MAGAZINE AUDIENCE ACCUMULATION: BASIC MODELING ISSUES

Joseph Agresti, Julian Baim, Martin R. Frankel. MRI

Introduction

This paper addresses two basic areas where mathematical models may be employed in the context of magazine audience accumulation.

Specifically it will address the following modeling issues:

I. The development of a mathematical model to characterize the time based accumulation of the audience of a single issue of a publication

II. The development of a mathematical model to characterize the time based accumulation of audience for a sequence of issues of a publication.

The recognition that mathematical models were a necessary tool in magazine audience accumulation studies can be traced to some of the earliest US Magazine Audience studies. In the 1960 study of "Advertising Exposure by Weeks" conducted by Alfred Politz Research on behalf of Reader's Digest, we find the following statement:

" The week-by-week estimates of cumulated issue exposures as derived from the survey showed in all cases a consistent growth pattern. However, to smooth the random variations in these estimates and hence to improve their reliability, it was decided to fit a curve of known characteristics to these data."

In the Newsweek sponsored study titled "Magazines' Daily Audience Accumulation Patterns and Inter-Media Activity Patterns" carried out by Audits and Surveys in 1977, it is noted that:

"Recognizing the fact that there is variability, more or less independent of the particular issue of a magazine, in the day-to-day pattern of reading, in the variability due to sampling, etc., the model would be required to describe the general pattern of audience growth."

We see two major benefits for using mathematical models to describe the accumulation of magazine audiences. The first benefit is exactly the benefit noted in the both the Politz and Audits & Surveys studies. That is the use of a mathematical model allows for the increase in reliability by smoothing out sampling variation. The second benefit associated with the use of models will be recognized in further research that is beyond the scope of the current paper. More specifically, the use of a mathematical model with a relatively small set of basic parameters will provide the dependent variables (i.e. the parameters) that are required in order to establish linkages between the "characteristics" of a magazine and its audience and the specific nature of the accumulation process.

1. A mathematical model to characterize the time based accumulation of the audience of a single issue of a publication

In our search for appropriate models that describe the accumulative audience of a single issue of a single magazine, we began by examining the models that were used in the Politz and Audits & Surveys studies.

Politz Study GOMPERTZ curve

The 1960 Politz Study noted that:

"A priori, it seemed reasonable to expect the one of the conventional asymptotic growth curves would be appropriate.."

Based on visual examination of graphs, felt that the Gompertz curve would serve as an appropriate model to characterize weekly accumulation for the four magazines studied (Reader's Digest, Saturday Evening Post, Life and Look) The form of the curve used was

$Y = ab^{c^X}$, where Y is the cumulated issue exposures, X is the age of the issue in weeks and a, b, and c are constants.

Audits & Surveys Study: Cumulative Gamma function (two parameter)

The 1977 Audits & Surveys study which focused on weeklies postulated that "...the initial reading behavior process tended to be Poisson in nature whereby any on of a large number of people (142,000,000) has a low probability of becoming a reader on any particular day. In addition the probability of this even is not independent of the age issue."

It went on to conclude among the various distributions considered... "the Gamma distribution seemed to be the most fitting, and possess the greatest ease of computation."

In conversations with the Lester Frankel, the principal author of the modeling phase of the Audits & Surveys study, he indicated that he had also considered Zipf's law as well as the Weibull distribution, but that the computational tractability associated with this two parameter Gamma, weighed heavily on its choice.

Other Models considered

The development of the personal computer has had a wide range of impacts in both the business and scientific community. One of the important benefits in field of statistics and mathematics has been the development and widespread availability of "computationally intensive" statistical methods for parameter estimation. By making use of iterative algorithms for non-linear regression, it is possible to consider mathematical models that do not have simple closed form parameter estimation forms. Thus in developing a mathematical model to describe audience accumulation over time, it was possible to greatly broaden the possible functional forms that were considered.

Given the computational flexibility available it was decided to consider the following possible model formulations in 5 general model families: Exponential, Power, Growth, Sigmoidal and Beta-Gamma. The functional form or the 25 different models considered is show given below:

I. Exponential Models:

Basic Exponential:	y	=	a*exp(b*x)
Modified Exponential:	y	=	a*exp(b/x)
Logarithm:	y	=	a+b*ln(x)
Reciprocal Logarithm:	y	=	1/(a+b*ln(x))
Vapor Pressure Model:	y	=	exp(a+b/x+c*ln(x))

II. Power Models:

Power Fit:	y	=	a*x^b
Modified Power:	y	=	a*b^x
Shifted Power:	y	=	a*(x-b)^c
Geometric:	y	=	a*x^(b*x)
Modified Geometric:	y	=	a*x^(b/x)
Root Fit:	y	=	a^(1/x)
Hoerl Model:	y	=	a*(b^x)*(x^c)
Modified Hoerl Model:	y	=	a*b^(1/x)*(x^c)

III. Growth Models:

Exponential Assoc (2):	y	=	a*(1-exp(-bx))
Exponential Assoc (3):	y	=	a*(b-exp(-cx))
Saturation Growth:	y	=	ax / (b + x)

IV. Sigmoidal Models:

Gompertz Model:	y	=	a * exp (-exp(b - cx))
Logistic Model:	y	=	a / (1 + exp (b - cx))
Richards Model:	y	=	a / (1 + exp(b - cx))^(1/d)
MMF Model:	y	=	(ab + cx^d)/(b + x^d)
Weibull Model:	y	=	a - b*exp(-cx^d)

V. Gamma and Beta Models:

Gamma	y	=	a*(x/b)^c*exp(a/b)
Incomplete Gamma	y	=	Integral [0,x] a*(x/b)^c*exp(a/b)
Beta	y	=	a*x^b*(1-x)^c
Incomplete Beta	y	=	Integral [0,x] a*x^b*(1-x)^c

The suitability of these 25 models was tested by obtaining model parameters for each model using 4 different accumulation data sets. The collection of these data sets is described in our companion paper: "Magazine Audience Accumulation: Development of a Measurement System and Initial Results." The four data sets described the audience accumulation for the following magazine and magazine groups:

Data Set I:	TV Guide
Data Set II:	Newsweeklies: Time, Newsweek, U.S. News
Data Set III:	People
Data Set IV:	7 Sisters:

Parameters for all 25 models were computed by the (L-M) Levenberg[1]-Marquardt[2] method for fitting nonlinear regression parameters. This algorithm, which is an iterative form of least-squares, if more fully described in Appendix A. The suitability of mathematical models is typically measured by two measures of "goodness of fit", the "Coefficient of Determination," which is also known as R- Squared, measures the proportion of variation in the dependent variable (in this case the percent of accumulated audience, that is explained by the model. This is a relative measure of model fit to the data. An absolute measure of model fit is given by the standard error of estimate. This measure is the square root of the average squared difference between the model predicted values and the actual data values, expressed in the actual units.

Table 1 shows the values of R-Squared and the standard error of estimate the 5 best performing models over the four data sets. The NA associated with the Weibull for the TV Guide data set indicates that the program encountered a numerical error in the evaluation of a required first derivative needed for the algorithm.

TABLE 1: Coefficient of Determination and Standard Error of Estimate for 5 Best Models

MODEL	TV Guide	N Weeklies	People	7-Sisters
Coefficient of Determination (R-Squared)				
MMF	0.99479	0.99228	0.98680	0.99738
Weibull	NA	0.98710	0.98030	0.99707
Exponential(3)	0.94299	0.93841	0.95358	0.99691
Gompertz	0.98913	0.91862	0.93439	0.99210
Logistic	0.99578	0.90282	0.91870	0.98454
Standard Error of Estimate				
MMF	0.01057	0.01096	0.01930	0.01152
Weibull	NA	0.01416	0.02357	0.01217
Exponential(3)	0.03484	0.03086	0.03607	0.01246
Gompertz	0.01522	0.03547	0.04289	0.01992
Logistic	0.00948	0.03876	0.04774	0.02787

[1] Levenberg, K. (1944). *A Method for the Solution of Certain Nonlinear Problems in Least Squares*. Qty. Appl. Math., v.2, 164-168.
[2] Marquardt, D.W. (1963). *An Algorithm for the Estimation of Non-Linear Parameters*, SIAM J., v. 11, 431-441.

This table shows a high degree of fit for all of the five models shown. As was expected, the two models that make use of four independent parameters generally show the best fit across the four data sets. We feel that on the basis of these data sets the MMF model of the form y = (ab + cx^d)/(b + x^d) provides the best overall fit. Figures 1-4 below show the plot of the this model and the actual diary based accumulation data.

Figure 1

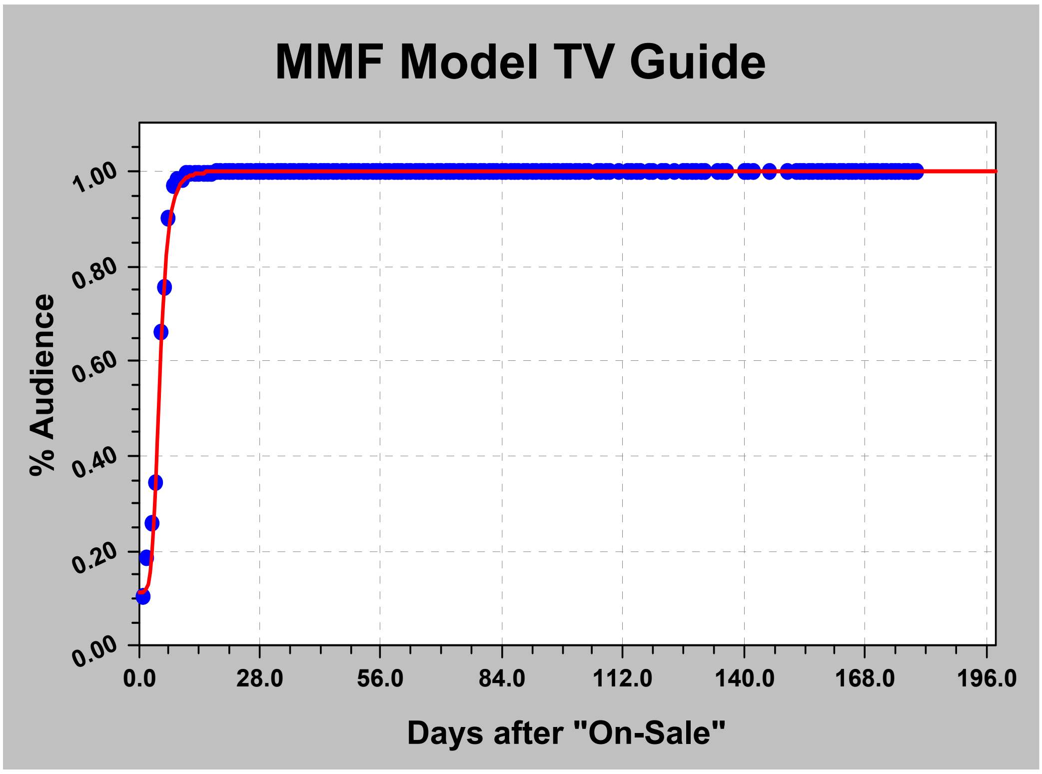


Figure 2

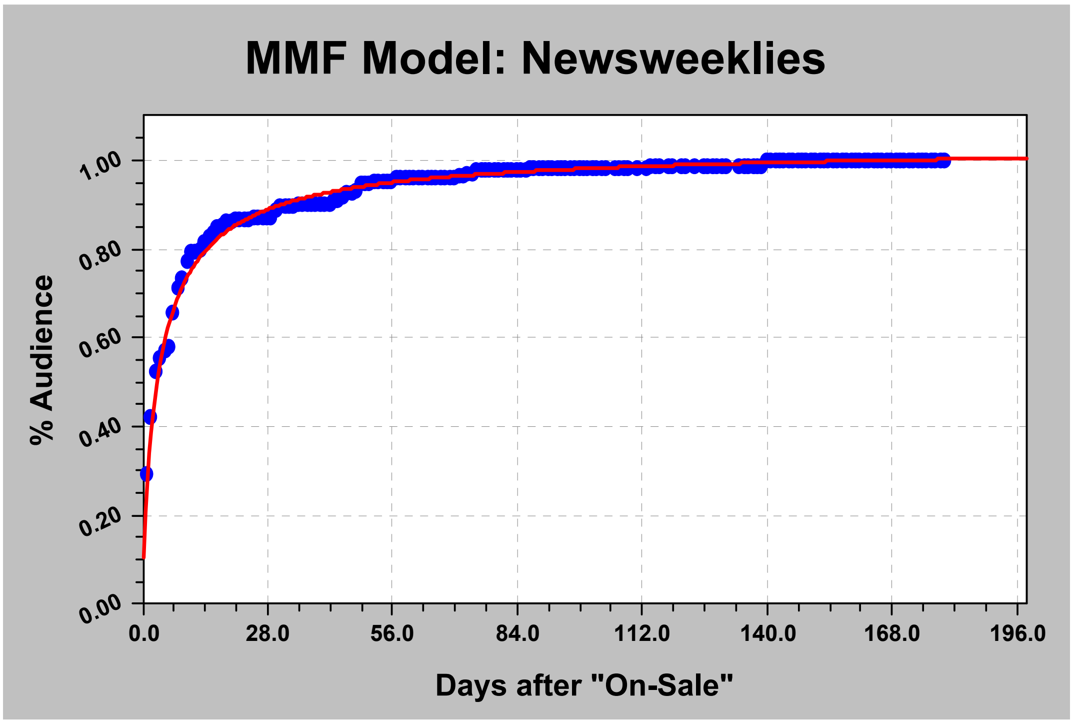


Figure 3

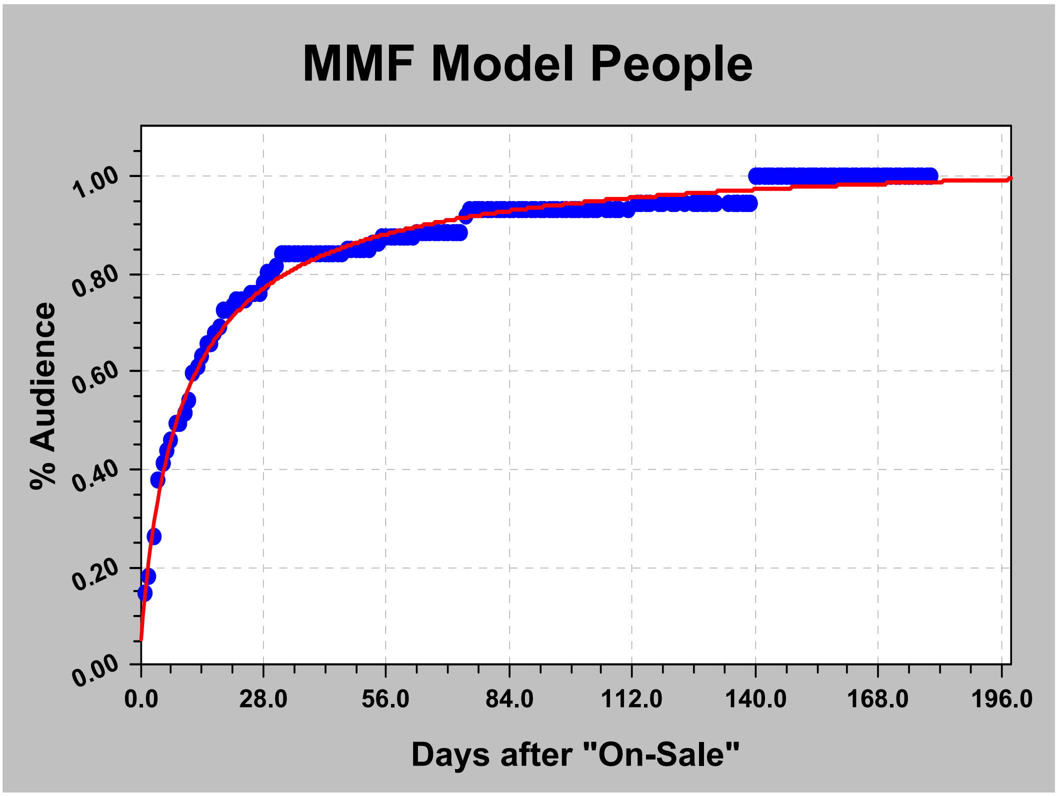
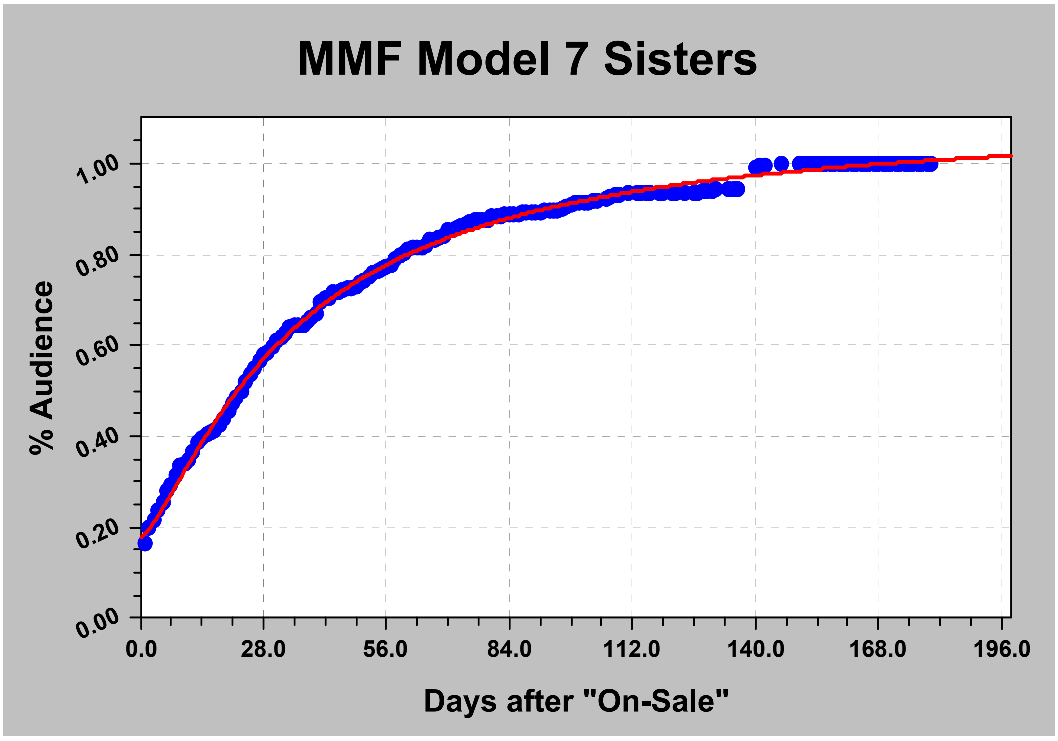


Figure 4



2 Development of a Mathematical Model to Characterize the Tme Based Accumulation of Audience for a Sequence of Issues of a Publication

Virtually all mathematical models that characterize the accumulation of total unduplicated (reach) audience of multiple issues of a single magazine title make use of the Beta-Binomial Model (BBM) first proposed by Hyett. In the first section of this paper we conclude that the MMF model is appropriate for describing the time based accumulation of unduplicated audience for a single issue of a magazine.

These two models, the BBM and MMF, may be combined in order to describe the time based accumulation of audience from multiple issues of a magazine.

Let R_1 denote the single issue audience (reach) of a magazine.

Further, let R_2, R_3, …R_n denote the cumulative reach (via the BBM) of n issues of the magazine.

Define IR_k as the incremental reach between issue k and k-1. That is

$IR_1 = R_1$ and $IR_k = R_k - R_{k-1}$ for k>1

For example, suppose we have a weekly magazine with a single issue total audience rating (average issue coverage percent of total population) of 10.0. Further, assume that the two issue reach (unduplicated net audience) is 16.0. Application of the BBM results in a three-issue reach of 20.2. Thus we have R_1, R_2, and R_3 are equal to 10.0, 16.0, and 20.2. Further we have IR_1, IR_2, and IR_3, are equal to 10.0, 6.0, and 4.2.

Let iaf[x|a,b,c,d] denote the accumulation function for a single issue. In our case

iaf[x|a,b,c,d] = (ab + cx^d)/(b + x^d).

Letting d, denote a specific date, define the function $ET(d_2 - d_1)$ as the number of days between date d_2 and d_1 if d_2 is after d_1, .assuming that d_2 occurs prior to d_1. The function $ET(d_2 - d_1)$ is defined to be zero, if d_1 is prior to d_2, or if d_1 and d_2 are the same. For example, if d_2 is equal to March 15, 1999 and d_1 is equal to March 8, 1999, then $d_2 - d_1$ is equal to 7. Note, the value of $d_1 - d_2$ is equal to 0, not –7.

When the accumulation function is evaluated at $ET(d_2 - d_1)$, the result is a value between zero and one the represents the proportion of an audience that is delivered in $ET(d_2 - d_1)$ days. For example, in the case of a typical newsweekly magazine, the issue accumulation function evaluated at 1, 14, and 21 days is equal to 0.104762, 0.798026, 0.855993[3]

Finally we define an indicator function:

δ(d2 , d1) = 1, when d2 ≥ d1 (when d2 is the same or after d1)
= 0, when d2 < d1 (when d2 is before d1)

For a given magazine, a particular schedule consists of a vector of K on-sale dates, one for each issue. This is denoted S={s_1; s_2; ..s_K } For example, suppose we consider the schedule for a newsweekly consisting of three issues with the first on-sale date July 1, 1999 (first two consecutive, and a one week skip between the second and third,). Thus we have S = { July 1, 1999; July 8, 1999; July 21, 1999 }.

The reach of schedule S as of date t, described by the following formula:

$$R[S,t] = \sum_{k=1}^{K} IR_k \times \delta(t - s_k) \times iaf[ET(t - s_k) \mid a,b,c,d])$$

For example on July 22, 1999, 21 days after the on-sale date of the first issue, 14 days after the on-sale date of the second issue and 1 day after the on-sale date for the third issue, the reach of the three issues would be

[3] The MMF parameters for this function are a= 0.104762, b= 3.623022, c= 1.03243 and d= 0.898683

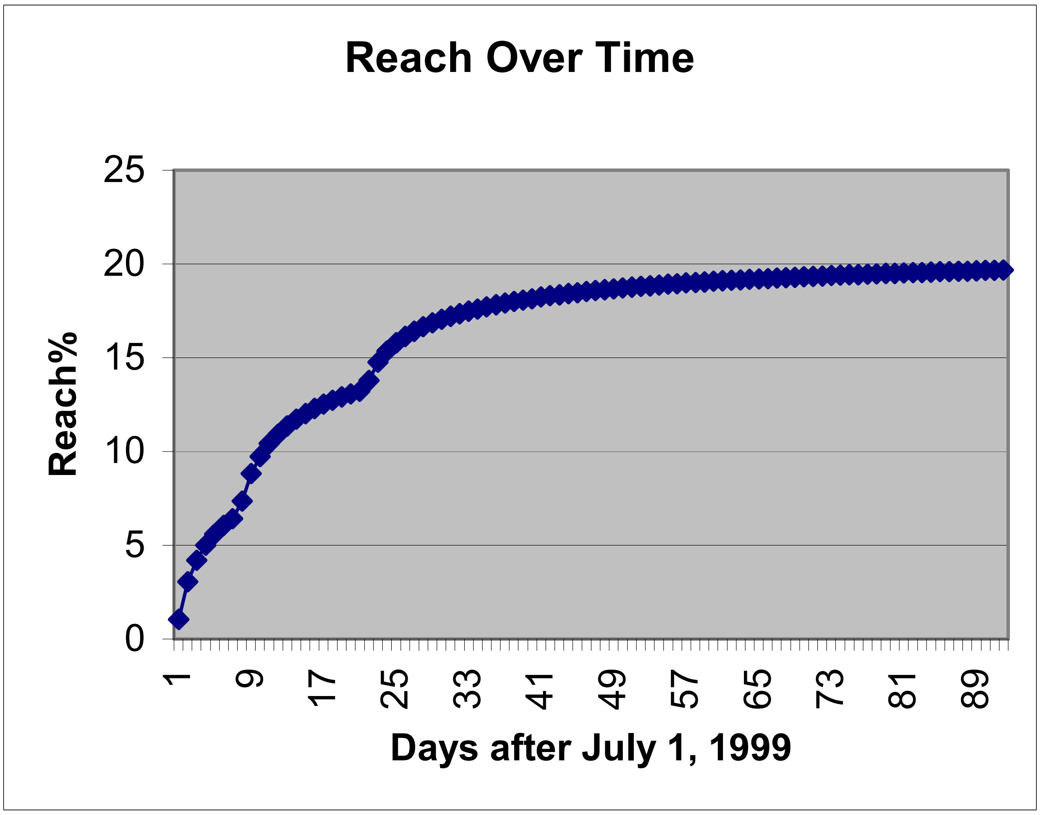
R [S,t] = (IR1 x δ(t , s1) x iaf [t - s1]
+ (IR2 x δ(t , s2) x iaf [t - s2]
+.(IR3 x δ(t , s3) x iaf [t – s3]

R[S,July 22, 1999] = (10.0 x 1 x 0.855993)
+ (6.0 x 1 x 0.798026)
+ (4.2 x 1 x 0.305425)

= 13.78809

Figure 5 shows a graph of the reach values by day (measured in days after July 1, 1999) is show below:

Figure 5



307

As is expected from the single-issue accumulation curve, the graph of the accumulation of 3 issues shows that the build in reach is somewhat wave shaped. The rate of increase is highest immediately following the on-sale dates for the three issues. These on sale-dales occur at 0, 7 and 21 on the horizontal axis. By the 65th day, the curve has almost reached the maximum, which is equal to 20.2.

III Summary and next steps

In this paper we have show that the MMF Model appears to provide an appropriate and practical mathematical model for characterizing the audience accumulation of magazines. This evidence is based on diary recording of first time reading for approximately 1,200 respondents. We plan to increase the sample size using the MRI daily diary to approximately 10,000 respondents. This increased sample will allow us to evaluate the model for a larger number of individual titles and magazine groups. In addition to this effort, we plan to extend our mathematical model to include the characterization of the time specific reach of a schedule of different magazines.