Planning and evaluating consumer promotions is facilitated by knowledge of the types of consumers who contribute to incremental sales. In particular, interest may focus on identifying the contributions of buyers segmented on the basis of their prior purchase history. When the distribution of the number of purchase occasions in a base period can be described by the negative binomial distribution (NBD), conditional trend analysis (CTA) is a simple and effective approach for identifying the sources of incremental sales during a test (promotional) period. As currently implemented, CTA assumes a stationary marketing environment. The authors propose an extension of CTA that explicitly incorporates varying marketing activities. They also show that the often observed underprediction of purchases in the test period by nonbuyers in the base period is a consequence of the skewness of the NBD and is not necessarily due to model misspecification. An illustration with scanner panel data is provided.

Nonstationary Conditional Trend Analysis: An Application to Scanner Panel Data

Suppose Tide steps up marketing activity in Atlanta. Aggregate sales increase. Is this increase attributable to the marketing activity or is it simply the result of random variation? Further, if the increase is attributable to the activity, what types of consumers contributed the most to the increase? Customers who do not habitually buy Tide, those who buy Tide infrequently, and frequent buyers are all potential sources of the increment. Each of these groups has different implications for the success of the activity.

To pinpoint the source of additional sales, we must predict what each group of customers would have purchased in the absence of the promotion. A simple prediction scheme uses each customer’s purchase rate in the prior period, the base period, as the expected sales rate in the promotional period, the test period. This prediction is misleading, however, because of the well-documented phenomenon of “regression to the mean” (Morrison and Schmittlein 1981). If customer purchase patterns are random, light or nonbuyers in the base period are expected to buy more in the following test period; similarly, heavy buyers are expected to buy less. More accurate answers are obtained by conditional trend analysis (CTA), a technique introduced by Goodhardt and Ehrenberg (1967) and extended by Morrison (1968). CTA is based on the negative binomial distribution (NBD) model of consumer purchase behavior (Ehrenberg 1959), which assumes that

— a household’s purchase occasions follow a stationary Poisson process with a constant rate and
— purchase rates are heterogeneous across households in a market. This heterogeneity is characterized by a gamma (mixing) distribution.

Under these assumptions, CTA predicts the expected number of purchases a household would make in a test period given the number of purchases in a base period. Sources of incremental business are identified by comparing the expected and actual purchases.

Though the NBD model and CTA have worked well in a variety of applications, two problems have been noted (Morrison and Schmittlein 1981):
1. The NBD tends to underpredict test period purchases by the zero class, the group of customers who bought nothing in the base period. This underprediction can be a serious problem as it leads to an overstatement of one of the key goals of marketing effort—attracting previous nonbuyers to the brand.

2. For accurate evaluation of the impact of marketing efforts in the test period, the base period must differ from the test period only in terms of those efforts.

Morrison and Schmittlein (1988) hypothesize that the stationary Poisson assumption causes the first problem and that the explicit introduction of marketing variables in the NBD framework is necessary to resolve the second problem. Because changes in marketing activity are a major cause of nonstationarity in purchase rates, the recommended introduction of marketing variables is also a call to relax the stationarity assumption of the NBD. Indeed, Sabavala (1988) remarks on the difficulty of finding a base period that differs from the test period only in the marketing effort of interest.

How serious is the nonstationarity? A convenient way to diagnose the presence of nonstationarity is with a control chart constructed for aggregate weekly sales. If individual household purchases are Poisson, aggregate sales—the superposition of the individual Poisson processes—are also Poisson. Hence, the appropriate control chart is the c-chart (Ryan 1989, p. 196–201).

Figure 1 is a graph of the number of weekly purchase occasions for two series: powder laundry detergents and Tide. Superimposed on each series is a c-chart. The center line is the mean weekly number of purchase occasions, and the control limits are plus and minus three times the square root of the weekly mean. Approximately 99.7% of the weekly purchases should be within the control limits if the process is stationary. Hence, the c-charts in Figure 1 indicate considerable nonstationarity in the series. Ignoring nonstationarity in consumer purchases when the observation periods are one week will seriously affect forecasting performance.

Part of the variation in purchase occasions may be explained by marketing variables and seasonal effects. Figure 2 is time plots of an aggregate measure of weekly advertising by the stores where the purchases were made. The plots demonstrate that weekly advertising by the stores is not constant over time. Moreover, the correlation of the logarithm of number of weekly purchase occasions and weekly advertising is .37 for powder detergents and .56 for Tide. Thus, weekly advertising is a potential predictor variable.

We hypothesize that nonstationarity is manifested as a multiplier of the household purchase rate. Hence, higher (lower) than usual marketing efforts would increase (decrease) the purchase rate of all households by the same factor. Using this formulation, we show that there is a simple relationship between the stationary and nonstationary forecasts: the nonstationary forecast is the stationary forecast multiplied by the ratio of the mean purchase rates in the test period to those in the base period. This procedure is similar to the traditional practice of adjusting the stationary forecasts by multiplying them by the ratio of the total number of purchases in the test period to those in the base period \textit{ex post} (Morrison and Schmittlein 1981). Modeling nonstationarity—explicitly including the impact of marketing, seasonal, and other variables in the purchase rate—enables us to make this adjustment \textit{ex ante}. Thus, we can forecast and compare the impact of marketing activities, customer segment by segment, rather than diagnose what happened after the fact.

Though introducing predictor variables improves the accuracy of the forecasts, underprediction persists in the nonstationary NBD. We show that underprediction is due, in part, to the skewness of the NBD. For distributions

---

1. The source of the data is the MSI Library supplied by A. C. Nielsen. The data consist of the purchase history for a panel of households in Springfield, Missouri, for 138 weeks from the beginning of 1986 to the thirty-fourth week of 1988.

2. The purchase occasions for Tide are for the 42-ounce package, which is the leading universal product code, accounting for 17.6% of total purchase occasions.
that are symmetric about the mean, such as normal distributions, the expected positive forecast errors are equal to the expected negative forecast errors. However, for positively skewed distributions, such as the negative binomial distribution, the expected positive forecast errors are greater than the expected negative forecast errors. Hence, the observed underprediction may not imply model inadequacy as it would for a linear model with normal errors.

Another major assumption of the NBD model is the gamma mixing distribution. Robbins (1977) relaxes the gamma assumption and analyzes the problem without specifying the distribution of the households’ purchase rates while maintaining the assumption that events follow a stationary Poisson process. His approach leads to a remarkably simple estimation and forecasting procedure. We show that nonstationarity affects Robbins’ model in the same way that it affects the NBD model. This result is intuitive after one realizes that the stationarity assumption applies to the underlying Poisson process, which is a common element to both models, and not to the distributional assumptions for the purchase rates, which differentiate the two models.

Related Literature

Previous efforts have been made to generalize the NBD model by relaxing the assumptions and including marketing and other predictor variables. Though a comprehensive literature review is beyond the scope of this article, representative research includes that of Gupta (1991), Morrison and Schmittlein (1981, 1988), Wagner and Taudes (1986), and the authors cited in their articles.

One research stream has focused on the interpurchase time implications of the NBD. For example, Chatfield and Goodhardt (1973), Morrison and Schmittlein (1981, 1988), and Schmittlein and Morrison (1983) consider a stationary process in which the interpurchase times follow an Erlang distribution of order two instead of the exponential distribution implied by a stationary Poisson process. In this article, we consider nonstationary Poisson processes, which can result in interpurchase times that are more regular than the exponential distribution (Lenk and Rao 1990). Another approach, by Vilcassim and Jain (1991), describes the times between a household’s purchases with general, nonstationary stochastic processes. Their model provides a detailed analysis of individual households, whereas CTA segments households by their purchasing behavior in the base period.

A second stream has focused on the incorporation of marketing variables. Wagner and Taudes (1986) incorporate heterogeneity and nonstationarity by assuming that (1) a household’s choice behavior follows a zero-order process, (2) a household’s purchase times can be described by a Poisson process, and (3) a household’s purchase rate is heterogeneous across the population, varies with time, and is a function of independent variables. They obtain a logistic regression model for a household’s choice probabilities, generalized Dirichlet distributions for aggregate purchase probabilities, and a multivariate Polya process (negative multinomial distribution) for the aggregate number of purchases. Gupta (1991) models purchase occasions with a nonstationary Poisson process with heterogeneity described by a gamma mixing distribution to obtain the nonstationary, negative multinomial distribution for the aggregate number of purchase occasions in disjoint time intervals. Our model is similar to those of Wagner and Taudes (1986) and Gupta (1991). We focus on the impact of nonstationarity on CTA and the resulting bias due to falsely assuming a stationary process.

Brand choice models provide a fundamentally different method for evaluating the effectiveness of consumer promotions. Guadagni and Little (1983) used random utility theory to derive a logistic regression model for the probability of brand choice at a purchase occasion. This model is designed to predict a household’s brand choice at a purchase occasion and not the total number of purchase occasions during a time period, which is the goal of CTA. Though Guadagni and Little introduce a time component in their logistic regression by using past purchases as an independent variable, brand choice models
are inherently cross-sectional, whereas CTA is a time series model. Additionally, standard logistic regression models do not explicitly consider heterogeneity in purchase behavior. One extension by Allenby and Lenk (1992) proposed a logistic regression model with random coefficients to describe heterogeneity in households' responses to marketing variables.

These and other extensions obtain their applicability at the expense of the simplicity and intuitive appeal of the NBD. We retain the structure of the NBD yet give the model greater flexibility by introducing predictor variables through the rate parameter or intensity function of the Poisson processes. This approach maintains the probabilistic structure of the NBD while providing a simple method to incorporate predictor variables.

**MODEL DEVELOPMENT**

In this section, we formulate the nonstationary negative multinomial distribution for purchase occasions and then consider the special case of the nonstationary NBD. The negative multinomial distribution is the multivariate analogue of the NBD for purchase occasions in more than one time period. Consider a panel of $H$ households. The purchase occasions for household $i$ are described by a Poisson process with proportional intensity, $\lambda_i \psi(t)$, at time $t$ where $\lambda_i$ is a component specific to household $i$ and $\psi(t)$ is a component common to all households. The specific component, $\lambda_i$, is a household's propensity to purchase the product and varies across the population, whereas the common component, $\psi(t)$, is due to a common marketing environment at time $t$. The common component changes over time for all members of the population and can include marketing variables, seasonal effects, and trends. Examples of proportional intensity models are given by Hausman, Hall, and Griliches (1984), Lenk and Rao (1990), and Wagner and Taudes (1986). In traditional NBD applications, $\psi$ is equal to one so that purchase occasions follow a stationary Poisson process with rate $\lambda$.

The aggregate purchase occasion process for the panel of $H$ households is the superposition of independent Poisson processes with intensity

$$
\sum_{i=1}^{n} \lambda_i \psi(t) = \lambda \psi(t).
$$

Under the assumption that $\{\lambda_i\}$ is a random sample from a gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$ (mean $\alpha/\beta$ and variance $\alpha/\beta^2$), $\lambda$ has a gamma distribution with shape $H\alpha$ and scale $\beta$ (Feller 1971, p. 47).

The negative multinomial is derived by considering the total number of purchase occasions during each of $K$ disjoint intervals, not necessarily of equal lengths. Let $N_i$ be the number of purchase occasions in the $k^{th}$ interval, which has length $\Delta_k$. Given $\lambda$, $N_i$ has a Poisson distribution with mean $\lambda \Delta_k$ where $\theta_i$ is the integral of $\psi$ over the $k^{th}$ interval and the $\{N_i\}$ are mutually independent (Ross 1983, p. 46-47). In the stationary case, $\psi(t)$ is equal to one and $\theta_i$ is equal to $\Delta_k$. The ratio, $\theta_i/\Delta_k$, is the mean intensity of the Poisson process during interval $k$.

The joint distribution of any subset, say $\{N_i, \ldots, N_j\}$, is obtained by mixing their independent Poisson distributions given $\lambda$ by the gamma distribution of $\{\lambda_i\}$:

$$
Pr(N_i, \ldots, N_j) = \frac{\Gamma(H\alpha + N(I,J))}{\Gamma(H\alpha) N_i!} \left( \frac{\beta + \Theta(I,J)}{\beta + \Theta(I,J)} \right)^{N_i} \prod_{i=1}^{J} \left( \frac{\theta_i}{\beta + \Theta(I,J)} \right)^{N_i},
$$

for $N_i, \ldots, N_j = 0, 1, \ldots$

where:

$$
\Theta(I,J) = \sum_{i=1}^{J} \theta_i \text{ and } N(I,J) = \sum_{i=1}^{J} N_i.
$$

The mean, variance, and covariances are:

$$
E(N_i) = H\alpha \beta^{-1} \theta_i,
$$

$$
\text{Var}(N_i) = E(N_i) \left[ 1 + (H\alpha)^{-1} E(N_i) \right],
$$

$$
\text{Cov}(N_i,N_j) = (H\alpha)^{-1} E(N_i) E(N_j).
$$

Marginal and conditional probability distributions are easily derived from equation 1. The negative binomial distribution is the marginal distribution of $N_i$:

$$
Pr(N_i) = \frac{\Gamma(H\alpha + N_i)}{\Gamma(H\alpha) N_i!} \left( \frac{\beta}{\beta + \Theta(I,J)} \right)^{N_i} \left( \frac{\Theta(I,J)}{\beta + \Theta(I,J)} \right)^{\theta_i},
$$

for $N_i = 0, 1, \ldots$

which is a special case of equation 1.

Following the analysis of Morrison (1968), after observing the number of purchase occasions in the first $J$ periods, we see that $\lambda$ has a gamma distribution with shape parameter $H\alpha + N(1,J)$ and scale parameter $\beta + \Theta(1,J)$. Then the conditional expectation of $N_{i+1}$ given $N_i$ to $N_j$ is

$$
E(N_{i+1} | N_{i}, \ldots, N_j) = \theta_{i+1} \frac{H\alpha + N(1,J)}{\beta + \Theta(1,J)},
$$

CTA specializes equation 2 to the number of purchase occasions in two periods, the base period, $N_1$, and the test period, $N_2$. After we observe $N_1$, CTA forecasts $N_2$ by its conditional mean:

$$
E(N_2 | N_1) = \theta_2 \frac{H\alpha + N_1}{\beta + \Theta(1,J)},
$$

which is a special case of equation 2.

An alternative to the gamma mixing assumption of the NBD has been provided by Robbins (1977). He considers arbitrary mixing distributions while retaining the Poisson purchase incidence assumption. Suppose the
distribution of \( \lambda \) is \( G \), and define the marginal distribution of \( N_1 \) as

\[
f(x) = \int_0^\infty Pr(N_1 = x|\lambda)dG(\lambda).
\]

Then the conditional expectation is

\[
E(N_2|N_1 = n_1) = (n_1 + 1) \frac{\theta_2 f(n_1 + 1)}{\theta_1 f(n_1)}.
\]

(See theorem 1 in the Appendix.)

**ESTIMATION AND PREDICTION**

Suppose \( M_x \) of the \( H \) households in the panel buy exactly \( x \) times in the base period. How many purchases do these households have in the test period? CTA forecasts the total number, \( T_x \), of purchases in the test period by the \( M_x \) households that purchased \( x \) items in the base period. Let \( N_{i1} \) and \( N_{i2} \) be the number of purchases by the \( i \)th household in a base and a test period, respectively. Then

\[
T_x = \sum_{i=N_{i1}+1}^{M_x} N_{i2}.
\]

Theorem 2 in the Appendix derives the joint distribution of \( \{T_x\} \) for the NBD model. The conditional mean of \( T_x \) is

\[
E(T_x|M_x, x) = \begin{cases} M_x \beta x + \frac{\alpha}{\beta} & \text{nonstationary negative binomial} \\ M_x(x + 1) \frac{\theta_2 f(x + 1)}{\theta_1 f(x)} & \text{nonstationary Robbins model}. \end{cases}
\]

For the NBD model the conditional expectation follows from equation 3 by identifying \( H = M_x \) and \( N_1 = xM_x \), or theorem 2. For Robbins' model the conditional expectation follows from equation 4 and the definition of \( T_x \).

Implementation of these forecasts requires parameter estimation. If \( \psi \) is unknown, it is estimated over a calibration period. Next, \( \theta \)'s, which are integrals of \( \psi \), are computed for the base and forecast periods. For a brand, this computation would be done by using planned marketing activity for the forecast period. This leaves estimating the parameters \( \alpha \) and \( \beta \) for the negative binomial distribution and \( f(x) \) for Robbins' model. Estimators of these parameters are given in the Appendix.

For both models, there is a simple relationship between the stationary and nonstationary forecasts. The forecasts for the nonstationary model adjust the forecasts from the stationary model by the ratio of the mean purchase rates in the test period to those in the base period. Consequently, the forecasts of the nonstationary model are easily obtained from those of the stationary model. The forecast of \( T_x \) for the nonstationary negative binomial model is

\[
NNB_x = \left( \frac{\theta_2/\Delta_x}{\theta_1/\Delta_1} \right) SNB_x
\]

where:

\[
SNB_x = M_x \Delta_1 \left( \frac{\theta_2 x + \alpha}{\beta_2 + \Delta_1} \right)
\]

is the forecast of \( T_x \) for the stationary negative binomial model, and \( \theta_2 \) and \( \beta_2 \) are maximum likelihood estimators for the stationary model. (See theorem 3 in the Appendix.) The forecast of \( T_x \) for the nonstationary Robbins model is

\[
NRM_x = \left( \frac{\theta_2/\Delta_x}{\theta_1/\Delta_1} \right) SRM_x
\]

where:

\[
SRM_x = \begin{cases} (x + 1) M_x \Delta_2/\Delta_1 & \text{if } M_x > 0 \\ 0 & \text{if } M_x = 0 \end{cases}
\]

is the forecast of \( T_x \) for the stationary Robbins model. (See theorem 4 in the Appendix.) These procedures are the *ex ante* versions of the *ex post* adjustments of the stationary forecasts by the ratio of total purchases in the test period to those in the base period.

If the average rates, \( \theta_i/\Delta_i \), in the base and test periods are approximately equal, the stationary and nonstationary models have similar forecasts even if there is considerable nonstationarity during the two time periods. This result formalizes Ehrenberg’s (1972) view of the domain of application for the stationary negative binomial model.

An example obtains in mature markets for frequently purchased nondurable goods where the household's purchases are sensitive to promotions. At any given time, one or more manufacturers have promotional campaigns, though there may be seasonal effects, over longer periods of time, such as a year, the average level of promotional activity and the total number of purchase occasions remain fairly steady. It is this feature that probably accounts for the robustness of the NBD model in applications where intervals are long—six months or more.

The domain of application of stationary CTA points to inherent problems with using stationary models to assess the impact of marketing efforts, especially if there are trends and seasonal effects or if the market is not mature. Promotions generally are of short duration, and aggregating over periods as long as six months or a year will mask the effects of a particular marketing effort. Additionally, if shorter periods are used, the effects of marketing effort and seasonal effects are confounded in stationary CTA. This situation is exacerbated if there are trends due to exogenous factors. An analysis based on a stationary model would attribute incremental purchases, or lack of them, to marketing effort when part
or all of the incremental sales may be due to seasonal effects or trends.

**APPLICATION TO SCANNER PANEL DATA**

We demonstrate the application of nonstationary CTA with data for powder detergents and Tide obtained from the MSI library of single-source data supplied by A. C. Nielsen.

First, we identify a model for the nonstationary intensity function by using standard data analytic techniques. This data analysis identifies variables to include in the model and the functional form for the intensity function. Second, using the model identification from the data analysis, we estimate the nonstationary negative multinomial distribution by maximum likelihood. The first two years or 104 weeks of data are used to estimate the parameters of the negative multinomial distribution, and the last 34 weeks provide a validation period. Third, using the estimated intensity function from the negative multinomial distribution and first 104 weeks of data, we perform "rolling" CTA, week-by-week, on the last 34 weeks.

**Specifying the Intensity Function**

We considered five aggregate marketing variables as potential predictor variables. These variables are derived from marketing activity at the retail level. For a given week and store, the marketing activity variables indicate whether a given universal product code was displayed in the store, was featured in a store advertisement, had an associated coupon, had a point-of-purchase display, or had a special price code. A promotion by a leading brand in a large store is likely to have a greater impact on purchase occasions than a similar promotion by a smaller brand or store. To account for this differential impact of promotions, we weighted the marketing activities by the share of the product/store combination.

Table 1 gives the means, standard deviations, and correlations for the variables in the study. Advertisements and special price have the highest correlations with the number of purchase occasions. They are also highly correlated with each other because stores often use advertisements to announce a special price.

**Regression approach to model identification and estimation**

The following log-linear model provided a fairly accurate description of the data.

\[
\log(N_i) = \phi_0 + X_i \beta + \epsilon_i
\]

\(N_i\) is the number of purchase occasions in week \(k\) and \(X_i\) is a vector of independent variables for week \(k\). One explanation for this empirical finding is that a standardized Poisson random variable is asymptotically normally distributed as its mean goes to infinity (Stuart and Ord 1987, p. 144–5). The observed mean purchase incidents are large—446 for powder detergents and 104 for Tide—and the normal approximation is fairly accurate. A second order expansion of \(\log\left(N_i\right)\) around the mean of \(N_i\) results in the logarithm of its mean plus an error term that is approximately distributed as a normal random variable. Thus, the log-linear model in equation 5 provides a reasonable approximation to the negative multinomial model.

Hausman, Hall, and Griliches (1984) found in a study of patent data that the ordinary least squares estimates of the log-linear model roughly approximated the maximum likelihood estimates of the negative multinomial distribution. This result implies that a "quick-and-dirty" method of performing nonstationary CTA is to use the ordinary least square estimates of \(\phi\) from the log-linear model instead of their maximum likelihood estimates from the negative multinomial distribution, though we use the latter in our study.

These observations are useful for at least two reasons. First, maximum likelihood estimation of the negative multinomial distribution is relatively time consuming in comparison with least squares estimation of the log-linear model; we rapidly examined different independent variables and tested model adequacy by using standard methods. Second, the form of the regression function in equation 5 implies the form of the intensity function for the Poisson process. Exponentiation of both sides gives

\[N_i = \exp(\phi_0 + X_i \beta) \exp(\epsilon_i)\]

Because the mean of \(N_i\) is \(\mu_0 \beta / \beta\), a model for the average purchase rates is

\[\mu_0 \beta / \beta = \exp(\phi_0 + X_i \beta) E[\exp(\epsilon_i)]\]

Other researchers, such as Kalbfleisch and Prentice (1980, p. 31) and Wagner and Taudes (1986), use the exponential function to model the intensity function in Poisson regression because the intensity function is constrained to be positive. In addition to this mathematical requirement, the empirical fit of the log-linear model provides further justification of the exponential intensity function.

We used the log-linear model and regression analysis to identify potential predictor variables. Table 2 presents the estimated log-linear models for powder detergents and Tide. The full model indicates that in-store display, point-of-purchase display, coupon, and special price may not help significantly to predict \(\log(N_i)\) given that advertising is in the model. Indeed, sequentially removing in-store display, point-of-purchase display, and coupon from the models did not substantially decrease the predictive power of the models.

Advertising and price are the most significant variables. Their coefficients should be positive, because they are weighted sums of zero/one variables that indicate the absence or presence of the marketing activity. A positive value of price indicates a special price, and one would expect sales to increase. For powder detergents, advertising and special price are statistically significant.

---

3 The data were obtained from File 9, Retail Tracking, of the MSI Library from A. C. Nielsen.
Table 1: SUMMARY STATISTICS FOR POWDER DETERGENTS AND TIDE
(138 weeks of weekly observations)

<table>
<thead>
<tr>
<th>Powder detergents Statistic</th>
<th>N</th>
<th>Log(N)</th>
<th>Display</th>
<th>Ads</th>
<th>Coupon</th>
<th>PPD</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>445.877</td>
<td>6.078</td>
<td>.024</td>
<td>.015</td>
<td>.001</td>
<td>.001</td>
<td>.018</td>
</tr>
<tr>
<td>S.D.</td>
<td>93.573</td>
<td>.212</td>
<td>.016</td>
<td>.016</td>
<td>.002</td>
<td>.002</td>
<td>.017</td>
</tr>
</tbody>
</table>

Correlations for powder detergents Variable

<table>
<thead>
<tr>
<th>N</th>
<th>1.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(N)</td>
<td>.988</td>
</tr>
<tr>
<td>Display</td>
<td>-.046</td>
</tr>
<tr>
<td>Ads</td>
<td>.406</td>
</tr>
<tr>
<td>Coupon</td>
<td>-.036</td>
</tr>
<tr>
<td>PPD</td>
<td>.019</td>
</tr>
<tr>
<td>Price</td>
<td>.118</td>
</tr>
</tbody>
</table>

Tide Statistic

| Mean            | 104.007 | 4.390 | .008 | .009 | .000 | .000 | .006 |
| S.D.            | 78.009  | .719  | .011 | .015 | .001 | .001 | .011 |

Correlations for Tide Variable

<table>
<thead>
<tr>
<th>N</th>
<th>1.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(N)</td>
<td>.928</td>
</tr>
<tr>
<td>Display</td>
<td>.208</td>
</tr>
<tr>
<td>Ads</td>
<td>.683</td>
</tr>
<tr>
<td>Coupon</td>
<td>.143</td>
</tr>
<tr>
<td>PPD</td>
<td>.040</td>
</tr>
<tr>
<td>Price</td>
<td>.388</td>
</tr>
</tbody>
</table>

N is number of weekly purchase occasions. Log(N) is natural logarithm of N. Display indicates displayed in store. Ads indicates featured in store advertisement. Coupon indicates store advertisement has a coupon. PPD is Point-of-purchase display. Price is special price.

However, the coefficient for special price is negative instead of positive. This fact and a large correlation (.659) between advertising and special price indicates multicollinearity. Consequently, special price was eliminated from the model, and the $R^2$ dropped slightly from .17 to .14. For Tide, both advertising and special price are statistically significant when they are the only independent variables in the model, but special price is not statistically significant when advertising is included. Therefore, of the five independent variables, only advertising is included in the final model for powder detergents and Tide.

A residual analysis indicates systematic, large positive and negative residuals that seem to indicate the presence of holiday and seasonal effects. For example, purchases of detergents tend to be unusually low during the week of Valentines Day, with surges before and after the holiday. This pattern also holds for holidays such as Thanksgiving and the Christmas–Hanukkah–New Year season. Also, there seems to be an unusually high number of purchases during the week of the Super Bowl. With only two years of data for estimation, we cannot estimate separate effects for each holiday and season.

Instead, we define two dummy variables that indicate negative and positive seasonal-holiday effects for the two years of data in the fit period. Though we view these dummy variables to be less satisfactory than separate variables for different seasonal and holiday effects, their use demonstrates the incorporation of seasonal effects into the model.

**Nonstationary Negative Multinomial Model**

After identifying the predictor variables and functional form of the intensity function by the log-linear model, we use the negative multinomial distribution in equation 1 to estimate the parameters from the first 104 weeks of data. The Appendix gives the stationary equations. The number of households, $H$, in the study is 2620. The expected number of purchases by the entire panel for a week in the fit period is

$$E(N_t) = H \alpha_0/\beta$$

$$= 2620 \exp(\phi_0) \exp(X_{t1} \phi_1 + X_{t2} \phi_2 + X_{t3} \phi_3)/\beta,$$

where:
Table 2
ESTIMATED LOG-LINEAR MODELS
(dependent variable is natural logarithm of number of weekly purchase occasions for 138 weeks)

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Display</th>
<th>Ads</th>
<th>Coupon</th>
<th>PPD</th>
<th>Price</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Powder detergents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>6.01</td>
<td>-1.79</td>
<td>6.80</td>
<td>-4.84</td>
<td>9.73</td>
<td>-2.37</td>
<td>.18</td>
</tr>
<tr>
<td>S.E.</td>
<td>.03</td>
<td>1.53</td>
<td>1.36</td>
<td>8.85</td>
<td>8.87</td>
<td>1.78</td>
<td>.19</td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td>.24</td>
<td>.00</td>
<td>.59</td>
<td>.23</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>6.03</td>
<td>6.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>.03</td>
<td>1.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>6.01</td>
<td>4.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>.02</td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>6.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.28</td>
<td>.01</td>
</tr>
<tr>
<td>S.E.</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td><strong>Tide</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>4.06</td>
<td>11.64</td>
<td>29.73</td>
<td>55.19</td>
<td>84.13</td>
<td>-7.61</td>
<td>.38</td>
</tr>
<tr>
<td>S.E.</td>
<td>.07</td>
<td>5.86</td>
<td>4.28</td>
<td>36.96</td>
<td>56.78</td>
<td>6.67</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td>.05</td>
<td>.00</td>
<td>.14</td>
<td>.20</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>4.15</td>
<td></td>
<td>28.23</td>
<td></td>
<td></td>
<td>.52</td>
<td>.34</td>
</tr>
<tr>
<td>S.E.</td>
<td>.06</td>
<td></td>
<td>4.27</td>
<td></td>
<td></td>
<td>5.64</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td></td>
<td>.00</td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>4.15</td>
<td></td>
<td>28.50</td>
<td></td>
<td></td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>.06</td>
<td></td>
<td>3.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td></td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>4.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.49</td>
<td>.12</td>
</tr>
<tr>
<td>S.E.</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.23</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Display indicates displayed in store. Ads indicates featured in store advertisement. Coupon indicates store advertisement has a coupon. PPD is point-of-purchase display. Price is special price.

\[ X_{4,1} = \text{aggregate advertising for week } k, \]
\[ X_{4,2} = \begin{cases} 1 & \text{if negative seasonal or holiday effect for week } k, \\ 0 & \text{otherwise}, \end{cases} \]
\[ X_{4,3} = \begin{cases} 1 & \text{if positive seasonal or holiday effect for week } k, \\ 0 & \text{otherwise}. \end{cases} \]

In the following data analysis, we identified \( \alpha \) with \( \exp(\phi_0) \) and \( \theta_0 \) with \( \exp(\phi_1 X_{4,1}) \).

Table 3 displays the results of fitting the negative multinomial distribution. The first part of the table reports the maximum likelihood estimators and their asymptotic standard errors for powder detergents and Tide. All of the estimated parameters are significantly different from zero at the .01 level.

The effect of advertising, \( \phi_1 \), is greater for Tide than for powder detergents. This fact is not surprising once we recognize that the powder detergent category includes various brands that rarely are advertised simultaneously by a particular store. Rather, stores tend to rotate the brands they advertise. Not infrequently, different stores advertise different brands in a given week so that aggregate advertising at the category level has proportionally less variation than advertising at the brand level. To obtain a quantitative measure, suppose the weekly advertising aggregated across stores for \( b \) brands is independent with common mean \( \mu \) and standard deviation \( \sigma \). Then the coefficient of variation for the weekly, aggregated advertising is \( \sigma/(\sqrt{b\mu}) \), which is less than the coefficient of variation, \( \sigma/\mu \), for any individual brand. Thus, the systematic variation in purchases due to advertising is partially masked at the category level. In fact, the coefficient of variation for weekly advertising is .84 for powder detergents in comparison with 1.53 for Tide.

The coefficients for negative and positive seasonal effects have practical significance. For example, fix the level of advertising and compare a week with positive
### Table 3
ESTIMATED NEGATIVE MULTINOMIAL DISTRIBUTION
(dependent variable is natural logarithm of number of weekly purchase occasions)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Powder detergents</th>
<th>Tide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MLE</td>
<td>S.E.</td>
</tr>
<tr>
<td>Log(shape)</td>
<td>ϕ_1</td>
<td>2.50</td>
<td>.108</td>
</tr>
<tr>
<td>Advertising</td>
<td>ϕ_2</td>
<td>5.01</td>
<td>.109</td>
</tr>
<tr>
<td>Negative seasonal</td>
<td>ϕ_3</td>
<td>-.19</td>
<td>.004</td>
</tr>
<tr>
<td>Positive seasonal</td>
<td>ϕ_4</td>
<td>.17</td>
<td>.012</td>
</tr>
<tr>
<td>Scale</td>
<td>β</td>
<td>79.58</td>
<td>.562</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Powder detergents</th>
<th>Tide</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>85.6</td>
<td>75.8</td>
</tr>
<tr>
<td>MAPE</td>
<td>14.6</td>
<td>88.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Powder detergents</th>
<th>Tide</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>117.9</td>
<td>83.4</td>
</tr>
<tr>
<td>MAPE</td>
<td>29.3</td>
<td>109.6</td>
</tr>
</tbody>
</table>

MLE is the maximum likelihood estimate. S.E. is the asymptotic standard error of the maximum likelihood estimate. RMSE is root mean squared error. MAPE is mean absolute percent error.

seasonal effects to a week without seasonal effects for Tide. Then the expected weekly sales with positive seasonal effects is 49% greater than that without seasonal effects.

The estimated household shape parameters (ϕ_n) for the two series are fairly close. However, the estimated scale parameters (β) differ considerably. In the absence of advertising and seasonal effects, the expected time between purchases for a randomly selected household is seven weeks for powder detergents and 41 weeks for Tide. As anticipated, the time between purchases is shorter at the product category level than at the brand level. The expected time between purchase occasions for Tide is large because many households infrequently or never buy Tide during the observational period.

The center of Table 3 reports fit statistics—root mean square errors (RMSE) and mean absolute percentage errors (MAPE)—for the stationary and nonstationary negative multinomial models. The stationary model does not include predictor variables and sets ϕ equal to one. The nonstationary model substantially improves on the stationary model: the predictor variables account for 34.6% of the total variation in the purchase occasions for powder detergents and 57.8% for Tide.

The lower part of Table 3 reports prediction statistics for forecasting the purchase occasions in weeks 105 to 138. Forecasts are given by equation 2. The nonstationary model improves the forecast MAPE by 19.1% for powder detergent and 50.5% for Tide. These observations are consistent with our findings on the relative impact of advertising for Tide and the powder detergent category.

### Conditional Trend Analysis

A goal of CTA is to predict the total number of purchases, T, in the test period for the M_i households that purchased x items in the base period. The nonstationary models use the estimated intensity function for the negative multinomial distribution, which is presented in the preceding section. The intensity function is estimated over the first 104 weeks of data. Then, we performed "rolling" CTAs from week 105 to week 138 for the stationary and nonstationary NBD and Robbins models using one-week and four-week base and test periods.

The lengths of the two periods do not have to be equal, but we selected equal periods for illustrative purposes. An anonymous reviewer noted that using longer base periods would lead to a more accurate segmentation of households based on their purchases. Then a short test period could be used to ascertain the impact of a promotion. In practice, the lengths of the periods should be selected according to managerial considerations.

One difficulty of the negative binomial model is that the zero purchase class includes households that never
purchase the product and households that purchase the product but did not do so in the base period. To compensate for the resulting “spike at zero,” we adapt the model of Morrison (1969) to the nonstationary case. These forecasts are given by equation A3 in the appendix.

Table 4 reports the MAPE for CTAs where the base and test periods are either one or four weeks. Purchase classes are reported in Table 4 only if $M_x$ is greater than zero for every CTA. The MAPEs are computed across the CTAs for each purchase class. The one-week forecasts have large MAPEs when $x$ is greater than zero because $T_x$ is small, so small absolute forecast errors have large percentage errors. The main message of Table 4 is that the nonstationary models generally perform better than the stationary models, and the MAPEs tend to be smaller for four-week periods than for one-week periods. Though the magnitude of improvement is modest in this particular case, once $\psi$ is estimated, nonstationary forecasts are obtained from the stationary ones without much additional effort by using theorems 3 and 4. Furthermore, Table 4 aggregates over weeks in which advertising levels in the base and test periods are similar and weeks in which they are widely different, thus masking the true contribution of the nonstationary model.

The true contribution of the nonstationary model is illustrated in Figure 3, a plot of the percentage improvement of the nonstationary model over the stationary model versus the ratio of the mean intensity in the test period to that in the base period. This ratio, $(\hat{\theta}_x/\Delta_x)/(\theta_x/\Delta_x)$, is the adjustment factor used to obtain the nonstationary forecasts from the stationary ones and measures the relative nonstationarity of the test and base periods. Figure 3 combines MAPEs across purchase classes within each CTA. When the adjustment factor is one, the stationary and nonstationary forecasts are identical, so the percentage improvement is zero. When $\hat{\theta}_x/\theta_x$ is greater (less) than one, while $N_x/N_x$ is less (greater) than one, the stationary model performs better than the nonstationary model. These weeks tend to have an adjustment factor close to one, so the results (negative percentage improvement in Figure 3) may be partially explained by the inherent variation of the data.

The plots for Tide in Figure 3 best illustrate the effects of nonstationarity. As the adjustment factor is farther from one, the percentage improvement increases considerably. Thus, the nonstationary model leads to substantially better forecasts when there is considerable difference between the base and test periods. Powder detergents have a pattern similar to that of Tide, but there is less variability in the adjustment factor because advertising at the brand level varies more than advertising at the category level.

**FORECASTING BIAS**

**Bias Due to Model Misspecification**

Morrison and Schmittlein (1981, 1988) note that CTA tends to underpredict the zero class, $T_0$. They conjecture that nonstationarity may be a cause. Their conjecture is partially true; incorrectly assuming a stationary process results in a forecasting bias. The direction and magnitude of the bias depend on the ratio of the average purchase rates in the base and test periods. Underprediction will result if the average rate in the base period is less than the average rate in the test period. (See theorems 5 and 6 in the Appendix.)

---

### Table 4

**CONDITIONAL TREND ANALYSIS FOR POWDER DETERGENTS AND TIDE**

(mean absolute percentage errors are computed within purchase class and across periods)

<table>
<thead>
<tr>
<th>Class</th>
<th>Powder detergents</th>
<th></th>
<th>Robbins</th>
<th></th>
<th></th>
<th>Tide</th>
<th></th>
<th>Robbins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NBD</td>
<td>Stationary</td>
<td>Nonstationary</td>
<td>Stationary</td>
<td>Nonstationary</td>
<td>NBD</td>
<td>Stationary</td>
<td>Nonstationary</td>
</tr>
<tr>
<td>One-week time periods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>21.6</td>
<td>18.6</td>
<td>21.0</td>
<td>16.8</td>
<td></td>
<td>75.6</td>
<td>38.4</td>
<td>73.2</td>
</tr>
<tr>
<td>1</td>
<td>56.1</td>
<td>52.0</td>
<td>37.9</td>
<td>32.2</td>
<td></td>
<td>75.3</td>
<td>65.7</td>
<td>76.9</td>
</tr>
<tr>
<td>2</td>
<td>73.9</td>
<td>68.3</td>
<td>101.5</td>
<td>92.2</td>
<td></td>
<td>103.5</td>
<td>102.4</td>
<td>151.9</td>
</tr>
<tr>
<td>3</td>
<td>105.3</td>
<td>102.4</td>
<td>195.4</td>
<td>190.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-week time periods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.2</td>
<td>42.5</td>
<td>58.2</td>
</tr>
<tr>
<td>0</td>
<td>17.4</td>
<td>13.5</td>
<td>19.1</td>
<td>14.6</td>
<td></td>
<td>33.8</td>
<td>30.4</td>
<td>35.9</td>
</tr>
<tr>
<td>1</td>
<td>13.9</td>
<td>10.9</td>
<td>14.8</td>
<td>12.3</td>
<td></td>
<td>40.9</td>
<td>38.1</td>
<td>41.9</td>
</tr>
<tr>
<td>2</td>
<td>14.2</td>
<td>13.5</td>
<td>18.7</td>
<td>20.0</td>
<td></td>
<td>47.9</td>
<td>55.0</td>
<td>61.5</td>
</tr>
<tr>
<td>3</td>
<td>12.5</td>
<td>12.0</td>
<td>21.9</td>
<td>21.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14.9</td>
<td>16.2</td>
<td>24.9</td>
<td>24.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28.2</td>
<td>24.7</td>
<td>53.2</td>
<td>49.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>32.9</td>
<td>31.6</td>
<td>55.2</td>
<td>48.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NBD is negative binomial distribution.
In our empirical study, though the nonstationary models provide more accurate forecasts than the stationary models, underprediction is a persistent phenomenon. Table 5 illustrates this result for the zero class. The forecast residuals, which are defined as the actual number of purchase occasions minus the predicted number, are stratified according to their sign. A negative residual indicates overprediction, whereas a positive residual indicates underprediction. The nonstationary models provide more accurate forecasts than the corresponding stationary models. However, all of the models tend to underpredict. Thus, the nonstationary model improves prediction accuracy, but underprediction persists.

Persistent underprediction indicates that more is involved than only incorrectly assuming a stationary model. If nonstationarity were the only factor, the forecasts would consistently underestimate only if the marketing activity in the test period were consistently greater than the marketing activity in the base period or if there were a trend in sales, which is not generally the case. Hence,
Table 5
RESIDUALS FROM PREDICTING THE TOTAL PURCHASES OF THE ZERO CLASS
(base and test periods are one-week in duration)

<table>
<thead>
<tr>
<th>Residuals</th>
<th>Statistics</th>
<th>NBD</th>
<th>Robbins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stationary</td>
<td>Nonstationary</td>
</tr>
<tr>
<td>Powder detergents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Number</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-25.7</td>
<td>-6.0</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>Number</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>77.2</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>Tide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Number</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-62.8</td>
<td>-30.2</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>52%</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>Number</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>78.8</td>
<td>39.6</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

NBD is negative binomial distribution. "Improvement" is the percentage change in the average residual of the nonstationary model in comparison with the stationary model.

additional factors must be contributing to underprediction.

Bias Due to Skewed Distributions

One possible explanation of the underprediction for the NBD is the choice of measurement scale. In the initial data analysis, we used the heuristic that the number of purchase occasions follows a log-linear model. If the log-linear model is specified correctly, the forecasts of the number of purchase occasions will tend to underpredict on the original scale. To see this point, define the forecast residuals on the logarithmic scale as

\[ r_k = \log(N_i) - \log(\hat{N}_i) \text{ for } J + 1 \leq k \leq K. \]

If the linear model for \( \log(N_i) \) is specified correctly, the mean forecast errors should be zero:

\[ \hat{r} = (K - J)^{-1} \sum_{k=J+1}^{K} r_k = 0. \]

Jensen’s inequality gives

\[ 1 = \exp(\hat{r}) \leq (K - J)^{-1} \sum_{k=J+1}^{K} \exp(r_k) \leq (K - J)^{-1} \sum_{k=J+1}^{K} \frac{N_i}{\hat{N}_i}. \]

Consequently, we should expect the forecasts \( \{\hat{N}_i\} \) to underpredict \( \{N_i\} \) even though the forecasts on the logarithmic scale are not biased. Thus, underprediction does not necessarily invalidate NBD in the same manner that it indicates an incorrectly specified linear model.

Theorem 7 in the Appendix provides a more detailed analysis and demonstrates that one should expect underprediction if the distribution of \( T \) is skewed to the right and overprediction if it is skewed to the left. Because the third central moment is positive for NBD, \( T \) is skewed to the right and underprediction is expected.

DISCUSSION

We extend the well-known conditional trend analysis (CTA) model, first proposed by Goodhardt and Ehrenberg, for evaluating the impact of marketing programs by including marketing variables, seasonal effects, and other time-dependent phenomena that contribute to nonstationarity. Traditional applications of CTA use relatively long base and test periods to “wash out” the impact of nonstationarity. For example, if seasonal effects are pronounced, one-year periods are desirable; a shorter test period is likely to bias the results. However, many marketing programs such as promotions are frequent and of short duration. Long test periods confound the effects of various programs in addition to introducing reporting delays that reduce the managerial usefulness of the results. By explicitly introducing nonstationarity, one can make the test periods as short as desired without biasing results. This approach allows evaluation and reporting of selected, short-term programs in a time frame that is managerially relevant. In addition to refining the traditional ex post evaluative function of CTA, modeling nonstationarity enables marketers to forecast and to evaluate ex ante the impact of alternative programs on customers segmented by prior purchase behavior.

Our approach to incorporating nonstationarity is via the proportional intensity model—each customer’s Poisson purchase rate is assumed to be multiplied by the same
function of the time-dependent variables. Though conceptually somewhat restrictive—for example, heavy dealing may induce heavy buyers to purchase more while having less impact on light buyers—this assumption leads to two consequences that make implementation of our model straightforward.

—A "quick and dirty" estimate of the proportional intensity function can be obtained by using log-linear models and ordinary least squares when the average number of purchase occasions is large. Thus, standard data-analytic techniques can be used to identify key variables. The quick and dirty estimates of the coefficients tend to be close to those obtained by using cumbersome maximum likelihood methods.

—Once the intensity function is estimated, the nonstationary forecasts are easily obtained by multiplying the standard, stationary forecasts by an adjustment factor that is the ratio of the average intensity in the test period to that in the base period.

Authors such as Wagner and Tandes (1986) and Vilcasim and Jain (1991) have used the proportional intensity model in marketing studies. However, their broader focus has generally resulted in greater complexity, both conceptual and computational. The approach we describe, in which attention is restricted to CTA, retains the elegance and simplicity of the NBD model of consumer purchase behavior while enlarging the domain of application of CTA.

Our empirical example demonstrates that introducing marketing variables substantially improves predictive accuracy, particularly as the average intensity of marketing activity in the test period diverges from that in the base period. Ignoring nonstationarity and proceeding along traditional lines results in substantially biased results in such cases. However, in our empirical examples, the well-documented phenomenon of underprediction of test period purchases by the zero class (persons who did not buy in the base period) persists. We prove that underprediction is a consequence of the skewness of the NBD and not principally due to model inadequacy. This underprediction is due to the forecasts in CTA being optimal with respect to squared-error loss, which is symmetric for over- and underprediction. Box and Tiao (1973, p. 310–315) comment on the inadequacies of using squared-error loss, and Hoaglin, Mosteller, and Tukey (1983) and Goodall (1983) review alternative loss functions. In CTA the loss function should explicitly recognize the asymmetry in costs for under- and overprediction. Blatter and George (1992) take such an approach in estimating demand functions. We are currently modifying CTA to incorporate classes of loss functions that are more realistic than squared-error loss.

APPENDIX

ROBBINS MODEL

Theorem 1 (Robbins): Suppose the distribution of λ is G, and let f(x) be the marginal distribution of N, evaluated at x:

\[ f(x) = \int_0^\infty \text{Pr}(N_1 = x | \lambda) dG(\lambda) \]

\[ = (x!)^{-1} \int_0^\infty \exp(-\lambda \theta_1)(\lambda \theta_1)^x dG(\lambda). \]

Then the conditional mean of N₂ given N₁ is

\[ E(N_2 | N_1) = (n_1 + 1) \frac{\theta_2 f(n_1 + 1)}{\theta_1 f(n_1)}. \]

The stationary case is Robbins' (1977) result.

Proof. After observation of N₁ = n₁, the posterior distribution of λ is

\[ dG(\lambda | n_1) = \exp(-\lambda \theta_1)(\lambda \theta_1)^{n_1} (n_1)!^{-1} dG(\lambda) / f(n_1). \]

The conditional mean is

\[ E(N_2 | N_1 = n_1) = E_{N_1 = n_1}[E(N_2 | \lambda)] \]

\[ = E_{N_1 = n_1}(\theta_2) \]

\[ = \theta_2 \int_0^\infty \lambda dG(\lambda | n_1) \]

\[ = (n_1 + 1) \frac{\theta_2 f(n_1 + 1)}{\theta_1 f(n_1)}. \]

The last line follows by

\[ (n_1!)^{-1} \int_0^\infty \lambda \exp(-\lambda \theta_1)(\lambda \theta_1)^{n_1} dG(\lambda) \]

\[ = \frac{n_1 + 1}{\theta_1} \int_0^\infty \exp(-\lambda \theta_1)(\lambda \theta_1)^{n_1 + 1} dG(\lambda) \]

\[ = ((n_1 + 1)/\theta_1) f(n_1 + 1). \]

\[ \Box \]

JOINT DISTRIBUTION OF \( \{T_i\} \)

Theorem 2: Assume that \( \lambda \) is a random sample from a gamma distribution with shape parameter \( \alpha \) and scale parameter \( \beta \). Given the purchases occasions \( \{M_i\} \) in the base period, \( \{T_i\} \) are mutually independent negative binomial random variables with probability mass function

\[ f(T_i | M_i) = \frac{\Gamma(\alpha M_i + x M_i + T_i)}{\Gamma(\alpha M_i + x M_i) T_i!} (\beta + \theta_1)^{\alpha M_i + x M_i} (\theta_2)^{-\beta + \theta_1 + \theta_2} T_i^{-\beta + \theta_1 + \theta_2} \]

for \( x = 0, 1, \ldots \) and \( T_i = 0, 1, \ldots \).

Proof. \( T_i \) given \( \lambda \) is a sum of independent Poisson random variables. Thus, \( T_i \) is a Poisson random variable with rate

\[ \lambda(x) \theta_2 = \left( \sum_{\{N_i = x\}} \lambda \right) \theta_2 \]

and \( \{T_i\} \) are mutually independent. Because \( \lambda \) is a random sample from a gamma distribution, \( \{\lambda(x)\} \) are mutually independent, and \( \lambda(x) \) has a gamma distribution with shape \( \alpha M_i \).
and scale $\beta$. After observation of purchases in the base period, \{\lambda(x)\} given \{M_i\} are mutually independent gamma random variables where $\lambda(x)$ has shape parameter $\alpha M_i + x M_i$ and scale parameter $\beta + \theta_i$. Thus, \{T_i\} are mutually independent negative binomial random variables with probability mass functions given by equation A1.

**ESTIMATION AND PREDICTION**

**Theorem 3.** Let $\hat{\alpha}_N$ and $\hat{\beta}_N$ be the maximum likelihood estimators for the nonstationary negative binomial model, and let $\hat{\alpha}_S$ and $\hat{\beta}_S$ be the maximum likelihood estimators for the stationary negative binomial model. The two sets of estimators are related by

\[
\hat{\alpha}_N = \hat{\alpha}_S \quad \text{and} \quad \hat{\beta}_N = \frac{\theta_i}{\Delta_i} \hat{\beta}_S.
\]

The forecast of $T_s$ for the nonstationary negative binomial model is

\[
\text{NNB}_s = \left(\frac{\theta_i/\Delta_i}{\theta_i/\Delta_i}\right) \text{SNB}_s
\]

where:

\[
\text{SNB}_s = M_i \Delta_s \left(\frac{\hat{\alpha}_S + x}{\hat{\beta}_S + \Delta_i}\right)
\]

is the forecast of $T_s$ for the stationary negative binomial model.

**Proof.** After the observation that $M_i$ households had $x$ purchase occasions in the base period, the likelihood for $\alpha$ and $\beta$ under nonstationarity is

\[
L_{\text{non}} = \prod_{x=0}^{\infty} \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)!} \left(\frac{\beta + \theta_i}{\beta + \theta_i} \right)^\alpha \left(\frac{\theta_i}{\theta_i}\right)^{\Delta_i} x^x e^{-\beta x}.
\]

whereas under stationarity it is

\[
L_{\text{st}} = \prod_{x=0}^{\infty} \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)!} \left(\frac{\beta + \Delta_i}{\beta + \theta_i} \right)^\alpha \left(\frac{\theta_i}{\theta_i}\right)^{\Delta_i} x^x e^{-\beta x}.
\]

Comparing the two likelihoods, we see they have the same functional form for $\alpha$ and that replacing $\beta$ by $\beta\theta_i/\Delta_i$ in the nonstationary likelihood results in the stationary likelihood. The first result follows because the maximum likelihood estimator of a function of a parameter is the function of maximum likelihood estimator of the parameter.

The estimated forecasts are found by substituting the unknown parameters by their maximum likelihood estimators:

\[
\text{NNB}_s = M_i \alpha_i \left(\frac{\hat{\alpha}_S + x}{\hat{\beta}_S + \Delta_i}\right)
\]

\[
= \left(\frac{\theta_i/\Delta_i}{\theta_i/\Delta_i}\right) M_i \Delta_s \left(\frac{\hat{\alpha}_S + x}{\hat{\beta}_S + \Delta_i}\right)
\]

\[
= \left(\frac{\theta_i/\Delta_i}{\theta_i/\Delta_i}\right) \text{SNB}_s,
\]

which proves the theorem.

The next theorem presents a similar result for Robbins’ model.

**Theorem 4.** The forecast of $T_s$ for the nonstationary Robbins model is

\[
\text{NRM}_s = \left(\frac{\theta_i/\Delta_i}{\theta_i/\Delta_i}\right) \text{SRM}_s,
\]

where:

\[
\text{SRM}_s = \begin{cases} 
(x + 1) M_{s+1} \Delta_s / \Delta_i & \text{if } M_i > 0 \\
0 & \text{if } M_i = 0
\end{cases}
\]

is the forecast of $T_s$ for the stationary Robbins model.

**Proof.** In Robbins' model $f(x)$ is unknown and is estimated by $M_s/H$ where $H$ is the number of households. Substituting this estimate of $f(x)$ into the forecasts results in

\[
\text{NRM}_s = \left(\frac{\theta_i}{\theta_i}\right) (x + 1) \frac{M_{s+1}}{H} \Delta_s/\Delta_i
\]

which proves the theorem.

**Maximum Likelihood Estimates for the Negative Multinomial Distribution**

Suppose $\psi$ is a function of $\phi_1, \ldots, \phi_p$. The log-likelihood is defined by equation 1:

\[
L = \log[\text{Pr}(N_1, \ldots, N_p)].
\]

Define $N = N(1, J)$ to be the total number of purchase occasions in the calibration period; $\Theta = \Theta(1, J)$, and

\[
\psi(x) = \frac{d}{dx} \log[\Gamma(x)].
\]

The stationary equations are:

\[
\frac{\partial L}{\partial \alpha} = \psi(\alpha + N) - \psi(\alpha) + \log(\beta) - \log(\beta + \Theta) = 0
\]

\[
\frac{\partial L}{\partial \beta} = \frac{\alpha + N}{\beta + \Theta} = 0
\]

\[
\frac{\partial L}{\partial \phi_i} = \sum_{k=1}^{J} \frac{N_k}{\phi_i} \frac{\alpha + N}{\beta + \Theta} \frac{\partial \phi_i}{\partial \phi_i} = 0
\]

for $i = 1, \ldots, p$. By solving for $\beta$ in $\partial L/\partial \beta = 0$, we obtain

\[
\beta = \alpha \Theta / N.
\]

Therefore, there is a restriction on the solution to the stationary equations that reduces the dimension of the problem from $p + 2$ to $p + 1$ parameters. Hausman, Hall, and Griliches (1984, p. 916) derive the normal equations for the negative multinomial distribution by assuming that $\alpha$ is equal to $\beta$, which is needlessly restrictive.

**Morrison’s Fix for the Zero Class**

Frequently, the NBD model performs poorly in forecasting $T_0$ because $M_0$ consists of households that never purchase the...
product and households that do purchase the product but did not do so in the base period. Morrison (1969) suggested a modification of the NBD model to adjust for nonbuyers. Let \( B \) be the set of buyers and let \( \bar{B} \) be the set of nonbuyers. Define \( p = P(B) \), and let \( f(x) \) be the negative binomial probability of purchasing \( x \) items in the base period. If a household is in \( B \), the number, \( N_i \), of purchases in the base period has a negative binomial distribution. For the zero class,

\[
g(0) = P(N_i = 0) = P(N_i = 0) \frac{\alpha}{\beta} P(B) + P(N_i = 0) \frac{\beta}{\beta} P(\bar{B})
\]

\[= f(0)p + 1 - p.
\]

For positive \( x \),

\[
g(x) = P(N_i = x) = x\beta P(B) + P(N_i = x) \frac{\beta}{\beta} P(\bar{B})
\]

\[= f(x)p.
\]

Note that \( g(\cdot) \) is a probability distribution.

The forecasts of \( T_i \) are

\[
E(T_i|x, M_x) = \begin{cases} 
\frac{pM_0 \theta_i \alpha}{\beta + \theta_i} & \text{for } x = 0 \\
M_x \theta_i \alpha + x \theta_i \beta + \theta_i & \text{for } x > 0.
\end{cases}
\]

The likelihood given by equation A2 becomes

\[
\prod_{x=0}^{\infty} g(x)^{M_x}.
\]

If \( f(0) \) is known and is less than one, the maximum likelihood estimate of \( p \) is

\[
\hat{p} = \min \left( 1, \frac{1 - M_0/H}{1 - f(0)} \right),
\]

where \( H \) is the total number of households in the panel. The normal equations and the Hessian are easily computed with sufficient perseverance. It is easily seen that the results of theorem 3 hold for Morrison’s fix if we redefine

\[
SNB_x = \begin{cases} 
\beta M_0 \Delta_x \frac{\alpha}{\beta + \Delta_x} & \text{for } x = 0 \\
M_x \Delta_x \frac{\alpha + x}{\beta + \Delta_x} & \text{for } x \geq 1.
\end{cases}
\]

Forecasting Bias

Bias Due to Model Misspecification

Suppose that, in reality, \( Z \) is a nonstationary Poisson process with intensity \( \lambda \xi(t) \), and that we incorrectly assume that the process is stationary with intensity \( \lambda \xi \). \( Z^* \) is our incorrect choice for the stationary process. Clearly, the forecasts for the two models can be widely different depending on the choice of \( \xi \).

A natural question is to find \( \xi \) such that our process \( Z^* \) is as “close” as possible to the true process \( Z \) on disjoint interval \( (s_k, t_k) \) (exclude the left endpoint and include the right endpoint) for \( k = 1 \) to \( K \). Intuitively, the optimal choice of \( \xi \) is the time average of \( \psi \) over the intervals. This choice is optimal with respect to minimizing the Kullback-Leibler information (Bickel and Doksum 1977, p. 226), which is the expected log-likelihood ratio. The Kullback-Leibler information is a measure of the “probabilistic closeness” of two distributions; small measures indicate similar distributions.

Consider \( K \) disjoint time interval \( (s_k, t_k) \) for \( k = 1, \ldots, K \). For each interval, define

\[
N_k = Z(t_k) - Z(s_k) \quad \text{and} \quad \Delta_k = t_k - s_k.
\]

Theorem 5: The \( \xi \) that minimizes the Kullback-Leibler information

\[
KL(\xi) = E_{N_1, \ldots, N_K} \left\{ \log \frac{Pr(N_1, \ldots, N_K)}{Pr(N_1^*, \ldots, N_K^*)} \right\}
\]

is the time average of \( \psi \) over the intervals:

\[
\xi^*_k = \frac{\sum_{k=1}^{K} \theta_k}{k}, \quad \sum_{k=1}^{K} \Delta_k
\]

Proof: The \( N_i \)'s are independent Poisson random variables, and the mean of \( N_i \) is \( \lambda \theta_i \). Also, the \( N_i^* \)'s are independent Poisson random variables, and the mean of \( N_i^* \) is \( \lambda \xi \Delta_k \). Consequently, their log-likelihood ratio is

\[
\log \left\{ \frac{Pr(N_1 = n_1, \ldots, N_K = n_K)}{Pr(N_1^* = n_1, \ldots, N_K^* = n_K)} \right\} = \lambda \left( \sum_{k=1}^{K} \xi \Delta_k - \theta_k \right) + \sum_{k=1}^{K} n_k \log \left( \frac{\theta_k}{\xi \Delta_k} \right).
\]

The Kullback-Leibler information is

\[
KL(\xi) = \sum_{n_1=0}^{\infty} \ldots \sum_{n_K=0}^{\infty} \log \left\{ \frac{Pr(N_1 = n_1, \ldots, N_K = n_K)}{Pr(N_1^* = n_1, \ldots, N_K^* = n_K)} \right\} \times Pr(N_1 = n_1, \ldots, N_K = n_K)
\]

\[= \lambda \left( \sum_{k=1}^{K} \xi \Delta_k - \theta_k \right) + \sum_{k=1}^{K} \lambda \theta_k \log \left( \frac{\theta_k}{\xi \Delta_k} \right).
\]

Setting the first derivative of \( KL(\xi) \) with respect to \( \xi \) equal to zero and solving for \( \xi \) results in \( \xi^* \). Because the second derivative is positive, \( \xi^* \) minimizes \( KL(\xi) \).

The next theorem compares the forecasts for \( Z \) and \( Z^* \).

Theorem 6: Let \( N_i \) and \( N_i^* \) be the number of events in the \( k \)-th interval for \( k = 1 \) and \( 2 \) for the processes \( Z \) and \( Z^* \). Given \( \lambda \), \( N_i \) has a Poisson distribution with mean \( \lambda \theta_i \), and \( N_i^* \) has a Poisson distribution with mean \( \lambda \xi \Delta_k \) where \( \xi^* \) is defined in equation A4. Under both the negative binomial and Robbins models,

\[
E(N_i^*[N_i^* = n] < E(N_i)[N_i = n] \quad \text{if and only if} \quad \theta_i / \Delta_i < \theta_i / \Delta_i^*
\]

Proof: \( \xi^* \) can be expressed as a convex combination of \( \theta_i / \Delta_i \) and \( \theta_i / \Delta_i^* \), that is,

\[
\xi^* = w \theta_i / \Delta_i + (1 - w) \theta_i / \Delta_i^*.
\]
where $w = \Delta_1/(\Delta_1 + \Delta_2)$. Therefore, the inequality $\theta_1/\Delta_1 < \theta_2/\Delta_2$ implies the further inequalities

$$0 < \frac{\theta_1}{\Delta_1} < \xi^* < \frac{\theta_2}{\Delta_2}.$$ 

Though the negative binomial model, which assumes that $\lambda$ has a gamma distribution, is a special case of Robbins’ model, which does not specify $\lambda$’s distribution, we prove the result for the two models separately.

**Negative binomial model.** The result for the negative binomial model follows from simple algebra. The forecasts are

$$E(N_2|N_1 = n) = \theta_2 \frac{(\alpha + n)}{(\beta + \theta_1)},$$

$$E(N_2^\alpha|N_1^\alpha = n) = \xi^* \theta_2 \frac{(\alpha + n)}{(\beta + \xi^* \Delta_1)}.$$

The bias is

$$E(N_2|N_1 = n) - E(N_2^\alpha|N_1^\alpha = n) = (\alpha + n) \left( \frac{\theta_2 \Delta_2}{\Delta_1 + \Delta_2} \right) \left( \frac{\theta_2 - \theta_1}{\Delta_2} \right).$$

Consequently,

$$E(N_2|N_1 = n) > E(N_2^\alpha|N_1^\alpha = n)$$

if and only if $\Delta_1/\Delta_2 > \theta_1/\theta_2$.

**Robbins’ model.** Let $G(\lambda)$ be the distribution of $\lambda$. The conditional expectations are

$$E(N_2|N_1 = n) = \theta_2 E(\lambda|\theta_1, n),$$

$$E(N_2^\alpha|N_1^\alpha = n) = \Delta_2 \xi^* E(\lambda|\Delta_1 \xi^*, n).$$

Because $\theta_2 > \Delta_2 \xi^*$, the proof will follow if we can show that for fixed $n$, $E(\lambda|\theta_1, n)$ is a decreasing function of $\theta$, which implies that

$$E(\lambda|\theta_1, n) > E(\lambda|\Delta_1 \xi^*, n)$$

because $\theta_1 < \Delta_1 \xi^*$ by hypothesis.

To proceed,

$$E(\lambda|\theta_1, n) = H(\theta_1, n + 1)/H(\theta, n)$$

where:

$$H(\theta, n) = \int_0^\infty \exp(-\lambda) \lambda^\alpha dG(\lambda).$$

Because $H(\theta, n)$ is the Laplace transform for $\lambda^\alpha dG(\lambda)$, it is finite, and we are able to interchange integrals and derivatives with respect to $\theta$. We show that the derivative of $log[E(\lambda|\theta_1, n)]$ is negative, so that it is a decreasing function of $\theta$.

$$\frac{d}{d\theta} \log[E(\lambda|\theta_1, n)] = \frac{d}{d\theta} \log[H(\theta, n + 1)] - \frac{d}{d\theta} \log[H(\theta, n)]$$

$$= \frac{H(\theta, n + 2)}{H(\theta, n + 1)} + \frac{H(\theta, n + 1)}{H(\theta, n)}$$

$$< 0$$

if and only if

$$H(\theta, n + 1)^2 < H(\theta, n + 2) H(\theta, n).$$

The last line is a consequence of the Cauchy-Schwartz inequality and the following identifications:

$$f(\lambda) = \lambda; g(\lambda) = 1$$

and

$$d\mu(\lambda) = \exp(-\lambda \theta) \lambda^\alpha dG(\lambda).$$

Then

$$H(\theta, n + 1) = \int_0^\infty f(\lambda) g(\lambda) d\mu(\lambda)$$

$$H(\theta, n + 2) = \int_0^\infty f(\lambda)^2 d\mu(\lambda)$$

$$H(\theta, n) = \int_0^\infty g(\lambda) d\mu(\lambda).$$

The Cauchy-Schwartz inequality is

$$\left( \int_0^\infty f(\lambda) g(\lambda) d\mu(\lambda) \right)^2 \leq \int_0^\infty f(\lambda)^2 d\mu(\lambda) \int_0^\infty g(\lambda)^2 d\mu(\lambda)$$

with strict equality if and only if $f$ and $g$ are linearly related. Consequently, we have a strict inequality in our application. The proof can be reversed so that the "if and only if" statement of the result is true. 

**Bias Due to Skewed Distributions**

**Theorem 7.** Let $T$ be a random variable with support on the non-negative integers and cumulative distribution function, $F$. Suppose the mean, $\mu$, of $T$ is finite. Then

$$E(T - \mu|T > \mu) = \frac{F(\mu)}{1 - F(\mu)} \left[ -E(T - \mu|T \leq \mu) \right]$$

**Proof.** Use conditional expectations to express the mean of $T$ as

$$\mu = E(T|T \leq \mu)P(T \leq \mu) + E(T|T > \mu)P(T > \mu)$$

$$= E(T|T \leq \mu)F(\mu) + E(T|T > \mu) \{1 - F(\mu)\}.$$ 

Then

$$0 = E(T - \mu|T \leq \mu)F(\mu) + E(T - \mu|T > \mu) \{1 - F(\mu)\},$$

which proves the theorem after we rearrange terms. \(\text{□}\)

In particular,

$$E(T - \mu|T > \mu) \begin{cases} > -E(T - \mu|T \leq \mu) & \text{if } F(\mu) > .5 \\ = -E(T - \mu|T \leq \mu) & \text{if } F(\mu) = .5 \\ < -E(T - \mu|T \leq \mu) & \text{if } F(\mu) < .5. \end{cases}$$

The distribution of $T$ is positively (negatively) skewed if $F(\mu)$ is greater than (less than) .5.

**REFERENCES**


Bickel, Peter J. and Kjell A. Doksum (1977), *Mathematical*


Lenk, Peter J. and Ambar G. Rao (1990), "Transition Times: Distributions Arising From Time Heterogeneous Poisson Processes," working paper, School of Business Administration, University of Michigan.


