*A Note on Alternative Methods of Logistic Projection**

COLM McCARTHY

ENGEL curves exhibiting high income elasticities at low levels of income which then decline with income to produce finite *per capita* saturation levels have been widely used in studies of consumer durable goods (Brown and Deaton [3]). The three parameter logistic growth function is one of several forms which satisfy these requirements and in its numerous applications to problems in biology and epidemiology an estimation method based on the logit transformation is popular (Berkson [1]). A simpler and less onerous method based on the observed rate of growth of *per capita* consumption is common in studies using the logistic to predict the level of ownership of motor vehicles. It appears however that the logit method should also be used in the latter application of the function.

1. *Alternative Logistic Projection Methods*

The three parameter logistic function given by

$$V_t = \frac{a}{1+be^{-ct}} \tag{1}$$

where V_t = ownership level at t,

a, b, c = positive coefficients, the first being the saturation level for V_t

and t = time

*The work reported here was undertaken at the Roads Division of An Foras Forbartha. The author would like to thank Trevor Rue for computational assistance.

has been found to describe closely the time-series data on *per capita* vehicle ownership. There are many objections to this specification but its simple time-dependent form and the close fits to the data which have been found in practice make it attractive for long-range projection. Blackwell, Tanner, Treacy and Tulpule [2], [4], [5], [6], use a method which reduces (1) to

$$V_t = \frac{aV_o}{V_o+(a-V_o)e^{-ar_ot\,(a-V_o)}} \tag{2}$$

where V_o = some past level of V_t,

r_o = rate of growth of V_t at $t = o$

and a = the saturation parameter.

Using (2) it is possible to predict V_t knowing only three inputs, the saturation level a, V_o, usually the most recent level of V_t and r_o, the rate of growth at V_o. In practice a discrete approximation to r_o will be necessary.

The derivation of equation (2) proceeds as follows. Equation (1) has the derivative

$$\frac{dV_t}{dt} = \frac{cV_t(a-V_t)}{a} \tag{3}$$

The time index may be arbitrarily set at o, so

$$\frac{dV_o}{dt} = \frac{cV_o(a-V_o)}{a} \tag{4}$$

Define $r_o = \dfrac{I}{V_o}\,\dfrac{dV_o}{dt}$ (5)

The substitution of (4) into (5) yields

$$r_oV_o = \frac{cV_o(a-V_o)}{a} \tag{6}$$

Hence $c = \dfrac{ar_o}{a-V_?}$ when $t = o$ (7)

$$\text{Furthermore } b = \frac{a-V_o}{V_o} \text{ when } t = o, \text{ from (1)} \tag{8}$$

so equation (1) may be rewritten as equation (2).

The Instantaneous growth rate of a logistic time series is given by

$$\frac{1}{V_t} \cdot \frac{dV_t}{dt} = \frac{c}{a}(a-V_t), \text{ from (3) and (5)} \tag{9}$$

which declines continuously with t. In using equation (2) any discrete approximation to r_o will overestimate the rate of growth for this reason. Furthermore, the partial derivative of V_t in equation (2) with respect to r_o is positive, hence any input value of r_o which represents a discrete approximation will, *ceteris paribus*, cause overprediction of V_t. The process of taking a longer period in approximating the growth rate will be less susceptible to short-run fluctuations in the data and this could compensate the bias resulting from overestimating the growth rate through a reduction in the variance of the predictions.

The estimating equation of the logit method is found by inverting (1), transposing and taking natural logarithms which yields

$$Ln\left(\frac{a}{V_t}-1\right) = Ln\, b-ct \tag{10}$$

This method exploits the full available time series, so its variance properties will presumably be superior to any variant of the growth rate method which uses only two observations. There would appear to be an obvious *a priori* case favouring the logit procedure, but the simplicity and computational economy of the growth rate method would justify its use in the absence of serious risks of prediction error. The customary difficulties in inferring, from purely analytical considerations, what the performance of alternative estimators might be in a specific context, are compounded in the present case since the data input requirements of the two methods are not the same. We will therefore compare them in a small sample context designed to reflect the vehicle ownership projection problem.

2. *A Monte Carlo Experiment*

In order to generate our data, the function

$$V_t = \frac{50}{1+24e^{-.1t}}+U_t \tag{11}$$

was used. The chosen parameters would roughly describe the postwar growth of car ownership per 100 persons in Great Britain with $t = 0$ at 1940, or in the Republic of Ireland with $t = 0$ at 1946. Giving t values from 1 to 50 and ignoring the disturbance term yields the $V_t - U_t$ series given in Table 1. Fifty sets of observations on the first 25 V_t values were obtained using 1,250 random drawings from a normal distribution with mean zero and standard deviation ·5. An additive normal disturbance distribution assumption has been made in order to generate the simulated series. A case can be made for a multiplicative disturbance or for a disturbance with a non-constant variance, but these issues are not pursued here. The parameters of the disturbance distribution were chosen so as to yield R^2 values of around ·99, roughly the kind of values found in fitting logistic functions to postwar British and Irish car ownership data.

TABLE 1: *The Basic simulated series*

t	$V_t - U_t$	t	$V_t - U_t$	t	$V_t - U_t$
1	2·2011	18	10·0660	35	28·9889
2	2·4214	19	10·8940	36	30·1991
3	2·6625	20	11·7698	37	31·3818
4	2·9261	21	12·6934	38	32·5334
5	3·2140	22	13·6642	39	33·6528
6	3·5282	23	14·6789	40	34·7299
7	3·8705	24	15·7367	41	35·7736
8	4·2431	25	16·8355	42	36·7647
9	4·6478	26	17·9696	43	37·7165
10	5·0869	27	19·1348	44	38·6255
11	5·5624	28	20·3298	45	39·4745
12	6·9764	29	21·5473	46	40·2836
13	6·6307	30	22·7795	47	41·0374
14	7·2271	31	24·0246	48	41·7530
15	7·8677	32	25·2750	49	42·4160
16	8·5534	33	26·5235	50	43·0381
17	9·2862	34	27·7642		

The fifty 25-observation series representing values of t from 1 to 25 were then used to prepare 50 sets of logit projections of V_t at $t = 30, 35, 40, 45$ and 50. Two variants of equation (2) were used to prepare the growth rate method projections, one using a one period growth rate input and the other using a five period geometric mean. These three projection methods are identified in table headings as methods 1, 2 and 3.

In Table 2 the population values of V_t at each forecast date are given alongside the means of the values forecast (MF) by each of the three methods. Three measures of forecast performance are also shown, the standard deviation of the forecast (SDF), the root mean square error (RMSE) and the percentage bias, which is simply the mean of the forecast, minus the population value expressed as a percentage of the latter.

TABLE 2: *Alternative predictions of the simulated series*

Forecast period	*Population value*	*Forecast outcomes*	*Forecast method*		
			1	*2*	*3*
		MF	22·8534	23·6155	23·5882
5	22·7795	SDF	·6592	3·7546	1·4280
		RMSE	·6632	3·8465	1·6876
		% bias	0·324	3·670	3·550
		MF	29·1068	30·0661	30·4129
10	28·9889	SDF	·9011	6·6475	2·3940
		RMSE	·9088	6·7342	2·7855
		% bias	0·407	3·716	4·912
		MF	34·8649	35·1479	36·3889
15	34·7299	SDF	1·0050	8·1368	2·8466
		RMSE	1·0140	8·1475	3·2948
		% bias	0·389	1·204	4·777
		MF	39·5963	38·7853	40·9930
20	39·4745	SDF	·9570	8·5664	2·7657
		RMSE	·9647	8·5941	3·1551
		% bias	0·309	−1·746	3·847
		MF	43·1347	41·3325	44·2296
25	43·0381	SDF	·8097	8·4841	2·3693
		RMSE	·8154	8·6538	2·6520
		% bias	0·224	−3·963	2·768

The pattern of the results in Table 2 is quite clear and conforms with expectations. The logit method gives predictions markedly superior to either of the growth rate method alternatives on all three criteria of forecast performance and for all forecast periods. The first alternative of the growth rate method, that using the one-period approximation for r_0, has, as expected, a far greater standard deviation of forecast than the method using the five-period approximation, but a generally larger bias. On the RMSE criterion, which takes account of both bias and variance properties (the mean square error being equal to the variance plus the square of the bias), the five-period method is preferred to the one-period, regardless of the length of the forecast horizon, its superior variance properties offsetting the greater bias where that occurs.

The three criteria of forecast performance shown in Table 2, which are those conventionally used to distinguish the properties of estimators and facilitate the choice between them, are not very instructive in certain respects. In dealing with point predictions from time series, more intuitive appeal may attach to the risks incurred of making errors above a certain tolerance. Since in the present instance the forecast errors may reasonably be taken to follow a normal distribution these probabilities can be calculated and are shown in Table 3 for tolerance levels of 5 and 10 per cent.

The relative positions of the three estimators are, of course, unchanged but it is interesting to note that the method using the one-period growth rate input will make 10 per cent or greater errors on about 60 per cent of the trials. The risk of 5 per cent errors with the logit method nowhere exceeds 11 per cent while the risk of 10 per cent errors is negligible.

It seems fair to conclude from the Monte Carlo experiment that the logit method is markedly superior to either of the growth rate method variants. If the growth rate method is used, the five-period variant is preferable to the one-period calculation.

TABLE 3: *Probabilities of forecast errors greater than 5 and 10 per cent*

Forecast period	*Probabilities of 5 per cent or greater errors*			*Probabilities of 10 per cent or greater errors*		
	Forecast method			*Forecast method*		
	1	*2*	*3*	*1*	*2*	*3*
5	·086	·767	·495	·0006	·555	·167
10	·110	·829	·611	·0014	·667	·304
15	·087	·831	·606	·0007	·670	·298
20	·041	·818	·538	·0001	·646	·214
25	·008	·804	·422	·0001	·577	·105

3. *Comparisons based on Actual Data*

In this section four studies undertaken by different authors are discussed and their projections, all of which were prepared with the aid of the growth rate method, compared with those which emerge from the application of the logit procedure.

Treacy [5] projected total vehicles *per capita* in the Republic of Ireland, using a ·5 saturation level, V_0 being the level in 1961 and r_0 the 1960–61 growth rate. His method is precisely method 2. Tanner's [4] predictions are for private cars *per capita* in Great Britain using ·45 as saturation level and 1964 as the base year. His growth rate input was an "assumed" rate whose precise derivation is not given. This latter feature also characterises Tulpule's [6] study of the British data, which has 1968 as base year and also used ·45 as the *per capita* saturation level. Finally Blackwell [2] used ·45 as saturation level for *per capita* private car ownership in the Republic of Ireland with 1967 as base year and the geometric mean 1962–67 growth rate. The "assumed" growth rate inputs of Tanner and Tulpule appear to be intuitive means of the previous several periods' growth rates, while Blackwell's rate is explicitly obtained in this fashion. All three sets of projections are therefore best viewed as representing method 3.

The logit projections calculated and offered for comparison use the same saturation levels as the original authors, data runs commencing in 1947 for Ireland and 1948 for Great Britain and terminating in each case in the year used as base year by them. In Table 4, the projections of Treacy, Blackwell, Tanner and Tulpule are numbered 1, 2, 3 and 4 respectively while the corresponding logit projections are numbered 1^1, 2^1, 3^1 and 4^1.

The projections of Blackwell, Tanner and Tulpule are higher for all forecast dates than those prepared by the logit procedure. This is quite consistent with the upward bias visible with method 3 in the simulation results. Treacy's results are quite close to the logit projections.

TABLE 4: *Alternative forecasts of British and Irish data*

Year	*1*	*1'*	*2*	*2'*	*3*	*3'*	*4*	*4'*
1965					·17	·158		
1970	·220	·224	·139	·134	·25	·216	·23	·223
1975			·188	·183	·32	·275	·30	·284
1980	·334	·339	·258	·236	·37	·327	·36	·336
1985			·316	·288	·41	·368	·39	·376
1990	·430	·434			·43	·398	·42	·404
1995					·44	·418	·43	·423
2000	·472	·474			·44	·431	·44	·434
2005					·45	·438	·44	·440
2010	·490	·491			·45	·443	·45	·444
2015								
2020	·498	·497						

4. *Summary and Conclusions*

Two alternative methods of obtaining logistic projections of vehicle ownership have been outlined, the logit and growth rate methods. Some reasons for expecting poor predictive performance from the latter have been discussed and corroborated in a simulation experiment designed to reproduce series similar to actual vehicle ownership data. In a direct comparison with existing studies, it was shown that the two methods can produce rather different predictions when applied to the same data.

The logit method requires a reasonably long run of past data, which is generally available in practice. It requires more computation than the growth rate method but appears to produce predictions appreciably less susceptible to estimation errors. It would appear particularly worthwhile to avoid these, since there are so many other and less remediable sources of forecast error in this field.

Institute of Public Administration, Dublin.

REFERENCES

[1] Berkson, J., "Why I Prefer Logits to Probits", *Biometrics*, Vol. 7, 1951.
[2] Blackwell, J., *Transport in the Developing Economy of Ireland*, The Economic and Social Research Institute, Dublin, 1969.
[3] Brown, A. & Deaton, A., "Models of Consumer Behaviour: A Survey", *Economic Journal*, Vol. 82, 1972.
[4] Tanner, J., "Forecasts of Vehicle Ownership in Great Britain", *Roads and Road Construction*, Vol. 43, 1965.
[5] Treacy, P., "A Precursory Review of Traffic Counting in Rural Areas", *Trans. Inst. Civil Eng. of Ireland*, Vol. 89, 1963.
[6] Tulpule, A., *Forecasts of Vehicles and Traffic in Great Britain, 1969*. Transport & Road Research Lab., Crowthorne, Report LR 288, 1969.