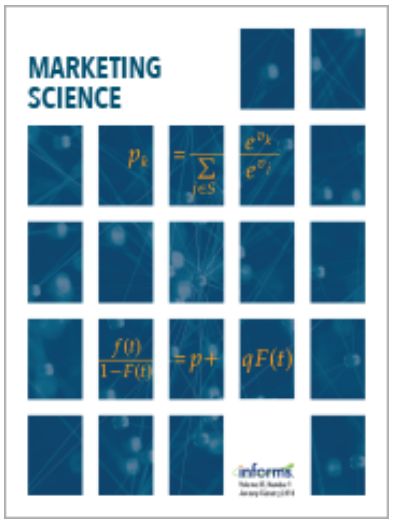


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares

Rajkumar Venkatesan, Trichy V. Krishnan, V. Kumar,

To cite this article:
Rajkumar Venkatesan, Trichy V. Krishnan, V. Kumar, (2004) Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares. Marketing Science 23(3):451-464. https://doi.org/10.1287/mksc.1040.0056

MARKETING SCIENCE
Vol. 23, No. 3, Summer 2004, pp. 451–464
ISSN 0732-2399 | EISSN 1526-548X | 04 | 2303 | 0451

DOI 10.1287/mksc.1040.0056

Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares

Rajkumar Venkatesan
School of Business, University of Connecticut, Storrs, Connecticut 06269-1041, rvenkatesan@sba.uconn.edu

Trichy V. Krishnan
Department of Marketing, National University of Singapore, Singapore, biztvk@nus.edu.sg

V. Kumar
School of Business, University of Connecticut, Storrs, Connecticut 06269-1041, vk@sba.uconn.edu

In this paper, we provide theoretical arguments and empirical evidence for how Genetic Algorithms (GA) can be used for efficient estimation of macro-level diffusion models. Using simulations we find that GA and Sequential Search-Based-Nonlinear Least Squares (SSB-NLS) provide comparable parameter estimates when the data including peak sales are being used, for a range of error variances, and true parameter values commonly encountered in the literature. From empirical analyses we find that the forecasting performance of the GA estimates is better than that of SSB-NLS, Augmented Filter, Hierarchical Bayes, and Kalman Filter when only pre-peak sales data is available for estimation. When sales data until the peak time period are available for estimation, SSB-NLS is able to obtain parameter estimates when the starting values provided are the estimates from using GA. The estimates from GA are not biased and do not change in a systematic fashion when post-peak sales data are used, whereas the estimates from SSB-NLS are biased and change in a systematic fashion. Summarizing, we find that GA may be better suited for diffusion model estimation under the three conditions where SSB-NLS has been found to have problems.

Key words: Bass model; starting values; systematic change and bias; closed-form solution; nonlinear least squares; genetic algorithms; pre-peak sales forecasting
History: This paper was received August 30, 2002, and was with the authors 4 months for 5 revisions; processed by Roland Rust.

1. Introduction

Data available for estimation of the Bass model or its extensions are usually restricted to a set of 12 to 15 observations. There are two reasons for this. First, sales data are typically collected annually to avoid fluctuations in sales within a year and seasonality issues. Second, sales of most of the new products tend to stop growing, and in fact start decreasing, after 7 to 10 years. Because a manager's interest is very likely to diminish after the growth stage of a new product, researchers have to work with a smaller dataset in many cases. A natural outcome of this problem is researchers' interest in exploring more and more sophisticated estimation techniques that can extract as much information as possible from smaller datasets with maximum efficiency. Alternatively, researchers have also investigated using information such as advance purchase orders (Moe and Fader 2002) and spatial dimensions of product adoption (Garber et al. 2003) for early prediction of new product sales.

The chronology of the various estimation techniques and the benefits and drawbacks of each method are outlined in Table 1. Of the three time-invariant estimation techniques employed in the diffusion literature, namely, Ordinary Least Squares (OLS), Maximum Likelihood (ML), and Nonlinear Least Squares (NLS), it is generally accepted that NLS is the best option among the current alternatives (Putsis and Srinivasan 2000). For the Bass (1969) model, NLS is applied to the equation

$$s(t) = m * [F(t) - F(t-1)] + \varepsilon(t), \tag{1}$$

where $s(t)$ is the sales function, m is the market potential parameter, $\varepsilon(t)$ is the normal additive error, $F(t)$ is the cumulative density function of time of adoption given by

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + (q/p)e^{-(p+q)t}}, \tag{2}$$

t = time period, p = coefficient of innovation, and q = coefficient of imitation.

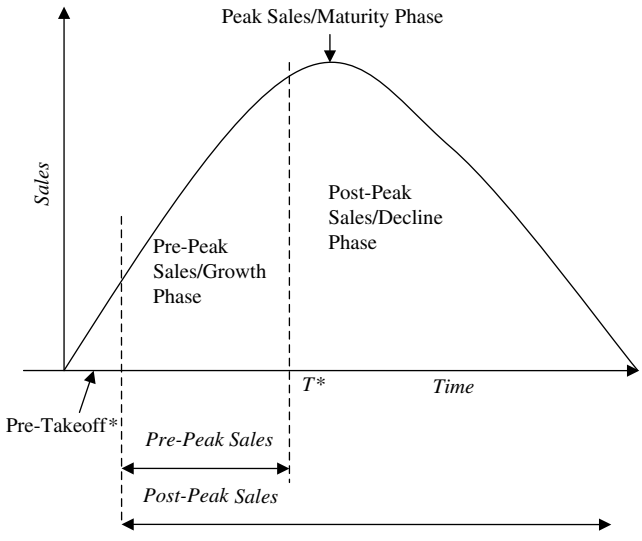
Table 1 Evolution of Econometric Estimation Techniques for the Bass Model

Author	Estimation method	Advantages	Drawbacks
Bass (1969)	Ordinary least squares (OLS)	Easy and straightforward implementation. Good fit.	Discrete form for a continuous process. No standard errors for the parameter estimates. Parameters may sometimes be outside the allowable range.
Schmittlien and Mahajan (1982)	Maximum likelihood (ML)	Continuous form operationalization. Minimizes sampling error.	Not efficient in reducing errors from sources other than sampling.
Srinivasan and Mason (1986), Jain and Rao (1990)	Nonlinear least squares (SSB-NLS)	Continuous form operationalization. Provides better fit than other methods, especially maximum likelihood.	Problems with convergence when data does not contain peak time period. Bias and systematic change in parameter estimates.
Lenk and Rao (1990)	Hierarchical Bayes (HB)	Provides good predictions of future sales including before peak time period when compared with ML. Utilizes variation in previous diffusion histories to avoid convergence in local minima.	Distribution assumptions about parameters. Lower accuracy when diffusion curves are skewed. Not easy for practitioners to use.
Xie et al. (1997)	Augmented Kalman filter (continuous observation—discrete measurement)	Provides better predictions of future sales including before peak time period when compared with adaptive filter, SSB-NLS, OLS, and ML.	Does not provide as good a forecast as the GA estimates. (Shown in the present study.) Not easy for practitioners to use.
Present study	Genetic algorithm (GA)	Addresses all three problems associated with SSB-NLS. Consistently performs better than other estimation methods.	Need specific software.

NLS used in popular computer packages employs a sequential search technique to obtain parameter estimates.[1] The widely used sequential search-based (SSB) NLS places three major restrictions on the estimation that span every stage of the product lifecycle. SSB-NLS estimation seems to have problems with data that covers three stages of a diffusion curve: pre-peak sales, peak sales, and post-peak sales (see Figure 1). With the pre-peak sales data, SSB-NLS has been repeatedly found to not achieve convergence (Srinivasan and Mason 1986, Lenk and Rao 1990). With the peak-sales data, it has been found that SSB-NLS's convergence largely depends on the initial values one provides for the parameters. With the post-peak sales data, it has been found that the SSB-NLS estimates of the Bass model are biased and change systematically as we add datapoints from later years (Van den Bulte and Lilien 1997, Bemmaor and Lee 2002, Venkatesan et al. 2000).

In §2, we provide theoretical arguments and intuition for how GA is, under certain circumstances, able to arrive at global optimal parameter estimates more efficiently even when the response surface is multimodal and noisy. We also show how a SSB-NLS has a probability of converging at a local optimal solution in these cases. In §3, using simulated data we show that the estimates from GA are similar to estimates from SSB-NLS under commonly encountered error variances and parameter estimate values, provided full datasets are used for estimation. Then, using empirical datasets we compare the performance of GA with SSB-NLS and other techniques proposed in the literature when the data does not contain peak sales, when there is data until peak sales, and when datapoints are added sequentially to post-peak sales data. Based on the results of our analyses in §4 and Appendix 2 we conclude that GA is able to produce better parameter estimates than SSB-NLS as evident in lower Mean Squared Errors (MSE) and Mean Absolute Deviation (MAD) under the three data related scenarios mentioned above.

Figure 1 Sample Diffusion Path of a New Product Innovation



*Typically not used in diffusion model estimation.

[1] Recently researchers have also found issues with statistical procedures implemented in Microsoft Excel (McCullough and Wilson 1999).

2. Intuitive Expectations for the Performance of GA

In this section, we provide intuitive reasons for why we expect GA to perform better than SSB-NLS and other search algorithms (e.g., a grid search) even though the estimation techniques use the identical objective function. Details on the estimation of the Bass model using GA are provided in Appendix A.

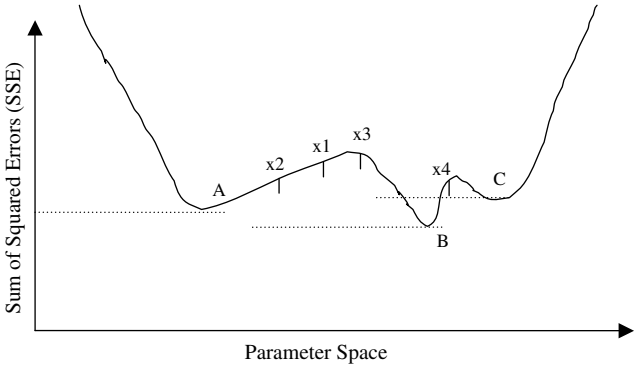
Comparison with SSB-NLS. The SSB-NLS and GA have several important and substantive differences due to which SSB-NLS and GA need not always provide the same solutions. A vast body of research outside of marketing has investigated the properties of parameter estimates from GA (Del Moral and Miclo 2001, Dorsey and Mayer 1995) and has also compared the performance of estimates from GA with traditional gradient search algorithms used in SSB-NLS (Salomon 1998, Dorsey and Mayer 1995). We summarize the major findings from these studies below:

- SSB-NLS techniques use single-point gradient search algorithms to locate the parameters that optimize the objective function (minimum sum of squared errors).
- GA uses parallel, evolutionary search algorithms to locate parameters that optimize the objective function (minimum sum of squared errors in this case).
- Theoretical expectations are that GA has a higher probability of convergence to global optimum solutions when datapoints are less, number of parameters is large, the parameter space is multimodal, and the model is inherently nonlinear (Del Moral and Miclo 2001). However, SSB-NLS is dependent on smooth, and mostly quadratic surfaces to ensure convergence to local optimal—with the expectation that if appropriate starting values are chosen the local optimum will represent the global optimal solution (Seber and Wild 1988).
- The estimates from GA provide better fit and forecasting performance as compared to SSB-NLS for highly nonlinear functions (Dorsey and Mayer 1995) and they represent inherently two different classes of optimization techniques that have different properties (Salomon 1989).

We further expand on these key features using an illustrative example shown in Figure 2.

Suppose that the solution space is not smooth and that it has a surface as shown in Figure 2. The optimal points A and C are local but are very dominant because a vast majority of starting values will move towards these two local optima. The global optimal point B is reachable only from a few starting values. In the SSB-NLS, unless you happen to be in those few spots, you would never reach the global optimum, however hard you try. This is because in current NLS packages there is SEQUENTIAL SEARCHING in that

Figure 2 NLS vs. GA



the next point in the search has to have a lower Sum of Squared Errors (SSE), i.e., there has to be a systematic continuity. For example, if we start at x1 (see Figure 2), SSB-NLS will definitely take us only to x2 and *never* to x3 or x4. But in GA, if we start at x1, it will not only consider x2 in the next step (as NLS does) but also consider x3 or/and x4! In other words, in GA, in each iteration, the program not only maintains the systematic search as the sequential search-based NLS does, but also searches for nonsequential points. So, it keeps trying random values as well simultaneously. This is the strength of GA. It combines the unique advantage of sequential search-based NLS and that of the random search to the maximum advantage. So, it has a much better chance to reach the global optima.

Also, one can argue that by trying NLS with different starting values, one should theoretically reach the global optimal. However, the issue is that we currently *do not* have any algorithm to help us carry out such "searching for initial values." Take the Bass model, where we typically have m in the order of thousands, p in the order of thousandths and hundredths, and q in the order of hundredths and tenths. This means that with just three parameters, there are literally millions of starting value combinations that are possible. As mentioned above, we do not have a systematic way to check all these values. This problem becomes multifold:

(1) if we have a model such as the Generalized Bass Model (Bass et al. 1994) or cross-national diffusion models as in Kumar and Krishnan (2002) that have more than six parameters, and/or

(2) if we have only few datapoints, in which case the solution space becomes very rough, and/or

(3) if we have those few points not in any "smooth" manner but in a very noisy manner as evidenced by the data used in Van den Bulte and Lilien (1997).

Comparison with Grid Search. A simple search algorithm, e.g., a grid search, should be able to perform better than SSB-NLS when the assumptions of SSB-NLS are violated. The grid search consists simply of constructing a three-dimensional rectangular block

of points to cover the region where the parameter estimates are known to exist, evaluating the objective function (least squares in this case) at each of the points, and taking the smallest value of the objective function as the optimal value. This can give precision of one "cell." Clearly, the grid must be fine enough so that the objective function is constant within each cell. Halving the cell edge length leads to 2^3 times as many function evaluations. The grid search method is extremely costly because it makes a large number of unnecessary function evaluations and there is not a scope for learning from past values. Also, the efficiency of grid search degrades as the number of parameter estimates required for estimation increases (e.g., the Nonuniform Influence models and the Generalized Bass Model). Hence, we do not expect the grid search to obtain parameter estimates that are closer to the truth when we restrict the time required for estimation to reasonable limits.

The GA has, however, two limitations. First, the search can entail many evaluations of the objective function and, consequently, a longer execution time when compared to SSB-NLS. However, with present day computational power this is not a major issue. The second limitation of the GA is that, one has to carefully select the convergence criteria to obtain an optimal solution (further details are in Appendix 1). Although the complexity of coding involved in implementing GA is a major impediment to its widespread applicability, many software packages (GA toolboxes for use with MatLab, S-Plus, C++, and Excel) are being released with wide-ranging functions and applications built into them, and these software packages can be run even on commonly available spreadsheets. A short technical description of how the GA works is given in Appendix 1.

3. GA as an Alternative Estimation Tool for the Bass Diffusion Model

In this section, we have two objectives. First, using simulated data we show that the estimates from GA are similar to estimates from SSB-NLS under commonly encountered error variances and parameter estimate values, provided full datasets are used for estimation. Then, using empirical datasets we compare the performance of GA with SSB-NLS and other techniques proposed in the literature when the data does not contain peak sales, when there is data until peak sales, and when datapoints are added sequentially to post-peak sales data. All the estimates reported in our analyses are median values based on 1,000 repeats of the GA estimation and the standard errors are computed from the standard deviation of the 1,000 estimates obtained from the repeats. The starting values (solution set in the initial iteration) for an estimation run is generated using uniform random values with the appropriate boundary conditions. The stopping rule for an estimation run is as follows: Terminate if the optimal solution string has changed by more than 0.1% in the last 10,000 iterations. The probability of crossover is set at 0.8, and the probability of mutation is set at 0.25. For each iteration, 200 sample solution vectors are generated (in other words, the population size for the estimation is equal to 200).

3.1. Properties of the Estimates from GA

Given that a GA is a simulation-based estimation technique, the finite sample properties of GA estimates[2] need to be established by conducting a Monte Carlo simulation study. The desirable properties of the estimates from NLS routines include (1) approximate normal distribution, and (2) variance of

$$\beta \approx 2\sigma^2 \left(\frac{d^2 f(x_t, \beta)}{d\beta^2} \right)^{-1} \tag{3}$$

(Griffiths et al. 1993). The estimates from GA are reliable if they exhibit these properties of estimates from SSB-NLS. The investigation of the asymptotic properties of the estimates from GA is important for deriving inferences based on the estimates from GA.

We conducted a simulation experiment to compare the properties of estimates from SSB-NLS, grid search, and GA. In the simulation experiment, we generated a total of 15 cells, where a dataset in each cell has a sample size of $T = 17$, by manipulating the q/p ratio over three levels (2, 5, 50) and the error variance (σ^2) of a mean zero normal random variable over five levels (0.06, 0.24, 0.40, 0.58, 0.78). We manipulated the values of q, p, and error variance to test the robustness of estimates from GA across the range of parameter values and error variances observed in the diffusion literature (Van den Bulte and Lilien 1997). We used the Srinivasan-Mason operationalization for generating the data and add a multiplicative proportional normal random error. Specifically, the data was generated as

$$s(t) = m * [F(t) - F(t-1)] * e, \tag{4}$$

where $e = (1 + \text{norm}(0, \sigma^2))$ and $\text{norm}(0, \sigma^2)$ is a random value from the normal distribution with mean $= 0$ and variance $= \sigma^2$. For each of the 15 datasets we estimated the values of p, q, and m using SSB-NLS, grid search, and GA. The values are reported in Table 2. We also plot a histogram of the 1,000 estimates for two datasets in our simulation

[2] The package "Evolver" is used for estimation using GA. Evolver is an Excel-based add-on package distributed by Palisade, Inc. for conducting analysis using GA.

Table 2 Results from the Simulation Study[a]

q/p	Error variance	True values			Estimates from SSB-NLS			Estimates from grid search			Estimates from GA		
		p^*	q^*	m^*	$p_{\text{SSB-NLS}}$	$q_{\text{SSB-NLS}}$	$m_{\text{SSB-NLS}}$	p_{GS}	q_{GS}	m_{GS}	p_{GA}	q_{GA}	m_{GA}
2	0.06	0.03	0.06	100	0.027	0.088	96	0.041	0.089	70	0.027	0.088	96
					(0.005)	(0.037)	(15.91)				(0.0002)	(0.002)	(0.95)
	0.24	0.03	0.06	100	n.a.	n.a.	n.a.	0.013	0.52	180	0.0065	0.010	181
											(0.0002)	(0.009)	(5.82)
	0.42	0.03	0.06	100	0.017	0.108	122	0.022	0.027	180	0.017	0.110	122
					(0.01)	(0.08)	(13.23)				(0.006)	(0.007)	(9.18)
	0.58	0.03	0.06	100	0.032	0.00001	171.18	0.026	0.099	180	0.032	0.001	168
					(0.006)	(0.0005)	(32.85)				(0.0004)	(0.0008)	(16.28)
	0.78	0.03	0.06	100	n.a.	n.a.	n.a.	0.016	0.017	180	0.016	0.040	191
											(0.0005)	(0.001)	(13.84)
5	0.06	0.015	0.075	100	0.007	0.059	204	0.01	0.057	150	0.012**	0.077	117
					(0.0001)	(0.01)	(56.73)				(0.0004)	(0.002)	(15.24)
	0.24	0.015	0.075	100	0.002	0.019	306	0.01	0.045	160	0.010	0.032	132
					(0.003)	(0.027)	(59.92)				(0.0008)	(0.02)	(5.21)
	0.42	0.015	0.075	100	0.026	0.248	39	0.017	0.299	60	0.027	0.240	39
					(0.001)	(0.06)	(10.25)				(0.0001)	(0.005)	(6.72)
	0.58	0.015	0.075	100	0.013	0.088	159	0.019	0.099	80	0.014	0.093	159
					(0.005)	(0.03)	(13.86)				(0.0004)	(0.08)	(2.78)
	0.78	0.015	0.075	100	0.007	0.0001	333	0.032	0.092	50	0.013	0.103	51**
					(0.0003)	(0.001)	(30.48)				(0.008)	(0.05)	(17.24)
50	0.06	0.015	0.85	100	0.021	0.752	103	0.01	0.88	120	0.021	0.752	103
					(0.003)	(0.05)	(4.69)				(0.0001)	(0.008)	(2.21)
	0.24	0.015	0.85	100	0.008	0.960	107	0.012	0.89	120	0.008	0.958	106
					(0.003)	(0.09)	(7.24)				(0.0005)	(0.004)	(3.85)
	0.42	0.015	0.85	100	0.020	0.683	110	0.01	0.69	70	0.020	0.683	110
					(0.03)	(0.15)	(13.43)				(0.008)	(0.08)	(5.87)
	0.58	0.015	0.85	100	0.004	0.999	115	0.016	0.89	24	0.004	0.999	114
					(0.001)	(0.08)	(21.09)				(0.0008)	(0.006)	(10.29)
	0.78	0.015	0.85	100	0.022	0.999	68	0.01	0.76	60	0.022	0.999	68
					(0.002)	(0.02)	(9.19)				(0.000002)	(0.0004)	(1.85)

[a]Values in parentheses represent standard errors. n.a. = not applicable, the algorithm did not converge.
**Significant at $\alpha < 0.05$.

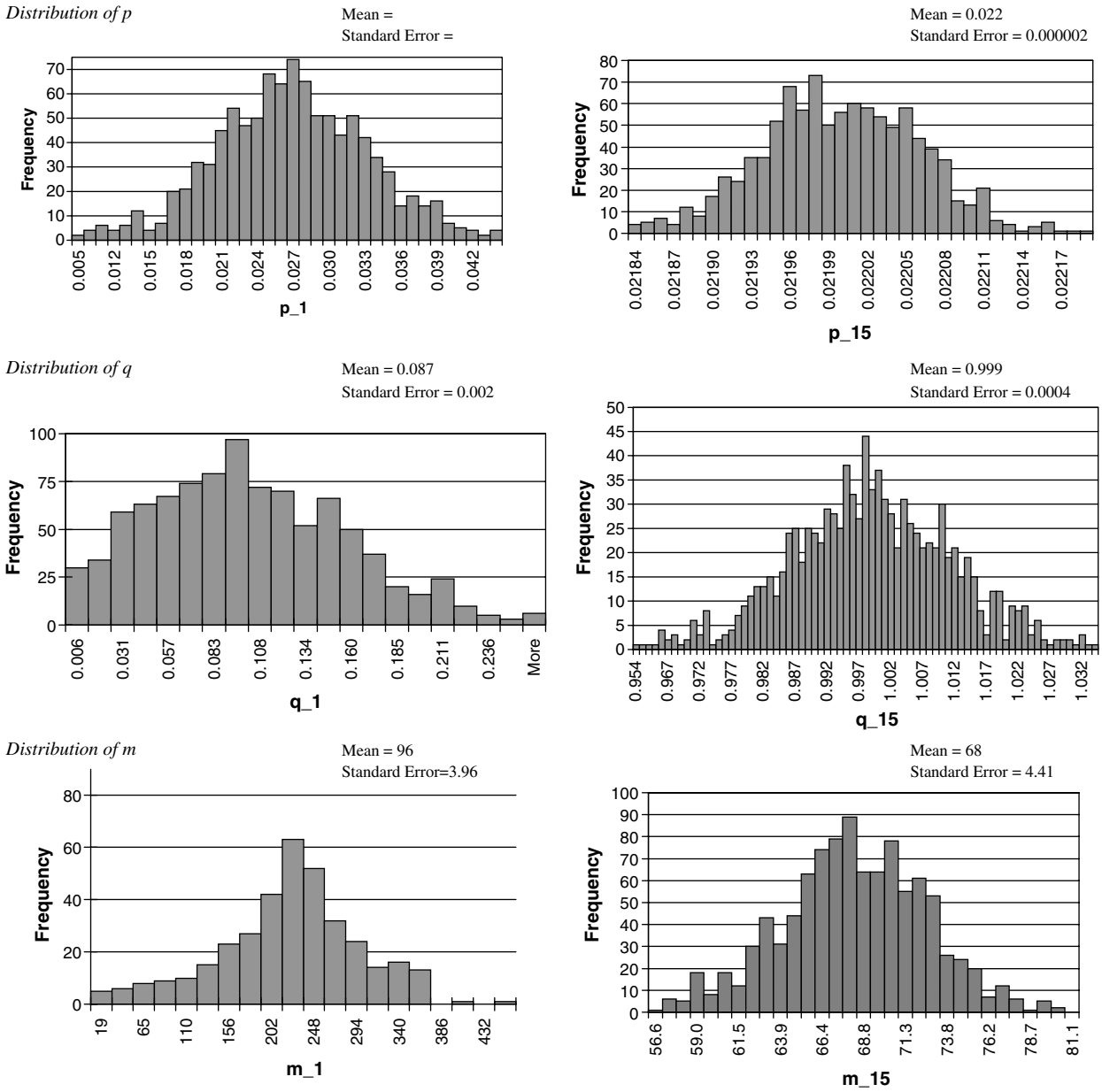
experiment: (1) $q/p = 2$ and $\sigma^2 = 0.06$, and (2) $q/p = 50$ and $\sigma^2 = 0.78$ as shown in Figure 3. These two datasets represent the two extreme ends in our simulation experiment and we use these to illustrate the distribution of estimates from GA.

Results. The estimates obtained from GA closely resemble the estimates from SSB-NLS and the grid search. The GA estimates also have acceptable standard errors that are proportional to the asymptotic standard errors obtained from the SSB-NLS method. The average deviations (from the true values) of the estimates from GA are 36%, 49%, and 33% for p, q, and m, respectively. Similarly, the average deviations from SSB-NLS are 45%, 58%, and 63% for p, q, and m, respectively, and 33%, 49%, and 45% for p, q, and m, respectively, for grid search. We also observe that the values of the standard errors from GA are consistently lower than the standard errors from SSB-NLS. This implies that GA is able to produce estimates that are more stable (consistent or robust) than SSB-NLS. However, we also need to acknowledge that the lower standard errors from GA as compared to SSB-NLS are an indication of the superiority of the estimation technique only when there is symmetric error in the data.[3] The estimates from SSB-NLS do not converge for two out of the 15 datasets. GA and the grid search produces estimates that are reliable for all the datasets even when SSB-NLS fails to converge.

To evaluate if there is any statistical difference between the estimates, the confidence intervals of the estimates from GA and the estimates from SSB-NLS are calculated. If the confidence intervals do not intersect, then the estimates from GA and SSB-NLS estimates are statistically different. The estimates from GA are significantly different from SSB-NLS in only two cases (for dataset 6, $q/p = 5$ and $\sigma^2 = 0.06$, p_{GA} is significantly different from $p_{\text{SSB-NLS}}$ at $\alpha < 0.05$, and for dataset 10, $q/p = 5$ and $\sigma^2 = 0.78$, m_{GA} is significantly different from $m_{\text{SSB-NLS}}$ at $\alpha < 0.05$). The histogram of estimates from dataset 1 ($p = 0.03$, $q = 0.06$, $m = 100$, and $\sigma^2 = 0.06$) and the estimates from dataset 15 ($p = 0.15$, $q = 0.75$, $m = 100$, and

[3] The authors thank a reviewer for bringing this to their attention.

Figure 3 Distributions of Parameter Estimates from GA for the Simulated Datasets 1 and 15*



*Dataset 1 has $q/p = 2$ and error variance = 0.06 and Dataset 15 has $q/p = 50$ and error variance = 0.78. All the standard errors are computed over 1,000 repeats of the GA estimation.

$\sigma^2 = 0.78$) are provided in Figure 3. This figure shows that the estimates obtained from GA are normally distributed even when the error variance is very high (for p_15, q_15, and m_15) and very low (for p_1, q_1, and m_1). Thus, it can be inferred that the parameter estimates obtained from GA are consistent and have the desirable properties of standard statistical techniques[4] and are comparable to SSB-NLS under most commonly encountered.

[4] The bias in the parameter estimates can also be checked from the contour plots of the residuals obtained from the parameter estimates.

Having shown that GA estimation is an appropriate alternative estimation technique for the diffusion models, we now compare the estimates obtained from GA with those from SSB-NLS with respect to the three issues discussed previously.

3.2. Estimation with Pre-Peak Sales Data

Although GA may perform better than SSB-NLS, we need another benchmark to evaluate the performance of GA. Hierarchical Bayesian methods (Lenk and Rao 1990) and Augmented Kalman Filter estimation techniques (Xie et al. 1997) have been proposed in the extant literature to overcome the

nonconvergence issue of SSB-NLS (with pre-peak sales period data). Xie et al. (1997) show the superiority of the Augmented Kalman Filter (AKF(C-D)) technique over the Hierarchical Bayes method. Hence, we use the Xie et al. (1997) study as our benchmark for evaluating the performance of GA. We also compare the performance of GA with a simple grid search. For comparability purposes, we use the same diffusion datasets—room air conditioners, color television, clothes dryer, ultrasound, mammography, foreign language, and accelerated program—that were used by Xie et al. (1997). We assess the effectiveness of one step ahead forecasts from each method using the MAD measure. To ensure comparability in the MAD measures we use the priors[5] published in Xie et al. (1997, Appendix A). Managers in general are interested in forecasting capability when using diffusion models during the growth phase. Hence, we compare the various techniques on only forecasting performance.

Sales at Each Time Period. We use the prior values as estimates of the diffusion curve before the product is launched (i.e., for time $t = 0$). The prior estimates predict sales/adoption at $t = 1$. We then repeat the above process by advancing one time period in each step until the peak is reached. By taking the mean of all the absolute deviations thus obtained we get the MAD for that dataset. The MAD measures for the AKF(C-D) and AF are obtained from Xie et al. (1997). The results of our analyses are shown in Table 3a. The reduction in prediction error as compared with AKF(C-D) ranges from 28% (clothes dryer) to 55% (mammography). In addition to reporting the MAD, we also test if the MAD from the GA estimates is significantly different from the other techniques for each product category using a pairwise test of the absolute deviations in each time period. We use a nonparametric bootstrap methodology for the pairwise tests because it accommodates for the time series nature of the data and does not make any assumptions regarding the distribution of the error terms (Davison and Hinkley 1997). The details of the statistical tests are provided in Appendix 2. We see from Table 3a that the MAD from GA is significantly lower from AF, and grid search across all product categories and the MAD from GA is significantly lower from AKF(C-D) for all product categories expect foreign language.

We also assess the prediction errors from GA for the post-peak sales for a diffusion curve. In this case, we compare the prediction errors from GA with AKF(C-D), AF, SSB-NLS, OLS, and grid search. For the post-peak one step forecasts we compute the absolute deviations from time periods $t^* + 1$ until the entire data available for estimation is used (note that t^* is the peak sales time period). The reduction in prediction error as compared with AKF(C-D) ranges from 56% (clothes dryer) to 71% (mammography). We also performed statistical tests similar to the pre-peak data on the post-peak sales data. We see from Table 3a that the MAD from GA is significantly lower from AF, grid search, AKF(C-D), and OLS across all product categories and is significantly lower than the MAD for SSB-NLS for all product categories except foreign language for post-peak data as well. Hence, we can infer that the GA provides a viable alternative to current estimation techniques with respect to forecasting of both pre-peak sales and post-peak sales.[6]

Time to Peak Sales and Sales at Peak Time Period. Using the estimates obtained for predicting each period sales we also compare the forecasting capability of GA with grid search and AKF(C-D) for predicting *time* until peak sales and the *sales* at peak time period. The results of our analyses are provided in Table 3b. With respect to time until peak sales, the reduction in prediction error as compared with AKF(C-D) ranges from 20% (ultrasound) to 60% (mammography). With respect to sales at peak time period, the reduction in prediction error as compared with AKF(C-D) ranges from 8% (clothes dryer) to 28% (foreign language). The MAD for time to peak sales and sales at peak time period are significantly lower from grid search for all the product categories. Again, the MAD of the GA estimates are significantly lower than the MAD from AKF(C-D) for all product categories except clothes dryer and ultrasound with respect to time to peak sales and is significantly lower than the MAD from AKF(C-D) for all product categories except mammography with respect to sales at peak time period. Next, we will explore the utility of GA in overcoming the initial (starting) values issues faced by SSB-NLS in estimating with data up to peak sales period.

3.3. Peak Sales Period Data and Initial Values

As stated earlier, GA starts by sampling n number of possible solutions from the search space as the population in the first iteration. Hence, by definition GA estimation is not affected by the starting point issues that SSB-NLS faces. To explain this phenomenon in more detail, we provide examples of situations where the SSB-NLS fails to converge (even if the peak sales exists in the data). We used three representative datasets: corn 1943 (Van den Bulte and Lilien 1997), air conditioner, and clothes dryer (Bass et al. 1994).

[5] In the case of GA, the prior is used to provide a prediction of sales/adoption at time period $t = 1$ before the product is launched.

[6] We also compared the performance of GA and SSB-NLS with multiple step ahead forecasts and other datasets such as cellular phone diffusion in Western Europe. GA performs better than SSB-NLS even here. The results can be obtained from the authors.

Table 3a Prediction Errors of One Step Ahead Forecasts from GA

Criterion	Method	Room air conditioners	Color television	Clothes dryer	Ultrasound	Mammography	Foreign language	Accelerated program
Pre-peak sales								
MAD*	AF	524.3[a]	2, 714.4[a]	368.2[a]	5.9[a]	5.1[a]	1.5[b]	3.6[a]
	Grid search	503.7[a]	1, 786.4[a]	305.9[a]	5.5[a]	4.8[a]	2.4[b]	3.4[a]
	AKF(C-D)	261.0[a]	971.0[b]	211.0[c]	4.8[a]	2.7[a]	1.0	2.8[a]
	GA	148.5	620.0	150.2	3.1	1.2	0.5	1.4
Post-peak sales								
MAD*	OLS	791.3[a]	2, 523.0[a]	401.7[a]	n.a.	n.a.	n.a.	n.a.
	SSB-NLS	334.3[a]	1, 083.8[a]	296.0[a]	4.8[a]	4.2[a]	1.0	2.5[c]
	AF	558.3[a]	3, 835.4[a]	103.6[a]	17.0[a]	2.9[a]	2.4[b]	5.0[a]
	Grid search	61.8[a]	267.1[a]	94.6[a]	4.2[a]	3.5[a]	2.2[b]	3.4[a]
	AKF(C-D)	55.0[b]	241.0[a]	77.0[b]	4.6[a]	3.1[a]	2.1[b]	2.2[a]
	GA	34.2	150.8	56.4	2.8	0.9	0.8	1.2

*$\text{MAD} = (1/k)\sum_{k=1}^{K} |x(k) - \hat{x}(k)|$, where k = number of forecasts for each product category, $x(k)$ is the observed number of incremental adopters in the kth time interval (t_{k-1}, t_k), and $\hat{x}(k)$ is the predicted value of $x(k)$. **The values for the AF and AKF(C-D) are sourced from Xie et al. (1997). n.a. = not applicable; reliable estimates were not obtained from OLS. The signs of the coefficients were negative.

[a]Significant at $\alpha < 0.01$; [b]Significant at $\alpha < 0.05$; [c]Significant at $\alpha < 0.10$.

Table 3b Prediction Errors for Time to Peak Sales and Sales at Peak Time Period

Criterion	Method	Room air conditioners	Color television	Clothes dryer	Ultrasound	Mammography	Foreign language	Accelerated program
Time to peak sales								
MAD*	Grid search	2.1[a]	3.8[a]	1.8	1.2	2.1[a]	14.3[a]	4.1[a]
	AKF(C-D)	1.9[b]	2.6[b]	1.2	1.0	1.7[b]	11.1[a]	2.9[a]
	GA	1.4	1.8	0.9	0.8	0.68	6.7	1.3
Sales at peak time period								
MAD**	Grid search	43.2[a]	130.8[a]	69.4[a]	9.7[a]	11.2[a]	5.9	11.9[a]
	AKF(C-D)	38.6[c]	104.1[b]	55.7	6.4[a]	9.1	4.5	7.8[b]
	GA	30.5	85.7	50.9	4.8	8.5	3.2	5.7

*$\text{MAD} = (1/k)\sum_{k=1}^{K} |\text{actpt} - \text{predpt}_k|$, where k = number of forecasts for each product category, actpt = actual peak sales time period, and predpt = predicted peak sales time period based on estimates using data up to time period t. **$\text{MAD} = (1/k)\sum_{k=1}^{K} |\text{actps} - \text{predps}_k|$, where k = number of forecasts for each product category, actps = actual peak time period sales, and predps = predicted peak time period sales based on estimates using data up to time period t.

[a]Significant at $\alpha < 0.01$; [b]Significant at $\alpha < 0.05$; [c]Significant at $\alpha < 0.10$.

We censored these datasets so that they only contain datapoints until the peak sales time period. The peak sales for the various datasets were $t = 11$, $t = 8$, and $t = 8$ for corn 1943, air conditioner, and clothes dryer, respectively. In SSB-NLS, when we first used the starting points recommended by Sultan et al. (1990), i.e., $p = 0.03$, $q = 0.38$, and $m = 1{,}000$, the SSB-NLS algorithm failed to converge as shown in Table 4.

Subsequently, we tried different combinations of starting values in a systematic manner. Specifically, the values of p were changed by 0.01 in the interval of $p = 0.001$ to $p = 0.09$. Similarly, the values of q were changed from 0.1 to 0.9 in the interval of 0.1 and the values of m were changed from 500 to 1,100 in intervals of 50 for air conditioner and clothes dryer. In a total of 1,053 combinations were used for (9 for p, 9 for q, and 13 for m) for air conditioner and clothes dryer. However, for corn 1943, we varied p from 0.0001 to 0.01 in increments of 0.0001, q from 0.1 to 0.9 in increments of 0.1, and m from 100 to 500 in increments of 100. Hence, we used a total of 4,500 combinations (100 for p, 9 for q, and 5 for m) for corn 1943.[7] SSB-NLS failed to converge in 25% of the cases for corn 1943, in 38% for air conditioner, and in 41% for clothes dryer.[8]

[7] We use different ranges for corn 1943 because it was based on adoption data and the cumulative adoption was in the range of 100 before peak sales. Also, the data for corn 1943 was highly left skewed indicating a low p estimate. SSB-NLS converged only 10% of the time when the 1,053 combinations used for air conditioner and clothes dryer were used for corn 1943, due to large starting values for m.

[8] We need to qualify that SSB-NLS would fail to converge less frequently if the starting values were based on experience (or management judgment). We use the above procedure for illustrative purposes only.

Table 4 Influence of Starting Values on SSB-NLS Convergence

	Starting value			SSB-NLS estimates		
Product	p^*	q^*	m^*	p	q	M
Corn 1943	0.03	0.38	500	n.a.	n.a.	n.a.
Air conditioner	0.03	0.38	500	n.a.	n.a.	n.a.
Clothes dryer	0.03	0.38	500	n.a.	n.a.	n.a.
After using starting values from GA*						
Corn 1943	0.00004 (0.00001)	0.9285 (0.039)	243 (10.72)	0.00004 (0.00001)	0.9396 (0.11)	241 (24.10)
Air conditioner	0.0071 (0.0006)	0.5230 (0.08)	13,442 (80.27)	0.0069 (0.002)	0.5212 (0.09)	13,509 (3,676.4)
Clothes dryer	0.00067 (0.00005)	0.2646 (0.03)	58,294 (130.23)	0.0034 (0.008)	0.2675 (0.09)	57,134 (15,015)

*We did not find any significant difference between the SSB-NLS estimates and the starting values provided from GA.
n.a. = not applicable; the estimates do not converge. Values in parentheses are the standard errors of the estimates.

We then estimated these datasets using GA and used the estimates from GA as starting values for SSB-NLS. We see that in all these cases, the SSB-NLS algorithm attained convergence and the estimates from SSB-NLS are similar to the starting values provided by GA. We compared the estimates from NLS and the starting values provided by GA using a t-test. The starting value provided to SSB-NLS is the mean value of 100 estimations based on GA (the standard error of the starting values can hence be calculated based on these 100 starting values). The SSB-NLS procedure provides the mean and standard error of the estimates. Hence, the confidence intervals of the starting values from GA and the estimates from SSB-NLS can be calculated. If the confidence intervals do not intersect, then the starting values and the SSB-NLS estimates are statistically different.

From these tests we found that the estimates from SSB-NLS and the starting values provided by GA are not significantly different from each other. Thus, we find that GA estimation allows us to overcome the starting value problem faced by the SSB-NLS in data until peak sales. Our analyses also highlight the severity of the starting value issues when using SSB-NLS to estimate the Bass model. Now, we will show how the GA estimation fares better in the post-peak sales data-based estimation as well.

3.4. Bias and Systematic Change in Parameter Estimates

To evaluate the performance of GA with respect to the bias and systematic change in parameter estimates, we performed two tests. We first employed the same datasets used in Van den Bulte and Lilien (1997) and replicated their analysis to ensure that we were able to obtain the same results. Second, we used the GA technique in place of SSB-NLS and repeated the exercise. While repeating what Van den Bulte and Lilien (1997) did, we made a minor, although materially insignificant, change to their original analysis. They used the following equation to check whether the dataset length affects the estimates systematically:

$$\log[Y_{irt}] = \delta_{ir} + \beta_{1r}\log[X(t_+)_{it}] + \beta_{2r}\log[(t_+)_{it}] + e_{irt}, \tag{5}$$

where $r = p$, q, or m.

In Equation (5), the subscript i indicates the innovation (i.e., the product), Y_{irt} is the estimate of parameter r (i.e., p, q, or m) corresponding to t_+ on the right-hand side of the equation, and δ_{ir} is the innovation-specific constant. After estimating the Bass model on the datasets of 11 products[9] (each with multiple truncations in the end), we used the following regression (in place of Equation (5)) to explore the impact of truncation of a dataset on the estimates of the three parameters, namely, p, q, and m:

$$\frac{Y_{irt} - Y_{ir}}{Y_{ir}} = \beta_{1r}X(t_+)_{it} + \beta_{2r}(t_+)_{it} + e_{irt}. \tag{6}$$

Our dependent variable $[Y_{irt} - Y_{ir}]/Y_{ir}$ directly measures for each dataset the extent of bias at the innovation level and, hence, we do not need the innovation specific constant used by Van den Bulte and Lilien (1997) on the right-hand side of their Equation (5). We use Equation (6) for our analyses because Equation (5) suffers from being confounded by level effects when estimates from various datasets are pooled. It is quite possible that the market potential, m, is high when the cumulative sales for a product is high, thus resulting in significant coefficients for $x(t_+)$ and t_+ irrespective of the amount of systematic change in the estimates. Similar arguments can be forwarded

[9] We have a sample size of 44 because we removed tetracycline from our analysis. Tetracycline had monthly data for 1.5 years without any indication of when the product was introduced relative to the time period available. We do not expect a new product to have a significant change in diffusion pattern over 1.5 years. Based on the findings of Putsis (1996), that the estimates of the Bass model are sensitive to the level of aggregation (monthly, quarterly, semiannual, and annual) in the diffusion data, we also aggregated the semiannual data for the CT-scanner to annual values, given that the rest of the data were at the annual values.

Table 5 Bias and Systematic Change

	Independent variables		
Dependent variables	Cumulative sales: $x(t_+)$	Time period: t_+	R^{2*}
Estimation technique = SSB-NLS ($N = 44$)			
$(m_{irt} - m_{ir})/m_{ir}$ $(r = 1)$	0.0001765[b]		0.23
		0.00435[a]	0.15
	0.000144[a]	0.00154	0.25
$(q_{irt} - q_{ir})/q_{ir}$ $(r = 2)$	−0.00035[c]		0.30
		−0.00971[c]	0.25
	−0.00025[a]	−0.0047	0.34
$(p_{irt} - p_{ir})/p_{ir}$ $(r = 3)$	0.000323[c]		0.21
		0.00442	0.05
	0.000367[c]	−0.00248	0.22
$(N = 44)$*			
Estimation technique = GA ($N = 44$)			
$(m_{irt} - m_{ir})/m_{ir}$ $(r = 1)$	0.00003616		0.014
		0.000951	0.020
	0.0000162	0.000745	0.022
$(q_{irt} - q_{ir})/q_{ir}$ $(r = 2)$	−0.0000968		0.026
		−0.004	0.078
	0.000015	−0.00414	0.090
$(p_{irt} - p_{ir})/p_{ir}$ $(r = 3)$	−0.0000277		0.001
		−0.000245	0.022
	0.0000584	−0.0032	0.026

*We have a sample size of 44 because we removed tetracycline from our analysis because it was monthly data for 1.5 years without any indication of when the product was introduced relative to the time period available. We do not expect a new product to have a significant change in diffusion pattern over 1.5 years. Based on the findings of Putsis (1996), that the estimates of the Bass model are sensitive to the level of aggregation (monthly, quarterly, semiannual, and annual) in the diffusion data, we also aggregated the semiannual data for the CT-scanner to annual values, given that the rest of the data were at the annual values.

[a]Significant at $\alpha < 0.01$; [b]Significant at $\alpha < 0.05$; [c]Significant at $\alpha < 0.10$.

for the coefficient of imitation and the time period t_+. The coefficient of imitation q is normally high for highly left-skewed data that require more datapoints for estimation with SSB-NLS (the peak sales occurs much later in the left-skewed data as compared to others). A more conservative test would be to normalize the estimates with respect to systematic change and then evaluate the influence of $x(t_+)$ and t_+. We hence believe that Equation (6) captures the concept we are trying to address better. The regression results are shown in Table 5. We obtain the same effects that Van den Bulte and Lilien (1997) report: (1) either β_{1m} or β_{2m} was significant and positive, (2) either β_{1q} or β_{2q} was significant and negative, or (3) either β_{1p} or β_{2p} was significant and positive.[10] This led them to conclude that the longer the dataset or the larger the cumulative sales at the last datapoint, the higher the m estimate, the lower the q estimate, and the higher the p estimate. In other words, if one adds datapoints from later years in the estimation, the estimates of m and p increase while that of q decreases systematically. The R^2 values for our operationalization range from 5% to 34% while the R^2 values from Van den Bulte and Lilien (1997) range from 10% to 42%.

In the next step, we performed a similar analysis on the 11 datasets with estimates obtained from GA estimates. The results are reported in Table 5. Overall, as can be seen from the table we do not observe the effects found from SSB-NLS. Note from Table 5 that although there are some changes in estimates as we change the length of the dataset, those variations are *not* related in any systematic fashion to the number of datapoints used for estimation or cumulative sales. In other words, *the coefficients in Table* 5 *indicate that neither of the independent variables* (β_{1r} = *cumulative sales and* β_{2r} = *time period*) *is significantly related to the variation in the parameters occurring as datapoints from later years are added to the dataset.* This clearly demonstrates that the findings of Van den Bulte and Lilien (1997) and later of Bemmaor and Lee (2002) regarding the systematic change do not show up when GA estimation is used in place of SSB-NLS estimation.

We also compared the SSE from SSB-NLS and GA for each product category and censoring level. We found that GA consistently outperformed or was at least equal to SSB-NLS in terms of in-sample fit. GA provided an improvement of at least 1% in the SSE when compared with the SSE from SSB-NLS.[11] When we subjected the GA estimates of the GBM model to the same experiments we conducted on the Bass model earlier, we still found no evidence of the systematic change and bias.

4. Conclusions and Future Research

In this paper, we have discussed in detail key problems that have been raised regarding SSB-NLS. These are: nonconvergence with pre-peak sales period dataset, need for appropriate starting values with data until peak sales, and systematic change and bias that are a function of the dataset length with post-peak sales data. We showed that GA is able to efficiently produce better parameter estimates as evident in the MSE and MAD when compared with SSB-NLS under the scenarios mentioned above. We find that

[10] Our choice for Equation (6) is questionable if we do not observe the same phenomenon as Van den Bulte and Lilien (1997) on the estimates from SSB-NLS. However, we do observe the same phenomenon as Van den Bulte and Lilien (1997).

[11] We need to compare the SSE of the estimates from GA with those from SSB-NLS in this case because the bias and systematic change in estimates from GA may not occur even if the estimates are suboptimal as compared to SSB-NLS (in other words, if the GA estimates have lower SSE as compared to the SSB-NLS estimates). However, we find that the estimates from GA have lower SSE as compared to the estimates from SSB-NLS, hence, ruling out the alternate explanation.

the estimates from GA perform better than current alternatives such as AKF(C-D) and HB in providing forecasts of new product sales before peak. Also, by definition, GA does not suffer from starting value issues. We also find that estimates from GA do not suffer from bias and systematic changes observed by Van den Bulte and Lilien (1997). Overall, the above benefits make GA an excellent candidate for time-invariant estimation of parameters of the highly nonlinear Bass model. In fact, reliable estimation of the diffusion curve before peak sales would allow managers to efficiently allocate marketing resources.

Researchers can further investigate the utility of GA as a tool for estimating marketing models that do not have the luxury of large datasets. Previous studies in marketing have used GA as an algorithm to either provide optimal advertising schedules (Naik et al. 1997), optimal product design (Balakrishnan and Jacob 1996), optimal competitive strategies (Midgley et al. 1997), or optimal customer contact and resource allocation rules (Venkatesan and Kumar 2004). The applicability of GA in estimation of continuous time mixture models is also an interesting venue for research.

GA can prove to be an excellent estimation technique for complex nonlinear models with less datapoints. Also, given the flexibility of GA, researchers can develop complex diffusion models (which may or may not have a closed-form solution) that address issues such as the influence of marketing-mix variables on multiple generations of technology and the diffusion of innovations across multiple countries (Parker 1994, Talukdar et al. 2002), take-off of new products (Tellis et al. 2003), indirect network externality effects (Basu et al. 2003), competitive marketing-mix reactions to new product introductions (Shankar et al. 1998), dynamics of brand level diffusion (Krishnan et al. 2000), and nonuniform diffusion of products (Easingwood et al. 1983). We fitted the NUI model using GA and a two-step iterative SSB-NLS technique and compared the parameter estimates from the above two methods. However, we did not find any significant difference in the performance of the estimates (measured in terms of MAD) between the two methods. Given that GA is able to produce better parameter estimates of the Bass model when there is less data and when the data are noisy, it would be worthwhile to investigate if the performance of GA improves when more complex models are estimated with less number of datapoints.

Also, the availability of GA even in commonly available spreadsheets makes GA very useful for managers who are interested in performing sensitivity analyses on the various parameters under their control, especially during the pre-peak sales time phase. Future researchers can use extensions of the Bass model that incorporate competitive effects and GA to assess the equilibrium state of various marketing strategies for new products in competitive markets. Future researchers can also investigate the extent to which the performance of GA depends on the fitness function that is used in the analyses. For example, in the case of the Bass diffusion models, do the parameter estimates obtained from GA differ if the fitness function is minimizing the first difference of each period sales function with respect to the parameter than if the SSE is minimized directly?

Acknowledgments

The authors thank the editor, area editor, and the four reviewers for their comments on this manuscript. They also thank the participants of the Texas Faculty Consortium-2000, Dallas, Texas, and the INFORMS Marketing Science Conference 2000, UCLA, Los Angeles, California, Frank M. Bass, Rajesh Chandy, Ed Blair, Scott Baggett, and Kumar Mehta for their comments on an earlier draft of this paper.

Appendix A. Estimation of Diffusion Models Using GA

GA (Goldberg 1989, Dorsey and Mayer 1995) is a parallel search algorithm that is based upon Darwin's theory of evolution. For our case, we use the Srinivasan and Mason (1986) operationalization of the Bass model (Equation (1)) for estimation using GA. The estimation proceeds as follows:

Solution String:[12] (p, q, m).
Fitness Function:

$$\text{SSE} = \frac{1}{T}\sum_{t=1}^{T}[x(t) - \hat{x}(t)]^2,$$

where $x(t) =$ actual sales and $\hat{x}(t) = m[F(t) - F(t-1)]$.

Step 1. Generate 1,000 random[13] candidate solution strings to create generation $i = 1$.

Step 2. Select candidate solution vectors for iteration $i+1$ from solution strings in iteration i. Solution vectors with a lower SSE have a higher probability of being selected for iteration $i+1$.

Step 3. Perform crossover and mutation[14] on the candidate solution vectors to generate a solution set for iteration $i+1$.

Step 4. Compute fitness value for solution vectors and the optimal solution string[15] in iteration $i+1$.

Step 5. Proceed to Step 2 if the optimal solution string has changed by more than 0.1% in the last 10,000 iterations; else terminate.

[12] The random numbers need to be generated accounting for restrictions on the range of each individual coefficient.

[13] We use uniform random numbers for this case for lack of any distributional assumptions on the parameters.

[14] Details on the crossover and mutation operations are provided in Appendix 1.

[15] This refers to the solution string that produces the minimal sum of squared errors (or fitness) in that particular generation.

Appendix 1. Iteration Mechanism for a GA

A simple GA is composed of three operators: reproduction, crossover, and mutation. Using these three operations, the GA iterates through the following three steps.

Step 1. *Reproduction* is a process in which individual strings of a generation (parent generation) are copied to the next generation (child) according to their objective function values, f. We can think of the objective function f in our case to be the SSE between the predicted values of sales and the actual value of sales. Selecting strings according to their fitness values means that strings with a higher fitness value have a higher probability of contributing one or more offspring in the next generation. The probabilities could depend on the proportion of solutions present in a parent generation, based on linear ranking system of the solutions or based on a tournament selection. This operator is an artificial version of natural selection, a Darwinian survival of the fittest among string creatures.

Step 2. After reproduction, simple *crossover* may proceed in two steps. First, members of the newly reproduced strings (or new generation) are paired at random. Second, each pair of strings undergoes crossing over as follows: an integer position K along the string is selected uniformly at random between one and the string length less one $[1, l-1]$. Two new strings are created by swapping all characters between positions $K+1$ and l. The mechanics of reproduction and crossover are simple, involving random number generation, string copies, and some partial string exchanges. Nonetheless, the combined emphasis of reproduction and the structured, though randomized, information exchange of crossover give GA much of their power.

Step 3. *Mutation* is the process of randomly changing a cell in the string or the solution vector. Mutation is the process by which the algorithm attempts to ensure a globally optimal solution. If the algorithm is trapped in a local minimum, the mutation operator randomly shifts the solution to another point in the search space, thus removing itself out of the trap.

The above steps are repeated until the algorithm is halted. The decision to halt the program can depend either on a pre-fixed number of generations, the time elapsed in the evolutionary process, or the change in the optimal solution in the previous n generations. The composition of the final generation of strings—the best strings—is the GA's solution to the problem. It should be noted that when new strings are created, the old ones (those belonging to the previous generation) are discarded. Because the reproduction process tends to choose the "fittest" members of a generation, the generations tend to evolve. Thus, an initial population of relatively undistinguished solutions evolves to yield a set of solutions that cover the optimal region in the final generation. To obtain the optimal solution, one can either use a gradient search algorithm such as hill climbing from the mean value in the final generation or use the solution with the best fitness in the last generation as the optimal solution. The decision on which method to choose to find the optimal solution vector depends on the criterion used to terminate the iterations. For example, if the iterations are stopped because the best solution string does not change by more than 0.1% in the last n iterations, then using the best solution in the last iteration is a good option. On the other hand, if we stop the iterations after a pre-fixed number n, then searching for the optimal solution using a hill-climbing algorithm would be a better option.

The operations crossover and mutation are not performed for every reproduction. The probability of a string being selected for crossover is proportional to the string's fitness. Each operation is assigned a particular probability of occurrence or application. For example, if the probability of crossover is 0.6, 60 out of every 100 strings undergo crossover. Likewise, if the probability of mutation is 0.03, 3 out of every 100 strings undergo mutation. The probability of mutation is traditionally lower than the probability of crossover, because the primary function of a mutation operator is to remove the solution from a local minimum. The probabilities are assigned based on the characteristics of the problem. For example, if the problem is characterized by a turbulent environment (i.e., if the solution space is not uniform all over), the probability of crossover and mutation are chosen to be high.

Appendix 2. Statistical Tests for Comparing Forecasting Accuracy Across Product Categories

Pre-Peak Sales Data

In Table 3a, we report the MAD of the one step ahead forecasts starting from the first datapoint until the peak sales datapoint for the pre-peak sales data. Specifically, we use prior values as estimates of the diffusion curve before the product is launched (i.e., for time $t=0$). The prior estimates predict sales/adoption at $t=1$. We then repeat the above process by advancing one time period each step until the peak is reached. By taking the mean of all the absolute deviations thus obtained we get the MAD for that dataset. We report the MAD for all estimation techniques across product categories in Tables 3a and 3b.

While performing statistical tests comparing the one step ahead forecast absolute deviations across different estimation techniques, we need to consider the following features:

(1) The scale of the absolute deviations may be related to the number of time periods used for estimation. For example, the absolute deviation obtained from forecasting the third time period based on two time periods for estimation may be higher than the absolute deviation obtained from forecasting the fourth time period based on three time periods for estimation.

(2) The estimated values of sales are correlated across time periods.

(3) The error distribution need not necessarily be normal.

To accommodate for the above features we perform a pairwise comparison of absolute deviations using a nonparametric test—bootstrap resampling for time series data (Davison and Hinkley 1997). The test is performed as follows.

Let us consider the case of comparing the absolute deviations of one step ahead forecasts from AF with those from GA for data with room air conditioners for pre-peak data.

Estimating Errors

Step 1. Estimate the sales data $[\hat{s}_{\mathrm{GA}}(t)]$ using GA with the room air conditioner pre-peak sales data $[s(t)]$.

Step 2. Compute the errors $[\varepsilon_{\mathrm{GA}}(t)]$ for each time period (10 time periods in this case) from the estimated sales data and the true sales data $[\varepsilon_{\mathrm{GA}}(t) = \hat{s}_{\mathrm{GA}}(t) - s(t)]$.

Generate Simulated Sales Datasets

Step 3. Assuming that the errors are exchangeable, sample with replacement, for example, 10 error values $[\varepsilon_{\text{sampGA}}(i)]$ from the errors computed in Step 2.[16] The values of the subscript i range from 1 to 10 in this case, corresponding to the 10 time periods of pre-peak sales data for room air conditioners.

Step 4. Generate simulated sales data as $S_{\text{simul}} = \hat{s}_{\text{GA}}(t) + \varepsilon_{\text{sampGA}}(i)$. For example, for the first time period, the simulated sales data is generated by adding the error value that was sampled first ($i = 1$) in Step 3 with the estimated value of sales from GA for the first time period.

Step 5. Repeat Steps 3 and 4 one thousand times to generate one thousand simulated datasets.

Computing the Distribution of MAD

For each of the simulated datasets compute the pairwise absolute deviations as follows:

Step 6. Compute the one step ahead forecast absolute deviations from AF by estimating the Bass model using AF and censoring the simulated datasets from time period = 0 until time period = 9. The absolute deviation (AD) for time period 3 for simulated data 1 is given as $\text{ADAF}(3,1) = |\hat{s}_{\text{AFsimul}}(3,1) - S_{\text{simul}}(3,1)|$. $\hat{s}_{\text{AFsimul}}(3,1)$ is the predicted value of sales for time period 3 based on the estimates from AF using the first two time periods of data from simulated dataset 1.

Step 7. Compute the one step ahead forecast absolute deviation from GA similar to the procedure followed for AF. The absolute deviation for time period 3 for simulated data 1 is given as $\text{ADGA}(3,1) = |\hat{s}_{\text{GAsimul}}(3,1) - S_{\text{simul}}(3,1)|$.

Step 8. Compute the pairwise deviation by computing the difference between the absolute deviations from AF and GA. For example, the pairwise deviation (PD) for time period 3 using simulated data 1 is given as $\text{PD}_{\text{AFGA}}(3,1) = \text{ADAF}(3,1) - \text{ADGA}(3,1)$.

Using the above process we can compute 10,000 pairwise deviations (10 pairwise deviations for each of the 1,000 simulated datasets). The confidence interval of the pairwise deviations will enable us to infer if the forecast errors from GA are similar (confidence interval contains zero), better (confidence interval is less than zero), or worse (confidence interval is greater than zero) as compared to the forecast errors from AF. The above process can be repeated for absolute deviations of post-peak sales data, predictions of time to peak sales, and predictions of sales at peak time period.

[16] We give equal probability for sampling each error value. In other words, we assume that the errors have a uniform distribution.

References

Balakrishnan, P. V., Varghese A. Jacob. 1996. Genetic algorithms of product design. *Management Sci.* **42**(8) 1105–1118.

Bass, Frank M. 1969. A new product growth model for consumer durables. *Management Sci.* **15**(January) 215–227.

Bass, Frank M., Trichy V. Krishnan, Dipak C. Jain. 1994. Why the Bass model fits without decision variables. *Marketing Sci.* **13**(Summer) 203–223.

Basu, Amiya, Tridib Mazumdar, S. P. Raj. 2003. Indirect network externality effects on product attributes. *Marketing Sci.* **22**(2) 209–221.

Bemmaor, Albert, Janghyuk Lee. 2002. The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Sci.* **21**(2) 209–220.

Davison, A. C., D. V. Hinkley. 1997. *Bootstrap Methods and Their Applications*. Cambridge Series in Statistical and Probabilistic Mathematics, No. 1, Cambridge University Press, Cambridge, U.K.

Del Moral, P., L. Miclo. 2001. Asymptotic results for genetic algorithms with applications to nonlinear estimation. L. Kallel, B. Naudts, A. Rogers, eds. *Theoretical Aspects of Evolutionary Computation*. Springer-Verlag, Berlin, Germany, 439–494.

Dorsey, Robert E., Walter J. Mayer. 1995. Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features. *J. Bus. Econom. Statist.* **13**(1) 53–66.

Easingwood, Christopher J., Vijay Mahajan, Eitan Muller. 1983. A non-uniform influence innovation diffusion model of new product acceptance. *Marketing Sci.* **2**(3) 272–296.

Garber, Tal, Jacob Goldenberg, Barak Libai, Eitan Muller. 2004. From density to destiny: Using spatial dimension of sales data for early prediction of new product success. *Marketing Sci.* **23**(3) 419–428.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.

Griffiths, W. E., R. C. Hill, G. G. Judge. 1993. *Learning and Practicing Econometrics.* John Wiley and Sons, New York.

Jain, Dipak, Ram C. Rao. 1990. Effect of price on the demand for durables: Modeling, estimation, and findings. *J. Bus. and Econom. Statist.* **8**(April) 163–170.

Krishnan, Trichy V., Frank M. Bass, V. Kumar. 2000. Impact of a late entrant on the diffusion of a new product or service. *J. Marketing Res.* **37**(May) 269–278.

Kumar, V., Trichy V. Krishnan. 2002. Multinational diffusion models: An alternative framework. *Marketing Sci.* **21**(3) 318–330.

Lenk, Peter J., Amber G. Rao. 1990. New model from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Sci.* **9**(1) 42–53.

McCullough, B. D., Berry Wilson. 1999. On the accuracy of statistical procedures in Microsoft Excel 97. *Comput. Statist. and Data Anal.* **31** 27–37.

Midgley, David F., Robert E. Marks, Lee G. Cooper. 1997. Breeding competitive strategies. *Management Sci.* **43**(3) 257–275.

Moe, Wendy W., Peter S. Fader. 2002. Using advance purchase orders to forecast new product sales. *Marketing Sci.* **21**(3) 347–364.

Naik, Prasad A., Murali K. Mantrala, Alan G. Sawyer. 1997. Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* **17**(3) 214–234.

Parker, Philip M. 1994. Aggregate diffusion forecasting models in marketing: A critical review. *Internat. J. Forecasting* **10** 353–380.

Putsis, Willam P., Jr. 1998. Parameter variation and new product diffusion. *J. Forecasting* **17**(3/4) 231–257.

Putsis, Willam P., Jr., V. Srinivasan. 2000. Estimation techniques for macro diffusion models. Vijay Mahajan, Eitan Muller, Yoram Wind, eds. *New-Product Diffusion Models*. Kluwer Academic Publishers, Boston, MA, 263–293.

Salomon, Ralf. 1989. Evolutionary algorithms and gradient search: Similarities and differences. *IEEE Trans. on Evolutionary Comput.* **2**(July) 1998.

Seber, G. A. F., C. J. Wild. 1989. *Nonlinear Regression*. John Wiley and Sons, New York, 92.

Schmittlein, D., V. Mahajan. 1982. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Sci.* **1**(1) 57–78.

Shankar, Venkatesh, Gregory S. Carpenter, Lakshman Krishnamurthi. 1998. Late mover advantage: How innovative late entrants outsell pioneers. *J. Marketing Res.* **35**(1) 54–70.

Srinivasan, V., Charlotte H. Mason. 1986. Nonlinear least squares estimation of new product diffusion models. *Marketing Sci.* **5**(2) 169–178.

Sultan, Fareena, John U. Farley, Donald R. Lehmann. 1990. A meta-analysis of diffusion models. *J. Marketing Res.* **27**(February) 70–77.

Talukdar, Debabrata, K. Sudhir, Andrew Ainslie. 2002. Investigating new product diffusion across products and countries. *Marketing Sci.* **21**(1) 97–114.

Tellis, Gerard J., Stefan Stremersch, Eden Yin. 2003. The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Sci.* **22**(2) 188–208.

Van den Bulte, Christophe, Gary L. Lilien. 1997. Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Sci.* **16**(4) 338–353.

Venkatesan, Rajkumar, V. Kumar. 2004. A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing*. Forthcoming.

Venkatesan, Rajkumar, Trichy Krishnan, V. Kumar. 2000. Systematic changes in parameters of macro-level diffusion models: A further inquiry. Marketing Science Conference, University of California at Los Angeles, Los Angeles, CA.

Xie, Jinhong, X. Michael Song, Marvin Sirbu, Qiong Wang. 1997. Kalman filter estimation of new product diffusion models. *J. Marketing Res.* **34**(August) 378–393.