

BASEMENT

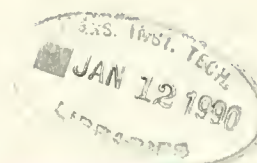


HD28
.M414

no. 1292-
82



3 9080 00583043 2



WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

•Pre-Test-Market Models: Validation and
Managerial Implications

Glen L. Urban*
and
Gerald M. Katz**

Working Paper No. 1292-82

April 1982

incomplete

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

Pre-Test-Market Models: Validation and
Managerial Implications

Glen L. Urban*
and
Gerald M. Katz**

Working Paper No. 1292-82

April 1982

incomplete

*Alfred P. Sloan School of Management, M.I.T.

**Management Decision Systems, Inc.

WIT, DOADIES

JAN 12 1990

INTRODUCTION

In 1978, Silk and Urban proposed a model for pre-test-market evaluation of new packaged goods - i.e. branded, low priced, frequently purchased consumer products. The model (named ASSESSOR) forecasts sales and/or market share for a new brand, provides a structure for evaluating alternative marketing strategies, and generates diagnostics that aid in improving the product. The forecast is based on the convergence of two models -- a trial/repeat model and a preference model. The trial/repeat model utilizes laboratory measures of trial and survey measures of repeat purchase of the new product. The preference model uses existing brand preference and the last brand purchased to parameterize a logit model. Preference for the new product after use is substituted into the logit model to predict share. This data, along with the effects of managements' introductory marketing plan (advertising, promotion, sampling, and couponing) on brand awareness and distribution, are used to derive a convergent forecast. See Silk and Urban for complete details.

Eskin and Malec have proposed an alternative trial/repeat procedure and it has been extended by Kalwani and Silk. Other pre-test-market models are commercially available, but they have not been documented in refereed journals as of this time.

Although pre-test-market models have been widely used in the packaged goods industry, there is a paucity of data available for determining their accuracy. Many advertising claims have been made, but none have been substantiated in professional journals.¹ These claims of exceptional accuracy are in contrast

¹For example, Elrick and Lavidge's "COMP" ad claims "correct to the nearest market share point more than 95% of the time" (Marketing News Nov. 27, 1981, section 2, p. 13). NPD's ad for "ESP" claims "forecasts of year one volume within 9.9% of in-market results" (Marketing News, Nov. 27, 1981, section 2, p. 5). Yankelovich, Skelly, and White claim "95% Record of Accurate Validations" for their "LTM" system (Laboratory Test Market-Record of Validations. New York, Yankelovich, Skelly, and White, undated).

are to be made, they should be made for all cases -- not just those where deviations between actual and predicted are large.

Another difficulty in validation is the long time lag between the pre-test-market forecast and the end of test market (usually 12-15 months). Many economic, political, and social events exogenous to the model may take place. During the time lag, the product may be subject to an organizational change such as a shift from the new products group to the brand group. In this case learning the actual share may be difficult because the brand manager is more concerned with confidentiality than model accuracy.

Comparison of the forecast to the actual test market share is straightforward, but the test market share is also subject to measurement error. In addition, competitive action in the test market may bias the share. Many test markets do not run long enough to obtain a steady state reading of the actual share. Furthermore, the market share itself is subject to variance due to audit procedures and may not be calculated based on the same category definition as in the pre-test-market analysis.

A final difficulty is sample size. Although many products are evaluated, only a fraction of these go to test and not all firms will disclose the market share for these products.

In this paper we will report a systematic attempt to deal with these difficulties. The sampling plan will be described; initial and adjusted shares will be reported; non-response bias will be examined; and an overall assessment of accuracy will be made.

Sampling Design

All firms in which ASSESSOR was applied were surveyed and all product forecasts examined.² 215 products were included in the sample. These

²Lists of firms were supplied by Management Decision Systems, Inc. and Novaction, Inc. The cooperation of these firms is gratefully acknowledged.

Results

Two hundred and fifteen questionnaires were sent and after two follow-up mailings and phone calls, 81 questionnaires or 38 percent of them were returned. Within these 81, 29 products did not go to test market, 5 were in test market but it was too early to observe the final share, and 47 had completed test market. In three of the 47 test market cases, market share could not be supplied because a competitor's product had been tested and the actual share was not known.

Table 1 shows the comparison of initial, adjusted, and actual shares. The average initial share forecast was 7.77 while the average actual share was 7.16. Thus, a positive bias of .61 share points is present and is significant at the 10 percent level ($t = 2.0$). The mean absolute deviation between the initial and actual shares is 1.54 and the standard deviation is 1.99 share points.

As expected, the comparisons between adjusted and actual shares show less error -- mean deviation of $-.01$, mean absolute deviation of $.83$, and standard deviation of 1.12 . Adjustments were made in 36 of the 44 cases. In most of these, the adjustments improved the accuracy, but in six of the cases the deviation increased. Figures 1 and 2 show histograms of the distributions of the errors. The initial versus actual distribution has skewness which is substantially reduced by the adjustment process.

Non response is a threat to the validity of the reported accuracy and must be examined. The initial share forecasts for all 215 studies has a mean of 7.13 and a standard deviation of 6.55. The 44 products in the validation sample have an initial share mean of 7.77 and a standard deviation of 5.72. They are not significantly different estimators of the true underlying population mean and variance ($t = .48$ and $F(214, 43) = 1.31$). Further analysis of non-response was done by comparing the first wave of 24 responses to the total sample of 44. Table 2 shows no significant differences in the mean shares or measures of

Table 1
COMPARISON OF INITIAL, ADJUSTED
AND ACTUAL MARKET SHARES

	INITIAL	ADJUSTED	ACTUAL	INITIAL- ACTUAL	ADJUSTED- ACTUAL
	7.20	4.20	3.30	3.90	0.90
	2.90	2.90	2.70	0.20	0.20
	2.00	2.00	2.20	-0.20	-0.20
	24.20	24.20	23.50	0.70	0.70
	4.40	1.30	0.60	3.80	0.70
	12.60	13.50	13.00	-0.04	0.50
	3.80	3.70	3.00	0.80	.70
	8.40	5.50	6.30	2.10	-.80
	16.50	14.70	12.90	3.60	1.80
	8.40	7.80	7.20	1.20	0.60
	5.20	2.80	1.60	3.60	1.20
	9.60	10.00	10.50	-0.90	-0.50
	3.00	3.00	3.20	-0.20	-0.20
	13.30	11.00	10.40	2.90	0.60
	0.80	2.30	2.50	-1.70	-0.20
	27.10	27.10	28.50	-1.40	-1.40
	13.50	10.00	10.80	2.70	-0.80
	1.23	0.63	0.30	0.93	0.33
	3.00	3.00	2.00	1.00	1.00
	9.00	7.90	7.20	1.80	0.70
	1.10	0.60	0.60	0.50	0.00
	4.90	4.90	5.00	-0.10	-0.10
	2.60	2.00	2.20	0.40	-0.20
	8.00	4.30	4.20	3.80	0.10
	6.00	5.00	4.40	1.60	0.60
	17.00	21.00	22.00	-5.00	-1.00
	7.80	7.90	7.60	0.20	0.30
	12.10	10.80	9.70	2.40	1.10
	5.40	5.80	5.10	0.30	0.70
	12.20	12.20	12.50	-0.30	-0.30
	13.40	14.50	17.20	-3.80	-2.70
	9.30	10.00	12.50	-3.20	-2.50
	1.80	1.80	1.90	-0.10	-0.10
	10.90	10.40	11.50	-0.60	-1.10
	9.00	6.70	8.40	0.60	-1.70
	9.60	9.30	8.20	1.40	1.10
	5.60	5.00	1.50	4.10	3.50
	4.90	2.70	3.80	1.10	-1.10
	14.40	15.00	17.00	-2.60	-2.00
	7.50	5.50	7.00	0.50	-1.50
	0.80	0.30	0.30	0.50	0.00
	0.43	0.43	0.20	0.23	0.23
	0.63	0.66	0.28	0.35	0.38
	<u>0.25</u>	<u>0.19</u>	<u>0.13</u>	<u>0.12</u>	<u>0.06</u>
AVERAGE	7.77	7.15	7.16	0.61	-0.01
AVERAGE ABSOLUTE DEVIATION				1.54	0.83
STANDARD DEVIATION OF DIFFERENCES				1.99	1.12

Table 2
EARLY VERSUS TOTAL SAMPLE RESULTS

<u>Shares - Means</u>	<u>First 24</u>	<u>Total 44</u>
Initial	7.9	7.8
Adjusted	7.5	7.2
Actual	7.3	7.2
 <u>Deviation - Means</u>		
Initial vs. Actual	0.6	0.6
Adjusted vs. Actual	0.2	0.0
 <u>Deviation - Mean Absolute</u>		
Initial vs. Actual	1.5	1.5
Adjusted vs. Actual	0.6	0.8
 <u>Standard Deviation</u>		
Initial vs. Actual	2.0	2.0
Adjusted vs. Actual	1.0	1.1

Table 3
CATEGORY RESULTS

<u>Initial versus Actual Share</u>	<u>HBA (n = 13)</u>	<u>Household (n = 11)</u>	<u>Food (n = 20)</u>	<u>Overall (n = 44)</u>
Mean				
Difference	0.43	0.61	0.73	0.61
Standard				
Deviation of				
Differences	2.08	1.71	2.06	1.99
<u>Adjusted versus Actual Share</u>				
Mean				
Difference	-0.29	-0.15	0.25	-0.01
Standard				
Deviation of				
Differences	1.09	1.23	1.02	1.12
<u>Average Actual Share</u>	7.35	10.14	5.40	7.16

Table 4
PAIRED COMPARISONS ACROSS CATEGORIES

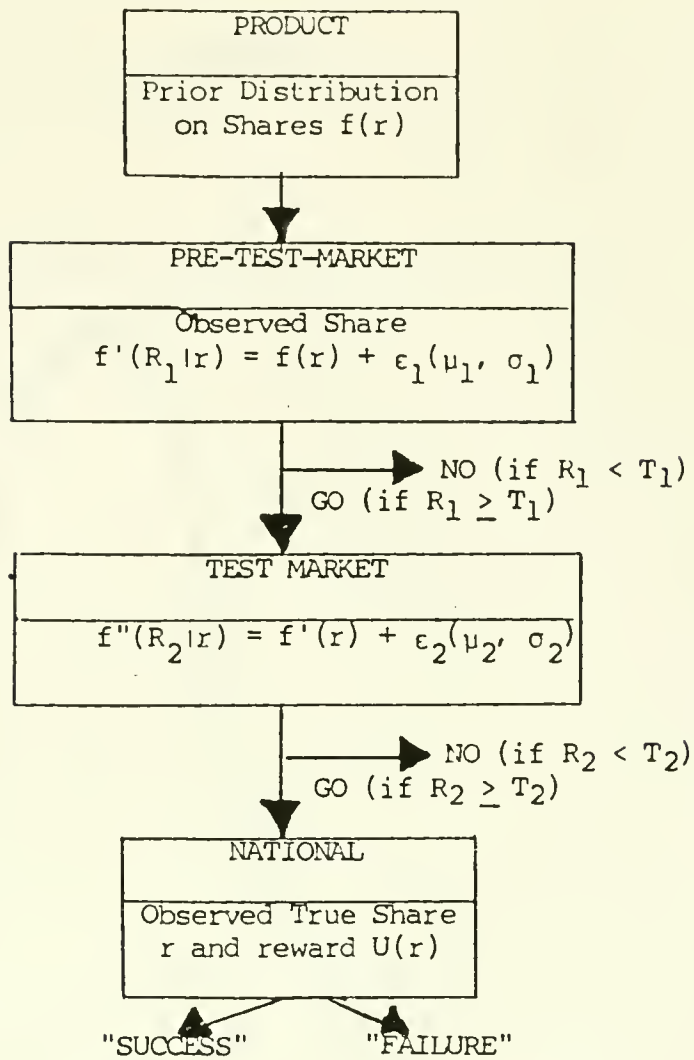
<u>Initial to Actual Share</u>	<u>HBA and Household</u>	<u>HBA and Food</u>	<u>Food and Household</u>
Difference in			
Means (t)	-0.18 (0.2)	-0.30 (0.4)	+0.12 (0.2)
Ratio of Variances			
F (df ₁ , df ₂)	1.47 (12,10)	1.02 (12,19)	1.44(9,10)
<u>Adjusted to Actual Share</u>			
Difference in			
Means (t)	+0.44 (0.9)	+0.04 (0.1)	-0.40 (0.9)
Ratio of Variances			
F (df ₁ , df ₂)	1.27 (10,12)	1.14 (12,19)	1.45 (10,9)

NOTE: at 10% t (df = 11) = 1.36
F (9, 10) = 2.35

is .69 $[(7 + 2 + 1 + 7 + 3(18/35))/(16 + 11) = .69]$. If the "borderline" class is also proportionately split, the proportion is 62 percent $[(7 + 2 + 1 + 7 + 3(18/35) + 1(27/33) + 1(27/33)(18/35))/(16 + 11 + 6(27/33)) = .62]$. Averaging these two estimates, we conclude that approximately two thirds (66%) of the products that pass the ASSESSOR pre-test-market screen subsequently succeed in test market. A. C. Nielsen Co. has systematically studied the success rates for products in test market over the past two decades. The most recent study (1977) included 228 test marketed items. Success was defined by whether or not the new product was subsequently launched nationally. Brands withdrawn from test markets or not launched nationally were considered failures. In this most recent study, they found a 35.5 percent success rate. This is a continuation of a trend toward lower success rates reported in their earlier studies (54.4 percent in 1961 and 46.6 percent in 1971). Other earlier studies by Buzzell and Nourse (1967) and General Foods (1973) suggest higher success rates that are consistent with the Nielsen observations in 1961 and 1971 in analogous years. If the Nielsen trend is correct it is becoming more risky to introduce new products. The success rate of 66 percent for products subjected to pre-test-model analysis and test marketed in the period 1973-1981 is 1.9 times as high as the rate of 35.5 percent observed by Nielsen in 1977.

Table 6 also shows that six products were test marketed despite a negative pre-test evaluation. All of these were subsequently judged by management as "big failures" in test market. However, if the "borderline" and "marginal" classes are allocated as above, we can conclude that 3.8 percent would succeed after failing ASSESSOR $[(1(6/33) + 1(6/33)(18/35))/(6(6/33) + 6) = .038]$. In most cases, after a "NO" forecast at the pre-test stage, a product is dropped and it is not known if it would have succeeded. In the six cases reported in Table 6, however, the products were either already in the test market or the model recommendations were over-ridden by other considerations.

Figure 3
SEQUENTIAL DECISION MAKING*

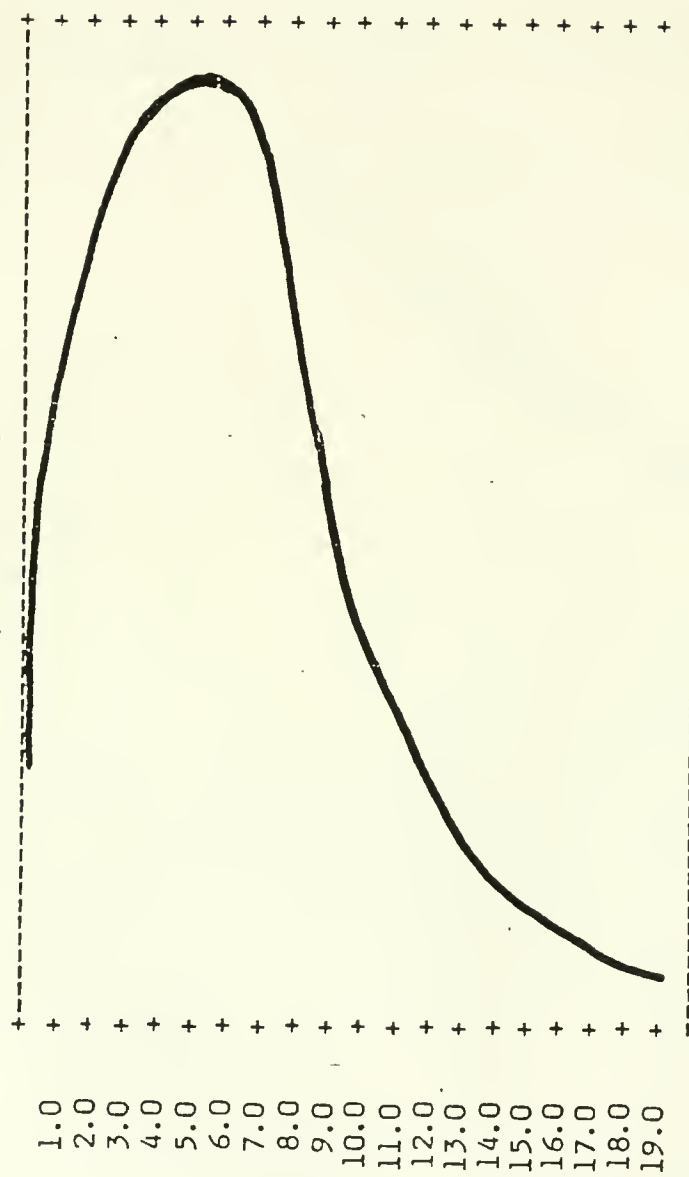


* Note: r = true share

R_1 = observed pre-test forecasted share

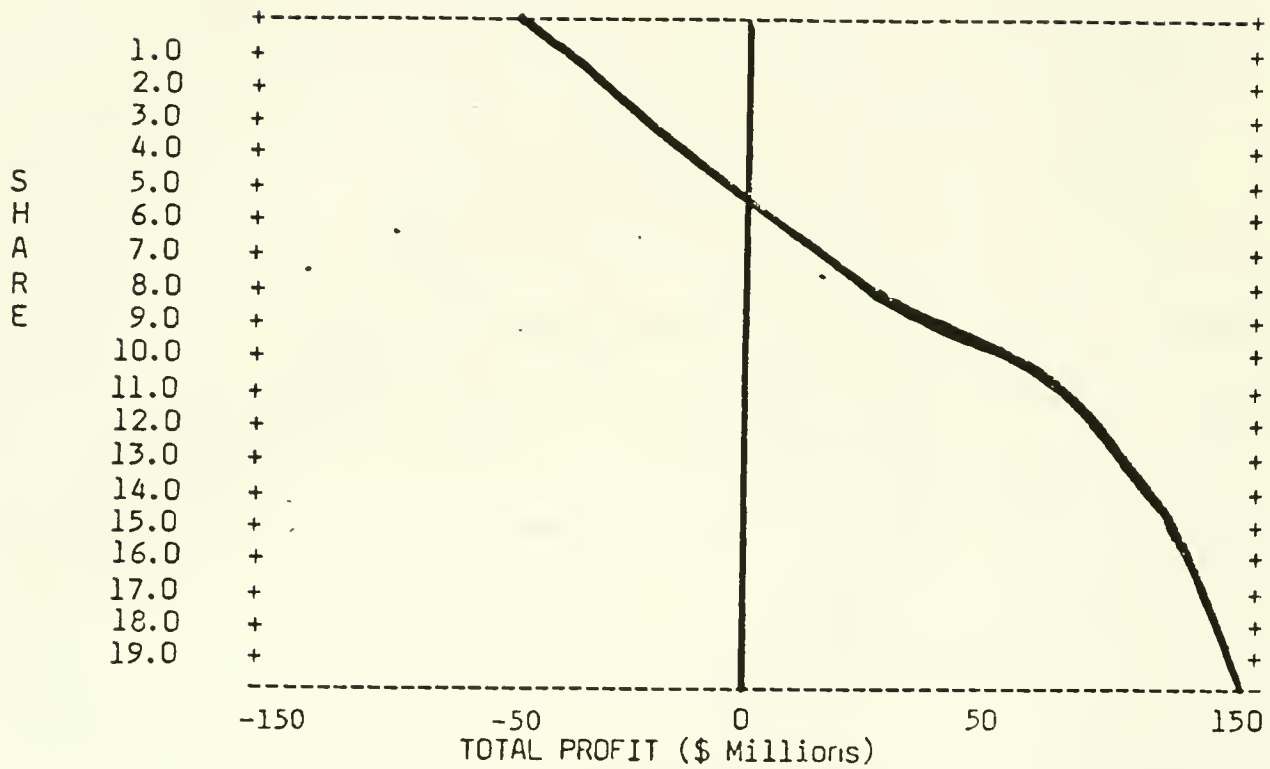
R_2 = observed test market share

Figure 4
Prior Distribution



SHARE

Figure 5
Illustration of Typical Reward Function $U(r)$



A tree of all possible outcomes is generated to enable calculation of these probabilities. See Figure 6. The prior distribution is described by discrete values (r_i) and the probability ($P(r_i)$) for each of these prior true share values. Similarly, the error distributions are made discrete to produce $\epsilon_{1,j}$ and $\epsilon_{2,k}$ values and their associated probabilities $P(\epsilon_{1,j})$ and $P(\epsilon_{2,k})$. For each prior value (r_i) we calculate a set of pre-test-market observations $R_{1,j}$ by adding each possible error value ($\epsilon_{1,j}$) to the r_i value and calculating the conditional probability $P(R_{1,j}|r_i)$. Test market shares are similarly produced by incrementing each associated r_i by the possible error values ($\epsilon_{2,k}$) and calculating the conditional probability $P(R_{2,k}|r_i)$. For each value of true share (r_i), the national reward is then calculated ($U(r_i)$). The tree is usually large. For example of the prior distribution has 40 points, the pre-test error distribution 20 points, and the test market error distribution 10 points, 8000 end point branches will be generated. Once the tree has been generated, the expected values in equation 3 can be calculated given T_1 and T_2 levels. Branches in which $R_{1,j} < T_1$ and $R_{2,k} < T_2$ are eliminated and the remaining probabilities used in equation 3.

We also can use the tree to calculate the probability of: (1) passing pre-test ($R_{1,j} \geq T_1$), (2) passing test market given that we passed pre-test ($R_{2,k} \geq T_2$ given $R_{1,j} \geq T_1$), and (3) the probability of national success ($r_i > \bar{r}$, where \bar{r} is the breakeven value in $U(r)$). Since we have enumerated all outcomes, we also can examine outcomes that were eliminated by the cutoff values. Specifically we can calculate the probability of a national success ($r_i > \bar{r}$) given a product: (1) failed pre-test ($R_{1,j} < T_1$), (2) failed test market after a successful pre-test ($R_{1,j} \geq T_1, R_{2,k} < T_2$), or (3) failed either test ($R_{1,j} < T_1$ or $R_{1,j} \geq T_1, R_{2,k} < T_2$).⁷

⁷Dynamic programming could be used to more effeciently maximize $E(T_1, T_2)$, but it would not allow calculation of these probabilities.

These are important managerial control variables and are meaningful measures of the two kinds of errors that can be made in the sequential decision system.

Typical Product Case

Now consider a typical launch of a major product (e.g. toothpaste, deodorant, aerosol cleaner, or cereal). Assume Figure 5 depicts the profit that would result with each level of national share accomplishment. The costs of test market are taken as \$1,500,000 and the cost of pre-test as \$50,000. The accuracy of the pretest (ϵ_1 (μ_1 , σ_1)) is approximated by the distribution of deviations between initial and actual shares shown in Figure 1, ($\mu = .61$, $\sigma = 1.99$).

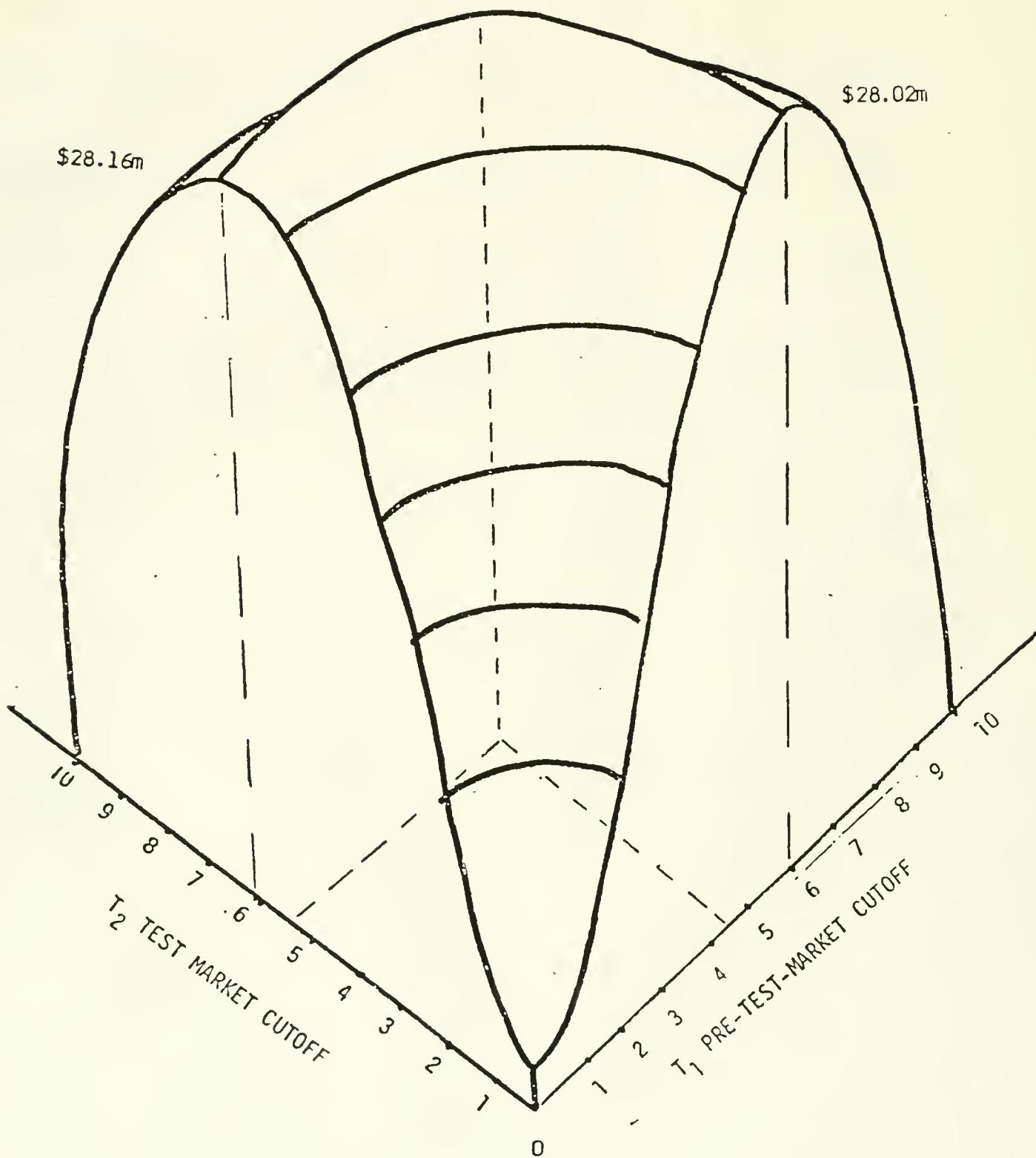
Little published data is available to determine test market accuracy, but in a comparison of first year national shares to test market shares for 50 new brands, Nielsen found 50 percent of the observations within 10 percent of the test market shares. Based on the 7.16 average test market share observed in our study (Table 1), this implies half the time share will be within .7 share points and assuming a normal distribution, the standard deviation of the error distribution is approximately one share point. Gold studied one source of error in test marketing - projecting test area sales to national. He analyzed seven established products in six market areas commonly used for test marketing and projected sales by three common methods. His comparisons of store audit values in the test areas to national sales showed considerable error. In projections based on only one test city, the standard deviation was 32.8 percent of the mean. Based on two test cities the deviation was 21.7 percent and based on three cities the deviation was 15.6 percent. Assuming a mean share of 7.16 (see Table 1), the standard deviation would be 2.3 share points for one test city, 1.6 for two cities, and 1.1 for three cities. The three city value is consistent with the Nielsen data. However, Gold emphasizes that his study only includes part of the total error. He did not include effects such as competitive efforts to distort test markets, the extra attention the sales force may lavish on the new product, or the interval of time between test market and

Figure 7
EXPECTED REWARD SURFACE

\$28.44m

\$28.02m

\$28.16m



TYPICAL PRODUCT RESULTS

	<u>NO TEST</u>	<u>TEST MARKET ONLY</u>	<u>TEST AND PRE TEST</u>	<u>PRE TEST ONLY</u>	<u>HIGH CUT OFF</u>	<u>LOW CUT OFF</u>
Expected Profit	\$16.74m	\$28.16m	\$28.44m	\$28.02m	\$21.46m	\$19.23m
T ₁ (pre-test cutoff)	-	-	4.5	6.0	9.0	1.0
T ₂ (test market cutoff)	-	6.0	5.5	-	9.0	1.0
Probability of Passing Assessor	-	-	69.%	54.%	28.%	94.%
Probability of Passing Test Market	-	47.%	72.%	-	70.%	96.%
Probability of National Success if Go National	49%	94%	91%	82%	99.9%	54%
Probability of National Success if Fail Assessor	-	-	5.%	11.%	31.%	.01%
Probability of National Success if Fail Test Market	-	10.%	5.%	-	10.%	.01%
Probability of National Success if Fail Either Assessor or Test Market	-	10.%	9.%	11.%	37.%	.01%

Table 8
RESULTS WITH PENALTIES FOR DELAY (Millions)

No Testing (no delay)	\$16.7
Pre-Test Only (3 month delay)	\$25.2
Test Market Only (12 month delay)	\$19.2
Pre-Test and Test Market (15 month delay)	\$16.6
Value of Pre-Test Only	\$ 8.5
Value of Test Market Only	\$ 2.5
Value of Test Market Given Pre-Test Market is Done	\$ -8.6
Value of Pre-Test Given Test Market is Done	\$ -2.6

utility" reward function. The expected reward is lowered by over 50 percent, the cutoffs are raised to improve the probability of national success, (95% versus 91% for the base case), and greater risk of eliminating a success is accepted (13% versus 9%). With the tighter screening criteria, fewer products pass pre-test and test market. The expected reward with no testing is -\$12.0 million and the total value of testing is now \$25.54 million ($13.54 - (-12)$). The total value of information for pre-test market is more than doubled relative to the "typical" case (25,070 versus 11,280). The incremental value of information for pre-test given that a test market is done also doubles and the incremental value of test market given that a pre-test market is done increases by 12 percent.

Now consider another variation from the base case. If a market is small in total sales volume, the monetary gains and losses associated with each share will be lower. We simulate this by multiplying the entire reward function by one half. See case 2 in Table 9. The best sequential testing procedure now does not include test market because its added information is not worth the test market cost. The best reward with test and pre-test is \$13.71 million while the reward with pre-test only is \$14.03 million. Even though the cutoff for pre-test is raised from its 4.5 base level to 6.5 more risk of error is accepted than in the "typical" case because the potential losses associated with failure are so much smaller (86% probability of national success instead of 91%).

A similar result is obtained when we simulate a 50 percent reduction in losses, while leaving the gains at the base level (case 3). This could arise if the cost of entry (advertising, promotion, and production) was low. The incremental value of test market is negative (-\$150,000) because the exposure to loss is low, and so the most efficient manner in which to manage such a brand is with a pre-test market analysis only.

achieved, and ensures that the product will not be modified before test market. Case 6 in Table 10 shows the results of improved pre-test accuracy. Relative to the typical case, the reward increases and test market is eliminated by the more accurate pre-test procedure. A higher cutoff is used, so fewer products move past the pre-test stage, but a similar national success rate is observed. The probability of eliminating a success is 1 percentage point above the base case. This simulation shows that the pre-test alone does a very good job in cases where adjustments are not required. This situation may be relatively rare, but it implies that the more accurate the pre-test, the more valuable it is and the more it can be relied upon as a screening procedure.

The "typical product" base case reflected an assumption of high test market accuracy ($\mu_2 = 0$, $\sigma_2 = 1.0$). As indicated earlier, work by Gold suggests that larger errors may be expected. Cases 7, 8, and 9 in Table 10 report evaluations with a mean error (μ_2) of .5 share points and increasing standard deviations for the distribution of differences between test market and true share ($\sigma_2 = 1.25, 1.5$, and 2.0). As the test market error increases, the total expected reward decreases, the incremental value of the pre-test market increases, the incremental value of the test market decreases, and the risk of national failure increases. At a standard deviation of 2.0 , the test market is no longer appropriate. The value of reducing the standard deviation can be calculated by the differences in the expected rewards. See Table 11. If the test market standard deviation were 2.0 , it would be worth \$330,000 to reduce it to 1.0 .

In reviewing Table 10, it is evident that the best testing program relies less on test market as test market errors increase and more on the pre-test market as pre-test errors decrease.

Table 11
VALUE OF INCREASED ACCURACY IN TEST MARKET

$\sigma = 1.25$ to 1.0	\$120,000
$\sigma = 1.5$ to 1.25	\$180,000
$\sigma = 2.0$ to 1.5	\$ 30,000
$\sigma = 2.0$ to 1.0	<u>\$330,000</u>

Managerial Use

The output of the sequential decision model can not be generalized because specific outcomes will depend upon the individual firm's: cost function, margins, investment policy, knowledge of the awareness it can generate, ability to achieve the targeted distribution, test market design, specific category prior distribution, and competitive practices. The model represents a tool for managers. If they can estimate the appropriate prior distribution (e.g. Figure 4), reward function (e.g. Figure 5), pre-test accuracy (e.g. Figures 1 or 2) and test market accuracy for their particular product and category, the sequential decision model can be used to set the best cutoffs and indicate which testing procedures should be used. Managers can examine the probabilities of passing each test, national success, and elimination of potential success, as well as the expected reward in order to evaluate the desirability of various test policies. Allowing the manager to change the probabilities by modifying the cutoffs is an alternative to direct assessment of the manager's utility function and the maximization of his expected utility (see Hauser and Urban for a discussion of some of the difficulties of utility assessment). The model gives decision makers a capability to manage the risks of new product development.

Decision makers must be careful in managing the enthusiasm for particular new products. Brand managers personally may have much to gain and little to lose by pushing their product even in the face of results that are below the cutoff values. The usual persuasion methods after rely on proposing changes that will be made to improve the product, identifying unrealistic conditions in the test market (e.g. competitor's high sampling and promotion), or arguing against the validity of the forecasting models. While these sometimes may be legitimate responses, care must be exercised not to drop the screening criteria

It has been assumed in the above analyses that the reward function is not changed by the test market. However, if the firm cannot produce the product in national volumes at its specified level of quality, the estimated reward obviously will not occur. Before considering elimination of test market, the firm must be sure it can produce the product within the design tolerances. If they are not sure, a test market may be required in any case. Another manner in which the test market can affect the reward function is by improvement of the product based on in-market experimentation and model analysis (see Urban and Hauser). In this case, test marketing may be advisable because it increases the expected reward. An example might be found in case 2 of Table 9, where if the test market improved the brand's profit by more than \$320,000, a test market should be included in the sequential testing system. The advent of pre-test market models implies that test market designs should allocate increasing emphasis to finding the best marketing mix for a product rather than on sales forecasting alone.

ACKNOWLEDGMENTS

Special thanks to Steve Shugan who helped formulate the sequential decision theory model and John Hauser and Leigh McAlister who made valuable criticisms and suggestions. Thanks to Management Decision Systems, Inc. for supplying the data for this study, to Barbara Petroulis, Mary Lyons, and Robert Wallace who made major contributions to its collection, and the cooperating companies who responded to the questionnaire.

BASEMENT

Date Due

6-18-90

Lib-26-67

MIT LIBRARIES



3 9080 00583043 2

