© Journal of Contemporary Issues in Business Research
ISSN 2305-8277 (Online), 2012, Vol. 1, No. 3, 69-77.

A GENERALIZED ADDITIVE LOGIT MODEL OF BRAND CHOICE*

SUNIL K SAPRA†
California State University Los Angeles, United States

ABSTRACT

The paper presents an econometric application of generalized additive Logit regression models (GALRMs) for brand choice. Our semi-parametric models are flexible and robust extensions of the Logit model. The GALRMs are fit to binary response data by maximizing a penalized log likelihood or a penalized log partial-likelihood. The GAMs allow us to build a regression surface as a sum of lower-dimensional nonparametric terms circumventing the curse of dimensionality: the slow convergence of an estimator to the true value in high dimensions. Four GALRMs are compared with a Logit model for brand choice and the best model is selected using various model selection criteria.

Keywords: Logit Model; Generalized Additive Model; Interactions; Backfitting Algorithm; Penalized Regression Splines.

INTRODUCTION

Logit models are pervasive in business and economics. The most common applications of Logit models have been to the analysis of brand choice data in marketing (Baltas 1997 and Guadagni and Little 1983) and transportation choice data in economics (Greene 2008 and Manski and McFadden 1983). Logit models belong to the class of generalized linear models, which relax the assumption that the response is normally distributed by allowing it to follow any distribution from the exponential family (for example, normal, Poisson, binomial, gamma etc.). Inference for GLMs is based on likelihood theory. McCullagh and Nelder (1989) provide an authoritative account of GLMs and Cameron and Trivedi (2005) and Greene (2008) provide econometric applications.

In recent years, semi-parametric extensions of linear regressions have been employed in economics and business. Nevertheless, there have been relatively few applications of semi-parametric extensions of generalized linear models. This paper is an attempt in this direction. We present an econometric application of a generalized additive model (GAM) to binary choice data, which is a semi-parametric extension of GLMs. A GAM is a semi-parametric GLM in which part of the linear predictor is specified in terms of a sum of unknown smooth functions of explanatory variables. Generalized additive models (GAMs) are a powerful generalization of linear, logistic, and Poisson regression models. GAMs are very flexible, and can provide an excellent fit in the presence of nonlinear relationships. GAMs and GLMs can be applied in similar situations, but they serve different analytic purposes. GLMs emphasize estimation and inference for the parameters of the model, while GAMs focus on exploring

* The views or opinions expressed in this manuscript are those of the author(s) and do not necessarily reflect the position, views or opinions of the editor(s), the editorial board or the publisher.
† Corresponding author Email: ssapra@exchange.calstatela.edu

data nonparametrically. The GAM approach offers more flexibility in model form than the GLM approach does. These models estimate an additive approximation to the multivariate link function. The key benefit of this approach is that each of the individual additive terms is estimated using a univariate smoother instead of a multivariate smoother for a high-dimensional non-parametric term circumventing the curse of dimensionality: the slow convergence of an estimator to the true value in high dimensions. The GAM formulation of the Logit regression model allows us to build a regression surface as a sum of lower-dimensional nonparametric terms. The two main approaches to estimation of GAMs are backfitting with local scoring algorithm (Hastie and Tibshirani (1990)) and penalized regression splines (Wood (2006)). Backfitting has the advantage that it can be used with any scatterplot smoother while the latter method has the advantage that estimation of the smoothing parameter using generalized cross validation is integrated into estimation.

This paper presents an econometric application of the GAM extension of the Logit model and demonstrates that it can overcome a serious weakness of the Logit model in failing to identify the nonlinearities in the link function. The paper is organized as follows. Section 2 introduces the generalized additive Logit regression model (GALRM). Section 3 presents the penalized regression method for the estimation of GALRMs. Section 4 presents an econometric application of GALRM to Coke data on choice between Coke and Pepsi. Section 5 provides some concluding remarks.

GENERALIZED ADDITIVE MODELS

Generalized additive models are nonparametric generalized linear models. Generalized additive models extend traditional linear models in another way, namely by allowing for a link between the nonlinear predictor f(x1, ... ,xp) and the expected value of y. This amounts to allowing for an alternative distribution for the underlying random variation besides just the normal distribution. While Gaussian models can be used in many statistical applications, for several types of problems they are not appropriate. For example, the normal distribution may not be adequate for modeling discrete responses such as counts, or bounded responses such as proportions.

Generalized additive models consist of a random component, an additive component, and a link function relating these two components. The response y, the random component, is assumed to have a density in the exponential family.

$$f_Y(y;\theta,\phi) = \left\{ \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y.\phi) \right\}$$

Where ϑ is called the natural parameter and Φ is the scale parameter. The normal, binomial, and Poisson distributions are all in this family. Unlike the generalized linear models, the mean $\mu = E(y | x1, x2, \ldots, xp)$ is not linked to the linear predictor $\sum xij\beta j$, but to the nonlinear nonparametric predictor.

$$\eta = g(\mu) = \alpha + \sum_{j=1}^{p} s_j(x_j),$$

Where s1(·), ... , sp(·) are smooth nonparametric functions defines the additive component. Finally, the relationship between the mean μ of the response variable and η is defined by a link function $g(\mu) = \eta$. The most commonly used link function is the canonical link, for which $\eta = \theta$.

A combination of backfitting and local scoring algorithms are used in the actual fitting of the model. In order to fit GAMs to the data, following Wood (2006), we use basis expansions of smooth functions and penalized likelihood maximization for model estimation in which wiggly models are penalized more heavily than smooth models in a controllable manner, and degree of smoothness is chosen based on cross validation, or AIC or Mallows' criterion.

The generalized additive models are fit to count data or binary response data by maximizing a penalized log likelihood or a penalized log partial-likelihood. To maximize it, the backfitting procedure is used in conjunction with a maximum likelihood or maximum partial likelihood algorithm (Hastie and Tibshirani (1990) and Wood (2006)). The Newton-Raphson method for maximizing log-likelihoods in these models can be presented in an IRLS (iteratively reweighted least squares) form. It involves a repeated weighted linear regression of a constructed response variable on the covariates: each regression yields a new value of the parameter estimates which give a new constructed variable, and the process is iterated. In the generalized additive model, the weighted linear regression is simply replaced by a weighted backfitting algorithm (Hastie and Tibshirani (1990)).

ESTIMATION OF GENERALIZED ADDITIVE LOGIT REGRESSION MODEL USING PENALIZED REGRESSION METHOD

Algorithm for Penalized Iteratively Re-weighted Least Squares (PIRLS) (Wood (2006)

The GALRMs are fit to binary response data by maximizing a penalized log likelihood or a penalized log partial-likelihood. The following algorithm is used to implement these methods. The R-package mgcv (Wood (2012)) was used for computations.

Step 1: Initialize $\hat{\alpha} = g(1/N\sum_{i=1}^{N} y_i), s_j^0 = 0, j = 1,2,..., p.$

Step 2: Construct an adjusted dependent variable z_{ij} as

$$z_i = \eta_i^0 - (y_i - \mu_i^0)(\partial\eta_i / \partial\mu_i)_0$$

$$\eta_i^0 = g(\mu_i^0) = \alpha^0 + \sum_{j=1}^{p} s_j^0(x_{ij}) \text{ and } \mu_i^0 = g^{-1}(\eta_i^0).$$

Step 3: Compute weights

$$w_i = (\partial\mu_i / \partial\eta_i)_0 (V_i^0)^{-1},$$

where V_i^0 is the variance of y at μ_i^0,

Step 4: *Penalized Spline Regression*

Minimize $\left\|\sqrt{W}(z - X\beta)\right\|^2 + \lambda\beta' S\beta$ with respect to β, where X is the matrix of data on basis functions used to represent the regression function, W is a diagonal matrix with i-th diagonal element w_i, S is a matrix of known coefficients in the penalty function $\beta' S\beta$ and λ is a smoothing parameter. Compute s_j^1, η^1, and μ_i^1, the second stage estimates of s_j, η, and μ_i.

Step 5: Repeat steps 2-4 replacing η^0 by η^1 until the difference between two successive values of η is less than a small prespecified number and convergence is obtained.

The Generalized Additive Logit Model

The generalized additive Logit model assumes that

$$g(\mu_i) = \text{logit}(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right) = \alpha + \sum_{j=1}^{p} s_j(x_{ij}),$$

where $\mu_i = p_i = E(y_i = 1|x_i) = \exp\{\alpha + \sum_{j=1}^{p} s_j(x_{ij})\} / \exp\{1 + \alpha + \sum_{j=1}^{p} s_j(x_{ij})\}$.

The adjusted dependent variable z and the weights w used in the algorithm above are

$z_i = \eta_i + (y_i - p_i) / p_i(1 - p_i),$

$w_i = p_i(1 - p_i)$

where $p_i = g^{-1}(\eta_i), \eta_i = \alpha + \sum_{j=1}^{p} s_j(x_{ij})$

The functions $s_1, s_2, \ldots, s_p$ are estimated by an algorithm like the one described earlier.

AN EMPIRICAL APPLICATION OF GAM LOGIT MODEL TO BRAND CHOICE

GAM Logit Applied to Coke Data

The dataset is from ERIM public data base, James M. Kilts Center, University of Chicago Booth School of Business and consists of data on 1140 individuals who purchased Coke or Pepsi.

Data Description

The dependent variable COKE is defined as follows:

COKE =1 if Coke is chosen,
= 0 if Pepsi is chosen
PR_PEPSI = Price of 2 liter bottle of Pepsi
PR_COKE = Price of 2 liter bottle of Coke
DISP_PEPSI = 1 if Pepsi is displayed at time of purchase, otherwise = 0
DISP_COKE = 1 if Coke is displayed at time of purchase, otherwise = 0
PRATIO = Price of Coke relative to price of Pepsi

TABLE 1
Summary Statistics

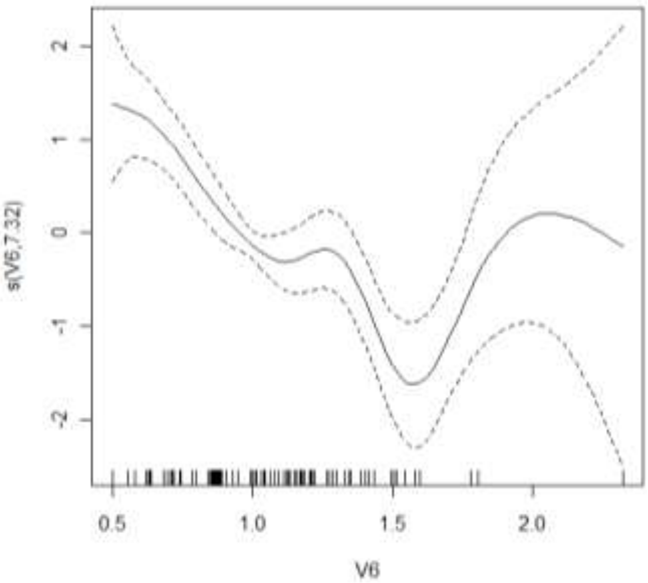
Variable	Obs.	Mean	Std. Dev.	Min	Max
COKE	1140	0.4473684	0.4974404	0	1
PR_PEPSI	1140	1.202719	0.3007257	.68	1.79
PR_COKE	1140	1.190088	0.2999157	.68	1.79
DISP_PEPSI	1140	0.3640351	0.4813697	0	1
DISP_COKE	1140	0.3789474	0.4853379	0	1
PRATIO	1140	1.027249	0.286608	0.497207	2.324675

Source. ERIM public data base, James M. Kilts Center, University of Chicago Booth School of Business

Models

The following models were fit to the data. Model 1 is a Logit model. Given the nonlinearity of the Logit link function in PRATIO as displayed in the partial residual plot of PRATIO in Fig.1, Model 2, a Generalized Additive Logit model introduces a nonparametric smooth term s(PRATIO). The possibility of interaction between PRATIO and DISP_PEPSI and between PRATIO and DISP_COKE was investigated in models 3 and 4 respectively by introducing a nonparametric interaction term in each model. Finally, Model 5 is a GLM Logit model, which introduces parametric interactions between PRATIO and DISP_PEPSI as well as between PRATIO and DISP_COKE.

FIGURE 1
Partial Residual Plot of V6



$Model\ 1: Logit\ Regression\ Model$

$$\eta = g(\mu) = \beta_1 + \beta_4 DISP\_PEPSI + \beta_5 DISP\_COKE + \beta_6 PRATIO$$

$Model\ 2: Generalized\ Additive\ Logit\ Regression\ Models$

$$\eta = g(\mu) = \beta_1 + \beta_4 DISP\_PEPSI + \beta_5 DISP\_COKE + s(PRATIO)$$

$Model\ 3: Generalized\ Additive\ Logit\ Regression\ Models\ with\ a\ Nonparametric\ Interaction\ Term$

$$\eta = g(\mu) = \beta_1 + \beta_5 DISP\_COKE + s(PRATIO * DISP\_PEPSI)$$

$Model\ 4: Generalized\ Additive\ Logit\ Regression\ Models\ with\ a\ Nonparametric\ Interaction\ Term$

$$\eta = g(\mu) = \beta_1 + \beta_4 DISP\_PEPSI + s(PRATIO * DISP\_COKE)$$

$Model\ 5: Logit\ Regression\ Model\ with\ Interactions\ between\ Pratio\ and\ DISP\_PEPSI\ and\ between\ Pratio\ and\ DISP\_COKE$

$$\eta = g(\mu) = \beta_1 + \beta_4 DISP\_PEPSI + \beta_5 DISP\_COKE + \beta_6 PRATIO + \beta_7 PRATIO * DISP\_PEPSI + \beta_8 PRATIO * DISP\_COKE$$

TABLE 2

Model 1: GLM Logit: Coke Data

Variable	Coefficient Estimate	Std. Error	t-ratio	p-value
Intercept	1.9230	0.3258	5.902	3.59×10^{-9}***
DISP_PEPSI	-0.7310	0.1678	-4.356	1.33×10^{-5}***
DISP_COKE	0.3516	0.1585	2.218	0.0266*
PRATIO	-1.9957	0.3146	-6.344	2.23×10^{-10}***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 1567.7 on 1139 degrees of freedom
Residual deviance: 1418.9 on 1136 degrees of freedom
AIC = 1426.9
Number of Fisher Scoring iterations: 3

TABLE 3

Model 2: GAM Logit: Coke Data

Variable	Coefficient Estimate	Std. Error	t-ratio	p-value
Intercept	-0.14249	0.08903	-1.601	0.109477
DISP_PEPSI	-0.60371	0.17453	-3.459	0.000542***
DISP_COKE	0.32180	0.16174	1.990	0.046635*

Approximate significance of smooth terms:

Variable	Estimated df	Refined df	Chi-squared	p-value
s(PRATIO)	7.323	8.258	59.48	7.77×10^{-10}**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
R-sq.(adj) = 0.135 Deviance explained = 11.1%
UBRE score = 0.24083 Scale est. = 1 n = 1140
AIC = 1414.544
Null Deviance: 1567.721 on 1139 degrees of freedom
Residual Deviance: 1403.554 on 1133 degrees of freedom
Number of Local Scoring Iterations: 8

TABLE 4
Model 3: GAM Logit Model with a Smooth Nonparametric Function and Interaction between V4 and V6: Coke Data

Variable	Coefficient Estimate	Std. Error	t-ratio	p-value
Intercept	-0.2583	6.8042	-0.038	0.96971
DISP_PEPSI	-0.4628	18.6839	-0.025	0.98024
DISP_COKE	0.4903	0.1729	2.836	0.00457**

Approximate significance of smooth terms:

Variable	Estimated df	Refined df	Chi-squared	p-value
s(PRATIO)	1.001	1.001	48.98	2.59×10^{-12}***
s(DISP_PEPSI*PRATIO)	4.861	5.723	23.88	0.000436***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) = 0.142 Deviance explained = 11.5%
UBRE score = 0.23287 Scale est. = 1 n = 1140
AIC = 1405.476
Null Deviance: 1567.721 on 1139 degrees of freedom
Residual Deviance: 1390.645 on 1129 degrees of freedom
Number of Local Scoring Iterations: 12

TABLE 5
Model 4: GAM Logit Model with a Smooth Nonparametric Function and Interaction between V5 and V6: Coke Data

Variable	Coefficient Estimate	Std. Error	t-ratio	p-value
Intercept	0.4717	3.5389	0.133	0.8940
DISP_PEPSI	-0.5088	0.1995	-2.550	0.0108*
DISP_COKE	-1.3885	9.3359	-0.149	0.8818

Approximate significance of smooth terms:

Variable	Estimated df	Refined df	Chi-squared	p-value
s(PRATIO)	7.356	8.225	40.898	2.67×10^{-6}***
s(DISP_COKE*PRATIO)	4.899	5.837	9.158	0.154

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) = 0.142 Deviance explained = 12%
UBRE score = 0.23751 Scale est. = 1 n = 1140
AIC = 1410.765
Null Deviance: 1567.721 on 1139 degrees of freedom
Residual Deviance: 1396.239 on 1129 degrees of freedom
Number of Local Scoring Iterations: 9

TABLE 6
Model 5: GLM Logit with Interaction: Coke Data

Variable	Coefficient Estimate	Std. Error	t-ratio	p-value
Intercept	2.2697	0.4209	5.393	6.92×10^{-8}***
DISP_PEPSI	-2.8349	0.6992	-4.054	5.02×10^{-5}***
DISP_COKE	1.4708	0.6831	2.153	0.03130*
PRATIO	-2.4247	0.4166	-5.820	5.88×10^{-9}***
DISP_PEPSI*PRATIO	1.9946	0.6069	3.286	0.00101**
DISP_COKE*PRATIO	-1.1171	0.7247	-1.541	0.12321

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 1567.7 on 1139 degrees of freedom
Residual deviance: 1400.7 on 1134 degrees of freedom
AIC: 1412.7
Number of Fisher Scoring iterations: 4

Comparing the Models
The estimation and results are presented in tables 2 through 6. A comparison of models using the AIC is presented in Table 7.

TABLE 7
Models and the AICs

MODEL	AIC
1	1426.9
2	1414.544
3	1405.476
4	1410.765
5	1412.7

Model 3, a modified GAM, which allows for nonparametric interaction between a discrete variable DISP_PEPSI and a continuous variable PRATIO, has the lowest AIC and UBRE score among the five models studied and is therefore the best model. The Logit regression model (Model 1) has the highest AIC. This is not surprising since the Logit model misses the nonlinearity in PRATIO in the link function. Model 3 also has the lowest deviance of 1390.645 on 1129 degrees of freedom, while Model 1 has the highest deviance of 1418.9 on 1136 degrees of freedom. The statistical significance of PRATIO differs markedly between the Logit regression model and the various GAM Logit models employed here. The signs of the coefficient estimates are all expected in all of the models but Model 4. For instance, the sign of DISP_PEPSI is negative and the sign of DISP_COKE is positive across models 1 through 3 and model 5 indicating that the odds of choosing Coke over Pepsi increase if Pepsi is displayed at the time of purchase and decrease if Coke is displayed at the time of purchase. The analysis of deviance in Tables 2, 3, and 4 indicates significant nonlinear contribution from the variable PRATIO. The nonparametric interaction term in Model 3 for nonlinear interaction between DISP_PEPSI and PRATIO turns out to be highly significant indicating another nonlinearity in the link function missed by the Logit model. The parametric linear interaction term for interaction between DISP_PEPSI and PRATIO in Model 5 also turns out to be highly significant. Nevertheless, Table 5 indicates a weak interaction between nonlinear effects of PRATIO and the variable DISP_COKE. The high degree of nonlinearity in PRATIO is also seen in the partial residual smoothing plot of PRATIO in Figure 1. The dotted curves around the solid curve represent +-2 standard errors around the solid curve. The only surprising result is the negative sign of the variable DISP_COKE in Table 4.

CONCLUSIONS

The paper has studied four generalized additive Logit regression models (GALRMs) as alternatives to the Logit regression model. In all of the empirical applications, each of the GALRMs provides a much better fit than the Logit regression model as reflected in lower AICs and lower deviances. The econometric techniques used in the paper are widely applicable to the analysis of count, binary response and duration types of data occurring frequently in economics and business. Nevertheless, the GALRMs are not without drawbacks. The computational algorithms are complex and interpretations can be difficult. These models are useful mainly when simple models for the linear predictor provide an inadequate fit for the data.

REFERENCES

Baltas, G. (1997). Determinants of store brand choice: a behavioral analysis. *Journal of Product & Brand Management, 6* (5), 315 – 324.

Cameron, A. C., and Trivedi, P. (1998). *Microeconometrics.* Cambridge University Press.

Greene, W. H. (2008). *Econometric Analysis*, Pearson/Prentice Hall, New York.

Guadagni, P. M., and Little, John, D.C. (1983). A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science Summer Vol. 2* (3), 203-238.

Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

McCullagh, P., and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London.

Manski, C., and Daniel McFadden (1981). *Structural Analysis of Discrete Data and Econometric Applications*. MIT Press, Cambridge.

Wood, S. N. (2012), R-package mgcv. Download From http://cran.r-project.org/web/packages/mgcv/mgcv.pdf

Wood, S. N. (2006). Generalized Additive Models: an introduction with R. CRC, Boca Raton.