KM 17(1985)
pag 89 -108

NETHERLANDS CENTRAL BUREAU OF STATISTICS

Department of Statistical Methods
P.O. Box 959, 2270 AZ VOORBURG, The Netherlands
Tel. 070 - 694341

MODELS FOR MEMORY EFFECTS

Dirk Sikkel*)

Abstract

A family of models for memory effects of answers to retrospective questions is discussed. These models can be used when a respondent is asked how many times he undertook a certain action in a time interval T. The models take into account that the memory effect may depend on the elapsed time since the actions took place and the number of actions that have taken place. The theory is illustrated by examples from the Dutch Health Survey 1981.

Keywords: stochastic process, negative binomial distribution, recall errors, medical consumption.

*) The views expressed in this paper are those of the author and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics. The author thanks the referees for helpful comments.

1. INTRODUCTION

In survey research it is a common situation that the respondents are asked to remember events that took place in some reference period. However, it is a well known fact that the respondent's memory is not completely reliable. Some textbooks, e.g. Som (1973), Sudman and Bradburn (1974) and Moss and Goldstein (1979) are completely or partly devoted to effects of memory on response as a function of time. There are two different types of errors that play a role in data that are based on retrospective questions:

1. Memory effects, which lead to underreporting of events because the respondent has forgotten that some events took place.
2. Telescoping effects, where the respondent does remember the events, but places them incorrectly on the time axis. This may lead to underreporting as well as overreporting. In Schneider (1981) this phenomenon is distinguished in forward-telescoping (placing an event forward in time) and backward-telescoping (placing an event backward in time).

In Sudman and Bradburn (1973) some 500 studies on memory and telescoping effects are examined and they conclude that both play a role, dependent on the subject of the study which may be medical, financial, voting or household expenditures. The larger the reference period, the stronger the memory effects. In medical studies, which is also the subject matter in this paper, the net effect of memory and telescoping effects did not result in overreporting, even when the reference period was small.

When in the above mentioned literature models were adopted, they were of the form $r=\phi(\tau)$, where r is the fraction of events that were reported in a given reference period, τ is the length of that reference period and ϕ is a function that generally declines when τ is large. An example of ϕ that was used in Sudman and Bradburn (1973) is

$$\phi(\tau) = ae^{-b_1\tau}\left[1 + \log \frac{b_2\tau}{\tau}\right] . \qquad (1.1)$$

Here a and b_1 are parameters that relate to memory effects and b_2 refers to telescoping effects. This three parameter model appeared to fit various data sets rather well. Apart from fitting such models the literature focuses on interviewing methods and extra devices that may reduce the memory and telescoping effects to some extent.

In this paper we take the point of view of a statistical office with a mass production of data and a decreasing budget. Our basic motivation consists of questions like: 'Does it make sense to use large reference periods, when events in the past are easily forgotten?', 'Does it make sense to ask the respondent data about more than m^+ events?' and, most important, 'How can data that are contaminated by memory effects contribute to efficient estimation?' (so that there is a trade of between sample size and length of the reference period, given the required precision of the estimates). We will try to achieve this by using a model that describes the stochastic process of the occurrence of events and then treating the memory effects as a part of the model.

We will concentrate on items of the following type: 'How many times did you undertake action a during τ time units immediately preceding the interview and when did these actions take place?' Questions of this type were asked in the Dutch Health Survey 1981, e.g. the number of contacts with the general practicioner (GP) or the number of hospitalizations. In the case of the GP, the total reference period was 3 months and of every contact (up to a maximum of six) it was asked how many weeks before the interview it took place. This made possible the following approach.

The reference period T (of length τ) is partitioned into t subintervals. It is assumed that in the first of these intervals there is no memory effect. In the other intervals there may be memory effects and their size is estimated. They may depend on the number u of the interval ($u=2,3, \ldots,t$) and the number of actions already reported before interval u.

In section 2 the basic process that generates the actions is described. In section 3 the models for the memory effects are given. Section 4 deals with the likelihood function. In section 5 some practical examples are given. Section 6 contains the main conclusions. In appendix A the derivation of the likelihood function is sketched.

2. THE PROCESS OF ACTIONS

We assume that on individual level the actions are generated by a Poisson process. Let M_i be the number of actions of individual i in reference period T with length τ, then

$$P\{M_i = m\} = e^{-f_i \tau} \frac{(f_i \tau)^m}{m!} , \tag{2.1}$$

where f_i is called the frequency of the actions of individual i. This amounts to assuming that the intervals between two consecutive actions of individual i are i.i.d. random variables, exponentially distributed with parameters $1/f_i$. This assumption may not be tenable in all situations, but when the process on individual level is 'random' enough it may be a good approximation to reality. We did, however, investigate the more general assumption that the intervals between two consecutive actions were gamma-distributed (see section 5). The results, however, suggest that this generalization leads to a lot of mathematical troubles and consumption of much computer time with hardly any improvement of the model.

When at random an individual is drawn he has a score f_i on the variable F, which is his frequency parameter of the Poisson-process. Thus in a random sample, F is a latent random variable, which has some distribution. We assume that this can be approximated by a gamma-distribution with parameters b and k, i.e. its density function is

$$g(f) = \frac{e^{-f/b} f^{k-1}}{\Gamma(k) b^k} . \tag{2.2}$$

When M is the number of actions of a randomly drawn individual it is well-known (see e.g. Johnson and Kotz, 1969) that M has a negative binomial distribution, i.e.

$$P\{M=m\} = \frac{k(k+1)\ldots(k+m-1)}{m!} \frac{(b\tau)^m}{(1+b\tau)^{m+k}} . \tag{2.3}$$

This is a widely used model e.g. for the case where the actions consist of buying non-durable consumption goods, see e.g. Chatfield et al. (1966).

A slight complication to the distribution of M is that questionnaires sometimes leave room for only a limited number of actions, due to lack of space. We define m^+ to be the maximum number of actions that can be registered in the questionnaire. Then $\{M=m^+\}$ means {a randomly drawn individual has undertaken m^+ actions or more}.

3. MODELS FOR MEMORY EFFECTS

Let the total reference period T be partitioned into t subintervals $T_1, T_2, \ldots, T_t$ with length $\tau_1, \tau_2, \ldots, \tau_t$ respectively. Because we are looking back in time from the moment of interview, T_u precedes $T_{u'}$ if u>u'. It is assumed that in T_1 no actions are forgotten. In $T_2, \ldots, T_t$ there may be memory effects. This implies that we disregard a possible telescoping effect in T_1. Given the results of Sudman and Bradburn (1973) this seems justified. There was, however, hardly a choice in this decision since only data on the reported actions and not those on the true actions were available.

We now assume that in an interval T_u the rate of actions that are forgotten is constant, but depends on j, the number of actions that are reported in the intervals $T_1, T_2, \ldots, T_{u-1}$. To be more precise, we assume that the probability that a respondent reports an action that he undertook in T_u is equal to $v_{u,j}$. If he has undertaken more than one action in T_u he reports each of them with probability $v_{u,j}$, independent of each other. It should be noted that the memory effect is in this way only related to the number of <u>reported</u> actions, not to the <u>true</u> actions. When no data on the true actions are available this is fundamentally impossible. But our definition covers the idea that a respondent may very well remember his first action but maybe not his third action preceding the interview, so that we can investigate if it makes sense to ask explicitly for this third action. A more fundamental drawback of this setup may be that the choice of the subintervals $T_1, T_2, \ldots$ is subjective. When these intervals are sufficiently small, and the memory effects are more or less smooth functions of u and j, this is no problem in practice. In our examples, partitioning into six subintervals and into twelve subintervals hardly made any difference for the results.

The model for memory effects, as presented here, implies that the number of reported actions in T_u of individual i who has parameter f_i for the true Poisson-process, is Poisson-distributed with parameter $\tau_u(1-v_{u,j})f_i$, given that he has reported j actions up to T_u. The rest of the derivation remaining the same, this amounts to measuring the number of actions as if the length of T_u is not τ_u but

$$\tau'_{u,j} = \tau_u(1 - v_{u,j}) \; . \tag{3.1}$$

This implies that we have $(t-1)*m^+ +2$ parameters (including b and k). When t and m^+ are moderately large, the total number can be very large. For this reason it is desirable to reduce, if possible, the number of parameters by making more restrictive models in order to estimate b and k efficiently. An additional advantage may be that restrictive models may give an extra insight into the structure of memory effects. The following models are investigated.

1. saturated time-number : $v_{u,j}$ has no functional form
2. additive time-number : $v_{u,j} = v_u + v'_j$ ($v'_0=0$ by definition)
3. time : $v_{u,j} = v_u$
4. number : $v_{u,j} = v_j$ ($v_0=0$ by definiton)
5. linear time-number : $v_{u,j} = \beta_1(\tau_1+\ldots+\tau_{u-1}+\frac{1}{2}\tau_u)+\beta_2 j$
6. linear time : $v_{u,j} = \beta(\tau_1+\ldots+\tau_{u-1}+\frac{1}{2}\tau_u)$
7. linear number : $v_{u,j} = \beta j$
8. 0-1 time-number : $v_{u,j} = v_u + \gamma.\text{sign}(j)$
9. 0-1 linear time-number: $v_{u,j} = \beta(\tau_1+\ldots+\tau_{u-1}+\frac{1}{2}\tau_u)+\gamma.\text{sign}(j)$
10. 0-1 number : $v_{u,j} = \gamma.\text{sign}(j)$
11. no memory effects : $v_{u,j} = 0$

Sign(j) equals 1 if j>0 and 0 if j=0.

The models 3, 6 an 11 are independent of the number of reported actions in $T_1\cup\ldots\cup T_{u-1}$. The models 4, 7, 10 and 11 are independent of time. The idea behind models 8, 9 and 10 is that the crucial distinction may be between the first and the second action preceding the interview and that there is no distinction in memory effect between action 2, 3, etc.

It should be noted that the models are partly nested. Model 1 implies every other model. Model 11 is implied by every other model. In table 1 it is indicated which model is implied by which.

Table 1. Nesting of the different models for memory effects

implied models	1	2	3	4	5	6	7	8	9	10	11
implying models											
1. saturated time-number	×	×	×	×	×	×	×	×	×	×	×
2. additive time-number		×	×	×	×	×	×	×	×	×	×
3. time			×			×					×
4. number				×			×			×	×
5. linear time-number					×	×	×				×
6. linear time						×					×
7. linear number							×				×
8. 0-1 time-number								×	×	×	×
9. 0-1 linear time-number									×	×	×
10. 0-1 number										×	×
11. no memory effects											×

4. LIKELIHOOD

The behaviour of an individual in the sample (that is assumed to be random) is characterized by a vector $\vec{M}_i=(M_i(1),M_i(2),\ldots,M_i(t))'$, where $M_i(u)$ is the number of reported actions of individual i in $T_1 \cup T_2 \cup \ldots \cup T_u$. The vector of actions of a randomly selected individual is in the same way given by $\vec{M}=(M(1),M(2),\ldots,M(t))'$. Such a vector, which we will call a 'profile', consists of t non-descending integers ≥ 0 and $\leq m^+$. The number of possible profiles is denoted by p^*. Then

$$p^* = \binom{m^+ + t}{t} . \tag{4.1}$$

The probability distribution of $\vec{M}$ can be derived from the multivariate negative binomial distribution, the multivariate version of (2.3), see appendix A.

We obtain maximum likelihood estimates by standard procedures. We differentiate with respect to the unknown parameters and set the partial first order derivatives to zero. Standard errors are computed by inversion of the matrix of second order derivatives. We were not able to prove that the likelihood function was globally concave. For numerical calculation of the estimates the standard Newton-Raphson algorithm appeared to work well. In our example the results did not depend on initial values in the sense that the algorithm always converged to the same values or did not converge at all. Moreover, the estimates of b and k under different models were more or less similar, which contributes to the confidence in the results. However, strictly speaking, it can not be guaranteed that the estimates in the examples correspond with global maxima of the likelihood function.

The likelihood approach allows us to test how well the models fit the data, but when t and m^+ are even moderately large, the standard goodness of fit statistic poses a problem. The estimation procedure is based on the incomplete t-way table m(1) by m(2) by...by m(t) of all possible profiles. As can be seen from (4.1), the amount of possible profiles can be very large, many of them being very unlikely. This amounts to analyzing a t-way table with many empty cells, on which we cannot apply asymptotic theory. The test statistic, which indicates the goodness-of-fit is

$$G^2 = 2 \sum_{\vec{m}} n_{\vec{m}} \log(n_{\vec{m}}/\hat{n}_{\vec{m}}) , \qquad (4.7)$$

where $n_{\vec{m}}$ is the number of individuals in the sample who have profile $\vec{m}$, and $\hat{n}_{\vec{m}}$ is the estimate of $n_{\vec{m}}$ under a model. G^2 has a χ^2-distribution with a known number of parameters when relatively many $n_{\vec{m}}$ are large. This, however, is very unrealistic. There are several solutions to this problem, e.g. assuming a prior distribution in the t-way table, see Bishop et al. (1975, pp. 410 ff). In the examples of the next section we will evade the problem by collapsing categories. In the model which is saturated with respect to the memory-effects, first we examine the G^2-statistic using only the categories '0', '1' and '2 or more' for the m(u) (u=1,2,...,6). In such a table the cells are well-filled, hence we may apply asymptotic theory. When there is a reasonable fit, we assume the model holds, and analyze the rest of the models on the basis of likelihood ratio test statistics, the alternative hypothesis being the saturated model 1. But then we can use categories '0', '1',...,'5', '6 or more' for the m(u). The likelihood-ratio test statistic G'^2 is χ^2-distributed, the number of degrees of freedom being the difference between the number of parameters of the saturated model and the more restricted model, see Cox and Hinkley (1979, pp. 321 ff).

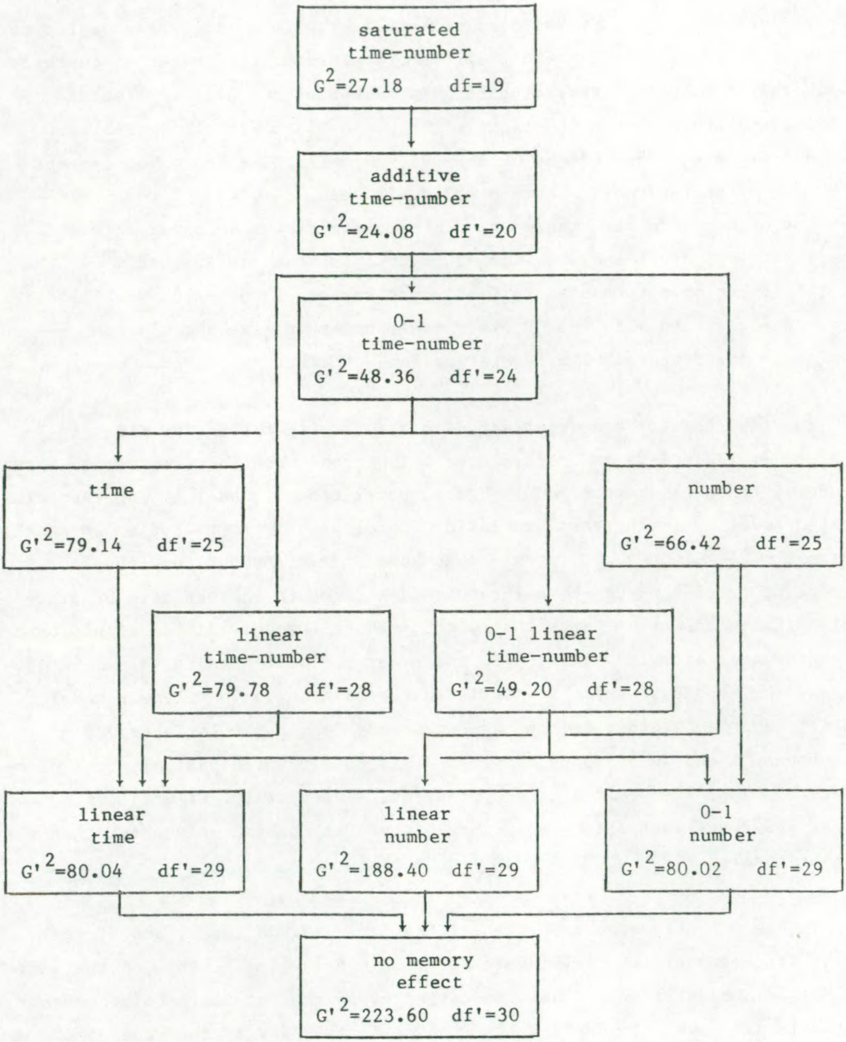
5. EXAMPLES

In the Dutch Health Survey 1981 among others the following three variables were analyzed with respect to their memory effects: the number of contacts with the GP (general practitioner), the number of contacts with specialists, and the number of contacts with the dentist. For all variables we used $t=6$. In the case of the GP and the specialists we had $m^+=6$, (except for the saturated model, where $m^+=2$) $\tau_1=1\frac{1}{2}$ week, $\tau_2,\ldots,\tau_6=2$ weeks. The number of contacts with the dentist was measured over a longer reference period: $\tau_1=1\frac{1}{2}$ month and $\tau_2,\ldots,\tau_6=2$ months. In that case we had $m^+=4$. The total number of respondents in the survey was equal to 10218. We will discuss the contacts with the GP extensively and then give shortly some results of the contacts with specialists and dentist.

The results of the model tests of contacts with the GP are given in figure 1. The models are represented in a graph. When there is a path from model i to model j, then model j is a restriction of model i. For the saturated model G^2 is the test statistic with alternative hypothesis 'no model at all'; G'^2 of the other models does have a χ^2-distribution with df' degrees of freedom where the saturated model forms the alternative hypothesis. The values of G'^2 show that with $\alpha=0.05$ all models differ significantly from the saturated model, with the exception of the additive model. However, the goodness of fit statistic indicates that the 0-1 linear model is still remarkably close to the saturated model given the large amount of respondents and we may hope that with this model, which has only four parameters, we still obtain a good approximation of reality. Figure 1 shows that the model that allows no memory effects is considerably worse than all other models.

In table 2 estimates are given of the parameters b and k, and $\bar{f}$, the mean frequency of the individuals for four models. In the case of the additive and the 0-1 linear model the estimates of the parameters of the memory effects are given. It appears that the estimates of $\bar{f}$ in the case of the 0-1 linear model is slightly different from the saturated and additive models, but it has a considerably lower standard error. This is a natural consequence of the fact that the memory effects are described in a rigid model which we a priori believe to be the truth. In the additive case we find that v_2 through v_6 are smoothly increasing, but the estimates of v_2'

Figure 1. Results of the model tests for contacts with the GP



through v'_5 are rather bizarre, which may be due to the fact that v'_4 and v'_5 have a high standard error. (Note that negative values for $v_{u,j}$ indicate overreporting.) In the sample there were few respondents who had four or

Table 2. Estimates in the saturated, the additive, the linear 0-1 and the 'no memory effects' model for contacts with the GP (time measured in weeks, t=6, m^+=6)

Parameter	Estimate	Standard error
saturated time-number		
b	0.0978	0.0085
k	0.6175	0.0491
$\bar{f}$	0.0604	0.0021
additive time-number		
b	0.0965	0.0080
k	0.6256	0.0469
$\bar{f}$	0.0604	0.0021
v_2	0.0352	0.0433
v_3	0.0997	0.0440
v_4	0.1578	0.0451
v_5	0.2021	0.0469
v_6	0.2760	0.0459
v'_1	0.2516	0.0456
v'_2	0.1347	0.0639
v'_3	0.2093	0.0751
v'_4	-0.2057	0.1424
v'_5	0.2322	0.1407
0-1 linear time-number		
b	0.1051	0.0078
k	0.5832	0.0385
$\bar{f}$	0.0613	0.0016
β	0.0227	0.0038
γ	0.2443	0.0424
no memory effects		
b	0.0594	0.0029
k	0.8009	0.0367
$\bar{f}$	0.0476	0.0008

more contacts with the GP. The estimates of the memory effects in the 0-1 linear model are much easier to interpret. The memory effect increases with 2.27% a week. After the first reported contact there is an additional effect of 24.43%. When no memory effects are assumed, the error of the estimate of $\bar{f}$ is further reduced, but this estimate is obviously very wrong.

Table 3. Empirical distribution after 11.5 week and estimates with standard errors for the saturated, additive and 0-1 linear model for contacts with the GP

Contacts	Measured %	Saturated %	Saturated s.e.	Additive %	Additive s.e.	0-1 linear %	0-1 linear s.e.
0	65.45	62.74	1.00	62.63	0.97	62.94	0.72
1	22.91	20.53	1.00	20.64	0.95	20.11	0.80
2	6.89	8.80	0.78	8.83	0.74	8.72	0.64
3	2.87	4.07	0.52	4.07	0.50	4.12	0.44
4	0.79	1.95	0.33	1.94	0.31	2.02	0.28
5	0.64	0.95	0.20	0.95	0.19	1.02	0.18
6	0.45[a)]	0.47	0.12	0.47	0.11	0.52	0.11
7		0.24	0.07	0.23	0.07	0.27	0.06
8		0.12	0.04	0.12	0.04	0.14	0.04
9		0.06	0.02	0.06	0.02	0.07	0.02
10		0.03	0.01	0.03	0.01	0.04	0.01
Total	100	99.97		99.97		99.96	

a) This percentage indicates '6 or more'.

The description of the memory effects is hardly an end in itself. But it may serve some higher purpose by identifying the parameters of the true process from which we can deduct statistics which are adjusted for memory effects. Such statistics are given in table 3, where the empirical distribution of contacts with the GP is compared to the adjusted distributions according to the three foregoing models. It is clear that the models are very similar compared to the empirical distribution. The categories '0' and '1' have higher empirical percentages, whereas these percentages in the other categories are considerably lower. The 0-1 linear model performs slightly better than the others with respect to the standard errors.

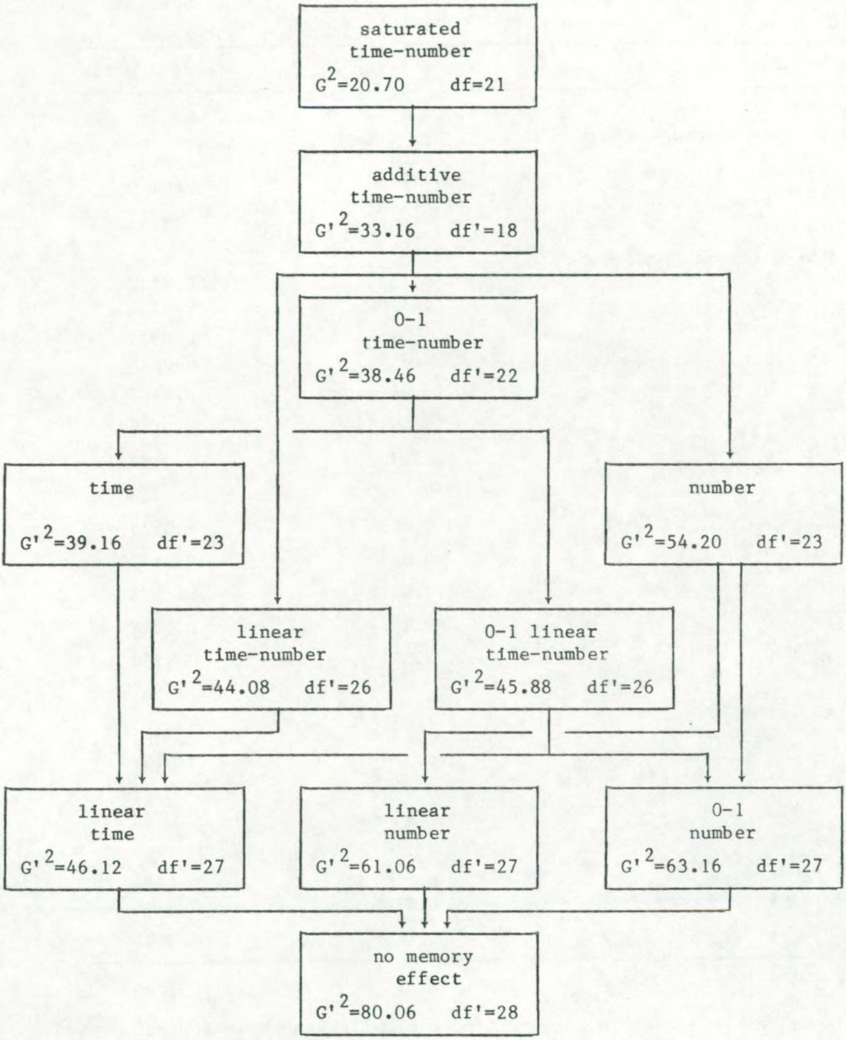
A result which was highly significant for the research project is that in the saturated model the standard errors of the estimate appeared to be more or less constant as a function of t, the number of two-week intervals.

Table 4. Estimates in the saturated, the additive, the linear time and the no memory effects model for contacts with specialists (time measured in weeks, t=6, m^+=6)

Parameter	Estimate	Standard error
saturated time-number		
b	0.1393	0.0140
k	0.2040	0.0176
$\bar{f}$	0.0284	0.0015
additive time-number		
b	0.1375	0.0132
k	0.2066	0.0166
$\bar{f}$	0.0284	0.0015
v_2	-0.0452	0.0676
v_3	-0.0173	0.0723
v_4	0.0986	0.0708
v_5	0.1922	0.0720
v_6	0.1690	0.0733
v_1'	0.0509	0.0809
v_2'	0.0753	0.0970
v_3'	0.0112	0.1208
v_4'	0.2914	0.1218
v_5'	0.1513	0.1938
linear time		
b	0.1379	0.0082
k	0.2194	0.0101
$\bar{f}$	0.0303	0.0011
β	0.0267	0.0038
no memory effects		
b	0.1181	0.0064
k	0.2180	0.0100
$\bar{f}$	0.0258	0.0007

This was not the case in the 0-1 linear model. When the reference period was artificially reduced to 3.5 weeks (by ignoring all earlier contacts), the standard errors increased to the level of the saturated model. This demonstrates the fact that when we use restricted models, events that are reported with memory effects, still contribute to the efficiency of the estimators.

Figure 2. Results of the model tests for contacts with specialists



The results of the second variable, contacts with specialists, are given in figure 2. All models are significantly different from the saturated model with $\alpha=0.05$. The best fitting model that uses few parameters seems to be the linear time model. The number effect does not seem to play a role. Parameter estimates are given in table 4 for the saturated, additive linear time and no memory effects model. Here the difference between the no memory effect model and the other models with respect to the estimate of $\bar{f}$ is much less than in the case of contacts with the GP. Apparently contacts with specialists are remembered better. The standard errors in the linear time model are smaller than in the other models which contain memory effects. This again indicates that it pays to use restrictive models.

The third variable under analysis is the number of contacts with the dentist. It could be expected that here the model would fail completely and so it did for the following reasons. We can split up the population in three groups: 1. those who often go to the dentist; 2. those who go regularly to the dentist every half year; 3. those who never go to the dentist. In group 2 and 3 we surely do not have the assumed Poisson-process on individual level. In group 2 the process is much too regular, in group 3 there is no process at all. Using $m^+=4$ and $t=6$ we found for the saturated model $G^2=7542$ with df=187. Therefore it does not make sense to look at more restricted models. The differences between model and reality are best illustrated by table 5, where the residuals of the profiles are given that in absolute value were larger than 100. When in a profile there is a sequence of exactly 3 ones (a six month period) the empirical frequency is much larger than predicted. When there is a shorter or longer sequence of ones, the empirical frequency is much smaller than predicted by the model.

We tried to improve the model by dropping the assumption that on individual level there is a Poisson process. A Poisson process implies that the interval between two consecutive actions has an exponential distribution. The most natural generalization is that this interval is gamma-distributed with parameters γ_i and ℓ, where γ_i is the scale-parameter and ℓ is twice the number of degrees of freedom. For $\ell=1$ we have the exponential distribution. We can interpret ℓ as a 'regularity' parameter. The higher ℓ, the more regular the process of undertaking actions. For integer values of ℓ,

Table 5. Frequencies, estimated frequencies and residual frequencies of some profiles[a)] in the saturated time-number model for contacts with the dentist

Profile	Observed	Estimated	Residual
0 0 0 1 1 1	605	358	247
0 0 1 1 1 1	413	516	-103
0 1 1 1 1 1	254	455	-201
1 1 1 1 1 1	199	392	-193
0 0 1 1 1 2	943	185	758
1 1 1 1 1 2	18	265	-147
0 0 0 1 2 2	15	131	-116
0 0 1 1 2 2	94	195	-101
0 1 1 1 2 2	867	204	663
1 1 1 2 2 2	754	176	578
0 0 2 2 2 2	27	168	-141
0 2 2 2 2 2	12	120	-108

a) In the profiles the cumulative number of contacts is given over the T_u ($u=1,2,...,6$).

and in particular $\ell=2$, the process was studied in the context of non-durable consumption goods in Chatfield and Goodhardt (1973). In our approach, ℓ was not restricted to integer values. Further it was assumed that there were individual differences in the parameter γ_i, which, again, could be approached by a gamma distribution with parameters b and k. The results, after going through a lot of mathematical trouble, were disappointing. For contacts with the dentist the estimated value of ℓ (on the basis of maximum likelihood) was 1.173 with hardly any improvement of the fit. We must therefore conclude that for processes which are rather regular and that are different for distinct groups, tailor-made models are necessary in which all special features of the process are accounted for.

6. CONCLUSIONS

The model presented here could give a good insight into the structure of memory effects. When it is possible to describe the memory effects with a limited number of parameters they result in efficient estimation procedures of statistics of interest, e.g. the average frequency of contacts. A drawback is that due to the basic assumptions they are not generally applicable. When the process is 'random' enough, the assumption that we have the Poisson property on individual level may be a satisfactory approximation. This seemed to be the case in contacts with the GP and specialists. However, the model performed very poorly when analyzing contacts with the dentist.

As far as the basic questions of this research project concerned, we may conclude that it does make sense to ask retrospective questions about events that are contaminated with memory effects, provided that we have a good model for 1. the process that generated these events and 2. the process of forgetting. When we have a plausible model, then events that lie relatively far in the past, or that are followed by many other (reported) events contribute to the efficiency of the relevant estimators.

APPENDIX A

The derivation of the likelihood function of $\vec{M}$ is a rather tedious excercise, although all steps consist of elementary operations. Here we give a rough sketch of how this function can be derived without proving all steps in detail. Our starting point will be the multivariate version of (2.3). When there are no memory effects, and $m(t)<m^+$ (i.e. the category 'm^+ or more' is not contained in the vector $\vec{m}$) we have:

$$P\{\vec{M}=\vec{m}\} = \phi_{\vec{m}}(\tau_1,\tau_2,\ldots,\tau_t,t)$$

$$= \prod_{x=1}^{t} \left[\frac{\tau_x^{m(x)-m(x-1)}}{(m(x)-m(x-1))!}\right] \frac{k(k+1)\ldots(k+m-1)\ b^{m(t)}}{(1+b\tau)^{m(t)+k}} \qquad (A.1)$$

If m^+ is contained in $\vec{m}$, matters are more complicated. Let us assume that $m(u)<m^+$ for $u=1,2,\ldots,s$ and $m(u)=m^+$ for $u=s+1,s+2,\ldots t$. Then

$$P\{\vec{M}=\vec{m}\} = P\{M(1)=m(1),\ldots,M(s)=m(s)\}\ \ P\{M(s+1)\geq m^+ \mid M(s)=m(s)\}$$

$$= \phi_{\vec{m}}(\tau_1,\tau_2,\ldots,\tau_s,s).\left[1-\sum_{m=m(s)}^{m^+-1} P\{M(s+1)=m \mid M(s)=m(s)\}\right]$$

$$= \phi_{\vec{m}}(\tau_1,\tau_2,\ldots,\tau_s,s).(1-\Psi_{\vec{m}}(\tau_1,\tau_2,\ldots,\tau_s,\tau_{s+1},s))\ , \qquad (A.2)$$

where

$$\Psi_{\vec{m}}(\tau_1,\ldots,\tau_{s+1},s) = \sum_{m=m(s)}^{m^+-1} \frac{(b\tau_{s+1})^{m-m(s)}}{(m-m(s))!} \cdot \frac{\left[1+b\sum_{x=1}^{s}\tau_x\right]^{m(s)+k}}{\left[1+b\sum_{x=1}^{s+1}\tau_x\right]^{m+k}} (k+m(s))\ldots(k+m-1)$$

$$(A.3)$$

If we define Ψ to be zero when m^+ is not contained in $\vec{m}$, then (A.2) gives the general expression for $P\{\vec{M}=\vec{m}\}$ when there are no memory effects.

Now let us assume that memory effects do occur. Then according to (3.1) this amounts to measuring as if the length of T_u is not τ_u but $\tau'_{u,m(u-1)} = \tau_u(1-v_{u,m(u-1)})$. When we have a restricted model, say the linear time-number model, we can substitute the appropriate expression for $v_{u,j}$. Thus

$$\tau'_{u,m(u-1)} = \tau_u\{1-\beta_1(\tau_1+\ldots+\tau_{u-1}+\tfrac{1}{2}\tau_u) + \beta_2 j\} \;. \qquad (A.4)$$

Now the general expression for the probability of a profile $\vec{m}$ is:

$$P\{\vec{M}=\vec{m}\} = \Psi_{\vec{m}}(\tau_1,\tau'_{2,m(1)},\ldots,\tau'_{s,m(s-1)},s)(1-\Psi_{\vec{m}}(\tau_1,\tau'_{2,m(1)},\ldots,\tau'_{s+1,m(s)},s)) \;. \qquad (A.5)$$

From here the derivation of the likelihood function is standard since the numbers of profiles $\vec{m}_1,\vec{m}_2\ldots$ in the sample have a multinominal distribution with parameters $P\{\vec{M}=\vec{m}_1\}$, $P\{\vec{M}=\vec{m}_2\}\ldots$.

REFERENCES

BISHOP, Y.M.M., FIENBERG, S.E. and HOLLAND, P.W. (1975), Discrete multivariate analysis: theory and practice, London : MIT Press.

CHATFIELD, C., EHRENBERG, A.S.C. and GOODHARDT, G.J. (1966), 'Progress on a simplified model of stationary purchasing behaviour', Journal of the Royal Statistical Society A, 129, 317-367.

CHATFIELD, C. and GOODHARDT, G.J. (1973), 'A consumer purchasing model with Erlang inter-purchase time', Journal of the American Statistical Association, 68, 828-835.

COX, D.R. and HINKLEY, D.V. (1979), Theoretical statistics, paperback edition, London : Chapman and Hall.

JOHNSON, N.L. and KOTZ, S. (1969). Discrete distributions, New York : Wiley.

MOSS, L. and GOLDSTEIN, H. eds. (1979), The recall method in social surveys, University of London : Institute of Education.

SCHNEIDER, A.L. (1981), 'Methodological problems in victim surveys and their implication for research in victimology', Journal Criminal Law and Criminology, 72, 818-838.

SOM, R.K. (1973), Recall lapse in demographic enquiries, New York : Asia Publishing House.

SUDMAN, S. and BRADBURN, N.M. (1973), 'Effects of time and memory factors on response in surveys', Journal of the American Statistical Association, 68, 805-815.

——— (1974) Response effects in surveys, Chicago : Aldine.

Ontvangen: 1-7-1984