MARKETING SCIENCE
Vol. 14, No. 3, Part 2 of 2, 1995
*Printed in U.S.A.*

# A LATENT LOOK AT EMPIRICAL GENERALIZATIONS

DONALD G. MORRISON AND JORGE SILVA-RISSO
*University of California, Los Angeles*

All empirical data and the resulting parameters are subject to error. In this paper we explicitly use the standard, but hardly profound, "model"

Observed Value = True Score + Error

as a lens for better viewing empirical studies in search of empirical generalizations in marketing. This lens is especially valuable when the unit of analysis is the individual consumer. However, even when a macro study contrasts price elasticities across cities, this $O = T + E$ framework can be very helpful.

**(Unobservable Constructs; Summary Statistics; Appropriate Comparisons; Shrinkage Estimators)**

## Introduction

Having given the concept of empirical generalizations in marketing considerable thought, we are nevertheless unable to give a sufficient condition. Any definitions we construct are either so general as to be meaningless or so situation specific to lack generality. We leave this task of sufficient conditions to our fellow authors. However, we do have what we feel is a solid necessary condition:

### *A Necessary Condition*

Whenever researchers are looking for empirical generalizations in marketing, they should explicitly consider their data as coming from the model:

$$\text{Observed Value} = \text{True Score} + \text{Error}.$$

The empirical generalization in question should then be based on the latent true scores. This necessary condition is especially important when the unit of analysis is the individual consumer. Thus, any empirical generalizations based on how much or when individuals purchase, e.g., all the "20 percent of the customers buy 80 percent of the product" type of generalizations, derived without using the $O = T + E$ lens, are likely to arrive at misleading conclusions, as our examples will illustrate. However, even when large cities are the unit of analysis, our $O = T + E$ framework is very helpful, e.g., comparing price elasticities across 10 major cities.



0732-2399/95/1403/0G61$01.25

*Contribution*

This paper makes one—and only one—contribution. The field of marketing science is maturing. Modeling and estimation issues per se, while important, will not advance the field and influence managerial decision making. We need to tell managers what we have learned. This knowledge will most often be in the form of empirical generalizations. This paper's contribution is:

> Researchers searching for empirical generalizations who use the $O = T + E$ framework will gain insights as they look at the world through a "new pair of eyes."

These new eyes will help speed the development of marketing science, since empirical generalizations are key pillars for the foundation of any science (to paraphrase Frank Bass' Guest Editorial in JMR (Bass 1993)). We will illustrate this point through a few examples: 80/20 concentration laws, NFL field goal kickers (and their relation to brand choice), a brief discussion of the work of Andrew Ehrenberg and his colleagues, and a study of price elasticities across cities.

## 80/20 Concentration Laws

Most readers have probably heard that old *empirical generalization*, "20 percent of the customers account for 80 percent of the purchases." Tables 1 and 2 come from Schmittlein et al. (1993) (hereafter SCM). These tables clearly show some of the problems in directly interpreting observed concentration levels. Table 1 presents the percent of unit sales accounted for by the top 20 percent of households (users and nonusers), for four product categories, as a function of the length of time these households were observed. The data show a substantial and systematic decline in observed concentration as the time period increases.

Table 2 presents the observed concentration among users of the product, defined as those households that happen to have made at least one purchase in the time period analyzed. The pattern here is different from that of Table 1. For catsup and yogurt, concentration increases with time, whereas for soup and detergent there is almost no change. In any event, for these product categories, the top 20% of consumers do *not* account for 80% of total unit sales.

In summary, Tables 1 and 2 show that:

(1) The 80/20 idea is not true;

(2) It matters a great deal whether or not the nonusers are counted;

TABLE 1

*The Percent of Total Purchases Due to the Top 20% of Entire Population*

| | Top 20% of the Entire Population | | | |
|---|---|---|---|---|
| Time Period of Observation | Catsup | Detergent | Yogurt | Soup |
| 1 month | 86 | 65 | 97 | 56 |
| 3 months | 63 | 55 | 85 | 49 |
| 6 months | 56 | 52 | 80 | 49 |
| 9 months | 54 | 50 | 78 | 48 |
| 1 year | 53 | 50 | 78 | 46 |
| 2 years | 52 | 49 | 73 | 45 |
| Average Number of Purchases per Household per Year | 4 | 11 | 16 | 49 |

$N$ = 3836 households.

Source: Schmittlein et al. (1993).

TABLE 2

*The Percent of Total Purchases Due to the Top 20% of Users*

| Time Period of Observation | Top 20% of Users | | | |
|---|---|---|---|---|
| | Catsup | Detergent | Yogurt | Soup |
| 1 month | 32 | 44 | 49 | 48 |
| 3 months | 42 | 47 | 58 | 47 |
| 6 months | 46 | 48 | 62 | 48 |
| 9 months | 48 | 47 | 63 | 48 |
| 1 year | 49 | 47 | 65 | 46 |
| 2 years | 50 | 48 | 65 | 45 |
| Average Number of Purchases per Household per Year | 4 | 11 | 16 | 49 |

$N = 3836$ households.
Source: Schmittlein et al. (1993).

(3) When nonusers are included, the concentration goes down as the observation time increases. The reverse is typically true when nonusers are excluded.

The following is a general model for developing a probability distribution of the amount purchased across individuals (or more commonly, households) that allows for use of our framework:

$$P(X = x) = \int P(X = x \mid \lambda) g(\lambda) d\lambda,$$

where

$P(X = x)$ = the *observed* distribution of purchases across individuals.

$P(X = x \mid \lambda)$ = the *conditional* distribution of purchases made during the period given the individual purchasing *rate* $\lambda$.

$g(\lambda)$ = the *unobservable* mixing distribution of purchasing *rates* across individuals.

This distribution $g(\lambda)$ can be continuous, can be discrete, can contain mass points and continuous components, and so forth (Robbins 1977, Morrison and Schmittlein 1988). Therefore, this mixing-distribution approach is not restricted to a parametric specification of $g(\lambda)$. Tables 1 and 2 are based on the observable histogram representing $P(X = x)$. However, $g(\lambda)$ is what is "true" and will remain invariant over observation times of differing lengths. Thus, all "laws" for concentration should be based on the unobservable mixing distribution $g(\lambda)$. SCM show how to do this within the NBD model when the conditional probability distribution $P(X = x \mid \lambda)$ is Poisson and the mixing distribution $g(\lambda)$ is gamma.

If the observation time is long enough *and* the consumers retain the same purchasing rate, then eventually "what you see" ($P(X = x)$) is "what you get" ($g(\lambda)$). However, these two conditions, long observation and stationary behavior, rarely are satisfied for the typical $P(X = x)$-based empirical concentration statistics. SCM show how to "back out" the unobserved $g(\lambda)$ from the observed $P(X = x)$. They also have a method for incorporating long-term, "hard core" nonbuyers. Because the mechanics have been presented elsewhere, the only message we wish to convey here is that if empirical generalizations on purchasing concentration are sought, then they should be based on the unobservable mixing distribution on the rates, $g(\lambda)$, and not on the observed amount purchased.

## NFL Field Goal Kickers: Skill or Luck?

The following is an even more tantalizing example than the 80/20 concentration laws. Morrison and Kalwani (1993) analyze how different in skill NFL kickers are. Table 3 shows statistics on field goals for the period 1989–1991. We notice that in the three years the "best" kicker has a percentage of successful kicks greater than 90%, while the "worst" is below or close to 60%. Furthermore, in the three years kickers in the bottom quartile are under 70%, while the ones in the top quartile are close to or over 80%. The data seem to tell us that there are clear skill differences. However, aggregating the data across the three years, those differences shrink.

Morrison and Kalwani (1993) construct a probability model for the performance of field goal kickers. Our previous example shows that the 80/20 law is not even true on observed data. For NFL kickers, however, there are "empirical" differences that stay "constant" across the years. Some candidates for "empirical generalizations" would be: (a) the percentage of successes for the bottom quartile of NFL kickers is under 70%, (b) the percentage of successes for the top quartile is close to 80%, (c) the best kicker has a percentage above 90%, (d) the percentage for the worst kicker is close to or below 60%, and (e) on average 74% of all field goal attempts are successful.

The Morrison and Kalwani (1993) results show that all these "empirical generalizations" are highly misleading. An analysis of the data through the $O = T + E$ framework reveals that the basis for such differences is the random component "$E$" and not true skill differences "$T$." In other words, the data are consistent with no differences in true skill, and higher and lower percentages due to positive or negative random errors.

This field goal study is an interesting example of how observed data may lead to conclusions and actions based only on random noise. The similarities of this football example with brand choice are apparent. Just substitute probability of buying a particular brand in a category shopping occasion for the probability of making each kick. Therefore, marketing scholars looking for empirical generalizations for brand switching or quantity purchasing behavior without using the $O = T + E$ framework have at least the potential risk of coming up with very misleading empirical generalizations.

## Ehrenberg, the Dirichlet and Empirical Generalizations

The marketing science community owes a big debt to Andrew Ehrenberg and his colleagues. While many of us were overly busy modeling, Ehrenberg et al. were collecting and organizing vast amounts of empirical data on consumer purchasing behavior. The key findings are presented elsewhere in this Special Issue and will not be repeated here.

TABLE 3
*NFL Field Goals Data*

| | 1989 | | | 1990 | | | 1991 | | | 1989–1991 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FG | FGA | PROP | FG | FGA | PROP | FG | FGA | PROP | FG | FGA | PROP |
| Best | 20 | 21 | 0.952 | 34 | 37 | 0.919 | 17 | 17 | 1.000 | 83 | 100 | 0.830 |
| Top 25% | 22 | 28 | 0.786 | 20 | 25 | 0.800 | 25 | 31 | 0.806 | 20 | 26 | 0.769 |
| Bottom 25% | 23 | 34 | 0.676 | 20 | 29 | 0.690 | 13 | 19 | 0.684 | 14 | 20 | 0.700 |
| Worst | 14 | 24 | 0.583 | 15 | 24 | 0.625 | 14 | 28 | 0.500 | 14 | 24 | 0.583 |
| Average | | | 0.732 | | | 0.748 | | | 0.737 | | | 0.740 |

Source: Morrison and Kalwani (1993).

The main model for the Ehrenberg work is the multinomial-Dirichlet (usually shortened to simply the Dirichlet). The first component of the model is the same as in our previous section on concentration.

(1) Consumers make purchases in the product category in a Poisson manner with an underlying rate $\lambda$.

(2) $\lambda$ is distributed gamma across consumers.

The remaining assumptions (in the spirit of the NFL kickers example) are:

(3) Given a purchase in the category, consumers are multinomial (zero order) processes with a probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ for choosing among the $k$ brands.

(4) The $\mathbf{p}$ vectors are distributed Dirichlet (multivariate beta) across consumers with distribution $f(\mathbf{p})$.

(5) $\lambda$ and $\mathbf{p}$ are independent.

If we let $n$ equal the number of purchases and $(x_1, x_2, \ldots, x_k)$ be the number of purchases of each of the $k$ brands (obviously $n = \sum_{i=1}^{k} x_i$), then we can characterize empirical generalizations by whether they come from the observed distribution of $n$s and $x_i$s or the unobservable $g(\lambda)$ and $f(\mathbf{p})$.

*The Ehrenberg School*

To the best of our knowledge all the law-like empirical generalizations produced by Ehrenberg and his colleagues come from observable $n$s and $x_i$s. Excellent discussions are found in Ehrenberg (1972), Ehrenberg (1975), Ehrenberg and England (1990), Ehrenberg et al. (1990), and Ehrenberg and Uncles (1993). An informative example is the well-known relationship between purchasing rate and market share:

$$w_i(1 - b_i) = \text{constant across all brands } i, \text{ where}$$

$$w_i = \text{the amount bought per buyer of brand } i,$$

$$b_i = \text{the proportion buying brand } i.$$

This is a very useful relationship. Companies who think they can develop a very strong "niche" brand do so at their peril. A niche brand has, by definition, a small $b_i$ which according to the empirical generalization causes $w_i$ to be *lower* than the typical brand in the category. Those companies planning for a *higher* than average $w_i$ with their small $b_i$ niche brand are bucking the odds more than they think. A good example of this is CPC International's failed attempt to establish Knorr soups as a strong niche brand in the United States. They easily achieved their $b_i$ penetration goal but never came close to reaching the targeted $w_i$, amount bought per buyer, objective.

*Brand Choice*

The Ehrenberg School starts with the Dirichlet framework, but again to the best of our knowledge, all their empirical generalizations are based on the observed $n$s and $x_i$s. This certainly makes these generalizations easier to calculate and even easier to communicate. Further, as shown above, the $w(1 - b)$ finding has real implications for brand strategies. Nevertheless, as our earlier discussion on concentration shows, basing findings on $n$s and $x_i$s as opposed to the unobservable mixing distributions $g(\lambda)$ and $f(\mathbf{p})$ can be very misleading. The SCM approach to 80/20 analysis seems to be unique (and much more complicated) for concentration studies by focusing directly on $g(\lambda)$.

The brand choice component of the Dirichlet is another matter. In quite similar modeling approaches, Bass et al. (1976), Sabavala and Morrison (1977), Kalwani and Morrison (1977), and Jeuland et al. (1980) develop the same latent measure of polarization for the unobservable brand choice probabilities.

*Observable vs. Latent*

When searching for empirical generalizations derived from individual purchasing histories, should we use the observable $n$s and $x_i$s or the unobservable mixing distributions $g(\lambda)$ and $f(\mathbf{p})$? As with all such questions the answer is "It depends!" There is an ever-present danger in using $n$s and $x_i$s. This is especially true when seeking empirical generalizations across categories with very different mean purchasing rates, variability of these rates, and/or lengths of observation. Tables 1 and 2 show just how dramatic these effects can be.

On the other hand, using $g(\lambda)$ and $f(\mathbf{p})$ can be very complicated when we deviate from the baseline NBD and Dirichlet models. Further, these latent approaches are very difficult to communicate to anyone not comfortable with probability mixing models, e.g., some researchers and virtually all managers.

Our advice when developing and interpreting the Ehrenberg School type of empirical generalizations is to think explicitly about:

$$\text{Observed Purchase Amount} = \text{True Rate} + \text{Error},$$

or

$$\text{Observed Share of Purchases} = \text{True Share} + \text{Error}.$$

In particular, are the interesting deviations from these laws "real" or due to a systematic error bias? An illustrative example of this approach is Kahn et al. (1988) where they looked at $w(1 - b)$ with the above advice. The Ehrenberg $n$s and $x_i$s based model was very "useful" and "worked" to detect "interesting" brands. However, the explicit $g(\lambda)$ and $f(\mathbf{p})$ analysis used by Kahn et al. (1988) discovered a market share effect which (at least to them) enhanced the interpretation of niche and variety seeking brands.

## Price Elasticities—Just How Different?

By now the reader may be thinking, "Sure, the $O = T + E$ approach makes sense for the NBD and Dirichlet-type empirical generalizations where the basic theory is captured by a probability mixing model. But what about the more important macro issues of marketing mix effects?" Consider the sets of price elasticities in Tables 4 and 5 from a Wittink et al. (1988) working paper, which was kindly provided to us by Dick Wittink.

These own price elasticities in Table 4 for the three brands and the ten geographical price elasticities in Table 5 are *measured* elasticities. Thus we have:

$$\text{Measured Elasticity} = \text{True Elasticity} + \text{Error}.$$

If we assume that the error terms are independent across brands and cities (an appealing assumption) and of equal size (in Table 4 the three standard errors of the estimated own brand elasticities are 0.10, 0.10, and 0.11), then we have the key result that the observed variance across brands (Table 4) or cities (Table 5), $\sigma^2_{OBS}$, is equal to the variance of

TABLE 4

*Own-Price Elasticities in Chicago*

| | |
|---|---|
| Star Kist | −3.30 |
| Chicken of the Sea | −3.62 |
| Bumble Bee | −4.19 |
| Mean = | −3.70 |
| Variance = | 0.20 |
| Standard Deviation = | 0.45 |

TABLE 5
*Price Elasticity for Star Kist*

| | |
|---|---|
| Boston | −2.53 |
| Chicago | −3.30 |
| Houston | −1.51 |
| Indianapolis | −1.42 |
| Jacksonville/Orlando | −1.30 |
| Kansas City | −1.93 |
| Los Angeles | −3.19 |
| New York | −2.67 |
| San Francisco | −1.90 |
| Seattle | −2.52 |
| Mean = | −2.22 |
| Variance = | 0.52 |
| Standard Deviation = | 0.72 |

the true scores *across* brands or cities, $\sigma_T^2$, plus the *within* variance of the error term, $\sigma_E^2$.

$$\sigma_{OBS}^2 = \sigma_T^2 + \sigma_E^2.$$

The observed variance (admittedly based on a sample of size 3) in Table 4 is 0.20. The error variance is 0.01 (i.e., $(0.10)^2$), making the true variance 0.19. Does the reader feel that 95 percent of the observed variance across brands is "real" while only five percent is measurement "noise"? We shall return to this later.

The standard errors of the city elasticities are not in Wittink et al., but they come from the same kinds of data as the Chicago results in Table 4. In a personal conversation Wittink told us the standard errors for Table 5 were "about the same" as in Table 4. So assuming this the observed variance, $\sigma_{OBS}^2 = 0.52$ is made up of the error variance, $\sigma_E^2 = 0.01$ and the residual true variance, $\sigma_T^2 = 0.51$. Does the reader feel uneasy with only two percent of the observed variance being estimation noise with a whopping 98 percent being "real" variance across cities?

### *Forcing Tough Questions*

Table 4 has so few entries that the estimated observed variance is a little shaky. However, the 10 geographical price elasticities in Table 5 range from −1.30 to −3.30. (The mean of −2.22 is "close" to Ehrenberg's generalization that frequently purchased package goods have elasticities of about −2.6.) However, it is hard to believe that the Chicago price elasticity for Star Kist tuna is 2.5 times greater that in Jacksonville/Orlando.

### *Are the Coefficients that Good?*

Most marketing mix coefficients (or their transformations) are estimated by maximum likelihood methods. The standard errors reported are usually based on asymptotic theory, i.e., on the curvature of the likelihood function at the maximum. In some cases, the sample size could be too small for the asymptotics to kick in. This is not typical given the large scanner data sample sizes. However, a perhaps more troubling situation exists.

In Schmittlein et al. (1987) (hereafter SMC) a maximum likelihood procedure was used to estimate the four parameters of the "Counting Your Customers" model. When using both real and simulated data the bootstrapped estimate of standard errors were typically *three times greater* than the estimates based on asymptotic theory. Or in our $O = T + E$ framework the error variance was *nine times larger* for the bootstrap vs.

asymptotic method. A personal conversation with Brad Efron, the originator of the bootstrap, indicated that these SMC (1987) results were "nonatypical."

*Generalizing Across Brands, Categories and Geography*

If marketing science is to advance, then some generalizations on the effects of marketing mix variables across brands, product categories, and geographical regions are vital. Thus, more work like Wittink et al. needs to be done *and published*. However, since all *measured* coefficients have error, we need to adjust for these measurement errors and potential biases. Explicitly thinking in the $O = T + E$ manner will help us do the appropriate adjustments. What we are really arguing is that some form of "shrinkage" estimators (empirical Bayes methods come in many slightly different "flavors") are required. See Blattberg and George (1991) and Rossi and Allenby (1993) for good examples of shrinkage estimators within the marketing mix arena.

The spirit of these shrinkage estimators is:

$$\text{Shrunken Estimate} = \alpha[\text{Estimate from the data}] + (1 - \alpha)[\text{Baseline Value}].$$

How much weight ($\alpha$) you put on the data is always a matter of conjecture. The appropriate "Baseline Value" has subjective components. Nevertheless, we feel that even with large sample size scanner data studies, some form of shrinkage estimation will help in the search for empirical generalizations. The data per se never tell the whole story. See Vanhuele et al. (1995) for an example of this in duration time modeling.

## Discussion

Useful generalizations can only come from empirical studies. We hope that we have now persuaded the reader to think in terms of

$$\text{Observed} = \text{True} + \text{Error},$$

when interpreting these results. For the Ehrenberg School of empirical generalizations this means explicitly considering the unobservable mixing distribution of purchasing rates $g(\lambda)$ and the equally unobservable mixing distribution on choice probability vectors, $f(\mathbf{p})$. Fader and Schmittlein (1993) is an excellent example of doing this. By explicitly using $g(\lambda)$ and $f(\mathbf{p})$ they develop new insights into Ehrenberg's Double Jeopardy result.

The 80/20 law and the NFL kickers examples illustrate that, if the underlying process that generates the observed data is not explicitly considered, erroneous empirical generalizations could be derived. We choose these two examples because they capture the dynamics of two important consumer's decisions: quantity and choice. For instance, the NFL kickers example clearly shows how easy it is to conclude that kickers widely differ in skill (or consumers in choice probabilities) when that is not the case. Marketing managers who ignore the fact that the data they observe are generated by a true component and random error are likely to base their decisions on overreaction to random events, as Morrison and Kalwani (1993) show is the case for decisions to hire and fire kickers.

It could be argued that "empirical generalization" means generalization based on empirically observed phenomena, and hence by definition ought to be based on "observed data." We do not disagree with this point. What we are saying is that in order to make valid empirical generalizations it is necessary to explicitly consider the data as coming from the model: $O = T + E$. In the examples we present the estimates of the parameters of the *underlying* probability distributions are obtained from the histogram of the *observed* data. In other words, we want to make clear that we are not proposing not to use observed data. We are just presenting a framework that has to be considered explicitly when deriving empirical generalizations from observed data.

Furthermore, when looking for empirical generalizations on marketing mix effects appropriate "shrinkage" may cause some generalizations to pop out that otherwise are lost in the error noise. This "noise" may also be considerably bigger than indicated by the standard errors derived (say) by asymptotic theory. Do any readers feel that 98 percent of the observed large variance in the price elasticities in Table 5 is "real"? Our $O = T + E$ model says the 98 percent figure is real if:

(1) We believe the standard errors, and

(2) We feel there are no systematic biases due to unobservable factors across these cities.

Our $O = T + E$ framework forces us to ask such questions and consider these assumptions.

These are leading questions that obviously would not be admissible in court. However, we will attempt to give our answer. Error estimates based on asymptotic properties come from a likelihood function that assumes an explicit functional form which is obviously not exactly correct. Therefore, it would not be surprising that error estimates are biased downwards due to model misspecification.

Furthermore, when we are comparing geographical areas it is very likely that we are omitting key variables that differ over those areas. The substantial difference in couponing activity across cities is one such example. We have no way of quantifying these distortions, but it is our opinion that a literal reading of the Wittink et al. results will vastly overstate the true variability of price elasticities across the ten cities.

Two overriding questions for any shrinkage estimator are: how much to shrink, and to what? Above we talk about how much to shrink. "Shrink to what" is in the spirit of any meta-analysis design. The solution involves which studies to include and how to weight them. We have no unambiguous answers to these questions. This ambiguity is reinforced by quoting Ingram Olken, a Stanford statistician. When asked about meta-analysis he said: "To do a meta-analysis is easy; now, to do it well . . . , that's a whole different issue."

In conclusion, the $O = T + E$ framework forces us to consider the latent rates and probability vectors for individual level marketing behavior. For macro studies, such as Wittink et al. (1988), $O = T + E$ focuses our attention on how to compare results within a study (e.g., the ten city elasticities) and to what extent we should adjust their results on the basis of similar studies done by others.

The reader who has stuck with us this far has discovered no new empirical generalizations in marketing. However, our objective will be met if this same reader now has a better understanding of how to construct his or her own empirical generalizations or to interpret the empirical generalizations proposed by others.

## References

Bass, F. M. (1993), "The Future of Research in Marketing: Marketing Science," *Journal of Marketing Research*, 30 (February), 1–6.

——, A. P. Jeuland, and G. P. Wright (1976), "Equilibrium Stochastic Choice and Marketing Penetration Theories: Derivations and Comparisons," *Management Science*, 22, 1051–1063.

Blattberg, R. C. and E. I. George (1991), "Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations," *Journal of the American Statistical Association*, 86 (June), 304–315.

Ehrenberg, A. S. C. (1972), *Repeat Buying, Theory and Applications*, Amsterdam: North Holland Publishing Company.

—— (1975), *Data Reduction: Analyzing and Interpreting Statistical Data*, Chichester, UK: John Wiley & Sons, Inc.

—— and L. R. England (1990), "Generalizing a Pricing Effect," *The Journal of Industrial Economics*, 39 (1), 47–68.

——, G. J. Goodhardt, and T. P. Barwise (1990), "Double Jeopardy Revisited," *Journal of Marketing*, 54, 82–91.

—— and M. D. Uncles (1993), "The Dirichlet Model: A Positive Critique," Working Paper.

Fader, P. S. and D. C. Schmittlein (1993), "Excess Behavioral Loyalty for High-Share Brands: Deviations from the Dirichlet Model for Repeat Purchasing," *Journal of Marketing Research*, 30 (November), 478–493.

Jeuland, A. P., F. M. Bass, and G. P. Wright (1980), "A Multi-Brand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes," *Operations Research*, 28, 255–277.

Kahn, B. E., M. U. Kalwani, and D. G. Morrison (1988), "Niching Versus Change-of-Pace Brands: Using Purchase Frequencies and Penetration Rates to Infer Brand Positionings," *Journal of Marketing Research*, 25 (November), 384–390.

Kalwani, M. U. and D. G. Morrison (1977), "A Parsimonious Description of the Hendry System," *Management Science*, 23 (January), 467–477.

Morrison, D. G. and M. U. Kalwani (1993), "The Best NFL Field Goal Kickers: Are They Lucky of Good?" *Chance*, 6, 3.

Robbins, H. (1977), "Prediction and Estimation for the Compound Poisson Distribution," *Proceedings of the Academy of Sciences*, U.S.A., 74, 2670–2671.

Rossi, P. E. and G. M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30 (May), 171–182.

Sabavala, D. J. and D. G. Morrison (1977), "A Model of TV Show Loyalty," *Journal of Advertising Research*, 17 (December), 35–43.

Schmittlein, D. C., L. G. Cooper, and D. G. Morrison (1993), "Truth in Concentration in the Land of (80/20) Laws," *Marketing Science*, 12 (Spring), 167–183.

——, D. G. Morrison, and R. Colombo (1987), "Counting your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (January), 1–24.

Vanhuele, M., M. G. Dekimpe, S. Sharma, and D. G. Morrison (1995), "Probability Models for Duration: The Data Don't Tell the Whole Story," *Organizational Behavior and Human Decision Processes*, 62 (April), 1–13.

Wittink, D. R., M. J. Addona, W. J. Hawkes, and J. C. Porter (1988), "Scan*Pro: The Estimation, Validation, and Use of Promotional Effects Based on Scanner Data," Working Paper.