Estimating the probability distribution of a travel demand forecast

John L. Bowman
Research Affiliate
Massachusetts Institute of Technology
5 Beals Street, Apt 3
Brookline, MA 02446
617-232-3478
John_L_Bowman@alum.mit.edu
(corresponding author)

Dinesh Gopinath
dgopinath@alum.mit.edu

Moshe Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Massachusetts Institute of Technology
77 Massachusetts Ave, Room 1-171
Cambridge, MA 02139
617-253-5324
mba@mit.edu

Abstract

We present a practical method for estimating the probability distribution of a travel demand forecast. Given a forecast of any variable of interest, such as revenue or ridership, the approach identifies independent sources of uncertainty, estimates a probability distribution of each source, estimates the sensitivity of the variable to each source, and then combines the effects. A case study is presented in which the probability distribution of a revenue forecast is developed for a new transit system.

Introduction

Travel demand forecasts are inherently uncertain because of assumptions about uncertain future events and errors in the forecasting procedure. Any forecast of expected outcome has a potentially large variance and may also be biased. In many cases, information about the probability distribution of possible outcomes is valuable, sometimes far more so than the expected outcome. In the case presented here, for example, the developer of a new public transportation facility needed to understand the lower tail of the distribution of possible revenue outcomes, in order to deal with the possibility that revenues might fall substantially short of expectations.

This paper presents a practical method for estimating the probability distribution of a travel demand forecast. Given a forecast of any variable of interest, such as revenue or ridership, the approach identifies independent sources of uncertainty, estimates a probability distribution of each source, estimates the sensitivity of the variable to each source, and then combines the

effects. We present the method in the next section. This is followed by the case study and a discussion of errors that can occur when the method is used.

Estimation Method

Starting with the predicted value, $r^{(p)}$, of a variable of interest, r, such as revenue or ridership, we first identify a vector of variables, $x = (x_1,...,x_k,...x_K)$, that induce error in the prediction. The error variables come from assumptions about uncertain future events, as well as limitations of the forecasting method, such as imperfect data, incorrectly specified models, and faulty procedures. For example, if the forecast depends on economic growth assumptions, one variable in the vector might be the assumed GDP. The error variables define a K-dimensional space of possible outcomes, each point being paired with a corresponding value of r. The accuracy of the method depends on including all sources of error and defining the error variables so that they are mutually independent.

Using x, we carve the space into a finite set of regions for which we can estimate the outcome of the variable of interest and its probability. For each dimension k we identify a small set of discrete outcomes $x_k^{n_k}, n_k = 1,...,N_k$, to which we can assign all the probability of x_k's possible outcomes, using reasoning and empirical evidence to approximate its true distribution, yielding $p(x_k^{n_k}), n_k = 1,...,N_k$ with $\sum_{n_k=1}^{N_k} p(x_k^{n_k}) = 1$. We then define a set of scenarios,

$S = \{(x_1^{n_1},...,x_k^{n_k},...,x_K^{n_K}); n_k = 1,...,N_k; k = 1,...,K\}$, covering all combinations of the discrete outcomes in all K dimensions. The number of scenarios in S equals $\prod_{k=1}^{K} N_k$. Using s as a one-dimensional index of the members of S, we also refer to a single member of S as $x^{(s)} = (x_1^{(s)},...,x_k^{(s)},...,x_K^{(s)})$. To estimate the probability of each scenario, $p(s)$, we assume the mutual independence of the error variables comprising x, so that

$$p(s) = \prod_{k=1}^{K} p(x_k^{(s)}), s \in S. \tag{1}$$

For each scenario s, r takes the unknown value $r^{(s)}$, which is assumed to approximate r for all points in the unidentified region surrounding s that has probability $p(s)$, with the regions together spanning the space of all possible outcomes.

To estimate the outcome $r^{(s)}$ for each region we assume that r depends on x according to a constant elasticity model, $r = \alpha \prod_{k=1}^{K} (x_k)^{e_k^r}$, where α does not depend on x, and e_k^r is the elasticity of r with respect to x_k. In the constant elasticity model, when x_k changes by $n\%$, r changes by $ne_k^r \%$. By dividing this model for the scenario outcome $r^{(s)}$ by the same model for the original forecast outcome $r^{(p)}$, and rearranging terms, we get

$$r^{(s)} = r^{(p)} \prod_{k=1}^{K} (\frac{x_k^{(s)}}{x_k^{(p)}})^{e_k^r}, s \in S. \tag{2}$$

We can thus estimate $r^{(s)}$ by estimating the elasticity of each variable in x separately, which we do by combining empirical tests with judgment.

Together, all pairs $r^{(s)}$ and $p(s)$ provide an estimated discrete approximation of r's probability distribution function.

Case study background

The authors estimated probability distributions of model-based revenue and ridership forecasts made by a transportation consulting firm for a new public transit subsystem in a major Asian city. The analysis relied upon interviews, site inspections, written descriptions of the transit system and forecasting model, and model forecasts produced by the consultant, including a battery of special test runs requested for this analysis.

The consultant's forecasting procedure consists of several basic components operating in sequence.

1. Assumptions are made about economic growth, changes in the transportation system, population, and number of households and employment. Model estimates made in subsequent steps depend on these assumptions.

2. Trip generation and distribution models estimate the total number of trips (total demand) made in a typical weekday.

3. Mode choice models estimate the proportion of the trips made by each of several modes, with the main modes defined as private and public, and with public sub-modes including standard and upmarket. The new transit service is classified as upmarket.

4. For each mode a network assignment procedure identifies the paths taken for all the mode's trips, resulting in flow estimates for each link and node in the network, including those representing the new transit service's line segments and stations. It is in the network model that choices between the new service and competing upmarket public mode alternatives are modeled.

5. The modeled network flows are adjusted using judgment to take into consideration anticipated effects which are not modeled, and the results are re-scaled to estimate peak period effects (for capacity analysis) and full-year effects (for revenue analysis).

Sources of uncertainty and bias

After studying the available information, the authors attempted to identify all potential sources of bias and uncertainty in the forecasts. Sixteen separate sources were identified within four major categories. In arriving at the list of 16 sources, some related sources were combined since the method of estimating the probability distribution assumes independence among sources. Here is a brief description of each identified source.

Factors affecting total demand prediction

1. Economic growth and related changes in income and employment. The model system is calibrated to match observed 1995 conditions, and its demand forecasts for a given year depend on the assumed growth rate from 1996 onward. Therefore, the forecasts are subject to uncertainty and bias in the growth estimates from all years back to 1996.

2. Total demand model errors. The models are built on four key assumptions: (1) The distribution of vehicle ownership, given average income, is stable over time; (2) the distribution of household size, given average household size, is stable over time, (3) household trip generation is constant with respect to household size and vehicle ownership, and (4) a 1995 roadside survey, though not a probability sample, is a reliable measure of the incidence of non-home-based and home-based trips for purposes other than work or school.

Model errors and uncertainty affecting mode choice and network model prediction

3. Network model sensitivity to congestion, and inaccurate main mode choice model composite generalized time. (a) The highway link performance functions may produce extremely large highway travel times under jam density conditions and cause the mode choice models to overreact to highway traffic congestion. (b) If generalized time for other modes changes relative to generalized time for the new transit service, then for some origin-destination pairs the path switches from 0% to 100% for the new service or vice versa. (c) The probability-weighted sub-mode generalized time used in the main mode choice model can cause counter-intuitive mode choice behavior.

4. Model error in elasticity of demand for the new service with respect to vehicle ownership. We expect the income elasticity of ridership on the new service to be negative. Even though ridership on the new upmarket service should increase relative to other public transit, this should be overshadowed by the shift from transit to private modes as income and auto ownership increase. In contrast to this reasoning, the consultant's model predicts an increase in ridership when vehicle ownership levels increase.

5. Transit captivity. The assumed transit captivity rates lack an empirical basis, and many different combinations of captivity rates and mode bias constants would have satisfied the consultant's calibration criteria, yielding substantially different overall relative sensitivities of mode share to travel time and cost.

6. Values of time. The distribution, mode choice and assignment models assume values of time based on 1995 measurements and an assumed stable relationship over time between income and value of time. The model predictions are sensitive to the assumed value of time because the new transit service is much faster and more expensive than the current downmarket sub-modes.

7. Relative values of wait time, walk time and in-vehicle time. The model assumes walk and wait time are twice as onerous as in-vehicle time. Although walk and wait time are generally more onerous than in-vehicle time, a range of factors, as small as 1.25 or less and as large as 3 or more can occur, depending on the circumstances and exact definitions.

8. Factors affecting attractiveness relative to other upmarket modes. The assumed boarding and transfer penalties are biased in favor of the new service; stairclimbing and crowding is likely to occur in some of the busiest stations, with an effect not incurred by existing services because of physical layout; differences in comfort, reliability, image and convenience are unaccounted for and may be substantial; public operators threatened by the new service may respond competitively; and forecasts do not account for the future opening of other planned transit services.

9. Errors in the measurement of network times and costs. The revenue and ridership predictions depend on the accurate estimation of travel times and costs for all mode and path alternatives, and the accurate calibration of the model so that its predictions match actual experience for the model's base year. Travel times were calibrated by matching predicted travel times with observed travel times on 30 selected routes. Mode volumes were calibrated by matching predicted mode-specific volumes crossing screenlines. The model was carefully calibrated, but it did not perfectly match prediction with reality, the measurement of 'reality' involved estimation and assumptions, and the number of calibrated counts was very small.

Competitive and operating factors affecting mode choice and network model prediction

10. Operating speeds and headways. The forecasts are based on unrealistically optimistic headways and operating speed assumptions that don't account for trainset down-time or suboptimal dwell times.

11. Roadway improvements. The competing travel times for private mode and other transit services may improve more than was assumed.

12. Feeder systems and fare coordination. The forecasts ignore anticipated addition of feeder service and fare coordination for the new service.

Factors affecting adjustments to model predictions

13. Induced resident demand. Model forecasts are arbitrarily adjusted upward by approximately 10% to account for the "release of suppressed demand". In principle, this is a valid adjustment, but its size has a large degree of uncertainty and, in our judgment, is biased upward.

14. Induced tourist demand. Similarly, 50,000 trips per day are added to the model forecasts to account for tourist demand over and above the modeled tourist demand.

15. Annualization factor. The consultants use a factor of 350 to annualize the estimated weekday demand, arguing that the rate for mass transit systems in Southeast Asia are generally in the range of 340 to 355. Uncertainty and bias in this factor propagate into annual revenue forecasts.

16. Ridership lost to startup time lag. The forecasts ignore the expected gradual build-up of patronage following opening of the new transit service, tied to delay in traveler response and achievement of expected service levels.

Unconsidered potential sources of uncertainty and bias

Despite the authors' efforts to comprehensively account for sources of uncertainty and bias, a few important potential sources were ignored. First, the pricing scheme and fare levels were not finalized at the time of the study, and some pricing flexibility is allowed over time. Second, unexpected variation in the spatial distribution of demand could activate capacity constraints. Third, errors in inflation assumptions introduce uncertainty in the forecasts. The estimated probability distributions that follow do not account for uncertainty and bias caused by these specific factors.

Estimating Probability Distributions of Revenue and Ridership

After an error variable was identified for each potential source of bias and uncertainty, revenue and ridership probability distributions were developed for each forecast year in three steps. First, the forecasting model was used to estimate demand elasticity for each error variable. Second, an approximate discrete probability distribution was estimated for each error variable and forecast year, using the best available information. Third, equations (1) and (2) were used to estimate the combined effect of the 16 error variables. Steps 1 and 2 are described below for 2001 for four of the 16 error variables, one from each category. This is followed by a figure illustrating the independent 2001 distributions (step 2) for all 16 error variables and another figure showing the 2001 estimated revenue distribution (step 3).

Total demand model error (error source 2)

In addition to uncertainty of economic growth forecasts, identified as source 1, uncertainty exists in the model's prediction of total demand, given the economic forecasts. That is, even if the economic forecasts are correct, the model can still predict the wrong total demand, because of errors in the model's specification of how total demand depends on the economic factors. The total demand models are built on four key assumptions: (1) The distribution of vehicle ownership, given average income, is stable over time; (2) The distribution of household size, given average household size, is stable over time, (3) household trip generation is constant with respect to household size and vehicle ownership, and (4) the 1995 roadside survey, though not a probability sample, is a more reliable measure than the household interview of the incidence of non-home-based and home-based trips for purposes other than work or school.

Of these assumptions, the one about auto ownership is most suspect. We have anecdotal evidence that, in recent years, income has concentrated in the top income quintiles. If this trend has occurred since 1995, then vehicle ownership may have declined relative to average income. On the other hand, if the cost of vehicle ownership rises (or falls) relative to the cost of living, then auto ownership will probably fall (or rise) relative to income. Assumptions (2) through (4) also introduce error, but again we have no strong evidence of systematic bias, and here the variance they introduce is probably small.

Ridership and revenue elasticities with respect to total demand were estimated using the consultant's model forecasts for 2001 under two different total demand assumptions. An examination of these and other model forecasts, in light of the identified weaknesses in the

models, indicated that the modeled elasticities were probably overestimated, so we discounted the raw elasticity estimates by 10%. The elasticity estimation results are in Table 1.

Table 1: Elasticity estimates for error source 2

Daily demand	case 1	case 2	% change	modeled elasticity	discounted elasticity (e_2^r)
Assumed Total (million trips)	17.05	18.16	6.51		
Modeled Ridership (million passengers)	.641	.591	8.46	1.30	1.17
Modeled Revenue (monetary units)	18.0	19.6	8.89	1.37	1.23

To estimate the distribution of the total demand model error, we assume a symmetric bell-shaped distribution with a standard error of approximately 5% for 2001, and approximate it with three probability mass points. Applying the elasticities and probabilities, we calculate ridership and revenue probabilities and expectation, given uncertainty in the total demand attributable to total demand model error.

Table 2: Year 2001 ridership and revenue distribution, considering uncertainty in consultant's estimate of total demand attributable to model errors (source 2)

Total demand, relative to consultant's estimate, (uncertainty attributable to model errors) ($x_2^{(s)} / x_2^{(p)}$)	Probability ($p(x_2^{(s)})$)	2001 Daily Ridership (000's)	2001 Daily Revenue (Monetary units)
.9 (i.e., total demand is only 90% as large as the consultants estimate)	.1	566	17.2
1.0	.8	641	19.6
1.1	.1	716	22.0
Expected Values		**641**	**19.6**

Values of time (error source 6)

In the model system, given the predicted number of trips produced at and attracted to each geographic zone, the model assumes values of time in distributing the trips among zone pairs and, for each pair, estimating each mode's share of trips and predicting the trip route. The model predictions are sensitive to the assumed value of time because the new transit service is much faster and more expensive than the current down-market service. The consultant estimated the 1995 values of time as 25% of average household hourly income, as calculated from a 1995 home interview survey. Future year values are indexed to average income with changes resulting from the assumed economic growth, so errors in income forecasts cause errors in assumed value of time. Although error source 1 accounts for this, there are additional sources of

error in the value of time estimates. These include measurement error in 1995 and changes over time in the relationship between income and value of time.

Using model forecasts, as we did for error source 2, we calculate **ridership and revenue elasticities** (e_6^r) **of .69 and .76**, respectively.

We judge that values of time could range between 10% and 40% of income, and we have no evidence of bias. so we use a discrete approximation of a triangular distribution, in which the peak occurs at the consultant's estimate of .25. Applying the elasticities and probabilities, we calculate ridership and revenue probabilities and expectation, given uncertainty in the values of time.

Table 3: Year 2001 ridership and revenue distribution, considering uncertainty in values of time (source 6)

Possible values of time, relative to consultant's estimates ($x_6^{(s)} / x_6^{(p)}$)	Probability ($p(x_6^{(s)})$)	2001 Daily Ridership (000's)	2001 Daily Revenue (Monetary units)
.17/.25 (i.e., VOT is 17% of income instead of 25%)	.22	436	14.8
1	.56	641	19.6
.33/.25 (i.e., VOT is 33% of income instead of 25%)	.22	783	24.4
Expected Values		**641**	**19.6**

Operating speeds and headways (error source 10)

Headways and operating speed are two of the most important service attributes affecting demand. The consultant's forecasts assume operating speeds of 35kph with 2 minute headways, systemwide, through all hours of operation. However, we expect lower service levels than this. The main reason is that the fleet size, which can increase only with a 2-year lead time, places a strict limit on achievable headways at a given operating speed. If two contractually required spare trainsets are kept on hand, and the rest of the fleet operates at 35 kph, average headways are between 2.65 and 2.8 minutes. If dwell times increase by 15 seconds, then speeds drop by 2.5 kph and headways increase by .2 minutes. The loss of one train would increase headways by .13 minute.

We estimate a one-sided probability distribution, approximated by a three outcome discrete distribution with a high probability of levels below but near the planned operating conditions, and low probabilities of actually achieving the estimates or falling far below the performance targets. From model elasticity tests of changes in speed and headway, we calculate **ridership and revenue elasticities** (e_{10}^r) **of .74 and .78**, respectively, measured with respect to the speed changes. Applying the elasticities and probabilities, we calculate ridership and revenue probabilities and expectation, given uncertainty in the operating speeds and headways.

Table 4: Year 2001 ridership and revenue distribution, considering the possibility of not meeting assumed speeds and headways.

Possible operating speeds ($x_{10}^{(s)}$) and headways	Probability ($p(x_{10}^{(s)})$)	2001 Daily Ridership (000's)	2001 Daily Revenue (Monetary units)
29.75 kph and 3.5 minute headways (speeds 15% slower than assumed)	.1	570	17.3
33.25 kph and 2.5 minute headways (avg. 5% slower than assumed)	.8	617	18.8
35 kph and 2 minute headways (as assumed)	.1	641	19.6
Expected Values		**615**	**18.7**

Induced resident demand (error source 13)

The consultant modified their modeled ridership forecasts upward by 9% each year through 2005 and 12.5% thereafter to account for the release of suppressed demand. In principle, this is a valid adjustment, since it is widely believed that the total demand for travel increases when the time and/or cost of travel goes down, and the total demand model does not take this into account. However, there is no evidence substantiating the choices of 9 and 12.5 percent. The assumption of suppressed demand has a large degree of uncertainty and, in our judgment, is biased upward.

We use the information at hand and judgment to estimate a probability distribution of the consultant's multiplier. First, we choose the interquartile range and, assuming a uniform distribution within it, we mass its 50% probability at its midpoint. We then identify worst and best cases and, for each, mass 25% of the probability one third of the way between the interquartile threshold and the extreme. Row 2 of Table 5 shows the assumed demand factors for the interquartile range, worst case and best case, as well as the factors for the 3 mass points. Since these factors act on the modeled forecast, we translate them into factors that act on the consultant's final forecast, since it includes several post-model adjustments. These factors are shown in row 3 of Table 5 and in Table 6 where, along with the probabilities of each mass point, they are used to estimate the distribution of ridership and revenue. Since the resulting multipliers act directly on the consultant's final forecast, we use elasticity of 1 in applying equation (2) for this error source.

Table 5: 2001 Uncertainty and potential bias attributed to consultant's estimates of induced resident demand

	Worst case	Lower mass point	Lower quartile thresh-old	Central mass point	Upper quartile thresh-old	Upper mass point	Best case
Demand with resident induced demand, as a proportion of the modeled demand	1.01	1.02	1.025	1.05	1.075	1.10	1.15
Demand with resident induced demand, as a proportion of the consultant's forecast		0.941		0.966		1.008	

Table 6: Distribution of 2001 demand attributed to uncertainty and bias in consultant's estimates of induced resident demand

| Demand multiplier $(x_{13}^{(s)} / x_{13}^{(p)})$	Probability $(p(x_{13}^{(s)}))$	2001 Daily Ridership (000's)	2001 Daily Revenue (Monetary units)
.941	.25	603	18.4
.966	.5	619	18.9
1.008	.25	646	19.8
Expected Values		**622**	**19.0**

Estimated 2001 revenue distribution

Figure 1 plots marginal probability masses for each of the error sources for year 2001 revenue, developed directly from the above estimates for four of the 16 error sources, and estimates similarly made for the other 12 sources. The figure is useful for visualizing the relative importance of the sources in causing bias and uncertainty.

The estimation of discrete approximations of the marginal probability distributions resulted in 3 factors with 2 mass points, 12 factors with 3 mass points and one factor with 4 mass points. The set *S* therefore includes $2^3 3^{12} 4^1 = 17,006,112$ **scenarios.**

Table 7 shows the values used to calculate probability and revenue of the most pessimistic scenario, drawn from above for four of the error sources, and from similar estimates made for the remaining 12 sources. Using these estimates in equation (1) yields the probability of the pessimistic scenario, $p(s) = 1.434*10^{-12}$. Using them in equation (2) with the consultant's revenue estimate of 6870 monetary units yields the pessimistic revenue estimate, $r^s = 1324$ monetary units. Making this calculation for all the scenarios yields an estimated cumulative distribution function, which is shown in Figure 2. In the case study, the developer was most interested in the 10th percentile of the CDF; that is, they wanted a conservative revenue

projection, with only a 10% chance of realizing less than the projected revenue. The analysis provided a 10th percentile revenue estimate of 3624 monetary units.
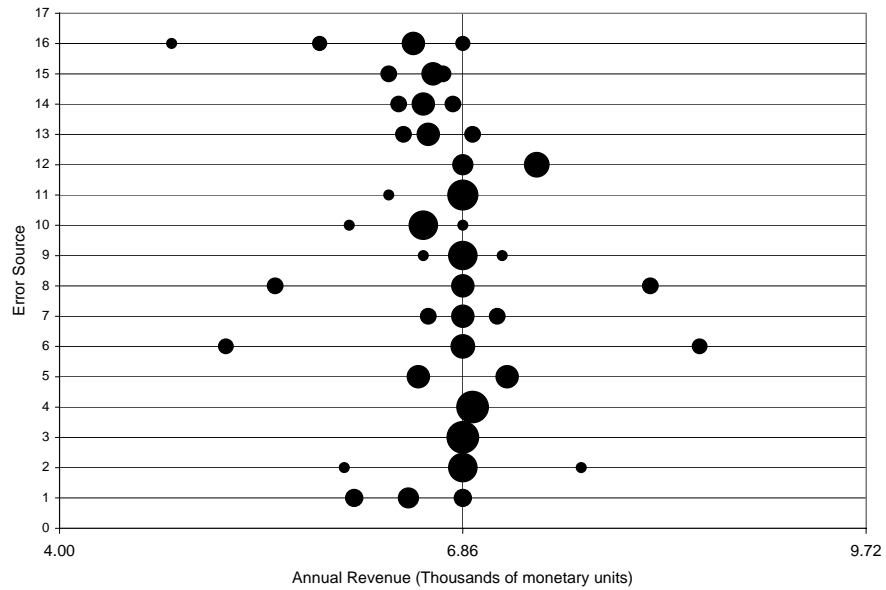
Note: Bubbles on a row represent the source's probability mass points. Each bubble's horizontal location corresponds to its revenue estimate and its area corresponds to its probability. 6.86 is the consultant's revenue estimate.

Figure 1: Estimated 2001 Revenue Probability Masses for Each Error Source

Table 7: Values used to calculate the probability and revenue of the most pessimistic scenario

Source index (k)	Elasticity (e_k^r)	Ratio of factor values ($x_k^{(s)} / x_k^{(p)}$)	Marginal probability ($p(x_k^{(s)})$)
1	1.4	.92	.3
2	1.23	.9	.1
3	1	1	.334
4	1	1.007	.333
5	0.48	.9	.05
6	0.76	.68	.22
7	-0.1	.625	.25
8	1	.806	.25
9	0.42	.9	.1
10	0.78	.85	.1
11	1.58	.95	.1
12	1	1	.4
13	1	.941	.25
14	1	.935	.25
15	1	.924	.25
16	1	.7	.1

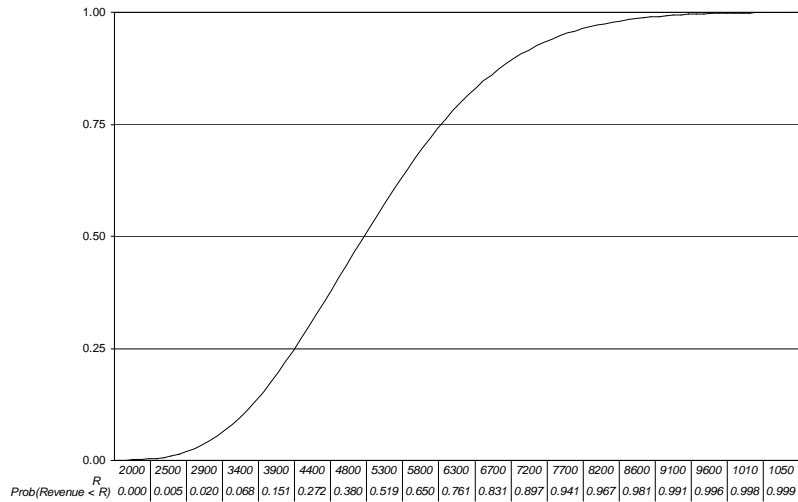| R	2000	2500	2900	3400	3900	4400	4800	5300	5800	6300	6700	7200	7700	8200	8600	9100	9600	1010	1050
Prob(Revenue < R)	0.000	0.005	0.020	0.068	0.151	0.272	0.380	0.519	0.650	0.761	0.831	0.897	0.941	0.967	0.981	0.991	0.996	0.998	0.999

Figure 2: Estimated cumulative distribution function of 2001 revenue

The effects of errors in the estimation of the probability distribution

The method used here to estimate the distribution of a variable of interest by quantifying sources of error is itself subject to error, especially of four major types. First, an important source of error might be missed. Second, the sources of error might be correlated, violating a basic assumption of the method. Third, the estimated elasticity for one or more error sources might be wrong. Finally, the probability distribution of an error source variable might be incorrect. We illustrate each of these potential problems for the case study.

The effect of missing an error source

If a source of error is missed then the variance of the probability distribution will be underestimated. The direction of bias will be opposite the bias caused by the missing error source. Figure 3 shows the estimated distribution when error source 13, related to induced demand, is missing from the calculation.
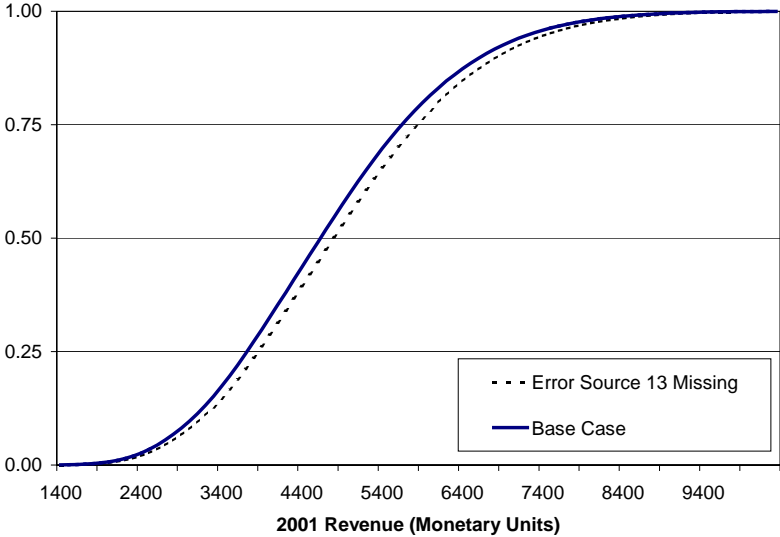

Figure 3: 2001 revenue CDF with and without error source 13 missing

The effect of correlation among error sources

If two sources of error are treated as independent but the errors they induce are actually positively correlated, then the variance of the probability distribution will be underestimated. Conversely, if they are negatively correlated, then the variance will be overestimated. We checked the effect of maximum positive correlation between sources 13 (induced resident demand) and 14 (induced tourist demand) by combining the two factors and assuming that their worst case, middle and best case scenarios coincide. Likewise, we checked the effect of negative correlation by combining them and assuming that 13's best case occurs with 14's worst case, and vice versa. Although the expected effect occurred, in this case it was so small that differences in the CDF graphs were not visually apparent, so no figure is provided.

The effect of incorrect elasticity estimates

If magnitude of the elasticity of the measure of interest with respect to an error variable is underestimated, then the variance of the probability distribution will also be underestimated, and vice versa. The effect on bias depends on the error source's distribution. Figure 4 shows the error distribution if the elasticity of revenue with respect to error variable 10 is 100% larger (1.56 instead of .78).
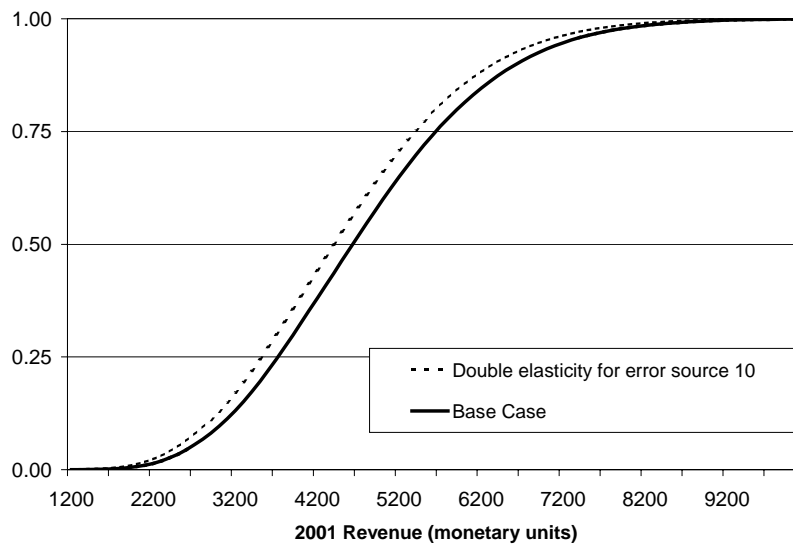

Figure 4: 2001 revenue CDF with doubled elasticity for error source 10

The effect of an incorrect error variable probability distribution

If the variance of an error variable is underestimated then the variance of the variable of interest will also be underestimated, and vice versa. If an error variable's estimated distribution incorrectly estimates bias, this will also bias the distribution of the variable of interest. Figure 5 shows the revised distribution if the distributions for error sources 13 and 14 are more pessimistic. In the pessimistic scenario, the lowest revenue mass points for errors 13 and 14 have probability of .75, the middle mass points have probability .2 and the highest have probability .05, instead of .25, .;75 and .25, respectively.
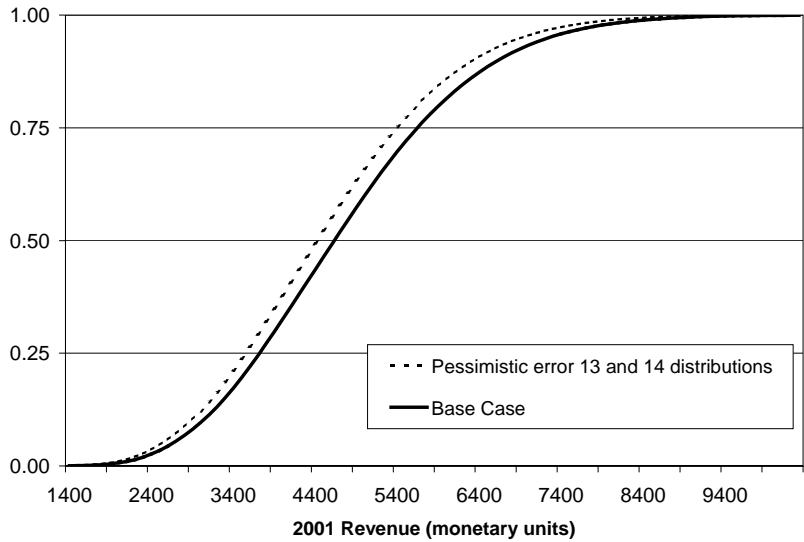
Figure 5: 2001 revenue CDF with pessimistic distributions for errors 13 and 14

Combined effects

If multiple errors occur in the estimation of error sources the combined effect can be larger than individual effects. Figure 6 presents the lowest and highest revenue combinations of the individual effects examined above. In the lowest combination, error source 13 is assumed to be present and positively correlated with error source 14. Error source 10 is doubled from the base case, and the pessimistic probability distributions of error sources 13 and 14 are assumed. The highest revenue case is like the base case, except error source 13 is assumed to be absent.
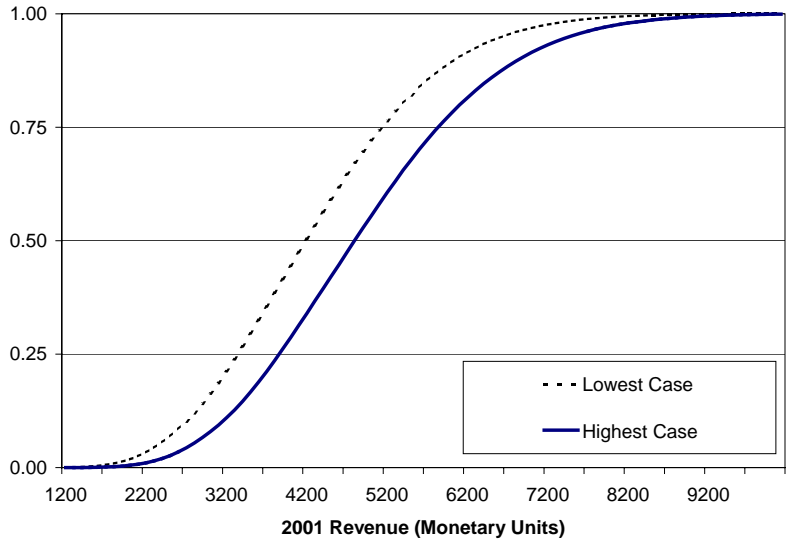

Figure 6: 2001 revenue CDF with lowest and highest revenue combinations from examples

Although the cited examples do not span the range of possible errors encountered in using the presented method, they illustrate the fact that the distribution is an estimate, and is itself subject

to error. The consequences can be significant. Table 8 shows the 10th percentile for all the above examples of errors that might occur in the application of the method. The 10th percentile ranges from 91% to 103% of the originally estimated 10th percentile.

Table 8: 10th percentile of the estimated 2001 revenue distribution for various cases of errors in application of the presented method

Case	10th percentile (Monetary Units)	10th percentile as a proportion of original 10th percentile estimate
Lowest case	3294	.91
Double elasticity for error source 10	3441	.95
Pessimistic distributions of errors 13 and 14	3481	.96
Positive error correlation of sources 13 and 14	3606	1.00
Original 10th percentile estimate	3624	1.00
Negative error correlation of sources 13 and 14	3629	1.00
Highest case (missing error source 13)	3738	1.03

Conclusions

We have demonstrated a practical procedure for using information about sources of uncertainty and bias in a travel demand forecast, or any other forecast, to estimate its probability distribution. This can be useful when information about the distribution, beyond the expected value, is needed. The procedure produces an estimated probability distribution that, like the original estimate, is subject to error that should not be ignored.