Forecasting from others' experience: Bayesian estimation of the Generalized Bass Model

Andrés Ramírez Hassan* Santiago Montoya Blandón†

June 28, 2016

Abstract

We introduce a Bayesian estimation procedure for the Generalized Bass Model used in product diffusion models. Our method attempts to forecast product sales early by virtue of previous similar markets. In a few simulation exercises, we compare our forecasting results to a frequentist nonlinear least squares method and a similar Bayesian alternative under different scenarios. We also apply our method to forecast the sales of room air conditioners using information from the market of clothes dryers. The results show that our Bayesian procedure provides accurate forecasts and is particularly useful when little to no historical data are available, that is, when sales projections are the most useful.

Keywords: Bayesian Estimation, Diffusion, Generalized Bass Model, Sales Forecast

*Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia, Carrera 49 - 7 S 50; Ph.: (+57)(4)261 95 00, ext. 9402; email: aramir21@eafit.edu.co

†Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia; email: smontoy4@eafit.edu.co

1 Introduction

Diffusion models have been remarkably important in several disciplines, and have been crucial in the study of the introduction of new products (Sultan et al., 1990). A considerable amount of academic research has addressed the subject, covering a wide range of different products, marketing strategies, forecasting procedures, and interactions between market forces (excellent reviews can be found in Mahajan et al., 1990; Bridges et al., 1991; Parker, 1994; Meade and Islam, 2006; Rao and Kishore, 2010). Among the many models used in studying product diffusion, the Bass model (BM) (Bass, 1969), introduced by Bass in 1969, stands out as the most extensively studied (Peres et al., 2010). This model is particularly interesting due to its ability to replicate and forecast sales data for a broad array of products through simple mathematical and statistical concepts. It was also the first model to explicitly include behavioral assumptions of consumers into the diffusion framework, using the ideas of Rogers (1962). The BM relies on some concepts from survival analysis, specifically the hazard rate. This quantity is defined as the rate of adopting a product or service, conditional on not having adopted it previously. The hazard rate of adoption is assumed to be linearly related to the fraction of total buyers from the potential market of a product.

The intuition is that a segment of the population, known as the "innovators," adopt the product regardless of who has bought it before, and then conveys the information through interpersonal communication (word of mouth) to the "imitators," who later adopt the product. The behavioral assumptions underlying the model are contained in three key parameters: the innovation coefficient (p), the imitation coefficient (q), and the short-run potential market (m).

The BM is regarded as one of the most important ideas of marketing research, as shown by its appearance on the list of the ten most influential articles of the magazine *Management Science* (Bass, 2004; Hopp, 2004). However, it has not been deemed flawless. One of its major shortcomings is that it initially does not allow the inclusion of variables such as price, advertisement, or promotion, which play an important role in determining the eventual adoption by consumers (Schmittlein and Mahajan, 1982). While many authors have introduced modifications to the Bass model or their own models that take this issue into account, Bass et al. (1994) introduced one of the most elegant solutions, referred to as the Generalized Bass Model (GBM). Their model extends the original Bass model to include *market effort*, which comprises a linear combination of the growth rate of the price, advertisement, and possibly other variables. One particular advantage is that the GBM reduces to the standard Bass model in case these variables are not significant in explaining the adoption rate.

The GBM has been found to perform well at forecasting sales curves for various products, and therefore will be the focus of our analysis (Chandrasekaran and Tellis, 2007; Dalla Valle and Furlan, 2011). Both models can be used to estimate the parameters given sufficient sales data, i.e., data up to, and including, the sales peak for a given product (Mahajan et al., 1986). Additionally, several estimation procedures have been proposed in the literature. Bass (1969) introduced a model that could be estimated by Ordinary Least Squares (OLS) after transforming the solution of the model from continuous to discrete time. Schmittlein and Mahajan (1982) proposed a Maximum Likelihood (ML) approach based on the unconditional probabilities of adoption that corrected for time aggregation bias and allowed the computation of standard errors for the parameters. Srinivasan and Mason (1986) introduced a nonlinear least squares (NLS) method that estimates better standard errors for the parameters and is usually the preferred estimation approach in the diffusion literature (see Mahajan et al., 1986, for a comparison of the three methods). Other methods have also been proposed in the literature: these include stochastic differential equations, different ways of discretizing the BM, and Kalman filter estimation (Skiadas and Giovanis, 1997; Xie et al., 1997; Satoh, 2001).

From the point of view of a manager planning on introducing an innovation into the market, demand forecasts are fundamental in the planning of production, distribution, and marketing efforts for the new product. Obtaining reliable and accurate forecasts prior to or during the initial years of a launch is therefore a crucial element in product planning. The main concern, however, is that in order to estimate the diffusion models to forecast the sales curve, one needs sufficient data, and by the time enough data has been collected, the product has reached maturity and the forecasts are useless (Lenk and Rao, 1990; Mahajan et al., 1990). This is the reason why most authors have turned to Bayesian techniques to forecast sales information, since it fits naturally the problem structure: it systematically incorporates prior information of possibly related products to make predictions, and then updates these predictions as data becomes available (Lilien et al., 1981; Lenk and Rao, 1990; Mahajan et al., 1990; Putsis and Srinivasan, 2000; Lee et al., 2003; van Everdingen et al., 2005).

The main objective of the present paper is to propose a Bayesian method for forecasting sales information of new products that is based on the GBM. The method allows great flexibility, by incorporating any number of covariates into the marketing effort component, allows for parameter estimation of the GBM before sales reach a peak, and is able to produce a forecast even when there is no data available, by using of prior information. The adaptation of the posterior distribution is rapid as data becomes available and we can obtain accurate forecasts before the sales peak is reached. That is, forecasts are feasible before frequentist methods, such as NLS, can be used to estimate precisely the parameters.

The remainder of this paper is structured as follows: Section 2 introduces the method and some of the innovations with which we propose to estimate the model, using a Bayesian approach. Section 3 presents some simulation results that compare our approach to others used in the literature, with some comments on their forecast accuracy. Section 4 presents the results of applying the method to the GBM data provided by Bass et al. (1994). Lastly, Section 5 presents our concluding remarks.

2 The Method

The usual representation of the Bass model is

$$\frac{f(t)}{1-F(t)} = p + \frac{q}{m}Y(t) \tag{1}$$

where the parameters p, q, and m are interpreted as previously stated. Here, $F(t)$ and $f(t)$ are the cumulative and non-cumulative proportion of adopters at time t, and $Y(t)$ is the total number of adopters up to but not including t. To derive the Generalized Bass Model, we incorporate market effort into the BM framework. This addition is introduced as a multiplicative term on the right hand side of Eq. (1)

$$\frac{f(t)}{1-F(t)} = \left[p + \frac{q}{m}Y(t)\right]x(t) \tag{2}$$

where $x(t)$ is the market effort, usually assumed to be (Simon, 1982; Bass et al., 1994)

$$x(t) = 1 + \beta_1\frac{P(t)-P(t-1)}{P(t-1)} + \beta_2\frac{\max\{0, A(t)-A(t-1)\}}{A(t-1)} \tag{3}$$

where $P(t)$ and $A(t)$ are the price and the advertising, respectively. These variables enter the market effort equation as percent increases. The inclusion of other variables is then trivial.

We propose a Bayesian method for estimating and forecasting the Generalized Bass Model, which fits naturally into the innovation diffusion context. As with all Bayesian models, we will need to supply a likelihood function and prior distributions, to be combined using Bayes' theorem (Rossi and Allenby, 2003). For the likelihood function, we will use the sales function proposed by Srinivasan and Mason (1986). By using an appropiate time aggregation due to Schmittlein and Mahajan (1982), the sales $S(t)$ between times t and $t-1$ are defined as

$$S(t) = m[F(t) - F(t-1)] + u_t \tag{4}$$

where u_t is an additive error term with variance σ^2, and $F(t)$ is the solution to the Generalized Bass Model, given by (see the Appendix)

$$F(t) = \frac{1 - \exp\{-X(t)(p+q)\}}{1 + (q/p)\exp\{-X(t)(p+q)\}} \tag{5}$$

where $X(t)$ is the cumulative market effort, found by transforming Eq. (3) into continuous time and integrating from 0 to t:

$$X(t) = t + \beta_1 \ln\left[\frac{P(t)}{P(0)}\right] + \beta_2 \ln\left[\frac{\hat{A}(t)}{A(0)}\right] \tag{6}$$

where $\hat{A}(t)$ is the last value for which there was a positive change in advertising. In more general terms, Eq. (6) can be written as

$$X(t) = t + \boldsymbol{\beta}' \ln\left[\frac{Z(t)}{Z(0)}\right] \tag{7}$$

where $\boldsymbol{\beta}$ is a vector of parameters and $Z(t)$ a vector of covariates at time t. We assume that u_t is normally distributed and so the likelihood function is also normal. With T intervals of data and writing $\boldsymbol{S}$, $\boldsymbol{F}$, $\boldsymbol{P}$, and $\boldsymbol{A}$ for the collection of sales, cumulative adoption, price, and advertising, respectively, the likelihood function can be expressed as

$$f(\boldsymbol{S}|p,q,m,\boldsymbol{\beta},\sigma^2,\boldsymbol{P},\boldsymbol{A}) = \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \exp\left\{-\frac{1}{2\sigma^2}[\boldsymbol{S} - m\Delta\boldsymbol{F}(p,q,\boldsymbol{\beta})]'[\boldsymbol{S} - m\Delta\boldsymbol{F}(p,q,\boldsymbol{\beta})]\right\} \tag{8}$$

Writing $\theta = [p, q, m, \boldsymbol{\beta}']'$, we assume $\pi(\theta)$ to be a multivariate t-distribution with location prior hyperparameters θ_0, prior scale matrix Σ_0, and $\nu_0 = 3$ degrees of freedom. This choice for the Bass model parameters, including only a few degrees of freedom, implies fatter tails and therefore the robustness of the outcomes (Berger, 1985). For the variance parameter σ^2, we decide against using the standard, inverse gamma $\mathcal{IG}(\epsilon, \epsilon)$ distribution with $\epsilon \to 0$

(Kelsall and Wakefield, 1999; Spiegelhalter et al., 2003). As shown by Gelman (2006), this distribution does not actually possess a proper limiting posterior distribution and should be avoided in empirical work. In contrast, we will use the Scaled-Beta2(a, b, κ) distribution proposed by Fúquene et al. (2014).[1] This distribution can be derived from a mixture of gamma distributions and is easily simulated from the scaled odds of a beta distribution (Fúquene et al., 2014, see Appendix). We follow Fúquene et al. (2014) and set the prior hyperparameters $a_0 = b_0 = 1$ and set κ_0 to a large number, to achieve a vague and robust distribution. Preliminary simulation exercises using the inverse gamma distribution showed poor performance in comparison to the Scaled-Beta2 distribution, which further motivated our selection.

The prior distribution for the GBM is

$$\begin{aligned}\pi(\theta, \sigma^2) &= \pi(\theta)\pi(\sigma^2) \\ &= \boldsymbol{t}(\theta_0, \Sigma_0, \nu_0)\text{SBeta2}(a_0, b_0, \kappa_0)\end{aligned} \tag{9}$$

For our model estimation, we set the prior parameters $p_0, q_0, m_0, \boldsymbol{\beta}_0$, as well as Σ_0, from previous, similar markets (Lilien et al., 1981; Lenk and Rao, 1990; Parker, 1994; Lee et al., 2003). Even in the event that no data are available, we can extract the values for the parameters from the priors and use them to predict future sales.

Combining the likelihood function and the prior distribution using Bayes' rule, we obtain the posterior distribution for the GBM parameters

$$\begin{aligned}\pi(\theta, \sigma^2|\boldsymbol{S}, \boldsymbol{P}, \boldsymbol{A}) = &\left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \exp\left\{-\frac{1}{2\sigma^2}[\boldsymbol{S} - m\Delta\boldsymbol{F}(p, q, \boldsymbol{\beta})]'[\boldsymbol{S} - m\Delta\boldsymbol{F}(p, q, \boldsymbol{\beta})]\right\} \\ &\times \frac{\Gamma(\nu_0/2 + 2)}{|\Sigma_0|^{1/2}(\nu_0\pi)^2\Gamma(\nu_0/2)} \left[\nu_0 + (\theta - \theta_0)'\Sigma_0^{-1}(\theta - \theta_0)\right]^{-\nu_0/2-2} \\ &\times \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \frac{1}{\kappa_0} \frac{(\sigma^2/\kappa_0)^{a_0-1}}{(\sigma^2/\kappa_0 + 1)^{a_0+b_0}}\end{aligned}$$

[1] For a long time in the statistics literature, a general version of this distribution has been known as the generalized beta prime distribution (Dubey, 1970). Only recently has it been introduced into Bayesian analysis as a natural prior in dynamic models.

3 Simulation Exercises

In order to test the goodness of fit for our method, we performed some simulation exercises. Defining population values for p, q, m, and $\boldsymbol{\beta}$, allows us to find $F(t)$, $X(t)$ and $S(t)$. We assume there are three covariates, and that these are controlled by the manager. Therefore, at any time, the complete time path of the covariates is known. The variables were generated from normal distributions with means of 4000, 1500, and 3000 and standard deviations of 100, 25, and 50, respectively. We then generated a few random previous markets and obtained the prior hyperparameters from them. The location vector and scale matrix were calibrated using the means and variances, respectively, of each parameter across previous markets.

The reference values can be observed in Table 1. The population values for the simulation generate sales data with a peak at $t = 9$ (see the Appendix for the formulas for the estimated sales peak). We began by estimating the model at $t = 0, 5, 8, 9, 10, 15, 20, 25$ using our Bayesian method and the NLS technique of Srinivasan and Mason (1986). We estimated the Bayesian model using a Metropolis–Hastings sampling algorithm with 400000 iterations, a burn-in period of 200000, and extracting samples every 50 iterations to avoid autocorrelation in the chains (Robert and Casella, 2004). All the chains were checked for convergence using standard tests and were found to fulfill all convergence criteria (Heidelberger and Welch, 1983; Geweke, 1992; Raftery et al., 1992; Brooks and Gelman, 1998).

The results are presented in Table 2. The point estimates, and standard errors are the medians and standard deviations for the final chains, respectively. We estimated the frequentist model using three different algorithms: unconstrained optimization based on the Nelder–Mead algorithm (Nelder and Mead, 1965), the PORT subroutine library for constrained optimization (Fox et al., 1978), and a genetic algorithm that does not require initial values (Sekhon and Mebane, Jr., 1998).[2] All of the procedures returned similar values, therefore we present only those obtained using the PORT algorithm.

[2]All computations were performed using the R open-source programming language (R Core Team, 2015). Specific packages include Plummer et al. (2006); Genz et al. (2014); Plummer (2015); Su and Yajima (2015).

Table 1: Values for Simulation Exercises

	p	q	m	β_1	β_2	β_3	σ^2	κ
Population	0.010	0.40	20000	-1.5	1.0	-2.5	1	2000
Market 1	0.010	0.35	20000	-1.3	1.0	-1.0	1	2000
Market 2	0.020	0.40	30000	-1.0	1.5	-4.0	1	2000
Market 3	0.025	0.50	50000	-0.2	3.5	-2.5	1	2000

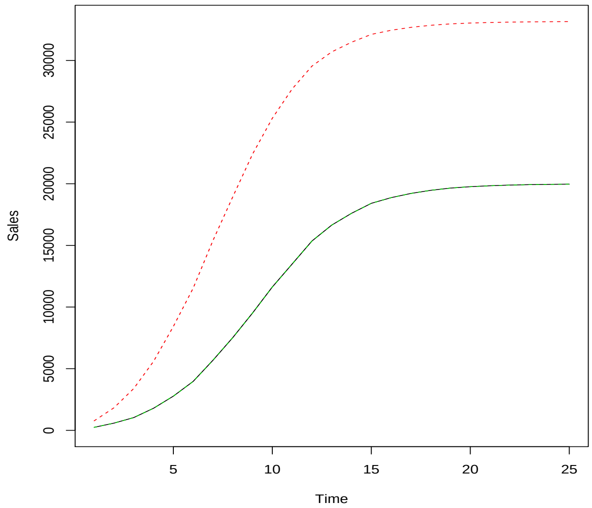
Source: Author's calculations

Table 2: Bayesian and NLS estimation with different amounts of information

	$t = 0$	$t = 5$	$t = 8$	$t = 9$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
	Bayesian estimation							
p	0.018	0.011	0.010	0.010	0.010	0.010	0.010	0.010
	(0.008)	(0.004)	(3.33e-04)	(7.14e-05)	(3.79e-05)	(1.38e-05)	(1.31e-05)	(1.16e-05)
q	0.412	0.405	0.401	0.401	0.401	0.400	0.400	0.400
	(0.077)	(0.031)	(0.007)	(0.003)	(0.001)	(2.19e-04)	(1.90e-04)	(1.73e-04)
m	34024.535	18514.368	19917.914	19980.627	19967.532	19993.028	19994.392	19995.073
	(15303.117)	(8347.344)	(905.511)	(232.882)	(65.088)	(7.007)	(6.482)	(5.823)
β_1	-0.793	-0.885	-1.387	-1.470	-1.503	-1.503	-1.486	-1.487
	(0.563)	(0.893)	(0.395)	(0.283)	(0.202)	(0.043)	(0.042)	(0.038)
β_2	1.986	1.118	1.083	1.046	1.029	1.009	1.010	1.009
	(1.304)	(0.663)	(0.250)	(0.161)	(0.107)	(0.014)	(0.014)	(0.012)
β_3	-2.520	-2.922	-2.804	-2.651	-2.597	-2.517	-2.521	-2.519
	(1.569)	(1.652)	(0.842)	(0.496)	(0.298)	(0.043)	(0.041)	(0.037)
	NLS estimation							
p	-	-	0.010	0.010	0.010	0.010	0.010	0.010
	-	-	(4.32e-05)	(1.70e-05)	(1.36e-05)	(1.03e-05)	(1.12e-05)	(1.04e-05)
q	-	-	0.401	0.400	0.401	0.400	0.400	0.400
	-	-	(0.001)	(0.001)	(4.83e-04)	(1.64e-04)	(1.65e-04)	(1.53e-04)
m	-	-	19917.063	19994.739	19971.402	19993.076	19994.519	19995.037
	-	-	(122.538)	(53.398)	(23.814)	(5.367)	(5.563)	(5.165)
β_1	-	-	-1.524	-1.541	-1.542	-1.506	-1.489	-1.490
	-	-	(0.093)	(0.082)	(0.074)	(0.032)	(0.035)	(0.033)
β_2	-	-	1.032	1.019	1.012	1.009	1.009	1.009
	-	-	(0.055)	(0.048)	(0.041)	(0.011)	(0.012)	(0.011)
β_3	-	-	-2.684	-2.586	-2.548	-2.516	-2.519	-2.518
	-	-	(0.213)	(0.149)	(0.116)	(0.032)	(0.035)	(0.033)

Standard errors in parenthesis.
"-" estimation was not possible.
Source: Author's calculations

Figure 1: Cumulative Sales forecast with information at different time periods



Source: Author's calculations

Curiously, in this simulation exercise, the frequentist approach yields good estimation results at $t = 8$, a period before the sales peak is reached. However, the impossibility of estimating the model in previous periods gives the Bayesian framework a clear advantage in terms of actual product planning. As we can see in Figure 1, even with only five data points, when achieving convergence for the frequentist estimation is impossible, the cumulative sales curve is approximated fairly well by the Bayesian GBM forecast. By the time eight data points have been collected, the forecast becomes almost perfect at that scale.

As a second exercise, we found the goodness of fit for our cumulative sales curve using one-step ahead forecasts. With each new data point, we computed the subsequent period of cumulative sales, and continued up to period $t - 1$. That is, a one-step ahead forecast at $t = 10$ would be the forecast of sales at $t = 11$ using the information collected up to $t = 10$. These results can be observed in Table 3 and Figure 2. Table 3 presents the mean squared error and the mean absolute error for the one-step ahead forecasts using both methods. The forecasting error remains small when more information is incorporated, as one would expect. Notice how going from $t = 0$ to $t = 5$, and from $t = 5$ to $t = 8$, the forecasting error is reduced by approximately 100% each time. It appears that both errors remain lower using

Table 3: Comparison of one step-ahead Mean Squared Error and Mean Absolute prediction error between Bayesian and Frequentist estimation with different amounts of information

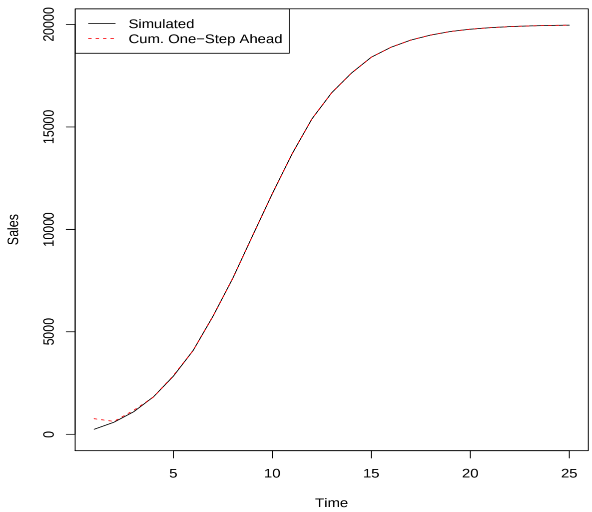
	$t=0$	$t=5$	$t=8$	$t=9$	$t=10$	$t=15$	$t=20$	$t=24$
Noncumulative Sales								
Bayesian estimation								
MSE	2.65E+05	5.44E+00	0.02	0.10	0.20	0.90	2.72	0.01
MAE	515.20	2.33	0.15	0.32	0.48	0.91	1.65	0.12
NLS estimation								
MSE	-	-	36.45	5.28	14.85	1.45	2.77	0.01
MAE	-	-	6.04	2.30	3.85	1.20	1.66	0.12
Cumulative Sales								
Bayesian estimation								
MSE	2.65E+05	12.5	0.12	0.15	0.02	0.49	0.63	0.35
MAE	515.20	3.53	0.35	0.38	0.13	0.70	0.79	0.59
NLS estimation								
MSE	-	-	35.72	7.00	13.58	0.25	1.19	0.40
MAE	-	-	5.98	2.64	3.68	0.50	1.09	0.64

Source: Author's calculations

Bayesian estimation in comparison to the NLS procedure. Figure 2 presents the one-step ahead forecast using information up to t to compute the observation at $t+1$. Other than a slight deviation at the starting periods, the one-step ahead forecasts seem to replicate cumulative adoption well. Before $t = 5$ is reached, the forecasting scheme seems to be adapted to the true underlying cumulative sales curve.

We then explored what would happen if the potential short-run market, m, was known. As noted by Trajtenberg and Yitzhaki (1989); Parker (1994); Van den Bulte and Lilien (1997), the estimation of the potential short-run market, either in terms of units or percentage of a fixed long-run market, is what introduces the most noise into the estimation and forecasting procedure. Parker (1994), particularly, notes that this parameter should not be estimated using early data on sales, and that there are several methods that work well in practice for approximating this value. Eliminating m from θ and fixing it at 20000 results in a vast improvement both in estimation and forecasting, as can be observed in Table 4 and Figure 3. Not having to estimate m makes both estimation procedures more accurate and computa-

Figure 2: Cumulative Sales one-step ahead forecast with information at different time periods



Source: Author's calculations

tionally faster. Bayesian estimation still has the advantage of providing sensible estimates when historical information is lacking. However, we can observe that NLS is now able to provide point estimates for the parameters at $t = 5$, albeit without standard errors. Table 5 presents the MSE and MAE for this exercise. In general, the trends from the previous exercises are maintained. Figure 3 shows how a similar m shapes the cumulative adoption curve. By having the endpoints of all forecast curves be the same, the forecasting accuracy is substantially improved.

We compare the results obtained in the previous exercises using a Scaled-Beta2 distribution with those generated by assuming the usual Inverse Gamma prior. Table 6 presents the point estimates and standard deviations from each approach. We observe that the estimations are quite similar in terms of point estimates. It would appear that using both priors, convergence to the population values is achieved around the sales peak. The standard deviations from the mean are also very similar, but seem to be lower using the Inverse Gamma prior.

Additionally, Table 7 presents the mean squared and mean absolute errors associated

Table 4: Bayesian and NLS estimation without estimating short-run market

	$t = 0$	$t = 5$	$t = 8$	$t = 9$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
				Bayesian estimation				
p	0.018	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	(0.008)	(4.82e-04)	(1.10e-04)	(5.32e-05)	(2.66e-05)	(1.16e-05)	(1.16e-05)	(1.00e-05)
q	0.417	0.401	0.400	0.400	0.400	0.400	0.400	0.400
	(0.079)	(0.012)	(0.002)	(0.001)	(3.72e-04)	(1.66e-04)	(1.60e-04)	(1.38e-04)
β_1	-0.800	-0.888	-1.488	-1.510	-1.549	-1.497	-1.485	-1.486
	(0.589)	(0.812)	(0.279)	(0.206)	(0.141)	(0.041)	(0.042)	(0.037)
β_2	1.926	1.081	1.048	1.036	1.000	1.008	1.010	1.009
	(1.304)	(0.560)	(0.166)	(0.122)	(0.075)	(0.014)	(0.014)	(0.012)
β_3	-2.473	-3.056	-2.679	-2.636	-2.509	-2.514	-2.521	-2.518
	(1.541)	(1.47)	(0.535)	(0.367)	(0.212)	(0.043)	(0.042)	(0.037)
				NLS estimation				
p	-	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	-	-	(2.16e-05)	(1.45e-05)	(1.21e-05)	(9.06e-06)	(9.95e-06)	(9.25e-06)
q	-	0.401	0.400	0.400	0.400	0.400	0.400	0.400
	-	-	(3.45e-04)	(2.36e-04)	(1.72e-04)	(1.29e-04)	(1.36e-04)	(1.26e-04)
β_1	-	-0.941	-1.544	-1.543	-1.577	-1.501	-1.487	-1.488
	-	-	(0.079)	(0.07)	(0.071)	(0.033)	(0.035)	(0.033)
β_2	-	1.065	1.022	1.019	0.989	1.008	1.009	1.009
	-	-	(0.048)	(0.042)	(0.038)	(0.011)	(0.012)	(0.011)
β_3	-	-3.169	-2.615	-2.589	-2.480	-2.513	-2.519	-2.518
	-	-	(0.168)	(0.126)	(0.106)	(0.033)	(0.035)	(0.033)

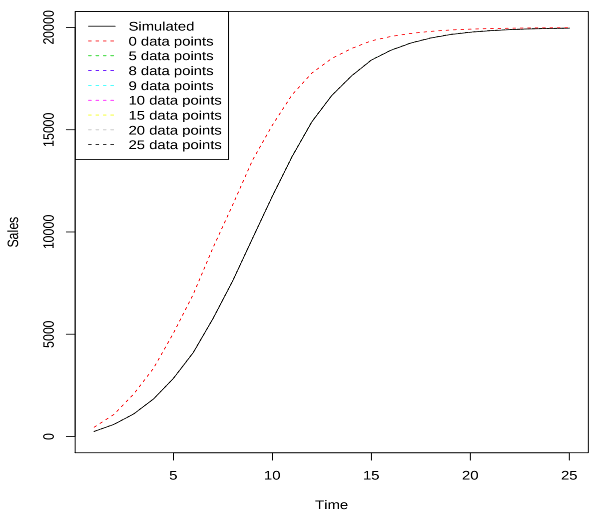
Standard errors in parenthesis.
"-" estimation was not possible.
Source: Author's calculations

Table 5: Comparison of one step-ahead Mean Squared Error and Mean Absolute prediction error between Bayesian and Frequentist estimation with different amounts of information without estimating short-run market

	$t=0$	$t=5$	$t=8$	$t=9$	$t=10$	$t=15$	$t=20$	$t=24$
Noncumulative Sales								
Bayesian estimation								
MSE	4.26e+04	23.18	0.54	9.21	4.23	0.52	2.52	0.01
MAE	206.41	4.82	0.74	3.04	2.06	0.72	1.59	0.10
NLS estimation								
MSE	-	-	36.45	5.28	14.85	1.45	2.77	0.01
MAE	-	-	6.04	2.30	3.85	1.20	1.66	0.12
Cumulative Sales								
Bayesian estimation								
MSE	4.26e+04	35.57	1.04e-03	13.82	0.01	24.18	17.53	18.43
MAE	206.41	5.97	0.03	3.72	0.11	4.92	4.19	4.29
NLS estimation								
MSE	-	23.46	1.05	7.23	8.77	0.54	2.52	0.01
MAE	-	4.84	1.03	2.69	2.96	0.73	1.59	0.10

Source: Author's calculations

Figure 3: Cumulative Sales forecast without estimating short-run market



Source: Author's calculations

Table 6: Bayesian Scaled-Beta2 and Inverse Gamma estimation without estimating short-run market

	$t=0$	$t=5$	$t=8$	$t=9$	$t=10$	$t=15$	$t=20$	$t=25$
				Scaled-Beta2				
p	0.018	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	(0.008)	(4.82e-04)	(1.10e-04)	(5.32e-05)	(2.66e-05)	(1.16e-05)	(1.16e-05)	(1.00e-05)
q	0.417	0.401	0.400	0.400	0.400	0.400	0.400	0.400
	(0.079)	(0.012)	(0.002)	(0.001)	(3.72e-04)	(1.66e-04)	(1.60e-04)	(1.38e-04)
β_1	-0.800	-0.888	-1.488	-1.510	-1.549	-1.497	-1.485	-1.486
	(0.589)	(0.812)	(0.279)	(0.206)	(0.141)	(0.041)	(0.042)	(0.037)
β_2	1.926	1.081	1.048	1.036	1.000	1.008	1.010	1.009
	(1.304)	(0.560)	(0.166)	(0.122)	(0.075)	(0.014)	(0.014)	(0.012)
β_3	-2.473	-3.056	-2.679	-2.636	-2.509	-2.514	-2.521	-2.518
	(1.541)	(1.47)	(0.535)	(0.367)	(0.212)	(0.043)	(0.042)	(0.037)
				Inverse Gamma				
p	0.019	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	(0.013)	(1.05e-04)	(3.27e-05)	(2.08e-05)	(1.56e-05)	(1.01e-05)	(1.06e-05)	(9.91e-06)
q	0.413	0.401	0.400	0.400	0.400	0.400	0.400	0.400
	(0.126)	(0.003)	(0.001)	(3.33e-04)	(2.19e-04)	(1.45e-04)	(1.45e-04)	(1.34e-04)
β_1	-0.832	-1.358	-1.510	-1.510	-1.517	-1.500	-1.496	-1.497
	(1.072)	(0.245)	(0.032)	(0.025)	(0.024)	(0.009)	(0.009)	(0.009)
β_2	1.995	1.065	1.026	1.020	0.993	1.009	1.010	1.009
	(2.081)	(0.245)	(0.074)	(0.058)	(0.051)	(0.012)	(0.013)	(0.012)
β_3	-2.487	-2.840	-2.560	-2.545	-2.495	-2.506	-2.510	-2.509
	(3.050)	(0.365)	(0.128)	(0.087)	(0.070)	(0.018)	(0.019)	(0.018)

Standard errors in parenthesis.
"-" estimation was not possible.
Source: Author's calculations

with the one step-ahead prediction from each model. While the downward trend of the errors as information is included in the analysis remains similar, it appears that the Scaled-Beta2 model still continues to be better in terms of forecasting. The one step-ahead forecasts are slightly improved from this base model in most scenarios. In particular, for most time periods before or exactly at $t = 10$, this difference is especially noticeable.

Lastly, in order to understand how the posteriors of our estimation process adapt to new information, we present some results in Figures 4 and 5. We plot the prior distribution and the posterior distribution including up to t periods in our Bayesian estimation framework. Essentially, what can be deduced from the figures is that after just five data points have

Table 7: Comparison of one step-ahead Mean Squared Error and Mean Absolute prediction error between Bayesian methodologies with different amounts of information without estimating short-run market

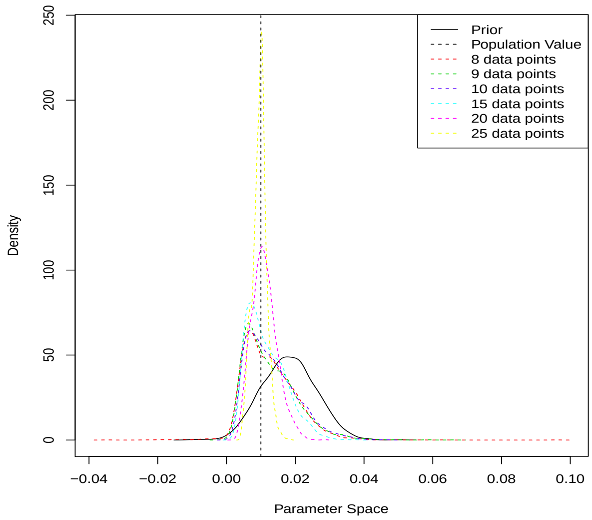
	$t=0$	$t=5$	$t=8$	$t=9$	$t=10$	$t=15$	$t=20$	$t=25$
				Noncumulative Sales				
				Scaled-Beta2				
MSE	4.26e+04	23.18	0.54	9.21	4.23	0.52	2.52	0.01
MAE	206.41	4.82	0.74	3.04	2.06	0.72	1.59	0.10
				Inverse Gamma				
MSE	4.30e+04	24.27	1.43	8.37	6.06	0.53	2.52	0.01
MAE	207.32	4.93	1.20	2.89	2.46	0.73	1.59	0.10
				Cumulative Sales				
				Scaled-Beta2				
MSE	4.26e+04	35.57	1.04e-03	13.82	0.01	24.18	17.53	18.43
MAE	206.41	5.97	0.03	3.72	0.11	4.92	4.19	4.29
				Inverse Gamma				
MSE	4.30e+04	24.55	0.73	10.71	0.03	24.05	17.52	18.43
MAE	207.32	4.95	0.85	3.27	0.18	4.90	4.19	4.29

Source: Author's calculations

been observed, the posterior shifts away from the prior and centers around the population value, i.e., the information contained in the data. This happens for all the parameters in the model. Furthermore, in these figures we can directly observe how including more data points in the estimation process increases the accuracy by reducing the width of the posterior. We conclude by noting that our model has several desirable properties, such as rapid adaptation to new information, an ability to make accurate forecasts with few data points, and sensible estimation results.

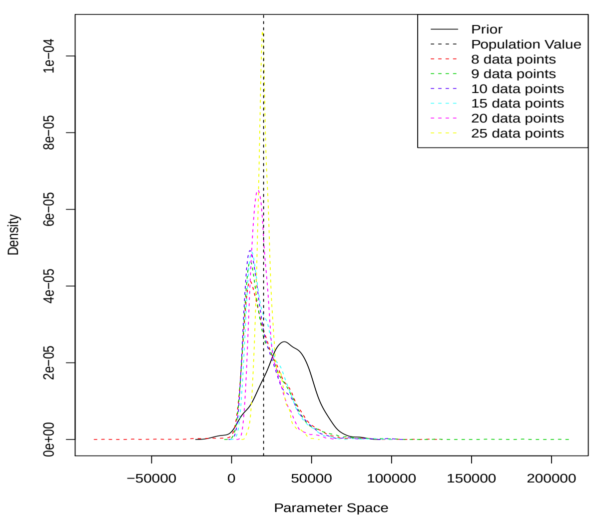
It is also important to note that even when varying the values for previous market parameters and the scaling factor of the variance, κ, over a relatively wide range, the results of our exercises remain fairly constant, providing evidence for robustness (available upon request).

Figure 4: Adaptation of posterior for p to new information



Source: Author's calculations

Figure 5: Adaptation of posterior for m to new information



Source: Author's calculations

4 Application Results

After corroborating that our model produces good results in terms of both estimation and forecasting, we apply it to the data provided by Bass et al. (1994). The present authors collected data on total sales, price, and advertising for three different products: room air conditioners, color TVs, and clothes dryers. Both Bayus (1993) and Jiang et al. (2006) classify room air conditioners and clothes dryers as part of the same product group (home appliances). Bayus (1993) performs a discriminant analysis and finds that these products are also classified together empirically. This means that these products share similar markets and therefore create a perfect setting for our Bayesian method. In Table 8 we present the descriptive statistics for the variables and products previously mentioned. Times series data is available for thirteen periods for both products. Sales are measured in thousands of units, prices are measured as average price in dollars, and advertising is measured in millions of dollars. The actual variable used as advertising in the estimation process is the product of Simon's operationalization, which amounts to taking the actual value when the change in advertising is positive and the last value for which there was a positive change otherwise.

Table 8: Descriptive Statistics

Room Air Conditioners				
Variable	Mean	Std. Dev.	Min	Max
Sales	1109.077	652.568	96	1828
Cum. Sales	5674.615	5098.302	96	14418
Price	324.154	48.775	259	410
Advertising	6.506	5.694	0	16.785
Clothes Dryers				
Variable	Mean	Std. Dev.	Min	Max
Sales	965.692	465.416	106	1523
Cum. Sales	5315.308	4336.768	106	12554
Price	214	17.459	185	245
Advertising	4.058	2.673	0	8.657

Source: Author's calculations based on Bass et al. (1994)

Our estimation procedure is as follows: first, we estimate the parameters of the GBM for clothes dryers via nonlinear least squares. Then, as in a kind of empirical Bayes setting, we use the resulting estimates as priors for our Bayesian estimation of the GBM for room air conditioners. This approach allows us to avoid using the same information twice, which is a criticism of empirical Bayes methods. Throughout both exercises we assume that the short-run market is fixed[3]. The initial values for the NLS and Metropolis–Hastings algorithms are randomly drawn from uniform distributions that fulfill the parameter restrictions. As before, we follow Fúquene et al. (2014) and take $a_0 = b_0 = 1$ as the hyperparameters for the Scaled-Beta2 prior, and set $\kappa = 2000$. We set the location vector for the multivariate t-prior equal to the point estimates of the NLS procedure from clothes dryers using the full sample; the scale matrix equals the asymptotic covariance matrix, and there are three degrees of freedom. Given that the sales peak for room air conditioners is around $t = 8$, we will estimate the model at periods $t = 0, 3, 7, 8, 9, 10, 13$. Table 9 presents the estimation results for both the Bayesian and NLS techniques. We observe that the signs of the estimates are consistent with the theory. In particular, price increases have a negative effect on sales, while advertisement increases have a positive effect. As more information is included, the estimation of all parameters is more accurate and the standard error tends to decrease. We see that it is not possible to obtain estimates before the sales peak using the frequentist approach, due to problems of convergence in the optimization process. All of the coefficients for the NLS estimation are statistically significant (except for β_1 at $t = 13$) and all of the chains converge to a stationary distribution with desirable properties according to several convergence diagnostics (available upon request).

Table 10 presents the mean squared and mean absolute forecast errors for our Bayesian method and the NLS algorithm. We observe that after $t = 3$, the errors are generally decreasing with t. Bayesian prediction errors dominate, particularly around the sales peak. By the time all of the data is used in the forecasting process, the NLS is generally better than the Bayesian approach.

[3] m is 18475 for room air conditioners and 15716 for clothes dyers.

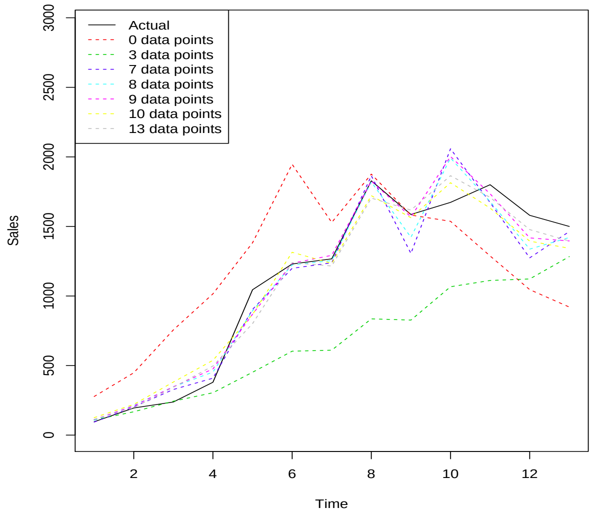
Table 9: Bayesian and NLS estimation for room air conditioners

	$t=0$	$t=3$	$t=7$	$t=8$	$t=9$	$t=10$	$t=13$
				Bayesian estimation			
p	0.013	0.005	0.004	0.005	0.005	0.006	0.005
	(0.090)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
q	0.307	0.253	0.320	0.332	0.355	0.336	0.354
	(0.249)	(0.154)	(0.040)	(0.038)	(0.019)	(0.021)	(0.016)
β_1	-0.853	-1.125	-3.467	-2.543	-1.814	-1.271	-1.079
	(1.277)	(1.165)	(1.908)	(1.327)	(0.540)	(0.539)	(0.467)
β_2	0.627	0.408	0.723	0.616	0.510	0.593	0.531
	(0.823)	(0.500)	(0.267)	(0.221)	(0.177)	(0.219)	(0.210)
				NLS estimation			
p	-	-	-	0.004	0.005	0.005	0.005
	-	-	-	(0.001)	(0.001)	(0.002)	(0.002)
q	-	-	-	0.282	0.355	0.332	0.354
	-	-	-	(0.036)	(0.020)	(0.024)	(0.018)
β_1	-	-	-	-5.026	-2.203	-1.606	-1.236
	-	-	-	(1.788)	(0.607)	(0.779)	(0.637)
β_2	-	-	-	0.920	0.535	0.654	0.556
	-	-	-	(0.306)	(0.194)	(0.279)	(0.256)

Standard errors in parenthesis.
"-" estimation was not possible.
Source: Author's calculations

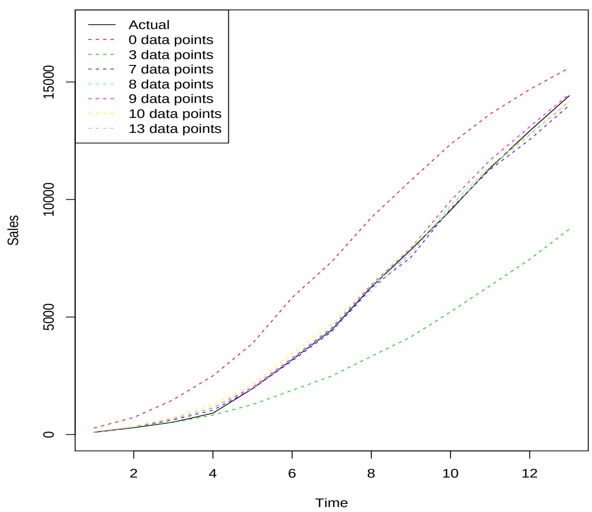
One of the most important results for our approach is the fact that it allows forecasts even when there are little to no data points. Figures 6 and 7 show the total and cumulative forecasts, respectively, with varying amounts of information. As is to be expected, more data points tend to improve the sales forecast; it appears that the cumulative forecast tends to be more accurate than the noncumulative forecast. Lastly, Figure 8 presents the one step-ahead forecast for cumulative sales. When most useful, at the introduction of a new product when there is no historical data, the one step-ahead forecast provides a fair approximation of the actual cumulative sales. As more periods pass, the forecast appears to adapt well to the shape of the distribution for cumulative room air conditioner sales.

Figure 6: Forecast of total sales with different amounts of data



Source: Author's calculations

Figure 7: Forecast of cumulative sales with different amounts of data



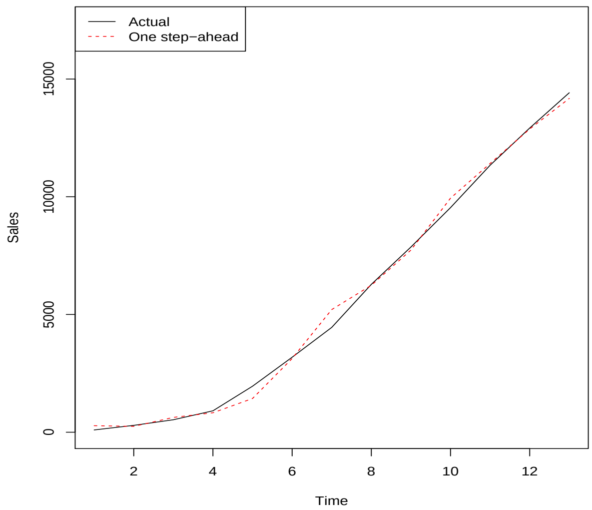
Source: Author's calculations

Table 10: Mean Squared Error and Mean Absolute Error with different amounts of information

	$t=0$	$t=3$	$t=7$	$t=8$	$t=9$	$t=10$	$t=13$
				Noncumulative Sales			
				Bayesian estimation			
MSE	182300	295481	28468	19182	15664	15828	13096
MAE	363	440	116	101	86	110	92
				NLS estimation			
MSE	-	-	-	53068	19052	16196	12948
MAE	-	-	-	151	94	111	89
				Cumulative Sales			
				Bayesian estimation			
MSE	4455177	10332902	34141	17506	31978	38385	15817
MAE	1892	2399	125	108	137	172	95
				NLS estimation			
MSE	-	-	-	117386	41728	34126	12423
MAE	-	-	-	225	149	161	85

Source: Author's calculations

Figure 8: One step-ahead forecast of cumulative sales



Source: Author's calculations

5 Concluding Remarks

In this study we introduced a particular Bayesian estimation process for the Generalized Bass Model for sales forecasting. We found, using several simulation exercises, that this process generates rapidly adapting forecasts with desirable properties. These forecasts accurately predict the total and one-step ahead sales information for a single product. As shown in the application to room air conditioners, if there is sufficient data on similar products or homogeneous markets, this information can be used to infer future adoption values even when no historical data on the product of interest is available. The implications for managers and decision makers are then substantial. Marketing planners could collect data on sales and other variables for products similar to the one of direct interest and have a systematical way of combining this information to obtain a reliable forecast for the future. Since marketing mix variables can be included in the analysis, managers could run simulations on different price and advertisement plans that would result in optimal adoption rates. This research is also of methodological interest since it explores the adaptation of posterior distributions to data availability and the effect of a Scaled-Beta2 prior distribution for the process variance that has not been widely used in the literature but that shows remarkable properties.

References

Bass, F. M. (1969). A New Product Growth for Model Consumer Durables. *Management Science*, 15(5):215–227.

Bass, F. M. (2004). Comments on "A New Product Growth for Model Consumer Durables The Bass Model". *Management Science*, 50(12 Supplement):1833–1840.

Bass, F. M., Krishnan, T. V., and Jain, D. C. (1994). Why the Bass Model fits without Decision Variables. *Marketing Science*, 13(3):203–223.

Bayus, B. L. (1993). High-Definition Television: Assessing Demand Forecasts for a Next Generation Consumer Durable. *Management Science*, 39(11):1319–1333.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Berlin: Springer-Verlag.

Bridges, E., Coughlan, A. T., and Kalish, S. (1991). New Technology Adoption in an Innovative Marketplace: Micro- and Macro-Level Decision Making Models. *International Journal of Forecasting*, 7(3):257–270.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Chandrasekaran, D. and Tellis, G. J. (2007). A Critical Review of Marketing Research on Diffusion of New Products. In Malhotra, N. K., editor, *Review of Marketing Research*, pages 39–80. Bingley, England: Emerald.

Dalla Valle, A. and Furlan, C. (2011). Forecasting accuracy of wind power technology diffusion models across countries. *International Journal of Forecasting*, 27(2):592–601.

Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika*, 16(1):27–31.

Fox, P. A., Hall, A. P., and Schryer, N. L. (1978). The PORT Mathematical Subroutine Library. *ACM Transactions on Mathematical Software*, 4(2):104–126.

Fúquene, J., Pérez, M.-E., and Pericchi, L. R. (2014). An Alternative to the Inverted Gamma for the Variances to Modelling Outliers and Structural Breaks in Dynamic Models. *Brazilian Journal of Probability and Statistics*, 28(2):288–299.

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis*, 1(3):515–534.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2014). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-0, http://CRAN.R-project.org/package=mvtnorm.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, pages 169–193. Oxford University Press.

Heidelberger, P. and Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31(6):1109–1144.

Hopp, W. J. (2004). Ten Most Influential Papers of Management Science's First Fifty Years. *Management Science*, 50(12 Supplement):1763–1763.

Jiang, Z., Bass, F. M., and Bass, P. I. (2006). Virtual Bass Model and the Left-Hand Data-Truncation Bias in Diffusion of Innovation Studies. *International Journal of Research in Marketing*, 23(1):93–106.

Kelsall, J. E. and Wakefield, J. C. (1999). Discussion of 'Bayesian Models for Spatially Correlated Disease and Exposure Data' by Best et al. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, page 151. Oxford University Press.

Lee, J., Boatwright, P., and Kamakura, W. A. (2003). A Bayesian Model for Prelaunch Sales Forecasting of Recorded Music. *Management Science*, 49(2):179–196.

Lenk, P. J. and Rao, A. G. (1990). New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures. *Marketing Science*, 9(1):42–53.

Lilien, G. L., Rao, A. G., and Kalish, S. (1981). Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment. *Management Science*, 27(5):493–506.

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

Mahajan, V., Muller, E., and Bass, F. M. (1990). New Product Diffusion Models in Marketing: A Review and Directions for Research. *The Journal of Marketing*, 54(1):1–26.

Meade, N. and Islam, T. (2006). Modelling and Forecasting the Diffusion of Innovations—A 25-year Review. *International Journal of Forecasting*, 22(3):519–545.

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.

Parker, P. M. (1994). Aggregate Diffusion Forecasting Models in Marketing: A Critical Review. *International Journal of Forecasting*, 10(2):353–380.

Peres, R., Muller, E., and Mahajan, V. (2010). Innovation Diffusion and New Product Growth Models: A Critical Review and Research Directions. *International Journal of Research in Marketing*, 27(2):91–106.

Plummer, M. (2015). *rjags: Bayesian Graphical Models using MCMC*. R package version 3-15.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Putsis, Jr., W. P. and Srinivasan, V. (2000). Estimation Techniques for Macro Diffusion Models. In Mahajan, V., Muller, E., and Wind, Y., editors, *New Product Diffusion Models*. Dordrecth, the Netherlands: Kluwer.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raftery, A. E., Lewis, S., et al. (1992). How many Iterations in the Gibbs Sampler? *Bayesian Statistics*, 4(2):763–773.

Rao, K. U. and Kishore, V. (2010). A Review of Technology Diffusion Models with Special Reference to Renewable Energy Technologies. *Renewable and Sustainable Energy Reviews*, 14(3):1070–1078.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Berlin: Springer-Verlag.

Rogers, E. M. (1962). *Diffusion of Innovations*. New York: The Free Press.

Rossi, P. E. and Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, 22(3):304–328.

Satoh, D. (2001). A Discrete Bass Model and its Parameter Estimation. *Journal of the Operations Research Society of Japan*, 44(1):1–18.

Schmittlein, D. C. and Mahajan, V. (1982). Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance. *Marketing Science*, 1(1):57–78.

Sekhon, J. S. and Mebane, Jr., W. R. (1998). Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models. *Political Analysis*, 7:189–213.

Simon, H. (1982). ADPULS: An Advertising Model with Wearout and Pulsation. *Journal of Marketing Research*, 19(3):352–363.

Skiadas, C. H. and Giovanis, A. N. (1997). A Stochastic Bass Innovation Diffusion Model for Studying the Growth of Electricity Consumption in Greece. *Applied Stochastic Models and Data Analysis*, 13(2):85–101.

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994, 2003). *BUGS: Bayesian inference Using Gibbs Sampling*. Cambridge: MRC Biostatistics Unit. www.mrc-bsu.cam.ac.uk/bugs/.

Srinivasan, V. and Mason, C. H. (1986). Nonlinear Least Squares Estimation of New Product Diffusion Models. *Marketing Science*, 5(2):169–178.

Su, Y.-S. and Yajima, M. (2015). *R2jags: Using R to Run 'JAGS'*. R package version 0.5-7.

Sultan, F., Farley, J. U., and Lehmann, D. R. (1990). A Meta-Analysis of Applications of Diffusion Models. *Journal of Marketing Research*, 27(1):70–77.

Trajtenberg, M. and Yitzhaki, S. (1989). The Diffusion of Innovations: A Methodological Reappraisal. *Journal of Business & Economic Statistics*, 7(1):35–47.

Van den Bulte, C. and Lilien, G. L. (1997). Bias and Systematic Change in the Parameter Estimates of Macro-Level Diffusion Models. *Marketing Science*, 16(4):338–353.

van Everdingen, Y. M., Aghina, W. B., and Fok, D. (2005). Forecasting Cross-Population Innovation Diffusion: A Bayesian Approach. *International Journal of Research in Marketing*, 22(3):293–308.

Xie, J., Song, X. M., Sirbu, M., and Wang, Q. (1997). Kalman Filter Estimation of New Product Diffusion Models. *Journal of Marketing Research*, 34(3):378–393.

6 Appendix

6.1 Bass Model

The main mathematical notions necessary to understand the Bass model come from survival analysis. Bass (1969) defines the probability of buying a product given that no purchase has been made as a linear function of the number $Y(t)$ of buyers until time t, the ultimate market potential m, and parameters of innovation p and imitation q, i.e.,

$$H(t) = p + \frac{q}{m} Y(t) \tag{10}$$

However, note that, by definition, $H(t)$ is the hazard function that expresses the likelihood of buying conditional on not having bought before. In that sense, Bass focuses only on initial purchases. Therefore, we can express $H(t)$ as

$$H(t) = \frac{f(t)}{1 - F(t)} = p + \frac{q}{m} Y(t)$$

In this context, $f(t)$ represents the likelihood that the product is purchased at any given time, $F(t)$ is the probability of buying the product before t, and $1 - F(t)$ is the probability that a purchase has not been made before time t. Denoting by $S(t)$ the total sales at time t, then

$$S(t) = mf(t)$$

That is, sales at any given moment are the amount of market potential that is reached. Total sales $Y(t)$ are then

$$Y(t) = \int_0^t S(\tau) d\tau = m \int_0^t f(\tau) d\tau = mF(t)$$

The original model is therefore reduced to

$$\frac{f(t)}{1 - F(t)} = p + qF(t) \tag{11}$$

This is a non-linear differential equation with an initial value of $F(0) = 0$; the cumulative likelihood of a purchase at $t = 0$ is equal to 0. Reordering Eq. (11)

$$f(t) = [p + qF(t)][1 - F(t)]$$
$$f(t) = p + (q - p)F(t) - q[F(t)]^2$$

In mathematics, this is known as a Riccati equation with constant coefficients. This simplification allows us to solve the equation exactly using the separation of variables method. Thus,

$$\frac{dF}{(p+(q-p)F-qF^2)} = dt$$

If we integrate both sides, we can complete the square in the denominator:

$$\begin{aligned}
\int dt &= \int \frac{dF}{-q\left(F^2+\frac{p-q}{q}F-\frac{p}{q}\right)} \\
&= \frac{-1}{q}\int \frac{dF}{F^2+\frac{p-q}{q}F-\frac{p}{q}+\left(\frac{p-q}{2q}\right)^2-\left(\frac{p-q}{2q}\right)^2} \\
&= \frac{-1}{q}\int \frac{dF}{\left(F+\frac{p-q}{2q}\right)^2-\left(\frac{p}{q}+\frac{p^2-2pq+4q^2}{4q^2}\right)} \\
&= \frac{-1}{q}\int \frac{dF}{\left(F+\frac{p-q}{2q}\right)^2-\frac{1}{4}\left(\frac{p^2}{q^2}+\frac{2p}{q}+1\right)} \\
&= \frac{-1}{q}\int \frac{dF}{\left(F+\frac{p-q}{2q}\right)^2-\frac{1}{4}\left(\frac{p}{q}+1\right)^2}
\end{aligned}$$

If we substitute $u = F + (p-q)/2q$ and $du = dF$, then

$$\int dt = \frac{-1}{q}\int \frac{du}{u^2-\frac{1}{4}\left(\frac{p+q}{q}\right)^2}$$

Factoring out $-\frac{1}{4}\left(\frac{p+q}{q}\right)^2$ from the denominator,

$$\int dt = \frac{1}{q}\int \frac{4du}{\left(\frac{p+q}{q}\right)^2\left(1-\frac{4q^2u^2}{(p+q)^2}\right)}$$

Substituting now for $s = 2qu/(p+q)$ and $ds = 2qdu/(p+q)$,

$$\begin{aligned}\int dt &= \frac{q}{(p+q)^2}\int \frac{4(p+q)ds/2q}{1-s^2}\\ &= \frac{2}{p+q}\int \frac{ds}{1-s^2}\end{aligned}$$

To calculate this last integral, we decompose $1/(1-s^2) = 1/[(1-s)(1+s)]$ using partial fractions:

$$\begin{aligned}\frac{1}{(1-s)(1+s)} &= \frac{A}{1-s} + \frac{B}{1+s}\\ 1 &= A(1+s) + B(1-s)\\ 1 &= (A+B) + (A-B)s\end{aligned}$$

which works by replacing $A = B = 1/2$. Then, we can rewrite this and find the integral as

$$\begin{aligned}\int dt &= \frac{2}{p+q}\int \left[\frac{1}{2(1-s)} + \frac{1}{2(1+s)}\right] ds\\ &= \frac{1}{p+q}\left[\int \frac{ds}{1-s} + \int \frac{ds}{1+s}\right]\\ &= \frac{1}{p+q}\left(\ln|1+s| - \ln|1-s|\right)\\ &= \frac{1}{p+q}\ln\left|\frac{1+s}{1-s}\right|\end{aligned}$$

Substituting back for u and s we obtain

$$\begin{aligned}t + C &= \frac{1}{p+q}\ln\left|\frac{1+s}{1-s}\right|\\ &= \frac{1}{p+q}\ln\left|\frac{1+2qu/(p+q)}{1-2qu/(p+q)}\right|\\ &= \frac{1}{p+q}\ln\left|\frac{1+2q[F+(p-q)/2q]/(p+q)}{1-2q[F+(p-q)/2q]/(p+q)}\right|\\ &= \frac{1}{p+q}\ln\left|\frac{p+q+2qF+p-q}{p+q-2qF-p+q}\right|\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{p+q}\ln\left|\frac{2(p+qF)}{2q(1-F)}\right| \\
&= \frac{\ln(p+qF) - \ln(q[1-F])}{p+q}
\end{aligned}
$$

The last equality follows from the fact that the absolute value function is multiplicative and that $p > 0, q > 0, 0 \le F(t) \le 1$. Solving for $F(t)$ leads to

$$
\begin{aligned}
F(t) &= \frac{qe^{(t+C)(p+q)} - p}{q(e^{(t+C)(p+q)} + 1)} \\
&= \frac{q - pe^{-(t+C)(p+q)}}{q(1 + e^{-(t+C)(p+q)})}
\end{aligned}
$$

The fact that $F(0) = 0$ allows us to evaluate the constant of integration:

$$
\begin{aligned}
0 &= q - pe^{-C(p+q)} \\
-C(p+q) &= \ln\left(\frac{q}{p}\right) \\
C &= \frac{\ln(p/q)}{p+q}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
F(t) &= \frac{q - pe^{-t(p+q)}e^{-C(p+q)}}{q(1 + e^{-t(p+q)}e^{-C(p+q)})} \\
&= \frac{q - qe^{-t(p+q)}}{q[1 + (q/p)e^{-t(p+q)}]} \\
&= \frac{1 - e^{-t(p+q)}}{1 + (q/p)e^{-t(p+q)}}
\end{aligned}
$$

The expressions for $f(t)$ and $S(t)$ are

$$
\begin{aligned}
f(t) &= \frac{\left[(p+q)e^{-t(p+q)}\right]\left[1 + (q/p)e^{-t(p+q)}\right] - (1 - e^{-t(p+q)})\left[-(q/p)(p+q)e^{-t(p+q)}\right]}{\left[1 + (q/p)e^{-t(p+q)}\right]^2} \\
f(t) &= \frac{[(p+q)^2/p]e^{-t(p+q)}}{\left[1 + (q/p)e^{-t(p+q)}\right]^2} \qquad (12)
\end{aligned}
$$

$$S(t) = \frac{[m(p+q)^2/p]e^{-t(p+q)}}{\left[1+(q/p)e^{-t(p+q)}\right]^2} \tag{13}$$

By maximizing $S(t)$, it is possible to find the moment at which the sales rate reaches its peak, i.e., by differentiating and setting the result equal to 0:

$$\begin{aligned}
S'(t) &= \frac{2\left[1+(q/p)e^{-t(p+q)}\right][m(p+q)^3/p](q/p)e^{-2t(p+q)} - \left[1+(q/p)e^{-t(p+q)}\right]^2[m(p+q)^3/p]e^{-t(p+q)}}{\left[1+(q/p)e^{-t(p+q)}\right]^4} \\
&= \frac{\left[1+(q/p)e^{-t(p+q)}\right][m(p+q)^3/p]e^{-t(p+q)}\left[2(q/p)e^{-t(p+q)} - 1 - (q/p)e^{-t(p+q)}\right]}{\left[1+(q/p)e^{-t(p+q)}\right]^4} \\
&= \frac{[m(p+q)^3/p]e^{-t(p+q)}\left[(q/p)e^{-t(p+q)} - 1\right]}{\left[1+(q/p)e^{-t(p+q)}\right]^3}
\end{aligned}$$

Setting this equal to 0 and solving for t,

$$\begin{aligned}
S'(t) &= 0 \\
\left[(q/p)e^{-t(p+q)} - 1\right] &= 0 \\
t^* &= -\frac{\ln(q/p)}{p+q}
\end{aligned} \tag{14}$$

For the Generalized Bass Model (GBM), the only necessary change is to add a multiplicative term on the right side of Eq. (11). This is referred to as the market effort, $x(t)$, and is assumed to be

$$x(t) = 1 + \beta_1 \frac{P(t) - P(t-1)}{P(t-1)} + \beta_2 \max\left\{0, \frac{A(t) - A(t-1)}{A(t-1)}\right\}$$

Transforming the market effort into continuous time and integrating between 0 and t, one obtains the cumulative market effort, $X(t)$:

$$X(t) = t + \beta_1 \ln\left[\frac{P(t)}{P(0)}\right] + \beta_2 \ln\left[\frac{A(t)}{A(0)}\right]$$

If we assume $X(0) = 0$, then all of the equations remain similar if one replaces t with $X(t)$.

6.2 The Scaled-Beta 2 Distribution

The beta distribution of the second kind, or *Beta2* distribution, can be obtained as the odds ratio from a beta distribution with parameters α and β. Using scaled odds, the *Scaled-Beta2* distribution can be obtained.

Both proofs use the univariate change of variables theorem. Let f_X be the probability density function of a random variable X and let $Y = g(X)$. The theorem states that the distribution of Y, denoted by f_Y, can be found using

$$f_Y(Y) = f_X[g^{-1}(Y)] \left| \frac{d}{dY} g^{-1}(Y) \right|$$

where $X = g^{-1}(Y)$. Applying this to this problem, let $X \sim \text{Beta}(\alpha, \beta)$ and $Y = g(X) = \frac{X}{1-X}$. This means

$$\begin{aligned} Y &= \frac{X}{1-X} \\ Y - XY &= X \\ Y &= X(Y+1) \\ X = g^{-1}(Y) &= \frac{Y}{Y+1} \end{aligned}$$

Recall the probability density function for the Beta distribution:

$$f_X(X) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} X^{\alpha-1}(1-X)^{\beta-1}$$

Then

$$\begin{aligned} f_Y(Y) &= f_X\left(\frac{Y}{Y+1}\right) \left| \frac{d[Y/(Y+1)]}{dY} \right| \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{Y}{Y+1}\right)^{\alpha-1} \left(1 - \frac{Y}{Y+1}\right)^{\beta-1} \left| \frac{Y+1-Y}{(Y+1)^2} \right| \end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{Y}{Y+1}\right)^{\alpha-1} \left(\frac{1}{Y+1}\right)^{\beta-1} \left|\frac{1}{(Y+1)^2}\right| \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{Y^{\alpha-1}}{(Y+1)^{\alpha+\beta}}
\end{aligned}
$$

which is the probability density function for a Beta2(α, β) distribution. If we scale the odds ratio by a constant κ, then $Y = \kappa \frac{X}{1-X}$ and

$$
\begin{aligned}
Y &= \kappa \frac{X}{1-X} \\
Y - XY &= \kappa X \\
Y &= X(Y+\kappa) \\
X = g^{-1}(Y) &= \frac{Y}{Y+\kappa}
\end{aligned}
$$

This implies

$$
\begin{aligned}
f_Y(Y) &= f_X\left(\frac{Y}{Y+\kappa}\right) \left|\frac{d[Y/(Y+\kappa)]}{dY}\right| \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{Y}{Y+\kappa}\right)^{\alpha-1} \left(1 - \frac{Y}{Y+\kappa}\right)^{\beta-1} \left|\frac{Y+\kappa-Y}{(Y+\kappa)^2}\right| \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{Y}{Y+\kappa}\right)^{\alpha-1} \left(\frac{\kappa}{Y+\kappa}\right)^{\beta-1} \left|\frac{\kappa}{(Y+\kappa)^2}\right| \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{Y^{\alpha-1}\kappa^{\beta}}{(Y+\kappa)^{\alpha+\beta}} \frac{\kappa/\kappa^{\alpha+\beta}}{\kappa/\kappa^{\alpha+\beta}} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{Y^{\alpha-1}\kappa^{-(\alpha-1)}}{[(Y+\kappa)/\kappa]^{\alpha+\beta}} \frac{1}{\kappa} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{\kappa} \frac{(Y/\kappa)^{\alpha-1}}{(Y/\kappa+1)^{\alpha+\beta}}
\end{aligned}
$$

This is the probability density function for the Scaled-Beta2 distribution with parameters κ, α and β.