REVIEW OF LADOT'S TAXICAB SERVICE LEVEL EVALUATION PROCEDURE

JAN DE LEEUW

Contents

1. Introduction

UCLA Statistical Consulting has been asked by taxicab companies in the city of Los Angeles and the city's Department of Transportation (DOT) to review DOT's procedure for estimating the proportion of calls to taxicab companies which are responded to within fifteen minutes. These estimated proportions are used to calculate tax payments to the city by cab companies. The taxi companies believe that the statistical sampling procedure used by the city is not sound; in particular that sample sizes are too small. Both sides

Date: 1993.

have agreed to seek UCLA's assessment of current procedures and recommendations for improvement.

DOT detailed its procedure for calculating sample size and arriving at an estimate of the proportion of calls responded to within fifteen minutes in *City of Los Angeles, Taxicab Service Level Evaluation Procedures; Minimum Sample Size Determination*, and Section 800 of Board Order number 471 describes how this estimate is used to determine cab company tax liability. We have reviewed these documents, and met several times with representatives of both the taxi companies and DOT. This report presents our understanding, analysis, and conclusions to date.

Section 2 presents some background information necessary to understand overall objectives of DOT's procedure, and aspects of taxicab operation in Los Angeles that impact it. Section 3 describes the test, and how taxes are computed. Section 4 presents our current assessment of the statistical theory behind the calculations. In Section 5, we discuss our tentative conclusions and recommendations. Finally, Appendix A discusses relationships between sample size, confidence level, and confidence interval width. Appendix B has some of the other computations our conclusions are based on.

2. Background

Eight taxi companies are licensed to operate in the city of Los Angeles. They are Valley Cab, Independent Taxi Owners Association (ITOA), United Independent Taxi Drivers (UITD), Bell Cab, Beverly Hills Cab, LA Taxi, LA Checker Cab, and United Checker Cab. LA Taxi and United Checker Cab are actually the same company; together with Long Beach Yellow Cab and South Bay Yellow Cab, they comprise Taxi Systems, Inc. Each cab company is authorized to operate in one or more of the eight zones (service areas) into which Los Angeles is divided. A license to operate permits a cab to pick up passengers in the zone in which it is licensed, and deliver those passengers anywhere. Cabs may not pick up passengers in service areas other than those in which they are licensed, and may have to make return trips empty. Outside the city of Los Angeles, taxis are subject to regulations of municipalities in which they travel. Boundaries of the eight zones, companies authorized to operate within them, and numbers of cabs licensed are show on the attached map (Taxicab Service Areas).

Operating arrangements of taxicab companies are varied. Taxi Systems and Valley Cab own, maintain, and insure all their vehicles. For a flat monthly fee, they lease cabs to drivers who keep all revenues less the monthly fee. These companies also provide dispatch services, and pay the DOT tax. At the other end of the spectrum are the associations: UITD and ITOA. These are taxicab owners who pool resources to share operational support. Members pay association dues which go toward dispatchers, other operating expenses, and taxes. Owners either drive their own cabs, or lease them out for a fee. In addition, the association itself may own and lease out some cabs. The remaining cab companies have operating arrangements which fall between these two extremes.

It is worth noting that under these arrangements, promptness of response time is in the hands of the drivers rather than the cab companies. Owner-drivers have the strongest incentive to be prompt since their association dues go directly towards paying city taxes. Lessors and hired drivers have little incentive, since they personally do not pay the tax. Cab companies will apply pressure on drivers by threatening to discontinue lease agreements, or with other penalties. Cab companies should be responsible for the performance of their drivers, and should perhaps consider ways to apply more direct economic pressure.

The Department of Transportation is the city's agency responsible for regulating taxicab operations. The purpose of regulation is to restrict the total number of cabs on the street in an effort to alleviate Los Angeles' traffic congestion problems. Restricting the number of cabs will create a natural monopoly for those operators receiving licenses, and thereby undermine competitive incentives for operators to serve the public efficiently. That is, without the threat of losing customers to new cab companies in their zones, currently licensed operators have little reason to respond promptly to service calls. DOT instituted its tax and testing procedures to create an incentive for these natural monopolists to provide good service. Such an incentive is thought to be more effective than merely threatening to revoke licenses.

The geographic configuration of the eight zones, and numbers of cabs allocated to each were determined in such a way that maximum response time to any point within a zone, at any time of day, is fifteen minutes. How this was done, and whether the DOT checks that it remains true, is not clear. It is, however, a crucial assumption since DOT does not otherwise compensate for factors causing response times to differ over times of day or locations. In other words, DOT does not wish to set different standards for response times for different times of day or areas, as they would have to do if zones were drawn arbitrarily, and cabs allocated uniformly. DOT agrees, one would hardly expect response time to be the same in Reseda at 2:00 on Saturday afternoon and at 5:00 pm on Friday in downtown LA. One standard of fifteen minutes allows DOT to pool information into one test per taxi company, rather than having to conduct as many small tests for each company as there were different standards.

DOT's regulatory tax is a monthly payment due the city of Los Angeles by the taxi companies. It is calculated as a base fee multiplied by a service factor. The service factor is one plus the sum of two tax increments: one related to telephone answering speed, and the other related to taxicab response times. If these service levels are satisfactory, the monthly fee is not affected by the service factor, and only the base tax is assessed (i.e. service factor = 1.0). These tax increment tables are listed under Board Order Number 471, Section 804.

Telephone answering time is rarely a problem for taxi companies; they seem to do a good job taking orders within 45 seconds. Taxes are usually determined by Schedule II, which gives tax increments for cab response time. If less than 70 percent of the orders are responded to within fifteen minutes, the Schedule II component of the service factor is 94 percent. If more than 76 percent of calls are responded to within 15 minutes, there is no Schedule II contribution. Between 70 and 76 percent the Schedule II contribution changes either 13 or 14 percent for each one percent difference in service level.

We summarize the procedure in the following table. DOT rounds down, i.e. a measured service level of 75.99% and a measure service level of 75.01% are both rounded to 75%.

measured service-level	tax-level
(.75-1.00]	100%
(.74-.75]	114%
(.73-.74]	128%
(.72-.73]	142%
(.71-.72]	155%
(.70-.71]	168%
(.69-.70]	181%
(.00-.69]	194%

DOT determined that 70 and 76 percent were appropriate bounds for this tax structure by conducting a survey of public opinion about taxi service. The tax structure between 70 and 76 percent seems to have been arrived at largely arbitrarily. Taxes go up quite sharply in this range. This means that even small variations in service level can have important consequences for the amount of tax that is paid. Variability in the estimate of service level can be the result of relatively small sample sizes used by DOT. Therefore, calculation of minimum required sample size deserves special scrutiny. We take this up in Section IV. DOT's testing procedure is described in the next section.

3. DOT's Testing Procedure

At least once a year DOT conducts tests to determine service levels of taxi companies. Results of those tests determine operators monthly tax liabilities until the next set of tests are conducted. Service level has two components: telephone answering time (percentage of calls answered within 45 seconds), and cab response time (percentage of requests for service met within fifteen minutes). DOT inspectors conduct tests by placing telephone orders for taxi service at prespecified locations and times, and recording response times.

To calculate an appropriate sample size, DOT follows guidelines that incorporate some important assumptions. DOT wants a large enough sample to produce a 95 percent level of confidence, and ten percent maximum error in estimating mean response time. The equation used to calculate sample size incorporates mean response time and standard deviation of response time as estimated by the previous test. Since mean response time from the current test will be used in a future test, sample size should be sufficiently large to ensure the accuracy of this estimate as well as that of the proportion of calls responded to within the prescribed time. (No mention is made in the DOT report of the necessity to have an accurate estimate of the standard deviation of response times.) Note that minimum sample sizes can vary among companies and from one test period to another.

Once required sample size is known, the next step is for DOT to decide on locations and times for its test calls. First, records of all calls (dispatch slips or computer printouts) for a given month are collected from each taxi company. At least 85 percent of test calls are randomly selected from these records. The remainder are randomly selected from complaint logs and hard to serve areas. With the record of calls in chronological order, the 400th call will provide the location and time of the first test call. From there, every 200th record will be selected to provide test call locations and times, until the required sample size is reached. When stacks of dispatch slips are used, DOT uses digital calipers to approximate the location within the stack, of every 200th record. If a test call location and time are ineligible, a new record is selected by going backwards through the stack from the ineligible record, until an eligible record is found. When this sampling is complete, daily assignment sheets listing the taxicab company and exact test locations and times are provided to inspectors.

Tests are made at the same times of day (to within 30 minutes), and on the same days of the week as those given by the original records. However, tests are not conducted on holidays, rainy days (or even days with a threat of rain), or under other abnormal circumstances such as earthquakes or when a major accident occur in the vicinity of the test location. If a test can't be conducted within 30 minutes of its prescribed time, it is postponed to a later date (but still made on the same day of the week and at the same time of day). If an exact address cannot be found, an address within 200 feet is used.

Tests are performed by DOT inspectors on a rotating basis so taxicab telephone answerers don't learn to recognize test calls. When test calls are placed, telephone answering time is measured from the time of the first ring until an order taker answers. Being placed on hold or getting a recording or an answering machine doesn't count as an answer. If the order is taken and service is promised, the tester waits at least 30 minutes for the cab to arrive, and records the response time. A tester may record "no show" in the following situations: 1) an order is taken, but service is refused, 2) service is promised but the cab doesn't arrive within 30 minutes, or 3) promised service time is greater than 30 minutes. In this last case the tester still waits 30 minutes just in case the cab arrives early. Upon completion of a test call, the tester goes on to the next address on the daily assignment sheet. This process repeats until all test calls on the assignment sheet are placed.

For each taxi company, DOT calculates the percentage of test calls answered in 45 seconds and the percentage of orders responded to in 15 minutes. These percentages are always rounded up to the next integral value. The tax increments are determined using Schedules I and II. The service factor results from summing the two increments. Finally, the monthly tax for each company is computed by multiplying the base fee by the service factor.

4. Statistical Analysis

The procedure DOT used to estimate service levels for taxi companies in Los Angeles requires sampling, inference, and decision making. Accordingly, there are three main steps in this analysis: calculating sample size, drawing the sample and conducting the tests, and making an inference about the true proportion of calls responded to within 15 minutes. After reviewing DOT's methodology we tentatively conclude that from a statistical point of view, there are unsatisfactory aspects of the sample size calculation and the inference procedure.

First, the sample size calculation seems unconnected to the decision rule. DOT's method (as described on pages five through seven of *City of Los Angeles, Taxicab Service Level Evaluation Procedures; Minimum Sample Size Determination*) mandates a sample size large enough to ensure that the estimate of mean response time is made with 95 percent confidence, and less than ten percent relative error. See Appendix A for a formal description of the procedure. It appears that after conducting a test with this sample size, DOT computes the proportion of calls responded to within 15 minutes, not the mean response time. In other words, the sample size calculation is for a different statistic than the one being used here for tax determination. Specifying a ten percent maximum error for mean response time bears no direct relation to maximum error for the proportion of calls responded to within 15 minutes. In fact, we could calculate the minimum sample size required to guarantee a maximum error of ten percent in the proportion directly from the usual formula for sample size, rather than via the complicated route used by DOT. This procedure is also outlined in Appendix A.

Second, the tax structure is much too sensitive for the accuracy of the estimated proportion using the sample size determination described above. This is acknowledged on page 28 of the DOT report, (*City of Los Angeles, Taxicab Service Level Evaluation Procedures; Minimum Sample Size Determination*). On the basis of the current number of calls (per period/per company) the true service level can only be determined very imprecisely. The sampling variability completely overwhelms the level of accuracy the tax was designed for. This second problem is illustrated in the calculations in Appendices A and B.

Finally, although we believe that the sampling procedure is essentially sound, we make two additional observations. First, there is a possibility that choosing every 200th dispatch record from a chronologically ordered set may introduce a time trend. For instance, if calls to taxi companies are cyclical over a period corresponding to 200 calls, then choosing every 200th call will cause all tests to be made at the same time of day and day of the week. Second, since proportion of calls responded to within 15 minutes is used to calculate tax obligation, there seems to be no need for DOT inspectors to wait 30 minutes for cabs to arrive. 15 minutes is long enough to determine whether a particular call was responded to in 15 minutes or less. Perhaps DOT could save some resources by allowing its inspectors to move on to the next test call after 15 minutes.

5. Conclusions and Recommendations

We have two main conclusions from our review of DOT's test procedure. First, DOT's assumption that maximum response time anywhere in Los Angeles is 15 minutes needs to be tested. If the existing zones and numbers of cabs allocated to them are such that maximum response times vary across parts of the city and times of day, then the assumption is obviously not justified. We should ask what is meant by maximum response time in this context. Maximum response time probably means maximum reasonable response time under ordinary conditions. If DOT could specify what ordinary conditions are, and give a condition for reasonable response time such as percentiles for responses under 15 or 20 minutes, we could compare estimates of the distribution functions for response times in the eight zones, and try to determine whether they are similar. If not, zones should be redrawn or cabs reallocated, to ensure that this assumption is justified. The entire analysis depends on this.

Second, the tax structure is too sensitive for the statistic upon which it is based. The variability in the estimate $\hat{p}$ of the true proportion π overwhelms the size of a tax bracket. The tax structure is a result of legislative action, and it is unclear whether or not it is changeable. The size of a tax bracket should be related to the accuracy of the statistic as well as the desired incentive effect.

At present, we have identified two alternatives to the present DOT calculation. Both require modifications to the existing tax structure. One is to base the tax on the proportion of calls responded to within 30 minutes rather than 15 minutes, and raise the proportion of successes required. Switching to proportion of responses within 30 minutes will reduce the standard error of the estimate from about three percent to about two percent. Since the standard error is a function of $\pi(1-\pi)$, cab companies with a high level of performance (high values of π) will require smaller sample sizes to accurately estimate their true performance. Sample size calculations could be made using the usual formula with π being estimated from previous tests. For example, for a company performing at a true $\pi = 0.99$, a 90 percent confidence interval of width plus or minus one percent could be obtained with a sample of 268 (see Appendix A, Table 2). This still requires that tax brackets be moved up and widened.

Another alternative is to base the tax on mean response time instead of proportion of successes. If the distribution of response times was normal, we know that mean response time could be accurately estimated with a relatively small sample size. In fact, we believe response times are not normal, but speculate that they may follow an Exponential, Log-normal or Gamma distribution. With additional analysis performed on the eight month Taxi Systems database, we may be able to determine the distribution of response times, and calculate the standard error of mean response time. This would imply an appropriate tax bracket width.

One difficulty with this second approach is that "no shows" would have to be counted as 30 minute responses (or assigned some other arbitrary value). This would lead to an underestimate of mean response time. A possible solution to this problem is not to include "no shows" in the calculation of mean response time, but rather to base the tax on a matrix of mean response times and the proportion of "no shows" in the sample. How to arrange such a matrix, and particulars of the tax structure will depend on the results of future analysis and discussions with DOT.

Appendix A. Computing Confidence Intervals

Suppose $\underline{x}_n$ is an asymptoticaly normal sequence of random variables. More precisely, we suppose there exist θ and ω^2 such that

$$n^{\frac{1}{2}}(\underline{x}_n - \theta) \stackrel{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, \omega^2).$$

This means that for each $\delta > 0$

$$\lim_{n\to\infty} \mathbf{prob}[-\delta \leq n^{\frac{1}{2}}(\underline{x}_n - \theta) \leq +\delta] = \Phi(\frac{\delta}{\omega}) - \Phi(-\frac{\delta}{\omega}) = 2\Phi(\frac{\delta}{\omega}) - 1.$$

To find an $(1-\alpha)$% confidence interval we compute δ in such a way that

$$\lim_{n\to\infty} \mathbf{prob}[-\delta \leq n^{\frac{1}{2}}(\underline{x}_n - \theta) \leq +\delta] = 1 - \alpha.$$

The solution is $\hat{\delta} = \omega\tau$, with $\tau = \Phi^{-1}(1 - \frac{1}{2}\alpha)$. Thus an $(1-\alpha)$% confidence interval for θ is $[\underline{x}_n - n^{-\frac{1}{2}}\hat{\delta}, \underline{x}_n + n^{-\frac{1}{2}}\hat{\delta}]$. The width of the interval is obviously $2n^{-\frac{1}{2}}\hat{\delta}$. Now suppose we want this width to be equal to 2ϵ. This means that

$$\hat{n} = \omega^2\frac{\tau^2}{\epsilon^2}.$$

DOT applies this formula to the mean μ, which then has $\omega^2 = \sigma^2$, the population variance. But, instead of fixing the absolute error, they want a width equal to $\epsilon\mu$, i.e. they fix the relative error. This leads to the formula

$$\hat{n} = (\frac{\sigma^2}{\mu^2})(\frac{\tau^2}{\epsilon^2}).$$

If we apply the formula to the proportion of successes π, for which $\omega^2 = \pi(1-\pi)$. For absolute errors the formula is

$$\hat{n} = \pi(1-\pi)\frac{\tau^2}{\epsilon^2},$$

and for relative errors it becomes

$$\hat{n} = \left(\frac{1-\pi}{\pi}\right)\frac{\tau^2}{\epsilon^2}.$$

Of course in actual applications of this formula we have to substitute estimates for μ, π, and σ^2.

Here we are interested in 95% confidence intervals for the probability of succes. We want to be able to determine this proportion accurately, and thus we need a confidence interval with a width of 1%. We have to choose $\alpha = .05$, and $\epsilon = .005$. For a true success rate of π this means that we need sample size $\hat{n} = \pi(1-\pi) \times 153,664$. This means that for true service level of 76% about 28,000 calls are required for the necessary stability. Even a true service level of 90% still requires a sample size of 13,000. This is clearly infeasible. If we go to the much more modest 90% level and a width of .02, we find $\hat{n} = \pi(1-\pi) \times 26,896$. This still means that for true service level of 76% DOT has to make 4900 calls. For 90% this would be 2400 calls, and for 99% 266 calls. Still infeasible. A more complete overview of the various combinations of α and ϵ is given in Tables 1 and 2.

In order to establish what can be done with sample sizes of $n = 100$ we rewrite our basic result as

$$\frac{\tau}{\epsilon} = \sqrt{\frac{n}{\pi(1-\pi)}}.$$

For true service level of 76% and a width of 1% this means that we can have only 10% confidence that the true value is in the interval. If we fix confidence at 95%, then we come up with an interval of about 17% width, which is of course quite useless.

TABLE 1. Sample size as a function of degree of confidence and width

true	conf=.95 width=.01	conf=.90 width=.01	conf=.95 width=.02	conf=.90 width=.02
0.50	38415	27055	9604	6764
0.51	38399	27045	9600	6761
0.52	38353	27012	9588	6753
0.53	38276	26958	9569	6740
0.54	38169	26882	9542	6721
0.55	38030	26785	9508	6696
0.56	37861	26666	9465	6666
0.57	37662	26525	9415	6631
0.58	37431	26363	9358	6591
0.59	37170	26179	9292	6545
0.60	36878	25973	9220	6493
0.61	36555	25746	9139	6436
0.62	36202	25497	9050	6374
0.63	35818	25226	8954	6307
0.64	35403	24934	8851	6234
0.65	34957	24620	8739	6155
0.66	34481	24285	8620	6071
0.67	33974	23928	8493	5982
0.68	33436	23549	8359	5887
0.69	32868	23149	8217	5787
0.70	32268	22727	8067	5682
0.71	31638	22283	7910	5571
0.72	30978	21818	7744	5454
0.73	30286	21331	7572	5333
0.74	29564	20822	7391	5205
0.75	28811	20292	7203	5073

TABLE 2. Sample size as a function of degree of confidence and width

true	conf=.95 width=.01	conf=.90 width=.01	conf=.95 width=.02	conf=.90 width=.02
0.76	28027	19740	7007	4935
0.77	27213	19166	6803	4792
0.78	26368	18571	6592	4643
0.79	25492	17954	6373	4488
0.80	24585	17315	6146	4329
0.81	23648	16655	5912	4164
0.82	22680	15974	5670	3993
0.83	21681	15270	5420	3818
0.84	20652	14545	5163	3636
0.85	19591	13798	4898	3450
0.86	18500	13030	4625	3257
0.87	17379	12240	4345	3060
0.88	16226	11428	4057	2857
0.89	15043	10595	3761	2649
0.90	13829	9740	3457	2435
0.91	12585	8863	3146	2216
0.92	11309	7965	2827	1991
0.93	10003	7045	2501	1761
0.94	8666	6104	2167	1526
0.95	7299	5141	1825	1285
0.96	5900	4156	1475	1039
0.97	4471	3149	1118	787
0.98	3012	2121	753	530
0.99	1521	1071	380	268

Appendix B. Stability Calculations

We now investigate the question of stability in more detail. Suppose the true service level of a company is π. The observed proportion $\hat{p}_n$ of successes in n trials (calls) will be approximately normally distributed. More precisely, the quantity

$$\underline{z}_n = n^{1/2}\frac{(\hat{p}_n - \pi)}{\sqrt{\pi(1-\pi)}}$$

has an asymptotic normal distribution with mean zero and variance one.

Thus we can approximate the probability of tax levels. For instance, the probability that a company with a true service level of π pays 114% tax, if we make n calls, is prob$(.74 < \hat{p}_n < .75)$. But

$$\hat{p}_n = \pi + n^{-1/2}\omega z_n,$$

with $\omega = \sqrt{\pi(1-\pi)}$. Thus the probability can also be writtens as

$$\mathbf{prob}[n^{1/2}\frac{.74-\pi}{\omega} < \underline{z}_n < n^{1/2}\frac{.75-\pi}{\omega}].$$

If $n = 100$ and $\pi = .80$, for instance, we see that we have to compute $\mathbf{prob}(-1.5 < \underline{z}_n < -1.25)$, which is approximately .933 - .894 = .039 (from the usual tables of the standard normal distribution). Thus we see that a company which has true service level of 80%, well within the safe service range, still has a probability of about 4% to get a 14% tax penalty if only a 100 calls are made. For borderline cases, such as service level 76%, we have about a 10% probability of a 14% tax penalty. Clearly, by symmetry, such a company actually has a 50% probability of a tax penalty, irrespective of size. Table 3 gives the tax level probability estimates for $n = 100$, and Table 4 does the same for $n = 1000$.

The numbers in Tables 3 and 4 can be summarized by computing expected values. We simply take the weighted sum of the tax levels, where the weights are given by the probabilities in Tables 3 and 4. Thus we compute the average (estimated) tax paid by a company with service level π if n calls are made. For true servicelevel between .50 and .75 the expected values are given in Table 5, for true level between .75 and 1.00 in Table 6.

If we look at the column with 100 calls, we see that a borderline company (true level 76%) pays on the average 119% tax. For a company with 100 cars this is, per year, $100 \times \$65 \times 12 \times .19 = \$14,820$ more than the company would have paid if the service level had been accurately determined. Or, to formulate it differently, $100 \times \$65 \times 12 \times (.19 - .04) = \$11,700$ more than the company would have paid if 1000 calls had been made.

We also emphasize that companies with relatively poor performace (true service level between .60 and .74) pay less tax, in some case up to 20% less tax, than they are supposed to pay. This is because of the asymmetry in the tax structure, in which companies are punished if they do not perform well, but not rewarded if they perform well.

TABLE 3. Tax bracket probability as a function of true service level (n=100)

true	100%	114%	128%	142%	155%	168%	181%	194%
0.650	0.018	0.012	0.017	0.024	0.033	0.043	0.054	0.799
0.660	0.028	0.017	0.024	0.033	0.043	0.054	0.064	0.737
0.670	0.045	0.024	0.033	0.043	0.054	0.064	0.074	0.665
0.680	0.067	0.032	0.043	0.054	0.064	0.074	0.081	0.585
0.690	0.097	0.043	0.054	0.065	0.074	0.082	0.086	0.500
0.700	0.137	0.054	0.065	0.075	0.082	0.086	0.086	0.414
0.710	0.189	0.065	0.075	0.083	0.087	0.087	0.083	0.330
0.720	0.252	0.076	0.084	0.088	0.088	0.084	0.076	0.252
0.730	0.327	0.085	0.089	0.089	0.085	0.077	0.066	0.184
0.740	0.410	0.090	0.090	0.086	0.077	0.066	0.054	0.127
0.750	0.500	0.091	0.087	0.078	0.066	0.054	0.041	0.083
0.760	0.593	0.088	0.079	0.067	0.054	0.041	0.029	0.051
0.770	0.683	0.079	0.067	0.054	0.040	0.029	0.019	0.029
0.780	0.765	0.067	0.053	0.040	0.028	0.019	0.012	0.015
0.790	0.837	0.053	0.039	0.028	0.018	0.011	0.006	0.007
0.800	0.894	0.039	0.027	0.017	0.011	0.006	0.003	0.003
0.810	0.937	0.026	0.016	0.010	0.005	0.003	0.001	0.001
0.820	0.966	0.016	0.009	0.005	0.003	0.001	0.001	0.000
0.830	0.984	0.008	0.004	0.002	0.001	0.000	0.000	0.000
0.840	0.993	0.004	0.002	0.001	0.000	0.000	0.000	0.000
0.850	0.997	0.002	0.001	0.000	0.000	0.000	0.000	0.000
0.860	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.870	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.880	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

URL, Jan de Leeuw: `http://gifi.stat.ucla.edu`

TABLE 4. Tax bracket probability as a function of true service level (n=1000)

true	100%	114%	128%	142%	155%	168%	181%	194%
0.650	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.996
0.660	0.000	0.000	0.000	0.000	0.000	0.003	0.019	0.977
0.670	0.000	0.000	0.000	0.000	0.003	0.018	0.067	0.911
0.680	0.000	0.000	0.000	0.003	0.018	0.067	0.161	0.751
0.690	0.000	0.000	0.003	0.017	0.066	0.161	0.253	0.500
0.700	0.000	0.003	0.016	0.065	0.161	0.255	0.255	0.245
0.710	0.002	0.016	0.063	0.161	0.257	0.257	0.161	0.082
0.720	0.017	0.062	0.161	0.259	0.259	0.161	0.062	0.017
0.730	0.076	0.161	0.262	0.262	0.161	0.061	0.014	0.002
0.740	0.236	0.265	0.265	0.161	0.059	0.013	0.002	0.000
0.750	0.500	0.267	0.161	0.058	0.012	0.002	0.000	0.000
0.760	0.770	0.160	0.056	0.012	0.001	0.000	0.000	0.000
0.770	0.934	0.054	0.011	0.001	0.000	0.000	0.000	0.000
0.780	0.989	0.010	0.001	0.000	0.000	0.000	0.000	0.000
0.790	0.999	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.800	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 5. Expected tax as function of true service level

true	correct	n=10	n=100	n=1000	n=10000
0.50	194	186	194	194	194
0.51	194	185	194	194	194
0.52	194	184	194	194	194
0.53	194	183	194	194	194
0.54	194	182	194	194	194
0.55	194	181	194	194	194
0.56	194	179	194	194	194
0.57	194	178	194	194	194
0.58	194	177	194	194	194
0.59	194	175	193	194	194
0.60	194	173	193	194	194
0.61	194	172	192	194	194
0.62	194	170	191	194	194
0.63	194	168	190	194	194
0.64	194	166	188	194	194
0.65	194	164	186	194	194
0.66	194	162	183	194	194
0.67	194	159	179	193	194
0.68	194	157	174	189	194
0.69	194	155	168	183	187
0.70	181	152	162	173	174
0.71	168	150	155	161	162
0.72	155	147	148	148	148
0.73	142	145	140	135	135
0.74	128	142	132	122	121
0.75	114	139	125	111	107

TABLE 6. Expected tax as function of true service level

true	correct	n=10	n=100	n=1000	n=10000
0.76	100	136	119	104	100
0.77	100	134	114	101	100
0.78	100	131	109	100	100
0.79	100	128	106	100	100
0.80	100	125	104	100	100
0.81	100	122	102	100	100
0.82	100	120	101	100	100
0.83	100	117	100	100	100
0.84	100	115	100	100	100
0.85	100	112	100	100	100
0.86	100	110	100	100	100
0.87	100	108	100	100	100
0.88	100	106	100	100	100
0.89	100	104	100	100	100
0.90	100	103	100	100	100
0.91	100	102	100	100	100
0.92	100	101	100	100	100
0.93	100	101	100	100	100
0.94	100	100	100	100	100
0.95	100	100	100	100	100
0.96	100	100	100	100	100
0.97	100	100	100	100	100
0.98	100	100	100	100	100
0.99	100	100	100	100	100