Volume 5 Number 1, 2002
ISSN 1094-8848

BUREAU OF TRANSPORTATION STATISTICS UNITED STATES DEPARTMENT OF TRANSPORTATION

JOURNAL OF TRANSPORTATION AND STATISTICS

JOURNAL OF TRANSPORTATION AND STATISTICS

JOHN V. WELLS Acting Editor-in-Chief
SUSAN LAPHAM Associate Editor
PEG YOUNG Associate Editor
MARSHA FENN Managing Editor
DORINDA EDMONDSON Desktop Publisher
CHIP MOORE Editor
MARTHA COURTNEY Editor
DARCY HERMAN Editor
LORISA SMITH Desktop Publisher

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 5 Number 1, 2002
ISSN 1094-8848

BUREAU OF TRANSPORTATION STATISTICS UNITED STATES DEPARTMENT OF TRANSPORTATION

Bureau of Transportation Statistics

Our mission: To lead in developing transportation data and information of high quality and to advance their effective use in both public and private transportation decisionmaking.

Our vision for the future: Data and information of high quality supporting every significant transportation policy decision, thus advancing the quality of life and the economic well-being of all Americans.

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

Bureau of Transportation Statistics
U.S. Department of Transportation
Room 7412
400 7th Street SW
Washington, DC 20590
USA
journal@bts.gov

Subscription information

To receive a complimentary subscription or a BTS Product Catalog:

mail Product Orders
Bureau of Transportation Statistics
U.S. Department of Transportation
Room 7412
400 7th Street SW
Washington, DC 20590
USA
phone 202.366.DATA [press 1]
fax 202.366.3197
internet www.bts.gov

Information Service
email answers@bts.gov
phone 800.853.1351

Cover and text design Susan JZ Hoffmeyer
Cover photo © PhotoLink/Getty Images

NOTE:
As of this issue, JTS will no longer include a month in the publication information.

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 5 Number 1
2002

Contents

Statistically-Based Validation of Computer Simulation Models in Traffic Operations and Management

JEROME SACKS
National Institute of Statistical Sciences
and Duke University

NAGUI M. ROUPHAIL
North Carolina State University

BYUNGKYU (BRIAN) PARK
University of Virginia

PIYUSHIMITA (VONU) THAKURIAH
University of Illinois at Chicago

ABSTRACT

The process of model validation is crucial for the use of computer simulation models in transportation policy, planning, and operations. This article lays out obstacles and issues involved in performing a validation. We describe a general process that emphasizes five essential ingredients for validation: *context*, *data*, *uncertainty*, *feedback*, and *prediction*. We use a test bed to generate specific (and general) questions as well as to give concrete form to answers and to the methods used in providing them.

The traffic simulation model CORSIM serves as the test bed; we apply it to assess signal-timing plans on a street network of Chicago. The validation process applied in the test bed demonstrates how well CORSIM can reproduce field conditions, identifies flaws in the model, and shows how well CORSIM predicts performance under new (untried) signal conditions. We find that CORSIM, though imperfect, is effective with some restrictions in evaluating signal plans on urban networks.

INTRODUCTION

The validation of computer simulation models is a crucial element in assessing their value in transportation policy, planning, and operational decisionmaking. Often discussed and sometimes

Jerome Sacks, National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4006. Email: sacks@niss.org.

informally practiced, the process is conceptually straightforward. Data representing both the input and the output of the model are collected, the model is run with that input, and the output is compared to field data. In reality, complications abound: field data may be expensive, scarce, or noisy; the model may be so complex that only a few runs are possible; and uncertainty enters the process at every turn. Even though it is inherently a statistical issue, model validation lacks a unifying statistical framework.

The need to develop such a framework is compelling, even urgent. The use of computer models by transportation engineers and planners is growing. Costs of poor decisions are escalating, and increasing computing power, for both computation and data collection, is magnifying the scale of the issues. The opportunity is as great as the need. Advances in statistical techniques for incorporating multiple types of information, while managing the multiple uncertainties, enable progress in quantifying validation (Berliner et al. 1999; Lynn et al. 1998).

The purpose of this paper is to set out key issues faced in the validation of transportation models and to advance a research effort to address these issues. Many of the issues we describe are common to models and modelers in all areas of science and engineering:

- give explicit meaning to validation in particular *contexts*
- acquire relevant *data*
- quantify *uncertainties*
- provide *feedback* to model use and development
- *predict* performance under new (untried) conditions

While easily outlined, the challenge is to meet these issues. This can be achieved by describing and developing approaches and methods that are effective and can be implemented. That there are many obstacles to surmount is no surprise to those who have attempted exacting validations. However, there are tools capable of overcoming the impediments.

In order to make our points clear, we will use a test bed that generates the questions a validation must address and, at the same time, accommodates analyses that respond to the main issues. The test bed we use is the microscopic simulator CORSIM in an application to the assessment and selection of signal timing plans on an important street network in Chicago, Illinois.

Several research issues emerge from this investigation, indicating the following needs:

- to formulate evaluation functions that capture transportation needs and are amenable to either direct or indirect observation in the field
- to measure and assess the impact of data quality on evaluation functions and performance
- to develop methods for treating a variety of problems connected with the analysis of uncertainties, especially predictions

The general conclusion from the test bed is that, despite imperfections, CORSIM is effective as a model for evaluating signal plans on urban street networks under some restrictions. The basis of the statement is the validity of CORSIM prediction of performance under new conditions assessed by a second data collection, the gold standard of validation. The simplicity of the conclusion belies the complexity of the process, particularly evident in the feedback step of tuning the model to the specific network using an initial data collection.

We introduce the test bed example and simulator in the second section, along with the specific evaluation functions we use. Acquisition of data and the two field collections are described in the third section. Estimation of the input to the model is described in the fourth section. The fifth section covers the range of validation questions and the analyses relevant to them, including tuning, based on the initial data collection. The next section discusses the prediction of performance under new conditions and the subsequent validation. Questions about uncertainty are discussed in the following section, and our conclusions appear in the final section.

THE TEST BED: CORSIM AND SIGNAL TIMING ON AN URBAN STREET NETWORK

CORSIM is a computer simulation model of street and highway traffic. It is the quasi-official platform used by the U.S. Department of Transportation (USDOT) to gauge traffic behavior and compare competing strategies for signal control before implementing them in the field (USDOT FHWA 1996).[1] For CORSIM to fulfill this purpose, two crucial questions must be addressed.

1. How well does CORSIM reproduce field conditions?
2. Can CORSIM be trusted to represent reality under new, untried conditions, such as revised signal timing plans?

The localized and complex behavior that signal plans induce on urban street networks makes answering these two questions a challenge. Flows on these networks, even on small sub-networks, are highly complex. They include a variety of vehicles, pedestrian-vehicle interactions, and driver behavior, as well as an assortment of network conditions, such as different lane arrangements, stop signs, parking lots, and one-way streets. Moreover, the traffic demands on the network are highly variable, changing month to month, day to day, hour to hour, and even minute to minute. Equally varied are the many movements (legal and otherwise) of vehicles and pedestrians.

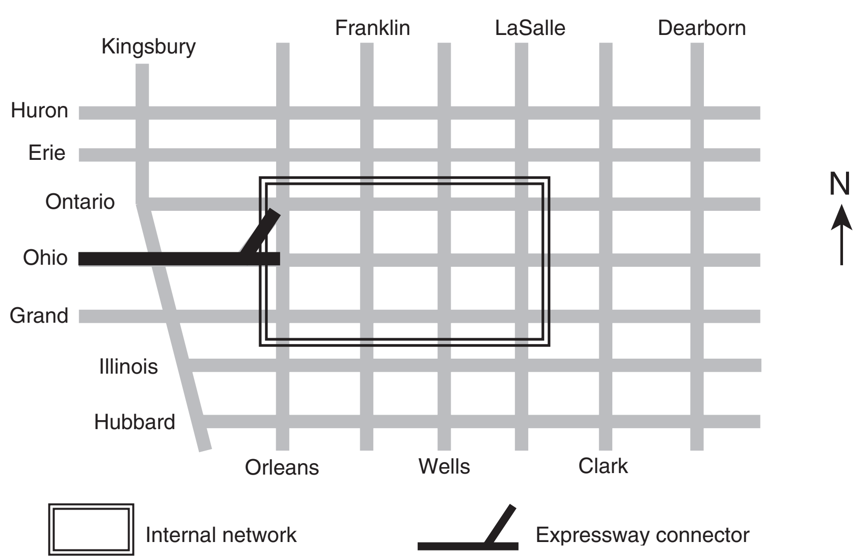
Since no simulator can realistically capture behavior exactly, formulating appropriate performance measures or evaluation functions is fundamental to the validation process. Variability, inherent in real traffic and also present in the computer model, compounds matters. Choices of performance measures introduce subjective elements and, thereby, potential sources of contention in assessment of the computer model.

[1] CORSIM version 4.32 is used in this paper.

To focus the issues, we undertook a case study with the cooperation of the Chicago Department of Transportation (CDOT) with the ultimate goal of optimizing the signal plans for a network more extensive than the one here. The test bed for the study is the network depicted in figure 1. The internal network (Orleans to LaSalle; Ontario to Grand) in figure 1 is the key part of a planned Real-Time Traffic Adaptive Control System (RT-TRACS) study to be carried out in the future. A different network was studied earlier (Park et al. 2001) and helped guide some of the decisions made in the current test bed.

Traffic in the network depicted in figure 1 flows generally south and east during the morning peak and north and west in the evening peak. This demand pattern is accommodated by a series of high-capacity, one-way arterials such as Ohio (eastbound), Ontario (westbound), Dearborn (northbound) and Clark and Wells (southbound), in addition to LaSalle (north- and southbound). For reference purposes, the Chicago central business district (CBD) is located southeast of the network.

FIGURE 1 Test Bed Network



CORSIM Characteristics and Inputs

CORSIM is a microscopic and stochastic simulator. It represents single vehicles entering the road network at random times moving (randomly) second-by-second according to local interaction rules that describe governing phenomena such as car-following logic (rules for maintaining safe distances between cars), lane changing, response to traffic control devices, and turning at intersections according to prescribed probabilities. CORSIM can handle networks of up to 500 nodes and 1,000 links containing up to 20,000 vehicles at one time. The figure 1 network has 112 1-way links, 30 signalized intersections, and about 38,000 vehicles moving through it in 1 hour. Streets are modeled as directed links with intersections as nodes.

There are a variety of inputs or specifications that must be made, either directly or by default values provided in CORSIM. Input that must be made directly include the following.

- specification of the network via fixed inputs describing the geometry, such as distance between intersections, number of traffic lanes, and length of turn pockets; the placement of stop signs, bus stops, schedules, and routes; and parking conditions
- probability distributions of interarrival times governing the generation of vehicles at each entry node of the network; the choices in CORSIM of arrival-time distributions are limited, in essence, to Gamma (Erlang) densities, λ = average interarrival time or $1/\lambda$ is the expected number of vehicles arriving in 1 second; k determines the shape of the Gamma density.

$$p(t \mid \lambda, k) = \frac{(\lambda k)^k}{(k-1)!} t^{k-1} \exp(-k\lambda t), \tag{1}$$

  assumed independent (vehicle-to-vehicle, node-to-node) but allowed to be different for each entry node
- vehicle mix (auto or truck) through independent Bernoulli trials with probabilities that can differ from entry node to entry node
- probability distributions of turning movements, assumed to be independent, vehicle-to-vehicle and link-to-link and different from link-to-link

CORSIM provides several default inputs. The chief inputs relate to driver characteristics, such as car-following behavior (how closely drivers follow other vehicles), left turn "jumpers" (drivers who "jump the gun" ahead of oncoming traffic), acceptance of gaps between vehicles (before making turns or lane changes), and lane-changing maneuvers. For example, gap acceptance is governed by a discrete distribution with 10 mass points. The default distribution can be accepted or altered. Other inputs with default distributions that can be altered are dwell times for buses, effects of pedestrians on turning vehicles, and short-term incidents, such as an illegally parked delivery truck.

Although altering the default distributions through data use is possible in some cases, data that would better determine driver characteristics are too elusive. For the test bed study, we assumed no pedestrian traffic (normally light on this network) and no incidents.

Signal settings are direct inputs. We single them out as controllable factors since altering these inputs to produce improved traffic flow drives the study. Signal settings consist of a cycle common to all signals, green times for movements at each intersection, and offsets (time differences between beginnings of cycles at intersections).

For validation, the signal plan will be the one in the field. For finding optimal fixed-time signal-timing plans[2] or for comparing alternative plans, the signal parameters will necessarily be manipulated. Comparisons are best done through the simulator since field experiments are not feasible. Relying on CORSIM to select an alternative to an in-place plan then raises our earlier-posed questions.

CORSIM Output

CORSIM comes equipped with an animation package (TRAFVU) allowing visualization of traffic movements, valuable when exploring the characteristics of the model and detecting problems and flaws. In addition to the visual output, CORSIM provides aggregated (over selected time intervals,

[2] Adaptive plans are under consideration as part of the RT-TRACS program and require extensive sensor capabilities to capture dynamic traffic conditions; models accommodating such plans are themselves subject to validation study.

such as the signal cycle) numerical output for each link. The numerical outputs include the following.

- throughput (the number of vehicles discharged on each link)
- average link travel time
- link queue time (the sum over vehicles of the times, in minutes, during which the vehicles are stationary, or nearly so)
- link stop-time (sum over vehicles of stationary time)
- maximum queue length (on each lane in the link over the simulation time)
- link delays (simulated travel time minus free-flow travel time, summed over all vehicles discharging the link)

Most of these statistics can be attached to movements or lane levels within each link, but we do not do so. We will take CORSIM performance measures from this output.

One hour of simulation for the test bed network takes about 40 seconds on a Pentium III-850 MHz PC. During this time, approximately 38,000 vehicles are processed through the network. While each run is quick, the need for many runs to deal with the substantial variability induced by the stochastic assumptions lengthens experimental time considerably. A detailed uncertainty analysis greatly increases computational demands. An advanced computing environment (for example, distributing the simulations across a network of machines) could, of course, substantially reduce computing time.

DATA COLLECTION

A crucial element in validation is designing and carrying out data collection both for estimating input to the model and for comparing model output with field data. The challenge lies in managing costs while obtaining useful data relevant to both estimation and validation.

For our test bed example, initial field data for the network were collected on a single day (Thursday, May 25, 2000) for three hours in the morning (7:00 am to 10:00 am) and three hours in the afternoon (3:30 pm to 6:30 pm). The processing of the data and the analyses were limited to the three one-hour periods, 8 am to 9 am, 4 pm to 5 pm, and 5 pm to 6 pm. This covered the peak periods as well as a "shoulder" period.

Acquiring data for the input to CORSIM is a formidable task. Input such as driver characteristics is extremely difficult to gather, and in the test bed example we relied mostly on CORSIM default values. There were very few pedestrians, and they had no discernible effect on traffic, leading us to ignore the pedestrian input. Incidents were not included, despite the fact that there were illegally parked vehicles that did affect traffic flow. Because illegal parking was an endemic condition, we coded the network to account for its effect. Other parameters, such as free-flow speed, were selected on the basis of posted speed limits. Signal timing plans and bus routes and stations were collected directly in the field and entered into CORSIM.

Traffic volume data were collected manually by observers counting vehicles and by video recording. Human observation is notoriously unreliable, but cost considerations did not allow video coverage of the full network. However, the video information, covering all the links of the internal network of figure 1, was rich enough to allow adjustment of the observers' counts that determined the flow rate of vehicles at entry nodes of the network. On the other hand, turning movements outside the internal network could neither be confirmed nor reliably adjusted by video information. Extracting the video information took a considerable investment of time and personnel, rivaling the cost of acquiring the raw video data.

Supplemental validation data were collected on a similar schedule on September 27, 2000. These were extracted primarily from video. The purpose was to answer our second question, if CORSIM accurately represents reality under new conditions, by analyzing its effectiveness of CORSIM in predicting traffic behavior under the September conditions.

It is most convenient to collect data for validation while collecting data for inputs. The use of the same or closely related data for both input and validation is an issue rarely confronted. The conventional wisdom says that such dual-use of the data is forbidden. In fact, it can be done but the attachment of computable uncertainties, essential to producing reliable results, is not straightforward. This issue is under study by a research team at the

National Institute of Statistical Sciences (NISS) and Duke University. A Bayesian approach based on Bayarri and Berger (1999) holds promise for producing methodology to treat the issue.

A problem as yet not addressed is assessing the impact of data of inferior quality. The problem is complicated by the need to specify the brunt of the impact; to quantify scenarios of alternative collections of data; and to design, execute, and analyze computational experiments to measure the consequences, or sensitivities, of model output to wrong data input, including incorrect signal settings or drifts in signal timing. This issue is not unique to transportation studies and research; it permeates virtually all sciences.

ESTIMATION OF CORSIM INPUT FROM INITIAL (MAY) DATA COLLECTION

The direct, fixed input required for CORSIM to run, including signal timing plans for each of the three one-hour periods, was obtained from the field and entered into CORSIM. The direct input requiring estimation was treated as follows.

- Vehicle mix at each entry node was estimated from one-hour (human-observer) counts for autos and trucks.
- Turning probabilities (left turn, right turn, through) at each intersection were estimated from one-hour video counts (where available) and from human-observer counts at other intersections.
- Inter-arrival rates (see equation 1) were estimated with the assumption that $k = 1$. The λ for each entry node and each of the 3 one-hour time periods was estimated as the total number of vehicles entering the (entry) link divided by 3,600.

Some λs were later adjusted to reduce discrepancies between downstream counts generated by CORSIM and those observed by video; the discrepancies were believed to be due to inaccuracy of human-observer counts and the effects of parking lots. Turning movements were left at their field estimates. Measuring the ultimate effect on uncertainty of these modifications is an issue that remains to be explored.

Validation Process

Validation without purpose has little utility. For example, our interest in CORSIM here is its value in assessing and producing good time-of-day signal plans. But, CORSIM could also be used to evaluate traffic operations under disruptions, such as a bridge closing, or to changes in the network, such as strict enforcement of parking laws or truck restrictions. A more subtle use could be in measuring the impact of driver decisions when faced with a network modification. Some objectives may only reflect changes in the network; others may also implicate induced changes in demand.

Navigating through this variety of issues requires multiple tools. For example, visualization and expert opinion give an overall assessment of whether the model output matches reality in a qualitative but highly subjective way. When video data are placed next to computer animations, discrepancies (and similarities) can be seen directly, particularly if viewers are experts familiar with the network and its characteristics.

However, the stochastic nature of CORSIM and of real traffic requires more than informal visualization. Questions remain, such as which random animation should be used to compare with the real traffic and is the single day of traffic recorded by video typical. More stringent comparisons based on a second tool, statistical analysis, become crucial in reducing the subjectivity, guiding the visualization through choices of animation, and pointing to model flaws responsible for aberrant behavior. The challenge is then to provide statistical analyses appropriate to the desired ends.

There can be many competing analyses, one for each evaluation criterion as defined in the following section. Treating the multiplicity of comparisons in a coherent way is often disregarded. Is the model flawed if it produces a poor match to reality at only one (five?) of one hundred links? Added complications come from comparisons based on evaluations of corridor and system characteristics as well as those of individual links.

Thus, the initial task is to select evaluation criteria. Comparison of the field and model through selected evaluation functions in the specific application of CORSIM to the network of figure 1 will touch on the concerns and issues raised.

Evaluation Functions

Selecting an evaluation function φ is crucial and sometimes complicated by competing practical and theoretical considerations. First, is φ relevant to the purpose? Choosing among many relevant φs is sometimes eased by requiring feasibility in both calculating model output for φ and collecting field data for calculating corresponding field value(s) of φ.

In our test bed example, a good criterion for judging a signal-timing plan may be average link travel time, complicated to obtain in CORSIM and costly to obtain in the field. The tactic of using probe vehicles, while possible in principle, is inhibited by the cost of using large numbers of vehicles and the need to account for the substantial variability connected with the use of probes. Computing vehicles' travel time from video is highly labor-intensive; useful, automatic area-wide detection methods, such as Mobilizer (Lall et al. 1994), are neither widely available nor fully adequate.

The evaluation function φ is likely to have versions at multiple time scales and at different levels of spatial aggregation. For example, total queue-time per cycle per link could be aggregated over cycles and over links to form evaluations based on behavior over selected corridors, over the whole system, and over distinct time periods. The choice of levels of space-time resolution adds to the determination of relevance and can be complicated by questions of feasibility.

Statistical analyses of the φs must treat the variability arising from the intrinsic stochastic structure of simulators such as CORSIM.[3] However, field variability is also consequential, and that cannot be so readily captured without elaborate and costly field-data collection. This is a confounding issue, partly addressed below.

Travel times are very hard to obtain in the field. Stop time per vehicle can be calculated for each link covered by video. Queue length per cycle can also be calculated, but queue time is very difficult to obtain in the field though a standard part of CORSIM output.

We chose stop time (stopped delay) on approaches to intersections as the primary evaluation function. It has been the typical measure by which intersection level of service (LOS) is evaluated (TRB 1994). The comparative ease of collecting stop time data from the video strongly affected our choice, reinforced by the fact that other criteria such as throughput, delay, travel time, and queue length are all highly correlated with stop time.[4] In addition, we believe that drivers on urban street networks are particularly sensitive to stop time, spurring traffic managers to seek its reduction. In fact, the Highway Capacity Manual's selection of stopped delay for LOS designation is meant to reflect the user's perception of the intersection's quality of service. We used V (the number of vehicles leaving an intersection, particularly exit nodes) as an auxiliary evaluation function. V is readily calculated from video and is also needed to calculate stop time per vehicle discharged (STV) at a link. At approach a,

$$STV(a) = \frac{\text{Total stop time}}{V(a)}.$$

$$V(a) = V_0(a) + V_s(a) \tag{2}$$

where V_0 is the count of vehicles that do not stop on a, while V_s is the count of vehicles that do stop on a. This raises the question of whether STV is an adequate reflection of the characteristics of the network (and signal plan) compared to the pair

$$P(a) = \frac{V_s(a)}{(V_0(a) + V_s(a))} \tag{3}$$

$$STVS(a) = \frac{\text{Total stop time}}{V_s(a)} = \text{stop time per stopped vehicle.}$$

We will see that these quantities provide a sharper understanding of the comparison between CORSIM and the field.

STV or *STVS* for aggregations of approaches (routes or corridors) is very difficult to obtain,

[3] Deterministic models will not have intrinsic randomness but will be exposed to variability either in assumptions about input parameters or from data used to estimate input parameters.

[4] Rejection of delay was also affected by CORSIM calculations that fail to include vehicles left in the system at the end of the one-hour simulation period, potentially resulting in misleading numbers under congested conditions.

requiring the tracking of individual vehicles. But some concept of performance on aggregation could be important. For example, a long delay on one link may be compensated by a short delay on the next link downstream, leaving the corridor and the system as a whole unaffected. By summing over the individual links forming a corridor, we create a "pseudo stop time" for the corridor. This will be close to a real stop time, provided vehicles turning off of or on to the corridor exhibit little or no difference from those traveling straight through. However, the value of such "pseudo stop times" is unclear, and here we only deal with individual links and approaches.

Multiplicity questions begin with the selection of links or approaches for comparison. We selected links on corridors that contained the heaviest traffic during the main peak period directions, east and south in the morning and west and north in the evening. A full treatment of multiplicity questions will not be presented here.

Tuning

Tuning and calibrating a model are general terms, often used interchangeably, sometimes yielding confusion. In the previous major section, we treated estimation of input to the model directly from field data. When model output data are used, either alone or with field data, to determine input parameters, the process is often called calibration. Tuning is a term commonly associated with adjusting input parameters to match model output. As in the usage of "calibration," the term tuning is frequently reserved for cases where the input parameters are unobservable or represent physical and other processes the model does not (or cannot) adequately incorporate.

The practice of tuning is not only common but often essential, especially for a long-range study of the model and its associated phenomena. Some input parameters may be neither well-specified nor capable of being estimated from the field data. One example is driver aggressiveness in our test bed. Some assumptions about input parameters may be found erroneous after viewing the data, and their modification may produce better simulations. Ultimately, the validation accompanying such tuning becomes problematic.

Two types of tuning were done in the test bed example. The first addressed the blockage of turns at two intersections and the subsequent gridlock. We altered the network by introducing sinks and sources that allowed the bypass of the blockage without affecting throughput. The second was stimulated by a substantial difference on one link (at the LaSalle/Ontario intersection in figure 1) between the field and CORSIM stop times. This difference was largely resolved by changing the free flow speed from 30 miles per hour (mph) to 20 mph. The input of 30 mph was induced by the speed limit; its revision to 20 mph is consistent with the observed (from video) speed of vehicles on the corridor (LaSalle Street).

Visual Validation

Where visualization is available, as it is with CORSIM animation and with video field data, a compelling approach to validation is visually comparing the two to see if traffic in CORSIM behaves like traffic in reality. To a great extent, this is a highly informal and subjective approach. Nonetheless, it is of great value in assessing CORSIM's capability to emulate reality as well as identifying sources of trouble or flaws in CORSIM, flaws that can sometimes be corrected by intervention in the coding.

The utility of visualization depends on the specifics of each application. What may be learned from the CORSIM example may pertain to other microsimulators but not necessarily to other computer models.

A sign of problems in an application of CORSIM is the presence, in several of the replicate simulation runs, of spillback and gridlock in situations where these do not occur in reality. Spillback will occur on networks such as in figure 1, where near saturation conditions are present during peak periods; however, recovery in the field usually takes place reasonably quickly. A difficulty with CORSIM is its apparent inability to recover readily from spillback, often resulting in gridlock. The effect on performance measures is usually to produce large outliers in a repeated set of simulations, sometimes indicated by large run-to-run variance. A histogram of outputs can identify large outliers. Following up with examination of the corresponding animations can often identify causes.

In two instances, it was apparent that the cause was an inability of CORSIM to allow driver adjustment to left (or right) turn blockage, resulting in a spillback that would never clear up.

Numerical Comparisons

Throughput Comparison

In table 1, we present test bed results on throughput for internal network. The net change indicates discrepancies showing less output in the morning and more output in the evening. This is due to the garage effect: vehicles disappear to the parking lots in the morning and reappear from them in the evening.

The means of 100 replicated CORSIM runs are close to the observed counts in table 2, except for eastbound Ohio/LaSalle in the morning and westbound Grand/Wells in the evening. The first can be explained in large part by the disappearance of vehicles in the morning into parking lots along Ohio Street, a major one-way, eastbound corridor. The second, correspondingly, can be attributed to the appearance of vehicles from parking lots on Grand during the evening. In addition, there is a high enough variability in CORSIM runs to account for a considerable part of the apparent discrepancy (see figures 2 and 3).

TABLE 1 Comparison of Throughput on Internal Network (vehicles per hour)

Period	Direction	Field (vehicle)	CORSIM (vehicle) average	CORSIM (vehicle) s.d.*
8–9 AM	in	11,805	11,895	48.1
	out	11,330	11,877	52.8
	net	–475	–18	–
4–5 PM	in	10,834	10,805	39.7
	out	10,990	10,796	40.6
	net	156	–9	–
5–6 PM	in	11,431	11,449	61.9
	out	11,756	11,422	71.9
	net	325	–27	–

Note: Field data were obtained from video taken on May 25, 2000. Averages are rounded to nearest integer.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.

It would be incautious to view the similarity of real data to the model runs as evidence of the model's validity. Whether these internal throughputs are good evaluation functions is unclear. They are, however, relevant to *STV* and *STVS* because they determine the denominators of those measures. Not taken into account is the tuning of the model to help match inputs to the model with the flows observed in the video. How to achieve this formally is a matter of some delicacy and is a research issue currently under investigation in a National Science Foundation sponsored research project at NISS.

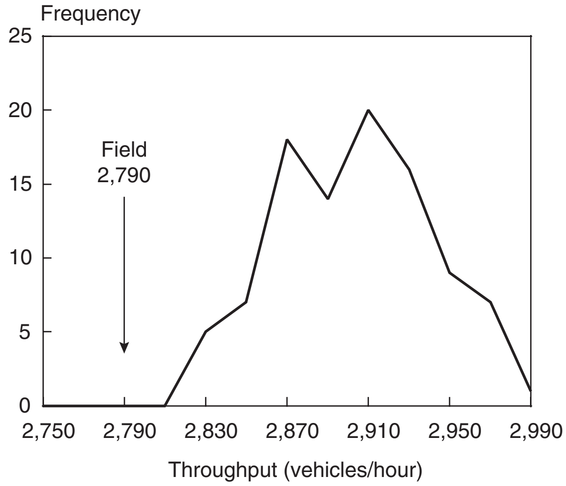
Though field variability cannot be adequately captured, we produced CORSIM and field time series of throughputs to examine whether CORSIM shows a degree of variability (over time) characteristic of the field data. Figure 4 presents such time series, obtained as follows. There are 48 signal

TABLE 2 Comparison of Throughput on Selected Key Links (vehicles per hour)

Period	Link	Field (vehicle)	CORSIM (vehicle) average	CORSIM (vehicle) s.d.*
8–9 AM	SB LaSalle at Ohio	1,651	1,641	30.3
	EB Ohio at LaSalle	2,790	2,894	38.9
	SB Wells at Ohio	693	694	17.4
4–5 PM	EB Ohio at Orleans	1,948	1,947	2.3
	NB Orleans at Ohio	1,498	1,489	25.0
	NB LaSalle at Ontario	1,500	1,478	28.0
5–6 PM	EB Ohio at Orleans	1,897	1,896	2.5
	WB Grand at Wells	1,204	1,133	21.9
	NB LaSalle at Ontario	1,636	1,617	26.3

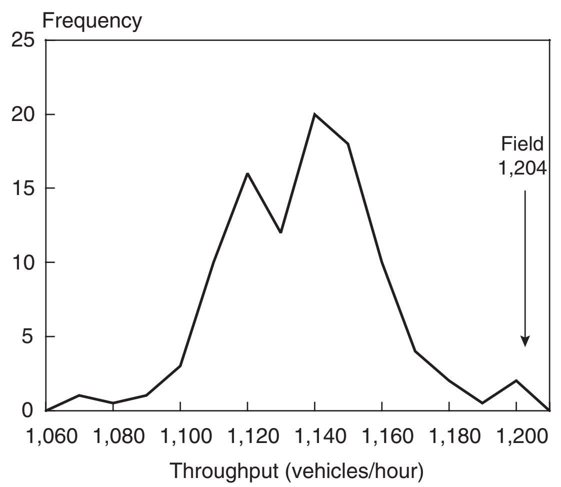
Note: Field data were obtained from video taken on May 25, 2000. Averages are rounded to nearest integer.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.
SB = southbound
EB = eastbound
NB = northbound
WB = westbound

FIGURE 2 Link Throughput at Eastbound Ohio/LaSalle (8–9 AM)



Note: Histogram is based on 100 CORSIM runs.

FIGURE 3 Link Throughput at Westbound Grand/Wells (5–6 PM)



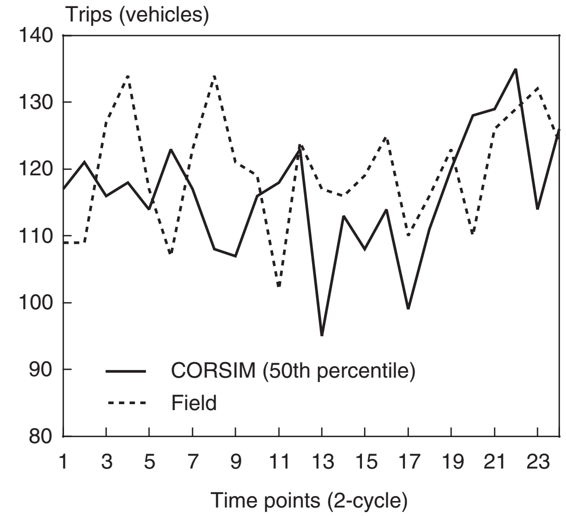
Note: Histogram is based on 100 CORSIM runs.

cycles during the 8 am to 9 am morning peak, and we combined throughputs over every 2 cycles, equal to 150 seconds of elapsed time in the 1-hour period. This leads to a time series at 24 time points. CORSIM was run 100 times, and the variation of each time series was computed as

$$\frac{\sum_{t=1}^{23}\left[Y(t+1)-Y(t)\right]^2}{23} \tag{4}$$

where *Y(t)* represents throughput during time interval *t*. We selected the representative CORSIM

FIGURE 4 Comparison of CORSIM with Field Variation (Eastbound Ohio at LaSalle, Morning Peak)



Note: Histogram is based on 100 CORSIM runs.

time series variation as the median of the 100 variations.

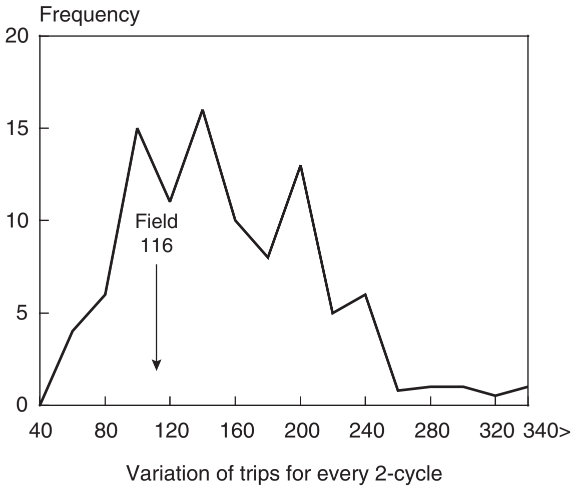
CORSIM variability, as shown in figure 4 (as well as on the link southbound LaSalle at Ohio), is close to that of the field. Indeed, the variation of the field series is 116 and is at the 30th percentile of the CORSIM distribution, as shown in figure 5.

Stop Time Comparisons

The distribution of stop time at each approach has some probability at zero (the proportion of vehicles that do not stop); this is singled out in the first part of table 3. Characteristics of the conditional distribution of stop time (given that a vehicle stops) are given in table 4. There are definite discrepancies on southbound LaSalle at Ohio during the morning, where CORSIM generates fewer stops but longer stop times for its stopped vehicles. On eastbound Ohio at LaSalle, a similar (though somewhat reduced) discrepancy is apparent. While there appear to be differences on some of the other approaches, none appear very significant. For example, CORSIM stops fewer vehicles on northbound LaSalle at Ontario in the 5 pm to 6 pm period, but the stop times are close.

These differences call for an explanation. Examination of video and CORSIM animation exposes the key cause: CORSIM does not fully reflect driver behavior. In particular, lane utilization

FIGURE 5 CORSIM Variation vs. Field (Eastbound Ohio at LaSalle, Morning Peak)



Note: Histogram is based on 100 CORSIM runs.

TABLE 3 Comparison of Stop Rates on Key Links

Period	Link	Field (percent)	CORSIM (percent) Average	CORSIM (percent) s.d.*
8–9 AM	SB LaSalle at Ohio	50	30	1.8
	EB Ohio at LaSalle	56	35	3.9
	SB Wells at Ohio	99	94	1.2
4–5 PM	EB Ohio at Orleans	50	59	1.0
	NB Orleans at Ohio	51	56	2.9
	NB LaSalle at Ontario	42	47	3.6
5–6 PM	EB Ohio at Orleans	48	59	1.0
	WB Grand at Wells	55	53	3.3
	NB LaSalle at Ontario	78	62	3.4

Note: Field data were obtained from video taken on May 25, 2000.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.
SB = southbound
EB = eastbound
NB = northbound
WB = westbound

TABLE 4 Comparison of Key-Links STVS (stop time per vehicle stopped)

Period	Link	Field (seconds/vehicle)	CORSIM (seconds/vehicle) Average	CORSIM (seconds/vehicle) s.d.*
8–9 AM	SB LaSalle at Ohio	27.8	32.3	1.8
	EB Ohio at LaSalle	15.4	18.6	0.8
	SB Wells at Ohio	33.1	39.6	0.4
4–5 PM	EB Ohio at Orleans	18.4	18.7	0.3
	NB Orleans at Ohio	20.6	20.6	1.7
	NB LaSalle at Ontario	33.5	27.9	3.0
5–6 PM	EB Ohio at Orleans	15.2	18.7	0.3
	WB Grand at Wells	8.3	10.5	2.1
	NB LaSalle at Ontario	33.4	34.2	2.9

Note: Field data were obtained from video taken on May 25, 2000.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.

in CORSIM is not consistent with lane utilization in the field. On some links, vehicles in the field more often join long queues where they are briefly stopped. These vehicles typically do not appear in CORSIM simulation as having stopped. This accounts for smaller *STVS* in the field than in CORSIM. So, even though CORSIM does not fully reflect the field, the key measure of how long truly stopped vehicles are delayed appears to match what is seen in the field quite reasonably.

PREDICTION AND VALIDATION

The most compelling form of validation is through confirmation by predictions in new circumstances. In the test bed example, a plan, different from the one in the field in May, was put in place in September 2000. Under these new circumstances (a new signal plan) predictions were to be made and data collection designed for September 27, 2000, a day expected to be similar to the date of the first data collection, May 25, 2000.

The simulator's performance prediction requires specification of the input expected at the time of the new data collection. Believing that the conditions in the field for the September data collection would be the same as in May, we ran CORSIM with the May input, except for signals.

After the data were collected in September, we compared the results, first for throughput (table 5) on several key links. Except for the 13% disparity on southbound LaSalle, the throughputs were close. Whether or not the disparity in demand on southbound LaSalle mattered awaited further analysis of stop time. The predictions of September stop time performance with the May input are in tables 6 and 7 (see also figures 6 and 7). Except for northbound Orleans to the freeway, the *STVS*s are reasonably close. For the reasons discussed earlier, we have several disparities on stop rates.

To clarify these matters, we first checked the effect of change in demand on southbound LaSalle during the morning peak. We decreased the input demand there by 10%, reran CORSIM 100 times, and obtained essentially no change in output. The stop rate on southbound LaSalle at Ohio went from 30.3% to 30.9%, while *STVS* went from 22.0 to 22.3 seconds per vehicle (sec/veh).

Next we explored the disparity on northbound Orleans at the freeway in the afternoon peak and observed, through video, that drivers effectively used green time of 20 seconds instead of the displayed green time of 16 seconds. Introducing this modification changed stop rates from 74% to 65%, and average *STVS* changed from 51.9 to 40.8

TABLE 5 Field-Measured Throughput Comparison at Key Links

Period	Link	May (vehicle)	September (vehicle)
8–9 AM	SB LaSalle at Ohio	1,650	1,441
	EB Ohio at LaSalle	2,790	2,798
4:30–5:30 PM	NB LaSalle at Ontario	1,607	1,696
	NB Orleans to Freeway	838	899
	NB Orleans at Ontario	1,051	1,107

SB = southbound
EB = eastbound
NB = northbound

TABLE 6 Comparison of Stop Rates on Key-Links

Period	Link	Field (percent)	CORSIM (percent) average	s.d.*
8–9 AM	SB LaSalle at Ohio	52	30	2.7
	EB Ohio at LaSalle	37	38	3.0
4:30–5:30 PM	NB LaSalle at Ontario	36	51	4.2
	NB Orleans to Freeway	53	74	4.1
	NB Orleans at Ontario	47	43	2.0

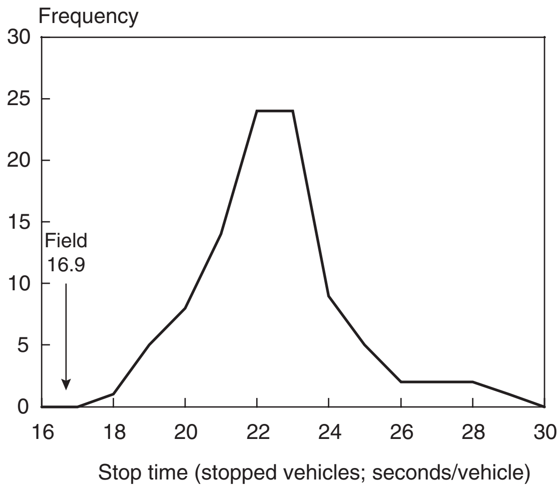
Note: Field data obtained from video taken on September 27, 2000.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.
SB = southbound
EB = eastbound
NB = northbound

TABLE 7 Comparison of *STVS* (stop time per vehicle stopped)

Period	Link	Field (seconds/ vehicle)	CORSIM (seconds/vehicle) average	s.d.*
8–9 AM	SB LaSalle at Ohio	16.9	22.0	2.0
	EB Ohio at LaSalle	15.2	21.6	1.2
4:30–5:30 PM	NB LaSalle at Ontario	26.4	24.8	1.8
	NB Orleans to Freeway	31.4	51.9	7.6
	NB Orleans at Ontario	21.9	24.0	1.0

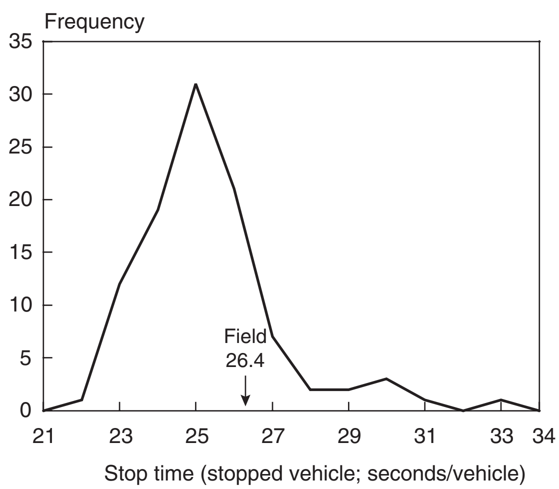
Note: Field data obtained from video taken on September 27, 2000.
* s.d. is the estimated (from 100 runs) standard deviation of a CORSIM run.
SB = southbound
EB = eastbound
NB = northbound

FIGURE 6 Link *STVS* at Southbound LaSalle/Ohio



Note: Histogram is based on 100 CORSIM runs.

FIGURE 7 Link *STVS* at Northbound LaSalle/Ontario



Note: Histogram is based on 100 CORSIM runs.

sec/veh with a standard deviation of 6.8. The difference between 31.4 (the field *STVS*) and CORSIM's average of 40.8 is neither statistically significant (within 2 standard deviations) nor practically significant (same level of service; see table 9). Nonetheless, we examined the northbound Orleans link more carefully. We noted that CORSIM has difficulty dealing with storage of vehicles on short, congested links just downstream of a wide intersection, exactly the characteristics of northbound Orleans at the freeway (the intersection at Ohio is 60 feet; the entire link is 240 feet; and the link is highly congested). We could have brought the CORSIM predictions more closely in line with the numbers in the field by altering the length of the link, but we regarded such tuning as potentially misleading.

TABLE 8 Δ CORSIM Compared with Δ Field

Link	Δ CORSIM	Δ Reality
SB Ohio at LaSalle	0	3
SB LaSalle at Ohio	–11	–10
NB LaSalle at Ontario	–9	–5
NB Orleans to Freeway	13	15
NB Orleans at Ontario	1	–2

Note: Δ = *STVS* [September] – *STVS* [May]
EB = eastbound
SB = southbound
NB = northbound

A highly informative evaluation function of CORSIM is the change in CORSIM predictions, Δ CORSIM (September *STVS* – May *STVS*), compared to the corresponding change in the field values, Δ Field. Even though the CORSIM predictions were not always accurate, the Δs are close and of the same sign (table 8). This is particularly important for comparing the performance of competing signal plans: predictions of improvements (in two links), no change (in two links), and degradation (on one link) in CORSIM jibes with the changes in reality.

ANALYSIS OF UNCERTAINTY

A more exacting treatment of validation requires closer attention to the following.

- uncertainties inherent to the simulator as well as from parameter estimates used to define input distributions
- multiplicity questions arising from the use of multiple evaluation functions (for example, the multiple link/approaches in tables 1 and 2)

The first item can be addressed through a Bayesian analysis. For instance, in the test bed example, the uncertainty question can be dealt with by specifying prior distributions for the λs in equation (1) as well as for the probabilities p of turning movements. Posterior distributions of λ, p can then be computed given field data. Before each CORSIM run, a draw from the posteriors can be made, leading to a selection of λ, p, which then provides the needed input for the run. The resulting variability in 100 runs, for example, will then incorporate both the inherent CORSIM variability as well as the uncertainty stemming from the use of the field data in estimating λ, p.

Bayarri, Berger, and Molina (2001) are carrying out such a Bayesian analysis. Preliminary results indicate that while the variability of *STVS* may increase, the qualitative behavior of CORSIM remains the same. Complications in the analysis derive from the complexity of the network and its impact on computing the posterior distribution. These results will appear elsewhere.

A fuller Bayesian treatment of uncertainty of prediction, now under study, can incorporate questions of systematic bias in CORSIM predictions of reality. One aspect of such an inquiry is the potential use of a "CORSIM adjusted by bias" predictor in place of CORSIM itself.

The treatment of multiplicity requires appropriate formulation. Methods described in Westfall and Young (1992) and Williams et al. (1999), as well as False Discovery Rate approaches (Benjamini and Hochberg 1995), are not clearly applicable due to the high level of dependence among evaluation functions.

Last, we note that the effect of the uncertainties will be felt in the evaluation functions or, equivalently, through loss structures that take practical significance into account. For example, a difference of 5 seconds in stop time can be minor, but a difference of 15 seconds may be major. One starting point may be a comparison of the field and CORSIM-predicted LOS. Table 9 shows criteria for LOS based on stopped time in the 1994 Highway Capacity Manual.

TABLE 9 LOS Designation in the Highway Capacity Manual (1994)

Level of service	Stopped time per vehicle (STV; seconds/vehicle)
A	STV ≤ 5
B	5 < STV ≤ 15
C	15 < STV ≤ 25
D	25 < STV ≤ 40
E	40 < STV ≤ 60
F	STV ≥ 60

CONCLUSIONS

We present conclusions about the validation process and the specific test bed model, CORSIM. The validation process has five key elements: *context, data, uncertainty, feedback,* and *prediction.* Context is critical. It drives the formulation of evaluation functions or performance measures that are ultimately the grounds on which validation must take place and affect interpretations of uncertainty. For example, statistically significant disparities may, in the context of an application, be practically insignificant. In addition, context and the specified evaluation functions can affect the selection or collection of data, both field and model output, to be used for evaluation. Conversely, the availability or feasibility of data collection can determine the choice of evaluation functions. These factors may then converge in the calculation of uncertainties stemming from noisy data and model imperfections. The outcome of the evaluations and the associated uncertainties points to possible flaws in the models and feedback to model adjustments that correct or, perhaps, circumvent the flaws. Ultimately, it is through prediction that validation of a model is reached.

The process we described is effective and generally applicable. Of course, implementing the particulars, done for the most part in the test bed example, will require filling in a number of gaps, most specifically in determining uncertainties but also in designing data collection, assessing the impact of data quality, and detecting flaws.

Test bed conclusions derive from the two questions we posed: Does CORSIM mirror reality when properly calibrated for field conditions? Does CORSIM adequately predict traffic performance under revised signal plans?

Comprehensive calibration of CORSIM is infeasible; there are too many parameters that can (and some that cannot) be calibrated with field data. Our approach was to focus on key input parameters, such as external traffic demands, turning proportions at intersections, and effective number of lanes (for example, due to illegal parking), using CORSIM default values for other inputs.

We found that CORSIM was effective but flawed. A major difficulty is CORSIM's propensity to turn spillback into gridlock; inadequately modeled driver behavior led to intersection blockage far too frequently. CORSIM does not accurately model lane distribution of traffic. Lane selection in reality was much more skewed than in CORSIM. CORSIM tends to stop more vehicles than indicated in the field. In reality, drivers coast to a near stop then slowly accelerate through the signal, but the behavior is much more abrupt in CORSIM.

The first of these flaws was corrected by modifying the network. The second flaw had some effect but was relatively minor. The third flaw manifested itself in disparate stop rates but did not seriously affect stopped time per vehicle stopped (*STVS*).

Overall, despite its shortcomings, CORSIM effectively represented field conditions. Even when the field observations lie outside the domain of the CORSIM distributions, as in figures 2 and 3, there is virtually no difference in the estimated levels of service (table 9) between the field and CORSIM, practically insignificant even if statistically significant.[5]

The predictability of CORSIM was assessed by applying revised (September) signal plans to the May traffic network. CORSIM estimates of *STVS* were reasonably close to field estimates, and the CORSIM LOSs were, for the most part, similar to those observed in the field. More importantly, CORSIM successfully tracks changes in traffic performance over time: on five links for which field data were available, two links exhibited a reduction in *STVS*, one link an increase, and two had no significant change; CORSIM's predictions were the same.

In summary, a candid assessment of CORSIM is that with careful calibration and tuning, CORSIM output will match field observations and be an effective predictor.

[5] The CORSIM distribution does not reflect the additional uncertainty induced by the field data estimates of model input parameters. Therefore, statistical significance here is overstated.

ACKNOWLEDGMENTS

The authors are very grateful for the support and assistance of Mr. Christopher Krueger and Mr. Thomas Kaeser of the Chicago Department of Transportation. We also thank the Urban Transportation Center at the University of Illinois at Chicago for their help in the data collection. This research was sponsored in part by grants DMS-9313013 and DMS-9208758 from the National Science Foundation to the National Institute of Statistical Sciences.

REFERENCES

Bayarri, M.J. and J.O. Berger. 1999. Quantifying Surprise in the Data and Model Verification. *Bayesian Statistics 6*. London, England: Oxford University Press.

Bayarri, M.J., J.O. Berger, and G. Molina. 2001. Fast Simulators for Assessment and Propagation of Model Uncertainty. *Proceedings of the Third International Symposium on Sensitivity Analysis and Model Output*.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B* 57(1):289–300.

Berliner, L.M., J.A. Royle, C.K. Wikle, and R.F. Milliff. 1999. Bayesian Methods in the Atmospheric Sciences. *Bayesian Statistics 6*. London, England: Oxford University Press.

Lall, B., K. Dermer, and R. Nasburg. 1994. Vehicle Tracking in Video Image: New Technology for Traffic Data Collection, in *Proceedings of the Second International Symposium on Highway Capacity,* Sydney, Australia, 365–83.

Lynn, N., N. Singpurwalla, and A. Smith. 1998. Bayesian Assessment of Network Reliability. *SIAM Review* 40:202–27.

Park, B., N. Rouphail, J. Hochanadel, and J. Sacks. 2001. Evaluating the Reliability of TRANSYT-7F Optimization Schemes. *Journal of Transportation Engineering* 127(4):319–26.

Transportation Research Board (TRB). 1994. *Highway Capacity Manual.* Washington, DC: National Research Council.

U.S. Department of Transportation (USDOT), Federal Highway Administration (FHWA). 1996. *CORSIM User Manual.* Washington, DC.

Westfall, P.H. and S. Young. 1992. *Resampling-Based Multiple Testing*. New York, NY: Wiley-Interscience.

Williams, V.S.L., L.V. Jones, and J.W. Tukey. 1999. Controlling Error in Multiple Comparisons, with Examples from State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics* 24(1):42–69.

LAURENCE R. RILETT
CLIFFORD H. SPIEGELMAN
Texas A&M University
and
Texas Transportation Institute

The authors are to be commended for a timely article. With the recent advances in intelligent transportation systems (ITS) deployment, the corresponding availability of large traffic databases, and the increased use of traffic microsimulation models by transportation engineers, the issues examined are important. However, we would like to take this opportunity to point out some additional issues and research questions related to the proposed methodology.

The success of this study is attributable to a number of factors. Determining the portability of the methodology to other situations, even similar situations, is a matter for further study. The model has several inputs: data, technical expertise, and network development. In the assessment of CORSIM, a large set of factors is evaluated, many of which are inputs to CORSIM. Thus, we wonder if a group with less expertise than the paper authors were to change the signal timing, would they have been as successful? If a junior engineer developed the network input to CORSIM, would the model have been as successful? If the data were collected at a different location, would it have been as accurate? How accurate do input data have to be for the model to make quality predictions?

In almost all areas of science, engineering, and life, making a prediction and then having that prediction judged accurate is powerful evidence that the method of making the prediction is a good one. Still one wonders how many successes and what proportion of successes is needed to validate a model. Babe Ruth gave a famous baseball prediction when he pointed to the Yankee Stadium center field stands and predicted that he would hit a home run. He hit a home run immediately after his prediction. Does this mean that he could repeat the process anytime he wanted? What percentage of times would he need to be right and out of what number of pointing tries would he need to be right? We think issues like those addressed in this paper can be handled by developing an appropriate statistical theory for field experiments.

In this study only a select number of parameters were calibrated while the majority of parameters were left to their default values. In fact, only three types of changes were made: new sinks/sources were added to the network, the free flow speed on another link was reduced, and some entry volumes (more correctly λ values) were adjusted to better reflect downstream measurements. Interestingly, the behavioral parameters (e.g., gap acceptance, car following headway) went untouched, which is typically not done. For example, it is often assumed that the field values are relatively accurate and the behavior parameters (e.g., gap acceptance, driver aggressiveness) are calibrated so that the modeled output and traffic data are similar (see references). Regardless, the question of transferability arises—is the validation methodology appropriate for all locations or just for those locations where the default parameters apply? At a minimum, further study is required before statements such as "... with careful calibration and tuning, CORSIM output will match field observations and be an effective predictor." In our opinion it is easy to intuit situations where no amount of expert manipulation of input will allow CORSIM to be used, because the behavior modeling (i.e., default behavior parameters) in CORSIM would not apply to the drivers in the traffic network being simulated. Two related questions to the above argument are: 1) which parameters should be calibrated and which should be left alone? and 2) when is it reasonable to cali-

Clifford H. Spiegelman, Department of Statistics, Room 3143, Texas A&M University College Station, TX 77845-3143. Email: cliff@stat.tamu.edu.

brate high-fidelity-type microsimulation models, such as CORSIM, with relatively low-fidelity data?

Lastly, the authors make two implicit assumptions in their prediction methodology that need to be explicitly identified. The first is that the entering volumes and turning movements are fixed, which implies that the origin-destination movements are fixed. For small networks, such as the test bed, this may be reasonable. However, for larger networks, better signal timing will lead to an increase in capacity—in congested networks, such as Chicago where there can be significant latent demand, this would normally lead to an increase in observed volume. This is one of the reasons before and after studies of major transportation improvements are so difficult. In fairness to the authors, they did perform a sensitivity analysis on demand based on observed volume counts after the change. However, the point remains that research is required on when this assumption of constant demand can be made. Intuitively, the relationship between demand and network capacity would be important for both the after analysis and the actual traffic signal optimization.

The second assumption is that the routes chosen by the drivers remain constant as evidenced by the constant turning percentages. Intuitively, if there is a significant change in signal timing, drivers will change their routes if they can find a faster way to get to their destination. The fact that the authors observed that the turning percentages changed after the new signal timing was implemented lends some credence to this argument. Similar to the demand assumption, the constant route assignment assumption needs to be studied so that the conditions under which it can be made will be known. Obviously, if either assumption is invalid then the potential of the proposed methodology could be limited.

In closing, we are pleased to see that traffic microsimulation model validation is receiving the research attention it deserves. Hopefully, the research discussed in this paper will spur further research in this area and the important questions raised in the article will be adequately addressed.

REFERENCES

Benekohal, R.F. 1989. Procedure for Validation of Microscopic Traffic Flow Simulation Models. *Transportation Research Record* 1320:190–202.

Cheu, R., X. Jin, K. Ng, and Y. Ng. 1998. Calibration of FRESIM for the Singapore Expressway Using Genetic Algorithm. *Journal of Transportation Engineering* 124, 6:526–35.

Davis, C.F. and T.A. Ryan. 1981. Comparison of NETSIM Results with Field Observations and Webster Predictions for Isolated Intersections. *The Application of Traffic Simulation Models.* Special Report 194. Washington, DC: Transportation Research Board.

Lee, D., X.Yang, P. Chandrasekar. 2001. Parameter Calibration for PARAMICS Using Genetic Algorithm. Paper presented at the 80th Annual Meetings of the Transportation Research Board, Washington, DC. January.

Prevedourous, P. and Y. Wang. 1999. Simulation of a Large Freeway/Arterial Networkwith CORSIM, INTEGRATION and WATSim. Paper presented at the 78th Transportation Research Board Annual Meetings, Washington, DC. January.

Radwan, A.E., F. Naguib, and J. Upchurch. 1991. Assessment of the Traffic Experimental and Analysis Simulation Computer Model Using Field Data. *Transportation Research Record* 1320:216–26.

Rakha, H., M. Van Aerde, L. Bloomberg, and X. Huang. 1998. Construction and Calibration of a Large-Scale Microsimulation Model of the Salt Lake Area. *Transportation Research Record* 1644:93–102.

Rilett, L.R., K. Kim, and B. Raney. 2000. A Comparison of the Low Fidelity TRANSIMS and High Fidelity CORSIM Highway Simulation Models Using ITS Data. *Transportation Research Record* 1739:1–8.

Wang, Y. and P.D. Prevedouros. 1998. Comparison of INTEGRATION, TSIS/CORSIM, and WATSim in Replicating Volumes and Speeds on Three Small Networks. *Transportation Research Record* 1644:80–92.

MAX D. MORRIS
Iowa State University

I congratulate and thank Sacks et al. for an interesting and thoughtful case study of model validation in an important application area. The authors offer an insightful description of the general process of testing a computer model against reality, but, more importantly, describe how they accomplished this task in a very specific, complex setting. In the development of new methodology, the "devil" is always in showing that the proposed ideas and techniques can be relevant to the "details" of real, important problems. Careful case studies, such as this one, are important steps toward improving the practice of model validation.

Each of the points I raise in this discussion has been addressed in some form and to some degree by the authors. I hope that my restatement and elaboration gives readers a useful alternative view of a few of the issues that must be faced when designing and interpreting a validation study.

I will focus my remarks on only a few aspects of the problem and model considered by Sacks et al. (at least in part to avoid the certain embarrassment that would otherwise arise because I do not have their extensive knowledge of traffic modeling). During any given period of time, real vehicles travel through the area studied by the authors, each experiencing some stop delay time at intersection approaches; the total of all such times across vehicles is a well-defined quantity ϕ. We have a clear general understanding of the physical process that gives rise to ϕ; individual vehicles arrive at the intersections corresponding to the entry nodes displayed in figure 1, negotiate their way through the grid, and exit or disappear into garages; given enough detail on the individual movement of each vehicle, it is a simple matter to calculate its contribution to total delay time. This simplified concept of reality might be denoted by

$$\phi \leftarrow \mathbf{R}(t,u;c)$$

where (with apologies to Sacks et al. for using notation not entirely consistent with their own) t denotes the exact and complete collection of arrival times at each entry node, u represents an extensive set of variables that fully characterizes each vehicle's destination and the rules it uses in reacting to its environment, and c represents the timing of the signal lights (that we will "control"). The notation "$\leftarrow$" rather than "=" in the above expression indicates that this is our idea of how reality works—not necessarily the same thing as reality itself. Envisioned in this way, **R** is conceptually simple. In fact, a model could at least in principle be written that does *exactly* what **R** does, given t, u, and c as inputs.

However, models that require detailed values of t and u that match reality are of limited practical value because t cannot in practice be known before the time period of interest (and then only if impractically extensive measurements could be recorded at each entry node during that time period), and realistically u can never be known. Instead, simulation models like CORSIM are written with the idea that these quantities can be regarded as random processes, fully specified by a comparatively very small number of parameters. Rather than demanding the unattainable t and u as inputs, we define a model as a sort of stochastic generalization of **R**:

$$\Phi \sim \mathbf{M}(\lambda, \pi; c)$$

where λ and π are vectors that characterize distributions of random variables T and U, intended to represent the uncertainty in t and u, and so serving as the definition of a random variable Φ. The practical distinction between t and u, as discussed by the

Max D. Morris, Departments of Statistics and Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011-1210. Email: mmorris@iastate.edu.

authors, is that it is sometimes possible to collect limited data directly related to the first, while distributions used to represent the second are usually set by "defaults." A computer program expressing **M** then serves two purposes: it generates a single realization of the random vectors T and U, and *then* evaluates **R** *as if* these were the actual values of t and u. It is worth rewriting **M** to emphasize this:

$$(T, U) \sim \mathrm{D}(\lambda, \pi)$$
$$\Phi = \mathrm{R}(T, U; c)$$

where **D** expresses the distribution of T and U, given the input parameters. One execution of the model produces one realization of Φ, rather than the exact response that would follow from the fully defined deterministic inputs. (Here, I will use the capital letters T, U, and Φ to denote the distributions defined when λ and π are selected, the random variables defined by these distributions, and the realizations of these random variables produced when the model is evaluated; corresponding lowercase letters are used to denote data collected from the physical system.) The outputs from a number of repeated runs initiated with the same parameter inputs λ and π yield simulation estimates of the characteristics of the distribution of Φ, for example, mean, standard deviation, and quantiles. We may wish to give this distribution a frequentist interpretation, hoping that values of ϕ observed on similar days will look like a random sample from this distribution.

Now, this formulation sounds fairly pedantic, but it may help in describing some very serious questions related to the model validation process.

WHICH INPUT VALUES?

Given validation data from a specific period, what distributions (T, U) should be used in generating the corresponding CORSIM outputs? If t or some portion of it can be collected along with ϕ during the test period, one possibility would be to calibrate **D** so that the distribution T is consistent with t, that is, compute an estimate of λ using these data. The authors note in the section on Data Collection that this is often regarded as forbidden. However, we make one fundamental operational simplification in writing **M**: we drop the demand to actually know the very extensive vector t, settling instead for the much easier-to-specify and hopefully "statistically similar" T. Unless the realizations of T generated in the process of evaluating **M** are in some sense credible relative to the value of t realized during the validation exercise, the apparent validation error between ϕ and Φ can be the result of

1. a lack of structural integrity in **R**,
2. problems with the distributional assumptions expressed in **D**,
3. unrealistic specification of distribution parameter inputs λ and π, or
4. any combination of these.

It seems to me that if the testing of model structure is of primary interest, T should be defined so as to correspond to t as closely as possible.

However, if the goal of validation is assessment of the *prediction capability* that would be obtained from using the model in realistic situations, then the use of t in specifying T should certainly be forbidden. In this case, the process of specifying T must be viewed as a "hard-wired" piece of the model, and validation must be carried out using T as it would be constructed in practice. Here, validation is actually a joint test against all three kinds of problems listed above, including those associated with the technique used to select inputs to characterize T for the prediction time period. The Bayesian sampling approach mentioned in the section on Analysis of Uncertainty would be one way to perform this joint assessment. However, improving any of the conceptual forms of **R**, the distributional form expressed in **D**, or the process used to select input parameters can potentially improve performance relative to this kind of validation. Because each kind of improvement requires a different kind of developmental effort, the "factoring" of predictive uncertainty corresponding to these sources is an important part of the validation process.

Practice that forbids fitting input parameters to data collected along with validation data may sometimes stem from fears of "overfitting" T, that is, customizing T so that t is "too typical" an outcome. This is certainly a legitimate concern, but may be secondary to the more fundamental issue of what is to be validated—the physically motivated

R, **R** plus the operationally necessary **D**, or the full model plus the process of setting inputs.

WHICH OUTPUT VARIABLES?

The authors stress the important fact that selection of the output to be used in validation requires a balance between the *relevance* to the important questions and the *feasibility* of collecting the measurements. (I've used ϕ and Φ here to denote the measured and computed quantities being compared, respectively, even if they are actually functions of what the programmer might ordinarily call the model's output.) Hence, the authors select *stopped delay time* as the basis for comparing the model with reality, even though the much more difficult-to-measure *average link travel time* might be more relevant to questions concerning the timing of signals. This quandary exists anytime a model is produced to simulate physical circumstances that are difficult to examine directly—a common situation because such difficulty is often a major motivation for writing the model in the first place. Sometimes the discrepancy between what can realistically be measured and what is of most interest is even greater, for example, for models written to evaluate the reliability of nuclear weapons.

In the last section, I suggest that the Sacks et al. model validation is really a joint validation of **R**, **D**, and the method by which distributional parameters are set as inputs. Continuing this process, and agreeing with Sacks et al. that any validation must be done in the context of the purpose of the model, I think we may also need to consider the relationship between the output selected for measurement and comparisons and the output variables most critical in evaluating the success of setting *c*. Sacks et al. note that average link travel time and stopped delay are highly correlated (see the section on the Validation Process) and so at least informally consider this point.

In some settings, validation based on simultaneous comparisons of several outputs to various kinds of measurements may be possible. There may be few (or, in the case of the modeling of a nuclear weapon, no) measurements available that would be judged to be most relevant for the purposes of model use, a considerable quantity of data available corresponding to outputs of less relevance, and an intermediate quantity of data lying somewhere between these on some scale of relevance. Methodology, which formally accounts for relationships between multiple sources of validation data, and the fact that some are more relevant to the purposes of modeling than others, will be useful in such contexts.

WHICH OUTPUT VALUES?

Given specification of the input parameters by whatever means, repeated executions of **M** lead to a simulated "reference distribution" Φ. The validation exercise may be considered successful if the observed ϕ is a credible realization from this distribution. So, for example, the authors compare the "field" values with corresponding computed average and standard deviation values in tables 1 through 4. This amounts to a test of the hypothesis that t and u are drawn from the joint distribution characterized by **D** and the selected input parameters, and that **R** faithfully represents reality given t and u. However, even given effective specification of T and U, the authors remind us that "no simulator can be expected to capture real behavior exactly" (see the second section of the article); various details are always omitted, some intentionally and others through incomplete knowledge. Thus, what I have called the **R** section of CORSIM may not (and probably should not) contain explicit representations of the effects of emergency vehicles, thunderstorms, short-term construction work, and the use of cell phones by the drivers of some vehicles. A more detailed concept of reality that includes such phenomena, and so is perhaps closer in some sense to what happens in the streets, might be denoted

$$\phi \leftarrow \mathrm{R}^*(t,u,v;c)$$

where v represents the additional specific deterministic details of these unmodeled subphenomena, and **R*** is the more elaborate understanding of reality that takes these into account. Suppose for simplicity that v is parameterized so that **R*** is the same as the simpler **R** when $v = 0$:

$$\mathrm{R}^*(t,u,0;c) = \mathrm{R}(t,u;c)\ \forall\ t,u,c.$$

Hence, even if T and U effectively represent the physical variability of t and u, and our model expresses $\mathbf{R}$ perfectly, Φ may be inconsistent with ϕ because of the particular value of v at the time of validation. A strict "frequentist" might wonder whether the average of real-world ϕ values from a large number of days with identical t and u, but with v varying over some implied distribution V, might look like a reasonable realization of Φ. Related to this, we would consider

$$\mathbf{R}(t,u,0;c) \overset{?}{=} E_V\left[\mathbf{R}^*(t,u,V;c)\right]$$

where equality would suggest that the model might be trusted to predict such averages. But this is likely not to be what the developer of the model had in mind, and in any case, the test would require data that are operationally or even theoretically impossible to collect. Still, if such omitted effects are actually present—and they nearly always are—they imply potential variability in ϕ, which is not represented by the random variables in our model. This could mean that when T and U faithfully represent variation in t and u, Φ suggests less variation than should be attached to ϕ. Alternatively, it could lead to a situation in which the specified distributions T and U must have unrealistically large variances if the observed ϕs are to "fill out" their matching calculated reference distributions.

Since the quantity and variety of data needed to fully answer these questions cannot typically be obtained experimentally, the pragmatic conclusion may be this: If it is important to predict both the mean *and* variability of ϕ for specified conditions, validation should be aimed at judging not only whether the observed ϕs are close enough to their predictive means, but also, for example, whether their squared deviations from that mean agree with predictive variance, with the understanding that this does not automatically follow from getting the input distributions right (physically). The authors do the next best thing to checking the day-to-day variation of ϕ by looking at how some output quantities and measured quantities vary over time within a single validation period (see figure 4). This may be as close to comparing day-to-day distributions as can be achieved within the constraints imposed by sampling in this particular problem.

WHAT EFFECTS?

In thinking about experiments for validation of any kind of computer model (whether stochastic or not), it may be useful to remember a basic tenet of physical laboratory experimentation. A model cannot be expected to contain all the details of reality, but our hope is that it faithfully represents the major influences and effects associated with important and interesting characteristics of the system (in this case, the timing pattern of the signals). So, while it may be too much to ask that a model precisely predict the activity of a certain condition, we may hope that it usefully predicts the effect of changing the important characteristics in the absence of any other changes. Classical experimental design and analysis recognizes similar concepts in its use of experimental blocks and focuses on systematic differences among treatments within a block, rather than attempting to predict the result of a specific treatment in an unspecified block.

The authors take this approach when discussing the values of Δ in table 8. A simplified view of the experiment described in this paper is a two-treatment design (signal timing settings) within a single block corresponding to a single definition of T (since the authors assumed that ". . . the conditions in the field for the September data collection would be the same as in May"). Viewed in this way, we realize that the field information addresses the effect of changing c at only one level of T. Would Δ be different at another T specification, for example, traffic conditions at another time of day? Traditional experiments are often designed under the assumption of additive block effects, in hopes that the answer to this question is "no." Additional experiments covering other $(T; c)$ combinations, for example, more blocks and treatments, may be too expensive for practical considerations in studies of this kind. But without them, we are left assuming that the effects of T and c are additive, or understanding that our validation pertains only to the T we have specified.

The authors have selected morning and evening rush hours—undoubtedly the most important conditions when setting signal timing; perhaps it is sufficient for their purposes to certify that the model can predict the effect of changes in the signal timing pattern for these conditions. If it is anticipated that T and c have important "interactions" in reality, and if studies can be extended to cover other traffic conditions, a validation exercise of broader scope might be considered.

NOT WHETHER, BUT WHERE?

Finally, returning to the authors' statement that "no simulator can be expected to capture real behavior exactly," the most natural question to ask will generally not be whether **M** can be thought of as a universal replacement for measurements that are difficult or impossible to make in reality; this simply will not be the case. With careful development and tuning, we may hope for a model that does a respectable job *within some range of conditions*. But just as good "weather" models do not produce good "climate" forecasts and vice versa, models that do a generally good job of modeling traffic in some circumstances may be entirely unrealistic in others. And so a more useful (but more difficult) eventual endpoint of model validation may be the solution to an inverse problem: Under what set of circumstances is **M** a reliable representative of reality, or where in the space of input values can **M** be trusted?

As with the selection of outputs for validation, our ability to usefully answer this question depends both on the range of circumstances of interest and the range of circumstances over which we can expect to collect physical data. It is of little use to consider model validity outside of the first range, and meaningful comparisons will be very difficult, perhaps indirect, and sometimes impossible outside of the second. As with other experiments, however, the goal should be not just a simple answer to one question, but a collection of answers that indicate the sort of situations for which the model (or model-and-sampling process) might be presently "certified," and the identification of other settings within which further study or development is needed. Hence the authors' conclusion that: "CORSIM, though imperfect, is effective in evaluating signal plans in urban networks, at least under some restrictions."

CONCLUSION

All the issues I have noted here can be framed in other ways, and each can be described from entirely different viewpoints. I have variously referred to quantities as random, fixed-and-unmeasurable, or altogether absent as it fits my purposes, while a mechanistic approach might ignore all randomness except that used to define the model, and a full Bayesian approach might always see all quantities as random. Regardless of the perspective, the issues of how and which data are used in setting input values, how validation data are collected and compared with outputs, and how the agreement between outputs and validation data is assessed raise difficult questions. Sacks et al. have done an excellent job of carefully considering and addressing these questions in the context of a specific and important problem. This and other thoughtful exercises of this sort will be the building blocks from which new and better methods for model validation may be developed.

JEROME SACKS

NAGUI M. ROUPHAIL

BYUNGKYU (BRIAN) PARK

PIYUSHIMITA (VONU) THAKURIAH

We thank the discussants for their comments. Their points are well taken and add focus to some key issues in planning or carrying out a validation.

COMMENTS AND QUESTIONS OF RILETT AND SPIEGELMAN

Rilett and Spiegelman raise the question of whether a junior engineer can adopt and implement our approach for validation. Our answer is yes. It would be a terrific learning experience, as it was for us!

Data Quality

They ask about the effect of data quality or accuracy of input data. This is a subject of great importance, wide open for study, and about which little has been done. An exception is Bayarri, Berger, and Molina (referred to in the Analysis of Uncertainty section), who have incorporated in their analysis the presence of observer error in the manually collected data.

Transferability

Is the strategy/methodology transferable to other networks? We think so, provided we stick to urban street networks with few pedestrians. Clearly, attention must be paid to driver behavior. As noted by Rilett and Spiegelman, it is typical to tune driver behavior input distributions. We have been deliberately cautious about doing so for fear of "over-tuning." We did tune in two instances: we changed the geometry by creating a sink and source to remove a CORSIM inability to cope with a congested intersection, and we adjusted the free-flow speed parameter on one corridor to conform to actual field conditions. In a third instance, we noted the effect of driver behavior in utilizing, at one intersection, more green time than ostensibly displayed and would incorporate this change if we were to proceed to a third stage.

A new network may require similar tuning. Our recommendation would be to do so very carefully, in a limited way and only after identifying the specific flaws that can be overcome with defensible tuning. We are some distance away from making this formal, but we are concerned that overly ambitious tuning masks flaws and can fail to account for natural variations.

Accuracy of Predictions

The discussants are correct that more than one instance of accuracy of predictions are needed to assess predictive validity lest the evaluation suffer from the Nostradamus or Babe Ruth effect—dubious though legendary. And somebody has to keep score. Our hope is that this is taken seriously and made part of any program that pursues the establishment and use of simulation models.

Variations in Demand

We agree with Rilett and Spiegelman that major changes in signal plans on large networks could well affect demand rates and lead to unexpected system characteristics. However, dramatic changes in an urban context are unlikely in the short run without major changes to the network geometry, at which point a new context must be faced. We would not advocate predicting characteristics under such new conditions on the basis of old demand rates.

Any changes in turning movements (we did not note any exceptional ones) after implementation of the new plan in September 2000 could not have been the result of adaptation to a new signal plan—the plan was in effect for less than 24 hours before data were collected.

Variations in demand rates and turning percentages are being accounted for in the current work of Bayarri, Berger, and Molina (cited above). Their results quantify a decrease in system performance and an increase in the variability of performance.

Further study of the system under scenarios of different demand (e.g., changing the input data by fixed percentages) could be done; how to make meaningful changes to turning percentages is less obvious. Simulators (unlike CORSIM) that induce routes based on origin-destination information may be more amenable to such study, but these are issues further down the road.

DISCUSSION BY MAX MORRIS

Morris points out the complexity inherent in pursuing a validation strategy that can distinguish among the multiple sources of uncertainty and their effect on validation goals. This can be done, as Morris notes, by a Bayesian formulation and analysis and has recently been carried out by a team of researchers at the National Institute of Statistical Sciences in an application to a deterministic computer model. The application to stochastic simulators such as CORSIM is, in principle, doable; the actual implementation will have considerable complexity and has not yet been done.

Morris notes that the effect of misspecification or inadvertent omission of details in the model can induce a bias that should be accounted for. This can be done by adapting the Bayesian formulation of calibration in Kennedy and O'Hagan (2001) to the current situation, modeling the field data as simulator + bias + measurement error and modeling the bias. How to incorporate issues of variability in the model output vis-à-vis the variability in the field is not clear without more extensive, and expensive, field data.

We agree with Morris that a more extensive data collection and study would be needed to assure validity under different contexts such as different time periods, days of the week, or weather conditions. This is also a point raised by Rilett and Spiegelman. In reality, validation must be an ongoing, and perhaps never-ending, process interacting with model development. At any point in time, we ought to be able to quantify the reliability of the model.

To conclude, we are gratified that the discussants agree with us about the value and importance of pursuing the multiple issues inherent to validation. We are possibly less skeptical than Rilett and Spiegelman about the utility of our approach in practice, but we seem to be in agreement with them and with Morris about what has to be done.

REFERENCE

Kennedy, M.C. and A. O'Hagan. 2001. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society B* 63:425–64.

Road Capacity and the Allocation of Time

DAVID M. LEVINSON
University of Minnesota

SESHASAI KANCHI
ICF Consulting

ABSTRACT

Additional highway capacity gained by increasing travel speed affects the share of time an individual allocates to daily activities, such as commuting and time spent at work, shopping, or at home. Some activities will be undertaken more, others less. Using the 1990 and 1995 Nationwide Personal Transportation Surveys and Federal Highway Administration data, this paper extends previous research that identified and quantified induced demand in terms of vehicle-miles traveled, by considering what type of demand is induced and which activities are consequently reduced. While total travel times did not significantly change between 1990 and 1995, there was a significant change in activity duration. Further, as a result of additional capacity, workers spent less time working and commuting and more time at home and doing other activities. Nonworkers, in contrast, traveled more and spent more time shopping and at home, but less time at other activities. This points out the differences in discretionary and nondiscretionary activities for workers and nonworkers. It also suggests increased highway capacity provides real gains for people, at least in the short term, because time, not vehicle-miles traveled, is the deciding factor for which activities are undertaken and which are eliminated.

David M. Levinson, Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455. Email: levin031@tc.umn.edu

INTRODUCTION

The impact of increasing the capacity of highways is a topic of recent interest. New and faster roads may attract more traffic than is currently diverted from existing roads. This *induced* or *latent* demand can be viewed as a boon or a bane. In the short term, highway expansion is expected to increase travel speeds. In the long run, traffic congestion may approach or exceed earlier levels. If the sole aim of capacity expansion is to reduce congestion, expansions that increase traffic may prove counterproductive. However, that same road construction may increase accessibility and affect people's daily activity patterns. Time savings in travel attained from increased highway capacity enables individuals to allocate more time to their activities and even increase their number, rather than spend time traveling. This ability to take advantage of new opportunities without increasing travel time enables people to achieve greater satisfaction from consumption, change to a better job, or move to a larger house. At a minimum, they should be no worse off. However, this additional travel may have negative environmental consequences, externalities that individuals do not usually consider in their travel decisions.

The objective of this research is to observe the nature of the changes in activity and travel patterns of individuals as a result of additional highway capacity. Time savings from travel due to highway expansion will give individuals more time to spend engaged in different activities. Travel is often referred to as a derived demand, as it reflects an individual's interest in taking advantage of a resource, in this case, increased road capacity. Carefully measuring changes in individuals' travel behavior will facilitate accurate travel forecasting. This research examines the nature of and change in activities that capacity expansion induces and develops a model to quantify the change (in terms of minutes per day) for workers and nonworkers.

A factor complicating the analysis of preferences is whether "time budgets" exist for work travel, all travel, or various activity categories. These time budgets, perhaps just very inelastic preferences, have appeared as empirical regularities in long-term examination of travel behavior. For instance, Levinson and Kumar (1994) found that in Washington, DC, commuting times from home to work averaged 28.5 minutes in 1958, 1968, and 1988. Similar results were found in the Twin Cities of Minneapolis and St. Paul (Barnes and Davis 1999). Furthermore, major changes in metropolitan population, demographics, female labor force participation, and suburbanization suggest that over the long term, individuals adjust their location to maintain approximate constancy in their commute durations, but not necessarily their distances.

Examining all travel, Levinson and Kumar (1995a) did not find the same kind of regularity. First, the share of workers increased, so more individuals traveled to and from work. Second, the additional proportion of workers had more nonwork travel. Previously, in the era of the one-worker, two-adult household, nonwork activities would have been the responsibility of the homemaker. Third, mobility and the near universal presence of a car for each licensed driver has changed the ability to perform nonwork activities outside the home, and as the cost of a favorable activity declines, the amount demanded increases. So while there may be a "commute travel budget," there is some evidence against a "comprehensive travel budget."

Despite the questions about commute and comprehensive travel budgets, there is one type of budget that is inarguable, the daily time budget. The 24-hour day, along with constraints associated with necessary daily activities (working, sleeping, eating, etc.), provide an upper limit on the possible amount of time a person can spend traveling. While the potential for induced time spent traveling may be large, it is not unlimited due to daily time budget constraints. We approached this question, using the Nationwide Personal Transportation Survey (NPTS) and Federal Highway Administration (FHWA) data to measure individual changes and activity patterns, controlling for network changes in each state. We used travel survey data to understand which types of activities and travel are being induced by capacity changes, and consequently, which activities and travel types are being reduced. We developed estimates of time spent in travel and at activities for major activity classifications (home, shop, work, and other) for 1990 and 1995. In our analysis, we controlled for socioeconomic and demographic strata including gender, work status, age, and income, as well as lifecycle categories and population density. For our capacity data, we adopted an approach similar to that used by Noland (1999), employing measures from the *Highway Statistics* series of FHWA. The significant independent variable is lane-miles of roadway, while other independent variables control for population growth, gasoline prices, and income.

This paper begins with a review of the key literature in the induced demand debate, which quantifies the effects of roadway capacity on some aspects of travel demand. This is followed by a brief description of the data used in the analysis. Then travel times and activity durations are compared between 1990 and 1995 using NPTS data. We discuss the theory of time use posed by economists and extend it to better account for the real spatial and temporal constraints that transportation analysts must consider. We pose a set of specific hypotheses concerning how time use should change with increased capacity. Then we develop a model to examine the change in time use between 1990 and 1995 as a function of growth in the highway network, controlling for demographic, spatial, temporal, and socioeconomic characteristics. This requires estimating a time-use model for individuals in the 1990 dataset. We then apply that model to the 1995 survey respondents as an approximation of the latter population's 1990 behavior. The subsequent section applies the difference model approach to determine the impact of highway capacity expansion on travel behavior using seemingly unrelated regression estimation models. A summary of study results concludes the paper.

INDUCED DEMAND RESEARCH

Researchers are trying to identify the extent to which trips are induced, shifted, and lengthened due to capacity expansion. The literature on induced demand suggests the overall elasticities of vehicle-miles traveled (VMT) with respect to lane-miles of capacity to be between 0.5 and 1.0, indicating that a 1% increase in capacity will increase the demand for VMT by between 0.5% and 1.0%.

Dunne (1982) used a representative individual approach to express aggregate demand but ignored the distribution of elasticities across the sample. He then determined point and arc elasticities and compared the weighted elasticity with the elasticity of a representative individual.

Goodwin (1996) conducted a study to verify the presence of induced traffic due to road capacity expansion. Comparing the observed and forecast traffic flows, taking into account the traffic reduction on alternative routes, he found the demand elasticity with respect to travel time (based on short- and long-term timeframes) to be –0.5 and –1.0, respectively.

McCarthy (1997) studied travelers' responses and attitudes toward market-based road pricing, showing that capacity expansion attracted diverted traffic and increased traffic growth induced by improved travel conditions. He found demand elasticity with respect to auto travel time using two different models for four primary modes of travel. By using linear logit and linear captivity models, he determined the demand elasticities to be –0.008 and –0.002, respectively.

Dowling and Colman (1998) studied behavioral change—including mode switch, rescheduling, trip chaining, destination change, and additional trips—responding to the travel time savings as a result of increased highway capacity for the San Francisco and San Diego metropolitan areas. They found that the existing travel forecasting practice probably resulted in an underprediction of 3% to 5% in the number of trips due to time savings that may have been induced by highway capacity expansion.

Hansen and Huang (1997) estimated the induced traffic as a consequence of adding capacity over the short or long run. At the area-wide county and metropolitan level, they found the elasticity of VMT with respect to lane-miles of capacity between 0.62 and 0.94 for periods of 2 and 4 years, respectively. Over a longer run of 10 years, they estimated the elasticity between 0.3 and 0.4 on the highway-segment level.

Noland (1999) studied relationships between lane-miles of capacity and induced VMT by specific road types and estimated long- and short-term elasticities using four different models. The results obtained corroborate the influence of induced travel, at the same time establishing a significant relationship between lane-miles of capacity and VMT. Induced travel was found to have varying influence by road type (Interstates, arterials, and collectors) and by region (urban and rural). He found that with a 1% increase in lane-miles of capacity, VMT grows annually from 0.79% to 1.73% over a period of five years. Using a distributed lag model, he also found that 28.7% of the VMT resulted from an increase in capacity expansion over the five-year period. The same model predicted that induced demand caused 23.7% of the increase in VMT. Noland and Cowart (2000) studied the impact of additional lane-miles on VMT growth using urbanized land area as the instrumental variable for lane-miles of capacity. They found that the impact of lane-mile additions on VMT growth is greater in urbanized areas that had a larger percentage of increases in total capacity and showed that lane-mile elasticities are smaller in the short run (0.284) as compared with the long run (0.904).

Barr (2000) studied Nationwide Personal Transportation Survey data to estimate relationships between average household travel time and VMT and found that individuals would spend 30% to 50% of the time savings from additional capacity on travel.

Fulton et al. (2000) studied county-level data from Maryland, Virginia, North Carolina, and Washington, DC, that related daily VMT to road capacity. They found the elasticities of VMT with respect to lane-miles of capacity to be 0.1 to 0.4 in the short run and 0.5 to 0.8 in the long run.

Marshall (2000), using the Texas Transportation Institute's urban congestion study data for 70 U.S. urban areas, found the elasticities for roadway demand relative to roadway supply as 0.85 for highways and 0.76 for principal arterials using simple regression techniques.

DATA

The travel behavior data used in this analysis come from the 1990/91 and 1995/96 Nationwide Personal Transportation Surveys. These telephone interview surveys collected data on household demographics, income, vehicle availability, location, and all trips made on the survey day. The 1990 NPTS survey was conducted between March 1990 and March 1991 and consisted of almost 22,000 household interviews and over 47,000 persons making almost 150,000 trips. The 1995 NPTS was conducted between May 1995 and June 1996 and consisted of 42,000 household interviews and over 95,360 persons making almost 409,000 trips. While the 1995 NPTS was conducted by giving the respondents a travel diary in advance of their scheduled interview, the 1990 NPTS was conducted over the telephone, which caused some problems. For example, identifying the origin and destination of trips was difficult. We assumed that all tripmakers began and ended their day at home. Due to some improbably high shopping times, we also excluded travelers with a daily shopping time greater than 420 minutes. Given the methodology adopted as a part of this paper, we have tried to minimize the biased nature across both the datasets. We did not drop any data on the basis of day of week, but rather considered both weekday and weekend trips in the analysis and use day of week as an explanatory variable. Table 1 summarizes the number of observations dropped and the reasons for dropping specific records for the 1990 and 1995 NPTS data.

TABLE 1 Summary of Data Analysis Adopted for the 1990 and 1995 NPTS

Description of constraints	1990	1995
Sample size—total trips	**159,832**	**381,388**
Reasons for dropping records		
Invalid destination	3,314	43
Trip in miles >200	23,372	27,455
Travel minutes >120	3,015	5,254
Age >65 years	9,210	35,399
Age <18 years	21,470	63,832
Shop duration >420	707	70
Total dropped	**61,088**	**132,053**
Subtotal at trip level (after records are dropped)	**98,744**	**249,335**
Subtotal at person level	**15,870**	**52,341**
Travel + duration minutes >1,440	7,652	656
Travel + duration minutes <1,440	2,643	17,119
Duration <0	654	2,237
Total dropped	**10,949**	**20,012**
Net total at person level	**4,921**	**32,329**

The time spent at each activity (excluding travel), defined as that activity's duration, was not reported directly in the NPTS. Only the times of the beginning and end of the travel portion of the trip

FIGURE 1 Activity Duration Calculations

Person ID	Origin	Destination	Travel time	Origin time	Destination time	Activity duration (minutes)
1	Home	Other	15	8:30	8:45	30
1	Other	Work	15	9:15	9:30	360
1	Work	Other	15	15:30	15:45	105
1	Other	Other	10	17:30	17:40	20
1	Other	Home	10	18:00	18:20	850
2	Home	Work	20	8:00	8:20	340
2	Work	Other	15	14:00	14:15	

were reported. The activity duration data were obtained by subtracting the destination time of a particular trip from the origin time of the next trip for the same individual, as shown in figure 1. All the activity durations and travel times for an individual add up to the daily time budget of 1,440 minutes (24 hours). The activity duration for the final return home requires that we assume the person's first activity the next day begins at the same time as today's. Thus, we subtracted the time the individual arrives home for the last time in the day from the time of origin of the first trip and add 1,440 minutes. Only those tripmakers whose daily time budget is equal to 1,440 minutes were considered for the study.

The highway data used in the analysis consist of roadway and state characteristics (e.g., lane-miles for all roadways, state's average fuel price, and state population) by state for 1990 and 1995. The data for VMT and lane-miles were obtained from *Highway Statistics* published by the Federal Highway Administration for each roadway type (Interstates, arterials, and collectors) by urban and rural region. We also used data on the population, per-capita income, and cost per energy unit (million Btu) of gasoline by state for all 50 U.S. states for 1990 and 1995. The income and fuel price data are in current year dollars.

TABLE 2 Time-Use Comparisons for 1990 and 1995 Data

	Home	Work	Shop	Other	Travel
FEMALE					
Nonworker					
1995	1,172		42	166	60
	(186)*		(64)*	(170)*	(44)
1990	1,220		35	127	58
	(209)		(70)	(172)	(61
Worker					
1995	944	313	25	93	65
	(226)*	(249)*	(49)*	(132)*	(44)
1990	928	284	30	132	65
	(357)	(357)	(69)	(191)	(64)
MALE					
Nonworker					
1995	1,171		30	177	62
	(200)*		(55)	(184)*	(46)
1990	1,222		29	130	59
	(211)		(60)	(183)	(65)
Worker					
1995	900	365	15	90	70
	(233)	(262)*	(37)*	(136)*	(48)
1990	903	338	20	110	69
	(360)	(367)	(59)	(189)	(71)

* denotes significance at 95% level by difference of means test between 1995 and 1990 results. Standard deviations are in parenthesis.

COMPARISON OF 1990 AND 1995 TIME-USE DATA

This research classifies activities into eight basic categories: time spent at and traveling to the activities of home, work, shop, and other. For a preliminary data comparison of activity patterns in 1990 and 1995, table 2 reports time use by gender and work status. To illustrate, the first row shows that the average female nonworker spent 1,172 minutes at home, 42 minutes at shop, 166 minutes at other, and 60 minutes of travel per day (averaged across all 7 days of the week). In our modeling, we used gender as an explanatory variable. Tables 3 and 4 elaborate the data for 1990 and 1995.

To determine whether these activity durations and travel times for 1990 and 1995 differ for each

TABLE 3 Summary of 1990 Time Use for Different Characteristics of Individuals

Description	Sample size	Time spent at: Travel	Home	Work	Shop	Other
Gender						
Male	1,590	68	929	319	18	107
Female	1,834	65	1,004	217	28	125
Work status						
Worker	2,740	68	906	328	21	117
Nonworker	684	61	1,225	0	31	124
Day of week						
Weekend	1,026	68	1,114	114	30	115
Weekday	2,398	66	907	329	21	117
Lifecycle (number of adults, age of youngest child)						
1, no children	807	73	930	278	22	137
2+, no children	915	65	935	309	21	111
1, 0–5	88	53	1,068	140	25	154
2+, 0–5	524	62	975	282	26	95
1, 6–15	184	76	934	235	26	169
2+, 6–15	423	62	966	277	26	109
1, 16–21	37	66	1,020	227	25	102
2+, 16–21	122	64	980	295	15	86
1, retired, no children	55	70	1,217	27	30	96
2+, retired, no children	269	61	1,128	123	26	102

of these categories, a difference of means (t-test) is performed. The following null and alternate hypotheses were tested:

$$H_o : E(X_1) = E(X_2)$$
$$H_a : E(X_1) \neq E(X_2) \quad (1)$$

The null hypothesis H_o tests for the population means of activity duration and travel time as equal whereas the alternate hypothesis H_a tests for the population means as not equal. Based on the hypothesis above, a t-statistic is calculated to infer whether two data samples differ from one another. It is defined as:

$$t = \frac{E(\overline{X}_1) - E(\overline{X}_2)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \quad (2)$$

where

$E(\overline{X}_1), E(\overline{X}_2)$ = the expected mean value of X_1 and X_2 for first and second dataset,

S_1, S_2 = the variance of the first and second sample set, and

N_1, N_2 = the number of observations in the first and second sample set.

Then, the decision rule is to reject the null hypothesis H_o if $|t|$ is greater than 1.96 (at a 95% confidence interval) and accept otherwise. That rule is applied to all the coefficients to compare the change in time use of individuals between 1990 and 1995.

The time spent at home decreased for nonworkers, remained essentially constant for male workers, and rose for female workers. The time spent at work increased for both male and female workers, which is consistent with the 1990–1991 recession and an expanding economy in 1995. For workers, particularly females, time at home in 1990 substituted for time at work in 1995. The time spent shopping decreased for male and female workers but increased for male and female nonworkers. Similarly, the time spent at other declined for workers but increased for nonworkers. Both are consistent with a strengthening economy in 1995, as workers chose to work more and nonworkers to spend more. The total travel time has either remained stable or slightly increased for all categories, as people in 1995 pursued more out-of-home activities.

TABLE 4 Summary of 1995 Time Use for Different Characteristics of Individuals

Description	Sample size	Time spent at: Travel	Home	Work	Shop	Other
Gender						
Male	12,687	72	917	333	15	103
Female	13,532	65	994	245	28	108
Work status						
Worker	21,512	69	911	351	19	91
Nonworker	4,707	63	1,169	0	37	172
Day of week						
Weekend	5,914	63	1,089	94	34	160
Weekday	20,305	70	918	344	18	89
Lifecycle (number of adults, age of youngest child)						
1, no children	2,084	66	927	321	19	108
2+, no children	8,598	68	936	318	20	97
1, 0–5	260	74	997	192	31	146
2+, 0–5	5,266	68	974	277	21	100
1, 6–15	505	68	950	286	25	111
2+, 6–15	5,227	71	945	296	22	107
1, 16–21	238	68	950	277	24	121
2+, 16–21	1,994	66	936	302	18	118
1, retired, no children	157	64	1,179	0	43	154
2+, retired, no children	1,890	66	1,073	145	33	123

Based on *t*-test values, although the change in activity durations (time spent at home, work, shop, and other) is significant for almost all categories, travel times are, interestingly, insignificant. This supports the "Rational Locator" hypothesis that people adjust their travel choices and relocate their homes and workplaces to maintain their travel commute over time (Levinson and Kumar 1994). The results obtained from a difference of means test showed that the value of $t < 1.96$ for travel, which means we cannot reject the null hypothesis. Thus, the 1990 and 1995 travel times by gender and work status are not different from one another and, thus, this conclusion does not contradict the Rational Locator hypothesis. The rest of the paper aims to determine how individuals reallocate their time due to increased capacity.

CONCEPTUAL MODEL

Becker (1965) proposed a model to study how households use time and market goods to produce useful commodities under the constraints of daily time budgets and income. He suggested that total time could be disaggregated into work and leisure (nonwork) time, but that while people earned money during work time, money was not only not earned but rather was spent in leisure time. Further, both money and time are required to produce household commodities (e.g., preparing dinner, washing dishes, and watching television). Additional time could be assigned to work to increase income or to leisure to increase pleasure. The value of additional income (requiring additional time) is diminishing because the amount of time available to produce household commodities decreases as time allocated to work increases. Jara-Díaz (2000) synthesizes much of the subsequent research on time allocation models, suggesting the following utility maximization equation, subject to separate money and time constraints:

$$\text{Max } U(G, T_L, T_W, t) \quad (3)$$

subject to

$$wT_W - G \geq 0 \quad (\lambda) \quad (4)$$

$$\tau - (T_L + T_W + t) = 0 \quad (\mu) \quad (5)$$

where

U = utility function,

G = aggregate consumption in money units,

T_W = time assigned to work,

T_L = time assigned to leisure,

t = exogenous travel time,

w = wage rate (work),

τ = total time available,

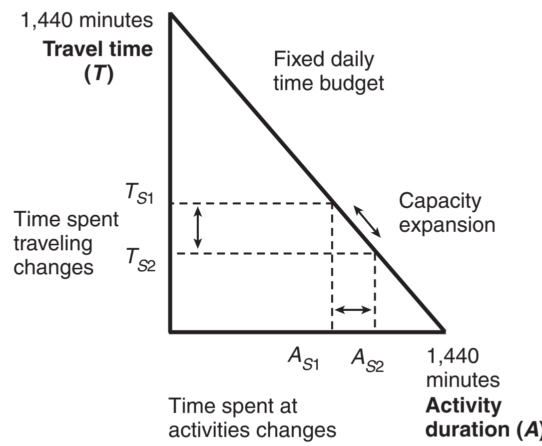
μ = Lagrange multiplier of time restriction,

λ = Lagrange multiplier of income restriction.

Such a model should be extended to separate out leisure from necessary non-income producing activities, such as shopping or going to school. In addition, scheduling of activities is also a critical factor (Small 1983). This model makes no representation of nonworkers and their time allocation, as nonworker revenue is independent of daily time spent in the paid labor force, although it may be a function of previous time in the labor force. While the economic framework is informative, it cannot tell us how individuals actually substitute travel and activity time, as that depends on the empirical valuations that people place on work, leisure, travel, etc. Thus, the allocation of time to activities is a complex phenomenon that does not lend itself easily to nonempirical analysis.

Nevertheless, we accept the premise that individuals balance time at and travel time to activities to maximize their utility, acting to attain economies in activity consumption. The relationship between travel times and activity durations is shown in figure 2, where the allocation is constrained so that travel times and activity durations sum to 1,440 minutes. In general, if the time/demand for travel is inelastic (i.e., if a reduction in the time required to travel a given distance does not increase the distance traveled enough to produce an increase in the total amount of time spent traveling), we would expect that greater highway capacity would cause travel time (T_{S1}) to fall to T_{S2}, leading to more time spent at nontravel (work, leisure) activities. However, if the additional capacity increases the range and quality of activities that users can reach, they may be willing to travel farther and spend more time to reach different (and better) activities, at the cost of less time spent at nontravel activities. While the theoretical possibilities depend on the distribution of activities (an individual's accessibility field), we expect that most travelers in metropolitan areas would use the capacity (faster speed) in part to increase distance and to reduce travel time, thus increasing time spent at activities.

FIGURE 2 Daily Travel and Activity Time Production Function



Moreover, if the journey is its own reward, or other benefits from travel alone are achieved, an increase in speeds may lead to more time spent traveling (just as speeds increase VMT and number of trips). This interpretation is consistent with Redmond and Mokhtarian's (2001) research that there is some positive benefit from traveling.

The mix between work and nonwork activities is indeterminate from the figure 2 analysis and requires empirical estimation. Furthermore, a distinction should be made between time at work and time spent working (presumably paid). As most employees know, payment is usually based on arrangements with employers to work a fixed number of hours. Time arriving early at work is not necessarily compensated, whether or not work is actually performed.

For obvious reasons, we evaluate separate models for workers and nonworkers. The total travel time to work and the time spent at work are zero for nonworkers, while they form a significant part of the daily time budget for workers, hence nonworkers will use this time for travel or spend it at other activities.

The methodology determines time use separately for workers and nonworkers with income as one of the independent variables in the model. We note that not all workers work full-time, but since ours is a daily analysis, we only know how much time a worker puts in on a given day, whether or not that is typical. Our analysis of the empirical data (shown in table 2) illustrates the differences in time spent at work between 1990 and 1995. Also, the estimation of different models for 1990 and 1995 embeds the differences in time spent at work in the coefficients, which is critical for our comparison of the effects of roadway capacity, discussed in subsequent sections. We conducted additional analysis to measure differences in underemployment between 1990 and 1995.[1]

The lane-miles of capacity increased for all road classes between years 1990 and 1995 (Interstates, arterials, and collectors). The research initiative proposed by this paper is to measure changes in individual time use with increase in capacity over a short-term period. We expect capacity increases are associated with positive time savings over a short time period.

HYPOTHESES

Workers

Our first concern is to determine the significance of additional capacity expansion on individual travel patterns for workers. Due to increasing highway capacity, the cost of travel drops as drivers attain higher speeds and reliability, which enables individuals to travel longer distances in the same amount of time. Since work travel is something workers would prefer to avoid (notwithstanding Mokhtarian and Redmond, since we are looking at short-term changes), we expect that every additional unit of highway capacity will decrease work trip travel times.

Time spent at work is somewhat more complicated. The economic models' suggestions are ambiguous as to where travel time savings will be spent. We assume that there is no concomitant income or work productivity change. Thus, we do not believe there will be any associated change in time spent in paid work. However, as noted above, paid work and time at work differ. Our hypothesis is thus related to scheduling and road reliability. Road capacity increases reliability (reduces variance in travel time). With increased capacity and faster speeds, the time spent at work will decrease due to reduced peak spreading. It is expected that the more reliable the roads, the more likely people get to work at their desired arrival time. This may lead to fewer early departures from home to avoid potential congestion. Thus, people will naturally spend less time early at work when they depart from home later.

In the evening, the desired departure time from work is unchanged, and due to time savings from travel (and increased reliability), workers will arrive home a few minutes earlier. (Some travelers may have departed earlier to avoid congestion in the evening; others may have departed later: these effects are thought to be offsetting). Thus, workers will be able to arrive at their work place later in the morning (but still on time), no longer needing to leave early to escape the brunt of traffic congestion, and will leave at about the same time in the evening. In all, it is expected that with increased capacity, there will be less variability in commuting travel times, resulting in less time spent at work.

Travel time to shop decreases with highway expansion because of faster roadways. Less time is spent shopping due to fewer shopping trips at larger more comprehensive stores. We expect road changes to be largely independent of income changes. However, time spent shopping is not simply a discretionary or nondiscretionary activity, and we have no reason to expect a priori that shopping is the province of high-income individuals. For instance, shoppers with less income should bargain hunt at more places in order to get the most value per dollar spent, which would increase their time spent shopping. On the other hand, individuals with higher income shop in part as a leisure activity. Thus, it is expected that income on average will have a largely nullifying affect for time spent at shopping activities among workers.

The 1990 to 1995 period saw the emergence of "big box" retailers that created scale economies on

[1] We used time spent at an activity (including time at work) as an instrument to estimate travel time to an activity (and vice-versa) but the coefficient estimates using instrumental regression were not found to be significantly different. This was done by using Hausman's specification error test to check whether a regressor is truly exogenous to the equation. (The results are detailed in Kanchi 2001). For both workers and nonworkers it is observed that the p-value corresponding to χ^2 was significantly higher than that of the $\alpha = 0.1$ (90% confidence interval). Thus, in order to keep our model simple we did not use instrumented variable regression in our analysis, because the results were not significant at the 90% confidence interval.

both the production and consumption side. These retailers were enabled by large, new truck-based just-in-time distribution systems and suburban freeways. So instead of many small stores, there are fewer but bigger retail stores, which sell a wider variety of goods. Time at shopping may be more often restricted to one big store rather than many smaller stores, and thus should decline as shoppers achieve economies of scale in consumption.

Travel time to other, as with travel time to work and to shop, decreases with capacity expansion because of time savings from faster roadways. Capacity expansion, which is mostly in fast growing suburbs, leads to the establishment of new activity centers. Because the nature of other activities for workers tends to be for pleasure and entertainment, the time spent at these activities will increase with highway capacity.

The flip side of travel time to work is travel time to home, which will similarly decrease with each unit increase in highway capacity. Workers are expected to spend part of the travel time saved at home. Travel is the cost associated with pursuing activities of interest and, hence, it can be considered the price (means) for undertaking activities (ends). Of the four activity durations (home, work, shop, and other), work and shop are necessary to fulfill an individual's daily needs and are "constrained" activities, while home and other are pleasure-maximizing "unconstrained" activities.

Nonworkers

In addition to the obvious difference in time spent at work, the major difference between the travel pattern of workers and nonworkers is that nonworkers spend more time at other activities (enabled by avoiding 300 minutes a day of work). This provides nonworkers more time and flexibility to take additional trips than workers. The qualitative meaning of some activities differs for nonworkers. In contrast to workers, nonworkers' shopping is a much more recreational or unconstrained activity. On the other hand, other activities may be less discretionary for nonworkers, as that population includes full-time students. School would be a primary activity, which can be considered similar to work for a worker. Hence, "other" is a more constrained activity. Time savings in transportation may relax the peak spreading for other activities for nonworkers as it did for work activities for workers.

On the whole for nonworkers, the frequency of home and shopping trips was higher than that for other activities. Thus, as capacity increases, nonworkers are expected to pursue more shopping-related activities. Hence, the destination travel times for home and shop tend to rise with increasing capacity, while the travel time to "other" decreases due to travel time savings associated with higher speeds. As with workers, time spent at home is a pleasure-maximizing unconstrained activity, and due to travel time savings from highway expansion, the time spent at home is expected to increase.

MODEL

Though we want to know how an individual in the 1995 survey would have behaved in 1990, unfortunately, the NPTS was not conducted as a panel survey. To compensate for this, we engaged in a two-stage procedure whereby we first estimated a model of 1990 individuals and then applied that model's coefficients to 1995 individuals. This enabled us to measure changes in behavior, controlling for as much variability as possible in socioeconomic, demographic, spatial, and temporal variations. The model to estimate time at each of the eight activities for a 1990 individual is:

$$T_{90i} = f(A,D,G,H,L,M,S,W) \tag{6}$$

Subject to

$$\left(\sum_{i=1}^{8} T_{90i} \right) = 1{,}440 \tag{7}$$

where

T_{90i} = time spent at activity i;

i = index of activities (travel time to and duration at home, work, shop, and other);

A = age;

D = local population density;

G = gender;

H = household income levels;

L = family lifecycle characteristics;

M = month of year interview was conducted;

S = state-specific variables;

W = day of week interview was conducted.

We selected these variables because of their availability and their significance in previous analyses of travel behavior by the authors (Levinson 1999; Levinson and Kumar 1995a, 1995b, 1997). The above model analysis was performed at the individual level rather than at the state level. This approach was employed because aggregation at the state level would yield 33 observations (1 for each state, with a number of states suppressed in the analysis because they had too few observations), which, due to many fewer degrees of freedom, would diminish the explanatory power of the model. We used states as explanatory variables to estimate the individuals' time use in 1990 and 1995. Dummy variables (0,1) were employed for each of the characteristics. The variables were entered linearly into the model.

The final model for T_{90i} was estimated using Zellner's seemingly unrelated regression subjected to the daily time budget constraint of 1,440 minutes. Seemingly unrelated regression estimation (SURE) models use asymptotically efficient, feasible, generalized least squares estimation (Greene 1997). The daily time budget constraint makes the covariance matrix of residual errors singular, which cannot be determined directly by SURE, so we dropped one equation and estimated the other seven simultaneously. The final dropped equation can then be calculated using the mathematical constraint equation, because the remaining coefficients and their sums are known. The SURE model is preferred over ordinary least squares (OLS) regression, because it overrules the assumption that error residuals are not interrelated. SURE estimates the whole model as a system of equations rather than one by one as in OLS. The coefficients from this model are shown in appendix table A1.

The equations for T_{90i} for 1990 obtained from the first stage were used to determine $\hat{T}_{90i}$ (an estimate of the travel times and activity duration that 1995 individuals had in 1990) subject to the reported socioeconomic, demographic, spatial, and temporal characteristics of each 1995 respondent. Simply put, we took the estimated 1990 time-use equations and applied them to the 1995 data.

We used $\hat{T}_{90i}$ to estimate a difference model of change in travel behavior between the 1995 individuals reported (or computed) activity times and the best estimate of their 1990 behavior. We evaluated two models (one for workers and one for nonworkers) in the form given below.

$$\begin{aligned}&\Delta T_i\\&= f(\Delta C / C_{90}, \Delta F / F_{90}, \Delta I / I_{90}, \Delta P / P_{90}, D_{95}, G_{95}, L_{95})\end{aligned} \tag{8}$$

Subject to

$$\sum_{i=1}^{8} \Delta T_i = 0 \tag{9}$$

where

$\Delta T_i = T_{95i} - T_{90i}$ Change in time at activity i between 1995 (reported) and 1990 (estimated),

i = index of activities,

ΔC = difference in lane-miles for all roadway types between 1995 and 1990,

C_{90} = sum of lane-miles for all roadway types in 1990,

ΔF = difference in state average fuel prices between 1995 and 1990,

F_{90} = state average fuel price in 1990,

ΔI = difference in state average per capita income between 1995 and 1990,

I_{90} = state-level per capita income in 1990,

ΔP = difference in state population between 1995 and 1990,

P_{90} = state population in 1990,

D_{95} = local population density estimates in 1995,

G_{95} = gender as noted in 1995 survey,

L_{95} = family lifecycle characteristics in 1995.

Since all eight activities in the 1990 and 1995 surveys are constrained by the individual daily time budget of 1,440 minutes, their differences sum to 0 minutes. A SURE is run on the above system of equations considering the ΔT_i (for each of eight activities (six for nonworkers)) as dependent variables. Again, all variables are entered linearly. Because the system of equations forms a singular error variance matrix, one of the equations is dropped and a SURE model is run on seven equations for workers (five equations for nonworkers) and the final dropped equation is obtained from the mathematical constraint. The full results are shown in appendix table A2 for workers and appendix table A3 for nonworkers.

TABLE 5 Elasticity of Time with Respect to Capacity

Dependent variable:	Workers		Nonworkers	
Change in:	Elasticity	Minutes	Elasticity	Minutes
Travel time to				
Home	–3.17E–04	–0.0108	1.48E–02	0.528*
Work	–7.06E–03	–0.123*	NA	NA
Shop	–4.71E–02	–0.190*	3.39E–02	0.235*
Other	–9.80E–03	–0.160*	–2.91E–02	–0.606*
Activity duration at				
Home	7.27E–03	6.56*	2.19E–03	2.60*
Work	–1.80E–02	–5.66*	NA	NA
Shop	–3.44E–02	–.767*	2.54E–02	1.19*
Other	2.72E–03	0.349	–2.83E–02	–3.95*

* Denotes significance of the variable at 95% level.

RESULTS

A summary of the final SURE results is displayed in table 5, which shows the elasticity of travel times and activity durations with respect to lane-miles of capacity. The elasticity η of independent variable x with respect to its dependent variable y is given by

$$\eta = \frac{dy / y}{dx / x} \tag{10}$$

The elasticities described here represent the percentage increase in change in time use with a 1% change in capacity. Thus, to illustrate table 5, for every 1% increase in capacity, workers decrease their travel time to home by 0.000317% or 0.0108 minutes, travel time to work by 0.00706% or 0.123 minutes, and so on. Hence, these represent the change in time use with respect to capacity.

While the numbers may appear small, a 1% increase in capacity increases time spent at home by over 6 minutes and reduces time at work by 5 minutes. As these numbers are estimated from state capacity data, it can be expected that local effects from a new or expanded roadway would be much greater. The results displayed in table 5 are consistent with the underlying hypotheses for both workers and nonworkers. The difference between worker and nonworker models is primarily due to the presence of an extra 300 minutes for nonworkers to pursue additional activities.

It is found that nonworkers, when given additional capacity, prefer shopping while workers pursue other activities. This is due to the qualitative shift in behavior between shop and other for workers and nonworkers, which yields such travel and activity behavioral patterns. Thus, it is important to model each category separately to determine its respective effect. Also, we found that with capacity expansion, individuals pursue more unconstrained activities (home and other for workers, home and shop for nonworkers), which presumably increases their utility. A somewhat surprising result is that additional roadway capacity leads to a net increase in time spent traveling by nonworkers (in contrast with workers). This lends credence to the idea that travel itself has a positive utility for nonworkers.

CONCLUSIONS

We observed that overall travel times have remained statistically unchanged between 1990 and 1995, while a significant change is observed in activity durations, both of which are in agreement with previous analyses. Linking a panel of highway data for the first time with time series travel behavior data suggests that while VMT may increase with capacity, the time spent traveling remains fairly stable. Furthermore, the effects on workers and nonworkers are different.

Using a simultaneous equation estimation difference model approach, this research shows how travel times and activity durations are affected by increasing highway capacity. We found that increases in highway capacity bring about small but statistically significant changes in individual daily travel behavior. Workers use the capacity expansion to spend more time at home and other activities, and spend less time at work. Nonworkers choose to use the additional capacity both for activities at home and for shopping. These observations may be somewhat surprising; however, we have found no alternative hypothesis consistent with the data, nor have we found (to date) any data that contradict the hypothesis. This analysis is the first to measure these variables as a function of road capacity. As such, it serves as a marker for future research to corroborate or refute. While there is clearly induced travel, we now have a better understanding of which travel and activities are induced with capacity and which are reduced.

ACKNOWLEDGMENTS

The authors of this paper would like to thank the U.S. Department of Transportation and DOT's Federal Highway Administration for providing the 1990 and 1995 NPTS data, and Dr. Bob Noland for his help in providing highway data to perform important analyses used as a part of this paper. Earlier versions of the paper were presented at the Induced Demand workshop at the University of California at Berkeley in June 2000; the International Association of Travel Behavior Research in Gold Coast, Australia, in July 2000; and the Western Regional Science Association meeting in Palm Springs, California, in February 2001: participants are thanked for their input. The authors would also like to thank Elva Chang, Gary Davis, David Gillen, and Gerard McCullough. The California Department of Transportation and the California PATH program at the University of California at Berkeley provided support as part of the project "Evaluation Methods for Measuring the Value of ITS Services and Benefits from Implementation." The University of Minnesota provided additional support. The authors are responsible for the analysis, opinions, and any errors.

REFERENCES

Barnes, G. and G. Davis. 1999. Understanding Urban Travel Demand: Problems, Solutions, and the Role of Forecasting, report #2 in the series: *Transportation and Regional Growth*, CTS 99-02. Center for Transportation Studies, University of Minnesota.

Barr, C.L. 2000. Testing for the Significance of Induced Highway Travel Demand in Metropolitan Areas. *Transportation Research Board 79th Annual Meeting Preprint CD-ROM*. Washington DC: Transportation Research Board, National Research Council. January.

Becker, G. 1965. A Theory of the Allocation of Time. *The Economic Journal* 75:493–517.

Dowling, R.G. and S.B. Colman. 1995. Effects of Increased Highway Capacity: Results of Household Travel Behavior Survey. *Transportation Research Record 1493*, 143–49.

_____. 1998. Effects of Increased Highway Capacity: Results of a Household Travel Behavior Survey. *Highway Capacity Expansion and Induced Travel—Evidence and Implications: Transportation Research Circular No. 481*. Washington, DC: Transportation Research Board, National Research Council. February.

Dunne, J.P. 1984. Elasticity Measures and Disaggregate Choice Models. *Journal of Transport Economics and Policy* 18(2):189–97.

Fulton, L.M., R. Noland, D. Mezzler, and J. Thomas. 2000. Statistical Analysis of Induced Travel Effects in the U.S. Mid-Atlantic Region. *Transportation Research Board 79th Annual Meeting Preprint CD-ROM*. Washington DC: Transportation Research Board, National Research Council. January.

Goodwin, P.B. 1996. Empirical Evidence on Induced Traffic. *Transportation* 23:23–54.

Greene, W.H. 1997. *Econometric Analysis*. New York, NY: Macmillan.

Hansen, M. and Y. Huang. 1997. Road Supply and Traffic in California Urban Areas. *Transportation Research: Part A, Policy and Practice* 31A:3 205–18.

Jara-Díaz, S. 2000. Allocation and Valuation of Travel Time Savings. Universidad de Chile, Santiago.

Kanchi, S. 2001. Time Use and Capacity Expansion, Master's thesis. Department of Civil Engineering, University of Minnesota.

Levinson, D.M. 1999. Space, Money, Lifecycle, and the Allocation of Time. *Transportation* 26:141–71.

Levinson, D.M. and A. Kumar. 1994. The Rational Locator: Why Travel Times Have Remained Stable. *Journal of the American Planning Association* 60(3):319–31.

_____. 1995a. Activity, Travel, and the Allocation of Time. *Journal of the American Planning Association* 61(4):458–70.

_____. 1995b. Temporal Variations on the Allocation of Time. *Transportation Research Record* 1493:118–27.

_____. 1997. Density and the Journey to Work. *Growth and Change* 28(2):147–72.

Marshall, N.L. 2000. Evidence of Induced Demand in the Texas Transportation Institute's Urban Roadway Congestion Study Data Set. *Transportation Research Board 79th Annual Meeting Preprint CD-ROM*. Washington DC: Transportation Research Board, National Research Council. January.

McCarthy, P.S. 1997. The Role of Capacity in Aggregate Share Models of Intercity Passenger Travel. *Journal of Transport Economics and Policy* 31(3):293–308.

Noland, R.B. 1999. Relationships Between Highway Capacity and Induced Vehicle Travel. *Transportation Research Board 78th Annual Meeting Preprint CD-ROM*. Washington DC: Transportation Research Board, National Research Council. January.

Noland, R.B. and W.A. Cowart. 2000. Analysis of Metropolitan Highway Capacity and the Growth in Vehicle Miles of Travel. *Transportation Research Board 79th Annual Meeting Preprint CD-ROM*. Washington DC: Transportation Research Board, National Research Council. January.

Redmond L.S. and P.L. Mokhtarian. 2001. The Positive Utility of the Commute: Modeling Ideal Commute Time and Relative Desired Commute Amount. *Transportation* 28(2):179–205.

Small, K. 1982. Scheduling of Consumer Activities: Work Trips. *The American Economic Review* 72:467–79.

TABLE A1 Coefficients from the Estimated Model of 1990 Time-Use Behavior

Independent variables	Workers								Nonworkers					
	Travel to				Time at				Travel to			Time to		
	Home	Work	Shop	Other	Home	Work	Shop	Other	Home	Shop	Other	Home	Shop	Other
States														
Alabama	4.08	8.22	0.10	–1.58	7.49	–10.65	–17.26	9.59	–7.75	–0.93	1.24	–42.54	0.73	49.26
Arizona	4.64	10.25	–0.61	–2.52	–69.08	63.95	–6.98	0.35	16.00	–0.07	4.02	–89.63	37.90	31.79
Arkansas	–4.45	–2.56	–0.69	–0.37	–13.45	–35.20	–11.95	68.67	–18.51	3.22	–5.88	37.64	–7.03	–9.44
California	5.79	3.78	0.53	–0.42	–28.74	–10.22	4.51	24.76	–7.19	–1.34	2.01	–5.71	2.82	9.41
Colorado	5.30	6.01	0.87	–5.15	–6.26	8.85	–3.93	–5.68	2.58	–2.96	–2.07	36.56	–11.58	–22.53
Connecticut	2.35	0.70	0.33	–0.18	–16.61	20.37	–5.23	–1.72	–6.72	–3.79	–4.30	53.85	–7.43	–31.59
Florida	0.89	2.53	–0.49	–1.97	–20.05	–3.71	–2.76	25.56	–9.10	–0.93	–4.40	15.50	–1.56	0.49
Georgia	0.24	3.63	–0.88	2.15	–62.20	59.71	–4.79	2.14	1.73	0.86	–1.21	20.93	–1.92	–20.39
Illinois	1.84	4.87	–1.73	–1.97	–52.87	49.98	–7.78	7.66	10.16	–1.49	2.50	–11.40	–5.59	5.83
Indiana	4.83	3.37	–0.85	–0.59	–52.60	22.70	–3.18	26.32	–0.14	–1.54	–2.74	31.41	–1.26	–25.73
Iowa	3.36	1.32	–0.66	–5.49	–9.14	–2.51	–13.13	26.25	–18.58	6.09	0.83	–26.06	38.96	–1.24
Kansas	–8.43	–5.48	–1.63	–1.33	117.28	–125.98	13.72	11.86	13.82	3.45	–7.14	–84.28	3.32	70.83
Kentucky	7.09	3.51	–2.77	4.33	–64.75	54.95	–16.49	14.13	–11.54	–0.88	18.34	30.37	–1.80	–34.49
Louisiana	–3.32	1.50	–1.20	5.07	–25.95	21.57	–12.05	14.38	–12.03	–3.46	–6.17	66.76	–4.13	–40.98
Maryland	8.23	7.90	–0.51	0.64	–74.10	14.91	–7.21	50.14	–9.23	–2.13	–4.82	41.47	–13.44	–11.84
Massachusetts	6.62	0.50	0.12	–0.98	25.53	–46.75	–8.60	23.56	–1.91	0.68	–10.04	73.53	5.18	–67.44
Michigan	11.67	0.42	0.09	–2.36	–55.63	–4.24	1.22	48.83	–9.94	–0.45	–8.35	46.06	20.95	–48.27
Minnesota	0.50	2.24	0.74	–3.68	–9.41	–12.18	1.80	20.00	8.83	–0.63	–8.08	–3.43	12.58	–9.27
Mississippi	–4.26	8.99	2.64	7.87	–81.30	16.14	6.07	43.85	–1.91	0.53	5.17	–4.10	0.22	0.10
Missouri	–2.06	8.29	–1.75	–4.17	–65.69	126.57	–18.81	–42.38	–2.26	6.54	–14.02	62.68	63.93	–116.87
New Jersey	8.14	10.02	–0.30	–2.09	–43.83	32.02	–11.02	7.05	–2.99	–1.45	–11.13	66.98	–12.07	–39.33
New York	8.02	6.24	2.40	–0.44	–49.95	33.45	–4.57	4.84	0.52	–2.25	–2.80	32.44	–2.72	–25.18
North Carolina	–5.12	3.69	–1.32	1.60	–42.68	70.62	–10.94	–15.85	–14.83	–2.01	3.39	41.56	–2.76	–25.36
Ohio	3.18	–1.02	1.69	1.14	–8.93	–22.73	–3.32	29.99	–7.08	–0.73	–4.04	0.37	1.38	10.10
Oklahoma	–5.36	0.89	–2.39	–2.00	62.88	–60.23	–14.28	20.50	1.03	0.53	–6.75	33.20	13.43	–41.44
Oregon	–5.24	0.00	–1.45	–3.67	–54.69	80.52	–6.32	–9.14	–18.90	7.77	–4.72	33.24	40.54	–57.93
Pennsylvania	–0.24	2.15	–1.06	–1.74	–33.18	35.11	–13.03	11.99	2.43	0.01	–4.46	28.57	9.22	–35.77
South Carolina	–2.81	–0.49	0.88	–4.82	–8.06	48.09	–12.18	–20.60	–14.55	–2.64	–10.12	46.80	2.35	–21.84
Tennessee	–5.19	0.52	0.44	–3.31	–26.33	41.12	–3.67	–3.58	–13.07	–0.69	–8.02	55.22	–6.33	–27.11
Texas	5.88	2.28	0.22	1.06	–16.49	–5.60	–7.20	19.86	–7.20	–1.02	–2.70	42.42	–10.12	–21.39
Virginia	2.76	1.86	–0.25	–2.52	2.10	–1.84	–1.96	–0.16	–12.43	–0.27	–9.36	37.95	–5.82	–10.08
Washington	4.94	7.82	3.79	1.12	–60.56	20.67	–7.39	29.61	0.00	–0.34	4.04	1.87	–14.80	9.22

continues

TABLE A1 Coefficients from the Estimated Model of 1990 Time Use-Behavior (*continued*)

Independent variables	Workers								Nonworkers					
	Travel to				Time at				Travel to			Time to		
	Home	Work	Shop	Other	Home	Work	Shop	Other	Home	Shop	Other	Home	Shop	Other
Population density														
0–99	6.27	1.50	0.69	–11.13	51.64	11.52	–1.56	–58.92	0.72	–3.20	–4.80	99.93	–42.69	–49.96
100–249	3.21	1.53	1.53	–12.05	45.45	14.39	4.31	–58.37	3.42	–3.04	–2.63	67.04	–36.60	–28.20
250–499	2.63	2.17	0.65	–11.81	46.23	12.74	6.19	–58.78	–8.99	–2.23	–7.62	80.54	–32.35	–29.35
500–749	4.82	1.11	0.32	–11.29	57.90	16.56	2.26	–71.68	4.48	–1.85	–0.66	55.25	–21.79	–35.42
750–999	–3.02	–0.59	2.03	–11.10	52.86	8.84	12.68	–61.70	2.15	1.89	3.94	35.78	–26.67	–17.09
1,000–1,999	0.68	1.91	0.29	–9.60	60.44	3.11	1.46	–58.29	–4.60	0.57	–1.34	54.86	–32.11	–17.38
2,000–2,999	1.28	0.09	0.91	–11.37	82.81	–26.88	6.13	–52.98	–9.63	–1.92	–2.55	78.60	–37.71	–26.78
3,000–3,999	2.66	1.27	0.26	–13.74	33.88	28.64	1.30	–54.27	1.07	–3.07	–7.54	54.27	–30.42	–14.30
4,000–4,999	0.37	2.26	1.17	–11.23	33.87	17.89	8.71	–53.04	–13.85	–1.35	–11.32	113.92	–31.33	–56.06
5,000–7,499	1.62	4.86	0.56	–10.17	56.93	–2.72	0.21	–51.29	–4.59	–1.51	–0.92	39.32	–21.93	–10.37
7,500–9,999	–1.39	2.42	1.45	–8.35	106.74	–41.52	1.81	–61.17	1.72	–1.70	–6.95	101.79	–38.47	–56.39
10,000–49,999	9.31	6.42	–0.18	–12.91	86.64	–8.91	2.29	–82.66	–5.14	0.92	–4.45	71.79	–24.54	–38.58
50,000+	4.63	18.45	–0.58	–12.53	36.88	24.46	12.18	–83.48	–6.63	1.17	3.17	124.52	–7.43	–114.80
Household income														
Less than $5,000	–1.21	–1.40	0.32	0.11	100.80	–103.15	6.70	–2.18	–7.77	0.46	–5.61	38.15	8.01	–33.25
$5,000–$9,999	–8.08	–0.65	0.29	1.74	–1.09	15.02	10.49	–17.71	–2.33	0.19	–4.55	20.55	–0.63	–13.23
$10,000–$14,999	–5.47	–0.58	0.28	1.65	18.65	–21.63	–1.09	8.19	1.33	–0.06	–1.43	22.08	2.00	–23.92
$15,000–$19,999	–5.34	0.04	0.70	0.54	25.48	–34.85	0.07	13.37	–1.81	1.31	–1.72	34.05	–0.97	–30.85
$20,000–$24,999	–3.89	–1.66	0.17	1.93	6.71	–27.35	2.27	21.82	0.09	1.49	–1.11	16.15	11.32	–27.93
$25,000–$29,999	–2.77	–1.12	0.64	5.19	–4.72	–36.87	7.87	31.78	1.42	2.66	8.55	–19.02	8.42	–2.03
$30,000–$34,999	–0.14	–2.00	1.44	4.14	–14.58	–42.35	10.77	42.72	–0.64	1.44	0.70	29.57	8.16	–39.23
$35,000–$39,999	–2.11	–0.21	1.10	4.05	–17.21	–17.94	8.11	24.19	1.57	0.38	2.28	–4.00	9.01	–9.24
$40,000–$44,999	2.50	1.98	0.22	4.30	–14.24	–22.04	–0.96	28.24	–2.42	2.19	2.53	–52.43	27.00	23.14
$45,000–$49,999	7.20	1.07	2.70	11.24	6.25	–83.07	10.01	44.59	3.26	2.19	1.43	–11.87	7.10	–2.12
$50,000–$54,999	4.36	2.83	1.43	3.62	0.70	–52.36	7.90	31.51	2.89	0.51	2.77	5.16	15.48	–26.80
$55,000–$59,999	7.30	3.82	1.14	6.85	15.51	–67.76	16.44	16.70	14.92	6.70	–0.12	–71.95	41.57	8.88
$60,000–$64,999	9.76	3.41	0.52	4.90	–21.25	–86.72	7.32	82.07	–10.14	2.45	–3.75	4.34	15.16	–8.06
$65,000–$69,999	3.87	–2.12	1.08	10.39	–32.03	–46.54	10.28	55.08	–3.01	–2.18	2.57	–20.73	24.28	–0.93
$70,000–$74,999	9.78	5.23	6.73	9.01	–49.11	–71.92	26.28	64.00	0.21	–3.20	9.88	–38.62	22.65	9.09
$75,000–$79,999	–2.96	0.35	–0.53	9.32	27.93	–83.91	–1.98	51.77	–20.04	1.91	–9.29	96.07	28.85	–97.50
$80,000+	6.39	2.63	0.78	8.88	–1.96	–67.55	–0.92	51.76	10.31	3.01	16.78	–37.46	1.29	6.06

continues

TABLE A1 Coefficients from the Estimated Model of 1990 Time-Use Behavior (*continued*)

Independent variables	Workers: Travel to Home	Travel to Work	Travel to Shop	Travel to Other	Time at Home	Time at Work	Time at Shop	Time at Other	Nonworkers: Travel to Home	Travel to Shop	Travel to Other	Time to Home	Time to Shop	Time to Other
Lifecycle														
(Adults, youngest child age)														
1, NA	–62.98	2.41	3.90	–34.23	22.65	74.38	–29.52	23.39	–10.53	–3.93	–9.25	7.20	–28.25	44.77
2+, NA	–68.04	4.03	2.70	–38.88	9.27	133.11	–37.27	–4.91	–5.88	–2.32	–8.54	12.26	–27.58	32.05
1, 0–5	–65.13	0.85	3.79	–33.00	36.75	31.22	–39.53	65.05	–12.46	–5.12	–11.00	35.16	–26.74	20.15
2+, 0–5	–69.53	6.55	2.68	–39.59	39.76	105.05	–36.00	–8.93	–15.23	–3.37	–11.76	39.04	–28.63	19.96
1, 6–15	–59.88	3.43	2.74	–33.36	13.98	59.15	–33.00	46.94	–11.05	–2.05	0.34	–78.44	–21.88	113.08
2+, 6–15	–66.99	3.07	2.81	–40.53	41.95	105.62	–35.33	–10.60	–8.21	–1.68	–10.14	–15.78	–20.75	56.56
1, 16–21	–66.99	3.03	2.81	–37.59	97.23	38.36	–37.64	0.78	–4.59	–3.07	–20.35	63.81	–31.67	–4.12
2+, 16–21	–67.34	–0.64	2.04	–39.22	45.49	109.62	–40.56	–9.40	–4.06	–3.63	–9.81	–34.44	–31.49	83.43
1, retired, NA	–35.88	–7.34	–0.11	–18.44	222.12	–77.56	–59.52	–23.28	–6.24	–2.62	–8.33	17.80	–26.21	25.58
2+ , retired NA	–64.05	4.55	1.92	–38.51	55.79	78.88	–37.43	–1.14	–8.44	–2.21	–16.74	61.69	–25.69	–8.59
Sex														
Male	2.40	3.26	–1.02	–2.04	–22.61	49.63	–10.63	–18.99	1.78	0.18	–1.04	3.78	–7.38	2.68
Month														
January	–0.30	–1.60	–1.43	–1.34	27.78	–15.43	–5.34	–2.34	–3.35	–0.44	–3.75	42.06	–16.66	–17.86
February	2.36	–0.29	–0.82	1.96	15.83	–34.58	–3.85	19.38	0.27	–0.45	–0.15	14.47	–13.59	–0.54
March	1.28	0.33	–0.30	2.64	–42.50	34.04	–5.88	10.41	–2.56	–1.96	4.78	–28.12	–16.65	44.50
April	3.53	1.08	–1.42	3.96	–13.38	–14.46	–7.68	28.39	2.41	–0.11	5.04	–16.04	–14.99	23.69
May	2.36	2.36	–0.41	6.00	2.09	–40.57	–4.75	32.92	–0.14	–0.16	5.11	–12.94	–16.69	24.83
June	3.16	0.14	–0.89	3.50	–13.64	–10.26	–5.48	23.49	1.31	0.17	12.74	–58.21	–17.13	61.11
July	10.68	2.91	–1.10	6.49	–21.10	–0.21	–8.28	10.61	0.78	–0.99	1.36	12.20	–8.08	–5.28
August	5.12	–0.10	–0.60	1.02	–7.21	–6.00	–7.60	15.36	–4.25	–0.17	–0.27	45.86	–13.86	–27.30
September	1.19	1.65	0.24	3.47	–20.52	1.06	–4.44	17.36	–3.20	–1.35	–1.21	53.01	–22.83	–24.42
October	0.67	0.44	–0.61	1.99	16.52	–7.91	–7.46	–3.64	2.55	–1.39	–2.73	27.57	–12.20	–13.80
November	4.12	–0.63	–0.34	1.78	15.82	–24.39	–0.62	4.25	–0.61	–0.20	1.21	–14.53	–6.24	20.37
Day of week														
Sunday	–0.48	–3.09	–2.10	–0.54	50.90	–40.69	–12.02	8.03	–6.11	–0.95	1.38	–1.75	–23.43	30.87
Monday	–8.48	12.15	–3.59	–8.98	–200.82	234.13	–17.30	–7.11	–2.22	–2.26	5.25	3.52	–27.37	23.08
Tuesday	–8.18	11.57	–3.26	–6.96	–223.99	254.35	–19.95	–3.57	–8.32	–0.51	0.48	–17.42	–16.67	42.43
Wednesday	–11.29	15.46	–3.89	–7.59	–239.79	257.03	–20.48	10.54	–5.51	0.49	4.45	–52.39	–11.03	63.99
Thursday	–8.82	12.22	–3.79	–7.57	–238.03	268.90	–16.65	–6.27	–12.26	0.81	–0.97	–11.81	–14.50	38.73
Friday	–5.59	12.29	–2.93	–4.75	–228.55	235.15	–21.02	15.41	–12.21	0.29	1.42	–15.73	–6.47	32.71
Constant	95.33	–4.48	3.10	64.60	1,023.01	41.81	82.25	134.38	55.96	10.77	33.67	1,103.21	118.11	118.27
r–squared	0.035	0.092	0.039	0.042	0.122	0.141	0.034	Derived	0.060	0.075	0.076	0.091	0.067	Derived

Note: Derived indicates the model was derived based on constraint equations, not estimated.

TABLE A2 Model for Change in Time Use Between 1990 and 1995: Workers

	Travel to								Time at							
Independent variables	**Home** Coefficient	*t*	**Work** Coefficient	*t*	**Shop** Coefficient	*t*	**Other** Coefficient	*t*	**Home** Coefficient	*t*	**Work** Coefficient	*t*	**Shop** Coefficient	*t*	**Other** Coefficient
% change in																
Lane-miles	–1.08	–0.23	–12.34	–2.16	–18.96	–6.47	–16.05	–2.83	656.17	12.63	–566.04	–10.27	–76.65	–6.96	34.95
Population	25.37	4.40	–0.58	–0.08	19.88	5.65	21.28	3.12	–105.22	–1.69	34.92	0.53	32.40	2.45	–28.06
Income	5.64	0.89	6.22	0.83	–13.81	–3.59	–9.66	–1.30	653.33	9.57	–688.50	–9.50	–17.95	–1.24	64.72
Gas prices	–44.33	–14.99	–7.53	–2.14	–20.39	–11.29	–15.38	–4.39	504.62	15.76	–354.00	–10.41	–26.57	–3.91	–36.43
Sex																
Male	0.75	2.90	4.22	13.59	–0.96	–6.06	0.01	0.05	–31.83	–11.28	11.84	3.96	–0.84	–1.41	16.80
Lifecycle																
(Adults, youngest child age)																
1, NA	–5.10	–3.36	0.57	0.35	–0.20	–0.24	–4.24	–2.64	1.30	0.08	7.79	0.45	–6.45	–2.07	6.34
2+, NA	–0.39	–0.27	–0.69	–0.44	0.51	0.64	–2.35	–1.53	30.98	1.95	–46.06	–2.72	1.59	0.53	16.42
2+, 0–5	0.32	0.22	–4.61	–2.94	0.01	0.01	1.33	0.85	22.09	1.37	–38.10	–2.23	0.43	0.14	18.53
1, 6–15	–5.55	–3.25	0.90	0.48	0.10	0.10	–1.61	–0.87	12.70	0.69	6.07	0.31	–1.69	–0.47	–10.92
2+, 6–15	–1.03	–0.70	–0.88	–0.56	0.38	0.48	2.10	1.35	4.39	0.27	–33.21	–1.95	1.40	0.47	26.86
2+, 16–21	–3.34	–2.18	0.88	0.54	0.90	1.08	–0.78	–0.48	–14.16	–0.85	–28.85	–1.64	3.29	1.05	42.06
2+, retired, NA	–3.22	–2.03	–3.87	–2.27	1.52	1.75	–0.90	–0.53	4.34	0.25	–31.25	–1.72	3.53	1.08	29.86
Month																
January	–0.28	–0.41	4.11	5.05	–2.10	–5.05	0.73	0.91	–24.83	–3.36	32.55	4.15	–5.70	–3.64	–4.48
February	–2.05	–3.34	2.20	3.01	–1.18	–3.14	–0.84	–1.16	–32.95	–4.96	65.12	9.23	–8.43	–5.98	–21.89
March	–0.93	–1.62	2.61	3.83	–1.41	–4.03	–2.43	–3.59	18.94	3.05	5.35	0.81	–5.60	–4.26	–16.53
April	–3.96	–6.43	0.88	1.20	–0.65	–1.72	–2.96	–4.06	–9.67	–1.45	49.56	7.00	–4.44	–3.14	–28.76
May	–2.33	–3.96	–1.46	–2.08	–1.06	–2.95	–5.84	–8.39	–11.04	–1.73	58.26	8.61	–4.03	–2.99	–32.50
June	–2.09	–3.34	1.64	2.21	–0.39	–1.03	–1.98	–2.67	–17.08	–2.52	42.83	5.96	–2.70	–1.88	–20.24
July	–9.54	–14.57	–2.34	–3.00	0.22	0.56	–3.95	–5.10	4.67	0.66	1.69	0.23	0.55	0.37	8.70
August	–2.55	–3.65	1.65	1.98	0.19	0.44	0.86	1.03	–16.40	–2.17	27.15	3.38	0.73	0.46	–11.62
September	0.51	0.78	–0.02	–0.03	–2.13	–5.28	–2.40	–3.08	6.33	0.89	15.29	2.02	–6.87	–4.54	–10.71
October	0.66	1.05	2.57	3.44	–1.05	–2.75	–0.91	–1.23	–44.24	–6.52	44.26	6.14	–2.96	–2.06	1.68
November	–3.42	–5.51	2.57	3.47	–0.39	–1.03	–1.98	–2.69	–35.80	–5.32	56.17	7.87	–4.68	–3.28	–12.46

continues

TABLE A2 Model for Change in Time Use Between 1990 and 1995: Workers (*continued*)

Independent variables	Travel to: Home Coefficient	*t*	Travel to: Work Coefficient	*t*	Travel to: Shop Coefficient	*t*	Travel to: Other Coefficient	*t*	Time at: Home Coefficient	*t*	Time at: Work Coefficient	*t*	Time at: Shop Coefficient	*t*	Time at: Other Coefficient
Day of week															
Sunday	–2.86	–5.55	–1.24	–2.02	–1.64	–5.23	1.74	2.86	10.65	1.91	–21.55	–3.65	–3.60	–3.05	18.49
Monday	8.19	17.68	3.31	6.00	–2.46	–8.70	0.63	1.16	39.87	7.95	25.56	4.80	–10.93	–10.28	–64.17
Tuesday	8.39	18.19	5.85	10.65	–2.54	–9.03	–1.34	–2.46	35.40	7.09	33.43	6.31	–7.48	–7.07	–71.70
Wednesday	11.09	23.79	1.91	3.45	–2.33	–8.19	–0.30	–0.54	46.79	9.27	33.28	6.21	–7.95	–7.43	–82.50
Thursday	15.40	2.19	5.54	0.66	–3.18	–0.74	–3.41	–0.41	–217.37	–2.86	97.46	1.21	–13.11	–0.81	118.67
Friday	7.46	15.96	3.64	6.53	–2.12	–7.42	–0.99	–1.79	36.42	7.20	32.67	6.08	–1.49	–1.39	–75.59
Population density															
250–499	–3.30	–6.69	–2.04	–3.46	–0.51	–1.69	11.84	20.23	–48.88	–9.14	–7.49	–1.32	–4.16	–3.67	54.55
750–999	1.33	2.34	1.17	1.72	–1.92	–5.50	10.86	16.06	–57.47	–9.29	–7.14	–1.09	–8.39	–6.40	61.55
1,000–1,999	–3.77	–7.01	–2.59	–4.05	–0.60	–1.83	8.82	13.87	–61.71	–10.61	3.11	0.50	0.94	0.77	55.79
3,000–3,999	–6.28	–12.41	–2.32	–3.85	–0.80	–2.60	12.37	20.64	–32.56	–5.94	–18.75	–3.22	1.88	1.62	46.47
5,000–7,499	–4.45	–9.28	–5.43	–9.50	–1.04	–3.54	9.21	16.22	–67.05	–12.91	18.07	3.28	3.40	3.09	47.29
Constant	–13.95	–8.50	–1.62	–0.91	3.42	3.77	–8.09	–4.59	92.41	5.20	–24.44	–1.30	4.96	1.45	–52.69
r–squared	0.106		0.089		0.0385		0.068			0.0813		0.1022		0.513	Derived

Note: Derived indicates the model was derived based on constraint equations, not estimated.

TABLE A3 Model for Change in Time Use Between 1990 and 1995: Nonworkers

Independent variables	Travel to: Home Coefficient	*t*	Travel to: Shop Coefficient	*t*	Travel to: Other Coefficient	*t*	Time at: Home Coefficient	*t*	Time at: Shop Coefficient	*t*	Time at: Other Coefficient
% change in											
Lane-miles	52.79	6.01	23.55	3.65	–60.64	–5.09	260.35	3.14	119.27	4.46	–395.32
Population	2.11	0.20	–17.46	–2.28	–5.13	–0.36	196.09	1.99	–64.83	–2.04	–110.78
Income	65.41	5.98	6.52	0.81	–20.29	–1.37	–60.35	–0.59	118.18	3.55	–109.47
Gas prices	6.06	1.25	8.06	2.26	–29.14	–4.43	41.98	0.92	55.71	3.77	–82.66
Sex											
Male	0.94	1.76	–2.33	–5.91	3.81	5.25	–28.94	–5.72	–6.92	–4.23	33.42
Lifecycle											
(Adults, youngest child age)											
1, NA	4.40	1.52	–3.66	–2.47	–12.35	–3.15	87.62	4.60	–9.35	–1.52	–66.65
2+, NA	–2.41	–0.88	–5.67	–4.28	–16.36	–4.40	80.81	4.74	–9.34	–1.70	–47.03
2+, 0–5	6.33	2.30	–5.35	–4.03	–12.54	–3.36	82.08	4.81	–13.17	–2.39	–57.34
1, 6–15	6.28	1.97	–8.07	–4.53	–16.24	–3.75	130.03	5.69	–16.19	–2.19	–95.81
2+, 6–15	0.95	0.34	–7.15	–5.27	–14.15	–3.77	117.73	6.76	–20.66	–3.67	–76.72
2+, 16–21	–3.66	–1.28	–5.12	–3.51	–13.64	–3.52	111.32	5.96	–9.37	–1.55	–79.54
2+, retired,NA	0.89	0.32	–2.97	–2.22	–9.14	–2.46	68.39	3.98	–5.38	–0.97	–51.80
Month											
January	2.09	1.60	–3.10	–3.23	5.48	3.09	–38.16	–3.10	–4.55	–1.14	38.24
February	0.51	0.44	–2.45	–2.86	5.16	3.26	–31.89	–2.89	–1.18	–0.33	29.85
March	3.11	2.81	–1.06	–1.31	–0.70	–0.47	14.21	1.36	–2.00	–0.59	–13.55
April	–1.66	–1.43	–2.97	–3.50	1.52	0.97	–11.80	–1.08	–4.74	–1.34	19.64
May	0.57	0.52	–2.31	–2.87	–0.93	–0.63	–0.61	–0.06	1.25	0.38	2.04
June	0.71	0.61	–3.25	–3.76	–6.34	–3.97	39.81	3.58	–0.70	–0.19	–30.24
July	0.71	0.58	–2.50	–2.77	3.65	2.19	–26.39	–2.28	–11.87	–3.17	36.39
August	5.27	4.00	–1.54	–1.59	5.35	2.99	–46.88	–3.77	0.14	0.04	37.67
September	3.94	3.16	–1.64	–1.80	5.60	3.31	–81.20	–6.90	7.61	2.00	65.70
October	0.46	0.39	–0.50	–0.57	8.38	5.22	–13.61	–1.22	–4.23	–1.17	9.49
November	1.17	0.99	–1.73	–1.99	1.91	1.19	–6.88	–0.62	–1.46	–0.41	6.99

continues

TABLE A3 Model for Change in Time Use Between 1990 and 1995: Nonworkers (*continued*)

	Travel to						Time at				
Independent	**Home**		**Shop**		**Other**		**Home**		**Shop**		**Other**
variables	Coefficient	*t*	Coefficient	*t*	Coefficient	*t*	Coefficient	*t*	Coefficient	*t*	Coefficient
Day of week											
Sunday	4.39	4.66	–3.20	–4.62	–2.54	–1.99	12.43	1.40	3.84	1.34	–14.91
Monday	–0.49	–0.53	–1.14	–1.68	–5.45	–4.37	19.28	2.22	11.36	4.05	–23.57
Tuesday	7.63	8.38	–2.41	–3.60	1.17	0.95	28.65	3.33	1.48	0.53	–36.53
Wednesday	5.93	6.48	–2.44	–3.64	–3.77	–3.04	63.25	7.33	3.24	1.16	–66.20
Thursday	11.82	12.73	–2.92	–4.27	2.27	1.80	20.50	2.34	1.90	0.67	–33.58
Friday	12.88	13.89	–0.59	–0.87	0.70	0.56	15.87	1.81	0.86	0.30	–29.72
Population density											
250–499	7.85	8.70	1.62	2.45	7.41	6.05	–78.93	–9.26	35.54	12.91	26.50
750–999	–5.04	–4.79	–2.10	–2.72	–7.81	–5.48	–25.28	–2.55	34.69	10.83	5.53
1,000–1,999	1.39	1.43	–1.91	–2.69	–0.71	–0.54	–50.09	–5.48	39.74	13.46	11.58
3,000–3,999	–5.20	–5.75	0.92	1.39	5.09	4.15	–60.48	–7.08	32.82	11.90	26.85
5,000–7,499	0.66	0.79	–0.23	–0.38	–0.76	–0.66	–40.22	–5.06	28.89	11.24	11.66
7,500–9,999	4.15	4.40	–2.06	–2.96	5.17	4.04	–92.32	–10.36	34.55	12.00	50.50
Constant	–23.79	–7.81	14.23	8.62	13.13	3.18	–56.75	–2.68	–22.63	–3.31	75.81
r–squared	0.161		0.101		0.123		0.138		0.148		Derived

Note: Derived indicates the model was derived based on constraint equations, not estimated.

Using Nonparametric Tests To Evaluate Traffic Forecasting Performance

STEPHEN D. CLARK
Morley Town Council

SUSAN M. GRANT-MULLER
HAIBO CHEN
University of Leeds

ABSTRACT

This paper proposes the use of a number of nonparametric comparison methods for evaluating traffic flow forecasting techniques. The advantage to these methods is that they are free of any distributional assumptions and can be legitimately used on small datasets. To demonstrate the applicability of these tests, a number of models for the forecasting of traffic flows are developed. The one-step-ahead forecasts produced are then assessed using nonparametric methods. Consideration is given as to whether a method is universally good or good at reproducing a particular aspect of the original series. That choice will be dictated, to a degree, by the user's purpose for assessing traffic flow.

INTRODUCTION

Many models attempt to predict the behavior of a system. These models may be physical, mathematical, statistical, or simulation representations of the system. Within the transportation field, physical models can be scale models of the geographical area of interest, mathematical models can be queuing models, statistical models can be platoon dispersion models, and simulation models can be meso- or microsimulation models. These models may operate on cross-section data, which represent

Stephen D. Clark, Morley Town Council, Town Hall, Morley, Leeds, United Kingdom LS27 9DY. Email: stephendclark@ntlworld.co.uk.

a snapshot of a system at a particular point in time, or on time series data, which represent the "movement" of a system through time.

If the appropriate model for the system is known, a dataset is used to calibrate the parameters in the model and then the model is applied. If the model is not known, then procedures are necessary to select from a range of models. As part of this selection process, a commonly used procedure is to split the dataset into two portions for training and testing purposes. The training portion, which is usually the larger, is used to calibrate the parameters in the model, and then the testing portion is used to assess the accuracy of the calibrated model in reproducing observed behavior. If the performance of the model with the testing dataset is deemed adequate, then the two datasets are pooled and the model recalibrated. Sometimes there is either insufficient data of acceptable quality to enable this partition to take place or no obvious way of dividing the datasets. In such cases, a with-replacement sampling approach may be adopted to construct the two datasets. To accurately assess without bias a model's goodness-of-fit, the modeler must first determine the values of the calibration parameters and then assess the performance of that model.

This paper, while incorporating forecasting models, is not concerned with a detailed study of the relative merits of these models, but with methods of assessing their ability to produce useable forecasts. In particular, this paper does not concern itself with the accepted iterative procedures of model identification, model estimation, and model diagnosis. It is assumed that these stages have been successfully completed and that the practitioner is now interested in how the model performs.

ASSESSING GOODNESS-OF-FIT

In the modeling processes and for models used for forecasting discussed in this paper, there are two types of discrepancy between the observed and modeled values. Within-sample discrepancies, which are typically generated during the model-fitting stages, are termed residuals in this paper. The outside-sample discrepancies are those that arise from applying the model to "unseen" data and are termed forecast errors in this paper. It is this latter form of discrepancy that is of most interest to practitioners and is the one considered in this paper.

A primary requirement is that a goodness-of-fit test be dependable. It should also be accurate and consistent in application. The fewer the number of assumptions that accompany the test the better the assessment of goodness-of-fit. Such assumptions may include the distribution of observations or existence of a sufficiently large sample size. Sometimes the test may be robust to departures from these assumptions, but a doubt may still exist over any measure that compromises any of these assumptions.

An additional modeler task is to communicate information to those who have the authority or influence to use it. Unfortunately, such individuals' expertise often differs from that of the modeler. This places a requirement that the metrics used in assessing goodness-of-fit are readily comprehensible and acceptable to specialists in other fields. Much of the motivation for this paper comes from earlier work by Dadkhah and Zahedi (1986), in which they propose various nonparametric tests to identify models that can predict turning points and directions of change in a time series. They also list a wide range of model evaluation tests in their appendix. In practice, however, not many of these evaluation tests outlined are actually used, because they would prove daunting when communicating results to a nonstatistically aware audience

The commonly used measures are those that involve an averaging of a simple function of the difference between the observed and forecast behavior. One such term is the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(f_t - v_t)^2} \tag{1}$$

Where f_t is the forecast at time t,
v_t is the observation at time t, and
N is the number of observations

while another is the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|f_t - v_t|}{f_t} \tag{2}$$

Both of these statistics have the advantage of being easily comprehended by most practitioners. Some disadvantages of these measures follow.

1. There is no criterion for assessing whether one value of the statistic is acceptable or not. Usually a range of forecasts are produced using either different methods or different datasets and a subjective opinion made as to whether one result is good or not in the context of the other results.
2. The RMSE or MAPE are often used in the model calibration stage to estimate the parameters in a model. Thus, there is the possibility that any calibrated model may be biased in producing estimates that give good performance on that measure but poor performance on other, equally valid, measures of goodness-of-fit.
3. While some forecasting methodologies specify several distributional requirements on the residuals from estimated models, and these requirements can be tested (but see 6 below), it is not usually necessary to place distributional requirements on outside sample forecast errors.
4. These statistics group all the observations together, losing the individual point-to-point relationship that exists. This drawback is particularly serious for time series data where the time element is important but lost in the aggregation.
5. The measures are not especially robust to outliers in the data, in particular the RMSE will exaggerate the impact of any outliers in either the observed or forecast series.
6. If any standard statistical tests are applied to these data, certain assumptions on the distribution of the difference between the modeled and observed values, termed the residuals, are required. These assumption can be (but are seldom) tested, but even when assumptions are found to be valid, there is still a remaining doubt (Type II errors).

NONPARAMETRIC METHODS OF ASSESSMENT

Nonparametric methods provide an alternative approach to assessing goodness-of-fit and pose certain advantages over parametric or averaging approaches, namely:

1. they do not assume any underlying distribution for the data used in the test,
2. they are able to provide objective methods for assessing whether a result is acceptable,
3. they are applicable with small sample sizes,
4. they can be robust to outliers, and
5. they are more readily comprehensible to specialists in other disciplines.

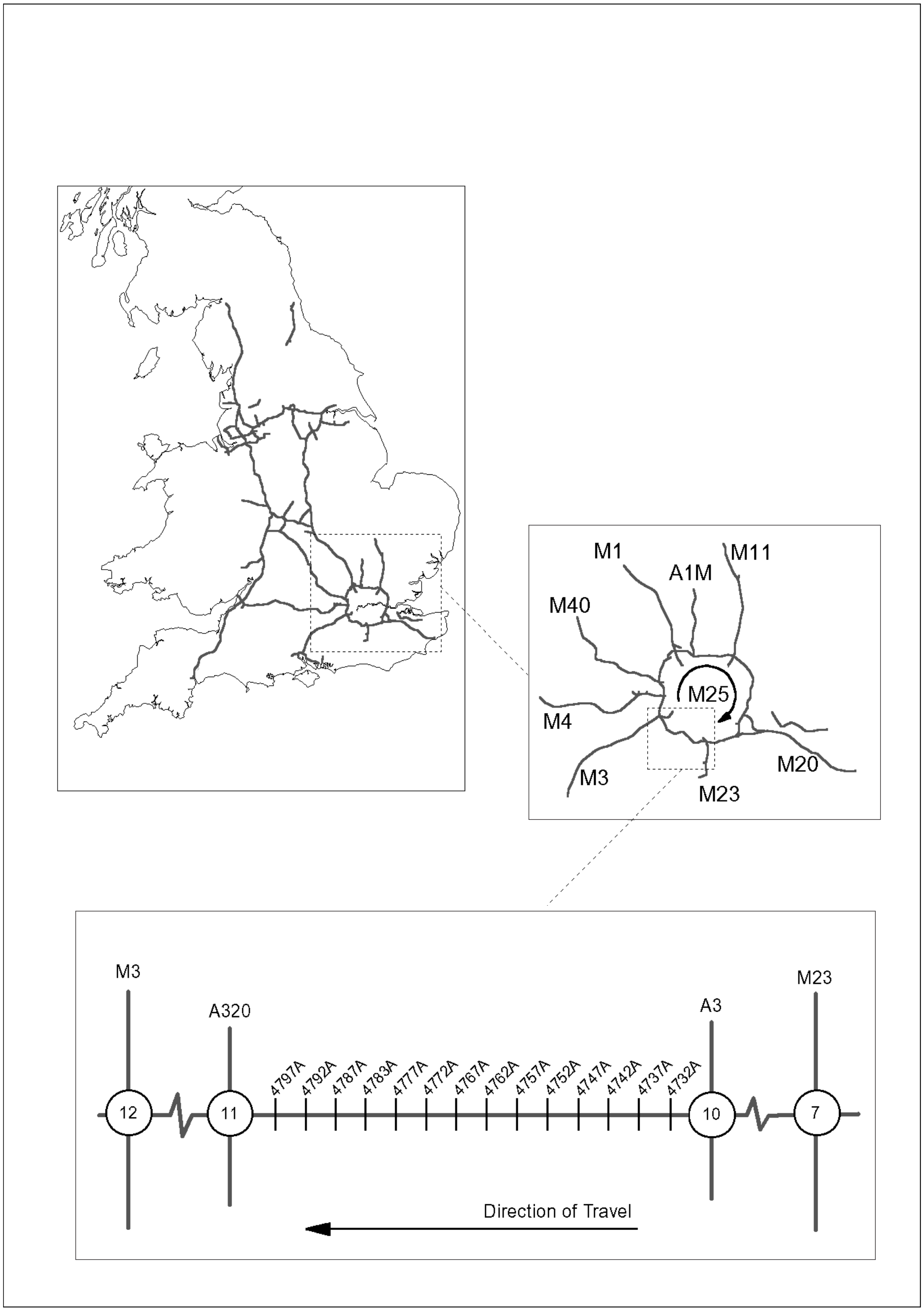
Three types of nonparametric tests are discussed in this paper. The first set are tests of the location of distributions based on signs, the second on the equality of shape of distributions, and the third on correspondence of distributions. These tests may be applied to the original and forecast data points and/or the original and forecast directions of change in a series.

DATA AND PRACTICAL CONTEXT

The English Highways Agency collects traffic information continuously at one-minute intervals on traffic flow (measured in vehicles), speed (km/hour), headway (seconds), and detector occupancy (percentage) on the M25 motorway (freeway). Detectors are typically located 500 meters apart and there is one in each traffic lane of the carriageway. One of the primary purposes for this infrastructure is to monitor traffic on the motorway with a view to activating a series of speed variable message signs as congestion builds (Maxwell and Beck 1996; Nuttall 1995). Currently, the Highways Agency uses the system in a reactive mode, that is, decisions on whether to activate the message signs are made on the basis of the most recent traffic situation. They are actively investigating whether an anticipatory mode may be more efficient, where traffic conditions are forecast for a short time horizon, typically less than one hour, and action taken to forestall anticipated congestion.

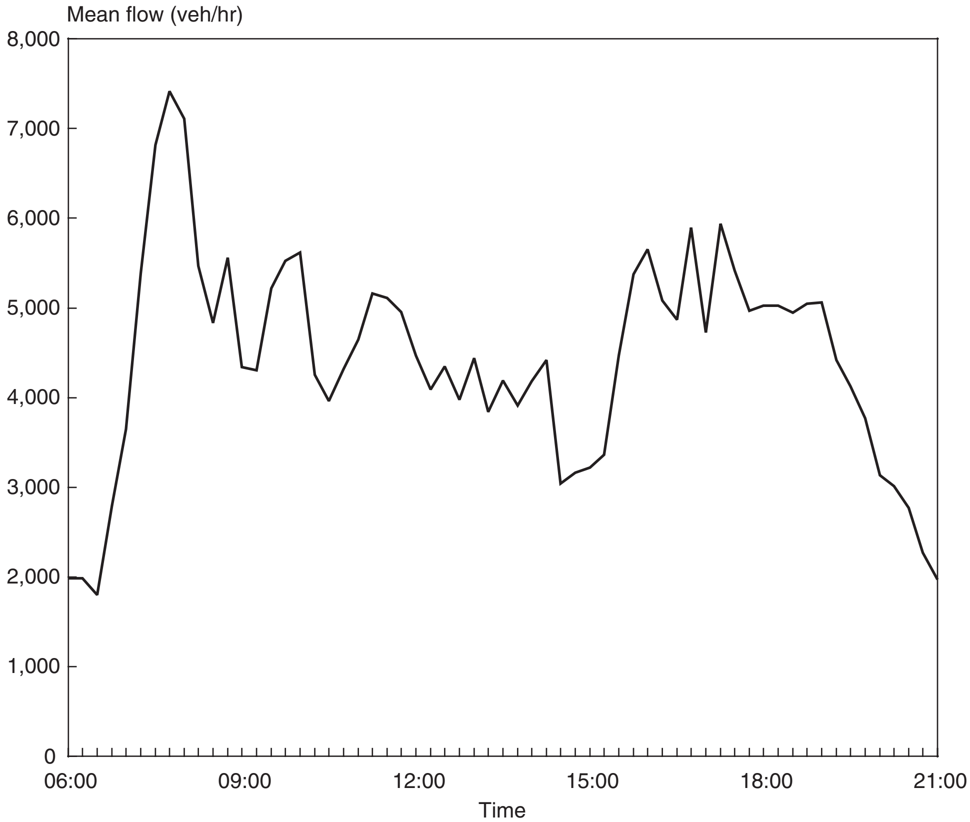
For the purposes of this study the 1-minute, 4-lane traffic flows have been aggregated into 15-minute carriageway flows (expressed as equivalent flows in vehicles per hour) starting at 6 a.m. and continuing until 9 p.m. The data were aggregated to overcome (or diminish) the effect of the few outliers or missing observations present in the one-minute lane measurements. Four sites were chosen

FIGURE 1 Location of Test Sites on the London Orbital Motorway (M25), England



for data sources, three are four-lane sections and between junctions, labeled as 4757A, 4762A, 4767A and the remaining site, 4802B, is a three-lane carriageway within a motorway junction site. Figure 1 shows the location of the three between-junction sites. Data were collected for all 4 sites for between 15 and 25 days in each of the months of August, September, and October 1997. This provided 184 days of traffic flows spread over 4 sites and 3 months. Figure 2 gives a typical flow profile for a day at one of the sites.

FIGURE 2 Typical Daily Flow Profile at Site 4762A



SUMMARY OF FORECASTING METHODS USED

Many studies have attempted to forecast traffic flows using a variety of techniques. Some have used computerized models of the network that represent the actual movement of traffic. On a simple level, the TRANSYT program (Vincent et al. 1980) contains a technique for predicting future downstream arrivals at signalized links in a traffic network. More complicated approaches involve the computerized simulation of individual vehicles moving through a traffic network (Morin et al. 1996; Algers et al. 1997).

The second group of work has attempted to model traffic flow as a time series of observations. Many well-recognized statistical models can be fitted to historical time series data and then used to produce short-term (usually one-step or two-steps-ahead) forecasts. Moorthy and Ratcliffe (1988) produced time series forecasts for an area of West Sussex, and Smith and Demetsky (1997) demonstrated application of a time series model (among others) to forecast traffic volumes on a freeway in Northern Virginia.

A third, more recent, direction is the use of artificial neural networks that can be trained to recognize complex (nonlinear) patterns in historic traffic flows and identify them in unseen data to produce "typical" follow-on conditions. This research has produced a large number of publications since the late 1980s, and Dougherty (1996) contains a review and extensive bibliography of such applications.[1]

In this paper the four forecasting methods used were selected in an earlier study (Clark et al. 1999) to encompass a range of time-series forecasting techniques.

Naive Model

The simplest forecasting technique is to assume that the currently observed level of flow will persist into the next time period:

$$f_{t+1} = v_t \tag{3}$$

Where f_t is the forecast flow at time t;
v_t is the observed flow at time t;

[1] For a detailed description of the technical aspects of artificial neural networks, the reader is directed to Bishop (1995).

This technique forms a benchmark that any competent forecasting methodology needs to exceed. No assumptions can be made about the distribution of the residuals or forecast errors from this model.

Long-Term Memory Model

A refinement is to forecast the future level of flow as an average of current and previous levels of flow. This method uses the arithmetic mean of four previous observations.

$$f_{t+1} = 0.25v_t + 0.25v_{t-1} + 0.25v_{t-2} + 0.25v_{t-3} \quad (4)$$

Where f_t is the forecast flow at time t; and

v_t is the observed flow at time t;

The structure of this model arises from the data format used in this paper, which comprises 15-minute observation periods, that is, a time lag of 4 provides 1 hour of data. Once again, no assumptions can be made about the distribution of the residuals or forecast errors from this model.

ARIMA Model

The next level is to assume a static structure for the period-to-period relationship in the data, but allow the strength of this relationship to vary over time. This may involve fitting a Box-Jenkins ARIMA-type model (Box and Jenkins 1976) to the series. Initial investigations indicate that in order to render the series stationary, a differenced logarithmic transformation is required.

$$(f_t - \mu_v) = \phi_{1,t-1}(v_{t-1} - \mu_v) + \varepsilon_t \quad (5)$$

Where f_t is the forecast flow at time t,

v_t is the observed flow at time t,

μ_v is the mean of the observed flow,

$\phi_{1,t}$ is a parameter to be estimated from data to time period t, and

ε_t is a random residual term.

This model is a general formulation of the previous two. Unlike the other models, the procedures used to estimate parameters in this model require certain normality assumptions for the residuals, but no assumptions are possible for forecast errors.

Nonlinear Model

Sometimes the assumption of an essentially linear relationship between two quantities, as in the previous three models, is not valid. In such cases, a nonlinear formulation of the model is required. The structure adopted here is to formulate a back-propagation neural network that relates previous levels of flow to future levels. Once again, it is not possible to explicitly derive a distribution for the residuals or forecast errors from this model.

TESTS

The application of nonparametric tests is well described in the statistical literature, and the reader is directed to these texts if further explanation is required.

Signs Test

One of the features of a series of errors from a well-behaved forecasting model is that it should contain a similar number of positive and negative observations. The assumption underlying this test is that the number of, say, positive errors is shown as a binomial distribution. The parameters of this distribution are the number of trials as (n–m), where n is the number of observations and m is the number of ties (i.e., the original and forecast values are the same) and the probability of success is half. The term success is commonly used when discussing the binomial distribution, but the term has no pejorative meaning here. Once an observed number of positive errors has been found, the two-tailed probability of obtaining this number of positive errors may be calculated. This probability may then be compared to some significance level to determine whether the assumption of an equal number of positive and negative errors is valid.

For a well-behaved forecasting methodology, one would hope to be able to accept the hypothesis that there are a similar number of positive and negative residuals. This ensures that the method does not tend to systematically over- or underpredict.

Wilcoxon Test on Location

When comparing the observed and forecast series, one of these two series should not be overrepresented when considering the magnitude of the values. To test for this, the two series are merged and the observations in the merged series given ranks. The ranks associated with observations from each of the series (original and forecast) are identified and summed. If the two series values are of similar magnitudes, then these two numbers should be similar, and tables are available to test for this. A modified Wilcoxon procedure may also be applied to establish whether the location of the differences between the observed and the forecast series is zero. Here the differences are ranked, and the sum of the ranks of positive differences should be similar to the sum of ranks of negative differences. The degree to which this is the case can be tested against tabulated values.

This test measures whether the location of two distributions are the same. In this case, the two distributions could be either the observed and forecast series or the differences between the observed and forecast series. In both cases, one would hope that the tests revealed that the location of the appropriate series was the same, or zero in the case of the modified Wilcoxon procedure.

Wilcoxon Test on Variance

Rather than test whether the location of two series are similar, this test measures whether the dispersion of two series are similar. Consider the case where one series occupied the lower and upper quartile of the merged series and the other, the middle two quartiles. Using conventional rankings, these two series would produce similar rank sum statistics and a conclusion that the location of the two series were similar would be made. It is clear, however, that in this extreme case the spread of observations is not the same. To test this, a different ranking method is deployed that spreads the lower ranks toward the ends of the series. The smallest value is given a rank of 1, the largest, 2, the second largest 3, the second lowest 4, the third lowest 5 and this pattern is repeated, moving into the center of the concatenated series. By adopting this ranking scheme, it is clear that in our extreme example the series at the extremes would have a significantly lower rank sum than the other series. This test should only be applied after determining that they have similar centrality locations.

Rank Correlation

This test enables a judgment to be made as to whether the same magnitude of observation is made at each time period. Ideally, the largest forecast is made at the same time the largest magnitude is seen in the original series and so on to the smallest magnitude of the two series. This statistic may be calculated on either the observed series or the differenced series. When applied to the differenced series, the test is focused on whether the magnitude of the changes in both observed and forecast series are seen at the same time.

In an ideal situation, the correlation would be +1. The "worst" case situation applies when there is an opposite relationship and the correlation would then be –1.

Direction of Change

Sometimes it is desirable to know whether a forecast series is generally moving in tandem with the original series. This is the equivalent of asking whether the successive differences in two series are the same. If the number of times that the direction of change for the forecast and observed series agree are counted, then this statistic should follow a binomial distribution. If the yardstick is to perform better than a random toss of a coin, then the probability of success is half. The probability of observing the number of agreements can then be calculated on this hypothesis. Before this test is applied, however, it is necessary to establish whether the occurrence of continuations or changes in direction are independent events through time (Dadkhah and Zehedi 1986). This may be tested for using a $2 \times 2\chi^2$ contingency test but, like tests on distributional assumptions, this outcome is subject to hypothesis errors and weakens the general utility of this test.

A good forecasting method should pass the test for independence and the number of times the direction of change agrees should be greater than what would be expected through chance.

EVALUATION OF THE FORECASTING METHOD

In this section, the three strands of data, forecasting method, and goodness-of-fit measure are brought together. For each forecasting measure, the performance over all 184 days is summarized in table 1. For the root mean square error, mean absolute percentage error, and rank correlation statistics, the mean and standard deviation (given in parentheses) of the statistic are presented. For the Wilcoxon tests, the number of times a significant difference is found at the 10% and 5% levels are presented. For the direction of change measure, four counts are provided and classified as to whether or not the observed changes are independent events $\left(p\left(\chi^2\right) > 10\% \text{ or } p\left(\chi^2\right) > 5\%\right)$ and if prediction of direction change is better than an even chance (p(Bin)<5% or p(Bin)<10%). For this last measure, the best possible performance for an individual day is $p\left(\chi^2\right) > 10\%$ and p(Bin)<5%.

The Wilcoxon location test on the differences between successive observations failed to produce any days with significant outcomes and has not been reported in table 1.

An assessment based on the root mean square and absolute percentage error indicators suggests that the nonlinear method performs best, followed by the naive and ARIMA models with the long-term memory model performing worst. This ordering is also preserved to some extent for the rank correlation statistic on the original and the first differenced series, although the rank correlation between the observed and forecast first differences has proved to be low across all forecasting methods. There is evidence from the test on the number of positive residuals and both the Wilcoxon tests that the distribution of one-step ahead forecasts for the nonlinear model is not in accord with those of the observed series. The naive and ARIMA models perform well at maintaining a similar distribution for the original and the forecast series. In the case of the naive method, this is not surprising since the forecast is the original series, only shifted by one time period. The test that emphasizes the ability of a forecast to predict correctly the direction of change in the original series shows the long-term memory model performs well.

TABLE 1 Statistical Performance of the Four Forecasting Methods

Model method		Naive model	Long-term memory model	ARIMA model	Nonlinear model
RMSE		597.8	904.5	637.7	471.5
		(154.7)	(158.8)	(184.7)	(107.9)
MAPE		9.16	13.92	9.41	7.59
		(1.96)	(2.33)	(2.08)	(2.11)
Rank correlation of original		0.768	0.590	0.757	0.809
		(0.087)	(0.093)	(0.090)	(0.071)
Significant number	5%	4	11	0	56
of positive errors	10%	11	22	4	70
Wilcoxon location	5%	0	0	0	1
on original	10%	0	0	0	7
Wilcoxon variance	5%	0	2	0	14
on original[1]	10%	0	4	0	34
Rank correlation of difference		0.064	0.113	0.082	0.098
		(0.138)	(0.130)	(0.121)	(0.119)
p(Bin)<5% (much	$p\left(\chi^2\right)$>10%	19	54	21	17
better than chance)	$p\left(\chi^2\right)$>5%	19	51	19	15
p(Bin)<10% (better	$p\left(\chi^2\right)$>10%	28	76	30	30
than chance)	$p\left(\chi^2\right)$>5%	28	72	25	28

[1] The number of days on which this test is valid is 184 minus the number of days on which there was a significant difference in the locations of the original and forecast series, e.g., for nonlinear at the 10% level this is 184–7=177 days.

ADDITIONAL DIAGNOSTIC TESTS

As mentioned in the first section of this paper, the focus here is on the evaluation stage of the performance of a forecasting method. It is correct to say that this evaluation should only be conducted once the modeler is satisfied and can demonstrate that the model chosen is appropriate for the data. This should not preclude, however, some form of ongoing model suitability evaluation.

In the earlier iterative model building process, residuals from the modeling are commonly examined to ensure that they adhere to some distributional assumptions. Of particular concern when dealing with time series data is that the residuals should not be autocorrelated and should have a constant variance. These issues are commonly covered in textbooks on econometrics (Maddala 1992; Gujarati 1995). There may be value in checking forecast errors when forecasting techniques are applied to ensure the errors have not acquired any of these features.

In performing these checks, a number of nonparametric techniques are available. As an illustration of this issue, autocorrelation may exist in the model residuals or the forecast errors. To test for first-order autocorrelation, one approach would be to establish whether there were an unreasonable number of runs of positive or negative values in the forecast errors. If there were too few runs, this would indicate positive autocorrelation, while too many runs would indicate negative autocorrelation. A slightly more complex but explicit nonparametric test for serial correlation of higher orders is given in Hoel (1984). Similarly, nonparametric approaches may be adopted to test for nonconstant variance in the forecast errors.

Returning to the example models and data used in the earlier section of this paper, the application of a runs test on the forecast errors shows that the number of days on which significant first-order autocorrelation at the 95% level was detected was low for the naive (9 days), ARIMA (7 days), and nonlinear (13 days) models but extremely high for the long-term memory model (179 days). The very high number of such days for the long-term memory model does not necessarily invalidate it because its parameter values were not estimated using a method that relies on uncorrelated residuals, but the reasons behind this feature would need to be explored.

CONCLUSIONS

Nonparametric tests are rarely used to evaluate the goodness-of-fit for a forecasting model. Given that such tests require fewer assumptions than parametric tests and that they can be correctly used with small samples, this appears to be a serious oversight. Nonparametric tests also allow for tests on the performance of a forecasting methodology without regard to the performance of other methods.

There are a wide variety of forecasting methods and tasks. It is unreasonable to assume that a forecasting methodology that is good at performing one task will necessarily be the best for other tasks. A modeler needs to make a judgment as to what is required from a forecasting method. The task is then to select or devise a goodness-of-fit measure that emphasizes the desirable properties of the forecast. Once the forecasts are known, the modeler is then able to make an objective judgment as to which method is the most appropriate. The nonparametric tests discussed in this paper are able to measure and compare different aspects of the performance of a forecasting method.

For the example given in this paper, each of the forecasting methods has its strengths. The nonlinear and naive methods are good at predicting the original level of the series, via low RMSE, MAPE, and high-rank correlation statistics. This may be important if it is necessary to predict when the level of flow crosses some form of traffic threshold, initiating the need for outside intervention. The ARIMA method is good at reproducing the distributional aspects of the original series. The long-term memory model is good at predicting the direction of change in a series—an ability that is useful for predicting a turning movement in a series. In the context of transportation, this has particular value in forecasting the beginning or end of a period of traffic volume growth.

ACKNOWLEDGMENTS

The authors would like to thank the English Highways Authority for supplying the traffic data used in this study. They would also like to thank Mr. Stuart Beale of the Agency for his help and

advice during the conduct of the research, of which this paper forms a part. The views expressed in this paper represent those of the authors only and should not be taken to be those of the Highways Agency or the Department of Transport.

REFERENCES

Algers, S., E. Bernauer, M. Boero, L. Breheret, C. Di-Taranto, M. Dougherty, K. Fox, and J. Gabard. 1997. Review of Microsimulation Models, SMARTEST Project Deliverable D3. European Commission.

Bishop, C.M. 1995. *Neural Networks for Pattern Recognition*. Oxford, United Kingdom: Clarendon Press.

Box, G.E.P. and Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day.

Clark, S.D., H.C. Chen, and S.M. Grant-Muller. 1999. Artificial Neural Network and Statistical Modeling of Traffic Flows—The Best of Both Worlds. *World Transport Research, Proceedings of the 8th World Conference on Transport Research*, vol. 2. Edited by H. Meersman, E. Van de Voorde, and W. Winkelmans. Oxford, United Kingdom: Elsevier Science, Ltd.

Dadkhah, K.M. and F. Zahedi. 1986. A Nonparametric Approach to Model Evaluation. *Journal of the Operational Research Society* 37(7):696–704.

Dougherty, M.S. 1996. Investigation of Network Performance Prediction: Literature Review, Technical Note 394. Institute for Transport Studies, University of Leeds, Leeds, United Kingdom.

Hoel, P.G. 1984. *Introduction to Mathematical Statistics*, 5th ed. New York, NY: Wiley and Sons.

Gujarati, D.N. 1995. *Basic Econometrics*, 3rd ed. New York, NY: McGraw-Hill.

Maddala, G.S. 1992. *Introduction to Econometrics*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Maxwell, H.A. and I. Beck. 1996. Traffic Control on the English Motorway Network, *Proceedings of the Eighth International Conference on Road Traffic Monitoring and Control*, Conference Publication No. 422, Apr. 23–25 1996, 136–44.

Moorthy, C.K. and B.G. Ratcliffe. 1988. Short-Term Traffic Forecasting Using Time Series Methods. *Transportation Planning and Technology* 12:45–56.

Morin, J-M., B. Baradel, and J. Bomier. 1996. Online Short-Term Simulation and Forecast of Motorway Traffic Patterns: Field Results Obtained on ASF Network in France. *Proceedings of the Third World Congress on Intelligent Transport Systems*, Orlando, Florida.

Nuttall, I. 1995. Slow, Slow, Quick, Quick, Slow: Taking the "Stop-Start" Out of the London Orbital. *Traffic Technology International*, 1995/Winter, 46–50.

Smith, B.L. and M.J. Demetsky. 1997. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering* 123(4): 261–66.

Vincent, R.A., A.I. Mitchell, and D.I. Robertson. 1980. User Guide to TRANSYT Version 8, Laboratory Report 888. Transport Research Laboratory, Crowthorne, Berkshire, United Kingdom.

The Impacts of Bypasses on Small- and Medium-Sized Communities: An Econometric Analysis

SIVARAMAKRISHNAN SRINIVASAN
KARA MARIA KOCKELMAN
The University of Texas at Austin

ABSTRACT

A relief route is a segment of a highway that moves traffic around the central business district of a city. Planners perceive it as a means of enhancing mobility and often associate regional economic progress with construction of bypasses. Though a bypass often means safer, quieter, less-congested downtowns, the communities receiving a bypass generally worry about potential negative impacts to the local economy. Hence, to make well-informed decisions on constructing relief routes, impact studies are needed. This paper examines the economic impacts of highway relief routes on small- and medium-size communities in Texas. Per capita sales in four different industry sectors were chosen as the indicators of impact.

The models developed suggest that the bypassed cities suffered a loss in per capita sales in all four industrial sectors considered. The magnitude of the traffic volume diverted appeared to be the greatest determinant of the impact. The overall impacts of the bypass were the most negative for gasoline service stations and the least for service industries. The impacts were less negative for cities that had high per capita traffic volumes. In addition, city demographics, regional trends, and proximity to a large city were estimated to have important impacts on

Kara Maria Kockelman, Assistant Professor of Civil Engineering, The University of Texas at Austin, 6.9 E. Cockrell Jr. Hall, Austin, TX78712-1076. Email: kkockelm@mail.utexas.edu.

the local economy. The industrial sectors considered for analysis represent only a portion of the total economy of the city. Therefore, negative impacts to these sectors do not necessarily mean that the economy as a whole suffers.

INTRODUCTION

Highway relief routes, also known as bypasses, move traffic around central business districts of cities. Relief-route users may experience travel time and cost savings, as well as increased safety. For some, rerouting through traffic is perceived as an advantage as it makes the downtown quieter, safer, and a more pleasant place for shopping (Otto and Anderson 1995). These routes may also affect the local economy negatively in terms of employment and income and sales volume. Thus, the overall impacts of a relief route on the city it bypasses are not obvious and cannot be easily generalized.

Planners view relief routes from a wider regional and state level and see them as one way to enhance intercity travel. On the other side, communities receiving bypasses may be concerned about potential negative economic impacts. These concerns can be critical for small- and medium-size cities that greatly depend on highway traffic. The challenge is to improve statewide mobility without hampering the local economies. Identification and quantification of economic impacts can inform decisionmaking regarding the construction of relief routes.

LITERATURE REVIEW

Considerable research exists on the impacts of bypasses. Some of the earliest studies date back to the 1950s. These and later studies examined the impacts on sales, employment levels, income, land use, land values, and other economic indicators. A wide array of methodologies has been employed, from the simplest forms involving before-and-after studies to more complex, indepth case studies and econometric modeling. Data at different levels of aggregation have also been used. A critique of some of the latest efforts to understand the economic impacts of highway bypasses is presented here.

A Wisconsin study by Yeh et al. (1998) used case study and survey and control-area methodology involving a nonpaired comparison of bypasses and control cities. An analysis of sales and employment data, travel surveys, and focus group interviews indicated little adverse impacts on the overall economy and little retail flight. The communities perceived their bypasses to be generally beneficial. This study shows that most of the medium and large cities bypassed were "natural destinations" and growth was one of the reasons these bypasses were needed. These may be cities where urban planners knew the impacts would not be too negative and so requested a bypass; hence, the sample may be biased. Since all communities that get bypassed will not fall into this category, results of this study cannot be generalized.

A survey and control-area method was adopted to study impacts in Iowa and Minnesota (Otto and Anderson 1995). The impact of the bypass, determined by a "pull factor" (defined as the ratio of the per capita sales in the bypassed community to that in the control group), indicated there was no significant difference in total sales reported by bypassed and control communities. Some redistributional effects were observed when the sales were broken down into components. This analysis, however, did not compare the sales levels before the bypass opened. It is possible that the bypassed cities had higher sales before the bypass, when compared with the control cities. A survey of the local business community helped identify perceptions, and it found that a majority of the respondents favored the bypass. However, results from such opinion-based data could be subjective and biased.

A classic example of a recent application of the before-and-after method is a study undertaken in Yass, Australia (Parolin and Garner 1996). Businesses were surveyed the year before the bypass opened. When surveyed a year later, 43.8% of the retail businesses reported a decrease in gross annual sales, 14% of jobs were lost (mostly in casual and part-time employment), and traffic surveys showed a 50% decrease in highway-generated trade. While these consequences appear significant, the study could not completely isolate the impact of the bypass on the local economy; there were other factors (like the withdrawal of construction workers and the opening of a service station near the bypass) that could not be entirely quantified but clearly affected the results.

Research undertaken for the Kansas Department of Transportation (Burress 1996) resulted in the development of a family of regression equations to explain a multitude of variables, such as city and county sales and employment levels. The effort also resulted in the calibration of a gravity model to obtain estimates of through and local traffic. The impact of the relief route was predominantly captured with indicator (dummy) variables, but several other variables, unrelated to relief routes (e.g., city demographics and regional trends in the industrial sector) were not controlled. The results indicated there were short-term negative impacts on some traffic-related businesses, but all such businesses did not suffer and effects were transitory. The study discovered that the impact of background effects—like the recession of 1990–91—were more significant than the bypass-related effects. Unfortunately, however, most of the models developed in this study had few explanatory variables.

A study of the economic impacts of highway bypasses on small Texas communities was done at the University of Texas at Austin (Anderson et al. 1992). Several methods, including projected development, multiple regression analysis, and cluster analysis, were used. The study concluded that highway bypasses might reduce business in small cities in rural settings. The models indicated a 15% drop in gas station sales and a 10% to 15% drop in sales at eating and drinking establishments. The study used pooled cross-sectional and time-series data; however, it did not use sophisticated methodologies like random-effects models that could extract more meaningful information from the panel data.

Work by Buffington and Burke (1991) used regression analysis on a panel dataset to examine the impacts of bypasses, loops, and radials on employment and wages. The impacts of bypass investments on manufacturing employment at the city level were positive. At the county level, the impacts were positive for both employment and wages. However, all the cities included in this study had some form of highway improvement (relief route, radial, or loop). There were no control cities in the dataset. This could lead to biased results, because all cities receiving highway improvements could have certain characteristics that are different from those cities not receiving such improvements. It is also possible that the general economy was improving, so that net effects appeared positive when in fact they were not as positive as they would have been without the bypass.

In 1996, the National Cooperative Highway Research Program consolidated the state of knowledge in the area of relief-route impacts (NCHRP 1996). Based on a literature review and responses to survey questionnaires sent to state departments of transportation, no conclusive evidence was found of a loss of sales, even in vulnerable locations, due to bypassing alone. This leaves open the possibility that relief routes can mean a loss in certain sales conditions.

The ability of multivariate regression analysis to isolate the marginal influence of a relief route from other factors that can possibly impact local economies makes it appealing for the current work. Studies that employed other methods failed to satisfactorily isolate the bypass impacts and those reviewed here that used regression analysis exhibit some deficiencies (e.g., excluding possibly relevant explanatory variables and sampling bias in the data used).

The work presented here takes a rigorous statistical and methodological approach to model the effects of bypasses on local economies. The impacts on four different industry types are examined using a panel dataset. Further, the impacts are modeled jointly, an approach that has not been adopted before. This research effort focuses on small- and medium-size cities in Texas. The results are intended to assist planners and engineers by providing reliable information on likely economic impacts in communities for which relief routes have been proposed.

While the strength of this study lies in the use of a sophisticated econometric methodology to model the impacts, its limitation lies in the fact that the impacts are measured primarily in terms of changes in the per capita sales in four different industrial sectors. Since the cities modeled here are of small and medium size, non-availability of data limited the extension of the methodology to model other sectors. For the same reasons, other factors characterizing the local economy, like income and employment levels, could not be modeled.

The paper is organized as follows. Data details are described, followed by a description of the modeling methodology. Analytical results are then presented and followed by a summary of conclusions and identification of areas of improvement to the current work.

DATA

We first created a list of cities in Texas with populations between 2,500 and 50,000, and then traffic maps were reviewed to classify the cities into those that are bypassed and those that are not. We further classified the bypassed cities based on the nature of their bypass(es), and only cities with a single bypass were considered for the study. This is the simplest form of bypassing, where the relief route splits from the old route at one side of the city and rejoins the same route on the other side. This exercise resulted in the identification of 23 bypassed cities[1] for analysis; 19 other, nonbypassed cities were chosen as "control" cities. For each of the 42 cities, 9 years of data (in years falling between 1954 to 1992) were collected. The sample, therefore, has a total of 378 data points.

We collected sales data for four industrial sectors from the U.S. Economic Census. These include total retail sales (Standard Industrial Classification (SIC) major groups 52–59), sales in gasoline service stations (SIC 554), sales at eating and drinking places (SIC 58), and service receipts (SIC major groups 70 through 89). Gasoline service stations and eating and drinking establishments are subcategories within the retail trade category. The data years are approximately five years apart. Both city- and state-level data were collected. All sales dollars were then adjusted for inflation and converted to current year 2000 dollars using the Consumer Price Index (University of Michigan 2000).

We obtained data on city demographics (e.g., overall population, unemployment, and elderly population), median household income, and average household size from the U.S. Census of Population. The data were used to derive an estimate of income per capita, as the ratio of median household income to the average household size, and incomes were converted to 2000 dollars. Census of Population data covered 1950, 1960, 1970, 1980, and 1990. These data were then linearly interpolated for the required data years.

The proximity of a city in this study to a large city was seen as an influential factor. A *large city* is defined here as the central city of a metropolitan statistical area (MSA) in 1990. The nearest large city was identified for each city sample, and distances were obtained from the *Texas Mileage Guide* (Texas Comptroller of Public Accounts 1999). The populations of these large cities were obtained from the U.S Census of Population and linearly interpolated for the required data years.

Using district traffic maps from the Texas Department of Transportation, we were able to infer the year when traffic first appeared on a city's relief route. These maps were used to determine the opening year of every relief route, and thus the number of years since opening for each data year. The traffic maps also provided data on the average annual daily traffic (AADT) at different count locations along the highways. Counts along the bypass were averaged to get an estimate of the traffic volume on the bypass. Counts on all state, U.S., and Interstate highways that pass through the city were averaged to get an estimate of the total traffic volume approaching the city.

Distances along the old and the new routes were obtained from county maps. The distances were measured from the point where the relief route branches off the old route to the point where it rejoins the old route. The county maps also provided information on the presence of frontage roads along the relief route.

ANALYSIS

Variable Specification

Per capita sales in four different industrial sectors were identified as indicators of the local economy. The industrial sectors are total retail (establishments that primarily sell merchandise for personal or household consumption), gasoline service stations (establishments that primarily sell gasoline and automotive lubricants), eating and drinking places (establishments that primarily sell prepared food and beverages), and service industries (establishments that provide a wide variety of services—e.g., lodging, repairs, health, amusement, legal, and

[1] These 23 cities were bypassed between 1965 and 1990.

technical—to individuals, businesses, government establishments, and other organizations).

We identified several variables to explain the four types of sales investigated. The impact of city demographics on a local economy is captured by introducing the fraction of population that is elderly (ELDERLY), the fraction of labor force that is unemployed (UNEMP RATE), and per capita income (INCOME PERCAP) as explanatory variables. Per capita income is expected to have a positive impact on per capita sales, while the unemployment rate is expected to have a negative impact. Elderly people may be more likely to shop locally, as opposed to driving out in search of more variety. Hence, an a priori expectation for this explanatory variable may be for a positive effect.

It also is hypothesized that the sales levels of small and medium cities are significantly influenced by the proximity of a large city, and the closer and more populated the large city is, the greater its influence. Thus, the ratio of the population of the nearest large city to its distance from the community under study is introduced as an explanatory variable (LARGECITY POP/DIST). More traffic moving through the city indicates a larger market for local goods and services. Since the models are developed at per capita level, the traffic volume approaching the city was normalized by the population of the city and this (TOT TRAFFIC PERCAP) was introduced as another explanatory variable.

Bypassed cities are identified by introducing an indicator variable (RELIEF ROUTE) that takes a value of one once the relief route is opened to traffic. The impact of a bypass depends on how much traffic and how far away traffic is diverted from a city's downtown. An estimate of the magnitude of the traffic diverted from the old route to the bypass is obtained as the ratio of traffic volume on the bypass to the total traffic volume approaching the city (TRAFFIC SPLIT). The greater the diversion, the greater the adverse impact on local sales is expected to be. The length of the bypass and the old route could be used as proxies for how far the diverted traffic is moved away from the old route. Variables, DISTOLD and DIST RATIO were introduced to capture this effect. The farther the diversion, the greater the expected negative impact.

The impacts of a relief route can be expected to change with time. The impacts may cease or there might be lagged effects on the community. The coefficient on the NUM YEARS variable captures this effect. A NUM YEARS SQ variable is also introduced to capture possible nonlinear time trends. The signs on these variables can be either positive or negative.

The per capita sales at the state level (STATE SALES PERCAP) for the specific industrial sector also is introduced as an explanatory variable to capture and control for more global trends in industry sales over time. The data year (YEAR) is used to capture other time-related trends.

The 1982 economic censuses provided sales data only for establishments with payrolls. Data for all other economic census years were available for all establishments. To characterize this data issue, an indicator variable (YEAR 1982) was introduced for observations in 1982 in all but the retail sales model (this problem was not observed for retail industry data). Sample means of the variables described are presented in table 1.

Model Specification

This section describes the econometric model structure and estimation method. A regression model developed on panel data from *N* cross-sections and *T* time periods can be represented as follows:

$$Y_{it} = \alpha + X_{1,it}\beta_1 + X_{2,t}\beta_2 + u_{it}, \quad i = 1 \text{ to } N, \, t = 1 \text{ to } T \qquad (1)$$

where

Y_{it} is the dependent variable,

$X_{1,it}$ are variables that vary over both cross section and time,

$X_{2,t}$ are variables that are time specific (the $X_{2,t}$ are cross-section invariant),

α, β_1, and β_2 are the model parameters to be estimated.

The error terms u_{it} can be broken down into unobservable cross-section-specific (i.e., city-specific) effects μ_i and a remaining term v_{it}. This is the conventional "one-way error components model" (see Baltagi 1995). Alternate model formulations arise depending on the assumptions made regarding the cross-sectional error term. One such formulation is the fixed-effect model, where the cross-sectional error term is estimated as a single

TABLE 1 Sample Characteristics

Dependent variables	Mean	SD
Per capita sales ($ per person)		
Total Retail (SIC 52 to 59)	1.226E+04	4.641E+03
Gasoline Service Stations (SIC 554)	1.113E+03	5.885E+02
Eating and Drinking Places (SIC 58)	6.469E+02	3.694E+02
Service Industries (SIC 70 to 89)	1.698E+03	1.487E+03
Independent variables specific to all cities	**Mean**	**SD**
State-level per capita sales ($ per person)		
Total retail (SIC 52 to 59)	8.447E+03	1.202E+03
Gasoline Service Stations (SIC 554)	6.450E+02	8.894E+01
Eating and Drinking Places (SIC 58)	6.486E+02	2.317E+02
Service Industries (SIC 70 to 89)	2.695E+03	2.006E+03
ELDERLY (percent)	15.54	5.61
UNEMP RATE (percent)	5.36	2.61
INCOME PERCAP ($ per person)	8.252E+03	1.970E+03
LARGECITY POP/DIST (persons per mile)	6.743E+03	8.174E+03
TOT TRAFFIC PERCAP (AADT)	1.772	1.16
Independent variables specific to bypassed cities	**Mean**	**SD**
TRAFSPLIT (fraction)	0.472	0.142
DISTOLD (miles)	5.109	1.524
DISTRATIO (fraction)	0.972	0.141

constant for each city. Another is the random-effects model, where the cross-sectional error term is assumed to be randomly distributed with a variance of σ_μ^2. The random-effects formulation has several statistical and practical advantages over the fixed-effects formulation (Maddala 1987) and hence is more suitable for the current work. This is the adopted error structure.

The specification just described models each industrial sector independently. In reality, there could be several unobserved characteristics of the cities that are impacting all modeled sectors of the economy. Therefore, the error terms can be correlated across equations. Estimation of the four equations separately ignores this correlation; hence, the resulting parameter estimates would not be as efficient as they could be (i.e., the standard errors of the unbiased parameters would not be minimized). This facet can be addressed by estimating the four regression equations as a set of "seemingly unrelated regression" (SUR) equations (Baltagi 1995).

In the case of SUR equations, we consider a set of M equations:

$$Y_j = Z_j' \ \delta_j + u_j, \quad j = 1 \ to \ M \qquad (2)$$

where

Y_j is the dependent variable,

Z_j is the set of explanatory variables,

δ_j is the vector of parameters to be estimated for equation j (of the M equations estimated jointly).

The error terms u_j can again be broken down into unobservable cross-section-specific (i.e., city-specific) effects μ_j and a remaining term v_j. The error structure, therefore takes the form

$$u_j = Z_\mu' \ \mu_j + v_j, \quad Z_\mu = I_N \otimes 1_T \qquad (3)$$

where

I_N is an identity matrix of size N,

1_T is a T*1 matrix of ones.

Since the equations are permitted to be correlated in their error terms, the cross-sectional error terms are distributed with a mean of zero and a variance-covariance matrix, $\sum_\mu \otimes I_N$, and the remainder error terms are distributed with a mean of zero and variance-covariance matrix, $\sum_v \otimes I_{NT}$ (I_{NT} is an identity matrix of size NT). In essence, the one-way random-effects model structure is extended to incorporate correlations across equations. The variance-covariance matrix for the set of equations takes the following form (see Baltagi 1995, p. 104):

$$\Omega = \sum_{\mu}(I_N \otimes J_T) + \sum_{v}(I_N \otimes 1_T) \quad (4)$$

where

J_T is a $T*T$ matrix of ones.

Defining transformation matrices P and Q as

$$P = I_N \otimes \bar{J}_T, \quad \bar{J}_T = \frac{J_T}{T} \quad (5)$$
$$Q = I_{NT} - P$$

the covariance-matrix can be rewritten as

$$\Omega = \sum_1 \otimes P + \sum_v \otimes Q \quad (6)$$

where

$$\sum_1 = T\sum_\mu + \sum_v$$

The set of regression equations can be estimated using feasible generalized least squares (FGLS) methods (Baltagi 1995). This requires an estimate of the covariance matrix. A methodology to estimate the variance components from the ordinary least squares (OLS) residuals was developed by Avery (Baltagi 1995) and can be summarized as the following:

$$\hat{\sum}_1 = \hat{U}^T P \hat{U} / N \quad (7)$$
$$\hat{\sum}_v = \hat{U}^T Q \hat{U} / N(T-1)$$

where

$\hat{U} = [\hat{u}_1 \hat{u}_2 \ldots \hat{u}_M]$ is an $NT*M$ matrix of disturbances,

$\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_M$ are the OLS residuals for the M equations.

An alternate way of estimating the model is by using maximum likelihood estimation (MLE). Asymptotically, MLE methods are more efficient than FGLS methods, but they require strong error-distribution assumptions and thus may render less robust predictors. Furthermore, Avery's FGLS estimation is as asymptotically efficient as GLS estimation (Prucha 1984).

If correlations across equations do not actually exist, then the independent estimation of equations is efficient. Therefore, it is useful to test the hypothesis that all the covariances are zero. This can be accomplished by a statistical test (detailed by Griffiths et al. 1993, p. 570). If the correlation matrix across the set of M equations is $\sum = [\sigma_{ij}]$, the null hypothesis that all σ_{ij} are zero (for $i \neq j$) can be tested against the alternate hypothesis that at least one σ_{ij} is non-zero using the test statistic λ.

$$\lambda = NT\sum_{i=2}^{M}\sum_{j=1}^{i-1}\hat{r}_{ij}^2, \quad \hat{r}_{ij}^2 = \frac{\hat{\sigma}_{ij}^2}{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}, \quad \hat{\sigma}_{ij} = \frac{u_i^T u_j}{NT} \quad (8)$$

where

u_j is the vector of OLS residuals for equation j.

Under the null hypothesis, the test statistic, λ, is chi-square distributed with $M(M-1)/2$ degrees of freedom. The results of this hypothesis test are described below.

RESULTS

The estimation methods were coded in the matrix programming language GAUSS (Aptech 1995). Random-effects models were developed independently for per capita sales in each industrial sector, and SUR models were developed to model the four industrial sectors jointly. In each case, the initial specification uses all available explanatory variables; statistically insignificant variables (t statistic < 1.6) were removed in a stepwise manner to arrive at the final specification. However, since the relief-route indicator variable is of fundamental interest, it is left in, regardless of its level of statistical significance.

The statistical test for the presence of correlation in the error terms across equations was performed, and the test statistic was estimated to be 239.4. This is significantly greater than the critical chi-square value; hence, the null hypothesis that all error correlations are zero is strongly rejected. The random-effects model was then extended to incorporate correlations across equations, and the system was estimated as a set of SUR equations. The correctness of the one-way error components structure was not tested statistically in the case of SUR. However, the null hypothesis that the variance of the city-specific error term is zero was rejected in the case of random-effects models. This may be expected to hold even for the SUR case. The SUR

TABLE 2 One-Way SUR Model for Per Capita Retail Sales

	Initial specification		Final specification	
	Coefficient	t-stat	Coefficient	t-stat
CONSTANT	5.48E+05	8.15	5.34E+05	8.92
STATE SALES PERCAP	1.46E+00	5.18	1.52E+00	5.76
YEAR	–2.85E+02	–8.05	–2.78E+02	–8.80
ELDERLY	1.58E+02	2.85	1.09E+02	2.32
UNEMP RATE	1.74E+02	2.12	1.21E+02	1.74
INCOME PERCAP	6.94E–01	5.19	6.49E–01	5.13
LARGECITY POP/DIST	1.18E–01	3.15	1.14E–01	3.26
TOT TRAFFIC PERCAP	2.77E+03	12.10	2.80E+03	12.58
RELIEF ROUTE	5.04E+03	1.85	5.35E+03	4.67
NUM YEARS	4.23E+01	0.32		
NUM YEARS SQ	–3.94E+00	–0.75		
TRAFFIC SPLIT	–1.76E+04	–6.11	–1.72E+04	–7.34
DIST OLD	2.15E+02	0.92		
DIST RATIO	–7.03E+02	–0.27		
ACCESS CONTROL	2.55E+02	0.30		
$R^2_{adj.}$	0.59		0.59	
σ^2_μ	3.31E+06		3.64E+06	
σ^2_ν	5.58E+06		5.47E+06	

TABLE 3 One-Way SUR Model for Per Capita Sales in Gasoline Service Stations

	Initial specification		Final specification	
	Coefficient	t-stat	Coefficient	t-stat
CONSTANT	4.45E+04	4.56	3.81E+04	6.05
STATE SALES PERCAP	3.04E+00	5.16	3.01E+00	7.10
YEAR	–2.33E+01	–4.54	–1.99E+01	–6.04
YEAR 1982	–3.89E+02	–4.55	–3.92E+02	–4.76
ELDERLY	4.64E+00	0.55		
UNEMP RATE	1.11E+01	0.76		
INCOME PERCAP	2.09E–02	0.99		
LARGECITY POP/DIST	–1.46E–03	–0.27		
TOT TRAFFIC PERCAP	2.95E+02	8.04	2.97E+02	8.72
RELIEF ROUTE	–2.80E+01	–0.06	–8.83E+01	–0.46
NUM YEARS	–2.18E+01	–0.92		
NUM YEARS SQ	7.98E–01	0.86		
TRAFFIC SPLIT	–4.20E+02	–0.86	–7.57E+02	–1.91
DIST OLD	2.12E+01	0.55		
DIST RATIO	–1.58E+02	–0.36		
ACCESS CONTROL	–1.41E+02	–0.97		
$R^2_{adj.}$	0.30		0.31	
σ^2_μ	5.31E+04		5.82E+04	
σ^2_ν	1.75E+05		1.73E+05	

models are presented in tables 2 through 5. Correlations between the error terms are presented in table 6.

The city-specific error term accounts for 40% of the total variance in the model for per capita retail sales. This fraction is 25% for the sales model for gasoline service stations, 36% for sales in eating and drinking places, and 27% for sales in service industries. Based on the estimates of the covariance matrix, it can be inferred that models for per capita sales in service industries and eating and drinking places are correlated the most in their unobserved

TABLE 4 One-Way SUR Model for Per Capita Sales in Eating and Drinking Places

	Initial specification		Final specification	
	Coefficient	t-stat	Coefficient	t-stat
CONSTANT	1.83E+04	2.11	1.50E+04	1.96
STATE SALES PERCAP	7.13E–01	3.30	6.97E–01	3.28
YEAR	–9.55E+00	–2.12	–7.81E+00	–1.97
YEAR 1982	–1.20E+02	–3.31	–1.16E+02	–3.27
ELDERLY	4.74E+00	1.03		
UNEMP RATE	7.89E+00	1.19		
INCOME PERCAP	3.17E–02	2.85	2.59E–02	2.67
LARGECITY POP/DIST	1.30E–02	4.25	1.44E–02	5.38
TOT TRAFFIC PERCAP	1.69E+02	8.76	1.66E+02	8.86
RELIEF ROUTE	–1.96E+02	–0.84	1.77E+02	1.79
NUM YEARS	–2.41E+00	–0.20		
NUM YEARS SQ	–1.82E–01	–0.39		
TRAFFIC SPLIT	–6.40E+02	–2.56	–6.75E+02	–3.32
DIST OLD	2.83E+01	1.42		
DIST RATIO	2.59E+02	1.17		
ACCESS CONTROL	–2.01E+01	–0.27		
$R^2_{adj.}$	0.58		0.59	
σ^2_μ	2.04E+04		2.38E+04	
σ^2_ν	4.34E+04		4.27E+04	

TABLE 5 One-Way SUR Model for Per Capita Sales in Service Industries

	Initial specification		Final specification	
	Coefficient	t-stat	Coefficient	t-stat
CONSTANT	1.32E+05	4.71	9.36E+04	4.05
STATE SALES PERCAP	7.68E–01	9.38	6.51E–01	9.61
YEAR	–6.82E+01	–4.73	–4.85E+01	–4.09
YEAR 1982	–2.12E+01	–0.16	3.18E+01	0.24
ELDERLY	2.54E+01	1.56		
UNEMP RATE	–3.06E+01	–1.12		
INCOME PERCAP	1.56E–01	3.89	1.53E–01	4.09
LARGECITY POP/DIST	3.50E–02	3.32	3.14E–02	3.32
TOT TRAFFIC PERCAP	3.97E+02	5.66	3.97E+02	5.89
RELIEF ROUTE	–1.17E+03	–1.35	6.54E+02	1.77
NUM YEARS	2.53E+01	0.57		
NUM YEARS SQ	–3.06E+00	–1.75	–1.52E+00	–2.96
TRAFFIC SPLIT	–2.29E+03	–2.47	–1.51E+03	–1.96
DIST OLD	8.64E+01	1.18		
DIST RATIO	1.63E+03	1.97		
ACCESS CONTROL	1.71E+02	0.62		
$R^2_{adj.}$	0.66		0.65	
σ^2_μ	2.04E+05		2.29E+05	
σ^2_ν	8.23E+05		6.27E+05	

error terms. The models for sales in retail and eating and drinking places are almost equally correlated in their unobserved error. The other correlations are much less.

The models developed indicate that the draw in traffic from the old to the relief route has a significant negative impact on the sales in the different industrial sectors. Characteristics of the relief route (access control, ratio of distance along the old route to the relief route) and time trends (NUM YEARS and NUM YEARS SQ) are not statistically significant in most of the models.The NUM YEARS SQ

TABLE 6 Error Correlations Across Equations

INITIAL SPECIFICATION

ρ_μ	Retail	Gas	Eat/Drink	Service
Retail	1.00	0.28	0.40	0.58
Gas	0.28	1.00	0.58	0.34
Eat/Drink	0.40	0.58	1.00	0.60
Service	0.58	0.34	0.60	1.00

ρ_v	Retail	Gas	Eat/Drink	Service
Retail	1.00	0.31	0.40	0.30
Gas	0.31	1.00	0.09	0.11
Eat/Drink	0.40	0.09	1.00	0.32
Service	0.30	0.11	0.32	1.00

FINAL SPECIFICATION

ρ_μ	Retail	Gas	Eat/Drink	Service
Retail	1.00	0.27	0.43	0.62
Gas	0.27	1.00	0.59	0.38
Eat/Drink	0.43	0.59	1.00	0.67
Service	0.62	0.38	0.67	1.00

ρ_v	Retail	Gas	Eat/Drink	Service
Retail	1.00	0.31	0.39	0.29
Gas	0.31	1.00	0.08	0.09
Eat/Drink	0.39	0.08	1.00	0.31
Service	0.29	0.09	0.31	1.00

variable is, however, negative and statistically significant for the service sales model. This suggests that the longer a relief route has been in place, the lower the per capita sales in the service industries. The coefficient on the relief-route indicator variable, which captures effects not picked up by other relief-route variables, was positive and statistically significant in all models except sales in gasoline service stations, where it was statistically insignificant.

Based on the coefficients estimated on the indicator variable and the percentage split in traffic, it can be inferred that the overall impact of the bypass on each of the sectors examined is negative when the traffic split exceeds a critical value. This critical traffic split is 31% for retail sales, 26% for eating and drinking places, and 43% for service industries. The impact on sales in gasoline service stations is negative irrespective of the magnitude of split. In 1992, the average traffic split was 47%.

We also found that per capita traffic levels in the city are major determinants of the per capita sales levels. Many of the city demographic variables were also estimated to be statistically significant. The nearness to a large city seems to have a positive impact on the sales in the different industrial sectors considered, except the gasoline service stations sector. Sales in the different industrial sectors also seem to be positively influenced by regional trends, as indicated by the coefficient on the state-level sales variables.

Thus, the models developed suggest that the marginal impact of the traffic split due to the bypass on the per capita sales in the four industrial sectors examined is negative. The net impact of the relief route, however, depends on the magnitude of all the variables considered in the model. To get a sense of this magnitude and to compare the impacts across the sectors, the estimated percentage difference in the per capita sales in the four industrial sectors before and two years after the opening of the relief route was calculated.

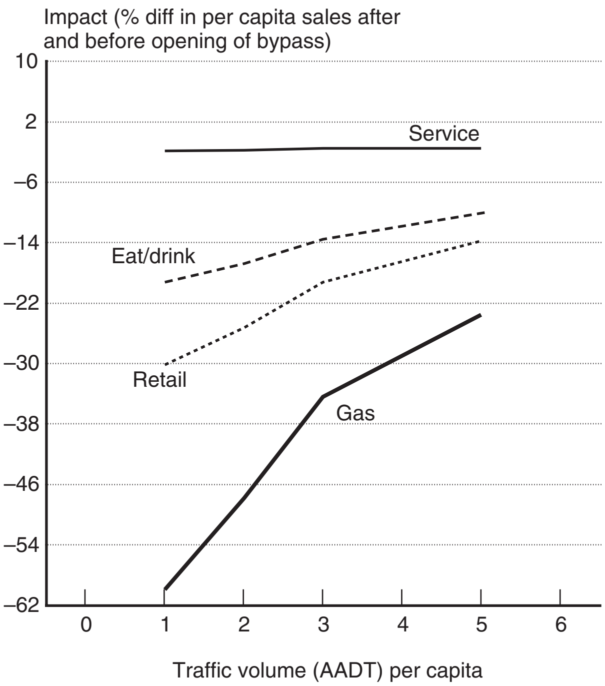
$$impact = \left(\frac{\hat{Y}_{relief\text{-}route} - \hat{Y}_{no\ relief\text{-}route}}{\hat{Y}_{no\ relief\text{-}route}} \right) \times 100 \qquad (9)$$

Four hypothetical cases were considered based on per capita traffic levels. The data for 1992 were divided into quartiles based on per capita traffic volumes. The mean per capita traffic volume was determined for each quartile; these were 1.326, 1.959, 3.180, and 5.132 AADT per person. The mean values for 1992 were used for the other explanatory variables. The impact measures derived are plotted as a function of per capita traffic volumes (figure 1).

The impact measure derived indicates an overall negative impact of the relief route on the per capita sales of the four industrial sectors analyzed. The impacts are most negative for gasoline service stations but negligible for service industries. The graph indicate that the negative impact decreases as the per capita traffic volumes in the city increase. Higher traffic levels can sustain the local economy even if a fraction of traffic is removed from the old route. It should be noted that average values of the explanatory variables were used to compute the measure and hence it represents the impact on an "average" city. For specific cases, the impact could be more or less severe, depending on the characteristics of the city in question.

As discussed earlier in this work, a higher traffic split to the bypass is estimated to have significant negative impacts on a city's per capita sales, after

FIGURE 1 Plot of the Impact Measure as a Function of Per Capita Average Annual Daily Traffic (AADT)



controlling for demographic and bypass-related variables. Several city and bypass characteristics were believed to influence traffic splits. Thus, an OLS model was run using the percentage split in traffic as the independent variable and the city and bypass characteristics as predictors. The 87 data points used to develop these models are those where a bypass exists in the larger dataset. The model results are presented in table 7. Results of this model elucidate some second-order effects of demographic and city variables on the economies of bypassed cities.

The population of the city was estimated to reduce traffic split (to the bypass). Cities that are highly populated carry higher fractions of approaching traffic. Larger cities were less affected in the original models, after controlling for traffic split, and they were also less likely to lose traffic to the bypass. Thus, a city's size provides a significant buffer, both directly and indirectly. A city's proximity to a large city, however, increases the traffic split. Nearby, large cities provide an alternative, and often more attractive, destination; so motorists may rather stop there (as opposed to stopping in the bypassed city). Therefore, proximity to a large city offers conflicting effects: it increases sales, after

TABLE 7 OLS Model for Percentage Traffic Split

	Coefficient	t stat
CONSTANT	1,048.192	2.709
YEAR	–0.52	–2.674
POPULATION	–1.93E–03	–8.481
LARGECITY POP/DIST	6.47E–04	5.023
NUM YEARS	1.677	3.509
NUM YEARS SQ	–3.40E–02	–1.893
DIST OLD	1.14E+00	1.767
DIST RATIO	1.32E+01	1.782
ACCESS CONTROL	1.62E+01	7.171
$R^2_{adj.}$	0.652	

controlling for traffic split, but also increases traffic split (which is estimated to reduce sales).

The longer the city has been bypassed, the greater is the estimated traffic split. However, the positive effect tapers off with time, as indicated by the negative coefficient on the NUM YEARS SQ variable. In addition, the longer the old route, the greater the estimated traffic split. However, perhaps to offset this, the coefficient on the distance ratio variable was estimated to be positive (and very statistically significant). This is not intuitive, since one would expect many motorists to avoid the bypass when it is longer than the old route. However, many bypassed cities have another highway passing through the city, and the bypass's location may facilitate traffic turning onto this other road (from the bypassed highway). The distance ratio may indicate the presence of such situations, and thereby be associated with a positive effect. Finally, from the model shown in table 7, the presence of frontage roads along the bypass was also estimated to increase the traffic split.

CONCLUSIONS

In this work, models were developed to study the influence of relief routes on several sectors of local economies of cities. Per capita sales in four industrial sectors (retail, gasoline service stations, eating and drinking establishments, and service industries) were considered as primary indicators of impact. Recognizing the panel nature of the data, a one-way random-effects error structure was chosen. The models were estimated as a system of *seemingly unrelated regression equations*, allowing correlation among unobserved factors impacting sales in the different sectors.

The models developed suggest that of the four sectors examined, the impact of a bypass is most negative on the per capita sales in gasoline service stations. The impact on the per capita sales in the other three sectors studied depended critically on the magnitude of the traffic diverted. When about half the approaching traffic was diverted to the bypass, all three sectors were negatively impacted. So, the better a relief route works from a traffic standpoint, the greater its adverse impact on local per capita sales. Of all the sectors studied, the service industries were minimally impacted by the bypass. As expected, per capita traffic volumes are estimated to strongly influence local sales. As the traffic levels per capita increase, the negative impacts due to the bypass are lessened.

The study also tried to identify the impact of city demographics and relief-route characteristics on the magnitude of traffic split. Larger cities lost less traffic to the bypass. Proximity to a large city increased the split. The magnitude of the split also increased with time after the opening of the bypass.

Though per capita sales were chosen as key indicators, random-effects models were also developed for the number of establishments per thousand population for each of the four industrial sectors studied (see Srinivasan (2000) for a detailed description of the methodology and empirical model results). These models again suggest that increasing traffic diversion to the new route had a negative impact on the number of establishments in all sectors but the gasoline service stations. Gasoline service stations were also negatively impacted, but this depended on the magnitude of the traffic split.

OLS models were also developed independently to study the impacts of the relief route on the population growth rates and the per capita income levels in the city. Once the city was bypassed, the population growth rate dropped 0.036% every year after the bypass opened. The bypass was also estimated to have a negative impact on the per capita income levels in the city (a decrease of approximately $50 every year after the bypass opened).

The sectors studied in this research can be expected to be the most vulnerable to a relief route. Gasoline service stations and eating and drinking places, respectively, account for only 7% and 8% of retail sales, and retail sales represent about 50% of the total sales (defined as the sum of retail, service, and wholesale industries). Sales in service industries constitute about 16% of total sales. Therefore, any negative impacts on these industrial sectors do not necessarily mean a significant negative impact on a bypassed city's overall economy.

The total traffic volumes approaching cities with relief routes were larger than those for control cities. This was probably an important reason for constructing the relief routes. Reduction in traffic volumes due to the relief route may have made the bypassed cities more like the less trafficked control cities. The average split in traffic to the new routes was about 47% in 1992. If not for the relief route, the entire traffic volume would have been carried by the old route and the congestion levels probably would have been high.

In light of these findings, transportation planners and cities should carefully consider proposals for relief routes, in order to determine if a bypass is in fact desirable and socially beneficial. Certain sectors of the economy, like gasoline service stations and eating and drinking establishments, could be critically impacted depending on the magnitude of traffic diverted. On the other hand, there are several non-economic benefits that can accrue to affected populations (e.g., safety and ease of movement downtown), and these need to be given fair weight when balancing any costs of concern. The models developed in this research provide means for assessing the magnitude of the impacts on certain sectors.

The current study used econometric modeling methods to study the impact of relief routes on the local economies at an aggregate city level. If spatially disaggregate data were available, similar methods could be extended to study the impacts along specific corridors. It would then be possible to examine possible relocation of businesses. A two-way random-effects model also would be a useful methodological extension to the current work, recognizing systematic variation in unobserved time-specific effects.

In this research effort, the impacts were measured primarily by changes in per capita sales in four different industrial sectors. Other impacts like changes in the number of establishments, population growth rates, and income levels were also modeled, but these were studied independently. A rigorous approach would be to develop a modeling framework that recognizes the dynamic interactions among the several economic indicator variables. Case studies and other investigative methods can illuminate issues like changes to quality of life, which are difficult to quantify and model statistically. Studies that employ a judicious mix of methodologies can help illuminate the different benefits and costs of bypasses. Findings from such studies will aid the planning of future bypasses in ways that improve service levels for through traffic while causing minimal distress to communities bypassed.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the Texas Department of Transportation and the research assistance of Dr. Susan Handy and Scott Kubly.

REFERENCES

Anderson, S.J., R. Harrison, M.A. Euritt, H.S. Mahmassani, M.C. Walton, and R. Helaakoski. 1992. Economic Impacts of Highway Bypasses, Report 1247-3F. Center for Transportation Research, University of Texas, Austin, Texas.

Aptech Systems Inc. 1995. *GAUSS 3.2*. Maple Valley, WA.

Baltagi, B.H. 1995. *Econometric Analysis of Panel Data*. New York, NY: John Wiley and Sons.

Buffington, J.L. and D. Burke, Jr. 1991. Employment and Income Impact of Expenditures for Bypass, Loop and Radial Highway Improvement. *Transportation Research Record* 1305:224–32.

Burress, D. 1996. Impacts of Highway Bypasses on Kansas Towns, prepared for the Kansas Department of Transportation. Policy Research Institute, University of Kansas, Lawrence.

Griffiths, W.E., R.C. Hill, and G.G. Judge. 1993. *Learning and Practicing Econometrics*. New York, NY: John Wiley and Sons.

Maddala, G.S. 1987. Recent Developments in the Econometrics of Panel Data Analysis. *Transportation Research A* 21A(4/5):303–26.

National Cooperative Highway Research Program (NCHRP). 1996. Effects of Highway Bypasses on Rural Communities and Small Urban Areas. *Research Results Digest 210*. Washington, DC: Transportation Research Board.

Otto, D. and C. Anderson. 1995. The Economic Impact of Rural Bypasses: Iowa and Minnesota Case Studies, Final Report. Midwest Transportation Center, Ames, Iowa.

Parolin, B. and B. Garner. 1996. Evaluation of the Economic Impacts of Bypass Roads on Country Towns, R&D Project TEP/93/6. New South Wales Roads and Traffic Authority, New South Wales, New Zealand.

Prucha, I.R. 1984. On the Asymptotic Efficiency of Feasible Aitken Estimators for Seemingly Unrelated Regression Models with Error Components. *Econometrica* 52 (1):203–8.

Srinivasan, S. 2000. Economic Impacts of Highway Relief Routes on Small- and Medium-Sized Cities: An Econometric Analysis. Masters Thesis, University of Texas, Austin, Texas.

Texas Comptroller of Public Accounts. 1999. Texas Mileage Guide. Available at http://www.window.state.tx.us/comptrol/texastra.html, as of May 2000.

University of Michigan. 2000. *University of Michigan Documents Center, CPI Calculator*. Available at http://www.lib.umich.edu/libhome/Documents.center/steccpi.html, as of May 2000.

Yeh, D., M. Gannon, and D. Leong. 1998. The Economic Impacts of Highway Bypasses on Communities, Technical Report SPR-0092-45-93. Wisconsin Department of Transportation, Economic Planning and Development, Madison, Wisconsin.

Rounding of Arrival and Departure Times in Travel Surveys: An Interpretation in Terms of Scheduled Activities

PIET RIETVELD
Vrije Universiteit

ABSTRACT

In travel surveys, most respondents apply rounding of departure and arrival times to multiples of 5, 15, and 30 minutes; in the annual Dutch travel survey, about 85% to 95% of all reported times are rounded. In this paper, we estimated rounding models for departure and arrival times. The model allowed us to compute the probability that a reported arrival time *m* (say *m* = 9:15 a.m.) means that the actual arrival time equals *n* (say *n* = 9:21 a.m.). Departure times appear to be rounded much more frequently than arrival times. An interpretation of this result is offered by distinguishing between scheduled and nonscheduled activities and by addressing the role of transitory activities.

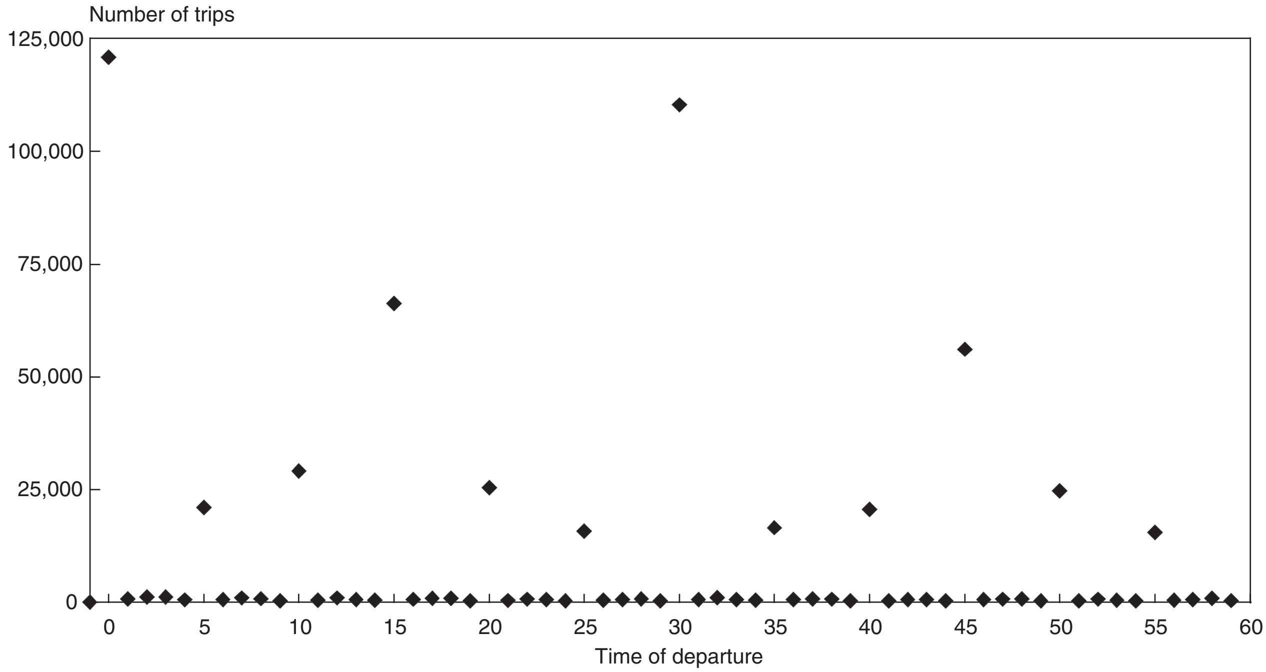
We argue that explicitly addressing rounding of arrival and departure times will have at least three positive effects. First, it leads to a considerably better treatment of reported travel time variances. Second, biases in the computation of average transport times based on travel surveys can be avoided. Third, it overcomes the problem of erratic patterns that appear in travel survey data for the minute-by-minute records of increases in the number of persons in traffic.

INTRODUCTION

Research on travel behavior is often based on travel times and distances reported by travelers. It is well

Piet Rietveld, Vrije Universiteit, Amsterdam, The Netherlands. Email: prietveld@econ.vu.nl.

FIGURE 1 Distribution of Departure Times



known that these reported values tend to be rather inaccurate. For distances, this is understandable, because there are many circumstances where travelers do not have instruments to measure distance. In the case of travel time, one might expect a more accurate measurement since most travelers wear watches, and, in particular, must pay attention to time in order to arrive at scheduled activities. Nevertheless, it is clear that inaccuracies occur (see, e.g., Rietveld et al. 1999). Some people take clock time more seriously than others, and there are also notable differences between cultures in the precision of timing activities (Levine 1997). In the present paper, we address the issue of rounding travel times—in particular, the rounding of arrival and departure times.

Consider the example of reported departure times of trips in the annual national transport survey in the Netherlands (CBS 1998). This survey is based on the travel diaries of about 144,000 randomly drawn Dutch citizens who reported their travel activities during one day in 1997. Respondents were requested to report the arrival and departure times of all trips on a certain day. Suppose a respondent j indicates that a trip started at departure time $[h_j{:}m_j]$, where h_j indicates the hour ($h_j = 0,1,...,23$) and m_j indicates the minute ($m_j = 0,1,...,59$). Let $q(m)$ denote the total number of respondents who reported their departure at minute m. Then figure 1 contains the observed distribution of the minute of departure m of all respondents ($m = 0,1,...,59$), where the hour h of departure has been deleted. The total number of reported departure times is 550,000 based on questionnaires filled out by 144,000 respondents. The figure shows extreme peaks in the distribution of reported departure times. It appears that about 22% of all travelers reported that they left at h o'clock sharp, ($h = 0,1,...,23$), whereas this figure is only 0.14% for travelers reporting that they left at 1 minute past h o'clock. Multiples of 5 and 15 minutes also get very high shares. The share of reported departure times of nonmultiples of 5 minutes is only 5%, whereas their share in multiples is about 80% (48/60). A similar pattern of reported departure times is observed in the U.S. Nationwide Personal Transportation Survey (see, e.g., Battelle 1997).

When analyzing travel statistics, it is important to be aware of rounding because unreliable data on travel times can result. For example, if departure and arrival times are normally rounded to multiples of 15 minutes, travel time will thus be rounded to multiples of 15, implying inaccurately reported travel time. Analysis of travel behavior will then be based on inaccurate travel time data. A similar conclusion holds for the analysis of travel time budgets (see, e.g., Zahavi 1977) and travel speeds. The rounding problem adds another error to the usual errors in statistical analysis (incomplete data, specification error, fundamental unpredictability of human behavior) and thus leads to larger variances of estimated coefficients.

Rounding does not only affect variances, it may even lead to a systematic bias for averages. As we will demonstrate later in this paper, there is no guarantee that in the case of travel times the probabilities of rounding upward and rounding downward are equal. Thus, rounding not only affects the reliability of individual observations, but it may also have an adverse effect on the reliability of national averages. We will demonstrate that rounding practices provide an explanation of the result reported by Battelle (1997) that the average of reported travel times is higher than the average of actual travel times.

Another example of the problem with rounding is found when departure and arrival time data are used to describe the development of traffic volumes during peak periods. Travel survey data of the type discussed here can be used to find out how many cars are on the Dutch roads from minute to minute (see, e.g., CBS 1996), but rounding can lead to erratic patterns.[1] The simplest way to overcome this would be to present data for time periods of 30 or 60 minutes, but this would imply that information is lost on how traffic volumes build up during the shoulders of the peak. This information is important for public and private decisionmakers who address congestion problems.

The above examples demonstrate how rounding departure and arrival times can affect data quality that influences transport analysis and policymaking. However, the relevance of the topic of rounding of departure and arrival times goes beyond data reliability. We will demonstrate that the rounding phenomenon sheds light on the nature of scheduling of transport-inducing activities. We develop a simple statistical model to analyze the propensity to round departure and arrival times and estimate it in the next section. An interpretation of differences between rounding in departure and arrival times is given in the discussion section in the context of scheduled activities.

FORMULATION AND ESTIMATION OF THE STATISTICAL MODEL

Formulation of the Statistical Model

As figure 1 shows, rounding departure times seems to take place toward certain anchor points such as:

- multiples of 5 minutes: 0, 5, 10, 15, 20,...,55
- multiples of 15 minutes: 0, 15, 30, 45
- multiples of 30 minutes: 0, 30
- multiples of 60 minutes: 0

Note that according to this approach the high outcome for the [*h*:00] o'clock departure time in figure 1 is the joint result of rounding to all multiples of 5, 15, 30, and 60 minutes. Another possibility is that people do not apply rounding but report the exact minute of departure.

Consider in more detail the possibility of rounding to the nearest multiple of 5. Let m be the actual minute of departure, and let d_{m5} be the absolute time distance to the nearest multiple of 5 ($d_{m5} = 1,2$). For example, when $m = 23$, the nearest multiple of 5 is 25 so that $d_{m5} = 2$. Note also that $d_{59,5} = 1$, since [$(h + 1)$:00] is the nearest multiple of 5 for [*h*:59]. The probability p_{m5} that the actual departure time m will be rounded to the nearest multiple of 5 is assumed to be:[2]

$$p_{m5} = a_5 + b_5 \cdot d_{m5} \qquad d_{m5} = 1,2$$

The coefficient a_5 is interpreted as a base value for rounding to a multiple of 5 minutes, whereas b_5 indicates the decrease of the probability of rounding as one moves away from a multiple of 5 minutes. We expect a_5 to be positive and b_5 to be negative; there is a tendency to round to the nearest multiple of 5 minutes, but this tendency decreases as one moves away from the nearest multiple of 5. For example, the probability of rounding

[1] Note that if the same level of rounding is used for both departure and arrival times, traffic volumes would be relatively stable from minute to minute. However, when rounding is greater in one of the two processes, irregular patterns will be found in the minute-to-minute records of traffic volume.

[2] Thus p_{m5} can be interpreted as the *conditional* probability that given the actual departure time m, the reported departure time is a multiple of 5 nearest to m.

11 to 10 is larger than the probability of rounding 12 to 10. Note also that as p_{m5} has to be positive, one must ensure that $a_5 + 2 \cdot b_5$ is positive.

In a similar way we formulate the rounding mechanisms for the other multiples of minutes:

$$p_{m,15} = a_{15} + b_{15} \cdot d_{m,15} \qquad d_{m,15} = 1,2,..,7$$
$$p_{m,30} = a_{30} + b_{30} \cdot d_{m,30} \qquad d_{m,30} = 1,2,..,15$$
$$p_{m,60} = a_{60} + b_{60} \cdot d_{m,60} \qquad d_{m,60} = 1,2,..,30$$

In the case of rounding to a multiple of 30 minutes, there are two nearest multiples when $m = 15$. In this case, the probabilities of rounding to [*h*:00] and [*h*:30] are assumed to be equal, so that the resulting probabilities of rounding are $(a_{30} + 15 \cdot b_{30})/2$. A similar case holds for the rounding to a multiple of 60 minutes.

After having defined these rounding probabilities, the probability that rounding of departure time m does not take place ($p_{m,0}$) equals:

$p_{m,0} = 1 - p_{m,5} - p_{m,15} - p_{m,30} - p_{m,60}$	for all m, not being multiples of 5
$p_{m,0} = 1 - p_{m,15} - p_{m,30} - p_{m,60}$	m = 5, 10, 20, 25, 35, 40, 50, 55
$p_{m,0} = 1 - p_{m,30} - p_{m,60}$	m = 15, 45
$p_{m,0} = 1 - p_{m,60}$	m = 30
$p_{m,0} = 1$	m = 0

Thus, there is only one case where we assume that rounding does not take place, that is, when $m = 0$. The resulting structure of transition probabilities can be found in table 1.

Example: when the actual time of departure m is 8:16, rounding can take place to 8:15 (via $p_{16,5}$; nearest multiple of 5), another time to 8:15 (via $p_{16,15}$; nearest multiple of 15), to 8:30 (via $p_{16,30}$; nearest multiple of 30), and to 8:00 (via $p_{16,60}$; nearest multiple of 60). The other possibility is that the actual and reported time of departure coincide (last column of table).

Consider now the distribution of actual departure times. Let g_m denote the probability that a trip made by the respondent actually starts at minute m. Then, given the conditional probabilities of rounding formulated in table 1, the *joint* probability of an actual departure time m and the reported value being its closest multiple of 5 is $g_m \cdot p_{m,5}$. Thus, we can derive the resulting probability that departures are reported to take place at time m. For example, the table demonstrates that the probability of a reported time of departure of [*h*:45], denoted as q_{45}, is the sum of probabilities of actual departures ranging from 38 to 52 minutes past h, each multiplied with its probability of rounding to 45 minutes:

$$q_{45} = [g_{38} \cdot p_{38,15} + \dots + g_{52} \cdot p_{52,15}] + [g_{43} \cdot p_{43,5} + \dots + g_{47} \cdot p_{47,5}].$$

For the other departure times, similar formulations can be derived. Note that for departure times m that are not equal to multiples of 5 we have simply:

$$q_m = g_m \cdot [1 - p_{m,5} - p_{m,15} - p_{m,30} - p_{m,60}].$$

We still have to formulate the distribution of actual departure times g_m. We will assume that all departure times within an hour are equally probable:

$$g_m = 1/60.$$

This assumption has to be made since we have no prior knowledge about the distribution of the exact minute in the hour during which departures take place.[3] Another assumption we make is that rounding is the only source of error. Thus, we will not consider other sources of error, such as mistakes made when filling out the survey questionnaire, inaccurate watches, etc. The possible implications of these assumptions are discussed at the end of the next section. These assumptions suffice for a specification of the likelihood q_m for all reported departure times m. Let N_m denote the actual number of times that departure minute m is reported by respondents. Then the resulting log-likelihood of the reported departure time m is:

$$lnL = N_0 \, ln \, q_0 + N_1 \, ln \, q_1 + \dots + N_{59} \, ln \, q_{59}$$

[3] Of course we have fairly accurate information about the distribution of departure times during the 24 hours of the day: during the night, the number of departures is much smaller than during the day. However, very little is known about the distribution between the minutes within the hour.

TABLE 1 Probability of Rounding the Actual Time of Departure *m* by a Respondent to the Nearest Multiple of 5, 15, 30, or 60 Minutes (below or above *m*), or to *m* Itself

	Time of departure in minutes reported by a respondent given his actual departure time *m*								
Actual time of departure: *m*	5— below *m*	5— above *m*	15— below *m*	15— above *m*	30— below *m*	30— above *m*	60— below *m*	60— above *m*	*m*— no rounding
0	0	0	0	0	0	0	0	0	1
1	$p_{1,5}$	0	$p_{1,15}$	0	$p_{1,30}$	0	$p_{1,60}$	0	$1-p_{1,5}-p_{1,15}-p_{1,30}-p_{1,60}$
2	$p_{2,5}$	0	$p_{2,15}$	0	$p_{2,30}$	0	$p_{2,60}$	0	$1-p_{2,5}-p_{2,15}-p_{2,30}-p_{2,60}$
3	0	$p_{3,5}$	$p_{3,15}$	0	$p_{3,30}$	0	$p_{3,60}$	0	$1-p_{3,5}-p_{3,15}-p_{3,30}-p_{3,60}$
4	0	$p_{4,5}$	$p_{4,15}$	0	$p_{4,30}$	0	$p_{4,60}$	0	$1-p_{4,5}-p_{4,15}-p_{4,30}-p_{4,60}$
5	0	0	$p_{5,15}$	0	$p_{5,30}$	0	$p_{5,60}$	0	$1-p_{5,15}-p_{5,30}-p_{5,60}$
6	$p_{6,5}$	0	$p_{6,15}$	0	$p_{6,30}$	0	$p_{6,60}$	0	$1-p_{6,5}-p_{6,15}-p_{6,30}-p_{6,60}$
7	$p_{7,5}$	0	$p_{7,15}$	0	$p_{7,30}$	0	$p_{7,60}$	0	$1-p_{7,5}-p_{7,15}-p_{7,30}-p_{7,60}$
8	0	$p_{8,5}$	0	$p_{8,15}$	$p_{8,30}$	0	$p_{8,60}$	0	$1-p_{8,5}-p_{8,15}-p_{8,30}-p_{8,60}$
9	0	$p_{9,5}$	0	$p_{9,15}$	$p_{9,30}$	0	$p_{9,60}$	0	$1-p_{9,5}-p_{9,15}-p_{9,30}-p_{9,60}$
10	0	0	0	$p_{10,15}$	$p_{10,30}$	0	$p_{10,60}$	0	$1-p_{10,15}-p_{10,30}-p_{10,60}$
11	$p_{11,5}$	0	0	$p_{11,15}$	$p_{11,30}$	0	$p_{11,60}$	0	$1-p_{11,5}-p_{11,15}-p_{11,30}-p_{11,60}$
12	$p_{12,5}$	0	0	$p_{12,15}$	$p_{12,30}$	0	$p_{12,60}$	0	$1-p_{12,5}-p_{12,15}-p_{12,30}-p_{12,60}$
13	0	$p_{13,5}$	0	$p_{13,15}$	$p_{13,30}$	0	$p_{13,60}$	0	$1-p_{13,5}-p_{13,15}-p_{13,30}-p_{13,60}$
14	0	$p_{14,5}$	0	$p_{14,15}$	$p_{14,30}$	0	$p_{14,60}$	0	$1-p_{14,5}-p_{14,15}-p_{14,30}-p_{14,60}$
15	0	0	0	0	$½p_{15,30}$	$½p_{15,30}$	$p_{15,60}$	0	$1-p_{15,30}-p_{15,60}$
16	$p_{16,5}$	0	$p_{16,15}$	0	0	$p_{16,30}$	$p_{16,60}$	0	$1-p_{16,5}-p_{16,15}-p_{16,30}-p_{16,60}$
·	·	·	·	·	·	·	·	·	·
29	0	$p_{29,5}$	0	$p_{29,15}$	0	$p_{29,30}$	$p_{29,60}$	0	$1-p_{29,5}-p_{29,15}-p_{29,30}-p_{29,60}$
30	0	0	0	0	0	0	$½p_{30,60}$	$½p_{30,60}$	$1-p_{30,60}$
31	$p_{31,5}$	0	$p_{31,15}$	0	$p_{31,30}$	0	0	$p_{31,60}$	$1-p_{31,5}-p_{31,15}-p_{31,30}-p_{31,60}$
·	·	·	·	·	·	·	·	·	·
59	0	$p_{59,5}$	0	$p_{59,15}$	0	$p_{59,30}$	0	$p_{59,60}$	$1-p_{59,5}-p_{59,15}-p_{59,30}-p_{59,60}$
60	0	0	0	0	0	0	0	0	1

Under the null hypothesis that reported departure times are equal to the actual departure times, all probabilities in table 1 are equal to zero, except the ones in the last column. This implies that

$$ln\ L_0 = N_0\ ln(1/60) + N_1\ ln(1/60) + \ldots + N_{59}\ ln(1/60) = N\ ln(1/60)$$

where N equals the total number of observations.

Estimation of the Model: Departure Times

The results of the maximum likelihood estimation for the departure minutes are reported in table 2. The likelihood values indicate strong support for rejection of the null hypothesis. The test statistic $\chi^2 = 2(lnL - lnL_0)$ is an asymptotically distributed chi-square with degrees of freedom equal to the number of restrictions on the parameters (8). The value of the test statistic corresponding to a 99% probability of rejection of the null hypothesis is 20.1 in this case. We found overwhelming evidence of the importance of rounding to multiples of 5, 15, and 30 minutes: their base values a_5, a_{15}, and a_{30} are clearly significant. Only rounding to the whole hour assumes a small value (a_{60} is less than 1%). The b values were very small, with the exception of b_5, indicating that the probability of rounding 4 to 5 equals 46.4%, whereas rounding 3 to 5 equals 42.8%. For rounding to multiples of 15, 30, and 60, the b values were positive, which was unexpected. Their levels were very small, however. The reason that some of them are significant is that the number of observations is large. Considering the magnitudes they assume, they can be ignored. Thus, we conclude that, with the exception of rounding to multiples of 5 minutes the rounding probabilities hardly depend on the distance to the reference value.

To illustrate the meaning of the estimates, we computed the implications for the rounding proba-

TABLE 2 Estimation of Rounding Model for Departure Times

Coefficient	Maximum likelihood estimate	Standard error
a_5	0.500	0.00142
b_5	–0.0360	0.00075
a_{15}	0.284	0.00142
b_{15}	0.0016	0.00017
a_{30}	0.177	0.00149
b_{30}	0.00015	0.00008
a_{60}	0.0093	0.00108
b_{60}	0.00055	0.00004
log-likelihood	$-1.376.10^{6}$	
log-likelihood (L_0)	$-2.252.10^{6}$	

bilities when the actual observation is 19 minutes after the hour. The following rounding possibilities and the corresponding probabilities are:

to 0 minutes after the hour (the nearest multiple of 60): 2.1%

to 15 minutes after the hour (the nearest multiple of 15): 29.0%

to 19 minutes after the hour (no rounding): 4.6%

to 20 minutes after the hour (the nearest multiple of 5): 46.4%

to 30 minutes after the hour (the nearest multiple of 30): 17.9%.

The estimation result in table 2 indicates that rounding to multiples of 5 minutes dominates when we consider an individual observation. Note, however, that rounding to a certain multiple of 5 (say n) only takes place for the 4 nearest neighbors (n–2, n–1, n + 1, n + 2). With the multiples of 15, 30, and 60, the numbers of these neighbors are 14, 29, and 59, respectively. Thus, the base values for a_5 to a_{60} must be multiplied by factors 4 through 59 to calculate the total number of reported departure times. In that case, the 30-minute multiple is used most frequently, and this is confirmed by the original data in table 1.

Estimation of Model: Arrival Times

A similar approach was applied to arrival time data. The raw data are presented in figure 2. It shows a pattern similar to the departure time figures, although the scores are less peaked in multiples of 5. The share of unrounded departure times is clearly higher (about 15% are rounded to a value like 1, 2, 3, 4, 6, 7, etc., as opposed to about 5% for arrival times).

Estimation results are shown in table 3. The results of the arrival time estimates are to some extent similar to the departure time roundings: the 60-minute rounding was the least important, and the b values were negligible, except b_5. A striking difference between departure and arrival times is that rounding to a multiple of 5 was much more dominant for arrival times. To illustrate, we again computed the rounding probabilities when the actual time of arrival was 19 minutes after the hour:

to 0 minutes after the hour (the nearest multiple of 60): 0.0%

to 15 minutes after the hour (the nearest multiple of 15): 9.3%

to 19 minutes after the hour (no rounding): 10.4%

to 20 minutes after the hour (the nearest multiple of 5): 76.0%

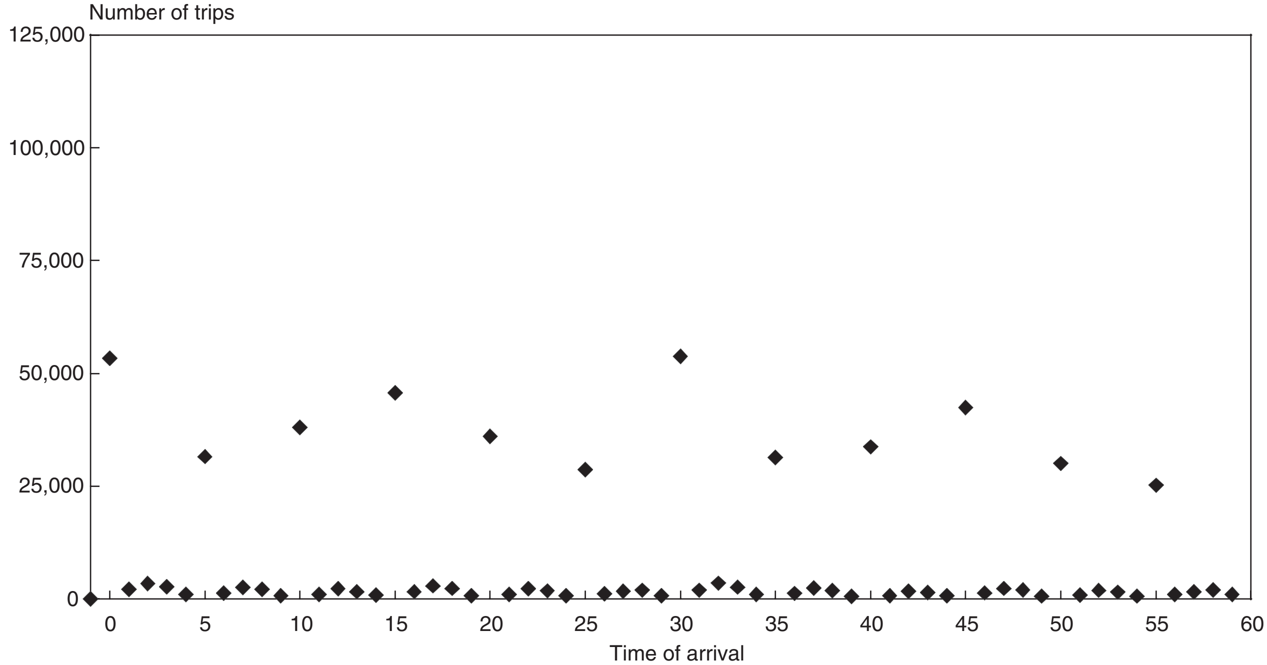
to 30 minutes after the hour (the nearest multiple of 30): 4.3%.

Thus, rounding to multiples of 5 minutes was dominant. Absence of rounding had the next highest shares and rounding to the nearest multiple of 15 was fairly unimportant. Rounding probabilities to multiples of 30 and 60 minutes were small.

Distribution of Actual Departure Times Conditional on Reported Departure Times

We conclude this discussion by noting that we have now derived the distribution of *reported departure time, conditional on the actual departure time*. It

FIGURE 2 Distribution of Arrival Times



may also be interesting to derive the reverse: the distribution of the *actual departure time, conditional on the reported departure time*. For example, when the reported time of departure m equals 15 minutes, what is the probability that the actual time n equals 8, 9, 10, and so forth? This can be achieved by using Bayes' formula (Hogg and Craig 1970). Let $p_{m,n}$ be the probability of the reported time m given the actual departure time n (estimated above), and let g_n be the distribution of actual departure times. Then the joint density $f(m,n)$ of m and n equals

$$f(m,n) = p_{m,n} \cdot g_n$$

Since we want to determine $k(n|m)$, the distribution of the probability of an actual arrival at n given a reported value m, we make use of the Bayes' formula

$$k(n|m) = [p_{m,n} \cdot g_n]/[p_{m,0} \cdot g_0 + p_{m,1} \cdot g_1 + \dots + p_{m,59} \cdot g_{59}].$$

Since we assume that the density of the actual departure time $g(n)$ is given as

$$g_n = 1/60 \text{ for } n = 0,\dots,59,$$

the Bayes' formula can be simplified as

$$k(n|m) = p_{m,n} / [p_{m,0} + p_{m,1} + \dots + p_{m,59}].$$

Application of this formula to, for example, $k(4,4)$ implies that $k(4,4) = 1$: when the reported time of departure equals 4, one can be sure that the actual departure time equals 4. On the other hand, we find the following probabilities (table 4) for the actual values underlying the reported observation $m = 15$. The table shows that a reported departure time of $m = 15$ means the probability that the actual departure time is indeed 15 is only 12.5%. The higher probabilities for the actual departure time are found in the range between 13 and 17 minutes, but the share for the remaining departure times is still substantial (43%).

Information of this type can be used in further statistical analyses of travel behavior data to give an adequate representation of errors in variables (see e.g., Johnston 1984). An important implication of our approach is that rounded observations of travel times have a much larger variance than unrounded ones. For example, in our approach, the reported duration of a trip of 32 minutes has a much smaller variance than a trip with a reported duration of 30

TABLE 3 Estimation of Rounding Model for Arrival Times

Coefficient	Maximum likelihood estimate	Standard error
a_5	0.900	0.00201
b_5	–0.1400	0.00127
a_{15}	0.065	0.00165
b_{15}	0.0071	0.00028
a_{30}	0.014	0.00146
b_{30}	0.0026	0.00014
a_{60}	–0.00005	0.00014
b_{60}	0.00006	0.00002
log-likelihood	$–1.615.10^6$	
log-likelihood L_0	$–2.252.10^6$	

TABLE 4 Probability (%) of Actual Departure Time (n = 8,…,22) Given a Reported Departure Time of m = 15

Actual departure time n	Probability of actual departure time given reported value of departure time m = 15
8	4.3
9	4.3
10	4.3
11	4.3
12	4.3
13	10.8
14	11.4
15	12.5
16	11.4
17	10.8
18	4.3
19	4.3
20	4.3
21	4.3
22	4.3

minutes.[4] Such differences in variance are not well captured in standard econometric methods.

DISCUSSION

One may wonder why the rounding rules applied to arrival times are more accurate than those for departure times (rounding to multiples of 15 and 30 minutes take place much less frequently). Various explanations exist.

The structure of the questionnaire. The question on the times of departure and arrival are posed in an identical way: "At what time did you depart/arrive? .... hour .... min." Note that these questions invite respondents to give an exact specification of the departure/arrival time. We conclude that the difference in the rounding practice for arrivals and departures cannot be explained by the way the questions are phrased.

Another point is that most respondents will fill out the questionnaire at the end of the day. Many of them will have forgotten their exact minute of departure and arrival for trips made 3 to 15 hours earlier. This explains the practice of rounding, but

[4] For example, in the most extreme case, a 2-minute trip with a departure at 8:14 and arrival at 8:16 may be reported as a 30-minute trip after rounding. The same holds true for a 58-minute trip that started at 8:16 and ended at 9:14. This illustrates the large range on which a trip with a reported duration of 30 minutes may be based. On the other hand, a trip starting at a reported time of 8:16 and ending at 8:48 will just have lasted 32 minutes according to our model, implying a 0 variance (remember that apart from rounding, all other data errors are ignored in our analysis).

it does not explain why it occurs more often with departures than with arrivals.

Structure of public transport timetables. A bias of public transport timetables toward multiples of 30 minutes as frequently used departure times might influence the reported departure times.[5] Such a timetabling practice, however, does not exist in The Netherlands. Note also that departure times reported here relate to the whole chain, so that the departure time would not indicate the time of departure of the train, but the time the respondent leaves to make a trip. A final observation is that in developed countries the only collective transport mode that does not use timetables at the one-minute level of precision is aviation (it uses multiples of 5).

As opposed to public transport time tables, most nontransport activities have a scheduled start at multiples of 15, 30, or 60 minutes: examples are hours at school, meetings, appointments, work, church services, sport events, cinema performances, etc. In some cases, both the *start* and *end* times are exactly specified, but often the beginning is more rigid and explicit than the end. This may create the perception that an important share of activities start at multiples of 15, 30, or 60 minutes and that a smaller share end at multiples of 15, 30, or 60 minutes. Consequently, the expectation is that the concentration of reported times at multiples of 15, 30, and 60 minutes is larger for arrivals than for departures. However, the data reveal that the opposite takes place. On the other hand, there are many activities that are not scheduled. For example, the arrival at home after an activity is usually not followed by an activity scheduled at an exact point in time. Thus, the share of scheduled activities in activity patterns must not be exaggerated.

Another point is that the start/end of an activity does not necessarily coincide with the arrival/departure of a trip. In many cases, there are *transitory activities* (e.g., relax, wait, talk to other participants, deposit one's coat at the cloak room, report at the entrance, find one's way to the exact place of the activity, wait for the elevator). The Dutch travel survey (like many other travel surveys) does not specify these transitory activities, so it is left to the respondent whether he considers them as part of the trip or of the activity carried out. Consider the case of a student whose lecture is scheduled to end at 12:45 sharp; in reality it ends at 12:47, the student talks to his classmates until 12:49, and he then leaves the university building at 12:53 to walk to his car, which he starts to drive at 12:56. Then he may answer the question "at what time did you leave" by filling out any of the above-mentioned times, plus rounded times such as 12:45, 12:50, 12:55, and 13:00 o'clock. A similar story, of course, holds true for transitory activities before a scheduled activity.

The question remains—why are people more inclined to round with departure times than with arrival times. Probably, the most important answer is that *scheduled activities force people to plan their trips in advance, which provides them with anchor points for their memory afterward*. At the end of the day, they will still remember whether they arrived long before the scheduled time, or whether they were late. Since, as mentioned above, scheduling takes place more often in terms of the start of an activity rather than the end, people will have more precise memories about the time of arrival and they will therefore also have a tendency to apply rounding less frequently than with departures. This sheds some light on the literature of scheduling. As put forward, for example, by Small (1982; 1992) and Wilson (1989), travelers face the challenge of arriving on time to scheduled activities (e.g., the start of work or the start of a business meeting). Given a high penalty for arriving late, travelers tend to take into account transport systems that are unreliable (congestion caused by nonrecurrent events, delays or missed connections in public transport) and thus plan their trip in such a way that delays can be accommodated. This means that we may expect travelers to arrive early in cases of scheduled activities with penalties and uncertainty in travel times. Because of the penalty for a late arrival, the traveler will have a keen eye on whether he really arrived early or late. When he arrives early, the traveler will have an additional type of transitory activity—waiting time, which is a cushion to avoid being late.

[5] Public transport maintains a 5% share of the total number of trips in the Netherlands. Its share in the total number of kilometers traveled is about 13%.

Thus, we arrive at several differences between the start and the end of an activity. First, the start is more often fixed in time than the end is. Second, the element of transport system uncertainty is present for the person who needs to meet requirements of being on time; it does not play a role at the end of the meeting. Third, the penalty for arriving late may be perceived to be larger than the penalty of leaving early.[6] These three differences imply that on average travelers will be much more concerned about the starting time of activities than the time they end.

We finish this section with a discussion of the possible implications of two assumptions on which the above estimations are based: uniform distribution of actual departure times during an hour and absence of measurement errors. The assumption that departure and arrival minutes are distributed uniformly was made since we have no prior knowledge about the distribution of the exact minute in the hour during which departures take place. One might argue that since scheduled activities usually end at 0, 15, 30, or 45 minutes after the hour, there will be a tendency that the density of actual departure times is higher at those times. This would offer an alternative interpretation for the empirical results. With the given data, this alternative interpretation cannot be falsified. However, it may be argued that it is not a very plausible explanation for several reasons.

First of all, we can make use of other data sources that include both actual and reported departure times. From a survey done in the United States (Battelle 1997) among car drivers in Lexington, it appears that the distribution of actual departure times is very close to uniform. The second reason is that transport statistics show that a considerable portion of human activities are not strictly scheduled: in the Netherlands more than half of all movement relates to activities such as shopping, recreation, and social visits (CBS 1998). Therefore, an outcome of 95% of actual departures taking place at round minutes (i.e., at multiples of 5) would be implausible. Another reason is that this explanation ignores the importance demonstrated above of transitory activities taking place between the end of an activity and the start of a trip. Another argument concerns trips where scheduled public transport services are used. The departure times at bus stops and railway stations tend to be distributed uniformly during the hour, so that one would expect a uniform distribution of departure times as described in the earlier section on formulation and estimation of the statistical model.[7] Also, the discussion above of the difference between the distribution of departure and arrival times strongly supports the view that the peaks in the distribution of reported times are due to rounding and not to peaks in actual times. We noted that if an activity is scheduled, the certainty about its starting time is usually higher than about its end time. Therefore, if the distribution of reported departure and arrival times were dictated by the actual start of these activities, one would expect larger peaks in the distribution of arrival times compared with departure times. In reality, however, the opposite is true.[8]

We conclude that with the given data we cannot test whether the distribution of actual departure minutes is uniform. It is highly implausible, however, that a non-uniform distribution is the sole reason for the peaks in the reported departure times. One cannot exclude, however, the possibility that there is a tendency for more people to arrive and

[6] We do not go into details about chaining activities with fixed start and end times. Travelers who are able to leave a sufficient amount of time between the end of one activity and the start of a second activity may then have spare time for an additional type of transitory activity. When the time is not sufficient, the traveler reveals which of the two activities will have the higher penalty (leaving early versus arriving late).

[7] What really matters is not the official departure time of the public transport services, but the departure time of the traveler from his origin, thus taking into account the access time to the public transport node. Thus, even if there is a tendency for public transport timetables to be biased toward departure times of the services in multiples of 5 minutes, the variance in the access times would make this invisible when departure times of travelers are considered.

[8] Another possibility with arrival times is that the distribution of actual times has high probabilities at times just before round minutes because most people try to arrive on time. However, inspection of the reported arrival times does not reveal such a tendency. For example, the data in figure 2 even demonstrate a slight tendency in the opposite direction: the share of respondents reporting they arrived between 1 and 15 minutes after the hour is somewhat larger than the share reporting they arrived between 45 and 59 minutes before the hour (26% versus 22%).

depart at round minutes rather than at other minutes. If this were true, it would imply that we have overestimated the rounding tendency. Given the above arguments, the possibility of overestimation is probably small.

The second assumption in the Formulation and Estimation section that may need some discussion concerns the premise that rounding is the only source of error when reporting departure and arrival times. In the statistical analysis, we ruled out the possibility that people report wrong departure times because of mistakes, inaccurate watches, or bad memory. Of course such errors will take place frequently in travel surveys and they will in part express themselves in rounding. For example, if a respondent does not remember the exact times at the end of the day, he may use proxies. In cases where these mechanisms do not express themselves via rounding, they contribute to the variance of error in observed data, but there is no reason to expect that they will lead to systematic distortions in the analysis of rounding.[9]

CONCLUDING REMARKS

Our analysis of departure and arrival times indicates that rounding is a rule, rather than an exception. About 5% to 15% of all reported times assume values that are not multiples of 5, whereas these are 80% of the possible clock times. In the case of scheduled activities, the reported times are probably more precise because scheduling implies the use of anchor points in the timeframe. With fixed schedules, there may be a high penalty for being late so that travelers will be more likely to remember the exact timing of trips. Since scheduling of start times takes place more often than for end times, it is plausible that reported times of arrival are more accurate than reported times of departure.

In the research on travel behavior, data on travel times usually play an important role. These travel times follow as the result of subtracting reported times of arrival and departure. Given the large rounding errors observed here, it is clear that errors in reported travel times (and related variables such as travel speeds) will be large. This "error in data" phenomenon will obviously hamper the analysis of data on individual travel behavior. In the present paper, we developed a method, based on a Bayesian approach to derive the probability that a reported arrival time m means that the actual arrival time equals n. This method can be used in "errors in variable methods" to give an adequate representation of the measurement error. We demonstrated that the variance of rounded travel times is much larger than that of unrounded times. This approach must be considered superior to the usual approach where all measurement error is supposed to be represented by a common variance.

Rounding has a larger impact than just affecting the variance of travel times, however. Given the large scale at which rounding takes place, it may also affect averages computed on the basis of national surveys when probabilities of rounding upward and downward do not cancel. Consider, for example, the distribution of reported trip duration in the Netherlands. This distribution is skewed: the most frequently reported trip duration (mode) is 10 minutes, the median value is 15 minutes, and the mean value is about 20 minutes. Therefore, the number of trips with an actual duration of between 15 and 30 minutes will be considerably larger than the number of trips with a duration between 30 and 45 minutes. As a result, the probability of rounding upward is considerably higher than the probability of rounding downward.[10] The conclusion is that in this case rounding of arrival and departure times leads to overestimates of average travel times.[11]

Finally, ignoring the rounding problem could lead to erratic patterns when the travel survey data are used to give a minute-by-minute record of the number of travelers in the transport system. Consider, for example, the 24-hour average number of people in transport in each minute for our sample of 550,000 respondents. The departure and

[9] Note also that without additional data, adding an error term ε_m with mean zero and variance σ^2 to the model, such that the reported departure time is equal to the actual departure time plus ε_m, will not yield meaningful estimates of σ.

[10] This implies that the figures of 20 and 15 minutes mentioned in the text for mean and median are biased. The effect on the mean is probably larger than on the median.

[11] In the Battelle study (1997), a comparison of reported and actual travel times indeed revealed that reported travel times based on recall generally overstate travel time. A similar conclusion was drawn about travel distances.

arrival data indicate that during the first minute of the hour 120,000 persons enter the transport system, whereas only 55,000 persons leave. This would imply a sudden net increase of 65,000 persons during 1 minute, which is much higher than the small net decreases during subsequent minutes of about 1,500 persons per minute. This obviously hinders a proper assessment of the development of the number of persons in traffic in the course of time. By using the Bayesian approach presented earlier, this problem can be overcome.

In our discussion of rounding, we touched on the importance of *transitory activities* in scheduled activity patterns. These transitory activities are often ignored in the analysis of travel behavior. A main reason for these transitory activities is that they emerge in a response to reduce the penalty of arriving late at a scheduled activity. They also result from infrequent public transport services. Transitory activities are important to reduce bottlenecks in internal and external transport systems. An example of an internal transport system is elevator capacity, which usually will not allow everybody to arrive just in time or leave immediately after a big event. Similarly, parking facilities do not function well under these circumstances. An example of external transport systems concerns the capacity to absorb visitors for large-scale events in stadiums, exhibition centers, etc. Transitory activities do not only keep bottleneck problems manageable, they may also have value per se for the travelers. They deserve more attention in transport behavior than they usually get. To properly analyze their presence and size, detailed questionnaires are needed.

A final point of attention is the possibility of linking reported time data to archived global positional data. The combination of geographic information systems and global positioning systems offers great potential for improving the quality of data on travel time and distance in passenger surveys. This holds true not only for automobile trips, but these systems may also provide useful data on other kinds of trips (Quiraga and Bullock 1998; Uchida et al. 2001).

ACKNOWLEDGMENT

The author thanks Uty Pang Atjok who provided computational assistance. In addition, David Greene and two anonymous referees gave constructive comments.

REFERENCES

Battelle Transportation Division. 1997. *Lexington Area Travel Data Collection Test: Global Positioning Systems for Personal Travel Surveys*. Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Central Bureau of Statistics (CBS). 1996. *Autos in Nederland (Cars in the Netherlands)*. Deventer, The Netherlands: Kluwer.

_____. 1998. *De Mobiliteit van de Nederlandse Bevolking (The Mobility of the Dutch Population)*. Voorburg, The Netherlands.

Hogg, R.V. and A.T. Craig. 1970. *Introduction to Mathematical Statistics*. London, England: Macmillan.

Johnston, R.J. 1984. *Econometric Methods*. New York, NY: McGraw-Hill.

Levine, R.V. 1997. *A Geography of Time*. New York, NY: Basic Books.

Quiraga, C.A. and D. Bullock. 1998. Travel Time Studies with Global Positioning and Geographic Information Systems: An Integrated Methodology. *Transportation Research C* 6:101–27.

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* 33:269–88.

Small, K.A. 1982. The Scheduling of Consumer Activities: Work Trips. *American Economic Review* 72:467–79.

_____. 1992. *Urban Transportation Economics*. Luxemburg: Harwood Publications.

Uchida, T., M. Pursula, A. Suzuki, Y.H. Lee, and D. Takehiko. 2001. Monitoring Personal Movement: Development of PEAMON (Personal Activity Monitor) for Automated Trip-Diary. Tohoku University, Japan.

U.S. Department of Transportation, Bureau of Transportation Statistics (BTS). 1997. *Nationwide Personal Transportation Survey*, Washington, DC.

Wilson, P.W. 1989. Scheduling Costs and the Value of Travel Time. *Urban Studies* 26:356-66.

Zahavi, Y. 1977. *The UMOT Model*. Washington, DC: The World Bank.

JOURNAL OF TRANSPORTATION AND STATISTICS

Guidelines for Manuscript Submission

Send all submission materials to

Marsha Fenn, Managing Editor
Journal of Transportation and Statistics
Bureau of Transportation Statistics
U.S. Department of Transportation
400 7th Street SW, Room 7412, K15
Washington, DC 20590
Email: journal@bts.gov

Please note: As the manuscripts are subject to peer review, submission of a paper indicates the author's intention to publish in the *Journal of Transportation and Statistics*. 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 editor if in doubt.

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.

The cover page should include the title of the manuscript, author name(s), and the telephone number and surface and email addresses of the contact author. The second page should contain a 100-word abstract briefly describing the contents of the paper, the research method, and the results. 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 although exceptions may be made in certain cases.

References follow the style outlined in the *Chicago Manual of Style*. All non-original material must be sourced.

Accepted papers must be submitted electronically in addition to a hard copy. Make sure the hard copy corresponds to the electronic version. As the publication date nears, authors will be required to return article page proofs to the Managing Editor within 48 hours of their receipt.

Acceptable software for text and equations is limited to Word, WordPerfect, and LaTeX. Data behind all figures, maps, and charts must be provided in Excel, Word, or DeltaGraph. Data in 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 GIS programs. Please save your GIS files using the following extensions: .eps or .pdf. If this is not possible, maps may be provided in ArcView3.x project files (.apr), ArcGIS8.x map files (.mxd), or Arc/Info metafiles (.map). Files using the .apr or .mxd extensions can be supplied on a CD-ROM.

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 5 Number 1, 2002
ISSN 1094-8848

CONTENTS