtained using methods outlined in Train (2003) and Hensher et al. (2005). Estimates can be constructed of individual-specific preferences by deriving the conditional distribution based (in-sample) on known choices (i.e., prior knowledge), as originally shown by Revelt and Train (2000). These conditional parameter estimates are strictly same-choice-specific parameters or the mean of the parameters of the subpopulation of individuals who, when faced with the same choice situation, would have made the same choices. Table 5 summarizes the VTTS based on individual parameters. Not all WTP distributions are in the positive range (figure 3); indeed, the percentage that is negative is small (up to 2.9%) but substantially higher when we assume that all attributes are relevant for all respondents.

Given the differences in variances of the VTTS distributions over the models for the same attribute, we conducted a Kruskall-Wallis test, which is the nonparametric equivalent to the ANOVA test (Siegel

TABLE 5 Values of Travel Time Savings (VTTS)
$ per person-hour noncommuter car driver

	Model 1: All attributes assumed to be attended to		Model 2: Deterministic attribute exclusion	
Attribute	**Sample mean**	**Sample standard deviation**	**Sample mean**	**Sample standard deviation**
Free flow time	17.80	9.12	18.89	8.08
Slowed down time	17.61	7.73	24.96	11.89
Range: free flow time	–8.30 to 67.31		–4.30 to 52.49	
Range: slowed down time	–66.91 to 35.69		–6.53 to 63.26	
Proportion of negative VTTS	2.90	2.32	0.78	0.78
Ratio slowed to free flow time	0.989	0.848	1.321	1.472
Sample size	3,568	3,568	3,072	2,944

FIGURE 3 Distribution of VTTS Under Alternative Attribute Processing Rules



and Castellan 1988). For the VTTS distributions obtained from the models, chi-square statistics were obtained for the free flow and slowed down time VTTS distributions, which we compared with a critical value of 5.99 (i.e., χ^2_2 at the 95% confidence level). We concluded that both the means and variances of the VTTS distributions for both attributes were statistically different between the two models.

Figure 3 shows the VTTS distributions for the free flow and slowed down travel time attributes estimated from the two models. When all data were used in the estimation process, the VTTS distribution had a much greater range than when the attribute inclusion/exclusion strategy was accounted for.

This evidence suggests a deflating effect on VTTS when one ignores the attribute processing strategy and assumes that all attributes are attended to. When the attribute exclusion rule was not included, the mean VTTS was 94.9% and 70.6%, respectively, of the VTTS under the attribution exclusion rule. Furthermore, when all attributes were deemed

relevant, the mean VTTS for free flow and slowed down time was almost identical, in contrast to a slowed down time VTTS that was 32.2% higher than the free flow time value when the exclusion rule was invoked. The latter relationship is intuitively more appealing. When converted to time savings benefits in road projects, these differences would make a substantial difference to the user benefits, given the dominance of travel time savings.

DISCUSSION AND CONCLUSION

In this paper, we show that accounting for individual specific information on attribute inclusion/exclusion results in significant differences in the parameter estimates of and hence the willingness to pay for specific attributes in choice models. These differences arise from a form of respondent segmentation, the basis of which is respondent attribute processing. By partitioning the log-likelihood function of discrete choice models based on the way individual respondents process each attribute, the outputs of the models we estimated represent the attribute processing segments only, rather than those of the entire sample population. In this way, we can detect the preferences for different segments in the sample population based on the attribute processing strategies existing in that population. In traditional choice models, such segments will likely go undetected.

Whether an attribute should be excluded from model estimation for a specific respondent is critical to the method and the results. We recognize that there may be other ways of defining the behavioral rule for including or excluding an attribute.[10] We also recognize that it is important to understand whether the attribute was excluded simply because of cognitive burden in the survey task in contrast to a genuine behavioral exclusion with respect to the relevance of the attribute in making such choices in real markets. It could be the case that the cognitive burden associated with the survey instrument may indeed be real, as it can be real in markets with information acquisition and processing; and so care is required in separating out and accounting for all these reasoning processes. Clearly, these conditions are all legitimate members of an individual's attribute processing strategy.

Ultimately, our preferred strategy would be to tailor the stated choice experiment to the individual based on the attribute processing strategy of the respondent. How best to do this is a matter of research. One question is whether the attribute processing strategy should be determined a priori and the SC experiment fixed for each respondent over the course of the experiment or whether the strategy is determined for each distinct choice set. The former approach is appealing for reasons of simplicity, the latter for completeness given that the attribute processing strategy may be linked not only to the attributes but to the attribute levels of the experiment.

The approach we outline here, whereby we employ an SC experiment derived from a single design plan, represents the more traditional approach to conducting SC experiments; however, we were able to account for the attribute processing strategy exogenously without having to tailor the SC experiment to each individual. Still, research is required as to whether it is best to ask each respondent which attributes were ignored at the end of the experiment, as we did here, or upon completion of each choice task. As with the tailoring of the SC task, the former approach is appealing for reasons of simplicity as well as the probable limiting of cognitive burden experienced by respondents, while the latter may represent a more complete approach, given that the attributes that are ignored or considered may be a function of the attribute levels of the alternatives as well as a function of experience or fatigue as the number of choice tasks completed increases.

We conclude by noting that the proposed approach discussed here applies equally to models estimated using revealed preference (RP) data. Researchers collecting RP data must prespecify the data collected and assume, as with SC data, that the attributes of RP data are processed homogenously over the sampled population. As with SC data, this need not be the case.

[10] Preliminary unpublished research by the authors in which we treat the exclusion rule as stochastic suggests that the mean VTTS is slightly higher than the evidence based on the deterministic application of the exclusion rule. This supports a position that suggests that failure to account for attribute processing rules tends to underestimate the mean VTTS.

ACKNOWLEDGMENT

The comments of two referees improved this paper materially.

REFERENCES

Arentze, T., A. Borgers, H. Timmermans, and R. DelMistro. 2003. Transport Stated Choice Responses: Effects of Task Complexity, Presentation Format and Literacy. *Transportation Research E* 39:229–244.

Bunch, D.S., J.J. Louviere, and D. Anderson. 1996. A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models, working paper. Graduate School of Management, University of California, Davis.

Carlsson, F. and P. Martinsson. 2003. Design Techniques for Stated Preference Methods in Health Economics. *Health Economics* 12:281–294.

Cattin, P. and D.R. Wittink. 1982. Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing* 46(3):44–53.

Gilboa, I., D. Schmeidler, and P. Wakker. 2002. Utility in Case-Based Decision Theory. *Journal of Economic Theory* 105:483–502.

Green, P.E and V. Srinivasan. 1978. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research* 5(2):103–123.

____. 1990. Conjoint Analysis in Marketing Research: New Developments and Directions. *Journal of Marketing* 54(4):3–19.

Greene, W.H., D.A. Hensher, and J. Rose. 2006. Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models (NW Transport Study Data). *Transportation Research B* 40(1):75–92.

Hensher, D.A. 2001. Measurement of the Valuation of Travel Time Savings. *Journal of Transport Economics and Policy* 35(1):71–98.

____. 2004. Accounting for Stated Choice Design Dimensionality in Willingness to Pay for Travel Time Savings. *Journal of Transport Economics and Policy* 38(2):425–446.

____. In press (a). How Do Respondents Handle Stated Choice Experiments? Attribute Processing Strategies Under Varying Information Load. *Journal of Applied Econometrics*.

____. In press (b). The Signs of the Times: Imposing a Globally Signed Condition on Willingness to Pay Distributions. *Transportation*.

Hensher, D.A., J. Rose, and W.H. Greene. 2005. *Applied Choice Analysis: A Primer*. Cambridge, England: Cambridge University Press.

Huber, J. and K. Zwerina. 1996. The Importance of Utility Balance and Efficient Choice Designs. *Journal of Marketing Research* 33(3):307–317.

Jou, R. 2001. Modelling the Impact of Pre-Trip Information on Commuter Departure Time and Route Choice. *Transportation Research B* 35(10):887–902.

Jovicic, G. and C.O. Hansen. 2003. A Passenger Travel Demand Model for Copenhagen. *Transportation Research A* 37(4):333–349.

Kahnemann, D. and A. Tversky. 1979. Prospect Theory: An Analysis of Decisions Under Risk. *Econometrica* 47(2):263–291.

Kanninen, B.J. 2002. Optimal Design for Multinomial Choice Experiments. *Journal of Marketing Research* 39:214–217. May.

Kuhfeld, W.F., R.D. Tobias, and M. Garratt. 1994. Efficient Experimental Design with Marketing Research Applications. *Journal of Marketing Research* 21(4):545–557.

Lam, S.H. and F. Xie. 2002. Transit Path Models That Use RP and SP Data. *Transportation Research Record* 1799:58–65.

Lazari, A.G. and D.A. Anderson. 1994. Designs of Discrete Choice Experiments for Estimating Both Attribute and Availability Cross Effects. *Journal of Marketing Research* 31(3):375–383.

Louviere, J.J. and H.J.P. Timmermans. 1990. Hierarchical Information Integration Applied to Residential Choice Behaviour. *Geographical Analysis* 22:127–145.

Louviere, J.J., D.A. Hensher, and J.F. Swait. 2000. *Stated Choice Methods and Analysis*. Cambridge, England: Cambridge University Press.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behaviour. *Frontiers of Econometrics*. Edited by P. Zarembka. New York, NY: Academic Press.

____. 2001. Economic Choices: Economic Decisions of Individuals, notes prepared for a lecture at the University of California, Berkeley. March 18.

Oppewal, H., J.J. Louviere, and H.J.P. Timmermans. 1994. Modeling Hierarchical Information Integration Processes with Integrated Conjoint Choice Experiments. *Journal of Marketing Research* 31(1):92–105.

Ortúzar, J. de Dios., A. Iacobelli, and C. Valeze. 2000. Estimating Demand for a Cycle-Way Network. *Transportation Research A* 34(5):353–373.

Revelt, D. and K. Train. 2000. Customer-Specific Taste Parameters and Mixed Logit, working paper. Department of Economics, University of California, Berkeley. Available at http://elsa.berkeley.edu/wp/train0999.pdf.

Richardson, A.J. 2002. Simulation Study of Estimation of Individual Specific Values of Time Using an Adaptive Stated Preference Survey, paper presented at the Annual Meetings of the Transportation Research Board, Washington, DC.

Rose, J.M. and M.C.J. Bliemer. 2004. The Design of Stated Choice Experiments: The State of Practice and Future Challenges, working paper. University of Sydney. April.

Rose, J.M. and D.A. Hensher. 2004. Handling Individual Specific Availability of Alternatives in Stated Choice Experiments, paper presented at the 7th International Conference on Travel Survey Methods, Los Sueños, Costa Rica.

Sandor, Z. and M. Wedel. 2001. Designing Conjoint Choice Experiments Using Managers' Prior Beliefs. *Journal of Marketing Research* 38(4):430–444.

Siegel S. and N. Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York, NY: McGraw Hill.

Starmer, C. 2000. Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk. *Journal of Economic Literature* 38:332-382.

Swait, J. and W. Adamowicz. 2001a. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research* 28:135–148.

____. 2001b. Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. *Organizational Behavior and Human Decision Processes* 49:1–27.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge, England: Cambridge University Press.

Toubia, R., J.R. Hauser, and D.I. Simester. 2004. Polyhedral Methods for Adaptive Choice Based Conjoint Analysis. *Journal of Marketing Research* 41(1):116–131.

Wang, D., L. Jiuqun, and H.J.P. Timmermans. 2001. Reducing Respondent Burden, Information Processing and Incomprehensibility in Stated Preference Surveys: Principles and Properties of Paired Conjoint Analysis. *Transportation Research Record* 1768:71–78.

Wittink, D.R. and P. Cattin. 1989. Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing* 53(3):91–96.

A Classification Tree Application to Predict Total Ship Loss

DIMITRIS X. KOKOTOS[1]
YIANNIS G. SMIRLIS[2,*]

[1] Department of Maritime Studies
University of Piraeus
80 Karaoli and Dimitriou Street
18534 Piraeus, Greece

[2] Department of Statistics and Actuarial Science
University of Piraeus
80 Karaoli and Dimitriou Street
18534 Piraeus, Greece

ABSTRACT

Ship accidents frequently result in total ship loss, an outcome with severe economic and human life consequences. Predicting the total loss of a ship when an accident occurs can provide vital information for ship owners, ship managers, classification societies, underwriters, brokers, and national authorities in terms of risk assessment. This paper investigates the use of classification trees to predict this type of loss. It uses a set of predictor variables that correspond to a number of factors identified as the most relevant to the total loss of a ship and sample data generated from a large database of recorded ship accidents worldwide. Through extensive tests of induction algorithms, Exhaustive CHAID was found to be more efficient in classifying the total loss accident cases. The predictive ability of the resulting classification tree structure can be utilized for risk assessment reporting.

INTRODUCTION

The analysis of ship accident cases is of great importance because of the economic costs (Goulielmos and Giziakis 1995; Bureau of Transport and Communications Economics 1994), the environmental impacts (Commission des Communautés Européennes 1992),

Email addresses:
* Corresponding author—jsmirlis@unipi.gr
D. Kokotos—dkokotos@unipi.gr

KEYWORDS: Classification trees, ship accidents, total ship loss.

and the loss of human lives. Causes of accidents include ships running aground; touching the sea bottom; striking wharves, drilling rigs, platforms, or other external substances; colliding with other ships; catching on fire; or suffering an explosion or other serious hull or machinery damage. The worst possible outcome of an accident may be the total loss of the ship. We define total loss of a ship here as a ship that is irretrievably damaged or sunk in a way that it cannot be salvaged (*actual total loss*) or as a ship that is so damaged that its recovery and repair would exceed the ship's insured value (*constructive total loss*) (Hudson 1996). Different factors determine the total loss of a ship. These factors are related to the quality of the construction, restoration or the resistance of the vessel, the violence and the severity of the accident, and the existing weather and sea conditions at the time of the accident.

Previous studies that look at the problem of predicting ship accidents and possible total ship loss use various datasets and data analysis techniques: discriminant analysis, logistic regression, stochastic models, and neural networks (Psaraftis et al. 1998). Giziakis et al. (1996) used logistic regression on accident data from the Greek Ministry of Mercantile Marine to predict ship failures based on several factors such as the age and the type of the ship, its gross tonnage, registration, etc. Le Blanc and Rucks (1996) proposed discriminant analysis to model ship accident classification in the Missisipi River region. Otay and Özkan (2003) proposed a stochastic prediction model to study the possibility of vessel accidents (collision, ramming, and grounding) in the Strait of Istanbul. Hashemi et al. (1995) developed a neural network structure to predict ship accidents under different conditions on the lower Mississippi River. Le Blanc et al. (2001) compared statistical analysis and neural network computing techniques (Kohonen networks) in a dataset of 900 ship accident cases on the same regions of Mississippi River. This comparison concluded that neural networks are significantly superior for classifying and predicting ship accidents over earlier statistical methods.

The above-mentioned research work and applications examine a relatively small number of accident cases restricted to a particular geographic area or a controlled region (rivers, ports, straits, canals, etc.). Traditional statistical methods and neural networks are the basic data analysis tools used to date for the development of applications. In this paper, we present a classification tree application for predicting total ship loss based on a dataset extracted from an existing large dataset of ship accident cases worldwide. We tested different algorithms for expanding classification trees and a number of values for initial parameters to conclude that Exhaustive CHAID[1] is the most effective algorithm that provides the best classification rates for accidents in which a ship is a total loss. This particular approach of using classification trees has not been investigated in depth in previous research efforts and applications for modeling ship accidents.

In the next section, we present a short overview of classification tree theory and the comparison tests of four classification tree expansion algorithms. The following section presents the predictive variables of the model, describes the data preparation procedures, and provides basic descriptive statistics. We then cover the application of classification tree theory to our test dataset. The last section presents conclusions and discussion of issues concerning potential applications based on the proposed classification tree.

CLASSIFICATION TREES FOR PREDICTING TOTAL SHIP LOSS

The classification tree is a data mining technique for predicting the membership of cases in classes defined by a dependent variable usually of the categorical type. Each case is measured along a number of predictor variables. The implementation of a classification tree is achieved through a training process (*induction*) in which a specific algorithm is applied to a sample dataset (*a training set*) composed of the predictor variables.

A typical induction algorithm works in two phases: the splitting phase and the pruning phase. The splitting phase is an iterative top-down process that expands the tree by defining *nodes* connected by *branches*. The nodes at the end of branches are

[1] CHAID stands for Chi-Squared Automatic Interaction Detection.

called *leaves*. The first node at the top of the tree is the *root node*. At every node, the splitting algorithm creates new nodes by selecting a predictor variable so that the resulting nodes are as far as possible from each other. The distance measurement used for the splitting depends primarily on the specific splitting algorithm and is determined by such statistics as gini, entropy, chi-squared, gain ratio, etc. One important feature of the splitting algorithm is the so-called *greedy*. This refers to the ability of the algorithm to look forward in the tree in order to examine if another combination of splitting could produce better overall classification results.

An alternative representation of the classification tree can be given by using a set of nested IF-THEN rules. Each IF-THEN rule identifies a unique path from the root to a leaf and describes a certain class of cases. This alternative representation of the tree is better for analysis, particularly when the tree is greatly expanded. The nodes at the lowest part of a branch that cannot be split further into other nodes because they contain cases with only one outcome are called *pure leaves*. The splitting phase terminates when a stopping rule, initially selected by the user, is satisfied. Stopping rules may include the maximum number of nodes, the number of variables in a node considered for splitting, a minimum number of cases per node, and so forth. Once the structure of the tree is developed, pruning may be required to make the tree more applicable to other similar datasets or to exclude nodes that seem inappropriate for the specific dataset or application.

The prediction accuracy of the classification tree is highly related to the misclassification *cost* (Fawcett 2001). The term cost is used to describe the situation when some predictions either occur more frequently than others or have more important consequences. Misclassification cost represents the percentage of cases that are incorrectly classified and it is frequently used as a typical measurement of the accuracy of the prediction. For a given class, misclassification cost is set to a specific value to denote the severity of a wrong prediction for that class. Another issue related to the cost is the *priors* or a priori probabilities that denote how likely a case will fall into one of the classes. Unequal priors are used in problems with specific knowledge about the size of the classes. Arrangements for defining misclassification costs and priors are confounded in complex problems (Ripley 1994).

To ensure that the tree will perform as well as in the training sample, a validation procedure can be applied. The most preferred type of validation is testing with a sample taken from the original dataset, especially when this dataset is large enough. The sample size can be approximately one-third to one-half of the learning dataset (Brieman et al. 1984). When no sample dataset is available, the validation can be done on subsets of the original training set. In all cases, the misclassification costs in the validation procedure must be close enough to those obtained by the learning procedure. This procedure verifies that the tree will perform equally well with other datasets. In the case when the misclassification costs are not close enough to the costs of the learning sample, the size and the splitting of the tree must be reconsidered.

A number of induction algorithms and software tools to implement classification trees appear in the literature. The various algorithms differ mainly in the statistical criteria used for splitting the nodes, in the types of dependent variables they support (scale, ordinal, nominal), in the number of nodes they allow for splitting, and in the elimination of redundancy during the generation of the rules. Among others, Classification and Regression Tree (known as CART or C&RT) (Brieman et al. 1984; Lee et al. 1997), CHAID (Kass 1980) and its extension the Exhaustive CHAID (Biggs et al. 1991), and QUEST (quick unbiased efficient statistical tree) (Loh and Shih 1997) are the most recently developed and more popular induction algorithms. A short description of these algorithms follows:

- **CART** generates only binary trees. It constructs the tree by examining all possible splits at each node for each predictor variable and uses the goodness-of-fit measurement criterion to find the best split. It assumes scale and ordinal or nominal types in the predictor and dependent variables.
- **CHAID** determines the best split at each node by merging pairs of categories of the predictor variable with respect to their distance from the dependent variable. The chi-square test measures this distance. It produces nonbinary trees and

assumes scale and ordinal or nominal types in the predictor variables.

- **Exhaustive CHAID** is an improvement over CHAID as it finds the optimal split by continuously testing all possible category subsets in order to merge related pairs until only one single pair remains.
- **QUEST** constructs the tree by examining the association of each predictor variable to the dependent variable and selecting the predictor with the highest association for splitting. Then Quadratic Discriminant Analysis (QDA) is used to find the best split point for the predictor variable selected. The association of a predictor to the dependent variable is measured by ANOVA F-test, Levene's test, or Pearson's chi-square test if the predictor is of the ordinal, continuous, and nominal type, respectively. QUEST like CART, yields binary trees.

QUEST is generally faster than the other techniques, but cannot be applied to regression type problems, that is, when the dependent variable is continuous. CHAID produces, at each split, a greater number of nodes than the other two algorithms, thus forming wider trees. To date, the literature does not give a recommendation for which algorithm to use to maximize the predictive accuracy of the tree. The practice usually followed is to test the different algorithms in order to find which one minimizes the misclassification costs and at the same time satisfies the restrictions of the dataset, such as the existence of missing values and the handling of ordinal or nominal variables (Witten and Frank 2000). The approach we take in this study is to identify the algorithm that will minimize the total loss accident classification rates.

We also directly compare classification trees to other traditional statistical methods such as logistic regression (Dillon and Goldstein 1984), because they can classify cases depending on classes defined by a dependent variable. Logistic regression is similar to other statistical explanatory and classification techniques such as linear regression, ordinal least squares and discriminant analysis, but it has less stringent requirements because it assumes no linearity of relationships between the dependent and the independent variables and does not assume normally distributed variables. As in the classification trees, the effectiveness of the statistical method is measured by the misclassification rate, that is, the percentage of cases that are not correctly classified to the total number of cases.

PREDICTIVE VARIABLES AND DATASET

In order to build an explanatory model to predict total ship loss, a preparatory phase of this study identified a number of factors that were conceptually grouped with those directly related to the vessel and with those that describe the particular conditions at the time of the accident. We initially identified these factors using accident reports (*Lloyd's Casualty Week* 1992–1999) and subsequently verified them from other references (Psaraftis et al. 1998; Giziakis 1996). The factors chosen include the type, size, age, and condition of the vessel at the time of the accident; its previous record of accidents; the weather and sea conditions; and the place and location of the ship when the accident occurred.

This study is based on an existing database of accident cases that was created for other projects (Giziakis and Kokotos 1996; Kokotos 2003). This database contains 27,664 records of shipping accidents worldwide between 1992 and 1999. The data were compiled mainly from textual ship accident references taken from *Lloyd's Casualty Week* (1992–1999) and validated against annual editions of *Lloyd's Register of Ships* (1992–1999), annual editions of *Lloyd's World Casualty Statistics* (1992–1999), and *Lloyd's Maritime Atlas* (1993). This reference database was further organized into predictive variables properly chosen to relate closely to the factors previously identified as the most relevant for explaining total ship loss, and it was prepared to conform to Lloyd's Casualty Information System (1980).

Through a data cleaning effort, a dataset of 4,619 shipping accident cases was generated by eliminating cases with identical accident information, missing values, and unreliable data. The final dataset used in this study contained only 352 accident cases (7.6%) where total ship loss was reported, while the rest of the cases (4,267 or 92.4% of the total) were related to accidents with no total ship loss. In this dataset, a small number of

ships were involved in more than one accident with one resulting in total loss.

The remainder of this section presents the predictive variables and the most important descriptive statistics for a better understanding of the problem.

Factors Related to the Ship

The **year when the ship was constructed** was used as an indicator of the general condition of the vessel. Most of the ships in the sample were built between 1967 and 1990; only about 20% were built before 1967 or after 1991. For the class of accidents with total ship loss, the average value for this variable was 1974.2; while for the class of no total ship loss, the corresponding value was 1977.8.

The **age of the ship** was calculated as the difference between the year of the accident and the year when the ship was built. Ships with ages between 15 and 20 years were more frequently involved in accidents.

The **gross tonnage of the ship** was used as a typical measure for the size of the ship, which strongly depends on the type of the ship (see below). The distribution of the values for the gross tonnage of the ships in the dataset was 13.6% below 1,000 tons; 67.4% between 1,000 and 24,500 tons; and 20% over 24,500 tons. A simple comparison of the average values of gross tonnage in the classes of accidents with or without total ship loss (12,084 and 18,234 tons, respectively) indicated that smaller ships were more frequently lost than bigger ships.

The **types of the ships** recorded in the dataset were tanker, general cargo, ferry, container, and bulk carrier. Containers appear to have the lowest accident rate where the ship is a total loss (about 3%), while for all the other types this rate was not significantly different from the average (between about 7% and 9% of the total number of ships).

The **number of previous ship owners** reflects the general condition of the vessel, because the practice followed by many ship owners is to sell the ship if its condition is declining and the involvement in serious accidents is expected to be more frequent. For simplicity, this variable was categorized into four groups corresponding to one, two, three, or four to five or more owners. The percentages of total accident cases included in these groups are 35.9%, 23.8%, 29.4%, and 10.9%, respectively.

The **number of previous ship accidents,** regardless of their type and their severity, served as an extra indication of the general condition of the ship. The maximum value in the dataset was seven. Total ship loss was not significantly correlated with this variable. Particularly for the class of accidents with total loss, 83% of these cases had only one previous accident and the rest (17%) had more than one.

The **registration society of the ship** at the time of the accident[2] was another variable. Registration societies certificate the condition of ships by adopting different survey standards. Sixteen distinct registration societies were recorded in the database and coded from R1 to R16 in random order. The average percentage of the accidents with total ship loss per society was 7.6% and the minimum and maximum were 0.5% and 24%, respectively.

Factors Related to the Accident

The **type of accident** variable describes what occurred, independently from the outcome of the accident (total ship loss or no total ship loss). The accident type was coded according to Lloyd's Casualty Information System and accident classification standards. Table 1 shows the percentage of accidents with total ship loss by accident type. It shows that fire/explosion and contacts with external substances are the accident types with the highest and lowest frequency of total ship loss.

The accident type is also very closely related to the type of the ship (figure 1). This dimensionless graph very closely plots the particular categories of the two variables that have strong relationships. The figure shows that tankers have frequent collisions because of their size and their lack of flexibility in maneuvering, while containers due to their cargo suffer from fires. Grounding accidents are more frequent for general cargo ship, and contacts are independent of the type of the ship.

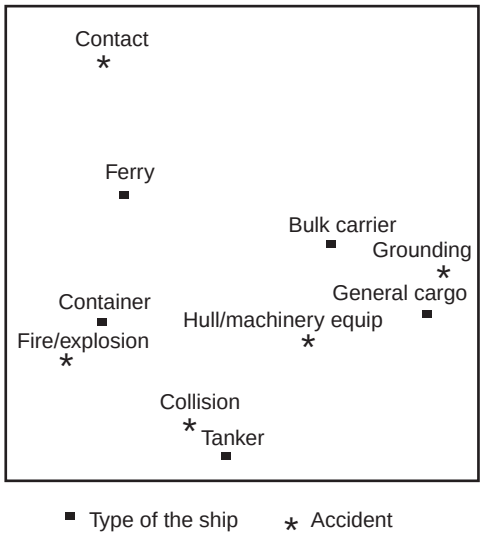
The variable for the **year and the month of the accident** covered 1992 to 1999. No significant differ-

[2] See *Lloyd's Register of Shipping* (www.mariners-l.co.uk/ResLloydsRegister.htm) for additional information on classification societies and their published registers. (Also see Haviland 1970.)

TABLE 1 Percentage of Accident Cases with Total Ship Loss for Different Types of Accidents

Accident type	Description	Percentage of accident cases with total ship loss
Grounding	Touching of the sea bottom or underwater wrecks for a significant period of time	9.50%
Contact	Striking an external surface substance (not another ship) such as drilling rigs or platforms, etc.	2.60%
Collision	Striking another ship, regardless of whether underway, anchored, or moored	7.40%
Fire/explosion	Fire or explosion, regardless of the cause	14.20%
Hull/machinery damage	Any case of hull/machinery damage or failure	6.40%

FIGURE 1 Relationship Between the Type of Accident and the Type of Ship



ences were found between the number of accidents in different years and months. The average number of accidents with total loss per month was 7.6%.

The particular **geographic area of the accident** coded into 12 major areas according to the standard classification of *Lloyds Maritime Atlas* areas. Figure 2 shows the 12 areas defined and table 2 presents the distribution of the number of accidents with total ship loss in the 12 areas. The greatest number of accidents with total ship loss occurred in the Indian Ocean and the fewest in the Gulf of Mexico-West Indies-Newfoundland.

For the specific **location of the ship** at the time of the accident, values are "port" for accidents that occurred within the region of a port, "overseas" for accidents that occurred at sea far from the coast, and "controlled seaways" for straits and canals. The number of accident cases and the associated percentages are 2,188 (47.4%) for ports, 1,548 (33.5%) for overseas, and 883 (19.1%) for controlled seaways. The number of accidents with total ship loss was equally distributed for all types of locations: ports, 6%; overseas, 9.8%; and controlled seaways, 7.8%. The most frequent accidents in ports were grounding and collisions and, in the overseas category, hull/machinery damage.

The variable for the **reported weather conditions** when the accident happened included: calm weather, poor visibility, storm, freezing conditions, and typhoon. Most of the accidents with total ship loss occurred during typhoons (8.8%), storms (8.3%), or in poor visibility (7.3%). In calm weather or freezing conditions, as expected, the percentage of the accidents with total ship loss was significantly lower (5.0% and 4.3%, respectively). In relation to the accident type, 43.8%, 49.2%, and 56.9% of the hull/machinery accidents occurred in calm weather, during storms, or in freezing conditions, respectively. During typhoons, the most frequent accident was grounding (50.1%), and in poor visibility, collisions (47.2%) were the most frequent accidents.

FIGURE 2 Classification of the Geographic Areas

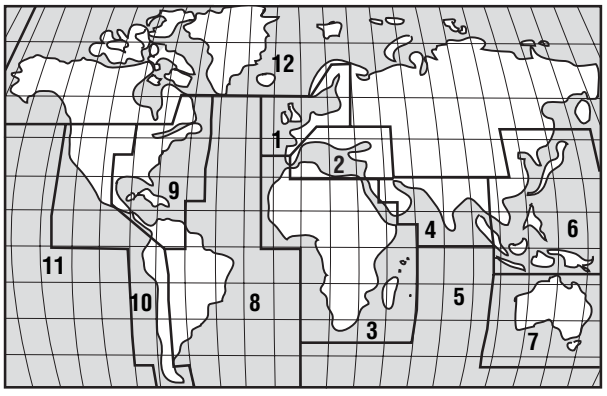


TABLE 2 Percentage of Accident Cases with Total Ship Loss in Different Geographic Locations

Code	Geographic area	Accident cases with total ship loss (%)
1	North Sea, Baltic	2.8
2	Mediterranean-Black Sea	4.7
3	Red Sea-West and East Coast of Africa	6.0
4	Persian Gulf-Bay of Bengal	2.6
5	Indian Ocean	10.7
6	China Sea-Japan	8.6
7	Australia	7.2
8	Atlantic Ocean	2.8
9	Gulf of Mexico-West Indies-Newfoundland	1.3
10	North and South America, Pacific Coast	4.0
11	Pacific Ocean	4.2
12	Alaska-Bering-USSR Arctic-Iceland	7.0

Total loss of a ship is the dependent variable for the analysis and is defined as a dichotomous variable accepting values of yes or no. Statistical tests performed (one-way ANOVA) to compare the average values of the above-mentioned variables for the classes of accidents distinguished by total loss showed no significant differentiation among them.

THE APPLICATION

In this section, we test different classification tree induction algorithms and logistic regression in order to identify the best-performing tree structure to predict total ship loss. We used the 12 variables described earlier as predictors with total loss as the dependent variable. Total loss is a dichotomous variable (it accepts values of yes or no), while the predictors are of various types: scale (e.g., gross tonnage, year ship was built), ordinal (e.g., number of previous ship owners), and nominal (e.g., location of the accident, weather conditions).

In a preliminary stage of the analysis, CART, CHAID, Exhaustive CHAID, QUEST, and logistic regression were applied to the dataset by defining equal misclassification costs and priors, assuming no previous knowledge of the problem. This effort, although it produced total misclassification rates of 93% due to the unbalanced training set (only 7.6% of the cases consisted of the class with total ship loss) resulted in very poor classification rates for the class of Total Loss = "yes." For that particular class, Exhaustive CHAID showed the best performance (a classification rate of 55.3%), the logistic regression showed the worst (only 9.99%), and the other two algorithms showed approximately 22%.

To resolve the problem of poor classification in the small class of accident cases with total ship loss, we experientially adjusted the misclassification costs in a second stage of the analysis. Logistic regression was excluded from this stage of the analysis because it cannot accept any further improvement. The misclassification cost for Total Loss = "yes" was set to a ratio of 12 to 1 so as to indirectly reflect the importance and severity of the total loss outcome compared with the damages and the consequences of a simple accident, a practice proposed in similar cost-sensitive classifications problems (Hollmen and Skubacz 2000).

Table 3 presents the rates of the correctly classified cases obtained from the four induction algorithms. From the table, it can be seen that Exhaustive CHAID retains its superiority over the rest of the algorithms and achieves the best rates in both cases.

In all the tests carried out in this analysis, the classification tree algorithms were applied in a sample training set of 3,079 cases (two-thirds of the total number of cases). The remaining 1,026 cases were considered as the test dataset used for validation. To ensure a uniform distribution of cases in every split, the child nodes were defined to include a

TABLE 3 Classification Rates After Adjustment of Misclassification Cost for Total Loss = "yes"

Induction algorithms	Total rate of correctly classified cases (%)	Rate of correctly classified cases of Total Loss = "yes"(%)
CART	78.3	59.9
CHAID	80.1	86.4
Exhaustive CHAID	90.7	87.5
QUEST	90.7	31.0

number of cases not greater than the half of the parent node. We used SPSS/Answer Tree software to implement the classification algorithms.

To confirm that the results of the Exhaustive CHAID were not dependent on the particular dataset and that this algorithm will perform well using other similar datasets, a validation procedure including three different tests was applied. First, Exhaustive CHAID was tested using the dataset of 1,026 cases (one-third of the initial dataset not included in the training set) and produced classification rates of 87.8% and 84.3% for the total number of test cases and the cases of Total Loss = "yes," respectively. A second used 10 subsets randomly selected from the initial dataset. This test gave the best classification rates—84.1% and 80.5%, respectively. A third test was a manual test of the classification tree structure for a small number of new accident cases not included in the initial dataset. Again, the classification rates were approximately similar to the outcome of the other two testing methods. This validation procedure verified that the tree structure produced by Exhaustive CHAID provides the best predictions of total ship loss.

Figure 3 presents the final classification tree structure produced using Exhaustive CHAID after the adjustments in misclassification costs during the training phase. For economy of presentation, only the first three levels of the tree are shown. Each node is identified by the node number and the number of cases included in the node for the classes of Total Loss = "no" and "yes," followed by the percentages and the totals. Table 4 gives, for every node presented in the tree, the condition applied for the expansion of the tree.

The hierarchy of the classification tree shows that the first split defined four nodes using the "year ship was built" predictor. Depending on the node of the first level, in the next splits the predictors "geographic area," "location," and "gross tonnage" were used to define nodes 5 up to 13. In the next levels, other predictors were included except for "weather condition" and "number of previous accidents," which had not been used anywhere in the tree structure.

The graphical representation of a classification tree, as in figure 3, may not be very convenient for analysts or decisionmakers, particularly when the tree is wide and contains a large number of nodes. An alternative, more suitable presentation of the tree can be given by describing each node by IF-THEN rules of the form:

"IF *condition* THEN *prediction*"

in which the *condition* part is a composite condition including the AND logical operator, and the *prediction* part is given in terms of a probability value for the *condition* to be true.

By using the alternative IF-THEN presentation of the classification tree produced in this study, different types of nodes can be located: those that contain cases in which total loss has a significant probability of occurring, those in which total loss of a ship is unlikely to occur, and those that present no clear conclusion. The most important nodes for this study are those that have significant probabilities of total loss. These types of nodes, although limited, reveal certain conditions of accidents in which total loss of the ship is a strong possible outcome. The following examples of selected nodes demonstrate the three types of nodes of the tree. The symbol "!=" which appears in the IF conditions is the "not equal" operator.

Example 1 describes a typical node for which the group of accident cases has very limited probability of total loss. Example 2 uncovers a group of accident cases with significant risk situations (39.3%). This is considered a valuable output of the analysis. Example 3 refers to a node where no clear distinc-

FIGURE 3 Classification Tree Produced Using Exhaustive CHAID after the Adjustments in Misclassification Costs

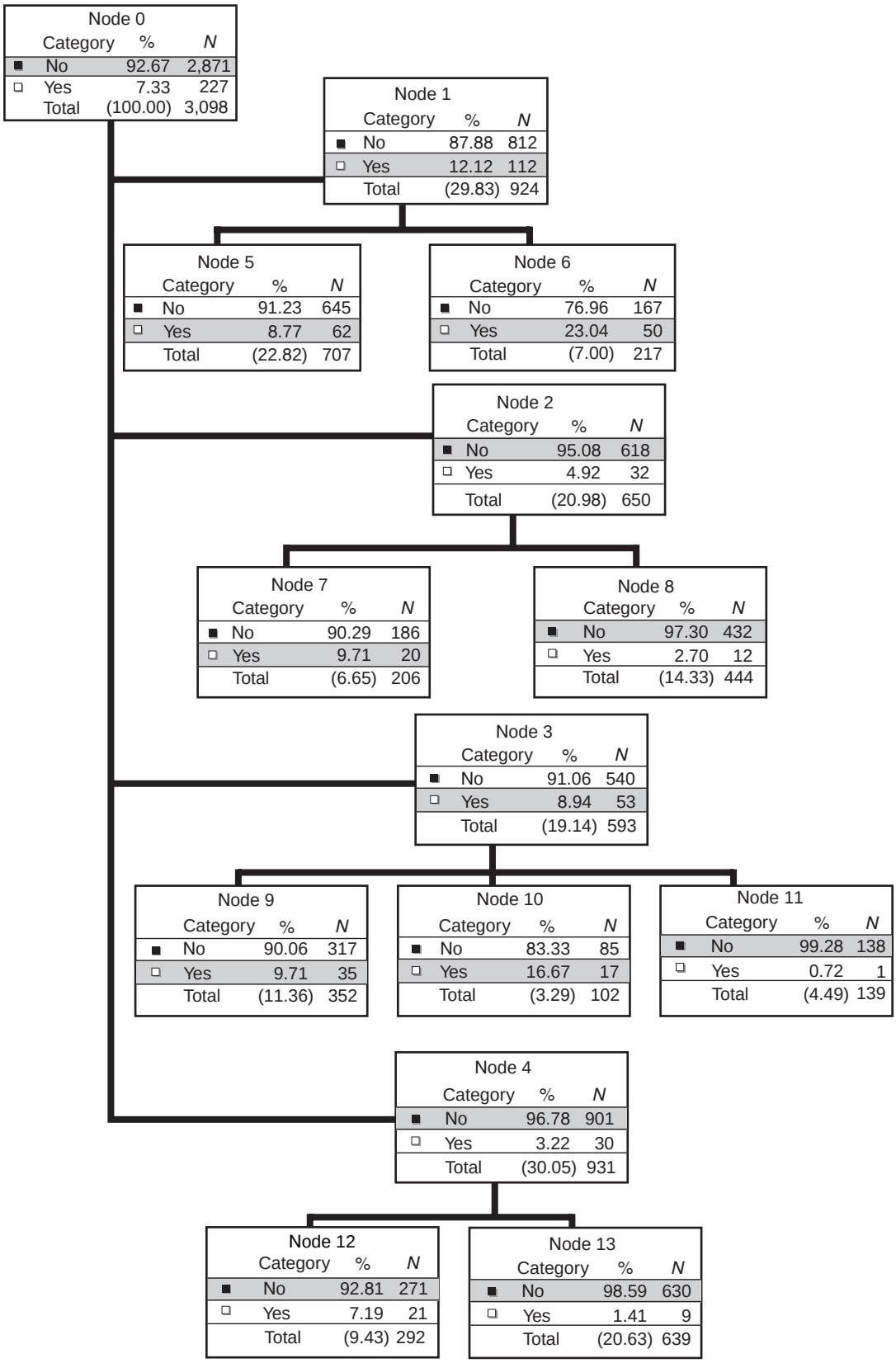


TABLE 4 Conditions per Node for Expanding the Classification Tree

Node	Condition
1	Year ship was built ≤1973
2	Year ship was built in (1973–1977]
3	Year ship was built in (1977–1981]
4	Year ship was built > 1981
5	Geographic area = {Mediterranean-Black Sea; China Sea-Japan; Persian Gulf-Bay of Bengal; North Sea, Baltic; N. & S. America, Pacific Coast; Alaska-Bering-USSR Arctic-Iceland; Atlantic Ocean; Gulf of Mexico-W. Indies-Newfoundland}
6	Geographic area ={Pacific Ocean; Atlantic Ocean; Red Sea-W. and E. African Coast}
7	Location = {overseas}
8	Location = {controlled seaways, ports}
9	Gross tonnage ≤12603
10	Gross tonnage (12,603–16,705]
11	Gross tonnage > 16,705
12	Geographic area = {Mediterranean-Black Sea; Australia; Alaska-Bering-USSR Arctic-Iceland}
13	Geographic area = {China Sea-Japan; Persian Gulf-Bay of Bengal; North Sea, Baltic; N. & S. America, Pacific Coast; Atlantic Ocean; Gulf of Mexico-W. Indies-Newfoundland; Pacific Ocean; Atlantic Ocean; Red Sea-W. and E. African Coast}

tion between the cases with total loss and no total loss can be seen, because the probabilities do not significantly differ from those obtained from the whole dataset. The condition associated with this node is very simple and not very specific.

Example 1

Rule : IF *(Year ship was built > 1981) AND (Geographic Area ! = "Mediterranean-Black Sea" AND Geographic Area ! = "Australia" AND Geographic Area ! = "Alaska-Bering-USSR Arctic-Iceland") AND (Accident ! = "Grounding" AND Accident ! = "Fire/Explosion") AND (Number of previous ship owners > 2)* THEN *Prediction = NO, Probability = 0.9783.*

Number of cases. Total number of cases = 94. Cases of total loss YES = 2 (2.17%), NO = 92 (97.83%). Probability for NO total loss = 0.9783.

Description. Any ship built after 1981 with two or more previous owners, involved in accident types "contact," "collision," and "hull/machinery damage" different from "grounding" and "fire/explosion," in areas other than "Mediterranean-Black Sea," "Australia," and "Alaska-Bering-USSR Arctic-Iceland."

Example 2

Rule. IF *(Year ship was built < = 1973) AND (Geographic Area ! = "Pacific Ocean" AND Geographic Area ! = "Australia" AND Geographic Area ! = "Atlantic Ocean" AND Geographic Area ! = "Red Sea – W & E African Coast ") AND (Accident ! = "Contact" AND Accident ! = "Hull/Machinery Damage") AND (Registration society ! = "R11" AND Registration society ! = "R2" AND Registration society ! = "R16" AND Registration society ! = "R12")* THEN *Prediction = YES, Probability = 0.393.*

Number of cases. Total number of cases = 79. Cases of total loss YES = 31 (39.3%), NO = 48 (60.7%). Probability for total loss YES = 0.393.

Description. Any ship built before 1973 registered in societies R1, R3 to R10, and R12 to R15 involved in accident types "collision," "fire/explosion," and "grounding" in areas other than "Australia," "Atlantic Ocean," and "Red Sea-West and East African Coast."

Example 3

Rule: IF *(Year ship was built > 1977 AND Year ship was built < = 1981) AND (Gross ship tonnage < = 12603)* THEN *Prediction = YES, Probability = 0.0994.*

Number of cases. Total number of cases = 352. Cases of total loss YES = 35 (9.95%), NO = 317 (90.05%). Probability for total loss YES = 0.0995.

Description. Any type of ship built between 1977 and 1981 having gross tonnage less than 12,603 tons.

The above-mentioned examples demonstrate the use of classification to identify groups of accident cases with significant or no significant possibilities of total ship loss.

CONCLUSION

In this paper, we presented a classification tree application to predict total loss of a ship as a consequence of an accident. The application was based on a large dataset of accident cases occurring in locations worldwide. Extensive tests indicated that

the Exhaustive CHAID induction algorithm minimized misclassification costs, a criterion that we defined as the most important for the particular application. Due to the unbalanced training set, the initial choice of setting equal costs resulted in poor classification rates for the class of Total Loss = "yes." To resolve this problem, initial information concerning misclassification rates was defined to reflect the importance of the total ship loss outcome in this particular application. The experiential comparison between different induction algorithms and logistic regression verified the superiority in classification of data mining techniques and especially of the classification trees over traditional statistical methods. Classification trees compared with statistical methods are also easily understood by both experts and non-experts and can provide a good illustration of the classification.

The prediction of total loss is of great importance for ship owners, ship managers, classification societies, underwriters, brokers, and national authorities, because it can provide valuable information for issuing risk assessment reports. In the case of a ship accident, by considering parameters such as the characteristics of the vessel, the geographic area and the particular location, the type of the accident, etc., through traversing of the tree or testing the IF-THEN rules, estimates can be made for the accident of the probability of total ship loss. In many accident cases, the prediction should be accurate and clear and can be used to activate different rescue plans so as to reduce the costs of damages to the vessel and, particularly, to save lives. The classification tree for predicting total ship loss may be utilized in the context of a potential decision support system and a risk management information system that will record, evaluate, and process data for ship accidents.

REFERENCES

Biggs, D., B. DeVille, and E. Suen. 1991. A Method of Choosing Multiway Partitions for Classification and Decision Trees. *Journal of Applied Statistics* 18(1):49–62.

Brieman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Bureau of Transport and Communications Economics. 1994. *Structural Failure of Large Bulk Ships*, Report 85. Commonwealth of Australia.

Commission des Communautés Européennes. 1992. L'Impact des Transports sur L'Environnement: Une Stratégique Communautaire pour un Development des Transports Respectueux de L'Environnement. *Livre Vert Relatif* COM(92) 46 final/29/4/92. Brussels, Belgium.

Dillox, W.R. and M. Goldstein. 1984. *Multivariate Analysis Methods and Applications*. New York, NY: John Wiley.

Fawcett, T. 2001. Using Rule Sets to Maximize ROC Performance, Proceedings of the 2001 IEEE International Conference on Data Mining, 29 November–2 December 2001, San Jose, CA, pp. 131–138.

Giziakis, K. 1996. *Criticism of the Content of Variables that are Used in the Analysis of Accidents in the Shipping Industry,* volume in honor of Professor Stavropoylos. Piraeus, Greece: University of Piraeus.

Giziakis, K., E. Giziaki, A. Pardali-Lainou, V. Michalopoulos, and D. Kokotos. 1996. Minimising the Risk of Failure for an Effective and Reliable European Shipping Network. *Proceedings of the 3rd European Research Roundtable Conference on Shortsea Shipping: Building European ShortSea Networks*. Bergen, Norway: Delft University Press.

Giziakis, K. and D.X. Kokotos. 1996. Needs and Benefits from the Development of Shipping Accident Databases, Proceedings of the Conference on Greek Coasts and Seas, Feb. 28–29, 1996, Piraeus, Greece.

Goulielmos, A.M. and K. Giziakis. 1995. Treatment of Uncompensated Cost of Marine Accidents in a Model of Welfare Economics, Proceedings of JAME Conference, Boston, MA.

Hashemi, R.R., L.A. Le Blanc, C.T. Rucks, and A. Shearry. 1995. A Neural Network for Transportation Safety Modeling. *Expert Systems with Applications* 9(3):247–256.

Haviland, E.K. 1970. Classification Society Registers from the Point of View of a Marine Historian. *American Neptune* 30:9–39.

Hollmen, J. and M. Skubacz. 2000. Input Dependent Misclassification Costs for Cost-Sensitive Classifiers. *Proceedings of the 2nd International Conference on Data Mining.* Edited by M. Taniguchi. Bellerica, MA: WIT Press.

Hudson, N.G. 1996. *Marine Claims Handbook, 5th Edition.* London, England: Witherby's Publishing.

Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* 29:119–131.

Kokotos, D. 2003. Data Mining: Decision Tree Analysis Upon Vessel Accidents, Ph.D. thesis. Department of Maritime Studies, University of Piraeus, Piraeus, Greece.

Le Blanc, L.A. and C.T. Rucks. 1996. A Multiple Discriminant Analysis of Vessel Accidents. *Accident Analysis and Prevention* 28(4):501–510.

Le Blanc, L.A., R.R. Hashemi, and C.T. Rucks. 2001. Pattern Development for Vessel Accidents: A Comparison of Statisti-

cal and Neural Computing Techniques. *Expert Systems with Applications* 20(3):163–171.

Lee, Y., B.V. Roy, C.D. Reed, R.P. Lippman, and K. Wadsworth. 1997. *Solving Data Mining Problems Through Pattern Recognition*. Upper Saddle River, NJ: Prentice Hall.

Lloyd's Casualty Information System. 1993. London, England: Lloyd's of London Press, Ltd.

Lloyd's Casualty Week. 1992–1999. London, England: Lloyds of London Press, Ltd.

Lloyd's Maritime Atlas, 10th ed. London, England: Lloyd's London Press, Ltd.

Lloyd's Register of Ships. 1992–1999. London, England: Lloyd's Register of Shipping.

Lloyd's World Casualty Statistics. 1992–1999. London, England: Lloyd's Register of Shipping.

Loh, W.Y. and Y.S. Shih. 1997. Split Selection Methods for Classification Trees. *Statistica Sinica* 7:815–840.

Otay, N. and S. Özkan. 2003. Stochastic Prediction of Maritime Accidents in the Strait of Istanbul, Proceedings of the 3rd International Conference on Oil Spills in the Mediterranean and Black Sea Regions, Istanbul, Turkey, September, pp. 92–104.

Psaraftis, C., G. Panagakos, N. Desipris, and N. Ventikos. 1998. An Analysis of Maritime Transportation Risk Factors, Proceedings of the 8th Conference ISOPE, Montreal, Canada, Vol. IV, pp. 484–492.

Ripley, B.D. 1994. Neural Networks and Related Methods for Classification (with discussion). *Journal of the Royal Statistical Society B* 56:409–456.

Witten, I.H. and E. Frank. 2000. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* San Francisco, CA: Morgan Kaufmann Publishers.

U.S. Transportation Models Forecasting Greenhouse Gas Emissions: An Evaluation from a User's Perspective

DAVID CHIEN

Office of Advanced Studies
Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 Seventh St., SW, Room 3430
Washington, DC 20590
Email: David.Chien@dot.gov

ABSTRACT

This paper briefly describes and evaluates some of the more important and frequently used models to estimate greenhouse gas emissions by a number of U.S. government agencies. Among the models covered are: National Energy Modeling System (NEMS); MARKAL-MACRO; MiniCAM; Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation (GREET) Model; and Transitional Alternative Fuels and Vehicles (TAFV) Model. These models have been used by the U.S. Congress and federal agencies to assess U.S. strategies to meet the Kyoto Accord, which would require the United States to maintain U.S. carbon emissions at 7% below 1990 levels between 2008 and 2012. In this paper, each model is described and its capabilities and limitations highlighted. Model perspectives are provided from a user's viewpoint, so that potential users will have a full understanding of the capabilities of these models and the resources needed to build, update, and maintain them.

INTRODUCTION

In December 1997, approximately 160 nations met in Kyoto, Japan, and developed the Kyoto Protocol, which would limit greenhouse gas (GHG) emissions

KEYWORDS: Greenhouse gases, transportation forecasting, modeling.

worldwide. In the protocol, the United States agreed to reduce GHG emissions levels to 7% below 1990 levels from 2008 to 2012 (although the U.S. government has not formally ratified the agreement). In considering the impacts of the Kyoto Protocol and other GHG emissions reduction policies or programs, the U.S. Department of Transportation (DOT), the U.S. Department of Energy (DOE), and the U.S. Environmental Protection Agency (EPA) have employed a number of useful models. These models range from pure energy and environmental analytical tools to integrated energy-environmental-economic models that capture the interactions of GHG reduction technologies and policies within the economy. Because such modeling efforts have significant impacts on policy and program decisionmaking, a critical review of those models is important.

This paper describes a number of models used by the U.S. government to better understand the impact of future technologies and policies on U.S. GHG emissions in the transportation sector.[1] In particular, the paper focuses on five models:

1. National Energy Modeling System (NEMS), maintained by the Energy Information Administration (EIA) within DOE;
2. Energy MARKAL-MACRO, maintained by Brookhaven National Laboratory and DOE;
3. MiniCAM, maintained by Pacific Northwest National Laboratory;
4. Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation (GREET), maintained by Argonne National Laboratory; and
5. Transitional Alternative Fuels and Vehicles (TAFV), maintained by Oak Ridge National Laboratory (ORNL) and the University of Maine.

This paper does not thoroughly evaluate detailed inputs to the models (i.e., data and assumptions). Therefore, the author highly recommends that interested readers refer to model documentation and visit the model websites for more detailed information on inputs and assumptions. Sources for further reading for each model are provided at the end of the paper.

Although this article provides potential users with brief descriptions of the GHG models employed by the U.S. federal government, it also evaluates the models based on key operational factors often overlooked by potential users. Topics discussed here include: the size of the data inputs and sourcecode of the models, the hardware and software platform and requirements, the run time or amount of time associated with execution of the models, the resources needed to develop and maintain the models, and examples of studies that have used the models extensively. Detailed model coverage with respect to the transportation sector is also evaluated. It should be noted that the evaluations contained in this article are a snapshot in time and were up to date at the time of writing. However, the reader should keep in mind that the models receive significant improvements over time.

BACKGROUND

Most of the models reviewed in this study originated before GHG modeling became prevalent. In fact, almost none of the models were originally designed to measure greenhouse gases, but rather operational aspects, criteria pollutant (nitrogen oxides—NO_X, sulfur oxides—SO_X, nonmethane hydrocarbons—HC, and particulates) emissions, and energy aspects of modeling and forecasting. Energy forecasting and emissions models are a natural fit for carbon emissions estimation and forecasting. Transportation energy and emissions models usually require two components, travel and fuel efficiency or emissions factors, that are related to the technology and its age. Given these components, fuel consumption can easily be converted into carbon emissions, because the burning of carbon emission compounds is in direct proportion to its consumption.

These models have become very popular and widely used because of their importance in formulating policies designed to reduce carbon emissions. With the ratification of the Kyoto Accord (minus the United States and Australia), many countries had to devise strategies to reduce carbon emissions rapidly. The economic and operational consequences for carbon emissions reduction policies have now emerged

[1] Although this article was written by the author, information presented here relies heavily on a more detailed document prepared by Kevin Greene of the Volpe National Transportation Systems Center (USDOT 2003).

as important factors to investigate when countries attempt to reduce carbon emissions to achieve the long-term Kyoto emissions levels.

DOT uses some of these models and integrates data from many sources. Many of the data sources used to develop these models reside at the federal agencies that created and currently maintain the models. Among those used most frequently for estimates of historical GHG emissions are the EPA[2] and EIA[3] models. Transportation and energy-related data can be found on the *Transportation Energy Databook* website run by ORNL and DOE,[4] DOT's Bureau of Transportation Statistics (BTS) website,[5] and the BTS TRANSTATS website.[6] The TRANSTATS website contains *National Transportation Statistics* data compiled from BTS surveys—Office of Airline Information databases, the National Household Transportation Survey (NHTS), the Commodity Flow Survey (CFS), and many more.

NEMS MODEL

Overview

> NEMS is a computer-based energy-economy modeling system of the U.S. energy markets for the midterm period through 2025. NEMS annually projects the production, imports, conversion, consumption, and prices of energy, subject to assumptions on macroeconomic and financial factors, world energy markets, resource availability and costs, behavioral and technological choice criteria, cost and performance characteristics of energy technologies, and demographics. The purpose of NEMS is to project energy, economic, environmental, and security impacts on the United States of alternative energy policies and of different assumptions about energy markets. (USDOE EIA 2003a)

Congress and other federal agencies have used NEMS[7] extensively to evaluate energy and transportation policies. The model has the advantage of extensive peer review by the U.S. transportation community including DOE, DOT, EPA, the Office of Management and Budget, the Government Accountability Office, and the National Academy of Sciences.

Structure

The structure of NEMS consists of an integrated modeling system representing all demand sectors of the economy (residential, commercial, industrial, and transportation), including a macroeconomic component and all energy supply sources (i.e., crude oil supply; oil refinery; oil distribution; natural gas including exploration, drilling, and distribution; electricity including nuclear, coal, natural gas, residual fuel, and small generators like wind and solar; coal; and renewable fuels).

The transportation sector is important, because it consumes over 27% of all energy, and approximately 98% of transportation consumption comes from petroleum use (USDOE 2002a). The NEMS Transportation Demand Module (TRAN) provides wide coverage of the aggregate transportation system including the following submodules: Light-Duty Vehicles (LDV), Aviation, Freight Transport (truck, rail, waterborne), Miscellaneous (transit, recreational boats, aviation gasoline), and Emissions (USDOE EIA 2003a).

The LDV Submodule covers 6 areas: Fuel Economy—6 car and 6 light-truck EPA size classes across 63 advanced subsystems and fuel savings technologies; Regional Sales—9 Census Divisions; Alternative Fuel Vehicles—12 types of vehicles; LDV Stock—vehicle retirement curves and capital stocks by 20 vintages and vehicle types; Vehicle-Miles Traveled (VMT)—by car and light truck as a function of income per capita and the cost of driving per mile; and LDV Fleet for business, government, and utility fleets (as part of the Energy Policy Act).

The Aviation Submodule includes two components. The Air Travel Demand Submodule forecasts revenue passenger-miles for international and domestic travel, revenue ton-miles for freight, and seat-miles demanded. The Aircraft Fleet Efficiency Submodule covers six fuel-saving technologies for regional, narrow, and wide-body aircraft: ultra-high bypass, propfan, improved thermodynamics, hybrid

[2] Available at http://yosemite.epa.gov/oar/globalwarming.nsf/UniqueKeyLookup/RAMR5CZKVE/$File/ghgbrochure.pdf.
[3] Available at http://www.eia.doe.gov/env/ghg.html.
[4] Available at http://www-cta.ornl.gov/cta/data/Index.html.
[5] Available at http://www.bts.gov/.
[6] Available at http://www.transtats.bts.gov/.
[7] NEMS Transportation Model contact: John Maples, USDOE, EIA, john.maples@eia.doe.gov (202-586-1757); also see http://www.eia.doe.gov/oiaf or http://www.eia.doe.gov/bookshelf/docs.html.

laminar flow, improved aerodynamics, and weight reduction. The submodule also contains 48 vintages of aircraft with aircraft survival curves and stock model representation.

The Freight Transport Submodule includes truck, rail, and waterborne. The Freight Truck Submodule uses macroeconomic gross outputs by North American Industrial Classification System (NAICS) code in determining VMT. The CFS and the Vehicle Inventory and Use Survey are used extensively to establish the connection between commodities and mode of travel. The truck stock model determines capital stocks by three truck size classes (Class 3, Classes 4 through 6, and Classes 7 and 8) and by vehicle age (20 vintages). Technology choice is based on future emissions standards, commercial availability, fuel prices, capital cost, efficiency improvement, and other cost-effectiveness criteria such as discount rates and payback period. There are 32 advanced subsystem and emissions control technologies (Argonne 1999). Gasoline, diesel, natural gas, and liquid petroleum gas are the fuels represented in the Truck Submodule.

Rail and Waterborne Submodules use ton-miles traveled estimated equations based on industrial output by NAICS code. Energy efficiency for old and new vehicles is estimated. A major drawback of the model is the lack of capital stocks and vintaging by age. Therefore, the growth rates of efficiency improvements must be made exogenously based on trends rather than an explicit endogenous calculation of the model. Specific technology representation and turnover cannot be endogenously determined, which limits the effect of advanced technologies over time, unless of course the modeler pre-determines this in the exogenous input file. Overall, the Rail and Waterborne Submodules have no sensitivity to fuel prices or the cost of travel in either the travel or efficiency forecasts.

The Miscellaneous Submodule includes mass transit, which covers six transit modes: three types of passenger rail—transit, commuter, and intercity; and three types of passenger buses—transit, intercity, and school. Travel is estimated for all six transit modes as a function of the relative historical growth rate of passenger-miles of travel relative to light-duty vehicle passenger-miles. Growth rates of efficiency improvements are calculated based on the growth rates of similar technology modes. This assumes that technology advancements will parallel those in modes using the same vehicles. For example, mass transit rail efficiencies would then be assumed to grow at the same rate as Class I freight rail. Therefore, the same caveats from the rail and waterborne models apply to the Mass Transit Module, because both lack explicit model responsiveness to fuel prices and travel costs.

TRAN also has a module that forecasts emissions of the criteria pollutants SO_X, NO_X, HC, carbon (CO), and carbon dioxide (CO_2). Most recently, TRAN incorporated the EPA Mobile 6.0 model, which is used by EPA and several state governments to calculate regional emissions. CO_2 and total CO emissions can be calculated by fuel type and by transportation mode, which allows the user to associate various policies or investments with an increase or decrease in carbon emissions.

Finally, the Macroeconomic Activity Module (MAM) currently consists of the Global Insight (formerly DRI/WEFA) Model of the U.S. economy, the Industry Model, the Employment Model, and the Regional Model. MAM uses the input-output (I-O) National Accounts data (from the Bureau of Economic Analysis of the U.S. Department of Commerce). One issue in using the I-O accounts data is that they undercount the effects of the transportation system on the economy, due to the exclusion of almost all private commercial businesses, which have their own private transportation and are currently counted under commercial operations. A potential improvement to the model would be to adjust the I-O data using the BTS Transportation Satellite Accounts (TSA). The TSA measures the private transportation associated with commercial operations to provide more detailed data for the I-O accounts.

Despite these issues, the MAM is a key element for measuring the impacts of potential GHG strategies on the economy. This makes it one of the most important components of NEMS, because it is essential to the convergence process and it fully integrates the economy with the modeling process, which many of the other GHG models reviewed in this article do not. Reaching equilibrium in a model

of this size is of paramount importance, especially because feedback effects of prices on transportation services have a tendency to be dampened significantly when macroeconomic feedback with the rest of the model is turned on. What does this tell us? The conclusion is that models that do not have this capability can overstate the effects of any given policy that may be implemented, because they do not account for economic changes and responses to those changes. Reaching equilibrium is critical to the accuracy in measuring costs and benefits of any policy or program.

Limitations

The NEMS model operates at a Census region and Census division level. Therefore, extrapolation and interpolation are needed to subdivide the estimates down to the state level. Local- or county-level forecasts are not applicable to the model. TRAN does not explicitly account for modal switching (shifting from one mode to another). Policies designed to shift ridership from one mode to another are currently not measurable nor easy to implement. The travel equations for most modes do use many of the same economic variables, which will result in simultaneous modal switching, but each equation contains a different set of variables that affect travel. Some modes (e.g., rail, waterborne, and all modes of transit) are not technology-based nor do they contain stock models, which make technology policy options limited for those modes of travel.

Resources

One of the drawbacks to using the model is also one of NEMS' greatest strengths. The size of the entire NEMS model is very large and detailed, requiring over 10 to 15 megabytes (MB) of storage just for the "restart file," which contains the starting values for the model each year. In order to do a "standalone" run, which consists of running only one module and keeping the others at reference case levels, would require 100 MB of storage space. Although NEMS can be installed on an individual personal computer (PC), the storage requirements are substantial. Hardware should consist of 512 MB of random access memory (RAM) and a 486 or Pentium processor. The model operates in Compaq Visual FORTRAN and requires the EViews software. If the user wanted to run the supply models also, then OML, a linear programming software, is also necessary.

When running in standalone mode with only one module endogenously active, the transportation module will return a solution within a minute. However, submitting a fully integrated run with all of the modules turned on or active would take about two to four hours depending on how many changes were made to the model. The current NEMS model at EIA employs approximately 40 full-time employees and 4 full-time contractors. Therefore, enhancing, updating, and maintaining the model requires significant resources. However, several agencies and national laboratories work with versions of the NEMS models and usually employ two to four people to operate and maintain the model for their uses. These NEMS model clones require EIA updates on an annual basis, which are posted on EIA's website at http://www.eia.doe.gov.

The MARKAL-MACRO MODEL

Overview

The MARKAL-MACRO Model[8] at DOE is an integration of two models, MARKAL and MACRO. MARKAL is the "bottom-up" technological model of energy and the environment, which includes depletable and renewable natural resources, processing of energy resources, and end-use technologies to meet the projected energy service demands in all sectors. MACRO is the "top-down" macroeconomic growth model that links MARKAL to the economy and maximizes utility (discounted sum of consumption). Top-down refers to models that are usually more aggregate in nature and estimate by forecasting a particular variable as a function of other aggregate causal factors. Bottom-up modeling approaches are more detailed at the individual equipment level and then sum up to the total in order to forecast variables.

MARKAL-MACRO finds the least-cost dynamic equilibrium under specific market and policy assumptions. DOE calibrates the MARKAL-MACRO Model

[8] DOE Energy MARKAL-MACRO Model contact: Philip Tseng, DOE, EIA, SMG Office, phillip.tseng@eia.doe.gov (202-287-1600); also see http://www.etsap.org.

to the NEMS outputs annually. The MARKAL-MACRO Model, used by over 45 countries, was developed by Brookhaven National Laboratory and then further improved by 18 Organization for Economic Cooperation and Development (OECD) countries.

Structure

MARKAL-MACRO optimizes the mix of fuels and technologies based on the consumer discount rate, technology characteristics, and fuel prices. Marginal costs for technologies and applications are used to determine the most efficient level of energy inputs along with technology costs and energy efficiencies. The model forecasts emissions sources and levels for CO_2, SO_X, NO_X, and any user-specified pollutants and wastes. The value of carbon rights (marginal cost of emissions) is one of the important outputs of the model. Outputs are solved in five-year intervals through 2050. Transportation coverage includes passenger cars, light trucks, heavy trucks, buses, airplanes, shipping, passenger rail, and freight rail.

The model can output a business-as-usual energy and carbon emissions profile. Identification of dynamic technology paths to meet emissions growth targets is one of the more common uses of the model outputs. The MARKAL-MACRO Model has facilitated the study by many countries of the costs of alternative approaches to reducing CO_2 emissions. Policy options would include fuel switching, substitution of capital and/or labor for energy services, demand reduction, emissions taxes, etc. MARKAL-MACRO can also identify opportunities for reducing CO_2 emissions through supply and demand technologies. Based on the technologies chosen, the model can calculate the cost of CO_2 emissions reductions.

The MARKAL-MACRO Model has proved useful in a number of areas. DOE has used it to analyze the Energy Policy Act of 1992. EIA has also built an international version of the MARKAL Model called SAGE to generate the annual *International Energy Outlook*. EPA is developing a national MARKAL database and scoping out a regional MARKAL representation of the U.S. economy. The MARKAL family of models is used by over 45 countries to support energy and environmental planning. The International Energy Agency (IEA) has a version of the Global MARKAL Model that they use to look into future energy technology perspectives. Most recently the model has focused on externalities measurement, hydrogen economy development, cost-competitive life cycle analysis, oil market response, technology learning, and country analysis.

Limitations

There are some limitations of MARKAL-MACRO. While it can provide an alternative and complimentary approach in longer term analysis (e.g., projection of renewable fuel penetration and reduction of CO_2 emissions), the model does not cover as much detail in all sectors as the NEMS model. The MARKAL-MACRO Model uses a simple approach to forecast energy service demands based on economic indicators such as housing stocks, commercial floor space, industrial production index, and VMT. Modeling at the individual equipment level would be difficult and require off-line analysis combined with aggregate implementation in MARKAL-MACRO. Individual sector modeling is relatively aggregate and may also require similar off-line analysis.

Resources

The data inputs to the model use about 7 to 20 MB of storage space, and the sourcecode is approximately 7 to 10 MB. The model can be run on a Pentium 4 processor with a 2 gigahertz (GHz) processor speed and 256 MB of RAM. Model execution is fairly quick at around five minutes. The model is quite complicated and requires special skills to run, similar to the NEMS model but with fewer people. Generally, two national laboratory analysts use and maintain the model for DOE. The MARKAL-MACRO Model is written in GAMS (General Algebraic Modeling System) programming language.[9]

[9]See http://www.gams.de/.

MiniCAM MODEL

Overview

The MiniCAM Model,[10] maintained by the Pacific Northwest National Laboratory (PNNL) forecasts CO_2 and other GHG emissions, and it estimates the impacts on GHG atmospheric concentrations, climate, and the environment. Although the model is a top-down agriculture-energy-economy model, it contains bottom-up assumptions about end-use energy efficiency. MiniCAM Model projections are made through 2100 and, therefore, the model has more futuristic technologies than NEMS. The model outputs forecasts in 15-year increments. Projections cover the entire planet in 14 global regions: the United States, Canada, Western Europe, Australia and New Zealand, Japan, the former Soviet Union, Eastern Europe, China, Southeast Asia, the Middle East, Africa, Latin America, South Korea, and India. Projections for Mexico, Argentina, and Brazil are under development.

Structure

MiniCAM is comprised of three larger models: the Edmonds-Reilly-Barns Model (ERB), the Agriculture Land Use Model (AGLU), and the Model for the Assessment of GHG Induced Climate Change (MAGICC). ERB represents the energy/economy/emissions system, including supply and demand of energy, the energy balance, GHG emissions, and long-term trends in economic output. AGLU simulates global land-use change from the production of composite crops, animal products, and forest products, and tracks GHG emissions associated with land use. MAGICC models the atmospheric/climate/sea-level system, which includes a gas cycle, climate, and sea-level model. MAGICC outputs atmospheric composition, radiative forcing, global mean temperature change, and sea-level rise.

Energy supply and demand are calculated in the model. Energy supply of renewable and nonrenewable sources is dependent on resource constraints, behavioral assumptions, and energy prices by region. Energy demand is a function of population, labor productivity, economic activity, technological change, energy prices, and energy taxes and tariffs. Transportation is one of three sectors (residential/commercial and industrial are the other two sectors) covered in the model, and passenger and freight technologies and modes are included. Model inputs consist of total service, service cost, vehicle technologies and their characteristics, price and income elasticities, technical change, percentage of population licensed to drive, and average speeds. The transportation system coverage includes automobiles, light trucks, buses, rail, air, and motorcycles for passenger modes; and trucks, rail, air, ship, pipeline, and motorcycles for freight modes. Seven major energy sources are modeled: oil, gas, coal, biomass, resource-constrained renewables, nuclear, and solar.

Limitations

Similar to the MARKAL-MACRO, individual equipment policies and certain detailed technology modeling would require off-line analysis and then aggregate implementation within MiniCAM. The model capabilities lie only at the national level and do not extend to the regional or state level. Due to the combination of three models within MiniCAM, the complexity of running the model may require specialized knowledge of the operations. In addition, because the model also contains scientific climate and atmospheric conditions, interpretation of the MAGICC results may require specialized knowledge in climatology. However, the three models within the MiniCAM can be run independently to narrow the analytical focus, and the results from each model can be interpreted separately.

Resources

The current MiniCAM Model[11] has an executable file size of about 1 MB and the data input files are about the same size. The sourcecode is approxi-

[10] MiniCAM Model contact: Son H. Kim, Pacific Northwest National Laboratory, skim@pnl.gov (301-314-6763) or Mariana Vertenstein, mvertens@ucar.edu (303-497-1349); also see http://www.pnl.gov/aisu/pubs/chinmod2.pdf and http://sedac.ciesin.org/mva/minicam/MCHP.html.

[11] As of 2005, the version of the MiniCAM Model described here has been superseded by ObjECTS-MiniCAM, a C++ version of the model that incorporates object-oriented programming designs for increased flexibility, maintenance, and modeling detail.

mately 903 kilobytes. Run time is approximately 30 seconds on a Pentium 4 (1.7 GHz), depending on the number of scenarios run at one time. MiniCAM can operate on a Pentium III or higher speed processor. FORTRAN is the modeling language using a MicroSoft Visual Studio compiler. However, there is a graphics user interface (GUI) front-end to the model if desired, which requires MicroSoft Acesss and Excel software. With the GUI, the user can run multiple scenarios at once and query, view, and chart results. Currently five or more people use MiniCAM at PNNL, and its maintenance requires two individuals.

GREET MODEL

Overview

> The GREET model is intended to serve as an analytical tool for use by researchers and practitioners in estimating fuel-cycle energy use and emissions associated with alternative transportation fuels and advanced vehicle technologies.[12]

GREET, maintained by Argonne National Laboratory (ANL), provides full fuel-cycle emissions analysis from wells to wheels, which represents emissions from all phases of production, distribution, and use of transportation fuels. Besides the fuel-cycle GREET model (GREET Series 1), there is a vehicle cycle model (GREET Series 2) that simulates emissions and energy use of direct input sources such as vehicle production, disposal, and recycling. ANL is currently finalizing a new version of GREET Series 2 for release. The strength of GREET is that it analytically compares energy use and emissions from vehicle technologies matched with many fuels, especially very advanced alternative fuels, over the entire fuel cycle. Emissions in the model include the following: GHGs (CO_2, methane, and nitrous oxide), NO_X, HC, CO, sulfur dioxide, and particulate matter. Other versions of the model have also included toxic pollutants, such as formaldehyde, acetaldehyde, 1,3-butadidine, and benzene (Winebrake et al. 2000).

[12] GREET Model contact: Michael Wang, Argonne National Laboratory, mqwang@anl.gov (630-252-2819); also see http://www.transportation.anl.gov/pdfs/TA/153.pdf.

Structure

The beauty of GREET is that it has a substantial combination of vehicle technologies and fuel types. GREET contains the following powertrains: conventional, direct injection, spark ignition, compression ignition, hybrid electric vehicles (which can be grid connected or not), electric vehicles, and fuel cell vehicles. Fuel types are also numerous: gasoline (reformulated or nonreformulated), diesel and low sulfur diesel, compressed natural gas, liquefied petroleum gas, liquefied natural gas, dimethyl ether, Fischer-Tropsch (FT) diesel, gaseous and liquid hydrogen, methanol, ethanol, biodiesel, and electricity. The GREET model contains more than 85 fuel production pathways and more than 75 vehicle/fuel combinations. These powertrains and fuel types can be produced from several feedstocks: petroleum, natural gas, flared gas, landfill gas, corn, cellulosic biomass, soybeans, and electricity. GREET is an excellent model to determine individual vehicle emissions and would be valuable in assisting evaluation of new transportation fuels and advanced vehicle technologies. EPA has incorporated GREET into their air emissions MOVES Model.

Limitations

GREET applies only to light-duty vehicles; however, this does not preclude it from being used for other vehicle types in the future. GREET does not include a vehicle choice model to forecast what people might purchase based on consumer preferences, but GREET output (total fuel-cycle emissions factors) can be used with future vehicle technology projections to get a more complete picture of the environmental impacts of these vehicle populations. Additionally, the model may be used in combination with policy options to reduce emissions and set emissions standards to achieve a goal.

Resources

The hardware requirements to run and operate the model are GREETGUI (GREET with a GUI interface or front end), which operates on PCs with Microsoft Windows 2000 or later. Minimum hardware requirements are a Pentium III processor at 166 megahertz (MHz) or higher, at least 64 MB RAM; and at least 30 MB of free space on the hard

drive. The recommended hardware profile is a Pentium processor at 400 MHz or higher, 128 MB or more of RAM, 100 MB of free hard disk space or more (Argonne 2001). GREET uses an Excel spreadsheet and Visual Basic. Use of GREET requires installation of MS Excel on a PC. Future plans are to convert it to C language in 2006.

GREET can also be run as a spreadsheet model that takes about 5 MB on an Excel spreadsheet. GREET recently added a Monte Carlo simulation module that stochastically generates a distribution rather than a point estimate. Running the model would normally be almost instantaneous, but for Monte Carlo simulations with Crystal Ball commercial software, run times may be approximately 3½ hours. Four people developed and are currently maintaining and running GREET at ANL.

TAFV MODEL

Overview

The Transitional Alternative Fuels and Vehicle Model[13] represents economic decisions among auto manufacturers, vehicle purchasers, and fuel suppliers, including distribution to end users. The model simulates decisions during a transition from current fuels to alternative fuels and traditional vehicles to advanced technology vehicles. Limited availability of alternative fuels, including refueling infrastructure, and availability of alternative fuel vehicle technologies are interdependent. TAFV assumes retail alternative fuel providers will maximize profits and spread capital costs across outlets to increase availability.

Structure

TAFV contains a model for predicting the choice of alternative fuel and alternative vehicle technologies for light-duty motor vehicles. The nested multinomial logit mathematical framework is used to estimate vehicle choice among technologies and fuel type combinations based on consumer preferences and vehicle attributes. Vehicle choice is dependent on prices, fuel availability, and the diversity of vehicle offerings (all endogenous) as well as luggage space, refueling time, vehicle performance, and cargo space (all exogenous parameters). Alternative fuel vehicles have three costs to vehicle manufacturers: capital costs, variable costs, and costs associated with diverse vehicle offerings. Calibration of the model through some key parameters, such as the value of time, the value of fuel availability, and discount rates, is based on existing literature. A spreadsheet model has been developed for calibration and preliminary testing. TAFV includes a range of vehicle- and fuel-related policies, including taxes or subsidies, federal mandates for vehicle acquisition (i.e., policies such as the Low Emission Vehicle Program and the Energy Policy Act). In addition, TAFV tracks GHG emissions from fuel production and vehicles using GREET-based emissions factors.

Limitations

Limitations of TAFV are that it includes only light-duty vehicles; growth rates in transportation demand and oil and gas prices are exogenous; it is national (U.S.) in scope, omitting regional detail and international trends in vehicle use or GHG emissions; and it assumes competitive behavior under complete foresight.

Resources

The model is quite small at 208 kilobytes, but inputs could be a few megabytes of spreadsheet data. The main program is written in the GAMS language. TAFV uses the MINOS5 and CONOPT2 nonlinear optimization solvers. The sourcecode is about 111 kilobytes in GAMS language, but the model requires more than 100 MB to execute. A model run takes approximately 30 to 60 minutes on a Pentium III 1,000 MHz PC to solve for the dynamic market equilibria (endogenous prices and quantities). Work files generated during a run can approach 1 gigabyte. Users should have 128 MB or more memory. TAFV can be run on Windows, Linux, and Unix, depending on which platform the licensed GAMS software resides. Maintenance currently involves a team of two; however, plans in the future are for a team of five over the next two years. TAFV has formed the foundation for an extended

[13] TAFV Model contacts: Paul Leiby, leibypn@ornl.gov (865-574-7720) and David Greene, Oak Ridge National Laboratory, and Jonathan Rubin, University of Maine; also see http://pzl1.ed.ornl.gov/altfuels.htm.

hydrogen vehicle transition model under development, HyTrans.

SUMMARY OF THE MODELS

Table 1 contains a short summary of the models reviewed in this study. The appendix provides documentation and more detailed information about each model, especially publications and studies that use the models. The appendix is meant to provide potential model users with a better understanding of how to apply the models to a specific problem and use.

CONCLUSION

Several very good models are available from which to choose when conducting GHG emissions studies, scenarios, or emissions estimates and forecasts for the transportation sector. Depending on the level of regionality and detail required, the model of choice will vary. Some of the models are more applicable at the aggregate level, such as MiniCAM and MARKAL-MACRO. Others such as NEMS, GREET, and TAFV are very detailed at the technology level. Maintenance, usability, resources, and analytical capabilities should be matched to the model choice.

REFERENCES AND BIBLIOGRAPHY[14]

Argonne National Laboratory. 1999. *Heavy- and Medium-Duty Truck Fuel Economy and Market Penetration Analysis for the NEMS Transportation Sector Model*, prepared for the U.S. Department of Energy, Energy Information Administration. Washington, DC. August.

_____. 2001. Development and Use of GREET 1.6 Fuel Cycle Model for Transportation Fuels and Vehicle Technologies, ANL/ESD/TM163. Center for Transportation Research, Energy Systems Division, Argonne, Illinois. June.

Energy Technology Systems Analysis Program. 2002. *ETSAP Newsletter* 8(2). August. Available at http://www.etsap.org.

Pacific Northwest National Laboratory, Advanced International Studies Group. 2001. *China-Korea-U.S. Economic Environmental Workshop Conference Proceedings, May 23–25, 2001.* Available at http://www.pnl.gov/aisu/pubs/chinmod2.pdf and http://sedac.ciesin.org/mva/minicam/MCHP.html.

U.S. Department of Energy (USDOE), Energy Information Administration (EIA). 2002a. *Emissions of Greenhouse Gases in the United States 2001,* DOE/EIA-0573(2002). Washington, DC. December.

_____. 2002b. *Transportation Energy Databook: Edition 22*, ORNL 69-67. Oak Ridge, TN: Oak Ridge National Laboratory. September.

_____. 2003a. *The National Energy Modeling System: An Overview 2003*, DOE/EIA-0581(2003). Washington, DC. March.

_____. 2003b. *The Transportation Sector Model of the National Energy Modeling System: Model Documentation Report*, DOE/EIA-M070(2003). Washington, DC. February.

U.S. Department of Transportation, Center for Climate Change and Environmental Forecasting. 2003. *Transportation Greenhouse Gas Emissions Data & Models: Review and Recommendations.* Cambridge, MA: USDOT, Volpe National Transportation Systems Center. March.

U.S. Environmental Protection Agency. 2002. *The U.S. Greenhouse Gas Inventory: In Brief*, EPA 430-F-02-008. Washington, DC. April.

Winebrake, J.J., H. Dongquan, and M. Wang. 2000. *Fuel-Cycle Emissions for Conventional and Alternative Fuel Vehicles: An Assessment of Air Toxics*, ANL/ESD-44. Argonne, Illinois: Argonne National Laboratory, Center for Transportation Research, Energy Systems Division. August.

APPENDIX A: MODEL USES

NEMS Model

General Topics of Energy-Related NEMS Studies

- Impacts of existing and proposed energy tax policies on the U.S. economy and energy system.
- Impacts on energy prices, energy consumption, and electricity generation in response to carbon mitigation policies such as carbon fees, limits on carbon emissions, or permit trading systems.
- Responses of the energy and economic systems to variations in world oil market conditions as a result of changing levels of foreign production and demand in developing countries.
- Impacts of new technologies on consumption and production patterns and emissions.
- Effects of specific policies on energy consumption, such as mandatory appliance efficiency and building shell standards or renewable tax credits.
- Impacts of fuel-use restrictions on emissions and energy supply and prices; for example, required use of oxygenated and reformulated gasoline or mandated use of alternative fuel vehicles.

[14] The appendix, which follows this section, includes additional bibliographic references.

TABLE 1 Technical Operating and Descriptive Characteristics for GHG Forecasting Models

	NEMS	Markal-Macro	MiniCAM	GREET	TAFV
Model size (data inputs and sourcecode)	Data inputs: 10–15 MB Sourcecode: >120 MB	Data inputs: 7–20 MB Sourcecode: 7–10 MB	Data inputs: 1 MB Sourcecode: 903 KB Executable: 1 MB	Excel spreadsheet~ 5 MB	Data inputs: several MB Sourcecode: 208 KB
Hardware requirements	512 MB RAM; Pentium processor	256 MB RAM; Pentium 4; 2 GHz processor	Pentium 4; 1.7 GHz processor	128 MB RAM; Pentium III; >400 MHz; need 100 MB of space on hard drive	128 MB RAM; Pentium III; 1,000 MHz; need 1 GB for model operational output
Software/ platform	PC platform; FORTRAN; Eviews software; OML linear programming software	Linux; DYNAMO modeling language	PC platform; FORTRAN; GUI Windows-based; MS Visual Studio; MS Access; MS Excel	PC platform; GUI with Windows 95 or higher but 98 recommended	Any platform; GAMS software
Run time	Standalone[1]: <1 minute Total Integrated[2]: 2–4 hours	5 minutes	<1 minute	Almost instantaneous; but if in simulation mode using stochastic distributions, then 3½ hrs.	30–60 minutes
Resources for maintenance	40 employees; 4 contractors	2 National Lab employees	2 National Lab employees	4 National Lab employees	2 National Lab employees
Transportation sector coverage	TRAN Module: LDV (car and light truck); Freight Truck (medium and heavy-duty); Aviation (wide and narrow-body, and general aviation; Rail (passenger and freight); Waterborne (passenger and freight); Miscellaneous (military, mass transit, recreational boats; criteria pollutant emissions and GHGs	End-user technologies by sector; LDV (car and light truck); heavy trucks; buses; airplanes; shipping; passenger rail; freight rail	Passenger mode: LDV (car and light truck), buses, rail, air, motorcycles. Freight mode: trucks, rail, air, ship, pipeline, and motorcycles.	Light-duty vehicle emissions for 8 advanced engine technologies (including hybrids and fuel cells) in combination with 15 fuel types including hydrogen, dimethyl ether and Fischer-Tropsch diesel	Light-duty vehicles; consumer choice model; auto manufacturers and fuel production and distribution sectors

[1] Standalone transportation mode—only the transportation module is operating while the other module components are static.
[2] Total Integrated mode refers to a model run, which has all of the modules active or operating and represents a dynamic equilibrium solution.

(continues on next page)

TABLE 1 Technical Operating and Descriptive Characteristics for GHG Forecasting Models *(Continued)*

	NEMS	Markal-Macro	MiniCAM	GREET	TAFV
Economic component	Uses Global Insight Macro Model and integrates all sectors of economy including employment and Census division regional models	Macro Growth Model fully integrated	ERB Model: 3 sector economy—residential/ commercial, transportation, industrial; long-term trends in economic output	None/not applicable	Uses macroeconomic inputs
Forecast period	2000–2025	Through 2050	Through 2100	Current year of operation using driving cycle	Through 2030
Time period	Annual	5 year intervals	15 yearly increments	Current year of operation using federal driving cycle	Annual
Optimizing solution	Dynamic equilibrium convergence with iterations	Constrained least cost dynamic equilibrium	Constrained dynamic equilibrium	None/not applicable but can be used with stochastic processes	Nonlinear optimization solvers
Regionality	U.S. by 9 Census divisions; however, some supply modules may be using industry regions also	United States	14 global regions: U.S., Canada, Western Europe, Australia and New Zealand, Japan, former Soviet Union, Eastern Europe, China, Southeast Asia, Middle East, Africa, Latin America, South Korea, and India	None/not applicable because it measures emissions from a vehicle type and not in aggregate	U.S. and some world energy supply areas
Emissions measured	Carbon and criteria pollutants: nitrogen oxides (NO_x), sulfur oxides (SO_x), carbon monoxide (CO), volatile organic compounds (VOC), particulates	Carbon dioxide (CO_2), SO_x, NO_x	CO_2, nitrous oxide (N_2O), methane (CH_4), CO, NO_x, VOC	CO_2, CH4, and N_2O, and criteria pollutants: VOC, CO, NO_x, particulate matter smaller than 10 microns (PM-10), and SO_x	CO_2 greenhouse gas equivalent; however, the model outputs are usually run through GREET to calculate other emissions

Sources: Personal c+ommunication with model authors—NEMS Model: John Maples; Markal-Macro Model: Phillip Tseng; Mini-Cam Model: Son H. Kim; GREET Model: Michael Wang; TAFV Model: Paul Leiby.

- Impacts on the production and price of crude oil and natural gas resulting from improvements in exploration and production technologies.
- Impacts on the price of coal resulting from improvements in productivity.
- Numerous energy-related studies for Congress or federal agencies: carbon and vehicle emissions modeling for Congress, EPA, and DOE.
- Transportation-specific model runs for the White House and other governmental agencies:
  * **Transportation gasoline tax model runs for the White House**, 1996. These model runs led to the 3¢ tax on gasoline implemented by the administration in 1996.
  * U.S. Department of Energy, Energy Information Administration, *Analysis of Corporate Average Fuel Economy Standards for Light Trucks and Increased Alternative Fuel Use*, SR/OIAF/2002-05 (Washington, DC: March 2002). This service report assesses the impacts of more stringent corporate average fuel economy (CAFÉ) standards on energy supply, demand, prices, macroeconomic variables where feasible, import dependence, and emissions. This study addresses the provisions of H.R. 4, S. 804, and S. 517 that pertain to light-vehicle fuel economy in the transportation sector. A qualitative discussion is provided for the alternative fuels provisions included in S. 1766 and H.R. 4 **at the request of the Senate Committee on Energy and Natural Resources.**
  * _____. *The Transition to Ultra-Low-Sulfur Diesel Fuel: Effects on Prices and Supply*, SR/OIAF/2001-01 (Washington, DC: May 2001). This study evaluates EPA's ultra-low-sulfur diesel fuel regulations for heavy-duty trucks **at the request of the House Committee on Science.**
  * _____. *The Impacts of Increased Diesel Penetration in the Transportation Sector*, prepared by the Office of Integrated Analysis and Forecasting (Washington, DC: August 1998). These model runs and scenarios were developed for the Office of Transportation Technologies within DOE.
  * Request from EPA on travel and emissions associated with various heavy-duty truck emissions standards levels for criteria pollutants.
  * Request from the U.S. Government Accountability Office (GAO) to estimate future alternative fuels penetration levels.
  * Request from GAO to estimate alternative fuel vehicle sales and stocks effect of the Energy Policy Act.
  * NEMS vehicle travel equations were used to develop a DOT, Federal Highway Administration (FHWA) VMT model. **The proposed VMT model development was an interagency effort between EIA, EPA, and FHWA**.

MARKAL-MACRO Model

MARKAL-MACRO was used for a project on "Policies and Measures for Common Action" conducted by the Annex I Expert Group on the United Nations (UN) Framework Convention on Climate Change.[15]

As part of a study by the OECD Secretariat on the environmental implications of energy and transport subsidies, the Italian participant used an "elastic" version of MARKAL to evaluate the impact of removing financial subsidies from the electric sector in Italy. The many ways in which financial interventions affect the electric supply industry were searched out, and MARKAL was used to assess their effect on electric and energy system costs and CO_2 emissions.

The Energy Technology Systems Analysis Programme (ETSAP) of IEA continues to provide a multinational capability to determine the most cost-effective national choices to limit future emissions of greenhouse gases by using consistent methodology that offers a basis for international agreement on abatement measures. The basic MARKAL Model continues to serve national interests, as illustrated by its use for a major national research and development (R&D) appraisal in the United Kingdom, its use to help develop the national least-cost energy strategy in the United States, and its acceptance by a wider international community. Outside ETSAP, MARKAL was used in Taiwan and (in the form of

[15] See http://www.etsap.org/annex5/main.html#3.1.

MENSA) in Australia to inform the debate on response strategies under the UN Framework Convention on Climate Change.

With the cooperation of the participants from Italy, Japan, the United Kingdom, and the United States, ETSAP contributed to the IEA study, "Electricity and the Environment." Detailed descriptions were provided of technologies available for electricity supply and demand in the short and medium term, including technical performance and engineering costs. Specific data were drawn from the MARKAL databases of the four cooperating countries.

Although a common set of runs among the ETSAP participants was delayed, four countries participated in CHALLENGE, a cooperative international project on energy and environment systems analysis. CHALLENGE consists of a network of scientists from Eastern and Western European countries. The project is intended to facilitate international negotiations and cooperation by providing a scientific basis for decisions on response strategies to reduce environmental stresses and climate risks due to energy use.

During Annex V, some participating countries provided inputs to major international studies by IEA, OECD, and the Annex I Expert Group on the UN Framework Convention on Climate Change.

ETSAP originated as an IEA program to help establish energy technology R&D priorities on the basis of the needs of all the IEA countries. A common methodology and comparable databases have been the touchstone of the program since its very beginnings. The standard MARKAL Model has continued to be the focus of the group's analyses, and recurring efforts have been made to assure reasonable consistency in the national databases.

GREET Model

The major applications of the GREET Model (reports available at www.transportational.gov) consist of the following:

1. Energy and GHG emissions effects of fuel ethanol (for the state of Illinois, DOE, the U.S. Department of Agriculture, and EPA).
2. Energy and emissions effects of natural gas-based transportation fuels for DOE.
3. Well-to-wheels analysis of energy and GHG emissions of advanced vehicle technologies and transportation fuels for General Motors (three volume report).
4. Fuel-cycle energy and emissions effects of the fuels petitioned to DOE under the Energy Policy Act.
5. Work with EPA to integrate GREET into EPA's next generation of motor vehicle emissions model (called MOVES).

APPENDIX B: MODEL BIBLIOGRAPHY AND PUBLICATIONS

NEMS Model

- Numerous energy-related studies for Congress or federal agencies:

U.S. Department of Energy, Energy Information Administration, *Analysis of Corporate Average Fuel Economy (CAFÉ) Standards for Light Trucks and Increased Alternative Fuel Use*, SR/OIAF/2002-05 (Washington, DC: March 2002).

_____. *Analysis of Efficiency Standards for Air Conditioners, Heat Pumps, and Other Products*, SR/OIAF/2002-01 (Washington, DC: February 2002).

_____. *Analysis of Strategies for Reducing Multiple Emissions from Power Plants: Sulfur Dioxide, Nitrogen Oxides, and Carbon Dioxide*, SR/OIAF2000-05 (Washington, DC: December 2002).

_____. *Impact of Renewable Fuel Standard/ MTBE Provisions of S. 1766*, SR/OIAF/2002-06 (Washington, DC: March 2002).

_____. *Impact of Renewable Fuel Standard/ MTBE Provisions of S. 517: Addendum*, SR/ OIAF/2002-06 (Washington, DC: April 2002).

_____. *Impacts of a 10-Percent Renewable Portfolio Standard*, SR/OIAF/2002-03 (Washington, DC: February 2002).

_____. *Impacts of the Kyoto Protocol on U.S. Energy Markets & Economic Activity*, SR/ OIAF/98-03 (Washington, DC: October 2002).

_____. *Measuring Changes in Energy Efficiency for the Annual Energy Outlook 2002* (Washington, DC: 2002).

_____. *Reducing Emissions of Sulfur Dioxide, Nitrogen Oxides, and Mercury from Electric Power Plants,* SR/OIAF/2001-04 (Washington, DC: September 2001).

_____. *Strategies for Reducing Multiple Emissions from Electric Power Plants with Advanced Technology Scenarios*, SR/OIAF/2001-05 (Washington, DC; October 2001).

- Carbon and vehicle emissions modeling for Congress, EPA, and DOE:

Interlaboratory Working Group on Energy-Efficient and Low-Carbon Technologies, *Scenarios for a Clean Energy Future* (Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory, National Renewable Energy Laboratory, and Argonne National Laboratory), ORNL/CON-476 and LBNL-44029 (Oak Ridge, TN: November 2000).

_____. *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy-Efficient and Low-Carbon Technologies by 2010 and Beyond* (Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory, National Renewable Energy Laboratory, and Argonne National Laboratory) (Oak Ridge, TN: September 1997).

U.S. Department of Energy, Energy Information Administration, *Analysis of the Climate Change Technology Initiative: Fiscal Year 2001*, prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/2000-01 (Washington, DC: April 2000).

_____. *Analysis of the Climate Change Technology Initiative*, prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/99-01 (Washington, DC: April 1999).

_____. *Analysis of the Impacts of an Early Start for Compliance with the Kyoto Protocol*, prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/99-02 (Washington, DC: July 1999).

_____. *Impacts of the Kyoto Protocol on U.S. Energy Markets and Economic Activity*, prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/98-03 (Washington, DC: October 1998).

_____. *Service Report: Analysis of Carbon Stabilization Cases*, prepared for the U.S. Department of Energy, Office of Policy and International Affairs, SR-OIAF/97-01 (Washington, DC: October 1997).

- Transportation-specific model runs for the White House and other governmental agencies:

U.S. Department of Energy, Energy Information Administration, *Analysis of Corporate Average Fuel Economy (CAFÉ) Standards for Light Trucks and Increased Alternative Fuel Use*, SR/OIAF/2002-05 (Washington, DC: March 2002).

_____. *The Transition to Ultra-Low-Sulfur Diesel Fuel: Effects on Prices and Supply*, SR/OIAF/2001-01 (Washington, DC: May 2001). This study evaluates EPA's ultra-low-sulfur diesel fuel regulations for heavy-duty trucks **at the** *request of the House Committee on Science.*

_____. *The Impacts of Increased Diesel Penetration in the Transportation Sector*, prepared by the Office of Integrated Analysis and Forecasting (Washington, DC: August 1998).

MiniCAM Model

Edmonds, J. and J. Reilly, *Global Energy: Assessing the Future* (New York, NY: Oxford University Press, 1985).

Edmonds, J.A., J.M. Reilly, R.H. Gardner, and A. Brenkert, "Uncertainty in Future Global Energy Use and Fossil Fuel CO_2 Emissions 1975 to 207," TR036, DO3/NBB-0081 Dist. Category UC-11, prepared for the U.S. Department of Commerce (Springfield, VA: National Technical Information Service, 1986).

Edmonds, J.A., M.A. Wise, and C.N. MacCracken, *Advanced Energy Technologies and Climate Change: An Analysis Using the Global Change Assessment Model (GCAM)*, PNL-9798 (Richland, WA: Pacific Northwest Laboratory, 1994).

Richels, R. and J. Edmonds, "The Economics of Stabilizing Atmospheric CO_2 Concentrations," *Energy Policy* 23(415):373, 1995.

GREET Model

Wang, M.Q., *GREET 1.5: Transportation Fuel-Cycle Model, Volume 1* (Argonne, IL: Argonne National Laboratory, 1999).

Wang, M.Q. and H.S. Huang, *A Full Fuel-Cycle Analysis of Energy and Emissions Impacts of Transportation Fuels Produced from Natural Gas* (Argonne, IL: Argonne National Laboratory, 1999).

Wang, M., C. Saricks, and M. Wu, "Fuel Ethanol Produced from Midwest US Corn: Help or Hindrance to the Vision of Kyoto?" *Journal of the Air and Waste Management Association* 49(7):756–772, 1999.

Winebrake, J.J., M.Q. Wang, and D. He, "Toxic Emissions from Mobile Sources: A Total Fuel Cycle Analysis of Conventional and Alternative-Fuel Vehicles," *Journal of the Air and Waste Management Association* 51(7):1073–1086, July 2001.

TAFV Model

Leiby, P. and J. Rubin, "Transitions in Light-Duty Vehicle Transportation: Alternative Fuel and Hybrid Vehicles and Learning," *Transportation Research Record* 1842:127–134, 2003.

_____. "Flexible Greenhouse Gas Emission Banking Systems," *Maine Agriculture and Forest Experiment Station Miscellaneous Report,* No. 427, April 2002.

_____. "Effectiveness and Efficiency of Policies to Promote Alternative Fuel Vehicles," *Transportation Research Record* 1750:84–91, 2001.

_____. "The Alternative Fuel Transition: Results from the TAFV Model of Alternative Fuel Use in Light-Duty Vehicles 1996–2010 Final Report, TAFV Version 1," *Maine Agricultural and Forest Experiment Station Miscellaneous Report,* No. 417, September 2000.

_____. "Dynamic Analysis of Achievable Potential and Costs for Alternative Fuel Vehicles," invited presentation at the International Energy Agency, International Workshop on Technologies to Reduce Greenhouse Gas Emissions, Washington, DC, May 1999.

_____. "Sustainable Transportation: Analyzing the Transition to Alternative Fuel Vehicles," *Transportation Research Board Circular* 492:54–82, August 1999.

_____. "The Transitional Alternative Fuels and Vehicles Model," *Transportation Research Record* 1587:10–18, 1997.

Leiby, P.N., J. Rubin, and D. Bowman, "Efficacy of Policies to Promote New Vehicle Technologies: Alternative Fuel Vehicles and Hybrid Vehicles," Proceedings of the 25th Annual IAEE International Conference, Aberdeen, Scotland, 26–29 June 2000.

Rubin, J. and P. Leiby, "An Analysis of Alternative Fuel Credit Provisions of U.S. Automotive Fuel Economy Standards," *Energy Policy* 28(9):589–602, 2000.

Estimating Confidence Intervals for Transport Mode Share

STEPHEN D. CLARK*
JOHN MCKIMM

Transport Policy Monitoring
Development Department
Leeds City Council
Leonardo Building, 2 Rossington Street
Leeds LS2 8HD England

ABSTRACT

One of the common statistics used to monitor transport activity is the total travel by a particular method or mode and, for each mode, this share is routinely expressed as a percentage of total personal travel. This article describes a simple model to estimate a confidence interval around this percentage using Monte Carlo simulation. The model takes into account the impact of both measurement errors in counting traffic and daily variations in traffic levels. These confidence intervals can then be used to test reliably for significant changes in mode share. The model can also be used in sensitivity analysis to investigate how sensitive the width of this interval is to changes in the size of the measurement errors and daily fluctuations. A bootstrap technique is then used to validate the Monte Carlo estimated confidence interval.

INTRODUCTION

The last 5 to 10 years in United Kingdom transport has seen the establishment of an increasing number of targets against which the performance of the transport system is to be measured. Many of these targets are expressed in precise numerical terms, and sophisticated monitoring regimes are in place

Email addresses:
* Corresponding author—Stephen.clark@leeds.gov.uk
J. McKimm—john.mckimm@leeds.gov.uk

KEYWORDS: Mode share, confidence intervals, Monte Carlo, bootstrap.

to determine the current value of the measure of interest. In some cases, this monitoring can provide complete information about the measure (the population), but more commonly only information on a sample of the measure is possible. Information from the sample is then used to infer the behavior of the population. Statistics tell us that all samples are subject to variation and in judging the value of an indicator (and in particular whether a target has been achieved) some account of this variability is necessary. Therefore, it is important to ensure that the precision of the monitoring regime that estimates the required indicator is compatible with the specified target level for the indicator.

The following section presents the background to the statistic to be modeled in this paper: the percentage of people who travel by a particular mode. The next section describes the Monte Carlo technique used to estimate the confidence interval around this statistic. The following section presents the survey methodology used by the city of Leeds in the United Kingdom to collect the base data. By using the information on how the base data were collected, ranges can be set for likely measurement errors and daily variation, which are detailed in the next two sections. A number of the implicit assumptions that result from this exercise are then highlighted. We next report on the application of the Monte Carlo technique and the issues surrounding the sensitivity analysis and sample size determination. The penultimate section uses the technique of bootstrap estimation to "validate" the Monte Carlo estimates of mode share deviation. The final section provides some suggestions on how the technique can be adapted for other purposes.

MODE SHARE STATISTICS

Local government authorities regularly undertake surveys to measure the volume of traffic and travel in their areas to aid in planning services and targeting investment. The measure of travel usually adopted is that by people rather than by vehicle. This allows for a more meaningful measure of travel to be estimated, because, for example, a fully loaded bus carries far more people than a single car. These surveys can range over a designated area (e.g., a town or city), be concerned purely with journeys across a designated cordon, or may result from an individual or household travel diary.

Because the volume of total travel in different areas varies, it is common to present, for each mode, these volumes as a share of the total travel volume in the area and to express this share as a percentage. This then enables a comparison to be made of mode shares between areas. Also, if such surveys are conducted at regular time intervals, then trends in each mode of travel can be identified.

Concerns arise when these surveys are based on a small sample size, maybe as few as one full day of observation (Royal Statistical Society 2005; USDOT 2003). These small sample sizes should not, however, be much of a surprise since, typically, a six-hour survey in a large metropolitan area may cost upwards of £10,000 (about $18,000). Obtaining a more reliable estimate of the mode share and the precision of this estimate would require more survey days; just to halve the standard error of the mean estimate requires three extra days, bringing the cost of the survey to £40,000 (about $72,000). But without an indication of this sampling variability, it is difficult to conclude that any observed changes are real and statistically significant.

Some survey techniques, such as stated preference surveys, attempt to estimate mode share, and, since they use well-understood statistical models, they are able to provide confidence bounds around any mode share estimates (Ortuzar and Willumsen 1994). Such surveys are, however, typically concerned with making a choice that involves at least one hypothetical alternative. Furthermore, they have other errors that may lead to greater imprecision than already present and are costly to administer and analyze.

The study described in this paper focuses on an alternative form of data, namely revealed preference data, where the modes actually used by individuals are recorded. Also, this information is provided in an aggregate form of travel data (i.e., the number of people traveling by the different modes) rather than the disaggregate form of household or individual travel diaries.

MONTE CARLO SIMULATION

Simulation is an attempt to replicate a real world phenomenon using a model and a set of simplifying assumptions. One form of simulation that involves the assessment of the behavior of random variables (e.g., observed traffic flow or vehicle occupancy) is the Monte Carlo approach. The method assumes that the traffic flow (or other variable) follows a statistical probability distribution. As part of the simulation process, repeated instances of random observations are taken from this assumed distribution, and the impact of these random draws on some output measure is recorded. Using this simple sampling approach, many replications can be made, and a reliable estimate of the output measure and its spread can be obtained.

This paper uses the Monte Carlo approach to simulate the observed differences that can occur as a result of measurement error and daily variation associated with the conduct of a cordon traffic survey. By obtaining a large simulated set of these errors and variations and using them to "correct" the observed count, it will be possible to calculate a set of confidence intervals around the output measure, in this case mode share.

While results from the statistical literature allow the distribution for a mode share to be established (see appendix), this closed-form distribution approach contains a number of disadvantages:

- reliable estimates of the parameters for the distributions are difficult to obtain, because few sample observations are available;
- incorporating sophisticated multivariate relationships into the model is necessary, because, for example, the estimate of the share of travel by rail will impact on the share by all other modes; and
- the model and the methodology need to be easily explainable to nonstatisticians; mathematical models involving Greek symbols are not useful to such an audience.

Monte Carlo approaches have been used previously in the transportation field. These include structural reliability (Pothisiri and Hjelmstad 2003; Zhao and Ang 2003), traffic modeling (Cassidy et al. 1994; Tarko 2000), network reliability (Chen et al. 1999, 2002; Lam and Xu 1999), and activity modeling (Kreihich 1979; Veldhuisen et al. 2000; Castiglione et al. 2003).

Perhaps the most similar study to the work described here is that reported in Williamson et al. (2002), where the Monte Carlo approach was used to investigate whether short period traffic counts (of 5-, 10-, and 20-minute duration) can accurately represent hourly traffic counts. The first stage was to assume a Weibull distribution for the count data and to estimate the scale and shape parameters of the distribution. In the next stage, 1,000 instances of 60 observations (1 observation for each minute) from the appropriate Weibull distribution were generated and used to construct a cumulative distribution plot. From this plot, 90% confidence intervals were estimated, and if the actual observed hourly count fell within this interval then the estimation was deemed a success. An application of this methodology showed that contiguous 20-minute counts were required in order to accurately estimate an hourly traffic count.

SURVEY METHODOLOGY

This section describes the survey methodology used to collect the data for the example application of the Monte Carlo simulation. A thorough understanding of the survey methodology is important, because this will later help in defining the ranges for measurement errors and daily variations. All the data here (except rail data) were obtained from on-street observation by a team of enumerators, where all movements in one direction, across a datum line, were recorded. A discussion of the methodology for each mode of travel follows.

Cars. Each enumerator was asked to count the number of cars, categorized by the number of occupants (1, 2, 3, and 4 or more). Depending on the volume of traffic on the road, they may also have been required to count goods vehicles and cyclists.

Goods vehicles and cyclists. If the person who was counting cars could not handle this category of traffic, another enumerator was used to count these vehicles. Cyclists using dedicated paths or the pedestrian pavement were included in the count.

Buses. An enumerator recorded the type of bus observed and made a roadside assessment, without

boarding the bus, of how full it was. Four types of buses were counted: mini, single deck, double deck, and articulated. The occupancy was recorded as empty, one-quarter full, half full, three-quarters full, full, and full with standing passengers.

Rail. The local Passenger Transport Executive (PTE) provided an estimate of the average volume of passengers arriving at the central train station. This estimate was based on onboard head count surveys conducted by train operator staff on three days during a year and was supplemented by additional PTE commissioned counts. They were then reconciled with other databases to provide an adjusted estimate.

Walk. The number of people walking across the datum line was recorded.

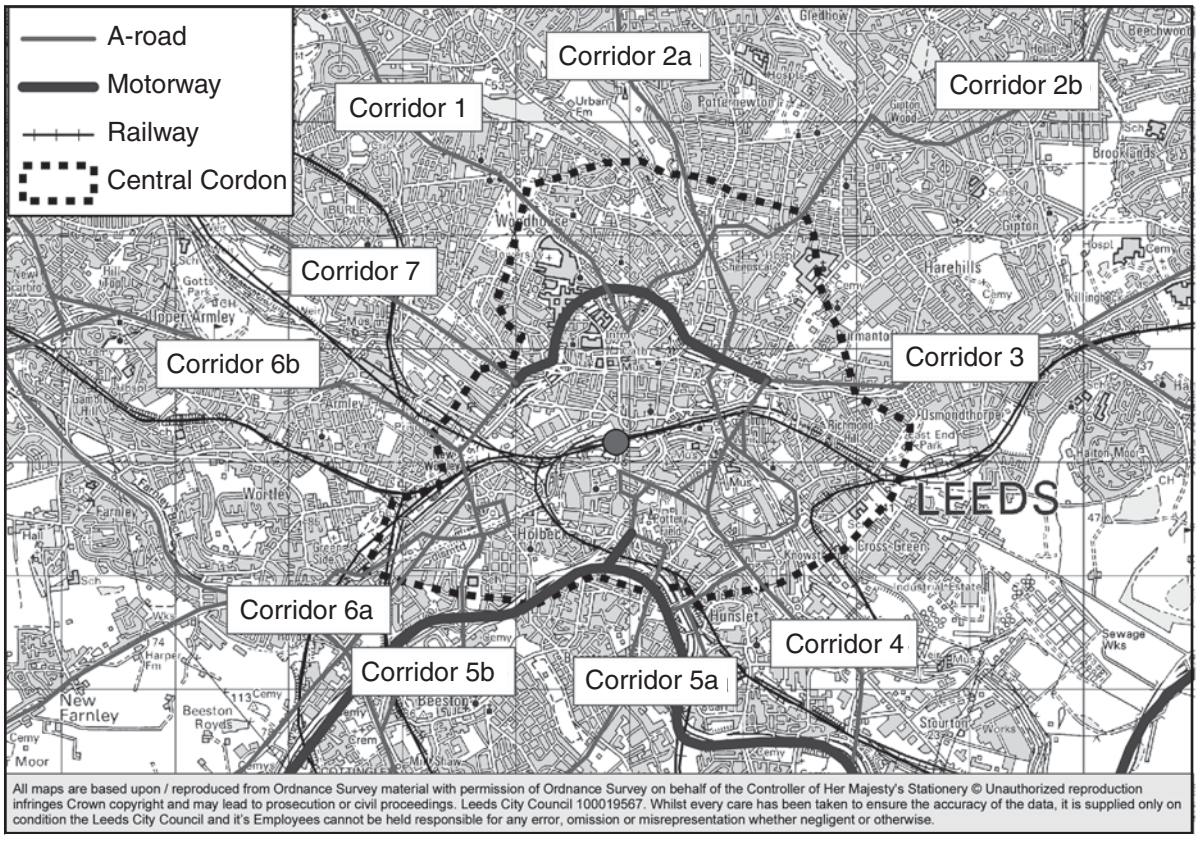
The surveys of the 34 radial roads into Leeds City Centre (figure 1) were conducted over 17 separate weekdays in May 2002 from 7:30 a.m. to 9:30 a.m. and 2:00 p.m. to 6:00 p.m. The number of radial roads surveyed on each day varied from one to up to five, but each road was counted only once.

The next two sections present ranges for the possible accuracy of the counts and the degree of daily variability during the morning peak. To a large degree, a Delphic approach (Dajani and Gilbert 1975), involving transport planners, survey managers, and statisticians, was used to arrive at a consensus opinion on the size of these error ranges. Other ranges may be used without invalidating the general Monte Carlo approach presented here.

MEASUREMENT ERROR

The measurement error assesses the accuracy of the enumerator counts. This is equivalent to comparing two (or more) counts of the same thing at the same time by different people to see how close they are in agreement. Clearly, this will depend on the skill and expertise of the staff involved.

FIGURE 1 Map Showing the Leeds Central Cordon and Arterial Corridors



Cars. The *Traffic Appraisal Manual* (DfT 2003) suggests that a skilled enumerator can achieve a 95% confidence interval accuracy of ±10%. Our own validation checks conducted by a second enumerator suggest that an interval of between ±5% and ±15% is usual. An error range of ±10% was selected for estimating the volume of single-occupant cars and a slightly larger range of ±12% for cars with more than one occupant, because this is a slightly more complex task.

Buses. It is likely that buses will be counted with more accuracy than cars, since they are a more visible presence on the road. Conversely, the measure of occupancy is likely to be inaccurate, because estimations of the occupancy must be made from the roadside. Table 1 gives the volume measurement and occupancy estimation errors for each type of bus. Minibuses have an error range similar to cars. All other bus types have a reduced error range, because they should be more noticeable. The error in estimating the occupancy of minibuses is low, because it is relatively easy for a quick and near precise estimate to be made. It is slightly more difficult to estimate the occupancy of single-deck buses. The most difficult task is estimating the vehicle occupancy of double-deck and articulated buses: with double-deck buses, it is very difficult to judge how full the top deck is; and with articulated buses, there is a large volume of information to assess visually. For these reasons, the occupancy error was set high at ±15%.

Rail. The PTE who provided the estimates for rail patronage judged the numbers to be accurate within a range of ±5%.

Walk and pedalcycle. Both these volumes are thought to be recorded at similar levels of accuracy to each other, near the ±10% mark.

Powered two-wheelers (PTWs). A PTW vehicle can be an inconspicuous part of the traffic. They do not necessarily keep to designated lanes and can easily speed along the carriageway or weave between lanes. This rational led to a high measurement error range of ±15%.

TABLE 1 The Range of Measurement Errors for Buses

Bus type	Volume errors	Occupancy errors
Minibus	±10%	±5%
Single deck	±5%	±10%
Double deck	±5%	±15%
Articulated	±5%	±15%

So far in this section, only the errors specific for each mode of travel have been quantified. In addition, it is not unreasonable to assume that there is a global error that affects all the modes counted on the same day. This may be due to generally unfavorable (foggy or wet) or favorable (dry and warm) roadside conditions. This global error is in addition to the mode-specific errors for all road-based volumes (i.e., not rail) and in this way modifies the mode-specific errors.

For the global volume errors, the range was set at ±5%. This means that, for example, a sample value for the error in estimating the volume of single-occupant car traffic was in the range of ±15% (a mode-specific element of ±10% and a global element of ±5%). Buses also have a global occupancy error to reflect the fact that in certain conditions (e.g., misty windows) occupancy in all buses will be difficult to estimate and also that rounding (to the nearest quarter) is involved. For buses, the global volume error was the same as for the other road-based modes, and the global occupancy error was set high at ±15%.

DAILY VARIATION

In addition to measurement error, taking into account the natural daily fluctuations that occur in traffic volumes is necessary. These variations can result from many causes; for example, a person may change his or her mode of travel or time of departure on successive days. Even if we were to count traffic with perfect accuracy, these daily variations will still be present in our data, and, in this section, estimates of the extent of these fluctuations are provided.

Cars. The daily variation in the volume of people traveling by car is specified for each category of car occupancy. These are set at ±5% for single- and double-occupant cars, ±8% for three-occupant cars, and ±12% for four or more occupants in a car. Some published evidence supports these ranges of variation. Phillips (1979) used a range of coefficients of variation of between 2.5% and 15% in determining the sample size for daily traffic flow estimation. Fox et al. (1998) suggested that a range for the coefficient of variation of 8% to 15% is appropriate, and, in the peak period, this value can be at the low end of this range (near 10%).

Buses. Buses run on a regular schedule each day, and, therefore, we would expect only small day-to-day variations in the number of buses counted. To quantify this, information reported in the 2003 West Yorkshire Local Transport Plan (WYPTA 2003) (which includes Leeds) shows that only 1.4% of all buses were canceled and of those that ran, 90% were less than 6 minutes late. In addition to this variation in the volume of scheduled buses, there was also variation in the average occupancy of buses. Both the volume and the occupancy variation are limited to ±5%.

Rail. Like buses, the volume of rail travel should be consistent from day to day. Statistics from the Strategic Rail Authority (2002) for the commuter rail operator in West Yorkshire show that the level of service reliability is comparable to that for buses. The percentage of train cancellations is 1.5%; however, the punctuality is slightly worse for trains, with just 83.8% of trains arriving within 5 minutes of their scheduled time (but 91.7% within 10 minutes). The range of variation was, therefore, set at ±5%, similar to the level for buses.

Walk. The volume of walk traffic is anticipated to vary slightly more than motorized methods of travel, because the traveler may easily substitute another mode (e.g., as a car passenger some days of the week or via bus on rainy days). The range was, therefore, set at ±10%.

PTWs. This mode is thought to be a highly variable form of travel. Statistics from the Department for Transport (1994) show that nearly 40% of motorcycle trips take place in the summer months and only 16% in the winter months. Many of these summer journeys will be for leisure purposes, and, because the primary concern here is with morning peak commuting trips, this suggests a range less than that indicated by the statistics. The range was set at ±12%.

Pedalcycle. Like walking and PTWs, this mode is thought to be highly variable on a day-to-day basis for many of the same reasons (DfT 1994, 1996). Cycling can, however, be even more unpleasant during adverse weather conditions than other modes (primarily for safety and comfort reasons) and so the variation range was set high at ±15%.

In addition to the mode-specific ranges of variation described here, an additional global element of variation was applied (in a similar manner to the global measurement error). This range of variation was set at ±5%. As a result, and referring to the values suggested for cars in this section, a compounded variation range of ±10% for single- and double-occupancy cars (a mode-specific element of ±5% and global element of ±5%) is possible.

MONTE CARLO SUMMARY

Before progressing to an illustrative example to show how this information is able to produce confidence intervals for mode share statistics, a few points are worth making.

- **Two distinct sources of uncertainty.** The measurement error represents the accuracy of the count, while the daily variability represents the fluctuation in these counts. Even if it were possible to count with 100% accuracy, there would still be daily variability, and, even if every traveler made the same journey by the same mode at the same time each day, there would still be differences in what enumerators counted.
- **Error structures.** Depending on the survey methodology adopted, the structure of the errors will change. If, instead of classifying cars by the occupants, one person counts both cars and people separately, it is likely that the measurement errors will be negatively correlated (i.e., they are able to count vehicles accurately but people inaccurately or vice versa).
- **Expertise required.** To set the ranges for the errors and variability requires some expertise and

assumptions. One approach is to start with a fairly well understood measure (e.g., the accuracy in enumerating cars) and set other rates relative to this.

- **Count duration.** The range of daily variation will depend on the schedule of when counts are conducted. The ranges for a survey of 25 locations, all conducted on 1 day, should be larger than an alternative survey where 5 locations are counted on 5 days and their values summed.
- **Correlation between days.** In the model specified here, no correlation exists in the errors or the variation between consecutive days. If it appears that, for example, high errors in counting at locations on one day would lead to a tendency to high errors on other days, then this could be accommodated within the model framework presented here.
- **Limitations on model use.** The model is purely concerned with travel behavior in an aggregate form and no information on the traveler's individual characteristics (e.g., gender, age, income) is required or used. The model cannot, therefore, anticipate the detailed results of policy interventions or produce forecasts of future behavior.

EXAMPLE APPLICATION

To apply the Monte Carlo technique to the problem of estimating confidence intervals, we used the Excel spreadsheet package. Excel provides all the facilities required to conduct the simulation (primarily the generation of random numbers, although some care is required; see Knusel 1998). It has the tools to interpret the output (i.e., produce graphs and tables) and is commonly available.

One aspect that still needs to be defined is the underlying distribution from which the sample errors and levels of variation are drawn. The simplest distribution available is the *uniform* distribution where each sample value within a range is equally likely. This does not appeal intuitively, because smaller error or variability values would be more likely than larger values. This requirement suggests that the *normal* distribution should be used. The normal distribution does not, however, have a limiting range; sampled values can extend between plus and minus infinity. Clearly, these more extreme values would not be expected to arise in practice, so we adopted the convention that 95% of the sampled error or variability rates should be within the set ranges for errors or variability as described above. The normal distribution is also symmetric. If it is thought that the measurement errors are one sided (i.e., either mostly under- or overestimates), then it is possible to sample primarily positive or negative values.

The sampling regime as described in this paper is built within a workbook.[1] A series of 17 worksheets hold the morning peak data collected on each of the 17 survey days. Each of these worksheets contains the following traffic information for all sites that were counted on that survey day:

1. the existing base case as surveyed during May 2002,
2. the sampled values for the measurement error; these errors are applied to the observed counts so that the measurement errors they contain are "corrected,"
3. the sampled daily variations; these are applied to the "error corrected" values calculated in step 2 to represent values that could reasonably be counted on a different survey day,
4. the measurement error calculations for buses; these calculations are more complex, because they are disaggregated by the four vehicle types and six occupancy levels,
5. the final results are the updated counts after the application of both the measurement errors and daily variations.

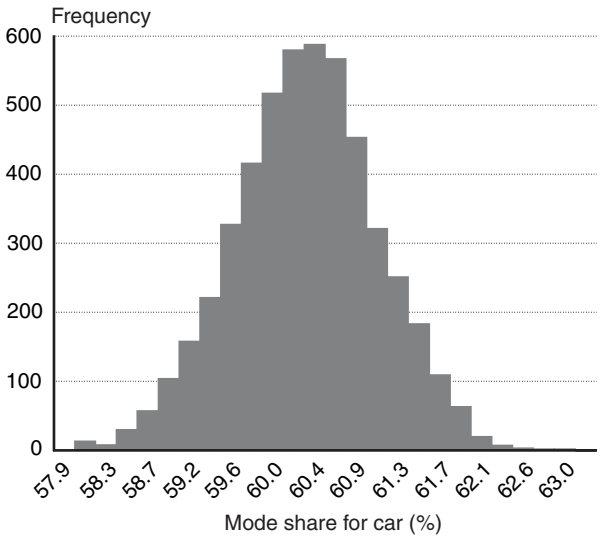
A summary spreadsheet accumulates the updated counts for all 17 sites around the cordon to produce an overall estimate of the mode share.

The process of generating repeated measurement errors and daily variations was achieved with the aid of a simple Visual Basic macro and the resultant mode shares recorded and graphed. Figure 2 shows the distribution of the mode share for cars after 5,000 such samples were conducted, which took less than 5 minutes to calculate on a 2GHz desktop PC.

The distribution has a mean of 60.3% and a standard deviation of 0.71%. The distribution appears normal with an estimated skewness of 0.01

[1] Available from http://www.stephenclark.clara.net.

FIGURE 2 Distribution of Car Mode Share



and an (adjusted) kurtosis of 0.06, both of which are close to the values expected for a normal distribution. It is, therefore, possible to estimate a 95% confidence interval for the car mode share between 58.9% and 61.7%. Similar confidence intervals can be calculated for the other modes. It should be noted that the resultant normal shape of this mode share distribution does not depend on the normality of the underlying sampling distribution; if a uniform sampling distribution is used, the same shape results, albeit, with a different spread.

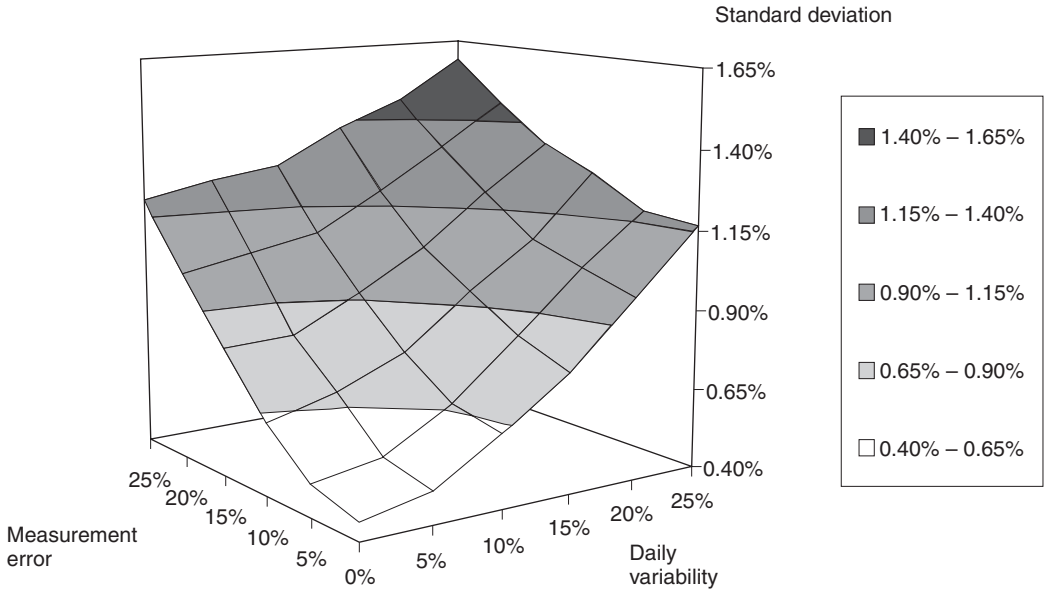
SENSITIVITY ANALYSIS

The measurement errors used here could be improved on if further resources were devoted to data collection. As an illustration of this possibility, the question is posed as to what degree of improvement would result from a halving in the mode-specific error with which single-occupant cars are counted and classified, from 10% down to 5%. When the Monte Carlo simulation model is re-run with this new error range, the interval reduces only slightly to between 59.0% and 61.6%.

A wider view of how sensitive the measure of spread in the mode share of car is can be obtained by graphing the standard deviation for a series of values for one or more of the assumed ranges. Figure 3 shows how the standard deviation of the single-occupant car mode share changes as the single-occupant car-specific measurement error changes from 0% to 25% and the daily variation in single car occupants changes from 0% to 25%. All other ranges stay at their default values.

As expected, the standard deviation increases as the ranges of variation increase. Even at a 0% value for both ranges, variation remains in the mode share for cars. This is due to the fact that the other modes are still varying at their old levels, and, since we are

FIGURE 3 Estimated Single-Car Occupancy Mode Share Standard Error
(sensitivity to changes in measurement error and daily variability)



dealing with a share, their variability will also impact on the variability of single-occupant travel by car.

SURVEY IMPLICATIONS

Information on the degree of variability of the car mode share statistic allows us to compute the minimum sample sizes required to reliably detect a specified level of change. Using the following equation for sample size estimation (Ortuzar and Willumsen 1994):

$$n' = z^2_{1-\frac{\alpha}{2}}\left(\frac{s^2}{\delta^2}\right)$$

where

n' is the required sample size,

$z^2_{1-\frac{\alpha}{2}}$ is the critical value of a α % standard normal distribution,

s is the estimated standard deviation of the measured quantity, and

δ is the minimum required change to detect,

and an example of an absolute one percentage point change as the target, the estimated sample size is:

$$n' = 1.96^2\left(\frac{0.71^2}{1.00^2}\right) = 1.94$$

which suggests that a sample size of two survey days is required to be 95% sure that an observed change of at least 1% in the average mode share for cars is significant. Table 2 shows the required sample sizes for a range of these changes for each of the three main modes of travel, using each of the Monte Carlo-derived estimates of the mode's standard deviation.

TABLE 2 Minimum Sample Size Requirements to Detect a Range of Target Reductions

	Mode of travel		
Target reduction	**Car**	**Bus**	**Rail**
0.5%	8	9	1
1.0%	2	3	1
1.5%	1	1	1
2.0%	1	1	1
2.5%	1	1	1
3.0%	1	1	1

BOOTSTRAP ESTIMATION

The technique of bootstrap estimation falls within the resampling family of techniques (Efron 1982; Efron and Tibshirani 1993). It is particularly useful when no simple expression is available to compute the summary statistics for a measure or only a limited sample size is available. The process essentially involves taking repeated subsamples from a larger sample (with or without replacement) and calculating the statistic of importance based on this subsample. The distribution of these subsample statistics is then used to infer information about the population as a whole.

The bootstrap technique has had some application within the transport field. Rilett et al. (1999) used the technique to estimate the variance of freeway travel time forecasts derived from an artificial neural network. This allowed predictions to be made of future confidence intervals for journey times along a freeway and then used as input to Advanced Traveler Information Systems. A study by Brundell-Freij (2000) focuses on assessing the accuracy in the estimates produced by complex transport models. This study used both Monte Carlo simulation and bootstrap techniques to show how different kinds of variation in the input data affect the quality of the final model estimates. The study suggests that these variations can be a large but unknown feature of transport models. Hjorth (2002) used the bootstrap technique to estimate the covariance structure of traffic counts conducted at pairs of sites. This information was then used to construct route flow proportions and probabilities.

Here we are interested in using the bootstrap technique to obtain estimates of the mode share confidence intervals from a limited number of surveys (DiCiccio and Efron 1996; Wood 2004). If we have a count of the traffic entering the city center on a limited number of days at each site, it would be possible to choose, at random, one day from each site and add them together to arrive at an estimate of the total volume of traffic entering the city center and hence calculate mode shares. So, for example, one bootstrap draw could combine the counts from

day five at site A, day two at site B, day one at site C, and so on, while the next draw would combine counts from day four at site A, day three at site B, day one (again) at site C, and so on. A large number of these draws could be taken and the distribution and summary statistics established for either the total volume or the mode shares.

Based on the Monte Carlo simulation work described earlier, additional surveys were conducted in May 2004, so that each radial road into Leeds City Center was surveyed on four days rather than the more usual one day. This sample size allows changes as small as 0.7% in the mode share for cars to be detected reliably. To increase the representative nature of the data, the survey was designed so that each of the 34 roads would be surveyed once on 4 different weekdays (excluding Fridays).

Aggregating the survey data together to produce a mode share for traffic crossing the entire cordon involved selecting 1 survey day from the 4 possible days at each of 34 survey locations. This produced a large number of possible combination of days and sites, 4^{34}, to be precise. To make this exercise more manageable, adjacent sites were grouped together to form seven corridors (see figure 1). This decreased the number of possible combinations to $4^7 = 16{,}384$. For the bootstrap exercise, just a fraction of these combinations were used: 4,000 selected at random from over 16,000 possibilities. The bootstrap mean of the 4,000 car mode shares selected was 57.3%, much lower than the Monte Carlo mean value calculated in 2002.

Table 3 gives the estimated standard deviations for the mode shares of car, bus, rail, and walk from the Monte Carlo and bootstrap techniques. The bootstrap-estimated standard deviation for car-based trips is 0.64%, compared with the Monte Carlo estimate of 0.71%. The bootstrap deviation could be expected to be different for a number of reasons:

- The range of daily variation selected for the Monte Carlo simulation was designed to account for the variety seen throughout the year, while the bootstrap estimate was based on just the variability observed within one calendar month. If the surveys used in the bootstrap estimation were conducted throughout 2004 rather than just in May, it is likely that a wider spread of observed variation would be present and the estimated standard deviation would increase above the 0.64% value found here.
- The survey enumerators knew there would be repeated surveys at each site. This ability to cross check counts may have encouraged them to be more accurate in their counting. A more accurate and consistent set of counts would produce a smaller deviation.
- The mean share was reduced significantly: the 0.71% Monte Carlo estimate is for a car share of about 60.3%, while the bootstrap estimate is a lower share of about 57.3%.

TABLE 3 Estimated Standard Deviations for Mode Shares

Mode	Monte Carlo (2002)	Bootstrap (2004)
Car	0.71%	0.64%
Bus	0.75%	0.76%
Rail	0.33%	0.27%
Walk	0.07%	0.12%

FURTHER IDEAS

In this paper, a Monte Carlo simulation regime was established to estimate the variability in mode share for a traffic cordon survey. While the illustrative example used a specific experimental methodology to collect the data and determine the structure of the model, the simulation approach proposed is flexible enough to allow the use of data that are collected through different survey designs. Of particular note here is that no "conservation of flow" principle has been applied to the changes (i.e., changes in one mode of travel are not mirrored with compensatory changes in another) but if thought necessary, this principle could easily be incorporated in the model.

There is nearly always a value in conducting more surveys to measure the important ranges that define both measurement errors and daily variation ("*Whenever you can, count.*" Sir Francis Galton). These surveys do, however, come at a cost. The model proposed in this study can help to identify

which survey methodology has the greatest impact on the accuracy of mode share and, therefore, provide the best value for the money.

ACKNOWLEDGMENTS

The authors would like to thank their colleagues, Ken Mason and Mohammed Mahmood, for their valuable contributions to the work reported here. We would also like to thank the three anonymous referees and Dr. Susan Grant-Muller for comments on earlier drafts of this paper. The opinions and ideas expressed in this paper are those of the authors alone and should not be taken to be those of Leeds City Council or its agencies.

REFERENCES

Brundell-Freij, K. 2000. Sampling, Specification and Estimation as Sources of Inaccuracy in Complex Transport Models—Some Examples Analyzed by Monte Carlo Simulation and Bootstrap. Proceedings of Seminar F, European Transport Conference, Cambridge, England, pp. 225–237.

Cassidy, M.J., Y.T. Son, and D.V. Rosowsky. 1994. Estimating Motorist Delay at Two-Lane Highway Work Zones. *Transportation Research A* 28(5):433–444.

Castiglione, J, J. Freedman, and M. Bradley. 2003. A Systematic Investigation of Variability Due to Random Simulation Error in Activity Based Micro-Simulation Forecasting Model, paper presented at the 2003 Annual Meetings of the Transportation Research Board, Washington, DC.

Chen, A, H. Yang, H.K. Lo, and W.H. Tang. 1999. Capacity Related Reliability for Transportation Networks. *Journal of Advanced Transportation* 33(2):183–200.

____. 2002. Capacity Reliability of a Road Network: An Assessment Methodology and Numerical Results. *Transportation Research B* (36):225–252.

Dajani, J.S. and G. Gilbert. 1975. Delphic Predictions and Cross Impact Simulation. *ASCE Journal of the Urban Planning and Development Division* 10(1):49–59. May.

Department for Transport (DfT). 1994. *National Travel Survey: 1991/93*. London, England. Chapter 6, Motorcycles and Their Users, pp. 37–43.

____. 1996. *Transport Statistics Report: Cycling in Great Britain*. London, England. Chapter 2, Cycle Traffic.

____. 2003. *Design Manual for Roads and Bridges. Volume 12, Section 1, Part 1: Traffic Appraisal Manual, Annex 10.3.6*. Available at www.official-documents.co.uk/document/deps/ha/dmrb/index.htm.

DiCiccio, T.J. and B. Efron. 1996. Bootstrap Confidence Intervals. *Statistical Science* 11(3):189–228.

Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall.

Fox, K, S. Clark, R. Boddy, F. Montgomery, and M. Bell. 1998. Some Benefits of a SCOOT UTC System: An Independent Assessment by Micro-Simulation. *Traffic Engineering and Control* 38(8):484–489.

Hjorth, U. 2002. Traffic Sub-Flow Estimation and Bootstrap Analysis from Filtered Counts. *Transportation Research B* 36(4):345–359.

Knusel, L. 1998. On the Accuracy of Statistical Distributions in Microsoft Excel 97. *Computational Statistics and Data Analysis* 26:375–377. See also http://www.stat.uni-muenchen.de/~knuesel.

Kreihich, V. 1979. Modelling of Car Availability Modal Split and Trip Distribution by Monte Carlo Simulation: A Short Way to Integrated Models. *Transportation* 8(2):153–166.

Lam, W.H.K. and G. Xu. 1999. A Traffic Flow Simulator for Network Reliability. *Journal of Advanced Transportation* 22(2):159–182.

Ortuzar, J.D. and L.G. Willumsen. 1994. *Modelling Transport*, 2nd ed. New York, NY: Wiley.

Philips, G. 1979. Accuracy of Annual Traffic Flow Estimates from Short Period Count. *TRRL Supplementary Report 514*. Crowthorne, Berkshire, UK: Transport Research Laboratory.

Pothisiri, T. and K.D. Hjelmstad. 2003. Structural Damage Detection and Assessment from Modal Response. *Journal of Engineering Mechanics* 129(2):135–145.

Rilett, L.R., D. Park, and B. Gajewski. 1999. Estimating Confidence Interval for Freeway Corridor Travel Time Forecasts. Proceedings of the 6th World Congress on Intelligent Transport Systems, Toronto, Canada.

Royal Statistical Society. 2005. Performance Indicators: Good, Bad and Ugly—The Report of the Working Party on Performance Monitoring in the Public Services, chaired by Professor S.M. Bird, submitted October 23rd, 2003. *Journal of the Royal Statistical Society A* 168(1):1–27. Available from www.rss.org.uk.

Strategic Rail Authority. 2002. *On Track*. London, England. Available at www.sra.gov.uk.

Tarko, AP. 2000. Random Queues in Signalized Road Networks. *Transportation Science* 38(4):415–425.

U.S. Department of Transportation (USDOT), Bureau of Transportation Statistics. 2003. Guide to Good Statistical Practice in the Transportation Field. Available at www.bts.gov/publications/guide_to_good_statistical_practice_in_the_transportation_field/.

Veldhuisen, J., H. Timmermans, and L. Kapoen. 2000. Microsimulation Model of Activity-Travel Patterns and Traffic Flow: Specification, Validation Tests and Monte Carlo Error. *Transportation Research Record* 1706:126–135.

West Yorkshire Passenger Transport Authority (WYPTA). 2003. *West Yorkshire Local Transport Plan: Annual Progress Report, 2002–2003*. West Yorkshire, UK.

Williamson, D.G., M. Yao, and J. McFadden. 2002. Monte Carlo Simulation in Sampling Techniques of Traffic Data Collection. *Transportation Research Record* 1804.

Wood, M. 2004. Statistical Inference Using Bootstrap Confidence Intervals. *Significance* 1(4):180–182.

Zhao, Y.G. and A.H.S. Ang. 2003. System Reliability Assessment by Method of Moments. *Journal of Structural Engineering* 129(10):1341–1349.

APPENDIX

Distributional Alternative

A question arises as to whether any results from the statistical literature will allow inferences to be made concerning the distribution of a share:

$$S_1 = \frac{X_1}{X_1 + X_2}$$

where

S_1 is the share of mode one,

X_1 is the volume of mode one,

X_2 is the volume of all other modes, and

the distribution of both X_1 and X_2 are known.

The γ distribution is one form of distribution that is quite flexible in the range of distributional shapes that it can represent. Another feature of the γ distribution is that if X_1 and X_2 are γ distributed random variables with $X_1 \sim \gamma(\alpha_1, \beta_1)$ and $X_2 \sim \gamma(\alpha_2, \beta_1)$, then S_1 has a β distribution, $S_1 \sim \beta(\alpha_1, \alpha_2)$. Critical to the use of this result is that both the γ distributions have similar values for the scale parameter, β_1.

In the context of the data used in this study, each mode of travel will have a different scale; the volume of travel by car is greater than that by bus and rail. This suggests that the β distribution approach to modeling the distribution of mode share may not be realistic.

Analysis of Work Zone Gaps and Rear-End Collision Probability

DAZHI SUN[1,*]
RAHIM F. BENEKOHAL[2]

[1] Texas Transportation Institute
Texas A&M University System
1100 NW Loop 410, Suite 400
San Antonio, TX 78213

[2] Newmark Civil Engineering Laboratory
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801

ABSTRACT

This paper studies platooning and headway/gap characteristics of traffic flow in highway short-term and long-term work zones under various car-following patterns. The relationship between traffic volume and the percentage of vehicles in platoons is developed, along with some statistical models for platoon size and headway/gap size distribution. An in-depth analysis of data reveals that vehicles in work zones with higher speed limits maintain shorter car-following time gaps than those in work zones with lower speed limits, even though more time is needed to stop a faster vehicle. This unusual combination of higher speeds and shorter car-following time gaps in work zones may contribute to the high proportion of rear-end collisions among all work zone-related accidents. This paper also presents a new method for evaluating rear-end collision potential, including the probability and the number of vehicles involved in rear-end collisions, by analyzing platoon and gap characteristics for locations without crash records during a construction period.

Email addresses:
*Corresponding author—d-sun@tamu.edu
R.F. Benekohal—rbenekoh@uiuc.edu

KEYWORDS: Car-following patterns, rear-end collisions, platoons, work zones.

INTRODUCTION

The number of fatalities in motor vehicle crashes in work zones has risen from 693 in 1997 to 1,181 in 2002 in the United States. Rear-end crashes are one of the most common kinds of work zone crashes and account for more than 30% of all crashes in work zones. By investigating gap characteristics of platooning vehicles in work zones, researchers may be better able to evaluate the risk of rear-end collisions for vehicles in platoons and understand driver behavior in this situation.

In the past, numerous investigations have looked at the headway characteristics of highway traffic, but limited studies exist on gap characteristics, particularly in work zones. Wasielewski (1979) reported on headway characteristics of highway traffic and concluded that the headway distribution was independent of the traffic volume. Luttinen (1992) studied the independence of consecutive headways using geometric bunch size distribution for two-lane highways in Finland. May (1990), Griffiths and Hunt (1991), Mei and Bullen (1993), and Akcelik and Chung (1994) applied different models to study the distribution of time headways in highway and urban traffic flow conditions.

Most existing headway studies investigated normal traffic flow conditions rather than work zone conditions. Work zone traffic flow has different characteristics due to lower speed limits, work activities, lane closures and channelization plans, and other geometric and traffic factors. The presence of queues and platoons of vehicles is more prevalent in work zones than on regular sections of highway. Therefore, it is necessary to explore drivers' car-following behaviors while they are traveling through construction areas.

Benekohal and Sadeghhosseini (1991) and Sadeghhosseini and Benekohal (1995) investigated the platooning and time headway characteristics of highway work zones. They examined the effects of traffic volume on distribution of time headways and on the percentage of platooning vehicles. Although headway characteristics have been used widely in these analyses, gap characteristics provide a better measure of car-following behaviors and safety-related issues. This paper focuses on quantifying the variations of time headways or gaps for different car-following patterns and work zone types and the relationship between car-following characteristics and the accident risks/safety performances for work zones.

To address these issues, we analyzed field data from 11 work zone sites. In the case study, we proposed and implemented a new gap-analysis-based safety performance evaluation methodology for work zones. Work zone safety performance is difficult to evaluate due to the lack of reliable work zone crash data. This new method provides an alternative approach to evaluating accident risk by analyzing crash predisposition under nonaccident situations.

DATA COLLECTION AND REDUCTION

Field data were collected at 11 work zones sites on Interstate highways in Illinois. Three of the sites investigated were short-term work zones and eight were long-term. In this study, a short-term work zone is defined as a construction or maintenance site that lasted less than a few days and the closed lane was delineated using cones, barrels, or barricades (but not barriers). A long-term work zone is defined as a construction or maintenance site that lasted more than a few days and the closed lane was delineated using concrete barriers. The short-term work zones studied had a posted speed limit of 45 mph, while all the long-term work zones had a posted speed limit of 55 mph.

All 11 sites had two lanes in each direction; one lane was closed due to construction and the other was open. A video camera was used to capture the times at which a vehicle passed over two specific markers placed at a fixed distance. The distance between the markers was about 250 feet, but varied for different sites. Data were collected over a time period of two to four hours for each site, depending on the traffic conditions. Initially, the videotapes were time coded. The time coding of the videotapes allowed us to read the travel time more accurately. Time headway, time gap, space, speed, and volume data were obtained from the tapes. The headway for each vehicle was computed based on the time measured at marker 2 (the marker closest to the camera) when the front bumper of a vehicle passed over the line of sight between the camera and the marker. The time headway for the following vehicle

was the time difference between the passing of the front bumper of the leading and following vehicles over the line of sight. The gap is the time difference between the passing of the rear bumper of the leading vehicle and the front bumper of the following vehicle over the line of sight. The time measurements are accurate to within 1/30 seconds.

Vehicles were classified into platooning or nonplatooning based on their speed and spacing. A platoon is a group of vehicles traveling close to one other with short headways. The literature gives four definitions of platoons based on either time or space headway, a combination of time headway and speed, or a combination of space headway and speed. Different thresholds in time headway, ranging from 2.5 to 6 seconds, have been used in the past to identify platooning vehicles. Keller (1976) and Benekohal and Sadeghhosseini (1991), for example, used five seconds as the threshold of time headway to separate platooning vehicles from the traffic flow. Sumner and Baguley (1978) used a gap of two seconds and speed differences of less than 10% as the platooning threshold. Horban (1983) suggested four seconds for the time headway threshold in a level-of-service study. Our analysis focuses only on vehicles in platoons and employs data on over 15,000 vehicles to investigate platoon and gap characteristics in work zones.

PLATOONING AND GAP CHARACTERISTICS IN WORK ZONES

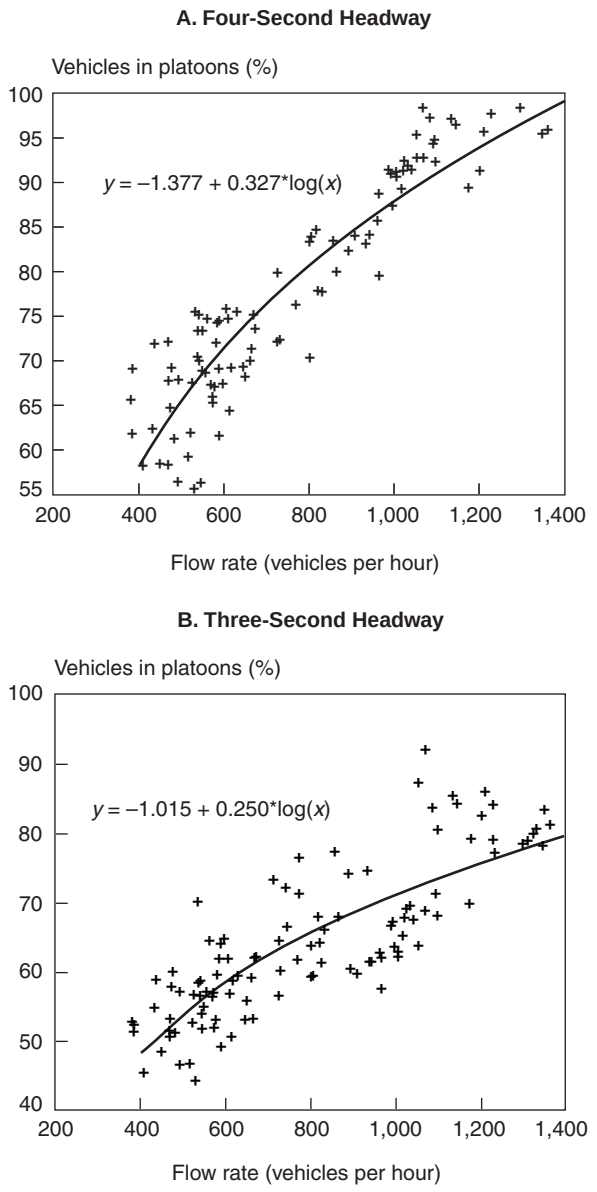
Analyses of platooning characteristics, including the percentage of vehicles in platoons and platoon size distributions, are discussed below. Then we analyze gap characteristics for platooning vehicles to determine the effect of different car-following patterns and work zone types on gap size, determine the effect of platoon size on gap size, and establish a gap size distribution.

Platoon Analysis

Percentage of Platooning vs. Volume

Figure 1 shows how the percentage of vehicles in a platoon varies as the traffic volume changes. Two different platooning criteria were applied to examine the effect of the threshold, using four-second and three-second headways.

FIGURE 1 Percentage of Traffic in Platoons vs. Flow Rate



We constructed 96 sets of data from the initial observation with each corresponding to a 15-minute period of observation. The average traffic volume and percentage of vehicles in the platoon were computed for each set (plotted in figure 1). Figure 1A shows that the percentage of vehicles in the platoon varied from 55% to 75% under low volume conditions (less than 600 vehicles per hour (vph)). The percentage increased to 95% when traffic volume reached about 1,200 vph.

Figure 1B uses a three-second headway as the platooning criteria, which presents a lower percent-

age of platooning than the four-second headway seen in figure 1A, yet there is still about 43% to 70% platooning at low volume conditions. Only 80% of vehicles were identified as platooning, with a volume of 1,400 vph using the three-second headway criterion. As the volume rises, the percentage increase in platooning slows, indicating that the three-second criterion does not accurately reflect the reality of platooning. Therefore, using a four-second headway as the platooning criterion is more appropriate and provides a greater margin of safety than the three-second headway. The methodology discussed in this paper is independent of the headway threshold; only the numerical values change. As a result, we define a platoon as a group of vehicles separated by a time headway of no longer than four seconds. The remaining discussion of platooning vehicles is based on this definition.

We examined the relationship between traffic volume and percentage of vehicles platooning and found that a logarithmic function fit the data better than other forms. As volume increases, the percentage of platooning vehicles increases and ultimately all vehicles will be considered as part of a platoon. The logarithmic function is expressed as

$$y = -1.377 + 0.327\,ln(x) \qquad (1)$$

where

x is the hourly flow rate (vph), $400 \leq x \leq 1400$, and

y is the percentage platooning (number of vehicles in a platoon/volume).

Platoon Size Distribution

The type of vehicle leading a platoon and the number of vehicles in each platoon were determined. Then platoons were classified into two groups: truck-leading platoons and nontruck-leading platoons. Truck-leading platoons have a large truck at the front of the platoon. Platoons with the same number of vehicles were further grouped into platoon size groups. The relative frequencies of the platoon size groups are shown in figures 2A and 2B for short-term and long-term work zones, respectively. These figures show that 70% to 80% of platoons had only two or three vehicles.

Difference models were evaluated to see which of the observed frequencies fit better. The goodness of fit was determined in terms of the root mean square (RMS) error. As a result, a shifted negative exponential function best fit the model in terms of having the least RMS error. For short-term work zones, equations (2) and (3) describe the relationship between platoon size (x) and the percentage of vehicles belonging to that platoon size $p(x)$. Equation (2) represents truck-leading platoons, and equation (3) represents nontruck-leading platoons.

$$p(x) = \frac{1}{1.6899}\exp(-(x - 1.5406)/1.6899) \qquad (2)$$

$$p(x) = \frac{1}{1.1993}\exp(-(x - 1.5075)/1.1993) \qquad (3)$$

Similar relationships were found for the long-term work zones as expressed by equation (4) for truck-leading platoons and equation (5) for nontruck leading platoons.

$$p(x) = \frac{1}{1.6244}\exp(-(x - 1.4853)/1.6244) \qquad (4)$$

$$p(x) = \frac{1}{1.3}\exp(-(x - 1.4916)/1.3) \qquad (5)$$

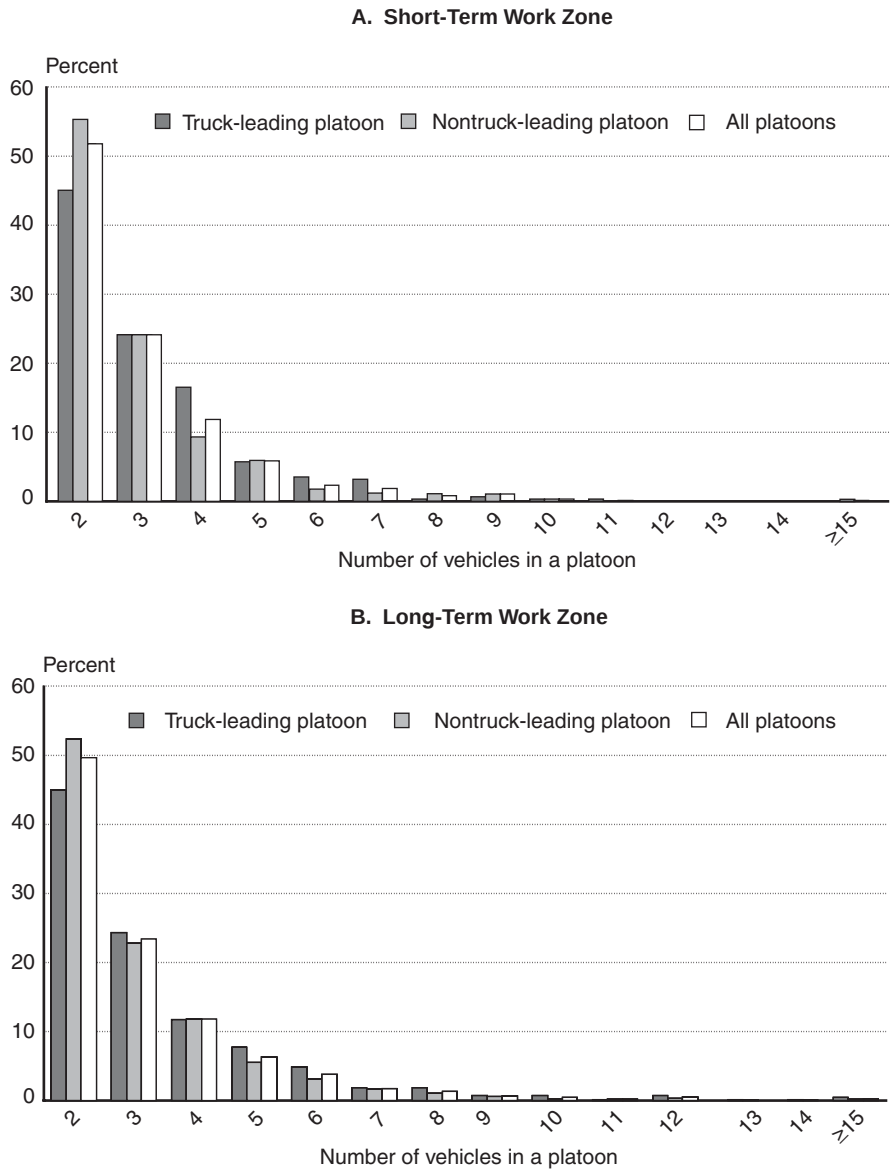
Table 1 shows the platoon size frequency, average headway, and average gap for short-term and long-term work zones. The table shows that more small platoons are led by cars. For example, the relative frequency of nontruck-leading two-vehicle platoons is 0.55 for short-term work zones and 0.52 for long-term work zones, while the relative frequency of truck-leading two-vehicle platoons is only 0.45 for both short-term and long-term work zones.

To determine if the platoon size distributions of the four cases shown in equations (2) through (5) differ significantly, the two most commonly-used two-independent-samples tests—the Mann-Whitney U test and the Kolmogorov-Smirnov z test in SPSS—were applied for the following combinations:

1. nontruck-leading platoon vs. truck-leading platoon in short-term work zones,
2. nontruck-leading platoon vs. truck-leading platoon in long-term work zones,
3. truck-leading platoon in short-term work zones vs. in long-term work zones, and
4. nontruck-leading platoon-in short-term work zones vs. in long-term work zones.

The results of these tests show that the significance is less than 0.05 for combinations 1 and 2;

FIGURE 2 Platoon Size Distribution



therefore, the hypothesis H_0: *the two distributions are identical* is rejected. For combinations 3 and 4, we cannot reject the null hypothesis because the *p*-value is considerably above 0.05. This indicates that the type of lead vehicle has a significant impact on the platoon size distribution, while the type of work zone does not.

Gap Analysis

To examine car-following safety in work zones, we analyzed the time gap instead of the time headway, because the time gap represents the actual time available for the following car to avoid a rear-end collision. We also studied the effects of the combination of the leader and follower on gap size as well as the effects of the leader of a platoon on platoon size. We determined the gap size distributions and used them to predict the probability of rear-end collisions.

Effect of Car-Following Patterns

Will the gap size be affected by the combination of leader and follower? For instance, a car following a truck may tend to keep a larger time gap than a car following a car. Also, the probability of a rear-end collision may depend on the brake features of the following vehicle and the time gap available to it. To answer our question, we studied the relationship

TABLE 1 Work Zone Platoon Characteristics

A. Short-Term Work Zone

	Truck-leading platoon			Nontruck-leading platoon		
Platoon size	Relative frequency	Headway (seconds)	Gap (seconds)	Relative frequency	Headway (seconds)	Gap (seconds)
2	0.451	2.819	1.991	0.553	2.045	1.680
3	0.241	2.593	1.857	0.241	2.098	1.699
4	0.165	2.398	1.751	0.093	2.167	1.735
5	0.057	2.201	1.627	0.059	2.129	1.693
6	0.035	2.109	1.577	0.017	2.333	1.852
7	0.032	2.263	1.733	0.012	2.109	1.685
8	0.003	2.344	1.510	0.010	2.212	1.740
9	0.006	2.394	1.575	0.010	2.243	1.741
10	0.003	3.348	2.732	0.003	2.598	2.082
11	0.003	2.321	1.714	—	—	—
12	—	—	—	—	—	—
13	—	—	—	—	—	—
14	—	—	—	—	—	—
15	0.003	2.267	1.649	—	—	—
≥16	—	—	—	—	—	—
Average	3.20 (size)	2.611	1.872	2.89 (size)	2.085	1.696

B. Long-Term Work Zone

	Truck-leading platoon			Nontruck-leading platoon		
Platoon size	Relative frequency	Headway (seconds)	Gap (seconds)	Relative frequency	Headway (seconds)	Gap (seconds)
2	0.450	2.356	1.739	0.524	1.719	1.416
3	0.243	2.194	1.575	0.228	1.756	1.439
4	0.117	2.017	1.447	0.118	1.853	1.514
5	0.077	1.974	1.482	0.055	1.828	1.467
6	0.049	1.856	1.371	0.031	1.844	1.438
7	0.019	1.915	1.427	0.016	1.705	1.303
8	0.019	2.077	1.529	0.010	1.533	1.216
9	0.007	2.074	1.448	0.006	2.242	1.667
10	0.007	2.194	1.755	0.003	1.690	1.304
11	0.001	1.560	1.293	0.002	2.135	1.726
12	0.007	1.991	1.532	0.003	1.956	1.546
13	—	—	—	0.001	2.305	1.864
14	—	—	—	0.001	1.933	1.660
15	—	—	—	—	—	—
16	—	—	—	—	—	—
17	0.001	1.588	1.246	0.001	2.185	1.608
18	0.001	1.566	1.249	0.001	2.155	1.758
19	—	—	—	—	—	—
≥20	0.001	2.084	1.744	—	—	—
Average	3.38 (size)	2.200	1.612	3.06 (size)	1.757	1.436

between car-following patterns and time gaps. We investigated the average gaps under different car-following patterns. The four possible car-following patterns analyzed are: car-car, car-truck, truck-car, truck-truck (leader-follower).

Table 2A and 2B detail the mean gap, the mean headway, and the frequency of four car-following patterns in short-term and long-term work zones. These tables show that the average gap is the shortest when a car follows another car. The next shortest gap is when a car follows a truck. The gap is longer when a truck follows a car or a truck. When a truck follows a car or a truck, the gap sizes are not as different as when a car follows a car or a truck. This would appear to indicate that car drivers are more sensitive to what type of vehicles they are following than truck drivers. The table also shows that the gaps in short-term work zones are longer than the gaps in long-term work zones for the same combination of leader and follower. Thus, it is important to know which one of these differences is statistically significant.

This study presents only the aggregate analysis of the mean headway/gap size for different car-following patterns. The data appear to show that each vehicle's car-following behavior is determined primarily by the vehicle directly in front of it, particularly in highway work zones with only one lane open. Of course, other vehicles may also impact driving behavior. The interdependence of different car-following patterns is a complicated problem that would benefit from a more extensive dataset and disaggregate analysis on numerous combinations. Our current dataset does not support this analysis, which we plan to address in a future study.

A "two-sample means *z*-test" was conducted to evaluate the difference between the mean time gap under different car-following patterns. We compared the gaps in short-term and long-term work zones for the same car-following pattern using a 95% confidence level to test the hypotheses. The results of the tests of 16 different hypotheses are presented in table 3. The *z*-test shows no significant difference in the time gap between truck-truck and car-truck following patterns in either short-term or long-term work zones. This further supports our findings in table 2. All the other null hypotheses were rejected indicating that, with a 95% confidence level, there is a significant difference in the time gap.

TABLE 2 Gap Size of Platooning Vehicles in Work Zones

A. Short-Term Work Zone

Pattern	Frequency (vph)	Mean gap (seconds)	Mean headway (seconds)
Car-car	1,087	1.610	1.986
Car-truck	209	2.030	2.429
Truck-car	374	1.805	2.672
Truck-truck	157	2.021	2.960

B. Long-Term Work Zones

Pattern	Frequency (vph)	Mean gap (seconds)	Mean headway (seconds)
Car-car	1,392	1.384	1.736
Car-truck	311	1.824	2.164
Truck-car	603	1.645	2.285
Truck-truck	384	1.865	2.578

Key: vph = vehicles per hour.
Note: The measured average speeds of platooning vehicles were 39.80 mph in short-term work zones and 50.78 mph in long-term work zones.

Safety Paradox

The analysis in the previous section shows significantly smaller time gaps for all car-following patterns in long-term work zones with a speed limit of 55 mph compared with gaps in short-term work zones with a speed limit of 45 mph. Although the measured average speeds in the two types of work zones that post the same speed limit vary slightly, the average speeds of nonplatoon and platooning vehicles in long-term work zones tend to be significantly higher than those in short-term work zone. For example, the measured average speed for a nonplatoon vehicle was 42.39 mph, with platooning vehicles averaging 39.80 mph in short-term work zones. In long-term work zones, the measured average speed was 53.29 mph for nonplatoon vehicles, and 50.78 mph for platooning vehicles. Table 2 shows that for a car-car pattern, the average time gap for vehicles with an average speed of 39.0 mph was 1.610 seconds, while the gap for vehicles with an average speed of 50.78 mph was 1.384 seconds.

TABLE 3 Results of *z*-test for Time Gap

	Hypothesis	*p* value (one tail)	Reject the null hypothesis?
Short term	$H_0: g_{c-c}^{st} = g_{t-c}^{st}; H_1: g_{c-c}^{st} < g_{t-c}^{st}$	2.53107E–08	Yes
	$H_0: g_{c-c}^{st} = g_{c-t}^{st}; H_1: g_{c-c}^{st} < g_{c-t}^{st}$	8.50431E–14	Yes
	$H_0: g_{c-t}^{st} = g_{t-t}^{st}; H_1: g_{c-t}^{st} < g_{t-t}^{st}$	0.454587	No
	$H_0: g_{t-c}^{st} = g_{t-t}^{st}; H_1: g_{t-c}^{st} < g_{t-t}^{st}$	0.000105	Yes
	$H_0: g_{c-c}^{st} = g_{t-t}^{st}; H_1: g_{c-c}^{st} < g_{t-t}^{st}$	8.50431E–14	Yes
	$H_0: g_{c-t}^{st} = g_{t-c}^{st}; H_1: g_{c-t}^{st} < g_{t-c}^{st}$	0.000236	Yes
Long term	$H_0: g_{c-c}^{lt} = g_{t-c}^{lt}; H_1: g_{c-c}^{lt} < g_{t-c}^{lt}$	0	Yes
	$H_0: g_{c-c}^{lt} = g_{c-t}^{lt}; H_1: g_{c-c}^{lt} < g_{c-t}^{lt}$	0	Yes
	$H_0: g_{c-t}^{lt} = g_{t-t}^{lt}; H_1: g_{c-t}^{lt} < g_{t-t}^{lt}$	0.212811	No
	$H_0: g_{t-c}^{lt} = g_{t-t}^{lt}; H_1: g_{t-c}^{lt} < g_{t-t}^{lt}$	7.38E–09	Yes
	$H_0: g_{c-c}^{lt} = g_{t-t}^{lt}; H_1: g_{c-c}^{lt} < g_{t-t}^{lt}$	0	Yes
	$H_0: g_{c-t}^{lt} = g_{t-c}^{lt}; H_1: g_{c-t}^{st} < g_{t-c}^{lt}$	0.000126	Yes
Short term vs. long term	$H_0: g_{c-c}^{st} = g_{c-c}^{lt}; H_1: g_{c-c}^{st} > g_{c-c}^{lt}$	0	Yes
	$H_0: g_{c-t}^{st} = g_{c-t}^{lt}; H_1: g_{c-t}^{st} > g_{c-t}^{lt}$	0.00107	Yes
	$H_0: g_{t-c}^{st} = g_{t-c}^{lt}; H_1: g_{t-c}^{st} > g_{t-c}^{lt}$	1.28E–05	Yes
	$H_0: g_{t-t}^{st} = g_{t-t}^{lt}; H_1: g_{t-t}^{st} > g_{t-t}^{lt}$	0.003949	Yes

Key:

g_{c-c}^{st} is the average gap for a car-followed-by-a-car pattern in a short-term work zone

g_{c-t}^{st} is the average gap for a car-followed-by-a-truck pattern in a short-term work zone

g_{t-c}^{st} is the average gap for a truck-followed-by-a-car pattern in a short-term work zone

g_{t-t}^{st} is the average gap for a truck-followed-by-a-truck pattern in a short-term work zone

g_{c-c}^{lt} is the average gap for a car-followed-by-a-car pattern in a long-term work zone

g_{c-t}^{lt} is the average gap for a car-followed-by-a-truck pattern in a long-term work zone

g_{t-c}^{lt} is the average gap for a truck-followed-by-a-car pattern in a long-term work zone

g_{t-t}^{lt} is the average gap for a truck-followed-by-a-truck pattern in a long-term work zone

Note: Level of significance (α) = 5%.

That is, the time gap decreased by 14% at an average speed increase of 22%.

The above numbers indicate a safety paradox: even though people know they need a greater safety buffer when they are driving at a higher speed, our data show that the actual time gap they maintained in a higher speed work zone was significantly shorter than that maintained in a lower speed work zone. The problem may be that people do not recognize this reduction in the safety buffer in terms of the time gap. Rather, people may judge their safety buffer in terms of the space gap. For example, for the car-car following pattern, the average space gap in a long-term work zone (103 feet) was still greater

than that for a short-term work zone (94 feet), but the real available time gap for the vehicle traveling in the long-term work zone was reduced by 22%.

The Effect of Platoon Size

The number of vehicles involved in a rear-end collision depends on the size of the platoon. The discussion here is limited to platoons of seven vehicles or less, because we had only a few observations for larger platoons. The variation in the time gap with respect to platoon size is illustrated in figure 3.

Figure 3 shows that in short-term work zones, gap size generally declines for truck-leading platoons as the platoon size increases, except for the slight upturn as the platoon size increases from six to seven. The declining trend also exists for truck-leading platoons in long-term work zones, although the trend is not as clear as in short-term work zones. For car-leading platoons, the average gap size does not seem to depend on platoon size. For platoons consisting of two or three vehicles, we found the average gap size of truck-leading platoons was greater than that of car-leading platoons in short-term and long-term work zones. For platoons with four or more vehicles, the gap sizes of car-leading platoons were, in general, greater than those of truck-leading platoons in short-term work zones, whereas no significant difference was observed for long-term work zones.

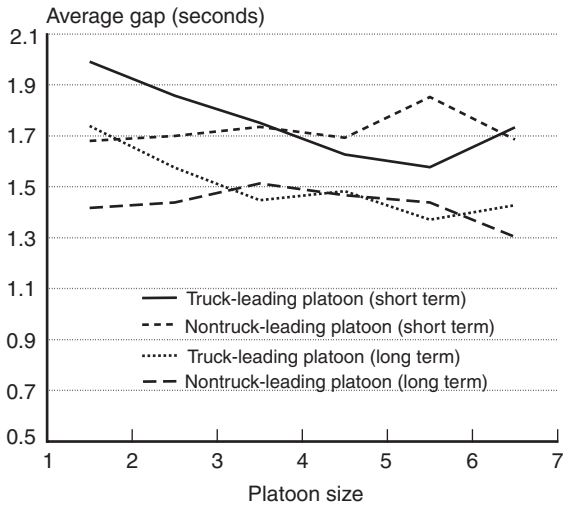
The above findings make it clear that as the platoon size of truck-leading platoons increases, drivers tend to follow more closely and thus become more vulnerable to a rear-end crash.

Gap Size Distribution

Gap size can better measure the probability of a rear-end crash than headway. The observed mean gap size was 1.73 in short-term work zones and 1.49 in long-term work zones. The median gap size was 1.68 and 1.42 for short-term and long-term work zones, respectively. In order to find the gap size distribution, we grouped the observed gaps into intervals of 0.25 seconds, starting from 0 seconds and ending at 4 seconds. Figure 4 is a relative frequency histogram of gap sizes for short-term and long-term work zones.

To find equations for the gap size distributions, the 10 widely applied mathematical models were

FIGURE 3 Effect of Platoon Size on Gap



assessed using BestFit with respect to RMS errors. Weibull offers the best fitted function followed by the BetaGeneral function. Gamma, InvGauss, and log-normal ranked as the third, fourth, and fifth best-fitted model, respectively. The appendix presents a brief introduction to these five models to show the density function as well as the model parameters.

The probability distribution function (PDF) of Weibull is

$$f(x) = \frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{(\gamma-1)} \exp(-((x-\mu)/\alpha)^{\gamma}) \qquad x \geq \mu;\gamma,\alpha > 0 \tag{6}$$

where

γ is the shape parameter; γ = 2.396 for short-term work zone and γ = 2.051 for long-term work zone;

μ is the location parameter; μ = 0.215 for short-term work zone and μ = 0.286 for long-term work zone; and

α is the scale parameter; α = 1.716 for short-term work zone and α = 1.373 for long-term work zone.

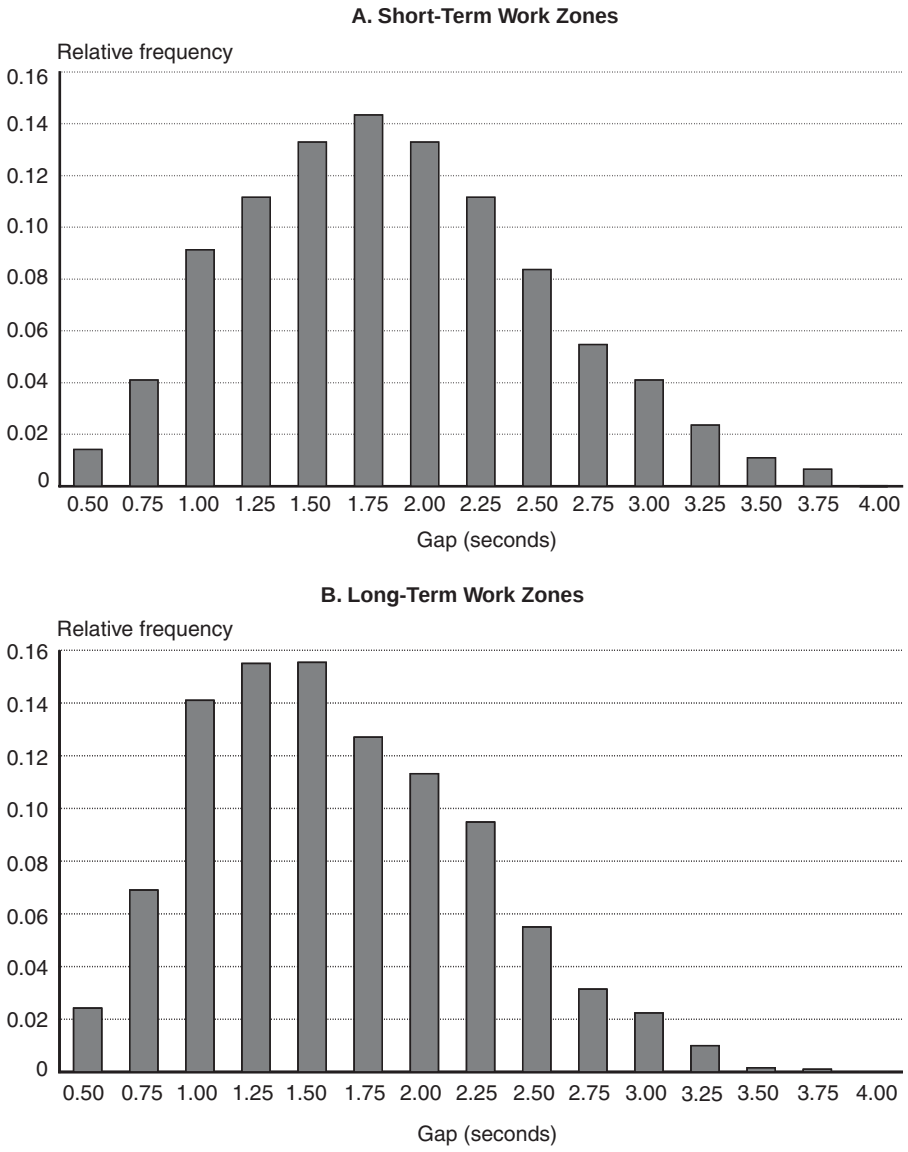
The cumulative distribution function (CDF) of the three-parameter Weibull is

$$F(x) = 1 - \exp\left(-\left(\frac{x-\mu}{\alpha}\right)^{\gamma}\right) \qquad \text{for } x \geq \mu \tag{7}$$

For short-term work zones, the PDF and CDF of the resulting Weibull models are

$$f(x) = \frac{2.396}{1.716}\left(\frac{x-0.215}{1.716}\right)^{(1.396)} \bullet \exp(-((x-0.215)/1.716)^{2.396}) \qquad x \geq 0.215 \tag{8}$$

FIGURE 4 Gap Distribution



$$F(x) = 1 - \exp\left(-\left(\frac{x - 0.215}{1.716}\right)^{2.396}\right) \quad \text{for } x \geq 0.215 \qquad (9)$$

For long-term work zones, the PDF and CDF of fitted Weibull models are

$$f(x) = \frac{2.051}{1.373}\left(\frac{x - 0.286}{1.373}\right)^{1.051} \bullet \exp(-((x - 0.286)/1.373)^{2.051}) \qquad x \geq 0.289 \qquad (10)$$

$$F(x) = 1 - \exp\left(-\left(\frac{x - 0.286}{1.373}\right)^{2.051}\right) \qquad \text{for } x \geq 0.286 \qquad (11)$$

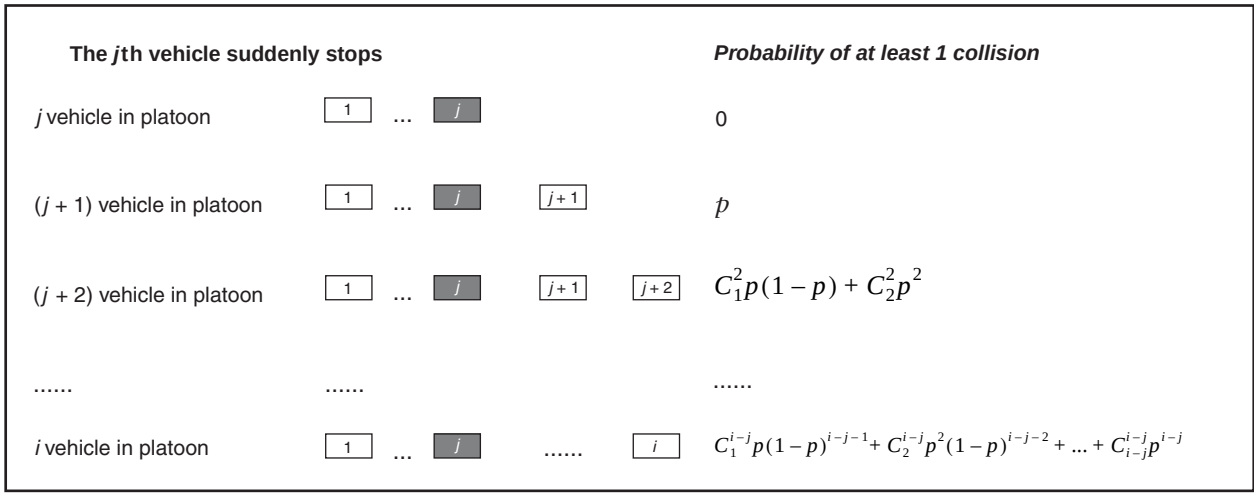
ESTIMATION OF SAFETY PERFORMANCE USING GAP AND PLATOONING ANALYSIS

Nearly one in three work zone crashes are rear-end crashes. Rear-end crashes in a work zone can occur when a vehicle suddenly decelerates due to an unexpected situation. The next section looks at the probability of one or more collisions as a vehicle suddenly decelerates.

Probability of at Least One Rear-End Collision

The risk of a rear-end collision is relative to the time gap, platoon size, and position of the problem vehicle in a platoon. In this section, we develop a model to compute the probability of rear-end collisions when a platooning vehicle suddenly decelerates.

FIGURE 5 Probability of a Rear-End Collision



The probability, (p) that a platooning vehicle has a less than critical gap can be obtained via the gap relative frequency histogram (figure 4) or the fitted Weibull CDF (equations (9) and (11)) introduced above. For example, if the critical gap takes a value of 1.0 seconds, we obtain a probability of 0.23 according to equation (11).

Here, we define a rear-end collision as the collision of a following vehicle with the leading vehicle due to an unsafe gap, when the leading vehicle suddenly decelerates. There are at most ($i-1$) rear-end collisions as the first vehicle in a platoon of size i makes a sudden deceleration.

Figure 5 shows our calculation of the probability of at least one rear-end collision as the jth vehicle of a platoon of i vehicles suddenly decelerates or stops. For a platoon of j + 1 vehicles, the probability of having a rear-end collision is p; and the probability of no rear-end collision is $1 - p$. For a platoon of j + 2 vehicles, the probability of one rear-end collision is $2 \times p(1-p)$; the probability of two rear-end collisions is p^2; and the probability of no rear-end collisions is $1 - p^2 - 2p(1-p)$. For a platoon of j + 3 vehicles, the probability of one rear-end collision is $C_1^3 \times p(1-p)^2 = 3p(1-p)^2$; the probability of two rear-end collisions is $C_2^3 p^2(1-p) = 3p^2(1-p)$; the probability of three rear-end collisions is $C_3^3 p^3 = p^3$; and the probability of no rear-end collisions is $1 - 3p(1-p)^2 - 3p^2(1-p) - p^3$.

As such, for a platoon of i (*where* $i \geq j$) vehicles, the probability of a rear-end collision is

$$C_1^{i-j} \times p(1-p)^{i-j-1} + C_2^{i-j} \times p^2(1-p)^{i-j-2} + ... + C_{i-j}^{i-j} \times p^{i-j}(1-p)^0$$

Thus, we are able to generalize the equation for calculating the rear-end collision probability when the problem vehicle is the jth vehicle in a platoon of size i as

$$p(y_i^j) = \begin{cases} \sum_{m=1}^{i-j} \frac{(i-j)!}{(i-j-m)!\,m!} p^m (1-p)^{i-j-m}, & \forall i \in \mathbf{N} \\ 0, \text{otherwise} \end{cases} \quad (12)$$

The probability of no rear-end collision is

$$p(\bar{y}_i^j) = \begin{cases} 1 - \sum_{m=1}^{i-j} \frac{(i-j)!}{(i-j-m)!\,m!} p^m (1-p)^{i-j-m}, & \forall i \in \mathbf{N} \\ 1, \text{otherwise} \end{cases} \quad (13)$$

where

$p(y_i^j)$ is the probability that at least one rear-end collision occurs as the jth vehicle of a platoon of size i makes a sudden deceleration.

$p(\bar{y}_i^j)$ is the probability that no rear-end collision occurs as the jth vehicle of a platoon of size i makes a sudden deceleration.

p is the probability that a vehicle has a gap less than the critical gap value (probability of having a rear-end collision);

j is the position of the problem vehicle in the platoon;

i is the platoon size;

m is the number of rear-end collisions that take place; and

$\mathbf{N}$ is the vector of all possible platoon sizes [2,3,4…].

When a nonplatooning vehicle keeps a headway of at least four seconds, the probability of rear-end collisions due to the sudden deceleration of a nonplatooning vehicle is considered to be zero $(p(y_1) = 0)$.

Probability that the *j*th Vehicle in a Platoon is a Problem Vehicle

The number of vehicles involved in a rear-end collision is relative to the platoon size and the position of the problem vehicle in the platoon. For example, if the leading car in a longer platoon is involved in an accident, the probability of a multivehicle rear-end collision is higher than that when any other vehicle in the platoon is involved.

Equation (14) gives the probability that a problem vehicle is in a platoon of size *i*:

$$p(x_i) = \frac{i\psi(i)}{V} \quad (14)$$

where

$p(x_i)$ is the probability that the problem vehicle belongs to a platoon of size *i*;

$\psi(i)$ is the number of platoons of size *i*, which is obtained from figures 1 and 2 for a given volume; and

V is the traffic volume.

The problem vehicle has an equal chance of being at any position within a given platoon. Thus, equation (15) was constructed to represent the probability that the problem vehicle is the *j*th vehicle in a platoon of size *i*:

$$p(x_i^j) = \frac{p(x_i)}{i} = \frac{i \bullet \psi(i)}{i \bullet V} = \frac{\psi(i)}{V}, \qquad \forall (j \le i) \quad (15)$$

Finally, equation (16) was developed to calculate the probability of having rear-end collision(s) when any vehicle in a traffic flow makes a sudden deceleration:

$$P(\textit{rear-end collisions}) = \sum_{i=1}^{N} \sum_{j=1}^{i} p(x_i^j) p(y_i^j) \quad (16)$$

As the probability of a rear-end collision caused by the sudden deceleration of a nonplatooning vehicle is considered to be zero $(p(y_1) = 0)$, equation (16) can be modified:

$$P(\textit{rear-end collisions}) = \sum_{i=2}^{N} \sum_{j=1}^{i} p(x_i^j) p(y_i^j) \quad (17)$$

where

$p(\textit{rear-end collisions})$ is the probability of one or more rear-end collisions as a problem vehicle suddenly decelerates;

$p(x_i^j)$ is the probability that the *j*th vehicle in a platon of size *i* is a problem vehicle, which is given by equation (15);

$p(y_i^j)$ is the probability of rear-end collisions when the *j*th vehicle in a platton of size *i* suddenly decelerates, which is given by equation (12).

Number of Vehicles in Rear-End Collisions

In order to predict the number of vehicles involved in rear-end collisions caused by the sudden deceleration of a problem vehicle in a work zone, it is necessary to know the mean number of vehicles (κ_i) involved in rear-end collisions for each platoon size *i*. We also need to know the probability (p_i) that the problem vehicle belongs to a platoon of size *i*. The mean number of vehicles in rear-end collisions can be computed by $\sum_{i \in \mathbf{N}} p_i \kappa_i$.

Finding κ_i

For a particular platoon $(i \in \mathbf{N})$, the number of vehicles involved in rear-end collisions depends on the position of the problem vehicle and platoon size. All vehicles in the platoon have an equal chance of being the problem vehicle, but each has a different number of following vehicles.

The number of vehicles involved in rear-end collisions also depends on the type of collision. If *m* rear-end collisions are continuous, there will be $m + 1$ vehicles involved. On the other hand, if these rear-end collisions are discrete, there will be at most $2m$ vehicles involved. To make a comprehensive prediction, we used the average value $(m + 1 + 2m)/2$ as the number of vehicles involved in *m* rear-end collisions.

We defined κ_i^j as the mean size of the rear-end collision if the problem vehicle is the *j*th vehicle in a platoon with *i* vehicles. Assuming that the problem vehicle is the first vehicle in a platoon, the following examples demonstrate how to find κ_i^1.

A platoon of two vehicles involves only one possible rear-end collision, thus $\kappa_2^1 = 2$.

For a platoon of three vehicles, the probability of a rear-end collision is $2 \times p(1-p)$; a probability of having two rear-end collisions is p^2; thus

$$\kappa_3^1 = \frac{2 \times 2 \times p(1-p) + 3 \times p^2}{2 \times p(1-p) + 3 \times p^2}$$

Likewise, for a platoon of four vehicles, the probability of having a rear-end collision is

$C_1^3 \times p(1-p)^2 = 3p(1-p)^2$;

the probability of having two rear-end collisions is

$C_2^3 p^2 (1-p)^2 = 3p^2(1-p)$;

the probability of having three rear-end collisions is

$C_3^3 p^3 = p^3$;

thus

$$\kappa_4^1 = \frac{2C_1^3 p(1-p)^2 + 3.5C_2^3 p^2(1-p) + 4C_3^3 p^3}{C_1^3 p(1-p)^2 + C_2^3 p^2(1-p) + C_3^3 p^3}.$$

We estimate the equation for κ_i^1 as follows:

$$\kappa_i^1 = \frac{\sum_{m=1}^{i-1} \left(\frac{3m+1}{2}\right) C_m^{i-1} p^m (1-p)^{i-1-m}}{\sum_{m=1}^{i-1} C_m^{i-1} p^m (1-p)^{i-1-m}} \qquad (18)$$

Similarly, the general formula for the mean number of vehicles in the rear-end collision when the *j*th vehicle is the problem vehicle in a platoon of size *i* is

$$\kappa_i^j = \frac{\sum_{m=1}^{i-j} \left(\frac{3m+1}{2}\right) C_m^{i-j} p^m (1-p)^{i-j-m}}{\sum_{m=1}^{i-j} C_m^{i-j} p^m (1-p)^{i-j-m}} \qquad (19)$$

Now, we can compute κ_i from the following equation:

$$\kappa_i = \frac{1}{i-1} \sum_{j=1}^{i-1} \kappa_i^j \qquad (20)$$

Therefore, the mean number of vehicles involved in a crash caused by sudden deceleration can be obtained from

$$\kappa = \sum_{i \in \mathbf{N}} p_i \kappa_i \qquad (21)$$

The p_i can be calculated easily using the percentage of platooning and the platoon size distribution.

CASE STUDY

Our case study attempts to predict the probability of rear-end collisions and the mean number of vehicles involved at a long-term work zone. We developed equations to make the prediction using two input variables. The input variables are: 1) work zone type (long-term or short-term), and 2) hourly volume.

Assume that there is a sudden deceleration in a work zone traffic flow, the proposed methodology presented here can be used to answer the following questions:

1. What is the probability of a rear-end crash?
2. How many vehicles might be involved in this crash?

The prediction uses equations (1), (5), (11), (17), and (21) to answer the above questions. Predictions are for a long-term work zone with volumes of 400 to 1,600 vehicles per hour at increments of 200.

Solution to Question 1

Assume the maximum platoon size is 15 vehicles.

$$P(\textit{rear-end collisions}) = \sum_{i=2}^{15} \sum_{j=1}^{i-1} p(x_i^j) p(y_i^j)$$

$$= \sum_{i=2}^{15} \sum_{j=1}^{i-1} \sum_{m=1}^{i-j} \frac{\psi(i)}{V} \frac{(i-j)!}{(i-j-m)!\,m!} p^m (1-p)^{i-j-m}$$

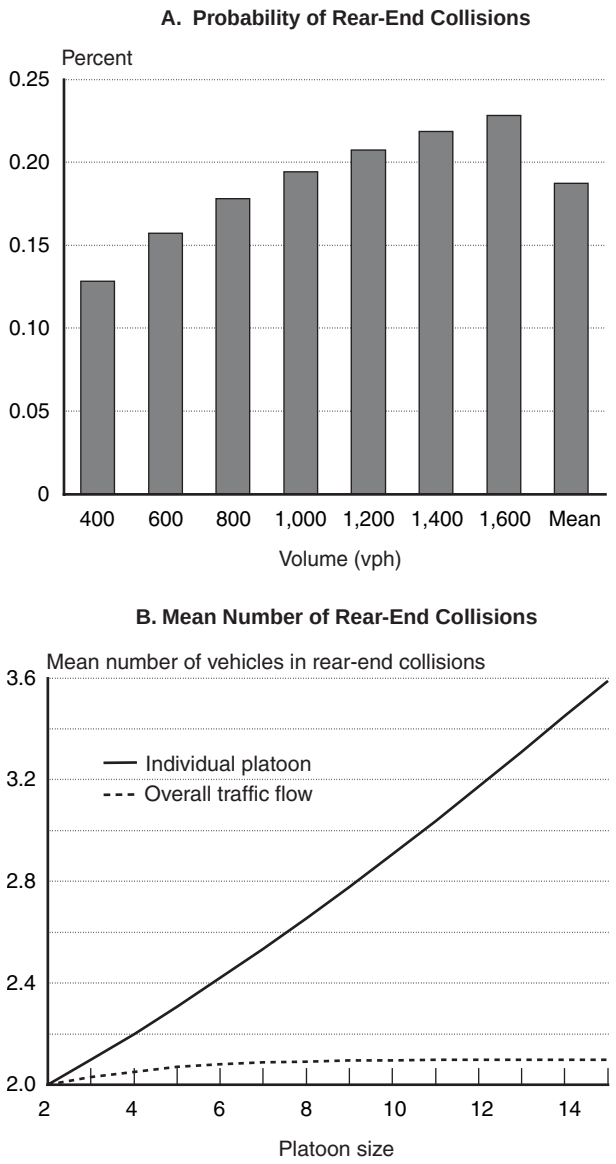
In a long-term work zone, with the assumption of a critical gap of 1.0 seconds, equation (11) gives us a p of 0.23. This is the conditional probability of having a rear-end collision given a sudden stop or deceleration of a platoon vehicle due to an incident, error maneuver, or some other unexpected reason. This probability may seem high; however, this rear-end collision probability is defined differently from the frequency of rear-end collisions in accident statistics. To get a real overall probability or frequency of real rear-end collisions on a given highway, this probability must be multiplied by the sum probability of all other types of accidents involving only a single vehicle at this location.

Using equations (1) and (7), and the average platoon size μ as 3.2, we can compute $\psi(i)$ from

$$\psi(i) = \frac{V}{\mu} \bullet [-1.377 + 0.327 \ln(V)] \bullet \left\{ \frac{1}{1.4079} \exp[-(x - 1.4856)/1.4079] \right\}$$

Now we can compute the conditional probability of rear-end collisions if one vehicle suddenly stops or decelerates. Figure 6A shows that the mean probability of a rear-end collision in a long-term work zone is 18.74% if a vehicle suddenly decelerates or

FIGURE 6 Probability and Mean Size of Rear-End Collisions



stops. The results also show that the risk of rear-end collisions increases as the volume increases.

Solution to Question 2

We also calculated the mean number of vehicles involved in rear-end collision(s) for different platoon sizes:

$$\kappa_i = \frac{1}{i-1}\sum_{j=1}^{i-1}\frac{\sum_{m=1}^{i-j}\frac{(3m+1)}{2}C_m^{i-j}p^m(1-p)^{i-j-m}}{\sum_{m=1}^{i-j}C_m^{i-j}p^m(1-p)^{i-j-m}}$$

Figure 6B shows that the mean number of vehicles, κ_i, will increase from 2.0 to 3.6 when the platoon size grows from 2 up to 15. The figure also shows that the mean number of vehicles involved for overall traffic will increase from 2.0 to 2.1 as the maximum platoon size of a traffic flow grows from 2 to 15. Obviously, the change in κ is not significant while the maximum platoon size changes significantly, even though the possible mean number of involved vehicles, κ_i, for a platoon size of 15 is about 1.8 times that for a platoon size of 2. This implies that using only the mean value may be misleading when we want to understand safety performance in work zones, because the change in the maximum platoon size will make a significant difference in the consequence of the worst case.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This paper presents an investigation of the platooning and gap characteristics in Interstate highway work zones, as well as of the gap sizes under different car-following patterns and work zone types. The study is based on data covering more than 15,000 observations. Models of the platoon size and gap size distribution for long-term and short-term work zones were developed. An in-depth analysis of the data reveals a safety paradox, which may indicate that drivers do not understand the safety implications of time and space gaps relative to speed limit increases at work zones. All the findings with respect to car-following characteristics provide practitioners a better understanding of drivers' behaviors in work zone areas.

We propose a new methodology to predict the probability of rear-end collisions in a work zone and the mean number of vehicles involved. Only two simple inputs are required to predict rear-end collisions using gap and platooning models. Because it is sometimes impossible to evaluate work zone safety performance using real crash data, this new methodology provides an alternative approach to assessing the safety performance in Interstate highway work zones. We present a case study to demonstrate the implementation of the new prediction methodology.

Some areas for future research include integrating the effect of heavy vehicle and work activity intensity on safety as an interesting extension to our methodology. It will also be important to conduct some disaggregate analysis to address the inter-

dependence of different car-following patterns. A study of this nature may need to consider the impact of various groups of drivers, such as age group, gender, driving habits, etc., which may require more extensive data collection.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees for their helpful comments during the development stage of this paper.

REFERENCES

Akcelik, R. and E. Chung. 1994. Calibration of the Bunched Exponential Distribution of Arrival Headways. *Road and Transport Research* (Australian Road Research Board) 3(1):43–59.

Benekohal, R.F. and S. Sadeghhosseini. 1991. Platooning Characteristics of Vehicles in Highway Construction Zones. *Modeling and Simulation* 22:16–23.

Griffiths, J.D. and J.G. Hunt. 1991. Vehicle Headways in Urban Areas. *Traffic Engineering and Control* 32(10):458–462.

Hoban, C.J. 1983. Toward a Review of the Concept of Level of Service for Two-Lane Rural Roads. *Australian Road Research*, September, pp. 216–218.

Keller, H. 1976. Effects of a General Speed Limit on Platoons of Vehicles. *Traffic Engineering and Control*, July, pp. 300–303.

Luttinen, R.T. 1992. Statistical Properties of Vehicle Time Headways. *Transportation Research Record* 365:92–98.

May, A.D. 1990. *Traffic Flow Fundamentals.* Englewood Cliffs, NJ: Prentice Hall.

Mei, M. and A.G.R. Bullen. 1993. The Log-Normal Distribution for High Traffic Flow. *Transportation Research Record* 1398:125–128.

Sadeghhosseini, S. and R.F. Benekohal. 1995. Space Headway and Safety of Platooning Highway Traffic, Traffic Congestion and Traffic Safety in the 21st Century, Chicago, Illinois, pp. 472-478.

Sumner, R. and C. Baguley. 1978. *Close Following Behavior at Two Sites on Rural Two Lane Motorways*, TRRL Report 859. Crowthorne, UK: Transport and Road Research Laboratory.

Wasielewski, P. 1979. Car-Following Headways on Freeways Interpreted by the Semi-Poisson Headway Distribution Model. *Transportation Science* 13(1):36–55.

Model	Density function and parameters
Weibull	Density function $f(x) = \frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{(\gamma-1)} \exp(-((x-u)/\alpha)^{\gamma}) \quad x \geq \mu;\gamma,\alpha > 0$ where γ is the shape parameter μ is the location parameter α is the scale parameter
BetaGeneral (Beta Generalized)	Density function $f(x) = \frac{(x-\min)^{\alpha_2-1}(\max-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)(\max-\min)^{\alpha_1+\alpha_2-1}}$ where α_1 is the continuous shape parameter $\alpha_1 > 0$ α_2 is the continuous shape parameter $\alpha_2 > 0$ **min** is the continuous boundary parameter min < max **max** is the continuous boundary parameter
Gamma	Density function $f(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}$ with gamma function $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}e^{-u}du$ where α is the continuous shape parameter $\alpha > 0$ β is the continuous shape parameter $\beta > 0$
Inverse Gaussian	Density function $f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\left[\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]}$ where both μ and λ are positive continuous parameters
Log normal	Density function $f(x) = \frac{1}{x\sqrt{2\pi}\sigma'} e^{-\frac{1}{2}\left[\frac{\ln x-\mu'}{\sigma'}\right]^2}$ with $\mu' = ln\left[\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right]$ and $\sigma' = \sqrt{ln\left[1+\left(\frac{\sigma}{\mu}\right)^2\right]}$ where both μ and σ are positive continuous parameters

Sampling and Estimation Techniques for Estimating Bus System Passenger-Miles

PETER G. FURTH

Civil & Environmental Engineering, 400SN
Northeastern University
360 Huntington Avenue
Boston, MA 02115
Email: p.furth@neu.edu

ABSTRACT

Most U.S. bus systems conduct on-off counts on a sample of vehicle-trips to estimate annual passenger-miles, which must be submitted to the National Transit Database. The required sample size depends on the techniques used. This paper reviews alternative methods, including simple random sampling, ratio estimation with a variety of possible auxiliary variables, stratified sampling, cluster sampling, and combinations of these approaches. Most of these alternatives take advantage of electronic registering fareboxes to obtain complete counts of boarding passengers.

Seven alternative estimation techniques are compared in a case study of Santa Cruz Metro. The most efficient approach combined two techniques, stratified sampling and ratio estimation using the combined ratio technique. The latter technique used on-off data from a sample of trips to estimate the ratio of passenger-miles to *potential passenger-miles*, a newly proposed auxiliary variable. This approach reduced the sampling burden by over 80% compared with both simple random sampling and a sampling method published by the Federal Transit Administration.

KEYWORDS: National Transit Database, passenger-miles, sampling.

INTRODUCTION

The Federal Transit Administration (FTA) requires that transit agencies benefiting from federal assistance report annual passenger-miles by mode to the National Transit Database (NTD).[1] Because transit agencies, unlike airlines, do not routinely capture passengers' origin-destination information, measuring passenger-miles is usually done at the level of a trip (i.e., a vehicle-trip), based on on-off counts made at each stop by an onboard surveyor called a checker. Because of the high labor cost involved, on-off counts are generally done on a sample of trips from which an estimate of annual passenger-miles is made. FTA specifies that passenger-miles estimates achieve ±10% precision at the 95% confidence level.

For the bus mode, FTA's precision requirement may be satisfied by following the sampling plan laid out in Circular 2710.1A (USDOT FTA 1990). This sampling plan, based on direct estimation of mean passenger-miles per trip from a random sample of at least 549 trips, is relatively burdensome, requiring roughly one full-time equivalent employee to conduct on-off counts. Alternatively, an agency may use a custom-made sampling and estimation plan, as long as it is applied with a sample size that achieves the specified precision level. By taking advantage of an agency's particular features—its size, route structure, and availability of data on other measures of passenger use that correlate strongly with passenger-miles—custom sampling plans can substantially reduce the sampling burden.

One particular development in the transit industry creates the possibility for more efficient sampling plans. It is the widespread adoption of electronic fareboxes, with which transit agencies can count boardings on every trip. Because boardings are correlated with passenger-miles, an alternative to directly estimating mean passenger-miles per trip from a sample of trips is to estimate the ratio between passenger-miles and boardings, and then expand this ratio by the annual boardings count.

Besides the ratio estimation techniques, a number of other sampling and estimation techniques can improve the precision of an annual passenger-miles estimate. This paper describes several approaches for estimating annual passenger-miles for bus systems, based on experiences developing sampling plans for more than 20 U.S. transit agencies. Numerical results for seven alternatives are compared in a case study of Santa Cruz Metro (California), a system with 47 routes that vary considerably in length and ridership. Alternative estimation techniques are shown to reduce the sampling burden by over 80% compared with Circular 2710.1A. The most efficient sampling approach is found to be one that combines stratified sampling with ratio estimation, estimating the ratio of passenger-miles to a newly proposed auxiliary variable called *potential passenger-miles*, defined for a trip as the product of (passenger boardings) * (route length).

This paper has four main sections. The first describes in more detail how passenger-miles are measured and why transit agencies are looking for more efficient sampling techniques. The second introduces the case study agency. The third describes alternative approaches to estimating passenger-miles, along with results from the case study. The final section compares the alternatives and offers conclusions.

SAMPLING ELEMENT AND THE COST OF MEASUREMENT

As mentioned previously, the measurement unit for passenger-miles is a trip, normally defined as the one-way movement of a vehicle from one terminal to another. To measure passenger-miles for a selected trip, one counts ons and offs, also called boardings and alightings, by stop. From the on-off count, passenger load on every interstop segment may be determined. Multiplying the load on each segment by segment length (a known quantity) yields the number of passenger-miles occurring on each segment, and summing over all segments yields passenger-miles for the trip. Note that with this measurement technique it is neither possible nor necessary to know the trip length of individual passengers. Also note that on-off counts also yield a measurement of trip-level boardings, so that paired measurement of boardings and passenger-miles is no more burdensome than measuring passenger-miles alone.

[1] See the NTD website at www.ntdprogram.com.

The industry norm is for on-off counts to be made by transit agency employees known as checkers. In all but the smallest transit agencies, it is too disruptive of operations for bus operators to make on-off counts. On-off counts can also be obtained using automatic passenger counters (APCs), but only a small number of U.S. transit agencies have APCs due to their high cost.

Anecdotal evidence indicates that the majority of U.S. transit agencies follow the sampling plan of Circular 2710.1A, checking 549 trips per year. Transit agencies find this sampling requirement a rather onerous burden. While sampled trips may last only 30 minutes on average, the checker time involved can run upwards of 2 hours per trip because of travel time to the start of the trip, slack time to ensure catching the right bus, and return time. Coordination and supervision are also difficult. When multiple trips per day are sampled, one may be early in the morning and another late at night; they may be separated by a large geographic distance; and two selected trips may be close enough in time that it is impossible for the same person to check both.

For the vast majority of U.S. transit agencies, then, sampling trips to estimate passenger-miles involves considerable labor cost, a cost that agencies are interested in reducing by employing more efficient estimation and sampling techniques. A recent National Academy of Sciences review of NTD legislation, while acknowledging that the federal requirement for reporting passenger-miles estimates is reasonable (because passenger-miles is part of the legislated formula for allocating federal funding), also acknowledges the burden of this type of sampling (Furth and McCollom 1987). It recommends the development of more efficient sampling plans, particularly plans that take advantage of boarding counts made using electronic fareboxes.

SANTA CRUZ METRO CASE STUDY

Santa Cruz Metro (SCM) affords an interesting case study of sampling methods to estimate annual bus passenger-miles. Its 47 bus routes include short and long local routes, long commuter routes, and one very long and heavily used express route along Highway 17 that crosses the mountains to San Jose. It is common for a route to follow several different routing patterns in the daily schedule, and many of its routes are loops.

Using electronic fareboxes, SCM counts all passenger boardings. Furthermore, its boardings counts are always associated with the route being served, making it possible to use estimation techniques based on route-level boardings.

Because sampling requirements are driven by weekday service, which typically accounts for 85% to 90% of passenger-miles, the case study analysis was confined to weekday service. Historic data available for analyzing estimation techniques was a single day's observation of every trip in the weekday schedule in fiscal year 2001. Each trip record indicates the trip's boardings and passenger-miles, as well as identifiers (route, date, and so forth).

In order to make the case study more representative of other U.S. transit systems, the Highway 17 route was omitted. Unlike SCM's other routes, that route operates mostly along a freeway using over-the-road coaches. Because nearly all its passengers travel express between Santa Cruz and San Jose, average passenger trip length is very high (about 30 miles) and has little variability, making it easy to estimate passenger-miles for this route. Because it accounts for 15% of SCM's passenger-miles, including the Highway 17 route would substantially distort the case study results from the perspective of representing a "typical" transit agency.

Table 1 provides a comparison of necessary sample sizes and other relevant statistics for SCM using alternative estimation approaches. Entries in the table are explained in the following sections of the paper as each estimation alternative is presented. The sample sizes used in these comparisons are not "finished." Their application would require accounting for weekends and the Highway 17 route, and would probably involve some rounding. In addition, all calculated sample sizes for SCM are inflated by 50% relative to the sample size formula given. This degree of oversampling is a reasonable precaution, because sample size calculations are based on historic data, and transit agencies typically use a recommended sample size for several years before recalibrating the sample size requirement using a more recent dataset. Including oversampling

TABLE 1 Comparison of Sampling Strategies

Sampling and estimation alternative		Unstratified ratio estimation statistics			Overall performance		
		cv of pass-miles	*cv* of auxiliary variable	Correlation coeff between pass-miles and auxiliary variable	Unit *cv*	Necessary sample size (SS)	SS reduction vs. Circular
—	Circular 2710.1A				n.a.	549	
A	Direct sampling and estimation				0.95	522	5%
B	Ratio estimation; auxiliary variable = boardings	0.95	0.77	0.67	0.72	296	46%
C	Stratified ratio estimation (4 strata); auxiliary variable = boardings				0.43	117	79%
D	Combined ratio estimation; auxiliary variable = boardings				0.45	117	79%
E	Ratio estimation; auxiliary variable = (boardings * route length)	0.95	0.91	0.89	0.44	112	80%
F	Ratio estimation; auxiliary variable = (boardings * adjusted route length)	0.95	1.09	0.89	0.51	147	73%
G	Combined ratio estimation; auxiliary variable = (boardings * route length)				0.39	86	84%

KEY: *cv* = coefficient of variation.

also makes comparisons with Circular 2710.1A more "fair," because the sample sizes called for in that circular, being intended for application nationwide, include a certain degree of oversampling.

SAMPLING AND ESTIMATION APPROACHES

Simple Random Sampling (Alternative A)

The population is all of the trips operated by a transit agency in a year; let N be the population size (number of trips operated in a year). Let Y = passenger-miles at the trip level, so that y_i = passenger-miles on trip i. Let $\overline{Y}$ and S_y be the population mean and standard deviation of Y; then $cv_y = S_y / \overline{Y}$ is the coefficient of variation (cv) of Y. Let n be the sample size, that is, the number of trips observed by means of on-off counts. Finally, let $\hat{Y}_{total}$ be the estimate of annual system-wide passenger-miles, the ultimate quantity being estimated.

If the sample of observed trips is drawn at random from the population of trips operated over the year, $\overline{Y}$ can be estimated by the sample mean

$$\bar{y} = \frac{1}{n}\sum y_i \tag{1}$$

The relative standard error (*r.s.e.*) of $\bar{y}$ is

$$r.s.e. = \frac{cv_y}{\sqrt{n}} \tag{2}$$

For unbiased estimators, *r.s.e.* is the *cv* of the estimator, and $(r.s.e.)^2$ is the relative variance of the estimator. Because the population size N is generally

large compared with the sample size, the finite population correction is ignored.

The estimate of total annual passenger-miles is found by direct expansion of the sample mean

$$\hat{Y}_{total} = N\bar{y} \quad (3)$$

Because N is a known constant, the relative standard error and precision of $\hat{Y}_{total}$ are the same as those of $\bar{y}$.

Precision at the 95% confidence level (*prec*) is given by

$$prec = 1.96(r.s.e.) \quad (4)$$

The necessary sample size to achieve a specified precision at the 95% confidence level is therefore given by

$$n = \left(\frac{1.96cv_y}{prec}\right)^2 \quad (5)$$

In our experience analyzing data from about 20 transit agencies, we have found the passenger-miles *cv* to almost always lie in the range 0.8 to 1.2, corresponding to a sampling requirement (without oversampling) of 250 to 550 trips. Two passenger-miles *cv* values already reported in the literature are 0.82 for greater Pittsburgh (Furth 1998) and 1.08 for greater Buffalo (Townes 2001). Only once have we encountered an agency with passenger-miles *cv* exceeding 1.2; the value of its *cv*, 1.3, would have required a sample size of 650 trips if the simple random sampling approach had been chosen.

SCM results for simple random sampling are shown in table 1 as alternative A. SCM's *cv* of passenger-miles was found to be 0.95. The corresponding necessary sample size, with 50% oversampling, was 522 trips.

Circular 2710.1A

The sampling plan in FTA Circular 2710.1A also uses the sample mean as an estimator. It varies slightly from random sampling because it uses a two stage sample, selecting n_1 days within the year in stage 1 and n_2 trips for each selected day in stage 2, with a resulting sample size of $n = n_1 n_2$. The circular offers a family of combinations of n_1 and n_2. Choices at stage 1 are sampling every day (n_1 = 365), every other day (n_1 = 183), every third day (n_1 = 122), and so forth. The combination with the smallest sample size, which is preferred by most agencies that follow the circular, is to sample 3 trips every other day, for a sample size of 549 trips per year. This sampling requirement is based on analysis done in the late 1970s of data from two transit agencies and was first published in 1978 as Circular 2710.1.

The standard error of a sample mean obtained using two-stage sampling involves variances at the two stages (Cochran 1977). However, it turns out that for passenger-miles, between-day variance of mean passenger-miles per trip is negligible in comparison with between-trip variance within a day, and the latter is essentially the same as between-trip variance over the entire population of a year's trips. Therefore, compared with simple random sampling, no advantage is gained by deliberately using a two-stage sample of the type used by Circular 2710.1A.

For the same reasons, the precision obtained using the two-stage approach of Circular 2710.1A is essentially the same as what would be obtained using simple random sampling with the same sample size. The range of passenger-miles *cv's* reported earlier, therefore, confirms the reasonableness of the Circular 2710.1A sample size, in the sense that most transit agencies following its sampling plan will achieve the specified precision.

Ratio Estimation

Ratio estimation (Furth and McCollom 1987; Cochran 1977) is a sampling and estimation technique that takes advantage of available data on an auxiliary variable that is closely correlated to the variable of interest. In order to use ratio estimation, two conditions must be met: the annual total of the *auxiliary variable* must be known, and sampled trips must provide paired measurements of the variable of interest (passenger-miles) and the auxiliary variable. Auxiliary variables that have been used for passenger-miles estimation include boardings and revenue.

Let X be the name of the auxiliary variable at the trip level; for the sake of definiteness, let X be trip-level boardings. Its population mean and total, $\overline{X}$ and X_{total}, are assumed to be known. Sampling yields a set of n-paired observations (x_i, y_i). Let $\bar{x}$ be

the mean of X from this sample. Also of interest are the statistics

S_x^2 variance of X, usually estimated using the sample variance s_x^2,

$cv_x = s_x / \bar{x}$ = estimated coefficient of variation of X,

r_{xy} = estimated correlation coefficient of X and Y.

Here, we are interested in estimating the ratio $R_{population} = \bar{Y} / \bar{X}$, which often has an intuitive meaning. When X represents boardings, this ratio is the average length of an unlinked passenger-trip, usually called average passenger trip length.

$R_{population}$ is estimated from the paired sample by statistic R, the ratio of sample means

$$R = \frac{\bar{y}}{\bar{x}} \tag{6}$$

The estimate of annual system total passenger-miles is then the product

$$\hat{Y}_{total} = X_{total} R \tag{7}$$

Because X_{total} is a known constant, R and $\hat{Y}_{total}$ have the same relative standard error and the same precision. The relative standard error of a ratio estimate is given by

$$r.s.e. = \frac{1}{\sqrt{n}} \sqrt{cv_x^2 + cv_y^2 - 2 r_{xy} cv_x cv_y} \tag{8}$$

"Unit cv" as a Measure of Statistical Efficiency

Equation (8) can also be expressed in the form

$$r.s.e. = \frac{ucv}{\sqrt{n}} \tag{9}$$

where the ratio estimator's *ucv*, standing for unit *cv*, is given by

$$ucv = \sqrt{cv_x^2 + cv_y^2 - 2 r_{xy} cv_x cv_y} \tag{10}$$

The concept of unit *cv* can also be applied to simple random sampling. Comparing equations (2) and (9), it is clear that, for simple random sampling, the unit *cv* is

$$ucv = cv_y \tag{11}$$

Unit *cv* is a convenient term, first proposed by Furth and McCollom (1987), for comparing the efficiency of estimation techniques. It summarizes the inherent variability in an estimation technique, because the relative variance of an estimate depends only on the unit *cv* and the sample size. By comparing unit *cv's* of various estimation techniques, we can readily see which one requires a greater sample size or yields the more precise estimate for a given sample size.

Using the concept of unit *cv*, a sample size formula that applies to all the estimation techniques presented in this paper is

$$n = \left(\frac{1.96 ucv}{prec} \right)^2 \tag{12}$$

and the precision (at the 95% confidence level) obtained for a given sample size is

$$prec = \frac{1.96 ucv}{\sqrt{n}} \tag{13}$$

With ratio estimation, bias can become a problem at low sample sizes (Cochran 1977). Equations (12) and (13) are only valid as long as the sample size is neither so small that bias becomes significant, nor so large that the finite population correction applies.

The efficiency of a ratio estimator depends strongly on the correlation coefficient, at the trip level, between the auxiliary variable and passenger-miles. Squaring equation (10) and rearranging, the square of the unit *cv* can be expressed as the sum of two terms

$$ucv^2 = (cv_x - cv_y)^2 + 2 cv_x cv_y (1 - r_{xy}) \tag{14}$$

For the kinds of auxiliary variables normally considered when estimating passenger-miles, the second term dominates. Therefore, as a general tendency, the stronger the correlation between the auxiliary variable and passenger-miles, that is, the closer r_{xy} is to 1, the smaller the unit *cv* of the ratio and the more efficient the estimation technique.

Boardings as the Auxiliary Variable (Alternative B)

Since the sampling plan in Circular 2710.1A was first published, nearly all buses in the U.S. transit fleet have been equipped with electronic fareboxes. Besides counting revenue, electronic fareboxes can also be used to count passenger boardings, making it possible to acquire a complete, systemwide count of boardings. Because boardings are correlated with

passenger-miles—trips with more boardings tend to have more passenger-miles—boardings can serve as a useful auxiliary variable for ratio estimation. As mentioned before, the ratio of passenger-miles to boardings is average passenger trip length. A study of Buffalo area data (Furth 1998) found the correlation of boardings to passenger-miles to be 0.59—not optimal, but enough to reduce the sample size requirement by 33% compared with direct estimation of the sample mean.

The approach of using boardings to help estimate passenger-miles can only be adopted by agencies that, like SCM, count all passenger boardings. While nearly every U.S. transit agency uses electronic fareboxes, they do not all get reliable boardings counts. Boarding counts using electronic fareboxes are partly automated and partly manual. Essentially, passengers who interact with the farebox by entering a standard fare or swiping a card through an attached magnetic card reader are registered automatically. To register passengers who do not have a standard farebox interaction (e.g., passengers using a nonmagnetic transfer or pass or those paying a reduced fare because they are seniors or pupils), bus operators have to push a button corresponding to the appropriate fare category. In many large cities, where bus operator duties are particularly demanding, the farebox is not always operated in a way that yields reliable counts of passenger boardings. Where this is the case, boardings cannot be used as an auxiliary variable to estimate passenger-miles. (The Chicago Transit Authority is a good example of a large city transit agency that gets reliable boardings counts using fareboxes. They use advanced fare-collection technology to maximize the fraction of passengers registered automatically and follow management practices that emphasize the need for operators to register remaining passenger boardings.)

Results for SCM are shown in table 1 as alternative B. The correlation coefficient between boardings and passenger-miles is 0.67. The resulting unit *cv* is 0.72; comparing it with the unit *cv* for simple random sampling (0.95), we can see how using boardings as an auxiliary variable reduces the variability inherent in the estimation technique. The necessary sample size, 296 vehicle-trips, represents a reduction of 43% compared with simple random sampling, and 46% compared with Circular 2710.1A.

Revenue as the Auxiliary Variable

The earliest applications of the ratio technique for passenger-miles estimation used farebox revenue as an auxiliary variable. Farebox revenue is correlated with passenger-miles (more revenue on a trip usually means more passengers and, therefore, more passenger-miles). Annual total revenue is certainly known. And, even before the invention of the electronic registering farebox, most transit agencies had mechanical registering fareboxes that allowed checkers making on-off counts to measure trip revenue by reading the revenue register at the start and end of the trip. With this approach, the ratio of passenger-miles per dollar of revenue can be estimated from a sample of trip observations and then expanded by annual revenue to yield an estimate of annual passenger-miles.

Furth and McCollom (1987) found a relatively strong correlation between revenue and passenger-miles using Pittsburgh area data from the early 1980s. Based on that analysis and a similar analysis of data from San Antonio, FTA published a revenue-based sampling and estimation method with a sample size requirement of only 208 trips (USDOT UMTA 1985). However, FTA later withdrew default approval for this sampling plan, because widespread adoption of monthly passes weakened the correlation of passenger-miles to farebox revenue. Agencies may still use this technique, but must justify the sample size they use by analyzing local data.

This technique was not tested as part of the SCM case study, because trip revenue data were not part of the available dataset. However, given the widespread use of passes at SCM, it is likely using revenue, because an auxiliary variable would be less efficient than using boardings.

Stratified Sampling

Stratification is another approach that can improve sampling and estimation efficiency. For passenger-miles estimation, stratification has been mostly used together with the ratio technique for estimating average passenger trip length. The goal of stratification is to divide the population of vehicle-trips in a way

that passenger trip length varies as much as possible between strata, rather than within strata. Stratification is usually done by route (Huang and Smith 1993), because routes can differ widely in their average passenger trip length (typically, average passenger trip length is small on short routes and large on long routes). Three variations of stratification by route have been followed, as described below.

Each Route a Stratum

In the first variation of stratified sampling, each route is a stratum. A sample of trips is observed in each stratum, measuring both boardings and passenger-miles on each observed trip. From each sample, the average passenger trip length ratio for the stratum is estimated and then expanded to annual passenger-miles by multiplying by the stratum's annual boardings (assumed to be known). Those annual passenger-miles figures are then aggregated over all the strata to yield the systemwide, annual estimate of annual passenger-miles.

In order to apply this approach, an agency needs not only counts of all passenger boardings during the year, but the ability to break out those counts by route. Among those agencies that get a reliable count of boardings using electronic fareboxes, some are still unable to stratify by route because bus operators do not register (by pushing a sequence of buttons) every time the bus changes route, and so recorded counts cannot be associated with a particular route.

Stratum-level parameters and statistics are defined as follows. Let N_h and n_h be population size and sample size for stratum h, respectively, both measured in trips. The unsubscripted variables N and n retain their meaning as overall population size and sample size; that is,

$$\sum N_h = N$$

$$\sum n_h = n$$

The relative size of stratum h, in terms of population size, is given by

$$w_h = N_h / N \tag{15}$$

Relative size serves as a weighting factor, since

$$\sum w_h = 1 \tag{16}$$

Let $\overline{X}_h$, assumed to be known, be the mean boardings per trip within stratum h, and let $\bar{y}_h$ and $\bar{x}_h$ be the sample means of Y and X within stratum h. Finally, let s^2_{yh}, s^2_{xh}, and r_{xyh} be the sample variance of passenger-miles, the sample variance of boardings, and the sample correlation coefficient between passenger-miles and boardings, respectively, within stratum h.

The ratio estimated within each stratum is

$$R_h = \bar{y}_h / \bar{x}_h \tag{17}$$

The estimate of total annual systemwide passenger-miles involves expansion by stratum, followed by aggregation over strata:

$$\hat{Y}_{total} = \sum N_h \overline{X}_h R_h = N \sum w_h \overline{X}_h R_h \tag{18}$$

The estimate of average passenger-miles per trip is

$$\bar{y}_{strat} = \frac{\hat{Y}_{total}}{N} \tag{19}$$

Because these final two estimates differ by only the known factor N, they have the same relative standard error, and consequently the same unit *cv* and the same precision.

The variance of an estimate made using stratified sampling depends in part on how the sample is allocated among the strata. In this paper, allocation is assumed to be proportional to stratum size, that is, for a given n,

$$n_h = w_h n \tag{20}$$

Proportional allocation is not, in general, the optimal (i.e., variance-minimizing) way of allocating a sample among strata. However, for the range of parameters typically encountered in passenger-miles estimation, proportional allocation is not much inferior to optimal allocation in terms of variance, and it has other desirable properties including ease in determining sample size and making certain types of estimators self-weighting.

With proportional allocation, the relative standard error of the annual systemwide passenger-miles estimate is

$$r.s.e. = \sqrt{\frac{1}{n}} \left[\frac{\sqrt{\sum w_h \left(s^2_{yh} + R^2_h s^2_{xh} - 2 R_h r_{xyh} s_{xh} s_{yh} \right)}}{\bar{y}_{strat}} \right] \tag{21}$$

The term in brackets is the unit *cv* of the estimator. Precision (at the 95% confidence level) for a given sample size, the sample size necessary to achieve a given precision, can be determined using equations (12) and (13).

While route-level stratification is a compelling concept, it has one serious drawback. Ratio estimators are biased when sample size is small (Cochran 1977). An analysis of transit trip-level ridership data found that in order to effectively limit bias, at least 10 trips should be observed per stratum (Furth and McCollom 1987). For even a mid-sized transit agency, this limitation makes stratification by route of no practical value, limiting the approach to bus systems with a small number of routes. Therefore, stratification by route was rejected as a sampling and estimation approach for SCM.

Stratification by Route Length (Alternative C)

One way to overcome the limitation of a minimum stratum sample size is to use a coarser stratification scheme, grouping trips into strata by route length. Correlation of boardings to passenger-miles can still be expected to be a good deal stronger within a stratum than systemwide, albeit not as strong as if each route were a stratum.

In the SCM case study, routes were grouped into four strata by length. Table 2 presents relevant statistics. Stratum 4 contained SCM's long express routes (but not the excluded Highway 17 route) and accounted for about 2% of the daily vehicle-trips; the other three strata, roughly equal in size, corresponded to short, medium, and longer routes. Within each stratum, correlation of boardings to passenger-miles (at the vehicle-trip level) was rather strong, with correlation coefficients ranging from 0.79 to 0.89. Of particular interest are the average passenger trip length ratios for the four strata: 2.8, 3.2, 7.8, and 10.5 miles, respectively. The large differences of the last two strata from the first two show the benefit of separating them into different strata.

Overall results are shown in table 1 as alternative C. The unit *cv* was 0.43, a large improvement over the previously described methods. The corresponding necessary sample size was calculated to be 109; constraining stratum 4 sample to at least 10 observations results in a required sample size of 117 vehicle-trips.

Combined Ratio Estimation (Alternative D)

A third approach to stratified sampling is to use the so-called combined ratio estimation technique (Furth 1998; Cochran 1977). It uses stratified sampling to select the trips that are observed, but then uses that data to estimate a single, systemwide ratio using the equation

$$R = \frac{\sum w_h \bar{y}_h}{\sum w_h \bar{x}_h} \tag{22}$$

Using a systemwide ratio is a disadvantage relative to conventional stratified ratio estimation (i.e., a ratio estimated for each stratum), weakening the correlation of boardings to passenger-miles. However, the method also offers two advantages. First, it

TABLE 2 Stratified Ratio Sampling Statistics

	Stratum				
	1	**2**	**3**	**4**	**All**
Stratum weight	37%	31%	31%	2%	100%
Mean passenger-miles per trip	41	123	237	197	129
cv of passenger-miles	0.68	0.63	0.62	0.45	
Mean of boardings per trip	15	39	30	19	27
cv of boardings	0.56	0.71	0.54	0.46	
Ratio of passenger-miles to boardings (average passenger trip length)	2.81	3.21	7.78	10.54	
Boardings to passenger-miles correlation coefficient	0.89	0.88	0.79	0.79	
Unit *cv* of the ratio	0.31	0.34	0.38	0.30	
Contribution to the sum in equation (21) (relative contribution to variance of the systemwide estimate)	60	549	2, 479	57	3, 144

is unbiased regardless of stratum sample size, and, therefore, permits every route to be a stratum. Second, it requires only knowledge of systemwide, not route-level, boardings, and, therefore, can be applied by transit agencies that routinely count all passenger-boardings, even if they cannot break out the counts by route.

In this technique, on-off counts are made for one or more trips on each route, providing paired observations of y (passenger-miles) and x (boardings), from which the combined ratio is calculated using equation (22). Allocation of the sample between strata (i.e., between routes) is again proportional to size. Relative standard error is estimated by

$$r.s.e. = \sqrt{\frac{1}{n}}\left[\frac{\sqrt{\sum w_h\left(s_{yh}^2 + Rs_{xh}^2 - 2Rr_{xyh}s_{xh}s_{yh}\right)}}{\bar{y}_{combined}}\right] \quad (23)$$

where

$$\bar{y}_{combined} = \bar{X}R \quad (24)$$

is the estimated mean passenger-miles per trip. Again, the quantity in brackets in equation (23) is the unit *cv* of the estimator.

The only previously published report that uses this technique for passenger-miles estimation found it to be very efficient. When applied to the eight-route transit system of Kenosha, Wisconsin, it called for a sample size of fewer that 50 vehicle-trips (Furth 1998). However, when SCM used this technique it was not as efficient. As shown in table 1, under alternative D, the unit *cv* (0.45) and the necessary sample size (117) are virtually the same as obtained for alternative C, conventional stratified ratio estimation.

Closer examination of the differences between conventional stratified ratio estimation and combined ratio estimation helps explain why the combined method did not perform as well at SCM as in Kenosha. Equation (23) is the same as equation (21), except that the former uses the combined ratio in place of stratum-specific ratios. In both formulas, the sum in the numerator represents the expected squared difference between observed and predicted passenger-miles. For conventional stratified ratio estimation, this difference for a paired observation (y_{ih}, x_{ih}) is $(y_{ih} - R_h x_{ih})$, while with combined ratio estimation the difference is $(y_{ih} - Rx_{ih})$. Naturally, differences tend to be smaller when using a stratum-specific ratio; the degree to which this factor hurts the performance of the combined ratio technique depends on how much average passenger trip length varies between routes. Because average passenger trip length is closely related to route length, one would expect the technique to be more effective when route length varies little within the network.

Not surprisingly, Kenosha's transit system, like those of many small cities, uses pulse scheduling based around a transit center. In this kind of network, routes are all designed to have roughly the same length. At SCM, in contrast, routes vary considerably in length, and so average passenger trip length varies widely between routes. This explains why for SCM the combined ratio technique holds no advantage over conventional stratified ratio estimation for estimating average passenger trip length. This is a significant finding that most likely extends to other transit systems whose route lengths vary considerably from one another.

Using Potential Passenger-Miles as the Auxiliary Variable (Alternative E)

In an effort to improve sampling efficiency further, a new auxiliary variable is proposed: the product of boardings and route length, which can be called *potential passenger-miles*. This formulation is motivated by the observation that trip-level passenger-miles tend to be proportional to not only the number of passengers on the trip but also to the overall length of the route. The ratio of passenger-miles to potential passenger-miles has an intuitive interpretation: it is the average fraction of a route's length that passengers travel. For example, a ratio of 0.6 would indicate that on average, passengers travel 60% of the length of their chosen route.

This estimation approach requires the usual sample of on-off counts and knowledge of annual boardings by route. Because route length is a known constant, potential passenger-miles can be calculated for both the sample data and the annual totals by simply multiplying every boarding count by the length of the route on which the count was made. In the SCM case study, on routes with multiple

routing patterns, "route length" was defined to be the length of the most often used pattern.

Mathematically, alternative E is simply ratio estimation, like alternative B, except that the auxiliary variable X is redefined to be potential passenger-miles. As indicated in table 1, alternative E, the correlation of passenger-miles with potential passenger-miles (r_{xy} = 0.89) turns out to be considerably stronger that the correlation with boardings alone (r_{xy} = 0.67 in alternative B); as a result, there is an impressive reduction in necessary sample size (from 296 to 112) when the auxiliary variable is changed from boardings to potential passenger-miles.

This result shows the value of the compound auxiliary variable (boardings * route length). However, as an overall approach, unstratified ratio estimation using this auxiliary variable still offers no substantial improvement to stratified ratio estimation using boardings as an auxiliary variable.

Using Adjusted Route Length to Calculate Potential Passenger-Miles (Alternative F)

Alternative F is the same as alternative E, except that in calculating potential passenger-miles, an adjusted measure of route length is used on loop routes. On loop routes—those that return to a main terminal by a substantially different path than the that taken when leaving that terminal—SCM defines route length as the length of the full loop rather than as the one-way distance between terminals. In alternative F, potential passenger-miles on loop routes were calculated using half the length of a loop as the route length.

It turns out that adjusting route length in this manner did not improve the correlation of potential passenger-miles with passenger-miles, as shown in table 1. Compared with alternative E, the correlation coefficient remained essentially unchanged while the *cv* of the auxiliary variable increased, resulting in an increased necessary sample size.

Combined Ratio Estimation Using Potential Passenger-Miles as the Auxiliary Variable (Alternative G)

The final alternative, alternative G, marries the two most efficient techniques found previously: ratio estimation using potential passenger-miles as the auxiliary variable, and stratification by route using the combined ratio estimation method.

Comparisons to this approach can be drawn against two other approaches: unstratified ratio estimation using potential passenger-miles as the auxiliary variable (alternative E), and combined ratio estimation using boardings as the auxiliary variable (alternative D). In alternative E, while both unstratified ratio estimation and combined ratio estimation involved a single, systemwide ratio, the stratification involved in the combined ratio method reduced inherent variability. In alternative D, the weakness was the large degree of variation in average passenger trip length between routes. When the auxiliary variable is potential passenger-miles, the ratio of interest becomes the fraction of route length covered by the average passenger trip, a ratio that does not vary nearly as much between routes.

Mathematically, alternative G is the same as alternative D, except that the auxiliary variable X represents potential passenger-miles. Again, we used proportional allocation between strata.

As indicated by table 1, alternative G turned out to be the most efficient, requiring a sample size of only 86. This represents a reduction of about 25% compared with alternatives D and E, confirming both the advantages of stratified sampling and of using combined ratio estimation with an auxiliary variable that varies little between routes.

Other Sampling Techniques

This overview of sampling techniques for estimating annual passenger-miles would not be complete without mentioning two other techniques that have been found to offer advantages.

Sampling Round Trips

The cost structure of on-off checks is such that it is almost always more efficient to sample round trips rather than independently selected trips: once a checker has surveyed a trip, the return trip can be sampled at nearly no additional cost because the checker usually has to be paid anyway to return to his or her starting point. Sampling round trips is an instance of cluster sampling, that is, selecting predefined clusters of, in this case, two trips for observation.

At transit agencies with labor agreements requiring eight-hour assignments for checkers, clusters lasting three to four hours are preferred, so that a checker can be assigned to one cluster in the morning and another in the afternoon. A cluster of this length is typically a chain of four, six, or eight trips performed by a single vehicle. The larger the cluster, the smaller the per-trip overhead related to getting to the start of the trip, supervision, and returning from the sampled trip.

However, when clusters tend to be homogeneous (which is certainly the case in this application, since the trips performed in a chain by a single vehicle are usually on the same route and take place during the same general time of day), variance per observed trip will be greater with cluster sampling than if trips are sampled independently (Cochran 1977). Therefore, the number of trips that would have to be observed to achieve a given precision using cluster sampling is greater than if trips are sampled independently. The *cluster effect* is defined as the ratio between these necessary sample sizes:

$$cluster\ effect = \frac{n_{cluster} * cluster\ size}{n_{SRS}} \qquad (25)$$

where n_{SRS} = necessary sample size in elementary units (e.g., one-way trips) using simple random sampling,

cluster size = number of elementary units per cluster, and

$n_{cluster}$ = necessary sample size (number of clusters) with cluster sampling.

In the literature (Cochran 1977), the cluster effect has been called Kish's d_{eff}, where d_{eff} stands for design effect.

The cluster effect can be used to convert a necessary sample size, obtained using a formula for simple random sampling, into a necessary sample size in units of clusters:

$$n_{cluster} = n_{SRS}\left(\frac{cluster\ effect}{cluster\ size}\right) \qquad (26)$$

A study of Los Angeles data (Furth et al. 1988) found that when sampling clusters of four trips to estimate the ratio of boardings to farebox revenue, the cluster effect was 2.2. Therefore, the number of four-trip clusters that would have to be observed is (2.2/4) = 55% as great as the number of one-way trips that would have to be observed if one-way trips were selected independently. Whether cluster sampling is cost-effective depends on if it is less expensive to do on-off checks on *n* trips selected independently or 0.55 *n* clusters of four trips.

A study of Madison, Wisconsin, data found that while sampling in round trip clusters was effective because of the small marginal cost of checking a return trip, sampling in larger clusters was not (Huang and Smith 1993). Our experience in analyzing cluster data for passenger-miles estimation from Dayton, Ohio, and Pittsburgh, Pennsylvania, confirms this finding. The larger the cluster, the greater the cluster effect, making clusters larger than a round trip rather ineffective as sampling units. Once a checker has measured passenger activity on a single round trip, little further information can be gained by sampling the next round trip operated by the same vehicle, since it will normally be operating on the same route and at the same period of the day. Large clusters are, therefore, recommended only when labor rules make it such that it costs little more to check multiple round trips than to check a single round trip.

To determine a sample size requirement using round trip clusters, it is often necessary to guess the magnitude of the cluster effect, because cluster data are rarely available for direct analysis. Experience suggests that a conservative estimate of the cluster effect is 1.5 for round trips. Using that value, equation (26) indicates that the number of round trips that would have to be sampled is 75% as great as the number of one-way trips that would have to be sampled using independent sampling. Therefore, if the cost of checking a round trip is no more than the cost of checking a one-way trip, sampling by round trip can reduce cost by 25% compared with sampling trips independently.

Two-Stage Sampling

In very small transit systems, the number of trips sampled over the year can approach the number of trips in the daily schedule. Sometimes, even in larger systems, the transit agency has a policy of checking every trip in the daily schedule once per year. If the finite population correction is accounted for, a two-stage design in which all (or most) of the trips in the daily schedule are observed eliminates all (or most) of the between-scheduled-trip variation (Cochran

1977). Because most of the variation in passenger-miles tends to be between scheduled trips (e.g., peak period versus offpeak) rather than between days for a given scheduled trip, such a two-stage approach can be quite efficient. This technique was demonstrated in a study done for the Los Angeles Blue Line light rail (Furth 1993) and has been applied in numerous bus systems as well.

COMPARISON AND CONCLUSION

Table 1 presents summary statistics for SCM comparing seven alternative sampling and estimation approaches using the Circular 2710.1A sampling plan. The key measure used to compare alternatives was the necessary sample size to meet the FTA precision criterion.

Circular 2710.1A, requiring a sample of 549 vehicle-trips, was the benchmark. This analysis found that it was a reasonable sample size to require, in the sense that passenger-miles variability at the trip level are such that, for most transit agencies, following it will achieve the FTA precision criterion. Alternative A used simple random sampling, where the only significant difference from the circular was that its sample size was based on a local *cv* of passenger-miles; for SCM, this alternative required almost as large a sample size (522) as Circular 2710.1A.

The remaining estimation techniques tested involved estimating a ratio between passenger-miles and an auxiliary variable where the annual total value is known. When boardings was the auxiliary variable, the ratio of interest was average passenger trip length. Compared with simple random sampling, this approach (alternative B) substantially improved efficiency, as the sampling requirement fell to 296.

When boardings were known by route, stratifying the population of trips by route length improved sampling efficiency, since average passenger trip length tended to vary systematically with route length, being greater on long routes and smaller on short routes. In alternative C, estimating the average passenger trip length ratio separately in four strata dropped the sampling need to only 117 trips. Making every route a stratum could further improve sampling efficiency; however, the constraint that ratio estimation be based on samples of at least 10 observations per stratum (in order to limit bias) makes it an impractical approach for an agency with a large number of routes.

The combined ratio method permits stratification by route in sample selection without concerns about bias, but it involves estimating a single, systemwide ratio. This technique, used to estimate the average passenger trip length ratio (alternative D), had the same sampling requirement as alternative C. It had the advantage that, unlike alternative C, it required only that an agency know annual system boardings, without requiring that annual boardings be known by route. The effectiveness of this technique was considerably greater for the transit agency in Kenosha, Wisconsin. Analysis of the technique suggests that its effectiveness will be greatest when the routes in a transit system vary little in length.

A new auxiliary variable was introduced, called potential passenger-miles, which is the product of boardings and route length. It performed better than boardings as an auxiliary variable. Without stratification (alternative E), it dropped the sampling requirement from 296 to 112; with stratification using combined ratio estimation (alternative G), it dropped the sampling requirement from 117 to 86. Attempts to use a modified definition of potential passenger-miles (alternative F) failed to improve efficiency. Alternative G, the most efficient approach, reduced the sampling burden by 84% compared with Circular 2710.1A. Conveniently, alternative G required only knowledge of system-level boardings, not route-level boardings.

While SCM's available dataset did not permit analysis of sampling trips in clusters, evidence from the literature and from unpublished studies indicates that sampling using round trip clusters improves cost-effectiveness, because a round trip cluster generally carries more statistical information than a single trip, while costing no more to observe due to the need of the checker to return to his or her starting point.

Evidence from only a few transit agencies is not sufficient to make a broad conclusion about the most efficient estimation and sampling method or sample size needed. Analysis of data from other transit agencies is needed to determine which results

are transferable. Nevertheless, the results of this study show a promising direction for any transit agency considering ways to reduce its passenger-miles sampling burden.

ACKNOWLEDGMENT

The author would like to acknowledge the careful and helpful reviews of three anonymous referees, whose comments guided the paper's revision.

REFERENCES

Cochran, W.G. 1977. *Sampling Techniques,* 3rd ed. New York, NY: John Wiley and Sons, Inc.

Furth, P.G. 1993. Ridership Sampling for Barrier-Free Light Rail. *Transportation Research Record* 1402:90–97.

____. 1998. Innovative Sampling Plans for Estimating Transit Passenger-Miles. *Transportation Research Record* 1618:87–95.

Furth, P.G., K.L. Killough, and G.F. Ruprecht. 1988. Cluster Sampling Techniques for Estimating Transit System Patronage. *Transportation Research Record* 1165:105–114.

Furth, P.G. and B. McCollom. 1987. Using Conversion Factors to Lower Transit Data Collection Costs. *Transportation Research Record* 1144:1–6.

Huang, W.J. and R.L. Smith. 1993. Development of Cost-Effective Sampling Plans for Section 15 and Operational Planning Ride Checks: Case Study for Madison, Wisconsin. *Transportation Research Record* 1402:82–89.

Townes, M., Chair of National Academy of Sciences Committee for the National Transit Database Study. 2001. Letter report to H. Walker, Acting Administrator of the Federal Transit Administration. June 1.

U.S. Department of Transportation (USDOT), Federal Transit Administration (FTA). 1990. *Sampling Procedures for Obtaining Fixed-Route Bus Operation Data Required Under the Section 15 Reporting System,* FTA Circular 2710.1A. Washington, DC. Also available at www.ntdprogram.com.

U.S. Department of Transportation (USDOT), Urban Mass Transit Administration (UMTA). 1985. *Revenue Based Sampling Procedures for Obtaining Fixed Route Bus Operation Data Required Under the Section 15 Reporting System,* UMTA Circular 2710.4. Washington, DC.

BOOK REVIEWS

Globalisation, Policy and Shipping: Fordism, Post-Fordism and the European Union Maritime Sector
Evangelia Selkou and Michael Roe
Edward Elgar Publishing
2004, 256 pages
ISBN 1-84376-934-4
$100 £59.95

This book contains nine chapters on a number of interrelated themes covering shipping policies, the European Union (EU), the impact of globalization, cohesion in European shipping policy, the case for tonnage tax, and neo- and post-Fordist developments in shipping policy. In a discussion of the impacts of globalization on the international shipping industry, the authors consider the role and relevance of national shipping policies and international bodies.

The book first examines policy objectives and structures and also illustrates the conflicts that can exist in policymaking. It then focuses on EU shipping policy and the different fiscal regimes applied to shipping, considering that the widespread adoption of tonnage taxation across the EU signifies an appreciation of shipping as a national asset. The final chapters discuss whether the changes in the maritime industry follow the series of structures recognized in other industries. In particular, chapter 8 identifies two partly contradictory tendencies in the industry, namely the neo-Fordist and post-Fordist (comparative advantage versus competitive advantage) dimensions, with close relationships to globalization in shipping.

I found the book to be well written and wide ranging, containing a wealth of references to the literature. It draws together the various strands of arguments, theories, and policies and weaves them into a composite picture that all students of shipping policy should appreciate. As such, this is a very welcome addition to the literature, particulary because it examines some of the core aspects of the globalized shipping arena and the extent to which global tendencies affect the formation of shipping policies. The authors recognize that maritime regulation must be international in nature (an observation that seems to have escaped some policymakers in the past) and illustrate their points with appropriate examples. In the final chapter, the authors point to some key shipping policy lessons.

The authors provide extensive references and one possible critique is that there are too many references (especially to Lloyd's List in chapter 7 which deals with the tonnage tax), but the points made are certainly well documented and justified. Copies of this book should definitely be held by university libraries and it could even be adopted as a textbook for some specialized shipping courses. This book is part of the series in *Transport Economics, Management and Policy* and it sits well in such company. At a price of £59.95, however, it might prove too expensive for individual students to purchase.

Reviewer address: Peter Marlow, Head, Logistics and Operations Management, Cardiff Business School, Colum Drive, Cardiff CF10 3EUm United Kingdom. Email: marlow@cardiff.ac.uk.

Competition and Ownership in Land Passenger Transport: Selected Refereed Papers from the 8th International Conference (Thredbo 8), Rio de Janeiro, September 2003
D.A. Hensher, editor
Elsevier
2005, 792 pages
ISBN 0-08-044580-2
$180 £110 €165

This very substantial volume, drawn from the 8th Thredbo conference, provides a wide range of papers dealing both with rail and road modes. The Thredbo series began in 1989 at the Australian mountain resort of that name and has subsequently been held every two years at a different location. It attracts a wide international audience, nowadays somewhat broader than the largely British, Australian, and American group at the first conference. Given the location of the conference,

the stronger contribution from South American authors is noteworthy.

Issues such as deregulation and competitive tendering have been a major factor since the initial impacts of local bus deregulation in Britain. A wider range of work, notably econometric studies, is now included. The volume is edited by Professor David Hensher of the University of Sydney, a co-founder of the series along with the late Professor Michael Beesley of the London Business School, to whose memory this volume is dedicated.

Individual papers are grouped into themes such as performance-based contracts, regulatory and planning tools, and performance data and measurement. Each of these served as the basis of a workshop in which intensive discussion took place, the main findings of which are summarized in a separate chapter. Selected papers from each workshop then follow. However, given the size of the volume, it is likely to be used as a reference work rather than read right through: for this purpose, a short abstract of each chapter would have been helpful.

Presentation is generally clear, although use of black and white print only means that some diagrams originally in color do not reproduce very well, and on some occasions references in the text to the color version are inconsistent with the version actually produced (e.g., figure 3.1 in chapter 10).

The issue of performance data and measurement has received greater attention than in previous conferences, being the focus of section 7 in this book. Chapters cover the current ownership structure in Britain (Charles Roberts), Brazilian railway privatization (Hostilio Neto), the Texas Governor's Business Council Performance Indicators for Urban Transport (Wendell Cox), efficiency changes in rail passenger operators since rail privatization in Britain (Jonathan Cowie), and bidding procedures for Brazilian urban bus systems (Alexandre Gomide).

Earlier work on deregulation and privatization used relatively simple performance indicators often aggregated at the whole operator or network level. Given the dramatic changes in productivity and costs—for example, through introducing competitive pricing into monopolistic systems—the outcomes were clear enough for measures such as cost per bus-kilometer run. However, with an increased focus on quality of service, more subtle indicators may have to be used, such as service reliability, patronage, and user attitudes. Even patronage may be measured only in crude terms and is dependent on operator ticketing systems' data. One consequence of deregulation and privatization is increased commercial confidentiality, resulting in detailed absolute figures not being readily available at the local level. Percentage changes may be quoted, such as those for ridership increases on "quality partnership" bus services cited by Roberts, but the absolute base from which such changes occurred is often unclear.

Chapter 13 by Berge, Brathen, Hauge, and Ohr offers useful examples of the wider range of performance measures now being used, in this case from Hordland County in Norway. These in turn are built into the contract mechanism, providing appropriate incentives for operators rather than simple cost minimization.

Contracts may also include other targets for performance, such as the revenue, volume, and accident levels set for the privatized Brazilian rail freight companies (although it can be presumed that the safety level figures are not targets in the same sense, but rather upper limits that companies would hope to fall within).

The sole U.S. contribution (chapter 41 by Wendell Cox) examines data from a wide range of countries for the respective roles of car and public transport. He also addresses the issue of how the greater use of public transport has the potential for avoiding road building to handle increased car flows. Very dramatic differences in urban densities, such as those between Hong Kong and U.S. cities, clearly correlate with different market shares for public transport modes. Cox suggests that the possibilities for modal diversion in expanding low-density U.S. cities are limited (except for some corridors into city centers), given the very high level of public transport services that would be needed. He indicates some limits to the higher density urban development now favored in some quarters, such as the greater volume of vehicle traffic per mile of road (despite the lower share taken by car), which may in turn result in congestion, lower speeds, and hence more pollution per vehicle-mile.

However, one must have reservations about some of the data and analysis presented by Cox. For example, figure 3 in his chapter is described as public transport vehicle-miles per square mile per annum. An average of 0.91 for western Europe seems implausibly low, unless what is meant is route-miles per square mile of area. This appears to be a simple error in stating the units used. He does not show *per capita* energy use nor pollution, which is generally far lower in high-density cities (as indicated in UITP's Millennium Database[1]) even if rates per vehicle-kilometer are higher. Having said this, given the existing low densities in the cities described (e.g., Houston), the scope for public transport shares on anything like the European level (let alone Asian) is obviously out of the question. However, the findings presented in this chapter may not be transferable to newly developed areas where the possibilities for higher density exist.

There is no doubt that the relevance of the topics covered in this conference series will remain strong, given continued interest in reducing costs and improving service quality in a wide range of countries. As Ian Wallis of Booz Allen and Hamilton (New Zealand) says in his review of developments in the main Australasian cities "…there should be plenty more to report on at Thedbo 9 [to be held shortly in Lisbon] and most likely at T10, T11, T12…."

Overall, this volume is undoubtedly of very great value as a work of reference, although given its size and cost it is likely to be a library acquisition rather than a personal one.

Reviewer address: Peter White, Professor, Transport Studies Group, University of Westminster, 35 Marylebone Road, London NW1 5LS, United Kingdom. Email: whitep1@wmin.ac.uk.

[1] The International Union of Public Transport (UITP), based in Brussels, compiles an extensive database of 100 large world cities. See http://www.uitp.com.

Industrial Location Economics
Philip McCann, editor
Edward Elgar Publishing Company
2002, 293 pages
ISBN 1-84064-672-1
$125 hardback ($50 paperback)

This must-read book will appeal to all students of the economics of location and the urban system including advanced undergraduates and graduate students, and faculty who work in the more general areas of planning and transportation, regional science, regional economic development, and location of economic activity. This is true if for no other reason than the first two chapters, by Phil McCann and John Parr, review the foundations of classical location theory (Weber and Moses; McCann, Losch, and Christaller; and Parr) and present more recent extensions and syntheses of the literature.

While those who teach industrial and firm location theory and practice are aware of most that is presented in these two chapters, others in the more general fields noted above will find these a useful way to update their background and learn about how the major research questions of today are being explored from this perspective. Such questions or topics include globalization, knowledge spillovers, urban and localization economies, agent-based perspectives, and complexity theoretic concepts. Throughout, we find McCann and Parr at their best when demonstrating close reasoning, thoughtful analysis, and reasoned but provocative insights. I have focused heavily on the McCann and Parr chapters here, because these two chapters by themselves make the book worth having in one's library. There is, however, much more of considerable interest and value.

The book is organized in three parts. Part I, Analytical Approaches to Industrial Location, in addition to the McCann and Parr chapters also considers location models from the perspective of the *new economic geography*, in chapter 3, and a review of firm migration literature with an assessment of prospective research for the future in chapter 4. Part II, Cities and Industrial Clusters, examines this topic of current interest in chapters 5 through 8. And Part III, consisting of chapters 9 through 11, examines the location behavior of multinational firms including

technology relationships between indigenous and foreign-owned firms (chapter 10).

Part I: Analytical Approaches to Industrial Location

While the McCann and Parr chapters have already been partly assessed, some additional notes will present a more complete picture. McCann observes there are important limitations to the applicability of the Weber-Moses framework. One is that market price or revenue of the output plays no role in the determination of the optimum location of the firm. Another is the strong role played in this analytical framework by transport costs when in fact these costs are for most firms a very small portion of total costs. He shows, however, that both of these limitations can be for the most part reconciled by substituting a total logistics cost variable for distance transport cost. This approach should be of particular interest to transport economists, planners, and practitioners. McCann goes on to show how it is possible to incorporate measures of the economies of distance and scale in the broader logistics reformulation.

The chapter by Parr identifies and deals with missing elements of the Central Place/Loschian urban systems perspective and places emphasis on issues related to innovation and knowledge spillovers in the context of localization, urbanization, and activity complex economies. He also emphasizes the need to develop or create a fuller understanding of how the urban system develops and evolves.

Dirk Stelder begins chapter 3 with the observation that "…a major empirical shortcoming of most NEG (new economic geography) models is their use of very abstract one-dimensional economic spaces like a circle or a horizontal line." By extending the approach to two dimensions Stelder shows that it is possible to produce complex hierarchical city distributions that approach reality to a significant extent. Of applied interest is a simulation of the European urban system that is accurate enough to be provocative. This work directly addresses Parr's concern about expanding our knowledge of the evolution of the urban system.

Chapter 4, by Piet Pellengarg, Leo van Wissen, and Jouke van Dijk, nicely reviews the literature of firm migration (relocation). At the outset, the authors recognize that firm relocation differs from firm location, because in the former case one location is substituted for another. In short, history is likely to influence the relocation decision and thus the outcome is conditional upon this history. This is important because it leads to the adoption of a stage approach to the study of the relocation decision: first the decision to move, then based on that comes the decision to relocate elsewhere. The relocation decision is also seen as different in that the focus is more on push factors rather than pull factors. The chapter provides three theoretical perspectives (neoclassical, behavioral, and institutional) that cover three major time periods: the 1970s, 1980s, and 1990s. The authors conclude that contemporary research focuses more heavily on the institutional nexus but the classical and behavioral approaches remain relevant and important. As a consequence, it is not possible to discuss or even formulate a general theory about the firm relocation decision at this time.

The firm relocation research shows distinct changes in motivation and orientation across the three decades leading up to the millennium. Rather than restate the various patterns of findings by the decade, it is perhaps more important to note here that much of the early literature was dominated by researchers in the United Kingdom. This persisted into the 1970s, because the United Kingdom's regional policy at that time focused on steering manufacturing industry to assisted areas by using such policy instruments as location controls and capital and workforce subsidies. Thus, the firm relocation studies of this era were part of a large bundle of studies aimed at determining the effect of these policy instruments on the economies of assisted areas. In the late 1970s and early 1980s, U.S. research began to appear in the literature along with significant publications from the Netherlands, Germany, France, and Italy.

The number of international firm migration studies decreased considerably in the 1980s. With this, the emphasis of the research changed to more of a focus on the relation of firm migration to urban decay and decline and policies designed to drive

renewal, and to hold or attract companies to such areas. Firm relocation research in the 1990s reflects the rise of the information and communications technologies sector and supporting services as emphasis on the policy side of this literature focused increasingly on the creation of innovative environments and new enterprises and industries. The chapter concludes with a discussion of modeling approaches for studying firm migration and how to operationalize various explanatory factors such as those both internal and external (context) to the firm and more specifically site- and situation-related forces. Throughout, the authors emphasize the importance and need for further research to reveal the role that the product life cycle plays in the relocation decision.

Part II: Cities and Industrial Clusters

Considerable interest exists in bringing more clarity to the industrial cluster literature and to the concepts that underlie its magnetism for both scholars and practitioners. And while the chapters in this part of the book certainly add to the debate and provide some clarification, they fall short of providing insight into best practices at both the methodological level and at the level of practice. But this is probably too harsh a view given that all research in this area remains subject to the need for considerable clarification, codification, and extension.

In chapter 5, Gilles Duranton and Diego Puga examine the concepts of diversity and specialization in cities. Probably the most important contribution of this chapter is the presentation of a set of stylized facts about the subject: for example, "specialized and diversified cities coexist," "larger cities tend to be more diversified," "the distribution of many urban system parameters such as relative city sizes are stable over time," "cities are increasingly specialized by function," and so on. These facts are useful in two ways. In their own right they offer a set of working hypotheses that already enjoy a fair amount of support. But they also create a template for assessing various urban theories from both static and dynamic perspectives and for advising policy. From a policy perspective, this leads to conclusions like: "…the link between innovation and diversity seems fairly robust…thus highly innovative clusters cannot be bred in previously highly specialized environments." Such observations if not fully vetted at least formalize patterns that many of us who have worked in this arena recognize and thus begin to lay the foundation for methodological and, perhaps, a more conceptual synthesis.

In chapter 6, Ian Gordon examines the issues of global cities, internationalization, and urban systems. He begins by observing that there has been a progressive internationalization of relationships at all levels and the revaluation of the advantages of urban agglomeration, especially core cities. Gordon also recognizes there is a good bit of "fuzziness" in the use of the concept of world or global cities, while effectively avoiding getting bogged down in an extended assessment to confirm this claim. He simply moves on to the main concern of the chapter—*globalcityization* and its relation to location, transportation, and trade functions within urban systems. Further analysis results in a conclusion that undue emphasis has been placed on the notion that global or world cities play an inordinate role as dominant nodes in the development and operation of the global networked economy. He suggests that an "…overstrong focus on the significance of the global city role can obscure the responsibility of other (often more traditional or older) factors for positive and negative developments in the city." He also argues that emphasis on the global city role "…reflects one particular set of interests from within a diverse economy, exaggerating the extent to which these are crucial to the wider economy." As such, this chapter presents an argument that is at odds with the prevailing notion of the importance of global cities in the operation of national urban systems and, therefore, national and global economic systems.

In chapter 7, Michael Steiner presents an institutional dynamics view of innovation and regional development as a framework for arguing the necessity of industrial clusters in new market economies and developing countries, especially those that have adopted institution-liberalizing policies. Yet these countries are often limited in their ability to guarantee the conditions required to fulfill the process of transformation and thus to use spontaneous market responses as strong or dominant drivers of development. Steiner extends this conclusion to apply as

well to the use of spontaneous cluster development as a driver. In short, a stronger regulatory or interventionist policy is necessary for steering industrial cluster development in the ascending and developing countries. He concludes that membership in the European Union (EU) and similar multinational organizations could serve as a guarantor in the external institutional building process and thus a way to dampen the need for strong intervention.

The final chapter of Part II, written by Edward Feser and Stuart Sweeney, assesses theory and methods used for comparing cross metropolitan business clustering. They recognize that the urban system is always in a state of flux, and this will impact dispersal or concentration tendencies in different ways in different industries and thus affect clustering behavior. In response, they develop a new analytical methodology for cluster analysis based on point process models that utilizes data for establishments by location. They go on to apply the methodology to 14 U.S. metropolitan areas and find considerable variation in clustering among the industries of a common value chain (cluster of related industries)—manufacturing. They, for example, find that the strongest intra-regional clustering occurs in "…paper and publishing and textiles/apparel…value…chains, and to a lesser extent electronics and computers, aerospace, canned goods and grain mill products." Clustering is weak for wood products and furniture, vehicle manufacturing, and chemicals. This interesting research is of considerable value in that it offers a direction for addressing an issue mentioned earlier in the review, namely the need for creating synthesis and agreement on methods and concepts that underlie the industrial cluster research agenda.

Part III: Multinational Firms and Location

Chapter 9 contains an essay by Ram Mudambi on location decisions of multinational enterprises (MNEs). He notes that such decisions have for the most part been considered in the context of the *eclectic paradigm* after Dunning, which provides a unifying framework for "…determining the extent and pattern of foreign-owned activities." He continues with the conclusion that three sets of forces or advantages drive MNE activities. These have to do with ownership of the enterprise, location of operations, and the internalization of advantages, or OLI. The ownership advantages come from the resource base owned or controlled by the firm, while the location advantages come from resources, networks, institutional structures, and so forth to a geographic entity that are immovable. Internalization brings activities that create transaction costs inside the structure of the firm.

The chapter presents literature on OLI in the context of a view that the multinational location decision can be modeled as a two-person game between the MNE and the host government. The analysis begins with the literature on the MNE location decision from 1945 to about 1980, a period dominated by market-oriented advantages that occurred under the "suspicious eye" of host governments. However, as globalization unfolded in the 1980s and subsequent liberalized approaches to development were more commonly adopted, a more interactive perspective between the MNE and the host government occurred. Then subsidiaries became much more linked to the MNE's international network, especially for mature MNEs. In this later period, life cycle effects of the firm became more important. This is viewed as mutually advantageous to both the MNE and the host government, because knowledge and technology spillovers within the firm and to other indigenous firms become more likely where such mature multinationals operate.

This issue of technology spillovers is central to the substance of chapter 10 by John Cantwell and Simona Iammarino. They observe that MNEs derive technology complementarities between related paths of innovation or corporate learning in distinct geographic or country settings. In this fashion, it is possible to spread the resulting competence base of the firm across its subsidiaries and, thus, more efficiently spread its technology assets. From this fact of MNE operations, the literature has considered the hypothesis that indigenous firms in the host country thus benefit from so-called technology spillovers of the MNE locations there. The authors next focus on the EU context arguing that the high level of cross national institutional congruence provides the most likely or strongest context for this hypothesis to thrive. Some support is found and a discussion of the barriers to generalizing the findings ensues.

The final chapter, by Tomokazu Arita and Philip McCann, examines the relationship between the spatial and hierarchical organization of multiplant firms through consideration of the global semiconductor industry. They observe that the phenomena of Silicon Valley and other renowned regional technology centers like Cambridge (England), Austin (Texas), Bangalore (India), and so on are only a small piece of the semiconductor industry and the small innovative nature of the economic structure of those economies can be and are misleading.

To examine this seemingly provocative hypothesis, Arita and McCann undertake an analysis of the structure of the semiconductor industries in Japan and the United States. They find that, like most industries, this industry with its forward and backward linkages is organized oligopolistically or nearly so. In Japan, the organizational structure is strongly vertical in nature and in a *keiretsu* style fashion, and nearly all of the R&D and semiconductor production, including basic and intermediate inputs, is located in Japan. In the United States, only a few major firms are producing semiconductors, but the supply chain is more diverse with significant outsourcing used to feed semiconductor production in plants based in the United States and elsewhere. Getting back to the point of the analysis, the authors conclude that the semiconductor industry is organized much like other intermediate to mature industries. Furthermore, only a small part tends to be organized in a network of small- to medium-sized enterprises (SMEs) like that found in Silicon Valley and the other regional technology centers cited above. While one might argue that this may be viewed as mixing oranges and apples, it focuses attention on the fact that the information technology industry is more than just the businesses of technology-intense regional economies. In fact, its backbone is organized much the same as other industries at the intermediate to mature stage of the life cycle.

Conclusions

This is an excellent book and for an edited volume it does a nice job of staying on message. The book's strong point is the several literature review-type chapters. These include chapters 2 and 3 on classical location theory, extensions, synthesis, and directions for future research, which are gems. Chapter 4 on firm migration or relocation does a nice job of covering the essentials of the literature in an historical context and provides an excellent guide to pressing research questions.

While Part II on cities and industrial clusters does not provide a literature review of either urban research or industrial clusters, it presents much insight into the thinking on localization and urban economies, concepts underlying the renewed interest in industrial cluster analysis and in particular in an urban context. Finally, chapter 9 provides a good assessment of the literature on the MNE location decision and the role of foreign direct investment. This review also provides historical perspective in addressing the major research questions and leaves the reader with a good sense of the engaging contemporary research questions.

Beyond the review-type essays, the book offers an introduction to interesting topics and to the frontiers of related research. For example, chapter 3 extends the classical perspectives on industrial location to agent-based modeling in the context of the new economic geography. All four chapters in Part II on cities and industrial clustering are of considerable interest ranging from Ian Gordon's examination of the idea that global cities play a superior nodal role in the global economic system to Feser and Sweeney's new methodological approach to the cross metropolitan comparison of industrial clusters.

In sum, the book makes a considerable contribution to the literature on industrial location economics. It comes at a time when many classical perspectives are both under attack and have undergone recent modification. Thus the review essays are very timely and the special topic chapters add flavor and perspective.

Reviewer address: Roger R. Stough, NOVA Endowed Chair and Professor of Public Policy, Associate Dean for Research, Development, and External Relations, George Mason School of Public Policy, 4400 University Drive, MS 3C6, Fairfax, VA 22030-4444, USA. Email: rstough@gmu.edu.

Transportation Services Index

INTRODUCTION

In 2002, a team of academics, under a research grant from the Bureau of Transportation Statistics (BTS), developed the Transportation Services Index (TSI) to measure the state of the transportation sector and its contribution to the economy. The TSI, which consists of three seasonally adjusted monthly indexes, reflects the changes in the output of services for the passenger, freight, and total transportation sector (figure 1 and tables 1 and 2). The *Journal of Transportation and Statistics* published the results of the research project in 2003 (Lahiri et al. 2003).

CALCULATION OF INDEX

BTS calculates three monthly transportation output indexes: the total TSI, the freight TSI, and the passenger TSI. Collectively, the three indexes use eight sources of component data that reflect output in terms of passenger-miles or ton-miles (or some proxy thereof).

The TSI freight index, calculated using tons or ton-miles depending on the modal data source, includes for-hire trucking, freight railroad services (including intermodal shipments), inland waterway traffic, pipeline movements, and air freight. The freight index does not include international or coastal steamship movements, private trucking, courier services, or the United States Postal Service.

The TSI passenger index, calculated using passenger-miles, includes local mass transit, intercity passenger rail, and passenger air transportation. The passenger index does not include intercity bus, sightseeing services, taxi service, private automobile usage, or bicycling and other nonmotorized forms means of transportation.

The process of converting the component data into the final indexes includes forecasting missing values, deseasonalizing raw data, indexing values, weighting, and chaining final modal estimates.

- Forecasting is used to estimate data that are not reported in a timely manner.
- All data used in the TSI are deseasonalized to remove the impact of seasonal patterns. Transportation data are highly seasonal, masking true changes from month to month and long-term changes through the years.
- All modal data are indexed to a base year, currently set to 2000.
- Indexed modal data are weighted based on the value added by each mode to the transportation service sector of Gross Domestic Product (GDP), provided by the Bureau of Economic Analysis in the National Income Product Accounts.
- Using the chained Fisher Ideal Index and the GDP weights, the modal data are aggregated into the three composite indexes.

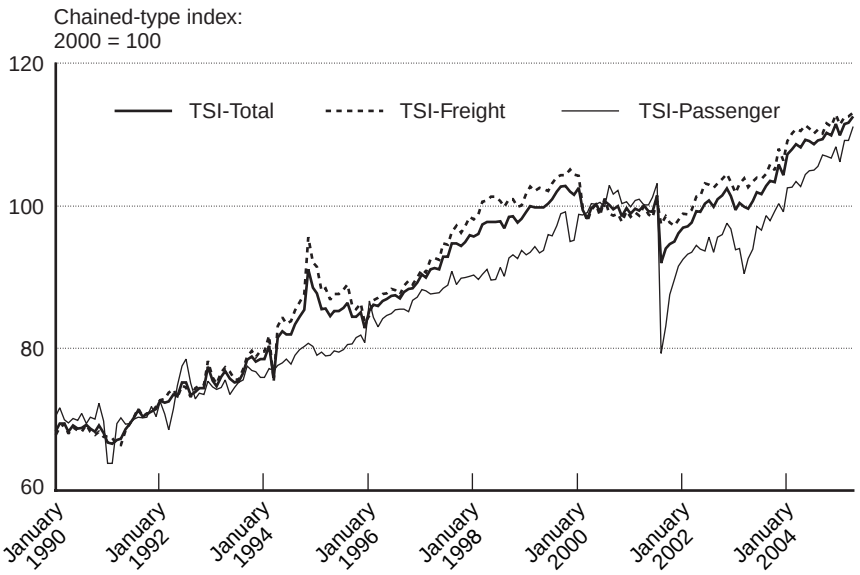
DATA TRENDS

Since its inception, the TSI has experienced an overall upward trend. During 1990, the TSI's first year of historical data, the total TSI ranged from 67.6 to 69.4. In comparison, during 2004, the total TSI ranged from 104.3 to 110.3.[1] The freight TSI and

[1] All numbers are indexed to a base year of 2000 (2000 = 100).

Review by: Jennifer Brady, Analyst, Bureau of Transportation Statistics, Research and Innovative Technology Administration, U.S. Department of Transportation, 400 Seventh St. SW, Room 3430, Washington, DC 20590. Email address: jennifer.brady@dot.gov.

FIGURE 1 Transportation Services Index: Total, Freight, and Passenger
January 1990–May 2005: seasonally adjusted



Source: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics.

TABLE 1 Percentage Change in the Transportation Services Index by Month Since January 2005
Seasonally adjusted, monthly average of 2000 = 100

	TSI		Freight		Passenger	
	Index	% change	Index	% change	Index	% change
January	111.6	1.5	112.8	1.4	108.4	1.5
February	110.0	–1.4	111.5	–1.2	106.2	–2.0
March	111.6	1.5	112.5	1.0	109.3	2.9
April[R]	111.6	0.0	112.5	0.0	109.2	–0.1
May[P]	112.6	0.9	113.1	0.6	111.2	1.8
June[P]	112.0	–0.6	112.6	–0.5	110.3	–0.8
July[P]	111.6	–0.3	111.9	–0.6	110.8	0.5

Note: P = preliminary; R = revised.
Source: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics.

passenger TSI, as components of the total TSI, experienced similar growth.

Minor directional changes in the TSI are frequent, with no period of more than three consecutive months of decline over the history of the index. Sustained and larger-than-normal change in the trend of the index occurs around the time of major events. Most noticeably, the events of September 11, 2001, caused a major change in all three indexes. Between August and September 2001, the total TSI declined 9.3%, the freight TSI declined 3.3%, and the passenger TSI declined 23.2%. However, the total and freight TSI returned to pre-September 2001 levels within a year; the passenger index took several years to return to pre-September 2001 levels.

BTS is currently researching the relationship of the TSI to the general economy to ascertain if the cycle of transportation output turns at about the same time as the economy turns (which would make this a coincident index) or if transportation turns in

TABLE 2 Percentage Change in the Transportation Services Index from Year-to-Year
May TSI (monthly average of 2000 = 100)

Year	TSI	Change from same month previous year (%)	Freight TSI	Change from same month previous year (%)	Passenger TSI	Change from same month previous year (%)
1996	86.7	2.6	87.6	0.9	84.2	6.5
1997	91.3	5.3	92.7	5.8	87.7	4.2
1998	97.8	7.2	100.7	8.7	91.1	3.9
1999	99.8	2.0	102.6	1.8	93.4	2.5
2000	99.7	–0.1	99.5	–3.0	100.3	7.4
2001	100.0	0.3	99.9	0.5	100.2	–0.2
2002	99.3	–0.8	101.3	1.4	94.4	–5.7
2003	99.7	0.4	102.7	1.3	92.6	–1.9
2004	108.3	8.6	110.5	7.7	102.8	11.0
2005[P]	112.6	4.0	113.1	2.4	111.2	8.2

Note: P = preliminary.
Source: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics.

advance of the turns in the economy (a leading index). However, due to the limited length of the TSI data series, it has only been possible to observe the TSI during one entire recession.[2]

RELEASE OF TSI

BTS recognized the benefits of producing the TSI on a monthly basis and released the first monthly estimates in March 2004, with an historic data series beginning January 1990. Currently, TSI is the only combined, multimodal, seasonally adjusted measure of transportation services made available on a monthly basis.

On approximately the 6th of each month, new and updated TSI numbers are released on the BTS website. The newest three months are preliminary numbers. Each month BTS releases the latest preliminary TSI, and replaces the oldest preliminary TSI with a revised TSI. After a number is revised, it will not change until the annual update is released in mid-year when most of the prior year data have been finalized.

To obtain the most current TSI data, as well as further documentation on the indexes, visit the BTS website: http://www.bts.gov/xml/tsi/src/index.xml.

REFERENCE

Lahiri, K., W. Yao, H. Stekler, and P. Young. 2003. Monthly Output Index for the U.S. Transportation Sector. *Journal of Transportation and Statistics* 6(2/3):1–27.

[2] The National Bureau of Economic Research designated March 2001 through November 2001 a period of recession.

New ASA Section on Transportation Statistics Seeks Members

This notice announces the tentative formation of the Section on Transportation Statistics of the American Statistical Association (ASA). The principal objective of this new section is to serve ASA members with special interests in: 1) developing and applying statistical methods to problems in transportation, 2) analyzing transportation data, 3) collecting transportation data, and 4) formulating mathematical models, whether deterministic or stochastic, which describe and explain underlying mechanisms and modes of action of fundamental processes in transportation. This section will cooperate with the Transportation Research Board and other professional organizations in order to sponsor joint meetings and sessions at professional meetings.

Contacts for the Section are Prem Goel (goel@stat.ohio-state.edu), Mike Griffith (mike.griffith@fhwa.dot.gov), Cliff Spiegelman (cliff@stat.tamu.edu), and me. Under the ASA Constitution, we need an expression of interest from 100 or more ASA members to form the Section so that we can provide the ASA Council with our initial mailing list of Section members. A draft of the Section charter is available and will be emailed on request. This announcement is for the purpose of adding names to our initial mailing list of founding members and for obtaining suggestions and comments.

All section members must be willing to pay an annual section fee of $5. This charge will appear on your next ASA membership dues statement, in the same way that charges for other section memberships appear. The first 100 persons who respond to this solicitation will be listed as founding members on the new section's website.

If you are willing to become a founding member, please send me an email with the following statement:

I, (insert your name), support the petition to form a new ASA section to be called the Section on Transportation Statistics.

Thank you very much for your help in establishing what we believe will be an active new section with a highly important mission in today's society and economy. If you have any questions or suggestions, please contact me.

Promod Chandhok, Ph.D.
Chair, ASA Interest Group on Transportation Statistics

promod.chandhok@dot.gov
Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
Washington, DC 20590
and
The George Washington University
Washington, DC

Guidelines for Manuscript Submission

Please note: Submission of a paper indicates the author's intention to publish in the *Journal of Transportation and Statistics* (JTS). Submission of a manuscript to other journals is unacceptable. Previously published manuscripts, whether in an exact or approximate form, cannot be accepted. Check with the Managing Editor if in doubt.

Scope of JTS: JTS publishes original research using planning, engineering, statistical, and economic analysis to improve public and private mobility and safety in all modes of transportation. For more detailed information, see the Call for Papers on page 112.

Manuscripts must be double spaced, including quotations, abstract, reference section, and any notes. All figures and tables should appear at the end of the manuscript with each one on a separate page. Do not embed them in your manuscript.

Because the JTS audience works in diverse fields, **please define** terms that are specific to your area of expertise.

Electronic submissions via email to the Managing Editor are strongly encouraged. We accept PDF, Word, Excel, and Adobe Illustrator files. If you cannot send your submission via email, you may send a CD by overnight delivery service or send a hardcopy by the U.S. Postal Service (regular mail; see below). Do not send CDs through regular mail.

Hardcopy submissions delivered to BTS by the U.S. Postal service are irradiated. Do not include a disk in your envelope; the high heat will damage it.

The cover page of your manuscript must include the title, author name(s) and affiliations, and the telephone number and surface and email addresses of all authors.

Put the **Abstract** on the second page. It should be about 100 words and briefly describe the contents of the paper including the mode or modes of transportation, the research method, and the key results and/or conclusions. Please include a list of **Keywords** to describe your article.

Graphic elements (figures and tables) must be called out in the text. Graphic elements must be in black ink. We will accept graphics in color only in rare circumstances.

References follow the style outlined in the *Chicago Manual of Style*. All non-original material must be sourced.

International papers are encouraged, but please be sure to have your paper edited by someone whose first language is English and who knows your research area.

Accepted papers must be submitted electronically in addition to a hardcopy (see above for information on electronic submissions). Make sure the hardcopy corresponds to the electronic version.

Page proofs: As the publication date nears, authors will be required to proofread and return article page proofs to the Managing Editor within 48 hours of receipt.

Acceptable software for text and equations is limited to Word and LaTeX. Data behind all figures, maps, and charts must be provided in Excel, Word, or DeltaGraph (unless data are proprietary). American Standard Code for Information Interchange (ASCII) text will be accepted but is less desirable. Acceptable software for graphic elements is limited to Excel, DeltaGraph, or Adobe Illustrator. If other software is used, the file supplied must have an .eps or .pdf extension. We do not accept PowerPoint.

Maps are accepted in a variety of Geographic Information System (GIS) programs. Files using .eps or .pdf extensions are preferred. If this is not possible, please contact the Managing Editor. Send your files via email or on a CD via overnight delivery service.

Send all submission materials to:

Marsha Fenn, Managing Editor
Journal of Transportation and Statistics
BTS/RITA/USDOT
400 7th Street, SW, Room 4117
Washington, DC 20590
Email: marsha.fenn@dot.gov

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 8 Number 2, 2005
ISSN 1094-8848

CONTENTS