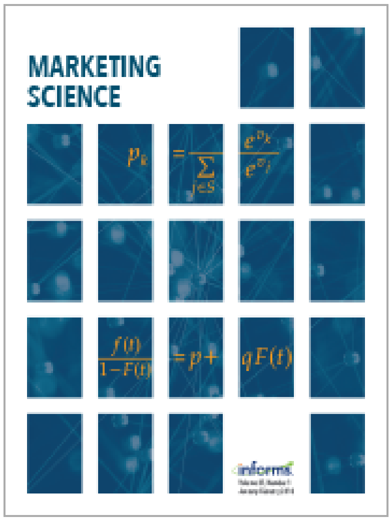

Marketing Science

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

Technical Note—Nonlinear Least Squares Estimation of New Product Diffusion Models

V. Srinivasan, Charlotte H. Mason,

MARKETING SCIENCE
Vol 5, No 2, Spring 1986
Printed in U.S.A

TECHNICAL Note

NONLINEAR LEAST SQUARES ESTIMATION OF NEW PRODUCT DIFFUSION MODELS

V. SRINIVASAN AND CHARLOTTE H. MASON
Stanford University
University of North Carolina

Schmittlein and Mahajan (*Marketing Science* 1982) made an important improvement in the estimation of the Bass (1969) diffusion model by appropriately aggregating the continuous time model over the time intervals represented by the data. However, by restricting consideration to only sampling errors and ignoring all other errors (such as the effects of excluded marketing variables), their Maximum Likelihood Estimation (MLE) seriously underestimates the standard errors of the estimated parameters. This note uses an additive error term to model sampling and other errors in the Schmittlein and Mahajan formulation. The proposed Nonlinear Least Squares (NLS) approach produces valid standard error estimates. The fit and the predictive validity are roughly comparable for the two approaches. Although the empirical applications reported in this paper are in the context of the Bass diffusion model, the NLS approach is also applicable to other diffusion models for which cumulative adoption can be expressed as an explicit function of time.
(New Product Models; Diffusion Models; Forecasting)

1. Introduction

In their paper "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," Schmittlein and Mahajan (1982) estimate the parameters of the Bass new product growth model using maximum likelihood estimation (MLE). Their approach yields an improvement in predictive validity over the classical approach (Bass 1969) of using ordinary least squares (OLS) on discrete time series. Schmittlein and Mahajan demonstrate that a bias is present in the classical approach because discrete time series data are used for estimating a continuous time model.[1] One advantage of the Schmittlein and Mahajan formulation is that the time interval bias is eliminated by using appropriate aggregation of the continuous time model over the time intervals represented

[1] Bass (1969, p. 223) recognizes the bias and provides a method for reducing it. However, a substantial amount of bias is likely to remain since his correction procedure makes the unrealistic assumption that the evolution of sales over time resembles the exponential density function.

0732-2399/86/0502/0169$01.25

by the data. Another cited advantage of the maximum likelihood approach is that approximate standard errors of the estimated parameters can be determined. However, because their maximum likelihood formulation considers only sampling error and ignores all other sources of error (Johnston 1984, p. 14), the computed standard error estimates may be too optimistic. Schmittlein and Mahajan make a distinction between data such as the adoption of consumer durables by U.S. households and survey data such as the adoption of medical innovations by a sample of hospitals. In the case of survey data, the sampling error may account for a major portion of the total error so that the downward bias in the standard error estimates may not be large. However, in the case of data for the entire U.S. households, the sample size is 50 million or more so that the sampling error is negligible.[2] Consequently, the standard errors are likely to be seriously underestimated.

§2 of this note proposes a nonlinear least squares (NLS) estimation of the Bass diffusion model with an additive error term representing total error. This model is estimated using the appropriate time aggregation of the continuous model as in the Schmittlein and Mahajan paper and the results are compared. The results reported in §3 indicate that for product adoptions by 50 million U.S. households the standard errors estimated by the MLE are biased downwards by orders of magnitude. However, in the case of survey data on the adoption of medical innovations by a sample of about 200 hospitals, the downward bias is negligible. The results indicate that the NLS is comparable to the MLE in terms of fit and predictive validity.

2. The Model Formulation

The formulation presented in this note begins with the adopters' probability density function $f(t)$ for adoption at time t (Bass 1969):

$$f(t) = (p + qF(t))(1 - F(t))$$

which has the following distribution function (Schmittlein and Mahajan 1982):

$$F(t) = \frac{(1 - e^{-(p+q)t})}{\left(1 + \frac{q}{p} e^{-(p+q)t}\right)}$$

where $p(>0)$ and $q(>0)$ are interpreted as coefficients of innovation and imitation. Letting m be the number of eventual adopters, the sales $X(i)$ in the ith time interval (t_{i-1}, t_i) is given by:

$$X(i) = m[F(t_i) - F(t_{i-1})] + u_i,$$

$$X(i) = m\left[\frac{1 - e^{-(p+q)t_i}}{1 + \frac{q}{p} e^{-(p+q)t_i}} - \frac{1 - e^{-(p+q)t_{i-1}}}{1 + \frac{q}{p} e^{-(p+q)t_{i-1}}}\right] + u_i \quad \text{for} \quad i = 1, 2, \ldots, T, \quad (1)$$

where u_i is an additive error term with variance σ^2.[3] The inclusion of the error term may be considered to represent the net effect of (i) sampling errors, (ii) the impact of excluded

[2] The maximum sampling error present in the "sample" of 50 million households adopting during a time period is $\sqrt{50 \times 10^6 \times 0.5 \times 0.5} = 3.54 \times 10^3$, corresponding to a penetration of 50% during a time period (Spurr and Bonini 1973, p. 143). The mean absolute deviation for one-step-ahead forecast, however, is at least 500×10^3 (Schmittlein and Mahajan 1982, Table 3). Clearly, the sampling error is an insignificant part of the total error for such data.

[3] A formulation with a multiplicative error term was also estimated but the additive error structure provides better results in terms of predictive validity. The additive error term has the theoretical problem of leading to negative sales if the usual assumption of normality of u is made. In the empirical results reported subsequently, the average value of $\hat{X}(i)/\hat{\sigma}$ over all observations was 7.7 and six of the 47 observations had $\hat{X}(i)/\hat{\sigma} < 3$. The fitted values $\hat{X}(i)$ were always nonnegative. Thus the problem of negative sales is only a minor issue.

variables (Johnston 1984, p. 14) which in this model may be factors such as economic conditions, technological improvements, advertising, pricing, and other marketing and competitive effects, and (iii) misspecification of the density function. (Easingwood, Mahajan, and Muller 1983 demonstrate that the density function in the Bass model is misspecified by showing that their nonuniform influence model, that relates rate of imitation q to cumulative adoption F, improves predictive validity.) Using nonlinear least squares (NLS), the parameters p, q and m and their standard errors (using asymptotic approximations) may be estimated directly.[4] The empirical results reported below used the maximum likelihood estimates as the starting values for p, q and m but sensitivity analysis indicates that different starting solutions have a negligible effect on the final estimates.[5] Consequently, ordinary least squares (OLS) estimates (see §3) which are easier to obtain may be used to provide starting values for p, q and m.

3. Empirical Results

The nonlinear least squares (NLS) estimation procedure is illustrated with the same time series data for the four consumer durables used by Schmittlein and Mahajan. The NLS estimates are compared with results from the maximum likelihood (MLE) approach. The distribution function for the time of adoption of a randomly chosen household from the entire population is given by:

$$G(t) = cF(t) = \frac{c(1 - e^{-bt})}{(1 + ae^{-bt})},$$

$$L(a, b, c, \{x_i\}) = [1 - G(t_{T-1})]^{x_T} \prod_{i=1}^{T-1} [G(t_i) - G(t_{i-1})]^{x_i} \qquad \text{where}$$

$a = q/p$,
$b = p + q$,
c = probability of eventually adopting = m/M,
M = total number of potential adopters,
L = likelihood function,

and using ordinary least squares (OLS) with improper time aggregation of the continuous model:

$$X(i) = \left[p + \frac{qN(t_{i-1})}{m}\right][m - N(t_{i-1})] + \epsilon(i),$$

$$X(i) = pm + (q - p)N(t_{i-1}) - \frac{q}{m} N^2(t_{i-1}) + \epsilon(i) \tag{2}$$

$$= \alpha_1 + \alpha_2 N(t_{i-1}) + \alpha_3 N^2(t_{i-1}) + \epsilon(i)$$

where $N(t_i)$ = cumulative number of adopters up to and including time period i and $\epsilon(i)$ is an additive error term. The above equation for $X(i)$ assumes that the time intervals (t_{i-1}, t_i) are equal. Otherwise, the right-hand side of the equation has to be multiplied throughout by $(t_i - t_{i-1})$.

[4] Dodds (1973, p. 310) makes a passing reference to a "nonlinear regression routine", but presumably the biased aggregation of the continuous diffusion model over discrete time periods would still be a problem with that approach.

[5] All numerical estimates were obtained using the BMDP3R nonlinear regression procedure in the BMDP Statistical Software package (Dixon 1983). The parameter constraints $\hat{p} \geq 0$, $\hat{q} \geq 0$ and $\hat{m} \leq M$ (total number of potential adopters) may be placed on the estimated parameters. In the empirical results reported subsequently, the estimated parameters satisfied the restrictions even without the imposition of the constraints. (The only exception was the estimation for dishwashers where the restriction $\hat{m} \leq M$ was essential, since otherwise the estimation would not converge.)

TABLE 1
Parameter Estimates for Consumer Durables[1]

Product	# of Observations	Annual Data for Years	OLS p	OLS q	OLS m (10^3)	MLE[2] p	MLE[2] q	MLE[2] m (10^3)	NLS p	NLS q	NLS m (10^3)
Clothes Dryers	13	1949–1961	0.0196	0 3481	15,281	0.0120 (0.000011)	0.3512 (0.00021)	15,948 (6.53)	0.0136 (0.0023)	0.3267 (0.0364)	16,497 (1,234)
Room Air Conditioners	13	1949–1961	0.0170	0 4049	17,091	0.0072 (0.000007)	0.4228 (0.00019)	17,581 (5.73)	0.0094 (0.0020)	0.3748 (0.0402)	18,713 (1,411)
Color T.V.	8	1963–1970	0.0357	0.6719	35,494	0.0160 (0.00001)	0.6566 (0.00021)	38,637 (7.34)	0.0185 (0.0032)	0.6159 (0.0541)	39,659 (2,267)
Dishwashers	13	1949–1961	Yielded an incorrect sign for the regression coefficient $\hat{\alpha}_3$			0.0035 (0.00011)	0.1282 (0.00054)	41,013 (1,359)	0.0026[3] (0.00029)	0.1303[3] (0.0115)	53,291 —[3]

[1] Numbers in parentheses are estimated standard errors (asymptotic approximations). Since p, q, and m are nonlinear functions of the OLS parameters, standard error estimates are unavailable for OLS.

[2] The standard errors for MLE reported in Schmittlein and Mahajan (1982, Table 1) are incorrect because of errors in equations (A3) and (A5) of that paper. The numbers reported here are based on the corrected equations (Schmittlein 1985).

[3] The NLS estimation for dishwashers was done under the constraint that $\hat{m} \leq M$, since otherwise the estimation would not converge. Since the constraint $\hat{m} \leq M$ was binding, no standard error is reported for $\hat{m}$.

3.1. *Parameter Estimates and Standard Errors*

Table 1 summarizes the NLS, MLE and OLS estimates for the parameters p, q and m and the standard errors for the NLS and MLE models. The standard errors for OLS are not available since the parameters p, q and m are nonlinear functions of α_1, α_2 and α_3.[6] The results are self-explanatory. We note that the standard errors for the NLS estimation are orders of magnitude larger than the MLE standard errors. The MLE approach, by taking into account merely the sampling error, seriously underestimates the standard errors of $\hat{p}$, $\hat{q}$, and $\hat{m}$. By taking the total error into account, the NLS approach leads to valid results.

For dishwashers, the NLS estimation had to be done under the constraint that $\hat{m} \leq M$, since the estimation would not converge otherwise. It may be noted that for this data set, the OLS procedure yielded an incorrect sign for the regression coefficient $\hat{\alpha}_3$ and the MLE procedure produces a standard error for m which is about 200 times that for the other three product classes. An explanation of the difficulty of estimating the model for dishwashers appears in Figure 4 of Schmittlein and Mahajan (1982) where it can be seen that the annual sales are continually growing with no signs of an imminent peak. (The sales data for the other three products include peaks.) Draper and Smith (1981, p. 467) provide a pictorial explanation of the difficulty of estimation under such data conditions.

3.2. *Fit Statistics*

In Table 2 the mean absolute deviation and mean squared error are reported for the four products. It is important to note that the statistics reported in Table 2 differ from the Schmittlein and Mahajan results because the fit statistics are computed here using only the actual and fitted number of adopters for each time period, rather than actual and fitted number of adopters for each time period and the actual and fitted number of nonadopters remaining. Some important comments on these results follow:

- In terms of mean absolute deviation, both MLE and NLS models provide a much

[6] One can, however, obtain asymptotic approximations to the standard errors of $\hat{p}$, $\hat{q}$, and $\hat{m}$ by using the approach outlined in Kendall and Stuart (1977, p. 247).

TABLE 2
Fit Statistics for Consumer Durables[1]

Product	Mean Absolute Deviation OLS	Mean Absolute Deviation MLE	Mean Absolute Deviation NLS	Mean Squared Error OLS	Mean Squared Error MLE	Mean Squared Error NLS
Clothes Dryers	111.80	102.46	101.52	20,818	17,380	16,367
Room Air Conditioners	173.20	145.46	144.59	41,265	30,951	26,267
Color T.V.	392.40	260.40	276.84	282,522	136,143	119,474
Dishwashers	—[2]	35.43	33.15	—[2]	1,813	1,751

[1] The fit statistics for MLE and NLS were computed using equation (1) in the text. For OLS, the fit statistics were computed using equation (2). Using equation (1) would make the OLS results considerably worse.

[2] OLS yielded an incorrect sign for the regression coefficient $\hat{\alpha}_3$.

better fit than OLS. The results are very similar for MLE and NLS with NLS slightly better in three cases and MLE better in one.

- In terms of mean squared error, again MLE and NLS significantly outperform OLS. By definition, NLS is at least as good as MLE since the objective of the NLS procedure is to minimize the error of squared residuals. For the four consumer durables estimated, the mean squared error for the NLS are 85%, 88%, 94%, and 97% of the MLE estimates for room air conditioners, color television, clothes dryers and dishwashers, respectively.

3.3. *One-Step Ahead Forecasts*

Table 3 summarizes the one-step-ahead forecast performance of the three models in order to assess their predictive validity. In terms of mean absolute deviation as well as mean squared error, MLE and NLS are much better than OLS. MLE and NLS yield very similar results with MLE better in two out of the four cases and NLS better in the other two cases.

3.4. *Empirical Results for Medical Innovations*

Schmittlein and Mahajan report empirical results for four medical innovations for which adoption time data were obtained from a survey of 209 hospitals. Unlike the four consumer durables studied earlier for which adoption time data were based on all potential adopters (numbering in excess of 50 million households), the medical innovation data are based only on a "small" sample. Consequently, for the medical innovations, sampling

TABLE 3
One-Step-Ahead Forecast Performance[1]

Product	Forecast Period	Mean Absolute Deviation OLS	Mean Absolute Deviation MLE	Mean Absolute Deviation NLS	Mean Squared Error OLS	Mean Squared Error MLE	Mean Squared Error NLS
Clothes Dryers	1955–1960	545[2]	285	323	464,299[2]	135,904	140,370
Room Air Conditioners	1956–1960	589	426	444	538,381	245,072	276,677
Color T.V.	1969–1972	3,211	2,560	2,369	11,664,233	8,529,629	7,490,945
Dishwashers	1955–1960	—[3]	67	40	—[3]	6,157	3,895

[1] Forecasts for MLE and NLS were computed using equation (1) in the text. For OLS, the forecasts were computed using equation (2). Using equation (1) would make the OLS results considerably worse.

[2] Results are for forecast periods 1957–60 because the OLS procedure yielded incorrect signs for the regression coefficients for the estimation periods 1949–1954 and 1949–1955.

[3] OLS yielded incorrect signs for the regression coefficient $\hat{\alpha}_3$.

errors may account for a major portion of the total error so that the downward bias in the MLE estimates of standard errors may not be large. Such an expectation is somewhat supported by the empirical results summarized below:[7]

- The relative magnitudes of the standard errors estimated by NLS and MLE for $\hat{p}$, $\hat{q}$ and $\hat{m}$ do not show any consistent pattern. Out of the twelve instances (four products, three parameter estimates), NLS had a larger standard error in five cases and MLE had a larger standard error in seven cases.
- Since NLS, by definition, minimizes sum of squared residuals, it would be expected to do better than MLE on this criterion. The mean squared residual and mean absolute residual for NLS were 29% and 50% respectively, of the corresponding MLE statistics.

4. Additional Considerations

4.1. *Other Diffusion Models*

Even though the presentation in this note has been limited to the Bass diffusion model, the MLE and NLS estimations can be applied to other diffusion models for which cumulative adoption can be expressed as an explicit function of time. (If cumulative adoption cannot be expressed as an explicit function of time, cumbersome numerical integration techniques have to be employed as part of the NLS and MLE estimation procedures.) Compared to the OLS, the NLS and MLE have the advantage that the model need not be linear in parameters. The OLS approach has the additional disadvantage of improper time aggregation of the continuous time model. As per the results presented earlier, the estimated standard errors and t-ratios are likely to be more valid in the NLS estimation compared to the MLE.

4.2. *Appropriateness of Asymptotic Approximations*

In the NLS approach the standard errors are based on asymptotic approximations. Given the small number of time periods for which data are usually available, one may wonder about the validity of the asymptotic approximations. Consequently, a small simulation study was conducted to determine whether there was a substantial bias in the reported standard errors (based on asymptotic approximations) in comparison to "true" standard errors.

The simulation study proceeded as follows:

(i) Based on some true parameter values p, q, m and σ^2 (variance of the error term u), data $\{x_t\} = \{x_1, x_2, \cdots x_T\}$ were randomly generated using equation (1). The normally distributed errors $\{u_t\}$ were randomly generated with mean zero and variance σ^2 and the times $\{t_i\}$ in equation (1) were taken to be 0, 1, 2, $\cdots T$. Applying NLS to the data $\{x_t\}$ yields estimates $\hat{p}$, $\hat{q}$, $\hat{m}$ and standard errors $s_{\hat{p}}$, $s_{\hat{q}}$, $s_{\hat{m}}$.

(ii) After repeating step (i) N times (N = number of replications in the simulation study), the averages $\bar{\hat{p}}$, $\bar{\hat{q}}$, $\bar{\hat{m}}$ and $\bar{s}_{\hat{p}}$, $\bar{s}_{\hat{q}}$, $\bar{s}_{\hat{m}}$ were computed. The parameters p, q, m, and σ^2 were held constant over the replications.

(iii) The bias in the parameter estimate of p is given by $\bar{\hat{p}} - p$. From the values $\hat{p}_1$, $\hat{p}_2, \ldots, \hat{p}_N$, an estimate of the 'true' standard error of $\hat{p}$ is given by

$$S_{\hat{p}} = \left[\frac{\sum_{i=1}^{N} (\hat{p}_i - \bar{\hat{p}})^2}{(N-1)} \right]^{1/2}.$$

The bias in the estimated standard error is given by $\bar{s}_{\hat{p}} - S_{\hat{p}}$. Similar calculations are done for q and m.

[7] Because of space limitations, detailed tabular results are not presented. Furthermore, given the small number of time periods for which data were available, no one-step-ahead forecast comparisons were made.

The simulation was done for the Clothes Dryer product class so that T was taken to be 13 (Table 1) and the true values for p, q and m were taken to be the numbers on the top row of Table 1 for the NLS model. The estimated residual variance from the NLS estimation of the Clothes Dryer data was taken to be equal to σ^2.

The results using $N = 50$ replications were that the biases in the standard errors for $\hat{p}$, $\hat{q}$ and $\hat{m}$ expressed as percentages (e.g., $|\bar{s}_{\hat{p}} - S_{\hat{p}}|/S_{\hat{p}}$) were less than 5%. The biases in the parameter estimates $\hat{p}$, $\hat{q}$ and $\hat{m}$ (e.g., $|\bar{\hat{p}} - p|/p$) were less than 7%.

The results of the small simulation study indicate that the asymptotic approximation used in calculating the standard errors has adequate accuracy and that the biases in the parameter estimates are small.

4.3. *Obtaining Standard Errors of Forecasts*

In addition to the parameters p, q, and m, other quantities such as peak sales, the timing of peak sales and forecasts of sales in future periods are of interest in practical applications. These quantities can be expressed in terms of the p, q, and m parameters as follows:

peak sales rate $= m(p + q)^2/4q$,

timing of peak sales $= \ln (q/p)/(p + q)$, and

total sales during time period $(\tau_1, \tau_2) = m[F(\tau_2) - F(\tau_1)]$,

where $F(t)$ is given by the formula at the beginning of §2. (The mathematical derivation of the first two quantities is given in Bass 1969.) Forecasts of the quantities can be obtained by substituting $\hat{p}$, $\hat{q}$, and, $\hat{m}$ in the corresponding expressions. To obtain the corresponding standard errors, we may proceed as follows. Let Q denote the quantity that is being forecasted. As in the examples above, it is a function g (say) of the parameters p, q, and m: $Q = g(p, q, m)$ so that $\hat{Q} = g(\hat{p}, \hat{q}, \hat{m})$. To obtain the asymptotic variance of $\hat{Q}$, we use the first-order Taylor series approximation of $\hat{Q}$:

$$\hat{Q} - Q \simeq (\hat{p} - p)(\partial g/\partial p) + (\hat{q} - q)(\partial g/\partial q) + (\hat{m} - m)(\partial g/\partial m).$$

Let $\mathbf{h}' = (\partial g/\partial p, \partial g/\partial q, \partial g/\partial m)$. It follows that (Johnston 1984, p. 162)

$$\text{variance of } (\hat{Q} - Q) \simeq \mathbf{h}'\mathbf{V}\mathbf{h} \tag{3}$$

where $\mathbf{V}$ is the variance-covariance matrix of the estimated parameters $(\hat{p}, \hat{q}, \hat{m})$. Using the estimated variance-covariance matrix $\hat{\mathbf{V}}$ obtained from the NLS computer program and by evaluating $\mathbf{h}$ at $(\hat{p}, \hat{q}, \hat{m})$, the above equation can be used to provide an approximate variance and hence standard error (square root of variance) for $\hat{Q}$.

The simulation procedure detailed in §4.2 was repeated to evaluate the adequacy of the approximate standard errors for the peak sales rate, timing of peak sales and the sales for the $T + 1$st time period. (The estimation data were for periods $1, 2, \ldots, T$.)[8] The results using $N = 50$ replications were that the biases in the standard errors were less than 10%. (The biases in the forecasted quantities were less than 3%.) These results again indicate that the asymptotic approximation is reasonable and the biases in the predicted quantities are small.

4.4. *Heteroscedasticity*

In the empirical studies reported so far, yearly data were used so that the time intervals (t_{i-1}, t_i) in equation (1) are equal. If the time intervals had not been equal, the errors

[8] Technically, a distinction needs to be made between the variance of the expected sales estimate and the variance of the actual sales around the forecasted value. The former variance was the one considered in the simulation study. It arises solely due to the estimation errors in ($\hat{p}$, $\hat{q}$, and $\hat{m}$) and can be estimated using equation (3). To obtain the latter variance, we need to add the additional variance created by the error term u which captures the difference between actual sales and expected sales. (See Johnston 1984, pp. 43–44 for this distinction in the context of the two-variable linear model.)

TABLE 4

Parameter Estimates as a Function of the Number of Years of Data Available—OLS

Product	# of Years of Data	p	q	m (10^3)	Magnitude of Peak Sales (10^3)	Timing of Peak Sales
Clothes Dryers	4	0.0807	1.2512[1]	1,791	635	2.06
	8	0.0139	0.4233	16,190	1,827	7.82
	12	0.0194	0.3845	13,980	1,483	7.39
Room Air Conditioners	4	0.0083	0.5032	13,808	1,795	8.03
	8	0.0151	0.6738	9,756	1,718	5.52
	12	0.0167	0.4425	15,731	1,874	7.13
Color T.V.	4[2]					
	8	0.0357	0.6719	35,494	6,613	4.15
Dishwashers	4[2]					
	8[2]					
	12[2]					

[1] This estimate of q violates the constraint $p + q \leq 1$ for the discretized version of the model.

[2] OLS yielded an incorrect sign for the regression coefficient $\hat{\alpha}_3$.

may be heteroscedastic, i.e., the $\sigma_{u_t}^2$ may differ from one observation to the other. By modeling $\sigma_{u_t}^2$ to be a function of $(t_i - t_{i-1})$, methods of weighted least squares can be applied to get around the problem of heteroscedasticity (Johnston 1984, p. 303).

Alternative models of heteroscedasticity could also be postulated. For instance, one would expect the sampling error is proportional to $\sqrt{P_t(1 - P_t)}$ where P_t is the probability of adoption during time period t. Since P_t is usually small in the present context, the sampling error is roughly proportional to $\sqrt{P_t}$. Leaving aside sampling error, the remaining errors may be expected to be proportional to P_t. To examine such possibilities, correlations were computed between the absolute value of the NLS residual $|e_t|$ and (i) $\sqrt{P_t(1 - P_t)}$, (ii) $\sqrt{P_t}$, (iii) P_t, and (iv) $P_t(1 - P_t)$ for the four data sets listed in Table 1. The correlation averaged over the four data sets was negative, small (≤ 0.10), and statistically insignificant for each of the four cases (i)–(iv) above. Consequently, the assumption of homoscedasticity appears to be a reasonable one in the empirical contexts considered here.

TABLE 5

Parameter Estimates as a Function of the Number of Years of Data Available—MLE

Product	# of Years of Data	p	q	m (10^3)	Magnitude of Peak Sales (10^3)	Timing of Peak Sales
Clothes Dryers	4	0.0307	0.9611	2,518	644	3.47
	8	0.0097	0.4017	16,052	1,691	9.04
	12	0.0118	0.3777	14,823	1,489	8.89
Room Air Conditioners	4	0.0109	0.4662	7,815	1,810	7.87
	8	0.0061	0.5955	11,333	1,722	7.61
	12	0.0069	0.4537	16,388	1,915	9.10
Color T.V.	4	0.0087	0.7311	56,347	10,545	5.99
	8	0.0160	0.6566	38,637	6,655	5.52
Dishwashers	4	0.0056	0.0461	33,913	492	40.72
	8	0.0039	0.0898	42,919	1,050	33.44
	12	0.0031	0.1219	47,398	1,519	29.30

TABLE 6

Parameter Estimates as a Function of the Number of Years of Data Available—NLS

Product	# of Years of Data	p	q	m (10^3)	Magnitude of Peak Sales (10^3)	Timing of Peak Sales
Clothes Dryers	4	0.0321	0.9178	2,618	643	3.53
	8	0.0087	0.3641	20,166	1,924	10.03
	12	0.0131	0.3560	15,268	1,461	8.95
Room Air Conditioners	4	0.0046	0.4260	19,408	2,112	10.52
	8	0.0070	0.5213	13,507	1,808	8.16
	12	0.0090	0.4009	17,537	1,838	9.26
Color T.V.	4[1]	0.0072	0.7339	65,444	12,244	6.24
	8	0.0185	0.6159	39,659	6,479	5.53
Dishwashers	4	0.1118	0.8260	991	264	2.13
	8[1]	0.0028	0.1127	53,291	1,577	31.99
	12[1]	0.0026	0.1302	53,291	1,804	29.49

[1] For these three cases, the NLS estimation was done under the constraint that $\hat{m} \leq M$, since otherwise the estimation would not converge.

4.5. *Autocorrelation*

For the empirical studies reported in Table 2, the serial correlation between the residuals in successive time periods was small (less than 0.25 in all cases). Had the serial correlation been larger and statistically significant, econometric procedures for handling autocorrelation problems need to be applied (Johnston 1984, p. 323).

4.6. *Number of Years of Data Required*

Tables 4–6 list the estimation results as the number of years of data available is systematically changed from 4 to 8 to 12. (The results in Table 1 were obtained with 13 years of data except for Color Television which used 8 years of data.) We find that the results are quite unreliable when only the first four years of data are available. Our results are consistent with the detailed study by Heeler and Hustad (1980, p. 1013) who recommend at least ten years of input data, including data on peak sales. Draper and Smith (1981, p. 467) provide a pictorial illustration of the difficulty of estimating such models with only a few years of data. An intuitive explanation in the OLS context is that the correlation between the cumulative sales $N(t_{i-1})$ and its square $N^2(t_{i-1})$ is so high that multicollinearity leads to very poor estimators.

A comparison of Tables 4, 5 and 6 indicates that the OLS procedure is considerably less stable than MLE and NLS. The stability of the MLE and NLS results are roughly comparable. Using the forecasted peak sales and peak periods with twelve years of data as the standard for comparison, the results seem to indicate that MLE may be more accurate when there are only four years of data and NLS may be more accurate when there are eight years of data.

5. Summary

An important contribution of Schmittlein and Mahajan's (1982) formulation is the appropriate time aggregation of the Bass continuous time diffusion model over the time intervals represented by the data. In their Maximum Likelihood Estimation (MLE), Schmittlein and Mahajan consider only sampling errors and ignore all other errors, such as the effects of excluded variables and the misspecification of the probability density function for adoption time. The results in Table 1 for product adoptions by the entire U.S. households show that the standard errors estimated by the MLE approach are biased downwards by orders of magnitude. The downward bias in the MLE standard error estimates appears to be negligible when the adoption data were based on a sample size of about 200. This note has proposed a Nonlinear Estimation (NLS) of the Bass diffusion model by modifying the approach taken by Schmittlein and Mahajan to take into account the total error. The key advantage of NLS is in obtaining valid standard error estimates. Furthermore, computer packages are more readily available for NLS (e.g., BMDP Statistical Software Package 1983) than for MLE. In terms of fit and predictive validity the NLS estimation performs comparably to the MLE formulation with both outperforming OLS (Ordinary Least Squares) regression which has a time aggregation bias. A more comprehensive study by Mahajan, Mason and Srinivasan (1986) concludes that NLS outperforms MLE in terms of one-step-ahead forecasts. Although the empirical results reported here are in the context of estimating the Bass diffusion model, the NLS (and MLE) approaches are also applicable to other new product diffusion models for which cumulative adoption can be expressed as an explicit function of time.[9]

Acknowledgements. The authors wish to thank Professor Vijay Mahajan of Southern Methodist University, Professor David Schmittlein of the University of Pennsylvania and the reviewers for their useful comments on earlier versions of this note.

[9] This paper was received in August 1984 and has been with the authors for 2 revisions.

References

Bass, F. M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (January), 215–227.

Dixon, W. J. (1983), *BMDP Statistical Software*. Berkeley, CA: University of California Press.

Dodds, W. (1973), "An Application of the Bass Model in Long Term New Product Forecasting," *Journal of Marketing Research*, 10 (August), 308–311.

Draper, N. R. and H. Smith (1981), *Applied Regression Analysis*. Second Edition, New York: John Wiley & Sons, Inc.

Easingwood, C. J., V. Mahajan and E. Muller (1983), "A Nonuniform Influence Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 2 (Summer), 273–295.

Heeler, R. M. and T. P. Hustad (1980), "Problems in Predicting New Product Growth for Consumer Durables," *Management Science*, 26 (October), 1007–1020.

Johnston, J. (1984), *Econometric Methods*. Third Edition, New York: McGraw-Hill.

Kendall, M. G. and A. Stuart (1977), *The Advanced Theory of Statistics*. Vol. 1, London: Griffin.

Mahajan, V., C. H. Mason and V. Srinivasan (1986), "An Evaluation of Estimation Procedures for New Product Diffusion Models." In V. Mahajan and Y. Wind (eds.), *Innovation Diffusion Models of New Product Acceptance*, Cambridge, MA: Ballinger, 203–232.

Schmittlein, D. C. (1985), Personal communication.

——— and V. Mahajan (1982), "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 1 (Winter), 57–78.

Spurr, W. A. and C. P. Bonini (1973), *Statistical Analysis for Business Decisions*. Revised Edition, Homewood, IL: Richard D. Irwin.