A REGRESSION MODEL OF THE NUMBER OF TAXICABS IN U.S. CITIES

Bruce Schaller
Principal

January 2005

Schaller Consulting
94 Windsor Place
Brooklyn, NY 11215
voice: (718) 768-3487
fax: (718) 768-5985
schaller@schallerconsult.com
www.schallerconsult.com

ABSTRACT

In cities that control the number of taxicabs by law or regulation, setting the number of cabs is one of the most important decisions made by taxicab regulators and elected officials. Licensing either too many or too few cabs can have serious deleterious effects on the availability and quality of service and the economic viability of the taxi business. Yet local officials often have difficulty quantifying the demand for taxi service or tracking changes in demand.

Multiple regression modeling of the number of cabs in 118 U.S. cities identifies three primary demand factors: the number of workers commuting by subway, the number of households with no vehicles available, and the number of airport taxi trips. These results can be used to identify peer cities for further comparison and analysis and to guide regulators in measuring changes in local demand for cab service.

INTRODUCTION

Most taxicab regulatory authorities control market entry into the taxi business, according to Gilbert et. al.'s 1998 survey of taxi operators. (2002) Taxi regulators' decision as to how many cabs to license is one of the most important decisions that they make. If regulators allow too few taxicabs, the resulting undersupply will create lengthy waits for cab service and sometimes prevent customers from obtaining service at all. Conversely, an oversupply of cabs can lead to service problems such as aging and ill-kept cabs and high turnover among underpaid and poorly qualified drivers.

Not only the general public but also social service providers and other transportation providers can be adversely affected by oversupply or undersupply of cab service. Social service agencies that subsidize taxi trips for seniors and disabled persons can find that these clients, who tend to take short trips and often need assistance, have difficulty obtaining cab service. Transit agencies that could achieve cost savings by contracting a portion of paratransit trips to taxi companies may find that the companies lack adequate capacity or are unable to provide the desired quality of vehicles and drivers.

Various methods are used in U.S. cities and counties to set the number of taxi licenses. The simplest (and most arbitrary) method is to freeze the number of cabs in operation at the time the decision is made – a policy adopted in New York, Chicago, Boston and other major cities during the 1930s. Another common approach is to require taxicab companies to show the "public convenience and necessity" (PCN) of increasing the size of the industry. Sometimes the PCN standard is married to a periodic review that may produce regular expansion of the industry in accord with growing demand. A related approach is to set a ratio between the number of cabs and an index based on population, taxi trip volumes or other factors.

Whichever method is chosen, taxicab regulators and elected officials need a means to objectively assess the appropriate number of cabs for their jurisdiction. This assessment should consider the availability of cab service, the effectiveness of company dispatch operations, the industry's financial condition and changes in taxi demand in recent years. It can also be valuable to compare the number of cabs locally with the number in comparable cities – a question often asked by elected officials.

This paper addresses two elements of this assessment. Using a multiple regression model of the observed number of taxicabs in 118 U.S. cities and counties, the paper identifies the primary factors that generate demand for taxicab service in the U.S. These results can help regulators build a time-series analysis of changes in local demand for cab service. Second, the model results can be used for benchmarking purposes to make comparisons between comparable cities or counties.

MODEL SPECIFICATION

The most obvious factor associated with taxi demand is population. In general, larger cities have more cabs. Regulators often compare the number of cabs in their jurisdiction with the number of cabs in cities or counties of about the same size. They may also compute the ratio of taxicabs to population and compare ratios among different cities.

The shortcoming of population and population ratios is the lack of a standard ratio of taxicabs to population. Cities with similar populations often have a quite different number of cabs: Houston (2,245 cabs) and King County, Washington (1,145); Detroit (1,310 cabs) and San Jose (465); New Orleans (1,608) and Cleveland (460); Salt Lake City (268) and Laredo (78); Alexandria, Virginia (645) and Ann Arbor, Michigan (85). Figure 1 graphs the wide variation in the ratio of taxis per 1,000 population in 118 U.S. cities.

Recognizing the need to take into account additional factors that are more closely related to taxi demand, regulators have used employment, transit ridership and indicators of tourism, business visitors and convention activity to evaluate the need for issuing additional taxicab licenses. Regulators also sometimes use factors endogenous to the taxi industry, principally the number of trips or ratio of trips to taxicabs.

A few formal studies have assessed determinants of taxi demand, although no previous study involves the number of cities in the model reported in this paper. Time-series models have been estimated for London, New York and Toronto. These models found population, employment, visitation, taxi fares, transit ridership, and seasonal factors to be statistically significant variables explaining changes in taxi demand. (Beesley 1979; Schaller 1999; Economic Planning Group 1998) Other studies using a sample of cities add low-income persons, motor vehicle operating costs and bus service miles as influences on taxi demand. (Hara Associates 1994; Fravel and Gilbert 1978) Notably, transit ridership is found to be a complement to taxi use rather than a substitute. (Economic Planning Group 1998; Fravel and Gilbert 1978)

Determinants of taxi demand can be organized into seven conceptual variables, each of which can be operationalized with one or more variables, as shown in Figure 2.

Several comments can be offered about this conceptual framework. First, city size measured by population or employment is not necessarily a good predictor of taxicab demand. As noted above, cities of about the same size may have a vastly different number of taxicabs.

Second, taxi users are often thought of as predominantly persons living in lower-income households that lack access to an automobile. In fact, this is not the case; only 39 percent of taxi trips are taken by members of no-car households. (Unpublished data from 2001 National Household Travel Survey. Persons in no-car households are in fact relatively heavy taxi users; taxicabs served 1.6 percent of trips for persons without a motor vehicle as compared with 0.15 percent for the entire population.)

Third, airport service is an important component of the demand for taxi service. Airport trips represent from one-third to one-half of taxi demand in some cities.

Fourth, transit ridership is expected to be a complement to taxicabs rather than a substitute. Transit usage creates demand for taxi service when transit riders use cabs to access transit stations or after exiting stations. In addition, the availability and attractiveness of transit can spur taxi demand for other parts of a journey. For example, one may use transit for the morning trip to work but later take a cab home after a night on the town.

Fifth, demand for service is also affected by the quality of service. Surveys of taxi users have found that customers would use taxicabs more often if cab service were more reliable and more readily available. Thus, poor service quality may reduce demand in some places.

Figure 1. Taxicabs per 1,000 population for 118 U.S. cities

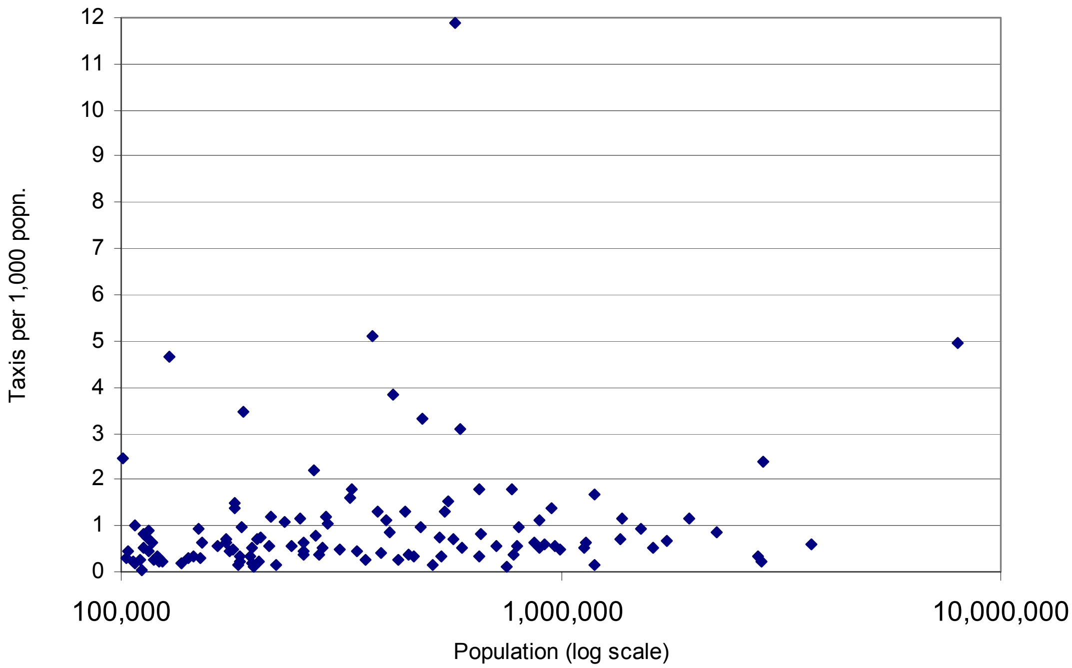
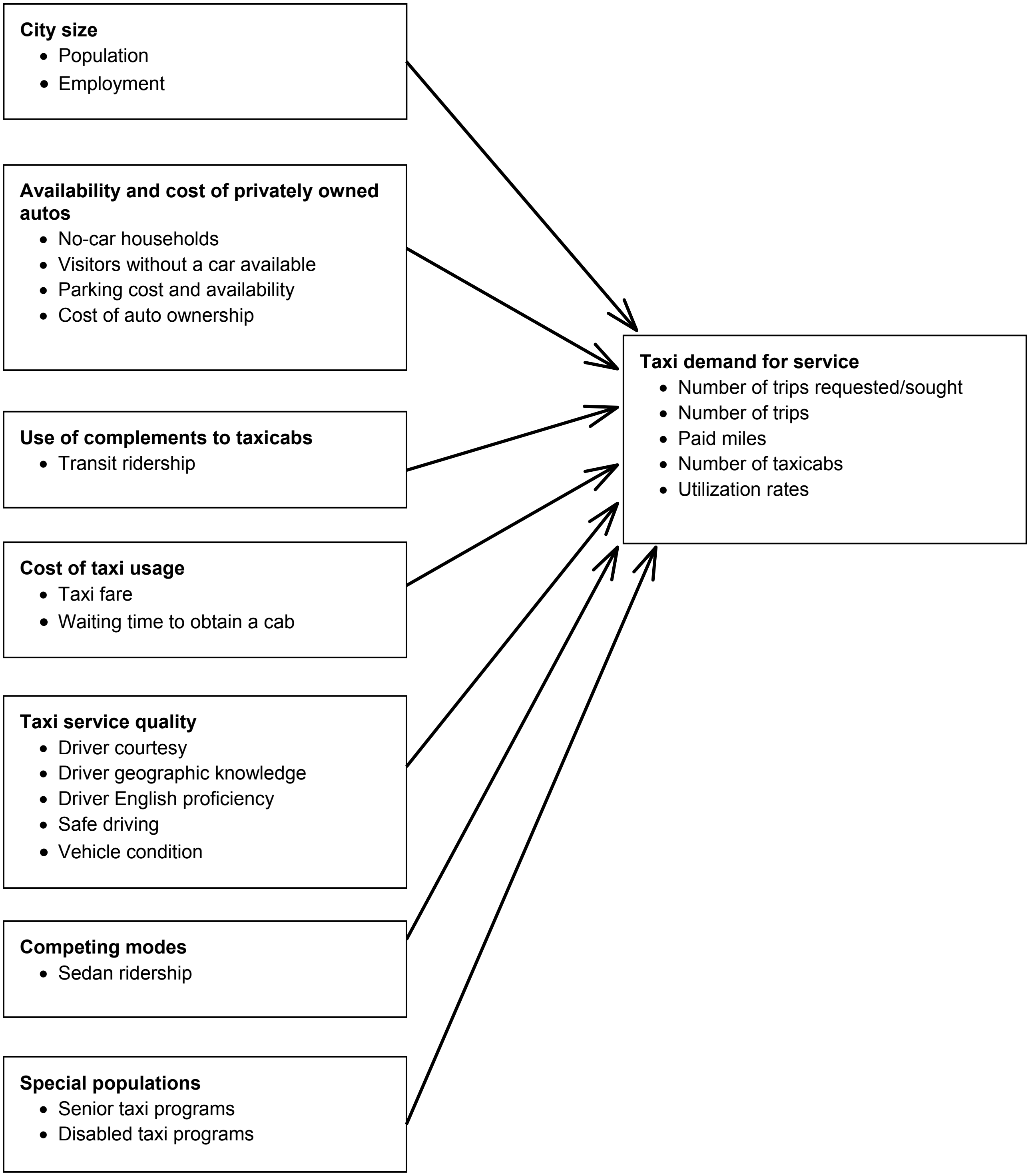


Figure 2. Conceptual model of factors influencing taxi demand



Finally, demand for cabs will be affected by programs for special populations and by competing services. Demand may be elevated by government-subsidized programs for senior citizens to use cabs or for taxi use by disabled persons. Conversely, sedan and limousine services may siphon a portion of the demand for taxicab-type service. This may be particularly important in areas where sedan and limo services are lightly regulated and/or the number of cabs is severely restricted.

DATA

The dependent variable in the model is the number of taxicabs in 118 U.S. cities and counties with 100,000 or more population. The primary data source is the Taxicab, Limousine and Paratransit Association's (TLPA) *2002 Fact Book* (2002), supplemented by newspaper articles and the author's first-hand knowledge. Livery cabs that serve the taxi market are included in the taxi vehicle counts for New York City, Newark, NJ and Phoenix.

Ideally, taxi demand would be measured by service miles, trips or trip requests rather than the number of cabs. Unfortunately, such data are not available for a sample of cities. Interpretation of results should thus bear in mind that the dependent variable (licensed cabs) is subject to government regulation and that taxi vehicle utilization levels vary significantly from one city to the next. The impact of using the number of cabs rather than a more ideal measure of taxi demand will be assessed later in the paper.

Independent variables tested in the model were:

- Population in 2000, based on the 2000 U.S. Census.
- Employment, measured as resident workers in each city. Source: 2000 U.S. Census.
- Vehicle ownership, measured as households with no vehicles available; households with one vehicle available, and aggregate vehicles available to households. Source: 2000 Census.
- Transit use, measured as resident workers commuting by public transportation, by subway, by light rail and by bus. For cities with a net inflow of transit commuters, commuters by place of work was substituted for resident commuters. Source: 2000 Census.
- Airport passenger volumes, measured as air travelers using taxicabs after arriving by air; and air travelers using shuttles or limousines after arriving by air. These variables include trips by both residents of the metropolitan area and visitors. The variable was calculated based on calendar year 2000 air passenger enplanements at U.S. airports (FAA 2004), estimated percentage of non-connecting passengers at each airport (based on various sources), and percentage of air passengers using taxicabs, shuttles and limousines (unpublished data from the 1995 American Travel Survey). A few airports (Reagan Washington National, Dallas/Fort Worth, Raleigh-Durham, Minneapolis and Cincinnati/Northern Kentucky) are served by taxicabs from multiple jurisdictions. Taxi, shuttle and limousine passengers are allocated among the jurisdictions in these cases.
- Taxi fares, calculated as the fare for a 5-mile trip with 5 minutes of waiting time, not including surcharges, based on the rates of fare in TLPA's 2002 Fact Book.

Except for taxi fares, the independent variables are measured in thousands.

These variables are available for the dataset of 118 cities and counties. Other logical independent variables, such as parking cost or availability, waiting time to obtain a taxi, taxi service quality, demand from programs for seniors and disabled persons, and sedan ridership are not available and thus are not tested in the model. It should be noted that the transit commuter variables are likely to capture, at least to a degree, variations in parking cost and availability since transit use is heavily correlated with parking supply and cost (Taylor and Fink 2003). Transit ridership variables thus serve both as a factor directly influencing demand for taxi service and as a proxy for parking cost and availability.

Due to lack of data availability, airport taxi trips is used as an independent variable even though, ideally, measures that capture underlying generators of airport demand for cab service would be used. Given the use of airport taxi trips to predict the total number of taxicabs citywide, airport taxi trips should be viewed as essentially a control variable in the equation, which is focused on explaining non-airport sources of taxicab demand.

The dependent variable (taxicabs) is for 2002 while independent variables are from 2000. This 2-year lag is intentional, recognizing that some jurisdictions issued taxi licenses in the context of the late-90s economic expansion. These issuances reflect attempts to meet demand at the peak of the economic boom in 2000.

MODEL ESTIMATION

As the first step in model development, the correlation between each independent variable and the number of taxicabs was examined. Three variables showed correlations of 0.77 or greater: subway commuters, airport taxi trips and no-vehicle households. A multivariate linear model with these three variables was then developed and tested with other potential variables. In testing, population, workers, one-vehicle households, aggregate vehicles owned by households, airport shuttle/limousine trips, light rail commuters, bus commuters, total transit commuters and taxi fares did not add to the explanatory power of the model.

Taxi fares were not statistically significant even though it is a precept of economic theory that price affects demand. The lack of statistical significance appears to stem from the relatively small variance in taxi fares relative to the other independent variables. It is also possible that the effect of fares on the number of cabs is confounded if, in some cities, low fares induce taxi owners to lobby for restrictions on the number of cabs in order to increase fare revenue on a per cab basis.

Finally, based on testing of the model, a dummy variable for cities with more than 19,000 no-car households was added to the model. The dummy variable is set to zero where the number of no-car households is less than 19,000 and set to one otherwise. The need for the dummy variable may owe to threshold effects. Cities with relatively few no-car households may lack the critical mass of demand for taxi service that is needed for a taxi company to provide reasonably prompt service. The threshold of 19,000 was determined based on examination of error terms when the model was run without the dummy variable. Without the dummy variable, the predicted number of cabs is consistently higher than the actual number in cities with fewer than 19,000 no-car households.

Table 1 presents summary statistics for variables in the model.

TABLE 1. Summary statistics

Including NYC (n=118)	Mean	Std. Deviation	Minimum	Maximum
Taxis	873	3,728	6	39,600
Subway commuters (thousands)	15.6	111.7	0.0	1,199.2
Airport taxi trips (thousands)	350.6	844.8	0.0	7,150.7
No-vehicle households (thousands)	43.2	158.0	1.1	1,682.9
Dummy for no-vehicle HH less than 19,000	0.4	0.5	0	1
Excluding NYC (n=117)	Mean	Std. Deviation	Minimum	Maximum
Taxis	542	988	6	6,900
Subway commuters (thousands)	5.5	20.4	0.0	131.3
Airport taxi trips (thousands)	292.4	563.8	0.0	3,078.8
No-vehicle households (thousands)	29.2	42.5	1.1	306.3
Dummy for no-vehicle HH less than 19,000	0.4	0.5	0	1

The final model is:

Number of taxicabs = f (no-car households; subway commuters; airport taxi trips; dummy)

Table 2 reports the results of the linear model for the entire dataset of 118 U.S. cities and counties. The adjusted R^2 for the model is 0.989, indicating that the model explains 98.9 percent of the variance from the mean. The F-statistic of 2,568.9 is also quite high.

Subway commuters, no-car households and airport taxi trips are statistically significant at the 95 percent confidence level. There is no indication of multicollinearity between independent variables, based on V.I.F. scores.

The coefficient for no-car households is 5.1, indicating that a change of 1,000 no-car households is associated with a change of 5 taxicabs in the observed cities and counties, other factors being held constant. The coefficient for subway commuters is 21.8, indicating that a change of 1,000 subway commuters is associated with a change of 22 taxicabs. For airport taxi trips the coefficient is 0.64, indicating that each 1,000 annual airport taxi trips accounts for 0.64 taxicabs.

These results indicate that an increment of 1,000 subway commuters is associated with four times more additional taxicabs as compared with an increment of 1,000 no-car households. The subway commuter variable is most likely playing a strong proxy role for parking costs and availability and, more generally, the degree of density and urbanization of cities with large subway systems, as well as direct demand from subway commuters' use of cabs.

The airport taxi trip coefficient implies a ratio of one cab per 1,562 airport taxi trips annually. Assuming 310 days of operation per year (85 percent utilization rate), 1,562 trips averages to 5 trips per day per cab. This figure is on the low end of the range of 5 to 8 airport trips per cab typically experienced. As expected, the airport taxi trip variable is reflecting not simply demand from airport-originating passengers but also demand for trips to the airport and trips around town during nonresidents' stay in the locale.

The dummy variable for jurisdictions with more than 19,000 no-car households is not statistically significant but improves the predicted values for cities with fewer than 19,000 no-car households and slightly improves the R^2 and thus is retained in the model.

The dataset includes one extreme value, New York City, which has 5 to 9 times as many taxicabs, no-car households and subway commuters as any other city in the dataset. There is thus a need to assess the impact of New York on the model.

TABLE 2. Estimation results – including New York City

Variable	Coefficient	St. error	t-statistic
Subway commuters	21.81	1.72	12.7
No-vehicle households	5.14	1.29	4.0
Airport taxi trips	0.64	0.09	7.2
Dummy for no-vehicle HH greater than 19,000	129.95	92.02	1.4
Constant	31.42	48.84	0.6
Observations	118		
Adj. R^2	0.989		
F statistic	2,568.9		

Table 3 reports the results of the model with New York City excluded. Coefficients in the non-NYC model are quite similar to those in the model with NYC included. The coefficient for subway commuters is almost identical, a rather remarkable outcome given New York's influence on the model when it is included. Coefficients for no-car households are slightly lower and for airport taxi trips are slightly higher with New York excluded. The R^2 drops to 0.84 and the F-statistic to 153.4, but these are still quite high considering that the number of cabs is an inexact proxy for taxi demand, as discussed earlier.

The model described here utilizes each variable without any transformations. As a check on the form of the equation, the model was run using two transformations. Transforming each variable (except the dummy variable) to logs produced results in which each variable is statistically significant with the expected sign, but with a somewhat lower R^2. A nonparametric regression using each city ranked from highest value to lowest for each variable also produced statistically significant coefficients for each variable and the expected signs, with about the same R^2 as for the model that excludes New York City.

EVALUATING MODEL RESULTS

How well does the model predict the number of cabs in various cities? Does the use of the number of taxicabs as the dependent variable, subject to local regulation and variations in utilization rates, bias the results?

Inspection of predicted and actual values for cities in the dataset suggests that the model works quite well. The predicted number of cabs reasonably closely matches the actual number in cities that based on separate information appear to have an appropriate number of cabs. These include San Diego, Los Angeles, St. Louis and several smaller cities or counties. Differences between predicted and actual number of cabs is within 12 percent in these jurisdictions, differences that could easily stem from differing vehicle utilization rates.

Notably, the model is quite good in predicting the number of taxicabs in jurisdictions that do not regulate the number of taxicabs (so-called "open entry" cities). These include Orange County, CA, Phoenix, Newark and New York City (the latter three cities including open-entry livery vehicles in the vehicle count). Thus, regulatory limits on the number of cabs do not appear to bias the coefficients of the independent variables.

The model also predicts substantially fewer cabs than are actually licensed in Washington DC, Dallas and Houston, cities in which there is reason to believe that an oversupply of cabs exists. Conversely, the model predicts a substantially larger number of cabs in San Francisco, Boston, and Montgomery County, MD, jurisdictions that have historically limited the number of cabs below market demand.

TABLE 3. Estimation results – cities excluding New York City

Variable	Coefficient	St. error	t-statistic
Subway commuters	19.72	2.47	8.0
No-vehicle households	4.47	1.41	3.2
Airport taxi trips	0.70	0.11	6.7
Dummy for no-vehicle HH greater than 19,000	145.79	92.85	1.6
Constant	36.10	48.92	0.7
Observations	117		
Adj. R^2	0.840		
F statistic	153.4		

DEVELOPING LOCAL MODELS FOR TRACKING TAXI DEMAND

One application of model results is to provide guidance for the development of a taxi demand model in specific locales, which regulators can use to assess changes in demand over time.

The model of U.S. cities indicates that the following variables should be included in the development of local time-series models. The variables used in a given locality will depend on data availability and local conditions. Several alternative measures are suggested for each conceptual variable.

- Households or residents without a car available. The cross-sectional model of U.S. cities uses the number of no-car households from U.S. Census data. Localities may not have this statistic available on an annual or monthly basis, as would be desirable for time-series modeling. Alternatives would be auto registrations or the ratio of automobile registrations to population. Another alternative would be bus ridership, which in the cross-sectional data is highly correlated with no-car households.
- Subway commuters. Local data on subway ridership is generally available at the city or county level. The cross-sectional model uses subway commuters, but total subway ridership may be an equal or better substitute. Analysis of ridership on weekdays versus weekends would help to distinguish the relevance of work versus non-work trips.
- Airport taxi trips. Airports sometimes keep exact counts of the number of taxi trips dispatched from their on-demand taxi lines. If that is not available, the number of airport enplanements is available for all U.S. airports. Care has to be taken, however, if there are changes in the percentage of connecting passengers or in taxis' share of the airport ground transportation market.
- Taxi fare for an average trip, adjusted for inflation.

Other variables that would be potentially valuable in the model are:

- Number of visitors, convention delegates or hotel room nights occupied in downtown hotels.
- Demand generated by programs for seniors or disabled persons.
- The ratio of parking spaces to downtown employment.
- Response times for taxi service.
- Average age of taxicabs in service.
- Number of cold-weather days, for northern climates.

Development of a model also requires identification of a variable for taxi demand. Measures of demand to consider are the number of calls received by cab companies requesting service, the number of taxicab pickups at cab stands, and taxi utilization indicators (e.g., paid miles or percent paid miles). Where the number of cabs is constrained by regulatory caps, vehicle utilization rates can capture changes in demand, as illustrated for medallion taxis in New York City (Schaller 1999).

BENCHMARKING WITH OTHER CITIES

The model is also a useful benchmarking tool for cross-city comparisons. These comparisons should not be the only basis for evaluating demand in a given city. One should not expect comparisons to suggest an exact number of cabs in a given locale. With these caveats, benchmarking can provide useful perspective on local industry size.

An example illustrates the use of model results to identify peer jurisdictions for detailed comparisons. The number of taxicabs in Montgomery County, Maryland, has been restricted to 580 cabs since the early 1990s. Analysis of public complaints and of cab company computerized dispatch data concluded that demand has been depressed due to unreliability of pickups and excessively long response times. (Schaller 2002)

The dataset of 118 jurisdictions identified three suburban jurisdictions for comparison: Fairfax County, VA; Prince Georges County, MD and Cambridge, MA. These jurisdictions are fairly densely developed suburbs with a substantial number of residents commuting by subway but without an airport. Due to its much smaller land area and the presence of two major universities, however, Cambridge was not felt to be a good comparison with Montgomery County, leaving Fairfax and Prince Georges counties for comparison.

Inspection of the independent variables shows that Montgomery County has the largest number of subway commuters of the comparison counties and is second to Prince Georges County in the number of no-car households. On this basis, one would expect Montgomery County to have substantially more cabs than Fairfax County and somewhat more cabs than Prince Georges County. The fact that Montgomery County has fewer cabs than either Fairfax or Prince Georges counties thus supported the other evidence that demand for cab service in Montgomery County was depressed by service quality problems.

CONCLUSIONS

The model presented in this paper identifies factors that explain most of the variation in the number of taxicabs among 118 U.S. cities and counties. Three strong factors were identified:

(1) The number of workers commuting by subway, which is both a direct generator of demand for cab service and also a proxy for parking costs and availability and overall urban density, factors that are not separately accounted for in the model.

(2) The number of no-car households.

(3) Taxi usage for airport taxi trips, which are themselves a direct measure of demand for service, and also captures demand for trips to return to the airport and local taxi trips by visitors.

Each of these independent variables measure the number of people not using privately owned vehicles. Notably, two oft-mentioned variables – population and employment – did not prove to be significant factors after subway commutation, no-car households and airport taxi trips are taken into account.

Results from the model show, for the first time, determinants of taxi demand for a broad cross-section of U.S. cities. Results are useful at the local level to identify variables for a time-series model of local taxi demand that can form a valuable analytic basis for assessing changes in demand for service. Results are also useful for identifying peer cities for further comparison and analysis.

REFERENCES

Beesley, M.E. Competition and Supply in London Taxis. 1979. *Journal of Transport Economics and Policy*. Vol. 13, No. 1, pp. 102-131.

Economic Planning Group. 1998. *Financial Analysis for Report to Review the Toronto Taxi Industry*.

Federal Aviation Administration. 2004. CY 2000 Enplanement Activity at U.S. Airports by State. http://www.faa.gov/arp/Planning/v2.htm. Accessed May 10, 2004.

Fravel, Frederic and Gorman Gilbert. 1978. Fare Elasticities for Exclusive-Ride Taxi Services. Publication UMTA-NC-11-0006, U.S. Department of Transportation.

Gilbert, G., T.J. Cook, A. Nalevanko and L. Everett-Lee. 2002. *The Role of Private-for-Hire Vehicle Industry in Public Transit.* Transit Cooperative Research Program Report 75.

Hara Associates. 1994. *City of Halifax Taxi License Limitation Study.*

Schaller Consulting. 2002. *Montgomery County Taxi Industry Study.*

Schaller, Bruce. Elasticities for Taxicab Fares and Service Availability. 1999. *Transportation*. Vol. 26, pp. 283-297.

Taxicab, Limousine and Paratransit Association. 2002. *2002 Taxicab Fact Book.*

Taylor, Brian D. and Camille Fink. 2003. The Factors Influencing Transit Ridership: An Analysis of the Literature, *Working Paper, UCLA Institute of Transportation Studies, UCLA.*

BIOGRAPHICAL SKETCH

Bruce Schaller (schaller@schallerconsult.com) is Principal of Schaller Consulting. He has consulted on taxi and transit issues for local governments, transit agencies, airports, the federal government, and for-profit, non-profit and advocacy organizations in the U.S., Canada and Moscow, Russia. He is a nationally recognized expert in taxicab regulatory issues and also specializes in providing market research to improve transit services and attract transit users. He has written extensively in both areas, with articles published in *Transportation, Transportation Quarterly, Transportation Research Record,* and the *New York Transportation Journal.*

Prior to establishing his consulting practice in 1998, Mr. Schaller was Director of Policy Development and Evaluation at the New York City Taxi and Limousine Commission and Deputy Director of Marketing Research and Analysis at New York City Transit. He has a BA from Oberlin College and Masters of Public Policy from the University of California at Berkeley.