MANAGEMENT SCIENCE
Vol. 16 No. 7, March, 1970
Printed in U.S.A.

MODELS OF FACT: EXAMPLES FROM MARKETING*†

A. S. C. EHRENBERG

London Graduate School of Business Studies

The first step in model-building is to note some simple observable patterns and to establish their degree of empirical generalisation. The second step is to integrate different empirical relationships of this kind. This is where the more complex mathematical models come in.

Two examples of model-building in marketing are discussed and the empirical approach is contrasted with the prevalent tendency either to try to develop mathematical models without any prior empirical knowledge ("sonking"), or to search for statistical significance as an end in itself. The basic question in model-building remains as follows:

> Take away the mathematical language and what generalised factual knowledge of the process in question still remains? If the answer is none, the mathematical symbol for *that* is very simple.

Sonking, or The Scientification of Non-Knowledge

What is a model supposed to model? There are two kinds of models, and two kinds of model-builders. On the one hand there is the empirical scientist who knows something and wants a model of what he knows.

On the other hand there is the man who does not know anything about the system in question and wants to build a model of that. Reasons for the popularity of this Scientification of Non-Knowledge [8] are threefold:

Methodological: Sonking is entrenched in the O.R. approach ("First, young man, construct a model of the whole system").

Social: Sonking is to provide instant mathematics for decision-makers in government and industry.

Psychological: Sonking allows the neo-cartesian modelling man to persuade himself of his own existence along well-trodden lines: 'I SONK, THEREFORE I AM'.

The alternative approach of basing oneself on facts tends nowadays to be dismissed as being deficient in theory:

"The result is *merely* empirical".

The trouble with models of fact is however only that empiricism has been given a bad name because of the modern statistical approach. Isolated facts are reported without any attempt at generalisation and integration, i.e. without even aiming to build a comprehensive model at all. Statisticians encourage the modern scientist to end his papers with one-, two-, or three-star levels of significance for his latest results.

This is a non-approach to model-building—to establish the statistical significance of an isolated finding is at best the *beginning* of analysis and not the end. As Gatty [21] has put it in his review of multi-variate statistical analysis:

> Even discovering a significant correlation may not be very helpful without a great deal of further study. Statistical significance of a correlation coefficient, or of a regression coefficient, merely means that there is a pretty good chance that it is in fact a number different from zero. One should not exaggerate the worthwhileness of a coefficient simply because it probably differs from zero.

---

* Received September 1967; revised January 1969.

† Based on a review paper of recent work which was given at the C.E.I.R. Symposium on Model Building in Economics and Industry, London, July 1967 [28]. The examples are mainly taken from work for the Independent Television Authority, JICTAR and J. Walter Thompson (TV audience duplication), and Esso, ICI, J. Walter Thompson and Unilever (buyer behaviour).

The drawback of isolated statistical findings is merely that they are isolated. To overcome this, one has to establish under what other empirical conditions the findings generalise to form a "law" [10, 12]. Once some generalisable findings have been established it becomes worth trying to discover how the different aspects of the system (i.e. the different laws) can be interrelated with each other, so as in fact to build up an empirically-based theory or model.

Two examples from marketing are used in this paper to illustrate these three standard approaches to model-building: firstly sonking itself, secondly the mere search for statistical significance as an end in itself, and thirdly the initial building-up and subsequent integration of empirical generalisations or laws.

A Practical Example: The Pattern of TV Viewing

One example of model-building concerns the extent to which the same people watch television programmes at different points in time. A model which provides a better understanding of TV viewing patterns is of potential value for programme-planning and sociological purposes, and in the efficient use of air-time for advertising.

Possible approaches to building a model of viewing patterns can be along the three lines already indicated.

The sonking approach is illustrated by postulating from the start a comprehensive stochastic model for the whole system. An example might read as follows:

> The audience at any time is regarded as generated by sampling the ith individual from the population at risk with a probability which is related first to the audience size r at the time t and second to the ith individual's general intensity of viewing v. Ignoring this, and assuming that, the proportion r_{st} of the population who view at two times s and t will then be given by $r_{st} = \cdots$, where $\cdots$.

In the absence of any generalised knowledge of actual viewing patterns (see also below), this *a priori* approach would however do no more than model the model-builder's probabilistic assumptions and empirical ignorance.

Real-life examples of the sonking approach to building media models are provided by the simulation and optimization procedures that have been developed and reported in the 1960's, as reviewed for example by Broadbent [2]. A recent specific case was described in *Management Science* by Charnes et al [4]. No generalisable knowledge of audience patterns is reflected by these models. All that is fed in are assumptions and maybe some undigested raw data [e.g. 5].

The multivariate statistical approach consists of taking some specific set of viewing data and deriving a new relationship with the complex of factors which could influence the duplication level, such as:

> The time of day and day of the week for each viewing occasion, whether these occasions are on the same day or on different days, the time of the year and the weather, the channel or station tuned to, the nature of the programmes that are being viewed, the programmes imediately preceding and following on that channel, the programmes on the opposing channels, feed-back effects of previous exposure to various types of programme, the audience levels at each point in time, the nature of the relevant population group, people's viewing and other needs as well as their attitudes and social habits, and so on and so on.

After making some "reasonable" assumptions about the form of the relationship between these explanatory variables and actual viewing behaviour, one would put the given set of data through a multiple regression programme or component analysis or the like and wait for significant coefficients to emerge. Examples of this approach are given by multiple regression analyses which have been referred to elsewhere [18, §7.5] and by various factor analytic studies [15, 20, 27, 30, 32, 34, 36].

Finally, the approach of building-up lawlike relationships is illustrated by first looking for any simple pattern that may exist in some given data, and then generalising by examining many other sets of data obtained under different conditions. An example is provided by the duplication of viewing data shown in Table 1 for Monday, the 24th January and Thursday, the 27th January 1966, in London (England). The table is taken from a recent paper with W. A. Twyman [18]. It gives firstly the percentage or "ratings" of the population who viewed the ITV channel for the first $\frac{1}{4}$-hour in the hour from 5.00 p.m. to 11.00 p.m. on each of the two days, and secondly the percentage of the population who viewed ITV for any two specific $\frac{1}{4}$-hours on both days.

It is then easy to see that viewing at time s on one day is positively correlated with viewing at time t on the other day. Thus the observed duplication figures r_{st} are generally higher than the simple "uncorrelated" cross-product $r_s r_t$. [It is also not difficult to see that these correlations can virtually all be accounted for by a single constant, by noting that r_{st} tends to exceed $r_s r_t$ by a fairly constant factor.] Thus the "duplicated" audience r_{st} at times s and t with ratings r_s and r_t can be described by the one-parameter relationship

$$r_{st} = kr_s r_t ,$$

where k is a constant—here about 1.3—for all pairs of times s and t on the two days. The values 1.3 $r_s r_t$ are shown in **bold** type in Table 1 and fit the observed duplication values to within a mean deviation of about ± 1 percentage points.

Nobody would be really interested in such a description of duplicated viewing in London on a certain Monday and Thursday late in January 1966, if it remained an isolated finding. But examination of several hundred other cases showed that the relationship held under the wide range of empirical conditions shown in Table 2, as was summarised at the time [10]. In addition, more recent data in Las Vegas [11] and in the London and Northern TV Regions in Great Britain [26] followed the same relationship, again within average limits of ± 1. The total number of cases covered now exceeds 200,000.

This duplication result $r_{st} = kr_s r_t$ is not so much a "model" of a complex system but a simple "law" describing one particular aspect of the system. Something more

TABLE 1

Duplication of Viewing: Observed and Predicted Values of the Percentage of Housewives Viewing ITV at any Two Times on Monday and Thursday, January 24th and 27th, 1966 in London*

MONDAY—$\frac{1}{4}$-hour Starting: and Rating r_s		THURSDAY—$\frac{1}{4}$-hour Starting: and Rating r_t						
		5 p.m. 8	6 p.m. 25	7 p.m. 43	8 p.m. 41	9 p.m. 34	10 p.m. 28	11 p.m. 18
5 p.m.	5	1 **1**	2 **2**	3 **3**	2 **3**	2 **2**	1 **2**	1 **1**
6 p.m.	24	5 **3**	14 **8**	14 **14**	12 **13**	10 **11**	10 **9**	6 **6**
7 p.m.	33	4 **3**	13 **11**	19 **19**	17 **18**	15 **15**	11 **12**	9 **8**
8 p.m.	43	4 **5**	14 **14**	23 **24**	22 **23**	19 **19**	16 **16**	9 **10**
9 p.m.	37	4 **4**	11 **12**	19 **21**	20 **20**	16 **17**	12 **14**	7 **9**
10 p.m.	28	2 **3**	8 **9**	15 **16**	14 **15**	14 **13**	10 **10**	8 **7**
11 p.m.	7	1 **1**	2 **2**	3 **4**	3 **3**	3 **3**	4 **3**	4 **2**

* Predicted values $r_{st} = 1.3 r_s r_t$ are shown in **bold** figures.

TABLE 2

Empirical Conditions under Which $r_{ts} = kr_t r_s \pm 1$ is Known to Hold

Any two programmes	1959
Any two days of the week	1965
Any two ratings levels from 0 to about 50	1966
Any two times of day from 2 p.m. to 12 p.m.	Summer
	Winter
Adults	
HW's	
Sets on	London, ITV
Continuous meter panels	Great Britain, ITV
Continuous diary panels	Alabama, WRBC
1-week diary surveys	Two-channel
1-week recall surveys	Poly-channel

like a fully-fledged "model" begins to emerge when one can account for (or "explain") the single parameter k in the law, for instance by relating it to some other aspect of viewing behaviour.

This kind of development has been described by Goodhardt in *Nature* [22]. It follows the standard line of starting with an empirical generalisation, positing an explanatory hypothesis, introducing an assumption, making a technical simplification to facilitate the analysis, skipping the mathematical details ("it can be shown that . . ."), checking that the theoretical result agrees with the empirical facts, discovering that the basic finding (that k is a function of people's viewing intensities) is a mathematical tautology anyway but all the more useful for that, and finally looking towards wider implications.

Because of its brevity, Goodhardt's note is reproduced in full (with a minor change of notation), the contrast with the sonking approach outlined at the beginning of this section being that Goodhardt's theory [only] explains a well-established empirical generalisation or law:

> Consider the proportion of the audience viewing television at some time t who also view a given programme on the same television channel at a certain time s on another day of the week. It is then known that for all different times t this proportion remains approximately constant.
>
> In its most general form, this finding is best expressed by the empirical relationship $r_{st} = kr_s r_t \pm 0.01$ between the proportion r_{st} of the total population who view the channel at both times s and t and the two audience levels or "ratings" r_s and r_t, where k is an empirical constant greater than 1. The part played by the content of a programme in attracting an audience therefore does not seem to act differentially across the population, but is summed up simply by the level of the audience which it attracts.
>
> This simple empirical finding suggests a stochastic model in which the audience at any time s is regarded as generated by sampling the ith individual from the population at risk with a probability which is related first to the audience size r_s at time s and second to the individual's general intensity of viewing v_i. The latter quantity does not vary with the programme being shown and can be defined as the daily total hours viewed by the ith individual divided by the hours viewed by the average individual.
>
> Ignoring that in this model the "sampling" should be without replacement from a finite population, we have that the probability p_{is} of the ith individual viewing the sth segment of time is given as a first approximation by $p_{is} \doteqdot v_i r_s$. Assuming now that two times s and t on two different days are sufficiently far apart for the "sampling" to be independent, the proportion r_{st} of the population of n individual who view at both times should be given by
>
> $$r_{st} = \sum_i (p_{is} p_{it})/n = [\sum_i (v_i' v''_i)/n] r_s r_t ,$$

where v_i' and v''_i are the ith individual's intensities of viewing on the two days. The summation term here is constant for all pairs of times s and t and this theoretical relationship therefore agrees with the empirical result $r_{st} = kr_s r_t$.

The constant k can be calculated either from the observed duplications r_{st}, as $k = \sum r_{st}/\sum (r_s r_t)$ where the summation is over all times s and t on the two days, or from the daily intensities of viewing v_i', and v''_i, as $k = \sum (v_i' v''_i)/n$ where the summation is over all individuals i. It can be shown that these two expressions are mathematically identical.

Examination of English and American viewing data obtained by TAM, Research Services and ARB under a variety of conditions—for example, both recent and five or more years ago, for all transmission times on different days of the week, for adults in general as well as for housewives and for the popular "set-on" type of audience measure, and for four different audience measurement techniques—has shown that the values of k for any pair of days can vary from as little as 1.2 to as much as 2.5, but the relationship $r_{st} = kr_s r_t$ still holds within the average limits of fit of ± 0.01.

This possible explanation of the empirical duplication of viewing law in terms of the general intensity of people viewing now provides a basis for examining the patterns of television viewing in general, and the wider implications of this finding are being investigated.

A Second Example: The Pattern of Consumer Purchasing

The study of consumer purchasing behaviour for non-durable branded household goods provides a second case-history where a wide range of simple empirical generalisations have been subsumed in a single theory or model. Consumer purchasing behaviour is of course central to marketing, and better understanding of it can lead to a variety of practical applications [e.g. 9, 13, 14, 16, 23–5].

The three approaches to model-building which were distinguished earlier can again be identified.

First, there is already quite a history of speculative model-building. For example, repeated attempts have been made to apply first-order homogeneous stationary Markov theory to repeat-buying and brand-switching concepts, references being given in an earlier review [7]. But the attempts appear to have been made without much recourse to real data, two possible reasons being that measurements of consumer purchasing behaviour are generally of the wrong form to feed into the Markov system, and that the assumptions of Markov theory do not apply to the empirical situation.

More recent instances of the kind of model which is based on what Kendall [29, §25] has more generally called "the writing down of hopeful relations full of stochastic elements and unknown functions . . . which merely express our ignorance" are the articles by Bell [1], Charnes *et al* [3], Lipstein [31] and Symonds [35] in *Management Science*. A quick look at such papers will show that if these authors actually knew anything factual about consumer behaviour, they have successfully kept it both from their readers and from their models. Sheth's extensive review of studies of buyer behaviour [33] gives other such references.

Secondly, the approach of explaining things by multiple regression or similar "statistical" means has led to many isolated statistics being reported. Typical examples were recently given by Frank *et al.* [19], who in effect reported findings like 'the multiple adjusted R^2 between the average price paid for Regular Coffee and 22 selected socio-economic and purchasing-behaviour variables was .48, in our data'. At a less sophisticated level we have innumerable market research reports setting out extensive but undigested numerical tabulations from sample surveys and consumer panels.

Thirdly, the empirical "look-and see" approach of searching for simple regular

TABLE 3

The Percentage of Total Sales Accounted for by Repeat-Buyers and by Buyers of at Least r Units, and Other Observed Statistics Approximately Stationary Brands over Two Equal Time-Periods: Twenty Varied Case-histories

Type of Statistic:			Basic	Proportional to w									Constant	Not Directly Related to w				
				% of Sales Accounted for														
					by Buyers of at least r units, for r =													
Product	Brand	Time & Place	Av. Units bght. per buyer: w	by Repeat-Buyers	2	3	4	6	8	12	16	20	Av. Units bght. per "new" buyer	Buyers %	Sales: Units per 100 Inf.	Period in Weeks	Buying of Other Brands	Marketing Activities
Soap Flakes	A	Midl. '65	1.3	45	41	21							1.2	2‡	3	12		
Clothing	B	U.K. '66	1.4	55	54	17	8	3					1.5	45	63	1		
Flour	C	U.S. '51	1.8	65	67	41	41	13					1.6	21‡	36	13		
Soap Flakes	A	Midl. '65	1.9	*	67	53	20	16					*	4‡	8	24		
Detergent	D	Lancs. '63	1.8	69	72	39	25						1.3	8‡	14	4		
Detergent	E	Lancs. '63	2.2	83	81	59	40	[6]					1.3	21	45	4	U	U
Soap Flakes	F	Midl. '65	2.5	82	78	70	58	38	19				1.3	4‡	10	12	N	N
Drink	G	U.K. '55	2.6	76	81	68	55	39	29	19	2		1.6	15	38	13	S	S
Detergent	H	Lancs. '63	2.9	84	88	80	54	[27]	[27]†				1.6	14	40	4	P	P
Fuel	I	U.K. '66	3.1	85	89	78	63	[36]	[21]	[6]	[4]†		1.8	27	85	4	E	E
																	C	C
Drink	G	U.K. '53	3.3	*	87	76	65	49	37	22	14	11	*	19	63	26	I	I
Detergent	D	Lancs. '63	3.3	78	88	74	67	48	38	[18]			1.7	13	43	12	F	F
Soap Flakes	F	Midl. '65	3.8	*	88	82	74	56	46	32	17		*	5	19	24	I	I
Soap Flakes	J	U.S. '51	3.9	89	90	82	67	62	46	33			1.5	16‡	65	13	E	E
Margarine	K	U.S. '51	3.9	87	91	84	72	72	53	33	33	15	1.8	26‡	100	13	D	D
Soup	L	U.K. '58	4.7	*	94	87	82	71	63	50	38	30	*	14	66	26		
Detergent	D	Lancs. '63	4.7	93	94	89	84	67	55	[20]	[3]†		1.6	30	140	12		
Fuel	I	U.K. '66	4.8	*	94	88	84	69	58	[32]	[15]	[8]	*	46	221	8		
Detergent	H	Lancs. '63	6.2	94	96	93	90	86	75	47	[26]	[26]	1.7	18	108	12		
Detergent	H	Lancs. '63	10.1	*	98	95	93	91	87	77	64	[49]	*	22	219	24		

* Repeat-buying information not available.
† "Shelving" pattern—see reference 3.
‡ Less than 50 buyers in sample.

patterns has led to a variety of empirical generalisations during the last ten years or so [cf. 16]. One example relates to consumers' repeat-buying behaviour for any given brand or pack-size in successive time-periods when there is no overall trend in sales. In practice it is useful to understand this kind of stationary equilibrium situation if one wants either to produce a change in it—e.g. an increase in sales—or to evaluate any change that has occurred.

A few of the empirical regularities that have been observed are illustrated in Table 3 which is taken from a recent paper [13]. Noting that the 20 different cases in the table are arranged in increasing order of the average number of purchases per buyer w (the first column of figures), it is easy to see

(a) that the percentage of the sales of a brand in a given period which is accounted for by so-called repeat-buyers (i.e. people who also bought in the preceding time-period) is monotonically related to this quantity w.

(b) That the percentage of sales which is accounted for by repeat-buyers is approximately the same as the percentage of sales accounted for by people who make more than one purchase in the time-period. (Compare the two **bold** columns in Table 3).

(c) That the average number of purchases per "new" buyer (buyers of the brand who did NOT buy the brand in the previous time-period) is roughly constant, at about 1.4 or so.

Various other such empirical regularities in repeat-buying behaviour have been discussed and documented elsewhere [e.g. 6, 16, 17].

Table 3 shows how these kinds of findings in fact hold under a wide range of different empirical conditions. For instance, reading the table from left to right, the results hold for a wide variety of product-fields, different brands in the same product-field, different times and different places, and (further over to the right) different market penetrations (% buying), different sales levels, different lengths of analysis-period, different buying patterns occurring for the *other* brands in the same product-field, and differing marketing conditions more generally. In all, a few thousand cases have been covered so far, the degree of empirical generalisation having been summarised in Table 4 [10].

TABLE 4

Empirical Conditions under Which the Various Aspects of Stationary Repeat-Buying Are Known to Hold

Percentage of buyers ranging from almost 0% to 50% or more
The 4 to 6 leading brands in each product-field
Large, medium and small pack-sizes
U.S.A.
Great Britain
Continental Europe
From 1950 to 1968
Summer/Winter
Different demographic sub-groups (size of household, etc.)
Buying behaviour in periods of 1 week to 6 months or more
Different product-fields including bread, breakfast cereals, butter, canned vegetables, cat and dog foods, clothing, cocoa, coffee, confectionery, cooking fats, dentifrice, detergents, disinfectants, food drinks, household soaps, household cleaners, jams and marmalade, margarine, petrol, polishes, processed cheese, sausages, shampoos, soft drinks, soup, toilet paper, toilet soap

Each of the factual relationships or laws illustrated in Table 3 represents a separate empirical generalisation. The important point to note is that no mathematical analysis has so far been needed—these patterns simply exist, and can be observed as such.

To build an integrated "model" for these various laws, each relationship must first of all be expressed in a convenient form, by some kind of mathematical equation.

One such way of describing the apparent relationships in Table 3 follows from first reformulating w, the average number of purchases per buyer, as a certain parameter q defined by the implicit relationship

$$w = -q/(1 - q)\ln(1 - q).$$

The observed proportion of sales which is accounted for by repeat-buyers (the first column of **bold** figures in Table 3) is then approximately equal to q.

An alternative representation for the most usual parameter-space $2 < w < 20$ can be given as an explicit function of w, to a close degree of approximation,

$$1.23(w - 1)^{1.23}/\{1.23(w - 1)^{1.23} + \tfrac{1}{2}\}.$$

Both these expressions hold for the data in Table 3 (and also more generally) to within average limits of about ± 3 percentage points.[1]

Next, the percentage of sales accounted for by buyers of at least 2 units—as shown by the second **bold** column in Table 3 for example—is similarly represented by the above "1, 2, 3" expression in w. The neater formulation in terms of q, viz. q, has the added advantage of being a special case of a more general descriptor q^{r-1} for the percentage of sales accounted for by buyers of at least r units, for any value of r greater or equal to 1.[2] In the case of the Table 3 data for example, this holds to within about ± 3 percentage points on average (except when the special "shelving" or "variance discrepancy" pattern operates [see 6, 13]).

Other empirical laws of stationary purchasing behaviour are a little more complex in their q-formulation, but some simplify on approximating in the mnemonic 1,2,3-type of language already used above. For example, the proportion of repeat-buyers from one period to the next is given by

$$1 + \ln(1 + q)/\ln(1 - q) \doteqdot 2(1 - w)/(1 - 2.3w), \quad \text{for } 2 < w < 20.$$

Again, the average number of units bought per repeat-buyer in each period is

$$-q^2/(1 - q)\ln(1 - q^2) \doteqdot 1.23w, \quad \text{for } w > 1.5.$$

The latter relationship simply says that under stationary conditions, the repeat-buyers' average rate of buying is about 20% higher than that of all buyers. This relationship holds empirically to within a mean deviation of about .1 or .2 for the various brands and product-fields and lengths of time-period which have been studied [e.g. 6, 17].

An even simpler finding is that the average number of units bought per "new" buyer is given by

$$q/\ln(1 + q) \doteqdot 1.4, \quad \text{i.e. a constant for } w > 2.$$

[1] The numerical values of $100q$ corresponding to the 20 observed w's in Table 3 are 40, 47, 66, 69, 66, 76, 80, 81, 84, 86, 87, 87, 90, 90, 90, 92, 92, 93, 95 and 97. (Note the need to multiply by 100 to convert proportions to percentages.)

[2] Goodhardt has noted recently that the expression $q^{-1-q/(1-q)\ln(1-q)}$ or q^{w-1} is approximately constant at .75 for values of w up to about 10 or so [24]. Buyers of at least the average amount w of a stationary brand in any time-period therefore account for roughly three-quarters of its total sales. This may be seen for the *empirical* data in Table 3 by interpolating for $r = w$.

This is one of the empirical patterns which was already referred to above. The observed values recorded in Table 3 agree to within a mean deviation of about $\pm.2$.

The form of these various equations makes it obvious that they did not spring to life merely by some kind of "empirical curve-fitting". Instead, in this particular form they are all derived from the so-called LSD or Logarithmic Series Distribution model of stationary purchasing behaviour. The importance of this model is that it has integrated various *separate* empirical generalisations or factual laws [see 6, and earlier references given there]. Many of these relationships had already been established previously, often by mere "curve fitting" and sometimes in quite a different mathematical form [see 12].

The one-parameter LSD model itself follows from a more general multivariate two-parameter NBD or Negative Binomial Distribution formulation as a convenient approximation which can be summarized for $i = 1$ to t stationary time-periods of lengths T_i by the probability generating function $G(\mathbf{u})$,

$$G(\mathbf{u}) = \{1 + a\sum T_i(1 - u_i)\}^{-k},$$

where $a = q/(1 - q)$, the u_i are dummy variables, and k is very small.

Validity, Meaning and Practical Usefulness

One reason for having discussed the models of TV audience duplication and of repeat-buying behaviour is that these models not only account for the illustrative data in Tables 1 and 3, but that several hundred million dollars have been spent in the U.S.A. and Western Europe on collecting similar data. This acts as a check both on the descriptive *validity* and on the practical *usefulness* of the two models.

As regards their *meaning*, the modern notion of model-building as a new art seems to seek for "real" meaning in some underlying stochastic formulation. As it happens, such stochastic formulations do exist for both the models here, namely Goodhardt's simulated sampling scheme by which television programmes pick their viewers [22], and a well-known compound Poisson formulation from which the NBD/LSD model of repeat-buying behaviour can be derived [see 6]. Such probabilistic gymnastics can be stimulating and of immense technical value to the model-builder himself, but they do not in fact appear to add to the validity of a descriptive model of fact. There is no short-cut in successful model-building to knowing and understanding what one wants to model.

The essential role in model-building of having well-structured prior empirical knowledge can be further demonstrated by considering situations where the models in the earlier sections are known *not* to apply.

Thus the duplication law does NOT apply to very high ratings, to viewing on the same channel at about 6 p.m. on two different days (see for example Table 1), to duplication patterns of *total* viewing on all channels, nor to viewing of different episodes of the same programme. These are all highly specific situations where the observable patterns are known to differ quite systematically from the particular viewing patterns which are summarised by the model $r_{st} = kr_sr_t$.

Again, the repeat-purchasing patterns for a total product-group, those for new brands [25], those which occur under nonstationary trend conditions [13], and those in certain relatively short time-periods [e.g. 14] all differ from the empirical regularities which occur under equilibrium conditions, as covered by the NBD/LSD model $G(\mathbf{u}) = \{1 + a\sum T_i(1 - u_i)\}^{-k}$.

In all these cases the empirical patterns are different from those already modelled, but there is no point in trying to postulate a mathematical "model" for these different situations until it is known what the different empirical patterns which need to be modelled are actually like (in generalisable terms). Are the 6 p.m. viewing duplications generally *higher* or *lower* than the values given by $r_{st} = kr_s r_t$? Is the percentage of sales of a product-field accounted for by repeat-buyers of the product generally *higher* than the theoretical LSD value of q which corresponds to the product-field value of w, or is it *lower*? Without first knowing which way the facts lie—in a generalisable sort of way—there is clearly no valid model of them which can be built.

The need of starting from generalisable empirical knowledge is emphasised here because of the contrary emphasis in our statistical, marketing, OR, and Management Science journals on developing models or theories *before* one has any empirical and generalisable knowledge about the system in question.

References

1. Bell, C. E., "The N Days of Christmas," *Management Science*, Vol. 14, No. 9 (May 1968), pp. 525–534.
2. Broadbent, S. R., "Media Planning and Models by 1970; A Review," *Applied Statistics*, vol. 15 (1966), pp. 234–256.
3. Charnes, A., Cooper, W. W., Devoe, J. K., and Learner, D. B., "Demon, Mark II," *Management Science*, Vol. 14, No. 9 (May 1968), pp. 513–524 and Vol. 14, No. 11 (July 1968), pp. 682–691.
4. ——, ——, ——, ——, and Reinicke, W., "A Goal Programming Model for Media Planning," *Management Science*, Vol. 14, No. 8 (April 1968), pp. 415–430.
5. ——, ——, Learner, D. B., and Snow, E. F., "Note on an Application of a Goal Programming Model for Media Planning," *Management Science*, Vol. 14, No. 8 (April 1968), pp. 431–436.
6. Chatfield, C., Ehrenberg, A. S. C., and Goodhardt, G. J., "Progress on a Simplified Model of Stationary Consumer Purchasing Behaviour," *J. Roy. Statist. Soc. A.*, Vol. 129 (1966), pp. 317–367.
7. Ehrenberg, A. S. C., "An Appraisal of Markov Brand-Switching Models," *J. Marketing Research*, Vol. 2 (1965), pp. 347–362.
8. ——, "America and the Rest—Some Comparisons," *Proceedings of 48th Annual Conference 1965* (F. E. Webster, ed.), Chicago: American Marketing Association, and *Commentary (Journal of the Market Research Society)*, Vol. 9 (1967), pp. 12–21.
9. ——, "Ten Questions about Consumer Purchasing and Some Answers," *Advertising Quarterly*, Vol. 9 (1966), pp. 3–8.
10. ——, "Laws in Marketing," *Applied Statistics*, Vol. 15 (1966), pp. 257–267.
11. ——, "A Necessary Step in Television Research," *Proceedings of the 50th Annual Conference 1967* (M. S. Moyer and R. E. Vosburgh, eds.), Chicago: American Marketing Association, and *Journal of the Market Research Society*, Vol. 9 (1967), pp. 203–212.
12. ——, "The Elements of Lawlike Relationships," *J. Roy. Statist. Soc. A*, Vol. 131 (1968), pp. 280–329.
13. ——, "The Practical Meaning and Usefulness of the NBD/LSD Theory of Repeat-Buying," *Applied Statistics*, Vol. 17 (1968), pp. 17–32.
14. ——, "Repeat-Buying of Textile Garments," *Op. Res. Quarterly*, Vol. 14 (1968), pp. 421–432.
15. ——, "The Factor Analytic Search for Programme Types," *J. Advertising Research*, Vol. 8, No. 1 (1968), pp. 55–63.
16. ——, "Towards an Integrated Theory of Consumer Behaviour," *J. Market Res. Soc.*, Vol. 11 (1969), pp. 305–337.
17. —— and Goodhardt, G. J., "A Comparison of American and British Repeat-Buying Habits," *J. Marketing Research*, Vol. 5 (1968), pp. 15–18.
18. —— and Twyman, W. A., "On Measuring Television Audiences," *J. Roy. Statist. Soc. A.*, Vol. 130 (1967), pp. 1–59.
19. Frank, R., Green, P. E., and Siebert, H. F., "Household Correlates of Purchase Price for Grocery Products," *J. Marketing Research*, Vol. 4 (1967), pp. 54–58.
20. Frost, A., "The Development of a Technique for TV Programme Assessment," *J. Market Res. Soc.*, Vol. 11 (1969), pp. 25–44.

21. Gatty, R., "Multivariate Analysis for Marketing Research," *Applied Statistics*, Vol. 15 (1966), pp. 157–172.
22. Goodhardt, G. J., "The Constant in Duplicated Television Viewing," *Nature*, Vol. 212, No. 5070 (1966), p. 1616.
23. —— and Ehrenberg, A. S. C., "Conditional Trend Analysis: A Breakdown by Initial Purchasing Level," *J. Marketing Research*, Vol. 4 (1967), pp. 155–161.
24. ——, "The Amount Bought by Above-Average Buyers," *J. Market Res. Soc.*, Vol. 10 (1968), pp. 157–171.
25. ——, "Repeat-Buying of a New Brand," *Brit. J. of Marketing*, Vol. 2 (1968), pp. 200–205.
26. ——, "Duplication of Viewing Within and Between Channels," *J. Marketing Research*, Vol. 6 (1969), pp. 169–178.
27. Green, P. E., Carmone, F. J., and Fox, C. B., "Television Program Similarities: An Application of Subjective Clustering," *J. Market Res. Soc.*, Vol. 11 (1969), pp. 70–90.
28. Kendall, M. G. (ed.), *Mathematical Model-Building in Economics and Industry*, Griffin, London, (1968).
29. ——, "On the Future of Statistics—A Second Look," *J. Roy. Statist. Soc. A.*, Vol. 131 (1968), pp. 182–204.
30. Kirsch, A. D., and Banks, S., "Program Types Redefined by Factor Analysis," *J. Advertising Research*, Vol. 2, No. 3 (1962), pp. 29–32.
31. Lipstein, B., "Test Marketing: A Perturbation in the Market Place," *Management Science*, Vol. 14, No. 8 (April 1968), pp. 437–448.
32. Rothman, L. J., and Rauta, I., "Towards a Typology of the Television Audience," *J. Market Res. Soc.*, Vol. 11 (1969), pp. 45–69.
33. Sheth, J. N., "A Review of Buyer Behaviour," *Management Science*, Vol. 13, No. 12 (August 1967), pp. 718–756.
34. Swanson, C. E., "The Frequency Structure of Television and Magazines," *J. Advertising Research*, Vol. 7, No. 2 (1967), pp. 8–14.
35. Symonds, G. H., "A Study of Consumer Behaviour by Use of Competitive Business Games," *Management Science*, Vol. 14, No. 7 (March 1968), pp. 473–485.
36. Wells, W. D., "The Rise and Fall of Programme Types," *J. Advertising Research*, Vol. 9, No. 3 (1969), pp. 21–27.