Investigating the Properties of the Eskin/Kalwani & Silk Model of Repeat Buying for New Products

Peter S. Fader
University of Pennsylvania

Bruce G. S. Hardie *
London Business School

1 Introduction

Central to diagnosing and/or forecasting the performance of a new FMCG product is the decomposition of aggregate sales into trial and repeat components, that is,

$$S(t) = T(t) + R(t)$$

where $S(t)$ is total cumulative sales for the new product by time t, and $T(t)$ and $R(t)$ are cumulative trial and repeat sales, respectively, for the same time period.

A given overall aggregate sales history could be the realization of very different purchasing scenarios. For example, a low sales level for a new product could be the result of: (i) many consumers making a trial purchase but very few of them making a repeat purchase (because the product does not meet their expectations), or (ii) a low trial rate but a high level of repeat purchasing amongst the triers (because the product meets a real need among a relatively small set of buyers). Without a proper trial and repeat sales decomposition, it is impossible to determine which of these scenarios (or the infinite other possible combinations of trial and repeat patterns that could result in the same aggregate sales curve) best describes the early sales data. Consequently, the product manager might have great difficulty in developing an accurate forecast and choosing the appropriate course of action. In short, any forecast of future sales based purely on the aggregate sales data is going to be highly suspect.

When the researcher is interested in developing a forecast of a new product's future sales on the basis of early test market results, mathematical representations of the trial and repeat components of sales (i.e., $T(t)$ and $R(t)$) will be developed and calibrated using

*Peter S. Fader is Associate Professor of Marketing, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6371, USA. Bruce G. S. Hardie is Assistant Professor of Marketing, London Business School, Sussex Place, Regent's Park, London NW1 4SA, UK.

household-level panel data. These individual models can then be used to forecast trial and repeat sales, from which an overall forecast of the new product's sales can be generated. [See Hardie, Fader, and Wisniewski (1998) for a review of models for forecasting $T(t)$.] While some researchers (e.g., Greene 1974) have proposed models for $R(t)$ based on an underlying counting process (i.e., directly modeling the number of repeat purchases made by each household by time t), it is more common to use the so-called 'depth of repeat' formulation in which $R(t)$ is decomposed in the following manner:

$$R(t) = \sum_{j=1}^{\infty} R_j(t) \tag{1}$$

where $R_j(t)$ is the cumulative number of consumers that have made at least j repeat purchases by time t. (For presentational simplicity, we assume that only one unit is purchased on each repeat purchase occasion. Consequently, repeat sales volume equals the number of repeat purchase occasions.) When such a framework is utilized, a model for $R_j(t)$ ($j = 1, 2, 3, \dots$) is developed, from which a forecast of $R(t)$ can be generated.

Perhaps the best known depth of repeat model is that proposed by Eskin (1973) — and further developed by Kalwani and Silk (1980) — which draws on early work by Fourt and Woodlock (1960). Originally developed for a test market environment, the basic structure of Eskin's model has been used in simulated test market models such as BASES (Lin, Pioche, and Standen 1982). And despite its age, the basic model form is still being used in commercial test market-based new product forecasting systems — see, for example, Fader, Hardie, and Stevens (1998). In terms of real-world impact, it is one of the most important models in the academic new product forecasting literature. However, little is known about the properties of this model. For example, what insights, if any, can be derived from the model parameters? Just how accurate and robust are its forecasting capabilities? What are the limits of its performance (e.g., when does it break down)? The purpose of this paper is to answer these questions by exploring the properties of this model.

The remainder of the paper is organized as follows. In section 2, we review Eskin's model of repeat buying for new FMCG products, along with Kalwani and Silk's extensions, describing the model structure and developing the log-likelihood function used to estimate model parameters. We then examine the issue of interpreting the model's parameters, asking what insights into the structure of the repeat buying process can be gained from examining their estimated values. In section 4, we explore the forecasting performance of the model; in particular, we examine how the amount of data used for model calibration impacts on forecast accuracy and seek to understand whether or not there are any market environments in which the model will always produce inaccurate forecasts. The paper concludes (section 5) with a brief discussion of the implications of the results of the empirical analyses for developers of new product forecasting models.

2 The Eskin/Kalwani & Silk Model

Starting with the depth of repeat decomposition of equation 1, Eskin (1973) contends that further insights can be gained by decomposing $R_j(t)$ in the following manner:

$$R_j(t) = \sum_{t_{j-1}=1}^{t} F(t_j \,|\, t_{j-1})\,[R_{j-1}(t_{j-1}) - R_{j-1}(t_{j-1}-1)],\; j = 1, 2, \ldots \tag{2}$$

where $F(t_j \,|\, t_{j-1})$ is the cumulative proportion of consumers who have made a j^{th} repeat purchase by time t_j, given that their $j-1^{th}$ repeat purchase occurred at time t_{j-1}.[1] ($j = 0$ denotes the trial purchase.) Note that $F(t_j \,|\, t_{j-1})$ can also be interpreted as the probability that a consumer who made a $j-1^{th}$ repeat purchase at time t_{j-1} has made a j^{th} repeat purchase by time t_j (i.e., the cdf of the time to the j^{th} repeat purchase, conditional on the time of the $j-1^{th}$ repeat purchase). We therefore let the random variable T_j denote the time (since the launch of the new product) at which a household makes its j^{th} repeat purchase, the realization of which is t_j.

After examining a number of empirical repeat-purchase curves (i.e., $F(t_j \,|\, t_{j-1})$), Eskin notes three characteristics:

i. The curve for each depth of repeat class (j) exhibits an exponential growth pattern leveling off to an asymptote — as is typically associated with a cumulative trial curve.

ii. Across depth of repeat classes, the curves are approximately parallel.

iii. The asymptote increases, at a decreasing rate, as j increases; this means, for example, that the proportion of consumers who have made their j^{th} repeat purchase within 52 weeks of their $j-1^{th}$ repeat purchase increases with j.

Based on these observations, Eskin proposes that the Fourt and Woodlock (1960) model of trial be used to model the cumulative depth of repeat curves. Using the continuous time representation of the Fourt and Woodlock model, we have

$$F(t_j \,|\, t_{j-1}) = p_{jt_{j-1}}[1 - e^{-\lambda_{jt_{j-1}}(t_j - t_{j-1})}] + \delta_{jt_{j-1}}(t_j - t_{j-1}) \tag{3}$$

Given this model formulation, Eskin makes three assumptions in order to come up with a parsimonious model:

[1] Time is measured on the positive real line, where the origin corresponds to the launch of the new product. By convention, integer values correspond to the end of each week; for example, $t = 2$ corresponds to the end of the second week since the launch of the new product. Current data collection technology is such that time of purchase is *recorded* on the real time line (e.g., on a daily or even hourly basis). However, in implementation, the model *processes* time in an integer manner (i.e., the data are treated as *grouped* data). Eskin (1973, footnote 2, p. 118) notes that when data are processed at a weekly level, a first repeat purchase cannot be made in the same week as its corresponding trial purchase. When such an event does occur in the data, the first repeat purchase must be coded as occuring in the following week. Following this convention, the earliest a j^{th} repeat purchase could occur is in week j+1. Consequently, the limits of the summation in equation 2 should really be j and $t-1$. The logic is as follows: the earliest time a $j-1^{th}$ repeat purchase could have occured is week j, while the latest time such a purchase could have occurred for a j^{th} repeat to occur in week t is week $t-1$. The summation limits of 1 and t in equation 2 are used to be consistent with the corresponding expression given in Eskin (1973).

i. $\lambda_{jt_{j-1}} = \lambda \; \forall j, t_{j-1}$. This implies that the average interpurchase times are constant across all depth of repeat classes.

ii. $\delta_{jt_{j-1}} = \delta \; \forall j, t_{j-1}$. As this linear 'stretch factor' is designed to capture heterogeneity in consumer buying rates, this implies that the level of heterogeneity is constant across all depth of repeat classes.

iii. The ultimate conversion proportions, $p_{jt_{j-1}}$, follow the pattern $p_{jt_{j-1}} = p_\infty(1 - e^{-\theta j})$ for $j = 2, 3, \ldots$. Initial analyses reported in Eskin (1973) indicate that p_∞ is less than 1. However, subsequent experience reported in Eskin and Malec (1976) indicates that $p_\infty \approx 1$. (This is consistent with the authors' own experience with such models.)

To use this model for forecasting purposes, the parameters λ, δ, p_1, p_∞, and θ are estimated using repeat purchasing data over a specified calibration period, pooling across all observed depth of repeat classes. Using equation 3, future values of $F(t_j \,|\, t_{j-1})$ are predicted, including for higher depth of repeat classes, from which forecasts of $R_j(t)$ and therefore $R(t)$ are derived for future periods using equations 2 and 1, respectively. (Clearly such forecasts are conditional upon a forecast of trial purchasing, which appears in equation 2 when $j = 1$.)

The δ term above was originally proposed by Fourt and Woodlock as a means of accounting for heterogeneity in consumer buying rates. An alternative — and less ad-hoc — approach is to explicitly model heterogeneity using a mixing distribution, such as the gamma distribution (Anscombe 1961). Such an approach is proposed by Kalwani and Silk (1980), who assume: (i) time to conversion from the $j-1^{th}$ to the j^{th} depth of repeat class is distributed exponentially with rate parameter λ_j, (ii) the rate parameter is distributed according to a gamma mixing distribution with shape parameter r_j and scale parameter α_j, and (iii) only a proportion p_j of those who made a $j-1^{th}$ repeat purchase will eventually make a j^{th} repeat purchase. The resulting mixture model is:

$$F(t_j \,|\, t_{j-1}) = p_j \left[1 - \left(\frac{\alpha_j}{\alpha_j + t_j - t_{j-1}} \right)^{r_j} \right] \tag{4}$$

While Kalwani and Silk (1980) examine the properties of this model, they do not fully explore its use as a replacement for equation 3. Eskin's assumptions regarding λ and δ are equivalent to assuming constant r and α across depth of repeat classes. Adding his assumption regarding the development of the ultimate conversion proportions, p_j, we arrive at the following model as a replacement to equation 3:

$$F(t_j \,|\, t_{j-1}) = p_j \left[1 - \left(\frac{\alpha}{\alpha + t_j - t_{j-1}} \right)^{r} \right], \; j \geq 1 \tag{5}$$

where

$$p_j = \begin{cases} p_1 & \text{if } j = 1 \\ p_\infty(1 - e^{-\theta j}) & \text{otherwise} \end{cases} \tag{6}$$

We will call this specification the E/KS model, after Eskin, and Kalwani and Silk.

Estimates of the parameters r, α, p_1, p_∞, and θ are obtained using maximum likelihood (ML) estimation in the following manner. Let $R_j(i, i')$ be the number of consumers who made their $j-1^{th}$ repeat purchase in week i' who have made a j^{th} repeat purchase by week i; it follows that $R_j(t) = \sum_{i'=1}^{t} R_j(t,i')$.[2] Let T_c be the length of the model calibration period (in weeks), and J_c the maximum depth of repeat class observed in the calibration period data. Recognizing the grouped nature of the data, the log-likelihood function is:

$$
\begin{aligned}
LL = \sum_{j=1}^{J_c} \sum_{t_{j-1}=j}^{T_c-1} \Bigg\{ & \sum_{t_j=t_{j-1}+1}^{T_c} [R_j(t_j,t_{j-1}) - R_j(t_j-1,t_{j-1})] \times \\
& \ln\left[p_j\left(\frac{\alpha}{\alpha+(t_j-1)-t_{j-1}}\right)^r - p_j\left(\frac{\alpha}{\alpha+t_j-t_{j-1}}\right)^r\right] \\
& + [R_{j-1}(t_{j-1}) - R_{j-1}(t_{j-1}-1) - R_j(T_c,t_{j-1})] \ln\left[1-p_j+p_j\left(\frac{\alpha}{\alpha+T_c-t_{j-1}}\right)^r\right]\Bigg\}
\end{aligned}
\tag{7}
$$

where p_j is given by equation 6. Maximizing this likelihood function using standard numerical optimization methods gives us the ML estimates of the model parameters.

3 Interpreting Model Parameters

How can we interpret the five parameters — r, α, p_1, p_∞, and θ —associated with the E/KS model? According to Eskin, and Kalwani and Silk, it appears that there is a nice behavioral story associated with them. For example, p_1 and p_j, $j = 2, 3, \ldots$ (as calculated using p_∞ and θ) should be interpreted as the proportion of consumers who will ultimately convert from one repeat class to the next. We may also try to interpret the estimated parameters of the gamma distribution in the usual manner. For example, the mean buying rate for the new product should be given by r/α. Furthermore, r is often interpreted as a measure of consumer heterogeneity — the coefficient of variation for the gamma distribution is $1/\sqrt{r}$, which means that the higher the value of r, the lower the degree of heterogeneity across the repeat buying consumers.

While we can try to ascribe such meanings to the model parameters, they do not reflect the true underlying process. To illustrate this, let us consider purchasing in a world with no dynamics in the purchasing process; that is, a stationary purchasing process as characterized by the NBD model. Modeling repeat buying as a Poisson *counting* process with gamma heterogeneity (i.e., the NBD model) is equivalent to modeling it as a (stationary) exponential *timing* process with gamma heterogeneity (Gupta and Morrison 1991). We create 20 simulated datasets, each containing 52 weeks of purchasing by 500 households, in which there are no dynamics in the purchasing process. One half of the simulated datasets correspond to the gamma parameter pair $r = 0.5$ and $\alpha = 10$, while the other half correspond to the parameter pair $r = 1.5$ and $\alpha = 30$.

Each simulated dataset is created in the following manner. Starting with household #1, a value of λ is drawn from the gamma distribution. Using this value of λ, an exponential interpurchase time is simulated, giving us the household's value of t_1, the time of its first repeat purchase. If $t_1 > 52$, the household is deemed to have made zero repeat purchases and the procedure moves on to the next household. On the other hand, if

[2] In light of the points raised in footnote 1, it would be more correct to write these summation limits as j and $t-1$.

$t_1 < 52$, another exponential interpurchase time is simulated and added to t_1 to give us the household's simulated value of t_2, the time of its second repeat purchase. If $t_2 > 52$, the household is deemed to have made only one repeat purchase and the procedure moves on to the next household. If $t_2 < 52$, the whole process continues for this household until $t_j > 52$, at which time the procedure moves to the next household. For the new household, a value of λ is drawn from the same gamma distribution, and so on, until the purchasing by all 500 households has been simulated.

For each simulated dataset, we estimate the parameters of the E/KS model by maximizing its log-likelihood function (equation 7). Four different calibration period lengths (T_c) are examined: 13, 26, 39, and 52 weeks. For each parameter set and calibration period, we compute the average of the estimated model parameters, along with the average of the estimated mean of gamma distribution. These numbers are reported in Table 1.

True	T_c	r	α	p_1	p_∞	θ	r/α
$r = 0.5, \alpha = 10$	13	6.503	40.189	0.420	1.000	2.936	0.159
	26	3.615	25.258	0.525	1.000	1.856	0.151
	39	2.100	14.011	0.591	1.000	1.383	0.152
	52	1.788	11.890	0.637	1.000	1.675	0.153
$r = 1.5, \alpha = 30$	13	14.715	164.208	0.618	0.995	4.440	0.089
	26	13.720	166.183	0.707	1.000	2.781	0.082
	39	6.525	88.232	0.792	1.000	4.106	0.077
	52	4.661	64.659	0.838	1.000	5.528	0.077

Table 1: Average Estimated E/KS Model Parameters

If the apparent behavioral story associated with the parameters of the E/KS model is true, we should find $\hat{p}_1 = \hat{p}_\infty = 1.0$ and $\hat{\theta} \to \infty$ in all cases, in accordance with the stationary Poisson process. In sharp contrast to this stationary data generating process, the parameter estimates strongly suggest the presence of dynamics in the consumers' repeat buying behavior for the new product. For example, referring to the row corresponding to $T_c = 52$ in the $r = 0.5, \alpha = 10$ condition, the following story would be told: Approximately 64% (p_1) of the triers will eventually make at least one repeat purchase. 96% of first repeaters ($p_\infty(1 - e^{-\theta j})$) will eventually make a second repeat purchase, and the conversion proportion for the third depth of repeat class is 99%. Within each depth of repeat class, the average buying rate (r/α) is 0.153 (i.e., every six and a half weeks).

However, the true nature of the underlying data is such that there are absolutely no dynamics (e.g., "dropping out" of the market) in the consumers' repeat buying behavior for the new product. The true, underlying average buying rate is low — 0.05 in both conditions — which means that there are a number of households that make zero purchases over a 52-week time frame. Rather than correctly capturing this pattern through the r and α parameters, the model is using the p_1 parameter to allow for these apparent non-buyers. (Given this phenomenon, it is not surprising that the estimated mean buying rate is grossly over-estimated.) Consequently, we are interpreting the lack of buyers in the calibration period as evidence of a structural characteristic of the market — that is, a group of households who did not like the new product and therefore will never make a repeat purchase — when in actual fact all households will eventually buy. It is simply that the true average buying rate is low.

This point is further illustrated by looking at the differences across the parameter set and calibration period conditions. As the calibration period lengthens, the percentage of households in the simulated panel that have not yet made a (repeat) purchase decreases. It is therefore not surprising to see the estimate of p_1 increasing as T_c lengthens. (It is interesting to note the relative stability of the estimated mean of the gamma mixing distribution (r/α) — despite the variation the estimates of r and α — as the length of the calibration period increases, further demonstrating the role played by the p_1 parameter.)

We also note that the estimates of p_1 are larger in the $r = 1.5, \alpha = 30$ condition. When $r < 1$, the gamma density is strictly decreasing from an infinite peak at 0. On the other hand, when $r \geq 1$, the mode of the gamma density occurs at $(r - 1)/\alpha$. Therefore, the $r = 0.5$ condition will see more households with low values of λ than the $r = 1.5$ condition, and therefore (relatively) more households not making any purchase in the 52 week simulation period. It follows that larger estimated values of p_1 (for a given T_c) are observed in the $r = 1.5, \alpha = 30$ condition.

In conclusion, what appears to be a plausible set of interpretations for the E/KS model parameters is in fact non-existent. Thus when evaluating this model in terms of its ability to provide insights into the underlying buying behavior of the market, it is very poor. Even for a completely stationary market, the parameter estimates tell a misleading story of dynamics in repeat buying behavior. And as we move from a purely stationary market (i.e., NBD) to a more realistic one that incorporates nonstationarity, we would expect these problems with parameter interpretability to become even worse. Furthermore, the nature and extent of the biases in the estimated values of the parameters are somewhat unpredictable (e.g., they vary based on characteristics of the market such as the degree of consumer heterogeneity).

But despite all these problems with parameter interpretability, we should not discard the model just yet. After all, the primary motivation for deriving and using a model such as E/KS is to generate a medium-term sales forecast for the new product. If the model can provide accurate forecasts — even if the associated parameters are completely useless — then it might still be a perfectly acceptable model from an 'engineering' perspective. With this objective in mind, we now turn our attention to the issue of assessing the forecasting performance of the E/KS model across a broad set of nonstationary (simulated) markets.

4 Examining Forecasting Performance

For any market researcher interested in using the E/KS model to generate repeat sales forecasts, a key implementation question concerns how many weeks of data are needed for model calibration in order to generate a 'good' forecast (i.e., what value T_c should take). The length of calibration period is of great interest to managers — test markets are a costly exercise, and the associated costs increase with the duration of the test, as do the opportunity costs of not "going national" earlier. Moreover, longer test periods give competitors more of a chance to evaluate the performance of the new product and possibly launch a similar product sooner than would normally be expected.

Eskin (1973) provides some initial insights into this problem. In his empirical analysis, he considers models based on calibration periods of 12, 24, and 52 weeks of data, and examines the forecasts for different periods out to a maximum of 52 weeks from product launch. As would be expected, predictions based on a 12 week calibration period are the worst performers — see Eskin (1973, pp. 123–4) for a complete discussion of his findings. However, we cannot draw any general conclusions from Eskin's work since only one new product dataset is examined.

In order to explore the overall forecasting performance of the E/KS model, along with the impact of T_c on forecast quality, we will use a set of simulated datasets. The primary reason for using simulated datasets is that it enables us to explore the robustness of the E/KS model in *many* different market settings. The simulated datasets are generated using an advanced model of repeat buying for a new FMCG product recently proposed by Fader and Hardie (1998). We now turn our attention to this model.

4.1 The NSEG Model

It is common to model repeat buying behavior in mature FMCG markets using the NBD model (Ehrenberg 1988), which assumes that a consumer's underlying buying rate is constant (i.e., stationary). If we think about the repeat buying of a new product during the early phase of its life, we expect there will often be a period of instability during which consumers' preferences for the new product are evolving; that is, there is some nonstationarity in consumers' underlying buying rates. As consumers gain more experience with the product, we would expect their preferences to "settle down" and therefore their buying patterns can then be better characterized by the usual NBD model.

Starting with the aforementioned exponential-gamma model, the *timing* counterpart of the *counting* NBD model, Fader and Hardie (1998) propose a nonstationary extension, which they call the nonstationary exponential-gamma, or NSEG, model. Nonstationarity is modeled using an individual-level renewal process for the buying rates — similar to that which lies at the heart of Howard's Dynamic Inference Model (Howard 1965).

The NSEG model is based on five assumptions:

i. The probability of a household that made a trial purchase ever making a repeat purchase is π.

ii. Interpurchase times are distributed exponentially with rate parameter λ.

iii. Household-level purchase rates, λ, are distributed across the population according to a gamma distribution with shape parameter r and scale parameter α. (The parameters of this distribution are time-invariant.)

iv. Following its j^{th} repeat purchase, a household renews its value of λ with probability γ_j. The depth of repeat-specific renewal probability — constant across households — is given by $\gamma_j = 1 - \psi(1 - e^{-\theta j})$.

v. At each renewal, a household receives a value of $\lambda = 0$ with probability ϕ. With probability $1 - \phi$, the household draws a new value of λ, independent of its previous one, from the same gamma distribution.

The first assumption is consistent with the notion of the p_1 parameter in the E/KS model. (However, because of the way the model is formulated, it does not spuriously capture the effects of heterogeneity, as is the case with the E/KS model.) The fourth assumption sees the use of a formula similar to that for p_j in equation 6, but the purpose is slightly different. The logic behind this expression for γ_j, the probability that a renewal occurs at depth of repeat level j, is as follows: we would expect that the probability of a consumer revising her preferences following a purchase (and consumption) experience would decrease as she gains more experience with the new product (i.e., as she moves to a higher depth of repeat level). Looking closely at the equation for γ_j, we note that as j increases, γ_j tends to $1-\psi$. Therefore, if $\psi = 1$, the probability of a renewal tends to zero as a consumer moves to higher depth of repeat levels; in other words, the model evolves to a stationary process which we interpret as the stabilization of consumer preferences. If $\psi < 1$, individual consumer preferences do not stabilize — which means there is long-term nonstationarity in the marketplace. The fifth assumption is presented as a paramorphic, as opposed to strictly behavioral, representation of how preferences for the new product evolve. The "spike at zero" — receiving a value of $\lambda = 0$ with probability ϕ — is simply a mechanism by which consumers can "drop out" of the market for the new product, even after making several repeat purchases. Drawing a value of zero upon a renewal is viewed as being equivalent to rejecting the new product from future purchase consideration.

Reflecting on assumptions (iv) and (v), we see that the probability of a randomly chosen household ever making a j^{th} repeat purchase, conditional on the fact they have made a $j-1^{th}$ repeat purchase, is $1 - \gamma_{j-1}\phi$ for $j = 2, 3, \ldots$. Thus the NSEG model can formally (and directly) capture a phenomenon the E/KS model attempts to capture indirectly via the p_j terms. Unlike the E/KS model, the parameters of the NSEG model are fully interpretable, so their estimates provide insights into the underlying buying behavior of the market. (Furthermore, Fader and Hardie (1998) show that the NSEG parameters are robust across calibration periods of different lengths, unlike the patterns for the E/KS model seen earlier in Table 1.)

The NSEG model is a very flexible model that can capture many patterns of repeat buying behavior. Probably the simplest model of repeat buying which it nests is the NBD model. If $\gamma_j = 0\ \forall j$, we have a stationary process. (This is associated with $\theta \to \infty$ and $\psi = 1$.) Furthermore, if $\pi = 1$ and $\phi = 0$, we have the two parameter exponential-gamma model of stationary repeat buying behavior.[3] Relaxing the assumption that $\pi = 1$ gives us the timing equivalent of NBD with "spike at zero" model where $1-\pi$ is the size of the structural "never buyers" segment.

Relaxing the constraint that $\gamma_j = 0\ \forall j$ results in a model which can capture various forms of nonstationarity in repeat buying behavior. If $\psi = 1$, the probability of a renewal occurring tends to zero as a consumer moves to higher depth of repeat levels. This means that the initial nonstationary repeat buying process evolves to a stationary process as the product transitions from being "new" to "established". Therefore the NSEG model is consistent with the notion of nonstationary buying behavior during the early stages of a new product's life and stationary buying behavior — as characterized by the NBD model — once it has become established in the marketplace. If $\psi < 1$, the repeat buying process is always nonstationary as $\gamma_j > 0\ \forall j$. In particular, if $\theta \to \infty$, the probability of renewal is constant $(1-\psi)$ across all depth of repeat levels, while for small θ, the probability of renewal tends to the constant $1-\psi$ as a consumer moves to higher depth of repeat levels.

[3]When we fit the NSEG model to stationary exponential-gamma data (such as that used in Section 3), the underlying parameters are recovered (e.g., the parameters that capture the effects of nonstationarity indicate the absence of nonstationarity) — cf. the E/KS model. The reason for this is that the "dropping out" phenomenon, along with other forms of nonstationarity, are formally modeled at the household-level, with the household-level likelihood function explicitly accounting for the dependence in buying rates across purchases. In contrast, the E/KS model ignores the dependency across purchases for any given household.

The ability of the NSEG model to capture a number of different repeat buying patterns makes it the ideal base for a simulation analysis designed to gain insight into the properties of the E/KS model.[4]

4.2 Simulation Design

In order to use the NSEG model to simulate new product purchasing, we must first specify the six model parameters — $\pi, r, \alpha, \psi, \theta, \phi$. The approach taken is as follows. We first specify two buying rate conditions, corresponding to a fast and slow purchase cycle. The 'fast' condition assumes an average interpurchase time of 4 weeks while the 'slow' condition corresponds to an average interpurchase time of 20 weeks (i.e., r/α of 0.25 and 0.05, respectively). We then specify three heterogeneity conditions, $r = 0.5, 1.0,$ and 1.5, representing (relatively) high-to-low heterogeneity, respectively. (This range of values is consistent with commonly observed levels of gamma heterogeneity (Fader and Hardie 1998, Morrison and Schmittlein 1988).) This implies six sets of values for r and α.

Beyond these two structural characteristics of each simulated market, we introduce different types of nonstationarity using the remaining four NSEG parameters, π, ϕ, ψ, and θ. Since the first three of these parameters can only take on values in the interval $[0, 1]$, we make three independent draws from a uniform distribution to obtain their values. The θ parameter is determined via a draw from the exponential distribution with scale parameter 1. We repeat this process 25 times. Each of these vectors of π, ϕ, ψ, θ are then combined with each of the six r, α combinations to yield 150 different sets of parameters from which the simulated datasets are created.

The specific values of π, ϕ, ψ, and θ are reported in Table 2. We see a wide variety of 'market conditions', with differing levels of nonstationarity and likelihood of ultimate rejection for the new product. For example, with draw 5, almost 63% of triers (π) will eventually make at least one repeat purchase. The probability of a renewal occurring after the first repeat purchase is 0.377 ($\gamma_j = 1 - \psi(1 - e^{-\theta j})$). By the fifth repeat purchase, this probability has declined to 0.079, close to the equilibrium probability of renewal of 0.076 ($1 - \psi$). Consequently, repeat buying behavior becomes relatively stationary. On any given renewal, there is less than a 5% chance ($\phi = 0.044$) of the consumer rejecting the new product. Therefore, 99.7% ($1 - \gamma_j\phi$) of those consumers who make a fifth repeat purchase will eventually make a sixth repeat purchase.

On the other hand, draw 19 represents a completely different type of new product. Almost 98% ($1 - \pi$) of triers reject the new product and never make a repeat purchase. Of those who do make a repeat purchase, the chance of subsequent rejection of the new product is high. The probability of a renewal occuring after the first purchase is a high 0.843. Whenever a renewal occurs, there is a 19% chance (ϕ) of the consumer rejecting the

[4]The reader may ask why a researcher would choose to use the E/KS model, when a model such as NSEG avoids the problems of parameter interpretability. The answer is due to differences in the ease of implementing each model. The parameters for the E/KS model can be obtained in a simple modeling environment such as Excel. On the other hand, the NSEG model requires the use of a more general software package such as MATLAB. In such a modeling environment, estimating the E/KS model parameters and generating sales forecasts takes a few seconds using a mid-level PC, compared to several hours for the NSEG model. Therefore, if the analyst is only interested in generating sales forecasts, the E/KS model merits serious consideration. Furthermore, incorporating covariate effects in the E/KS model is a relatively straightforward exercise, and the resulting model can still easily be implemented in Excel. This is not the case for the NSEG model.

Draw	π	ϕ	ψ	θ	Draw	π	ϕ	ψ	θ
1	0.615	0.792	0.922	0.304	14	0.067	0.468	0.564	2.326
2	0.331	0.583	0.492	1.312	15	0.964	0.601	0.093	1.622
3	0.866	0.180	0.480	1.943	16	0.426	0.825	0.123	0.074
4	0.472	0.370	0.914	0.150	17	0.663	0.890	0.631	2.793
5	0.626	0.044	0.924	1.121	18	0.070	0.098	0.730	0.400
6	0.956	0.581	0.143	0.400	19	0.026	0.189	0.235	1.106
7	0.158	0.572	0.058	0.513	20	0.427	0.482	0.266	0.726
8	0.400	0.878	0.742	0.871	21	0.932	0.910	0.405	0.435
9	0.776	0.303	0.517	1.839	22	0.539	0.048	0.047	1.446
10	0.497	0.002	0.518	5.956	23	0.457	0.495	0.643	0.537
11	0.387	0.523	0.003	0.299	24	0.172	0.746	0.622	0.084
12	0.061	0.089	0.091	0.383	25	0.835	0.578	0.287	0.415
13	0.904	0.477	0.155	0.046					

Table 2: Simulation Model Parameters

new product. Even in equilibrium ($j \to \infty$), the probability of a (surviving) consumer rejecting the product after a repeat purchase is almost 15% ($\phi(1-\psi)$). Clearly this simulated product is a failure.

For each of the 150 parameter sets, we create one dataset of simulated purchasing by 500 households over a 52 week period. Each simulated dataset is created in the following manner. Starting with household #1, the simulation begins by drawing a uniform random variate to determine whether it will ever make a repeat purchase (with probability π). If this is the case, a value of λ is drawn from the gamma distribution. Using this value of λ, an exponential interpurchase time is simulated, giving us the household's simulated value of t_1, the time of its first repeat purchase. (We assume $t_0 = 0$ for all households.) If $t_1 > 52$, the household is deemed to have made zero repeat purchases and the procedure moves on to the next household. If $t_1 < 52$, a uniform random number is drawn to determine whether the household retains its value of λ (with probability $1-\gamma_1$) or whether a renewal occurs (with probability γ_1), in which case a new value of λ is drawn. Another uniform random number is drawn in the process of determining the new value of λ. With probability ϕ, a value of $\lambda = 0$ is drawn and the household is deemed to have rejected the new product (and the procedure moves on to the next household). If the new value of λ is drawn from the gamma distribution (with probability $1-\phi$), or no renewal has occurred, another exponential interpurchase time is simulated and added to t_1 to give us the household's simulated value of t_2, the time of its second repeat purchase. If $t_2 > 52$, the household is deemed to have made only one repeat purchase and the procedure moves on to the next household. If $t_2 < 52$, the whole process continues for this household until $t_j > 52$ or a value of $\lambda = 0$ is drawn when a renewal occurs, at which time the procedure moves to the next household. For the new household, a uniform random variate is drawn to determine whether it will ever make a repeat purchase (with probability π), and so on, until the purchasing by all 500 households has been simulated.

Across the 150 simulated datasets, the average number of repeat purchases is 487, ranging from a maximum of 2943 to a minimum of 14. This broad range suggests that we have captured a wide variety of market conditions, which should provide a rigorous and generalizable test of the forecasting performance of the E/KS model.

4.3 Results

For each of the 150 simulated datasets, we estimate the parameters of the E/KS model by maximizing its log-likelihood function (equation 7) for each of three different calibration period lengths (T_c = 12, 24 and 52 weeks) chosen to be consistent with Eskin (1973). In all cases, the model parameters are used to create forecasts for the 52 week period using equations 1, 2, 5, and 6. The accuracy of each forecast is measured by the absolute percentage error in the year-end sales estimate — hereafter APE_52.[5] (For $T_c = 52$, we are actually measuring fit.) In total, 450 different models are estimated and the associated forecasts generated.

We divide the results into two groups, corresponding to the fast and slow purchase cycle conditions. Within each group, we compute the average APE_52 for each level of T_c and r, averaging across the 25 simulated datasets associated with each cell. These numbers are reported in Table 3.

		Slow Purchase Cycle			Fast Purchase Cycle		
		$T_c = 12$	$T_c = 24$	$T_c = 52$	$T_c = 12$	$T_c = 24$	$T_c = 52$
	0.5	23.1%	9.0%	1.5%	11.7%	6.7%	2.0%
r	1.0	22.4	6.9	1.3	17.9	5.8	2.1
	1.5	26.2	9.5	1.6	14.6	5.9	1.9

Table 3: Average APE_52 by Product Characteristics and T_c

The first thing we note is how well the E/KS model fits the 'true' data when using the full 52 weeks to calibrate the model (i.e., $T_c = 52$), with an average APE_52 of 1.6% and 1.9% for the slow and fast purchase cycle conditions, respectively. As we look across the entire table, we can draw three key conclusions:

- *Impact of T_c:* Not surprisingly, the quality of the forecasts deteriorates as the calibration period length shortens. However, the forecasts generated with 24 weeks of data are quite accurate. It is very encouraging that the slow-cycle products (with a mean interpurchase time of 25 weeks between repeat purchases) can be forecast so well with only 24 weeks of calibration data.

- *Purchase Cycle Effects:* When shorter calibration periods are used, the E/KS model generates much more accurate forecasts in fast moving categories. In and of itself, this is not too surprising, as we will generally observe more repeat purchases than for products in slower moving categories. What is surprising, however, is that the forecast quality increases at a faster rate for slow-cycle (compared to fast-cycle products) as the calibration period lengthens. In fact, when using the full 52 weeks of data for model calibration, the slow-cycle products actually show a slightly better fit than the fast-cycle ones, although this observation has little practical value from a forecasting perspective.

- *Impact of Heterogeneity:* There are no significant main effects or interactions involving the heterogeneity (r) conditions. Thus despite the fact that different levels

[5] A commonly used alternative measure is Mean Absolute Percentage Error (MAPE). As the qualitative results mirror those of APE_52, we only report the numbers associated with the latter measure.

of heterogeneity can generate markedly different biases in the estimated E/KS parameters (as seen in Table 1), there is no associated impact on the E/KS model forecasts.

The moderate size of most of these errors offers some encouragement about the usefulness of the E/KS model for forecasting purposes. Beyond the accuracy of the model's forecasts, another relevant issue is the degree of bias (i.e., systematic over- or under-prediction). The APE_52 error measure does not give us any insight as to whether or not any such biases exist in the forecasts produced by the E/KS model. In order to examine this, we determine, for each of the 450 datasets (150 parameter combinations $\times$ 3 levels of T_c), whether the forecast generated by the E/KS model under- or over-predicts year-end sales. As with Table 3, we divide the results into two groups, corresponding to the fast and slow purchase cycle conditions. Within each group, we compute the percentage of cases in which the E/KS model under-predicted week 52 sales for each level of T_c and r, and report these numbers in Table 4.

		Slow Purchase Cycle			Fast Purchase Cycle		
		$T_c = 12$	$T_c = 24$	$T_c = 52$	$T_c = 12$	$T_c = 24$	$T_c = 52$
	0.5	56%	44%	100%	56%	48%	100%
r	1.0	64	48	100	60	36	100
	1.5	60	60	100	48	36	100

Table 4: Percent Of Models Under-Predicting by Product Characteristics and T_c

The most notable bias in this table is also the least consequential. Specifically, the models utilizing 52 weeks of calibration data underestimate the actual year-end sales level in every case. But because the magnitude of these errors, as seen in Table 3, is so small, there is little reason for concern. For the remainder of the table, there is no substantial bias in the predictions. There are some minor trends evident: (i) the 12-week models tend to show slight over-forecasts, (ii) the 24-week models tend to produce under-forecasts, and (iii) the slow purchase cycle products under-forecast slightly more often than the fast-cycle ones. But these differences are not very large in magnitude, and they do not seem to combine together in any harmful ways. And as before, there are no clear associations with the different levels of heterogeneity.

In summarizing these forecast results, we see a dramatically different picture than the one shown in our earlier discussion of parameter interpretability. The E/KS model tends to produce accurate and unbiased forecasts that appear to be quite robust across a variety of product characteristics and nonstationarity conditions.

Beyond our reporting of average values across market conditions, it is also instructive to examine some specific cases to help understand the factors that may limit the model's forecasting performance. An attractive feature of using simulated, as opposed to real, data is that it provides us with a sufficient diversity of market environments so that additional insights into the capabilities of the E/KS model can be obtained. In particular, are there any market environments in which the model breaks down?

At the heart of the simulation are the two buying rate conditions (a slow versus fast purchase cycle), three different levels of consumer heterogeneity, and the 25 sets of the parameters π, ϕ, ψ and θ. For each of the 150 'market conditions,' we compute the average value of APE_52 across the three model runs associated with each of the three calibration conditions (T_c). The distribution of these averages is presented in Figure 1. While the E/KS model performs very well overall, it appears that there may be some market environments where this is not the case.

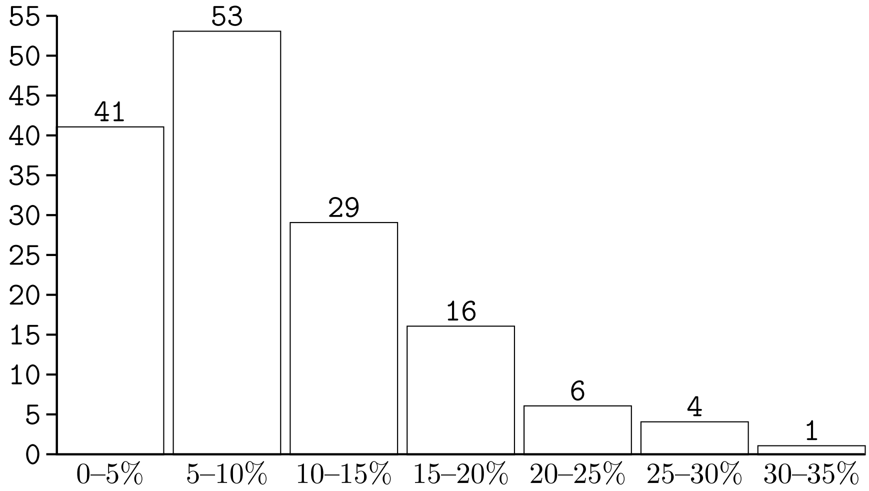

Figure 1: Distribution of Average APE_52 Across Simulation Conditions

To gain further insight into this issue, we investigate the simulation draws that consistently result in bad forecasts. For instance, the five environments that comprise the right tail of Figure 1 (i.e., APE_52 $\geq$ 25%) stem from draws 7, 12, 18, and 19 (twice) in Table 2. In carefully examining these parameter values, one striking characteristic emerges — in all four cases, the value of the π parameter takes on small values ($\pi < 0.16$). Furthermore, four of the five datasets with 20% $\leq$ APE_52 $<$ 25% also have $\pi < 0.2$. Recall that this parameter conveys the proportion of triers who will eventually make at least one repeat purchase. Since the majority of triers in these scenarios fail to make a repeat purchase, these simulated 'products' are all destined to be failures.

A low value of π is a sufficient (but not necessary) condition for a low level of overall repeat purchasing, especially for $T_c < 52$.[6] Indeed, beyond the association observed for the π parameter, there is a fairly strong relationship between APE_52 and the inverse of the number of repeat purchases associated with each of the 150 unique datasets ($\hat{\rho} = 0.467$). In other words, successful products are easier to forecast than unsuccessful ones. This suggests an interesting type of 'double jeopardy': not only do successful products have better prospects for long-run profitability, but they will also tend to receive more accurate forecasts than products with lower sales levels.

In any event, this relationship is the only pattern that emerged in our investigation of the datasets with inaccurate forecasts. There appear to be no other systematic factors that impact the ability of the E/KS model to make reasonable estimates of a new product's sales performance. Once again, the absence of such factors points to the robustness and flexibility of the E/KS model specification.

[6] This limited amount of data can lead to relatively unreliable parameter estimates, thus resulting in poor forecasts. Consequently, the problem of poor forecasts in such an environment should not be viewed as a weakness of the E/KS model; we would expect any model for $R(t)$ to have similar problems.

5 Conclusions

The so-called E/KS model of repeat buying for new FMCG products presented in this paper is a natural integration of early work by Eskin (1973) and Kalwani and Silk (1980). The basic structure of this model has been widely used in various real-world new product forecasting systems. However, very little has been known about the properties of the model. This paper has sought to address this knowledge gap.

As originally presented, the parameters of the E/KS model seemed to offer appealing interpretations that could convey useful diagnostics to managers. For example, the p_j parameters were intended to be interpreted as the percentage of consumers who will eventually make a j^{th} repeat purchase, given a $j-1^{th}$ repeat purchase. However, our analysis demonstrates that this is clearly *not* the case — even for a completely stationary market, the parameter estimates tell a misleading story of dynamics in repeat buying behavior. Therefore, from the perspective of providing insights into the underlying buying behavior of the market, the E/KS model fails.

However, this does not mean that the model should be discarded. The analysis presented in this paper demonstrates that, *from a forecasting perspective*, the E/KS model is a robust and flexible model capable of generating good forecasts of a new product's repeat sales. It only appears to produce poor forecasts for new products that exhibit high levels of post-trial consumer rejection. This leads to low repeat sales levels, making it difficult to reliably estimate the model's parameters, which in turn leads to poor forecasts.

For any market researcher interested in using the E/KS model, a key question is how many weeks of data are needed for model calibration. As would be expected, forecasting accuracy increases with the number of weeks of data available for model calibration. Furthermore, for a given number of calibration weeks, more accurate forecasts will be generated for products with a fast purchase cycle, compared to those with a slow purchase cycle. Our analysis suggests that sufficiently accurate forecasts — from a manager's perspective — can be generated using 24 weeks of data, regardless of the product's purchase cycle.

Therefore, despite the problem with parameter interpretation, the E/KS model appears to be a very good model for forecasting the repeat sales for a new FMCG product, provided the analyst is willing to use a model which provides an 'engineering' solution to the forecasting problem, as opposed to one that provides insight into the underlying buying behavior in additional to providing accurate forecasts.

A key issue to be addressed in future research concerns the incorporation of marketing mix covariate effects in the basic model, and an examination of the associated impact on forecasting performance and amount of data needed for model calibration.

References

Anscombe, F. J. (1961), "Estimating a Mixed-Exponential Response Law," *Journal of the American Statistical Association*, **56** (September), 493–502.

Ehrenberg, A. S. C. (1988), *Repeat-Buying: Facts, Theory and Applications*, 2nd edn., London: Griffin; New York: Oxford University Press.

Eskin, Gerald J. (1973), "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," *Journal of Marketing Research*, **10** (May), 115–29.

Eskin, Gerald J. and John Malec (1976), "A Model for Estimating Sales Potential Prior to the Test Market," *Proceeding 1976 Fall Educators' Conference*, Series No. 39, Chicago, IL: American Marketing Association, 230–33.

Fader, Peter S. and Bruce G. S. Hardie (1998), "Modeling the Evolution of Repeat Buying for a New Product," Wharton Marketing Department Working Paper.

Fader, Peter S., Bruce G. S. Hardie, and Robert Stevens (1998), "Trial/Repeat Modeling Revisited: A New Product Forecasting System Using Scanner Panel Data," presented at the 1998 A/R/T Forum (Keystone Resort, CO), June 7–10.

Fourt, Louis A. and Joseph W. Woodlock (1960), "Early Prediction of Market Success for New Grocery Products," *Journal of Marketing*, **25** (October), 31–8.

Greene, Jerome D. (1974), "Projecting Test Market 'Trial-Repeat' of a New Brand in Time," *1974 Combined Proceedings*, Series No. 36, Chicago, IL: American Marketing Association, 419–22.

Gupta, Sunil and Donald G. Morrison (1991), "Estimating Heterogeneity in Consumers' Purchase Rates," *Marketing Science*, **10** (Summer), 264–69.

Hardie, Bruce G. S., Peter S. Fader, and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, **17** (June–July), 209–29.

Howard, Ronald A. (1965), "Dynamic Inference," *Operations Research*, **13** (September–October), 712–33.

Kalwani, Manohar and Alvin J. Silk (1980), "Structure of Repeat Buying for New Packaged Goods," *Journal of Marketing Research*, **17** (August), 316–22.

Lin, Lynn Y. S., Alain Pioche, and Patrick Standen (1982), "New Product Sales Forecasting: Recent BASES Model Experiences in Europe and in the United States," in *Fitting Research to Turbulent Times*, 35th ESOMAR Congress, Vienna, 29th August–2nd September, 441–59.

Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, **6** (April), 145–59.