MARKETING SCIENCE
Vol. 4, No. 1, Winter 1985
Printed in U.S.A.

A BAYESIAN CROSS-VALIDATED LIKELIHOOD METHOD FOR COMPARING ALTERNATIVE SPECIFICATIONS OF QUANTITATIVE MODELS

ROLAND T. RUST AND DAVID C. SCHMITTLEIN
University of Texas at Austin
The Wharton School, University of Pennsylvania

There are many situations in marketing in which several alternative quantitative models may be built to model a particular marketing phenomenon or system. Few methods exist for comparing the fit of such models if the models are not nested, especially if their performance on each of several criteria is important.

This paper proposes a Bayesian cross-validated likelihood (BCVL) method for comparing quantitative models. It can be used when the models are either nested or nonnested, and is especially useful for nonnested models. A simulation based upon a typical marketing modeling situation shows the incremental benefit of using the BCVL method rather than existing techniques, and explores the circumstances under which BCVL works best. The applicability of the BCVL method is demonstrated using several typical marketing modeling situations.
(Model Comparison; Cross-Validation; Marketing Models)

1. Introduction

Decision areas in marketing are seldom so well understood that only one reasonable model may be constructed that describes the phenomena at hand. Rather, it is generally the case that two, three, or several such models may be constructed. Normatively, it is often argued that several alternative models *should* be constructed (Wilson 1979). This is the case in empirical studies where competing hypotheses concerning marketing phenomena are tested using the performance of models consistent with those hypotheses (Bass 1969). Several models are also commonly constructed in simulation studies. They may be compared on their ability to represent or approximate a "true" (and perhaps more complicated) process which is generating the data (Carmone and Green 1981). Thus an important problem encountered by builders of marketing models is that of selecting the best model from several reasonable ones.

For example, consider a typical situation in which a researcher is modelling the effectiveness of the marketing mix. Sales (Y) is assumed to be a function of advertising spots (X_1), number of sales representatives (X_2), and wholesale efficiency index (X_3). Let us suppose that the two sales prediction models below are the models under consideration:

$$\begin{aligned} &\text{Model A:} \quad \hat{Y} = A_0 + A_1X_1 + A_2X_2 + A_3X_3, \\ &\text{Model B:} \quad \hat{Y} = B_0 + B_1\ln(X_1) + B_2\ln(X_2) + B_3\ln(X_3). \end{aligned} \tag{1}$$

0732-2399/85/0401/0020$01.25

Model A is a simple linear regression model and model B is the linear-in-logs model (Carroll, Green and DeSarbo 1979).

Many criteria may be used to compare quantitative marketing models. These criteria include such things as underlying assumptions, data requirements, and theoretical implications (Larreche and Montgomery 1977; Little 1979; Mahajan and Muller 1979; Shocker and Srinivasan 1979; Narasimhan and Sen 1983). We are concerned with identification of the best functional or structural form for a quantitative marketing model. Here, depending on the aims of the modeler, "best" refers to the specification that is truly generating the data observed, or most closely represents data that will be observed.

Such a comparison of models has been traditionally treated in terms of statistical significance. It is usually possible using classical statistical inference to compare two models when one model is a constrained version of the other. This perspective is of use when the functional form of one model is a special case of a more general model, as is the case, for example, in determining the appropriate degree of a polynomial regression. Many other such examples exist in the marketing literature (e.g., see Blattberg and Sen 1973). This nested model approach requires that the mathematical forms already be specified and that they be ordinally related, with one model being a more restrictive case of the other. But often, as seen in the preceding example, the mathematical forms of the competing models will be unrelated. In such a case the usual statistical tests for nested models cannot be used.

This paper examines the case in which two or more models are to be compared, where the models are quantitative descriptions of the relationships between variables and have well-defined likelihood functions. A new method of model comparison, termed the Bayesian Cross-Validated Likelihood (BCVL) method, is proposed. Unlike classical statistical inference methods such as the likelihood ratio test, the BCVL method may be used even if the competing models are nonnested.

§2 provides a brief review of recent advances in the comparison of nonnested quantitative models. §3 introduces the BCVL method, §4 compares the BCVL method to existing approaches, and §5 presents a simulation which explores the performance of the BCVL under varying conditions for one potential application. §§6, 7, and 8 describe applications to marketing problems involving the aforementioned marketing mix models, innovation of diffusion models, and a logit analysis of business failures. §9 provides discussion and conclusions concerning usage of the BCVL method.

2. Methods of Model Comparison

Many disparate methods for comparing the structural or functional forms of quantitative models have been proposed in the last 20 years. These methods may be classified in terms of methodological emphasis as supermodel methods (Atkinson 1969), cross-validation methods (Mosteller and Tukey 1968), likelihood methods (Akaike 1974), or Bayesian methods, although there is inevitably considerable overlap and interrelation between types.

The supermodel approach involves the building of a more general model, for which each competing model is a special case. This general approach is similar in form to discrete mixtures in distribution theory (Johnson and Kotz 1977) and to supermodels in applied probability (Stacey 1962).

Atkinson (1969, 1970) proposed that if $f(x)$ and $g(x)$ are competing models, then a new function $h(x)$ should be defined, such that:

$$h(x) = f(x)^{\lambda} g(x)^{1-\lambda}. \tag{2}$$

When $\lambda = 1$, $h(x)$ reduces to $f(x)$ and when $\lambda = 0$, $h(x)$ reduces to $g(x)$. Inferences

about λ would then seem to imply conclusions about the relative acceptability of the two models.

Atkinson also mentioned a natural, intuitively appealing alternative to the above exponential formulation:

$$h(x) = \lambda f(x) + (1 - \lambda) g(x). \tag{3}$$

This formulation was rejected by Atkinson since λ less than zero or greater than one might result. Rust (1981) constrained λ to the [0, 1] interval, and used Kuhn–Tucker nonlinear optimization to obtain the parameter estimate. Unfortunately, the λ value arising from the exponential approach is difficult to interpret (Atkinson 1969) and this linear approach is difficult to implement. Also, it is difficult to use these methods when the researcher is interested in the models' performance on more than one criterion.

Traditionally, split-half cross-validation has been a widely used method of model comparison, estimating the competing models on one sample, and then compiling error statistics on the other (Mosteller and Tukey 1968; Stone 1974). It is an easy method to implement and its results are easily understandable. Mosteller and Tukey (1968) discuss some of the shortcomings of split-half cross-validation and describe alternative cross-validation methods that involve breaking up the sample into several groups, and then using each group sequentially as a holdout sample. At the extreme, the groups hold just one case, and the result is a jackknife-like cross-validation. Stone (1974, 1977) details cross-validation methods of this type. The disadvantage of jackknife-like cross-validation is that in many situations it is overwhelmingly time consuming to run the required number of analyses.

With these methods the choice of a particular quantity whose prediction is to be cross-validated is also important. The natural quantities to choose vary with the desired prediction from the model. Consequently, when the approach will be used to make several kinds of predictions (e.g., using a model of brand choice to forecast both overall market share and also the most preferred brand for each of a set of individuals) the appropriate quantity to validate may not be obvious. Clearly, several quantities might be selected and used to cross-validate competing models. But for these results to identify one model as "best" requires two difficult decisions. First, all important criteria (i.e., quantities to validate) must be identified. And second, unless one model dominates on all evaluation dimensions, the relative importance of the criteria must be specified.

Likelihood approaches to the comparison of nonnested models were first proposed by Cox (1961, 1962). His methods were applicable to only a limited set of cases, and thus were not widely applied. However, Akaike (1973, 1974) introduced a likelihood method of startling simplicity in application. Derived from information theoretic assumptions, the Akaike criterion of model comparisons is (using the notation of Stone 1979):

$$A = \log(\text{maximum likelihood}) - (\text{number of estimated parameters}). \tag{4}$$

Here, larger values of A indicate more preferred models.

This method penalizes models having a large number of parameters and is very easy to apply whenever the maximum likelihood estimation procedure is applicable. Recently, Akaike's criterion for model selection has been applied in a wide variety of settings, including ARMA time series (Akaike 1974, 1979; Shibata 1976; Neftci 1982), regression (Gorman and Toman 1966; Stone 1977), conjoint analysis (Takane 1982), multi-sample cluster analysis (Bozdogan and Sclove 1982), multidimensional scaling (Takane 1981) and in the detection of outliers (Kitagwa 1979). It has also been proposed for choosing among models in meteorology (Jones 1977).

In addition to its simplicity, Akaike's information criterion has another remarkable property. Inagaki (1977) has shown how a model's total error can be expressed as the sum of a "modelling error" component and an "estimation error." Modelling errors refer to the discrepancy between the general parametric model being estimated (e.g., the linear or linear-in-logs model in the introduction) and the true model which is operating. Estimation error stems from the fact that the particular parameter values which are used in a model must be estimated from a dataset. Both of these errors are based on Kullback–Liebler information measures. As the sample size is increased the estimation error for a model may be decreased to any desired level, but the modelling error cannot be reduced in this way. Using an incorrect parametric form (i.e., introducing modelling error) means that there will always be a discrepancy between the estimated model and the true process, regardless of the sample size. When compared with parsimonious models, complex models which include more effects (and more parameters) will generally have a smaller modelling error and a larger estimation error. Inagaki shows that Akaike's information criterion is a Bayes solution for trading off these modelling and estimation errors in choosing models. A special case of the method proposed in this paper is asymptotically equivalent to Akaike's criterion.

Another simple likelihood method was proposed by Schwarz (1978). A measure theoretic argument resulted in the criterion:

$$B = \log(\text{maximum likelihood}) - (\tfrac{1}{2}\log n) \times (\text{number of estimated parameters}), \quad (5)$$

where n is the number of independent observations. Like the Akaike criterion, the Schwarz criterion severely penalizes models with large numbers of parameters. A critical discussion of the asymptotic properties of the Akaike and Schwarz criteria is given in Stone (1979). Hauser (1978) proposed a likelihood method for comparing probabilistic choice models which was based upon information theoretic criteria. Unlike the Akaike and Schwarz approaches this method makes no allowance for model parsimony.

Bayesian methods of model comparison are used to provide posterior probabilities for the competing models. These posterior probabilities are roughly interpretable as probabilities of model correctness (given that the correct model is known to be a member of the specified set).

Bayesian estimates as an alternative to least squares estimates in the General Linear Model have been developed (Box and Tiao 1973; Lindley and Smith 1972; Smith 1973; Zellner 1971), and the application to the calculation of posterior probabilities for competing regression models has been reported by Atkinson (1978). Atkinson's Bayesian method assumes the specification of a multivariate prior distribution for the coefficients, a requirement that limits its practical utility. Other Bayesian methods arising from Econometrics are reviewed by Gaver and Geisel (1974).

Applications of Bayesian methods for the comparison of marketing models include work by Blattberg and Sen (1975) and Barry and Wildt (1977). Blattberg and Sen employed Bayesian model discrimination (BDIS) techniques (Zellner 1971) to compare stochastic brand choice models, while Barry and Wildt demonstrated that for some decision-making situations Bayesian techniques could be developed which bypass the model selection process.

The interrelatedness of the above categories of methods is exemplified by the fact that using Bayesian arguments Smith and Spiegelhalter (1980) obtained criteria closely related to those of Akaike and Schwarz. The method proposed in the next section combines advantages of the general cross-validation, likelihood and Bayesian approaches, so it may be helpful to review the relative strengths that each of these methodologies brings to our BCVL procedure.

As in cross-validation the criterion here will be computed on a holdout sample. This

usually means that the researcher is more concerned with choosing a model that will perform well with future datasets rather than picking one which perfectly mirrors the underlying process. In other words simpler models with less estimation error may be chosen over more complex, and more correct, models. This use of cross-validation is appropriate in most marketing situations where the chosen model is to be used for making actual managerial decisions (e.g., designing products, setting advertising budgets, selecting salespeople).

But models are also used to test theories about marketing phenomena, with statistical significance serving as the criterion for rejecting theories. Here, errors in estimation of parameters are not used as a criterion. Instead, more complex models are chosen if they significantly reduce modelling error. This is as it should be, and we do not propose the model selection method below as an alternative to statistical significance tests where those tests both *should* be used (i.e., in testing theories) and *may* be used. However, as noted earlier, the common tests for statistical significance apply only to nested models. When those significance tests cannot be performed, we argue that the approach proposed here, based on cross-validation, is a reasonable alternative. Simulation results in §5 show that our approach *is* effective in identifying the true underlying model among several nonnested competitors, at least when the models are of roughly equal complexity.

The advantages which the likelihood and Bayesian methodologies bring to our proposed method are somewhat interrelated. As described in the next section, cross-validation of the likelihood is appealing in itself, is equivalent in large samples to Akaike's criterion, appears also to be similar to Akaike's criterion in small samples (see §8) and makes it easy to use a Bayesian approach. That is, prior probabilities for each of the competing models can be specified and combined with the likelihood to yield posterior probabilities. In particular this enables the researcher to perform pseudo-hypothesis tests for the models. In other words, instead of specifying an α-level of 0.05 for rejecting a model in a statistical significance test, a prior probability of 0.95 for that model may be given. The procedure's description in the next section, together with the simulation and empirical examples which follow, should further illustrate the benefits of combining these three general methodologies.

3. Formulation of the BCVL Procedure

This section proposes a Bayesian cross-validated likelihood (BCVL) method; a new method for comparing alternative specifications of quantitative models. The method provides a criterion for model comparison that may easily and usefully be applied in a large number of model selection situations in Marketing.

3.1. *Theoretical Justification*

The BCVL method is Bayesian, permitting the specifications of prior probabilities for the competing models. These priors may incorporate managerial judgement, intuition, or other factors relating to model plausibility. Much has been written concerning criteria other than sheer performance which should be used in selecting a model (see, for example, Lilien 1970; Little 1970; Armstrong and Shapiro 1974; Wilson 1979). The Bayesian nature of the method also produces posterior probabilities that are interpretable as probabilities of model correctness, given that the set of alternative models is known to contain the correct one.

Cross-validation is an important part of the BCVL method, enabling the models to be tested on data other than those used to estimate the parameters of the models. A split-half procedure could be used, or (whenever possible) entirely separate data sets could be used.

The BCVL method uses the likelihood function as its indicator of model fit.

Cross-validation of the likelihood function is an alternative to the use of situation-specific fit criteria, and is thus of more general applicability. For example, maximum likelihood estimation methods have been formulated for a vast number of statistical techniques for which alternative fit techniques vary widely. Also supporting the use of likelihood functions is the Likelihood Principle, which states that the likelihood function is sufficient for inferences and decisions in any experiment (Raiffa 1970, p. 286). While the Likelihood Principle is not accepted by all statisticians, it *is* accepted by all Bayesians.

In addition marketing models often have multiple objectives. Several important parameters or measures may arise from the same model, each with its own unique practical relevance. For example, the well-known 2-parameter NBD model (and several competitors) can predict the distribution and moments of the number of brand purchases, the amount of brand switching, and the expected number of future purchases by individuals who have made 0, 1, 2, etc. purchases in an earlier period. Similarly, the 2-parameter beta binomial model of media viewing, and its competitors, predict the distribution of the number of exposures for a media vehicle as well as the reach and frequency generated by any given number of exposure opportunities. It also predicts the probability that an individual will see the vehicle on the next occasion given exposure to X out of the last N opportunities. (See Greene 1982 for an introduction to both of these marketing models.) We already know that maximizing the likelihood function is an especially effective method for choosing the particular parameter values, both for the NBD (Shenton and Bowman 1977) and for the beta binomial model (Kleinman 1973). It is simply suggested that the same idea is useful for choosing among models with different parametric forms, e.g., between the NBD and other models of brand choice. In such circumstances the use of the likelihood function provides a standard method of assessing fit which should provide an effective surrogate for a variety of more specific fit criteria.

3.2. *Posterior Probabilities for Competing Models*

The BCVL method of model comparison is summarized by the following steps:

1. Specify the mathematical forms of the competing models.
2. Choose prior probabilities for the models.
3. Randomly split one sample or obtain two distinct samples of empirical observations on which the models may be tested.
4. Estimate the parameters of the models on the first sample.
5. Using the parameters from step 4, obtain likelihoods for the models from the second sample.
6. Calculate posterior probabilities for each of the models.

The specification of competing models is accomplished ideally by the application of established theory. Where existing theory is insufficient, exploratory methods may be employed to seek out plausible models (e.g., see Tukey 1977; Rust and Bornman 1982).

Prior probabilities are then assigned to the models, perhaps incorporating such factors as attractiveness, perceived goodness, or nonperformance criteria. If there is inadequate advance basis for differentiation between the models, equal priors may be assigned. Using model parameters estimated on the first sample, likelihoods are calculated for each model on the second sample. Then posterior probabilities may be calculated for each of the models.

Expressed more formally, let S_j $(j = 1, \ldots, M)$ be the competing models, which may be nested or nonnested, with prior probabilities $P(S_j)$. Let $\boldsymbol{\alpha}_j$ denote a parameterization of S_j, constrained to a feasible parameter space $\mathbf{A}_j$. Let $\mathbf{D}_1$ and $\mathbf{D}_2$ represent the empirical data from samples one and two respectively.

First, model parameter estimates $\boldsymbol{\alpha}_j^*$ are obtained from the first sample:

$$\boldsymbol{\alpha}_j^* = \max_{\boldsymbol{\alpha}_j \in \mathbf{A}_j} L(\mathbf{D}_1 ; S_j(\boldsymbol{\alpha}_j)) \tag{6}$$

where L represents likelihood and $S_j(\boldsymbol{\alpha}_j)$ represents model S_j with parameters $\boldsymbol{\alpha}_j$. Then posterior probabilities are obtained for each of the models:

$$\begin{aligned} P(S_j \mid \mathbf{D}_2) &= \left[L(\mathbf{D}_2 ; S_j) \cdot P(S_j) \right] / L(\mathbf{D}_2) \\ &\cong \left[L(\mathbf{D}_2 ; S_j(\boldsymbol{\alpha}_j^*)) \cdot P(S_j) \right] \Big/ \sum_{i=1}^{M} \left[L(\mathbf{D}_2 ; S_i(\boldsymbol{\alpha}_i^*)) \cdot P(S_i) \right]. \end{aligned} \tag{7}$$

To facilitate computation, it is useful to reexpress equation (7) in terms of the likelihoods at each of the data points:

$$P(S_j \mid \mathbf{D}_2) = \sum_{x=1}^{N} L(\mathbf{D}_{2x} ; S_j(\boldsymbol{\alpha}_j^*)) \cdot P(S_j) \Big/ \sum_{i=1}^{M} \left[\prod_{x=1}^{N} L(\mathbf{D}_{2x} ; S_i(\boldsymbol{\alpha}_i^*)) \right] P(S_i). \tag{8}$$

Taking logarithms and exponentiating for each of the products in (8) gives:

$$= \exp\left[\sum_x \log L(\mathbf{D}_{2x} ; S_j(\boldsymbol{\alpha}_j^*)) + \log P(S_j) \right] \Big/ \sum_{i=1}^{M} \exp\left[\sum_x \log L(\mathbf{D}_{2x} ; S_i(\boldsymbol{\alpha}_i^*)) + \log P(S_i) \right] \tag{9}$$

$$= \left\{ \sum_x \left[\exp\left\{ \sum_x \log L(\mathbf{D}_{2x} ; S_i(\boldsymbol{\alpha}_i^*)) + \log P(S_i) - \sum_x \log L(\mathbf{D}_{2x} ; S_j(\boldsymbol{\alpha}_j^*)) - \log P(S_j) \right\} \right] \right\}^{-1} \tag{10}$$

where $\mathbf{D}_{2x}$ represents the data corresponding to observation x.

It should be noted that the division of data into an estimation sample ($\mathbf{D}_1$) and prediction sample ($\mathbf{D}_2$), performing steps 3–5 above, can be repeated to obtain more stable values for the predicted likelihood function. For example, with a total of N observations one can leave out the xth one, estimate each model based on the remaining $N - 1$ points, and compute the likelihood (call it $L(x)$) of the omitted observation. This procedure can be performed N times for $x = 1, 2, \ldots, N$. Then, the cross-validated likelihood is

$$L^* \equiv \prod_{x=1}^{N} L(x) \tag{11}$$

and one would use this value in step 6 to calculate the posterior probabilities for each model.

Using L^* as the cross-validated likelihood for predictive models has several conceptual advantages. First, a relatively stable value for L^* results since all N data points are held out in turn. Second, since $N - 1$ observations are used to predict each point the cross-validated likelihood should be close to the likelihood actually obtained with future data (when presumably all N observations will be used in estimation). This corrects a serious drawback of split-half cross-validation in which, due to only half the sample being used, the estimation error will be larger than it would be using the entire sample. Finally, if the proposed model is true Stone (1977) showed that $\text{Log}\, L^*$ is asymptotically equal to Akaike's Information Criterion A. Naturally, however, computation of L^* may be quite time consuming since for each model the parameter estimates $\boldsymbol{\alpha}_j^*$ must be obtained N times.

Strictly speaking L^* is an approximation, because some dependence exists between the density estimates at the different data points. However as the sample size increases this dependence decreases in importance, and may be ignored for all practical purposes in large samples. As a rule, the effect of any one data point on resulting parameter estimates should be very slight with sample sizes which are typical for marketing studies.[1] Thus we believe that the decrease in estimation error with the L^* method, and the resulting increase in stability, make the L^* method preferable to the sample split-half version of the BCVL whenever the computational requirements for L^* are feasible. Further justification for this form of cross-validation is given by Stone (1974), and the method is empirically reinforced by the simulation results in §5.

In summary, the BCVL method described here provides a mechanism for choosing between general functional forms for marketing models; e.g., choosing between Models A and B in the introduction. Once the general form has been chosen the entire set of available data would be used to estimate the particular parameters (A_0, A_1, A_2 and A_3 for Model A) which completely specify that selected model. The BCVL is employed by choosing priors for the competing models and then, for each model,

—estimating any parameters using an initial sample,

—computing the likelihood of the observations in a holdout sample, and

—computing the model's posterior probability.

Of course if equal priors are assigned, the BCVL chooses the model which has the highest likelihood for seeing the data actually observed in a holdout sample.

4. Comparison of the BCVL to Other Approaches

The BCVL method incorporates the ideas from several methods for comparing models. As the discussion in §3 suggests, using the likelihood of one or more holdout samples to aid in model selection (in particular as computed by L^*) has been proposed previously (Stone 1977). One purpose of this paper is to increase the awareness of these methods among marketing model-builders, and §§6–8 show their applicability to some marketing problems. The other objective here is to demonstrate how prior probabilities can be included in a natural way with approaches that cross-validate the likelihood. It is perhaps useful to briefly point out some of the ways in which this BCVL method is different from related methods which have been proposed in the past.

The BCVL method is different from traditional cross-validation methods in two respects. First, the cross-validation is performed using the likelihood function, instead of a situation-specific criterion such as the mean square error of a predicted quantity. Second, the Bayesian format permits the incorporation of prior probabilities for the competing models and produces posterior probabilities for the models.

The BCVL method also differs from the likelihood methods of model comparison which have been proposed in recent years. Aside from the incorporation of prior and posterior probabilities, the BCVL method assesses validity through direct cross-validation rather than by subtracting an inferred penalty term from the likelihood function.

The BCVL method is much more general than the Bayesian methods which have been developed for such specific applications as regression. If the likelihood functions may be computed, then no additional specialized packages are needed. Thus the BCVL method may be applied even to situations for which no traditional Bayesian estimation methods have been developed. Some Bayesian methods do lead to posterior distributions for model parameters, which are not available with the BCVL. But those

[1] An exception to this rule would be a regression model with a great deal of multicollinearity in the predictor variables. But even there, while the effect of one data point on the parameter estimates can be large, the effect on the likelihood of a held-out observation will be much smaller.

posteriors are generally available (and appropriate) for estimating a selected parametric model, rather than for choosing between different models.

Since Blattberg and Sen (1975) and Barry and Wildt (1977) have proposed other Bayesian methods in the marketing literature the ways in which the BCVL method is different will be described. The Blattberg and Sen approach requires the *a priori* construction of prior distributions for the parameters of each model. This requirement, which is typical of the traditional Bayesian approaches, is bypassed by the BCVL method. The only priors needed to employ the BCVL method are the prior probabilities for each of the models (which must also be supplied in the Blattberg and Sen approach).

Barry and Wildt make the interesting suggestion that if a decision maker is interested in a particular marketing variable Y then what is really important is the posterior distribution for Y, and thus the model selection step may be unnecessary. The BCVL approach is more general in the sense that by employing likelihoods it can easily also treat situations in which several variables are important. The Barry and Wildt approach might conceivably be extended to multiple criteria, but the computational problems (which for a simple application might involve multiple integration of a multivariate normal distribution) appear formidable. It should be noted that in situations where a single criterion is important it may sometimes be possible to improve upon the BCVL by directly minimizing the appropriate loss functions. However, the simulation of the next section implies that such gains may be small.

5. A Simulation of the Comparative Performance of the BCVL

A simulation was performed to test the performance of the BCVL in comparison to the performance of existing methods of comparing nonnested quantitative models. The BCVL is a model comparison method of great generality, and it would be impossible to design a simulation which would evaluate the performance of the BCVL under all possible conditions and applications. We therefore employ what we feel is a typical application, with the hope that the findings of this simulation may be suggestive of the findings which might result from many practical applications.

5.1. *Design*

We investigate the case in which three nonnested models of seasonal demand are to be compared:

$$\text{MODEL A:}\quad Y = a + b\,\text{Sin}(\theta) + e, \qquad \theta \in [0, 2\pi),$$

$$\text{MODEL B:}\quad \begin{cases} Y = a + b(\pi\theta - \theta^2) + e, & \theta \in [0, \pi), \\ Y = a + e, & \theta \in [\pi, 2\pi), \end{cases}$$

$$\text{MODEL C:}\quad \begin{cases} Y = a + e, & \theta \in [\pi/4, 3\pi/4), \\ Y = b + e, & \theta \in [5\pi/4, 7\pi/4), \\ Y = c + e, & \text{otherwise}, \end{cases}$$

with e normal and identically and independently distributed with zero mean, and θ (representing time) is uniformly distributed.

Model A models the seasonality of demand by a simple sine wave. Model B assumes a base rate a of sales, above which demand peaks in the strongest season. Model C is analogous to using seasonal dummy variables to model sales, which in effect assumes constant demand across a given season. When A is the true model we use $a = 15$ and $b = 5$, when B is true we use $a = 10$ and $b = 20/\pi^2$, and when C is true we use $a = 18$,

$b = 10$, and $c = 12$. Thus the models look similar over the time parameter θ, and may reasonably be used to approximate one another.

Seven model comparison methods are used, and their performance is evaluated on two criteria. One of the advantages of this particular application is that mean square error, a common criterion for evaluating regression models, may be used. If the BCVL method performs well against mean square error in this well-explored application, then that may make us more comfortable about using BCVL in applications which are not so well explored. Three mean square error methods are used: mean square error on the total sample, on a holdout sample, and on a "jackknifed" holdout set (one point deleted at a time). The Akaike and Schwarz criteria are also evaluated, as are two forms of the BCVL: the simple split-sample form and the conceptually preferable L^* form.

There are 5400 random data sets generated, reflecting 25 replications each of all combinations of three error variance levels (10, 20 and 30), four sample sizes (15, 30, 50 and 100), six orders of model likelihood (with the most likely model occurring with probability 0.6, next most likely with probability 0.3, and the least likely with probability 0.1), and three levels of knowledge about the model priors (priors known, priors unknown (and thus assumed equal) but all three models in the consideration set, and the true model omitted from the consideration set). All observations in each of the 5400 data sets come from one single true model. But for each experimental condition (i.e., variance level, sample size, knowledge about priors, etc.) the true model varies across the 25 replications, in accordance with the prior probabilities specified for that condition. These variables are chosen to reflect a wide range of possibilities relating to such factors as model quality, quantity of data, and the prior knowledge of the researchers. In total over 263,000 data points are employed in the simulation.

The two criteria used to evaluate the model comparison methods are mean square error (MSE) and the proportion of the time the true model was correctly selected (PROP). Mean square error is a convenient fit measure which is specific to the regression example, but PROP is a general measure which applies to all potential applications of the BCVL.

Each of the seven methods was used to select the best model specification, and that model was then estimated using the entire sample.

5.2. *Results*

The overall performance of the comparison methods is compared in Table 1. Pairwise z tests were performed on the means, testing the two-sided hypothesis that the mean (or proportion) for the L^* form of the BCVL is equal to the competing mean (or proportion). On the MSE criterion the L^* form of the BCVL is superior to all other

TABLE 1

Simulation Results: Overall Comparative Performance of Model Comparison Methods

Method	MSE	PROP
MSE (Total)	2.49**	0.44**
Akaike	2.35**	0.46
Schwarz	2.31*	0.45*
MSE (Split)	2.51**	0.37**
BCVL (Split)	2.48**	0.37**
MSE ("Jackknife")	2.34**	0.45
BCVL (L^*)	2.28	0.46

*Worse than BCVL(L^*) at the 0.05 level.
**Worse than BCVL(L^*) at the 0.01 level.

TABLE 2

Simulation Results: Comparative Performance of Model Comparison Methods By Error Variance

Method	Low Variance (10) MSE	PROP	Medium Variance (20) MSE	PROP	High Variance (30) MSE	PROP
MSE (Total)	1.45**	0.51	2.57**	0.42**	3.45**	0.39**
Akaike	1.41	0.53[b]	2.43**	0.44*	3.21**	0.41
Schwarz	1.39[a]	0.52[a]	2.41**	0.44*	3.14	0.39**
MSE (Split)	1.54**	0.43**	2.62**	0.35**	3.37**	0.32**
BCVL (Split)	1.60**	0.41**	2.56**	0.36**	3.28**	0.34**
MSE ("Jackknife")	1.42	0.51	2.42**	0.44*	3.18**	0.41
BCVL(L^*)	1.41	0.51	2.35	0.46	3.10	0.42

*Worse than BCVL(L^*) at the 0.05 level.
**Worse than BCVL(L^*) at the 0.01 level.
[a]Better than BCVL(L^*) at the 0.05 level.
[b]Better than BCVL(L^*) at the 0.01 level.

methods at least at the 0.05 level. On the PROP criterion only the Akaike criterion and MSE ("Jackknife") are not significantly worse.

Interesting patterns emerge when results are compared for different levels of error variance (Table 2). For low variance BCVL(L^*) shows little advantage over most other methods and in fact performs uniformly worse than the Schwarz criterion. But as variance increases the comparative performance of BCVL(L^*) improves. For medium and high variance BCVL(L^*) is never worse than the other methods and is always significantly better on at least one of the criteria.

Sample size variation also produces interesting (and, in this case, unexpected) results. One of the concerns about the use of a likelihood-based technique such as the BCVL is its performance on small sample sizes, where its consistency property is of little use. However, from Table 3 it can be seen that for the small sample size of 15 BCVL(L^*) outperforms all other methods on the MSE criterion at the 0.01 level. At larger sample sizes its performance is also good, as it never does significantly worse than any other method on either criterion.

Table 4 examines the effect of prior knowledge on the comparative performance of the methods. The ability of the BCVL approach to incorporate directly and simply this prior knowledge makes BCVL(L^*) much preferable when there is knowledge of the priors. This is manifested by significant advantages (at the 0.01 level) over all other methods. Interestingly even when the priors are *not* known BCVL(L^*) performs quite well. Another potential concern for a Bayesian model is the problem of the true model being left out of the consideration set. This is apparently not a serious problem (at least in this application) because on the MSE criterion BCVL(L^*) performs statistically better (at the 0.01 level) than all but the Schwarz criterion for this case.

TABLE 3

Simulation Results: Comparative Performance of Model Comparison Methods by Sample Size

Method	Very Small (15) MSE	PROP	Small (30) MSE	PROP	Large (50) MSE	PROP	Very Large (100) MSE	PROP
MSE (Total)	4.58**	0.30	2.59**	0.41**	1.68**	0.48**	1.10**	0.56
Akaike	4.31**	0.34	2.42	0.43	1.60	0.48	1.07	0.57
Schwarz	4.22**	0.34	2.37	0.42**	1.58	0.48*	1.09	0.56**
MSE (Split)	4.38**	0.26**	2.60**	0.34**	1.81**	0.40**	1.25**	0.47**
BCVL (Split)	4.25**	0.27**	2.63**	0.33**	1.79**	0.41**	1.26**	0.47**
MSE ("Jackknife")	4.26**	0.32	2.43*	0.44	1.60	0.49	1.07	0.57
BCVL(L^*)	4.10	0.33	2.38	0.45	1.58	0.50	1.07	0.57

*Worse than BCVL(L^*) at the 0.05 level.
**Worse than BCVL(L^*) at the 0.01 level.

TABLE 4

Simulation Results: Comparative Performance of Model Comparison Methods By Knowledge of Priors

Method	Priors Known: Correct Model in Consideration Set		Priors Unknown: Correct Model in Consideration Set		Priors Unknown: Correct Model Not Considered	
	MSE	PROP	MSE	PROP	MSE	PROP
MSE (Total)	2.17**	0.66**	2.17**	0.66	3.12**	n.a.
Akaike	2.02**	0.69**	2.02	0.69[a]	3.01**	n.a.
Schwarz	2.01**	0.67**	2.01	0.67	2.93	n.a.
MSE (Split)	2.25**	0.55**	2.25**	0.55**	3.04**	n.a.
BCVL (Split)	2.15**	0.60**	2.28**	0.52**	3.01**	n.a.
MSE ("Jackknife")	2.02**	0.68**	2.02	0.68[a]	2.98**	n.a.
BCVL(L^*)	1.92	0.72	1.99	0.66	2.95	n.a.

*Worse than BCVL (L^*) at the 0.05 level.
**Worse than BCVL(L^*) at the 0.01 level.
[a]Better than BCVL(L^*) at the 0.01 level.

5.3. *Conclusions*

The L^* ("Jackknife") version of the BCVL method performs very well across a wide variety of conditions. It should be stressed again that this is just one of a vast number of possible applications, and that any sweeping conclusions concerning the performance of the BCVL method would be premature. Nevertheless for this application the results appear generally favorable for the preferable L^* version of the BCVL.

In general BCVL(L^*) appears to perform particularly well for situations involving high variance, low sample sizes, or extensive prior knowledge. The Akaike and Schwarz criteria also performed well, although generally somewhat worse than BCVL(L^*).

6. Regression Models of the Marketing Mix

To exemplify the usefulness of likelihood methods in comparing regression models, we first examine the case in which two nonnested regression models are compared. Churchill (1979) reported territory data for the sales (Y), advertising spots (X_1), number of sales representatives (X_2), and wholesale efficiency index (X_3) for a brand of ball point pens (Table 5). Let us suppose that the two sales prediction models below are the models under consideration:

$$\text{MODEL A:} \quad \hat{Y} = A_0 + A_1X_1 + A_2X_2 + A_3X_3, \tag{12}$$

$$\text{MODEL B:} \quad \hat{Y} = B_0 + B_1\ln(X_1) + B_2\ln(X_2) + B_3\ln(X_3). \tag{13}$$

Model A is a simple linear regression model and Model B is the linear-in-logs model (Carroll, Green and DeSarbo 1979). Model B has the property of modelling diminishing returns for each variable separately.

Assuming normal error variance, the likelihood function for each model is calculated as follows:

$$L(\mathbf{D}) = \prod_{i=1}^{N} \left[\left(1/\sqrt{2\pi V}\right) \exp\left\{ -(Y_i - \hat{Y}_i)^2/2V \right\} \right] \tag{14}$$

where $\mathbf{D}$ represents the data and V is the error variance.

The Akaike and Schwarz criteria may be applied to the entire set of data. Since least squares regression estimates are maximum likelihood, standard least squares estimation programs may be used to provide the estimates. It is worth noting, though, that the maximum likelihood estimate of the error variance is the mean square error, which is a biased estimator (Theil 1978, p. 150).

TABLE 5

*Territory Data for Click Ball Point Pens**

Territory	Sample Assignment	Sales (Y) (in Thousands of $)	Advertising (X_1) (TV Spots Per Month)	Number of Sales Reps (X_2)	Wholesale (X_3) Efficiency Index
005	Post	260.3	5	3	4
019	Pre	286.1	7	5	2
033	Post	279.4	6	3	3
039	Post	410.8	9	4	4
061	Pre	438.2	12	6	1
082	Post	315.3	8	3	4
091	Post	565.1	11	7	3
010	Post	570.0	16	8	2
115	Pre	426.1	13	4	3
118	Post	315.0	7	3	4
133	Post	403.6	10	6	1
149	Pre	220.5	4	4	1
162	Pre	343.6	9	4	3
164	Pre	644.6	17	8	4
178	Pre	520.4	19	7	2
187	Post	329.5	9	3	2
189	Post	426.0	11	6	4
205	Pre	343.2	8	3	3
222	Post	450.4	13	5	4
237	Pre	421.8	14	5	2
242	Pre	245.6	7	4	4
251	Pre	503.3	16	6	3
260	Post	375.7	9	5	3
266	Post	265.5	5	3	3
279	Pre	620.6	18	6	4
298	Post	450.5	18	5	3
306	Post	270.1	5	3	2
332	Post	368.0	7	6	2
347	Pre	556.1	12	7	1
358	Pre	570.0	13	6	4
362	Pre	318.5	8	4	3
370	Pre	260.2	6	3	2
391	Post	667.0	16	8	2
408	Post	618.3	19	8	2
412	Pre	525.3	17	7	4
430	Pre	332.2	10	4	3
442	Post	393.2	12	5	3
467	Pre	283.5	8	3	3
471	Pre	376.2	10	5	4
488	Post	481.8	12	5	2

*This table is adapted from Churchill (1979, p. 505).

The results of the Akaike criterion are as follows for Models A and B:

$$C_A(A) = -153.60 - 5 = -158.60, \tag{15}$$

$$C_A(B) = -160.21 - 5 = -165.21. \tag{16}$$

Thus, the Akaike criterion would select Model A, the linear model, as the preferred model.

The results of the Schwarz criterion are:

$$C_B(A) = -153.60 - \tfrac{5}{2}\ln(40) = -162.82, \tag{17}$$

$$C_B(B) = -160.21 - \tfrac{5}{2}\ln(40) = -169.43. \tag{18}$$

The Schwarz criterion also prefers Model A. In fact, whenever the number of estimated parameters is the same for two competing models the two criteria will unavoidably prefer the same model, since the log likelihood is adjusted by the same amount for each model. In effect one is just comparing the likelihoods.

Alternatively, posterior probabilities may be computed by employing the BCVL method. Here, instead of using L^* the sample will be randomly split, as in Table 5. The coefficients for each of the models are obtained using the estimation sample. Likelihood functions are then calculated for both models on the validation sample. Assuming priors of 0.5, posterior probabilities of 0.91 and 0.09 are obtained for Models A and B, respectively, using equation (7).

Whereas comparison of log likelihoods and other error criteria are often difficult to make intuitive, since the magnitudes vary with the application, these posterior probabilities provide a measurement of model comparison that is easily grasped, and whose magnitudes are comparable across application. Therefore, the BCVL method should facilitate tradeoffs between the objective and subjective components of model selection.

In the interest of parsimony, it may be desirable to consider models obtained by eliminating unnecessary independent variables from Models A and B. One approach to dealing with this issue involves employing the usual nested F tests to determine the significance of the variables. Then, with the insignificant variables removed, Model A with its remaining variables could be tested against Model B with its remaining variables.

In this case, using the 0.05 significance level and the data from all observations, X_3 is deleted from both Models A and B. The resulting criteria based on 2-variable versions of the models are:

$$C_A(A) = -154.39 - 4 = -158.39, \tag{19}$$

$$C_A(B) = -171.80 - 4 = -175.80, \tag{20}$$

$$C_B(A) = -154.39 - 4/2\ln(40) = -161.77, \tag{21}$$

$$C_B(B) = -171.80 - 4/2\ln(40) = -179.18. \tag{22}$$

Using each criterion Model A, the linear model, is still preferred. It is interesting to note the comparison between the criteria for the two variable models and those for the three variable models. Using both criteria the performance of Model B is judged to have deteriorated with the loss of X_3, which is inconsistent with the implications of the hypothesis tests. However, Model A is judged to have been improved by the elimination of X_3, which is consistent with the hypothesis tests.

The construction of pseudo-hypothesis tests using the BCVL method may be illustrated by reconsidering the number of independent variables question. In Model A, assuming equal priors, the posterior probability for the three-variable model was 0.53 versus 0.47 for the model deleting X_3. A pseudo-hypothesis test giving the more parsimonious model a prior of 0.95 would make the posteriors in this case 0.06 for the three-variable model and 0.94 for the two-variable model, leading to the two-variable model being selected. It is equivalent to set the prior at 0.95 and then use 0.5 posterior probability as the critical value, as may be obtained from equation (7).

The selection of the best model using BCVL does not negate the usefulness of conventional model testing procedures. The researcher will still find it appropriate to perform the usual diagnostic tests (e.g., on normality of the error term). If the "best" model is found to be invalid on these grounds, then this is an argument for expanding the consideration set of model specifications.

7. Choosing an Innovation Diffusion Model

Innovation diffusion models of new product acceptance are generally developed and applied with one of two uses in mind. First, the inferred characteristics of the population (e.g., proportion of "innovators" vs. "imitators") may be of interest. Or secondly, one might desire forecasts of some quantities before they are observed. These can include the ultimate number of adopters, the number of adopters in some future time interval, the time at which the adoption rate will peak, and the adoption rate in that period.

When these and other predictions are of interest cross-validating the likelihood is a natural summary measure of model effectiveness. Also, since equal priors are assumed for the models considered, the criterion for selection is simply L^* (defined in §3). That is, one observation at a time is dropped for cross validation.

The general class of models examined here is described by Schmittlein and Mahajan (1982). The probability that any individual adopts by time t is given by

$$F(t) = c(1 - e^{-bt})/(1 + ae^{-bt}) \tag{23}$$

where $p \equiv b/(a + 1)$ is a coefficient representing innovativeness, $q \equiv ab/(a + 1)$ represents imitation, and c is the probability of ever adopting. So, in a sample of size M, cM is the expected number that will eventually adopt.

Heeler and Hustad (1980) discuss an interesting issue regarding innovation diffusion models, namely the use of management judgement or other external sources instead of early adoption data to estimate the parameters. For example, the average parameter estimate using the entire adoption histories of several similar products might be used. This decision can be viewed in terms of the two sources of error described in §2. The use of externally acquired parameter values obviously reduces estimation error but since the choice of analogous products is unlikely to be perfect it increases modelling error. This section shows how the BCVL can be used to choose between the use of external parameter values and values estimated from the early adoption data at hand.

The diffusion pattern for four medical innovations—ultrasound, *CT* head, *CT* body, and mammography—will be of interest here. Yearly adoption data from a sample of 209 hospitals are available for the estimation of (a, b, c) and are described in detail elsewhere (Schmittlein and Mahajan 1982). However, instead of estimating all three parameters from these data, the notion of estimating $a = q/p$ separately using the adoption pattern for several durable goods will also be investigated. Specifically a will be obtained as the average maximum likelihood estimate from data on four consumer durable adoptions—clothes dryers, room air conditioners, color TV and dishwashers. Those estimates are given in Schmittlein and Mahajan (1982). Taking this approach reduces the instability of the estimates since only b and c must be estimated from the hospital adoption data. Consequently, it will perform better in prediction than the model with all three parameters estimated if the relative sizes of the innovative and imitative effects are nearly the same for consumer durables and medical innovations.

The results from fitting the two and three parameter models are given in Table 6. The model with all three parameters estimated is effective in representing the actual adoption curves, as the figures in Schmittlein and Mahajan (1982) show. The first four columns list, for the four innovations, the maximum likelihood parameter estimates, logarithm of the maximum likelihood l, logarithm of the cross-validated likelihood L^*, Akaike's information criterion A, and the BCVL posterior probabilities derived from $\text{Log}\, L^*$. Based on both the values of $\text{Log}\, L^*$ and A the three-parameter approach is preferred for *CT* head, *CT* body, and mammography; with the two-parameter model chosen for ultrasound. So only in this last case is the accuracy of the judgmental estimate of a and the added stability of the two-parameter model enough to offset the

TABLE 6

Diffusion Model Results for Medical Innovations

	Ultrasound	*CT* Head	*CT* Body	Mammography Case 1	Mammography Case 2
3 Parameter Model					
$\hat{a}$	70.39	393.6	1191.	249.8	249.8
$\hat{b}$	0.4406	1.399	1.671	0.6474	0.6474
$\hat{c}$	0.9377	0.5608	0.4956	0.5944	0.5944
$l(\hat{a},\hat{b},\hat{c})$	− 511.3	− 321.0	− 274.0	− 417.3	− 417.3
Log $L^*(\hat{a},\hat{b},\hat{c})$	− 514.4	− 323.9	− 276.9	− 420.5	− 420.5
A	− 514.3	− 324.0	− 277.0	− 420.3	− 420.3
Posterior prob.	0.35	1.00	1.00	0.99	0.38
2 Parameter Model					
$\hat{a}$	41.41	41.41	41.41	41.41	551.7
$\hat{b}$	0.3830	0.8922	0.8334	0.4475	0.7348
$\hat{c}$	0.9784	0.5936	0.6062	0.6245	0.5886
$l(\hat{a},\hat{b},\hat{c})$	− 512.1	− 329.3	− 285.1	− 423.2	− 418.2
Log $L^*(\hat{a},\hat{b},\hat{c})$	− 513.8	− 330.8	− 286.7	− 424.8	− 420.0
A	− 514.1	− 331.3	− 287.1	− 425.2	− 420.2
Posterior prob.	0.65	0.00	0.00	0.01	0.62

modelling error from the imperfect analogy of consumer durables with medical innovations. Note that in all cases the values of Log L^* and A are very close. In addition to the result on their asymptotic equality, it seems that one can also hope to use A as a substitute for L^* in approximating BCVL posterior probabilities, for large sample sizes.

Finally, the last column of Table 6 results from considering, for mammography, a different external source for the parameter a in the two-parameter model. Given that the estimate $\hat{a}$ has already been found for ultrasound, *CT* head and *CT* body, the average of these three values can provide a in the two-parameter model for mammography. Presumably there is a closer analogy between mammography and the other medical innovations than with consumer durables. In fact, the results support this contention since the approach where only $\hat{b}$ and $\hat{c}$ are obtained from the early mammography adoptions outperforms, in terms of Log L^* and A, the model with all three parameters. The posterior probability for the two-parameter model increases from 0.01 to 0.62 when the more appropriate innovations are used to estimate a. So the analysis demonstrates that, in some cases, using externally derived quantities for diffusion models can indeed constitute a better strategy than estimating them from initial data on adoptions. But more importantly, the BCVL approach or other likelihood criteria enable the researcher to make this tradeoff between modelling and estimation error any time that the use of external parameter values is under consideration.

8. Modelling Business Failures

In this final application a model is developed to predict business failures in the years before they occur. Data for 23 failed and 23 nonfailed firms that were originally analyzed by Sharma and Mahajan (1980) are used to calibrate logistic regression models for this binary response. Those researchers found two variables to be significantly related to failure—return on assets and the firm's current ratio. So the question here concerns the effectiveness of using both of these quantities to predict business failures.

As was the case with the previous applications, a variety of predictions may be of

interest. In addition to making a "failed/nonfailed" binary prediction for each of a set of firms, one may wish to identify those firms whose probability of failure is greater than q as a risk assessment. Alternatively, a set of firms might be rank ordered in terms of likelihood of failure. As discussed in §3.1, with these multiple predictions cross-validating the likelihood function is again a natural measure of predictive effectiveness.

In this section two specifications of the logit model will be compared, the first using both return on assets and the current ratio as predictors, and the second using only return on assets. With both variables in the model one has *Equation* (1):

$$\text{Log}\left(\frac{P_t}{1-P_t}\right) = \alpha_0 + \alpha_1 ROA_{t-K} + \alpha_2 CR_{t-K} \qquad \text{where} \tag{24}$$

$$P_t = \text{the probability that a firm doesn't fail in year } t, \tag{25}$$

$$ROA_{t-K} = \text{the firm's return on assets in year } t-K, \text{ and} \tag{26}$$

$$CR_{t-K} = \text{the firm's current ratio in year } t-K. \tag{27}$$

Since the data were available, business failures were predicted in years 1 to 5 before failure. So equation (1) was estimated by maximum likelihood separately for the cases

TABLE 7

Logit Analysis of Business Failures

	Years Before Failure				
	1	2	3	4	5
Equation (1)					
Constant	− 3.69	− 3.96	− 3.56	− 2.63	− 5.44
	(2.07)	(1.63)	(1.38)	(1.13)	(2.36)
Return on Assets	25.04	31.65	19.18	18.22	33.13
	(9.38)	(11.49)	(7.47)	(6.93)	(16.19)
Current Ratio	1.23	0.58	0.72	0.19	0.66
	(1.08)	(0.81)	(0.64)	(0.48)	(0.83)
Log Likelihood l_1	− 5.31	− 15.32	− 22.35	− 24.59	− 11.53
Log L^*	− 6.70	− 19.24	− 25.65	− 27.69	− 17.71
A	− 8.31	− 18.32	− 25.35	− 27.59	− 14.53
Posterior prob.	0.67	0.28	0.43	0.35	0.04
Equation (2)					
Constant	− 1.90	− 3.21	− 2.45	− 2.33	− 4.64
	(0.95)	(1.11)	(0.85)	(0.85)	(2.20)
Return on Assets	− 28.83	36.15	22.82	19.25	37.80
	(10.26)	(11.05)	(7.40)	(6.66)	(17.96)
Log Likelihood l_2	− 6.33	− 15.73	− 23.18	− 24.68	− 12.01
Log L^*	− 7.43	− 18.32	− 26.35	− 27.08	− 14.42
A	− 8.33	− 17.73	− 25.18	− 26.68	− 14.01
Posterior prob.	0.33	0.72	0.57	0.65	0.96
$2(l_1 - l_2)$	2.04	0.809	1.65	0.172	0.957
Sample Size	36	44	46	44	26

$K = 1, 2, 3, 4, 5$, using the BMDP statistics package (Dixon 1981). The results are listed in Table 7, with the estimated standard error for each coefficient in parentheses.

Similarly, when only return on assets is used, the model becomes *Equation* (2):

$$\text{Log}\left(\frac{P_t}{1 - P_t}\right) = \beta_0 + \beta_1 ROA_{t-K}. \tag{28}$$

Results for equation (2) are also given in Table 7. As can be seen by comparing Log L^* and Akaike's A for equations (1) and (2), the model using only return on assets is chosen for prediction with four of the five time horizons. That is, whenever predicting failures two years to five years in advance, the estimation error (with these sample sizes) is increased too much when the current ratio is included, more than offsetting its contribution to the model. The degree to which the simpler model is superior is further illuminated by examination of the BCVL posterior probabilities in Table 7. Equal priors were used in computing these posterior probabilities.

9. Issues in Application

9.1. *Choice of Samples*

The strongest tests of model predictive validity require the use of two or more independent samples. Unless multiple independent samples are available, the researcher is forced to cross-validate on subsamples of a single sample. One advantage of the BCVL method over the other likelihood methods of model comparison is its ability to utilize explicitly two or more independent samples.

9.2. *Using the Posterior Probabilities*

In any practical situation the selection of a model may involve many criteria, and thus the ability to analyze the corresponding tradeoffs becomes important. The posterior probabilities obtained from the BCVL method may be useful as a means of making more informed decisions on these tradeoffs. The posterior probabilities provide an indication not only of which model is better, but also to what degree, and thus facilitates the tradeoffs involving other criteria.

One possible use of the posterior probabilities would be to form an acceptance region. For example, management may have a strong preference for Model A as opposed to Model B, on the basis of subjective criteria. A possible decision rule would be to accept Model A if its posterior probability were, say, greater than 0.2. This reasoning is similar to that employed in traditional hypothesis testing.

9.3. *Specification of Good Models*

The specification of good models is essential to the successful application of any model comparison. In the BCVL model, posterior probabilities will be apportioned across the models, even if all of them are bad. Thus, careful consideration should be given to relevant theoretical and empirical considerations in the model specification step. It should be noted however that at least for the application in the simulation the BCVL model performed well even when the true model was not in the consideration set.

9.4. *Nested Models*

When nested models are included among the competing set, the assumptions of the BCVL model would appear to be violated. This problem has been noted by Atkinson (1978) in the context of the Bayesian comparison of linear regression models. In particular, if Model A included Model B as a special case, then *both* Model A and Model B may be true. A simple solution to this dilemma is achieved by deleting from the parameter space of Model A all points in which Model A is identical to Model B.

This constraint forces the two models to be distinct, and the assumptions are no longer violated. It should also rarely become an issue in operationalizing the BCVL method. When the parameter spaces are continuous, as will typically be the case, the probability of any finite collection of points (such as the set normally deleted with the above constraint) will be zero. Thus, nested models will generally provide no theoretical or operational difficulties.

There may be some concern over the fact that classical hypothesis testing and this Bayesian approach may produce results for nested models that are apparently unrelated. However, this is not a condemnation of the Bayesian approach, but rather a reflection of basic differences in philosophical orientation between the Classicists and the Bayesians. A discussion of the theoretical basis for Bayesian statistics is given by Box and Tiao (1973).

As alluded to previously, decision rules may also be incorporated within the Bayesian approach, which resembles classical hypothesis testing. For example, when Model B is more complex (i.e., more difficult to use) than (nested) Model A, then a typical decision rule used in hypothesis testing is to choose Model B only if the researcher is 95% sure that it is better. An alternative decision rule might be to choose Model B over Model A only if B's posterior probability was greater than 0.95. This alternative pseudo-hypothesis testing procedure has the advantage of being applicable even to models that are not hierarchically related.

9.5. *Strengths and Limitations*

The BCVL method described in this paper permits the comparison of nonnested, quantitative models, even when models are functionally unrelated and/or have different numbers of parameters. The posterior probabilities obtained for each of the models facilitate tradeoffs in which objective criteria may be more conveniently weighed against the subjective criteria. As is illustrated in the previous sections, applications are possible for a wide array of typical predictive models in marketing.

On the other hand, the BCVL method may not be applied to models for which likelihood functions may not be computed. This is not a serious limitation, however, since the use of likelihoods is now widespread, and the use of models for which likelihoods may be calculated is the typical case in marketing. Also, as stated previously, if all of the competing models are bad, then the BCVL method will appear to support bad models. This may be overcome through the exercising of care in model specification.

A more serious problem arises in the choice of samples for cross-validating the likelihood function. Split-half cross-validation will overstate the model's estimation error, since only half of the available sample is used. Thus, the jackknife-like procedure used to obtain L^* is preferred. Note that when the sample size N is relatively small, estimating the parameters N times to calculate L^* is not too difficult. When N is larger, direct calculation of L^* may be infeasible (too time consuming), but here one can use the asymptotic equality of $\text{Log}\, L^*$ and Akaike's Information Criterion A. Since the latter quanity is easily computed, the cross-validated likelihood may be replaced by e^A in very large samples.

In summary, for choosing between competing predictive marketing models, the BCVL approach offers three distinct advantages:

1. it allows incorporation of prior predispositions regarding model choice, with the prior and posterior probabilities facilitating tradeoffs between objective and subjective model criteria;

2. through cross-validation it reflects the total error (due to modelling plus estimation) in the model's predictions; and

3. by cross-validating the *likelihood functions* as opposed to single forecast quantities

(i.e., usually an expected value) the ability of the chosen model to make a variety of predictions is more accurately assessed.

Over a wide range of conditions in a large simulation, for at least one typical application, the BCVL approach generally outperformed the existing methods for comparing nonnested quantitative models.[2]

Acknowledgement. The authors thank Donald G. Morrison, Yoram Wind, William Perreault, and Mukesh Bajaj for many helpful suggestions and comments, and Vijay Mahajan for providing data used in this study, as well as a number of valuable suggestions.

[2]This paper was received June 1981 and has been with the authors for 3 revisions.

References

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the 2nd International Symposium of Information Theory*, N. Petrov and F. Csadki (Eds.), Budapest: Akademiai Kiado, 267–281.

——— (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716–723.

——— (1979), "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting," *Biometrika*, 66, 237–242.

Armstrong, H. Scott and Alan C. Shapiro (1974), "Analyzing Quantitative Models," *Journal of Marketing*, 38 (April), 61–66.

Atkinson, A. C. (1969), "A Method for Discriminating Between Models," *Journal of the Royal Statistical Society* (B), 32, 323–353.

——— (1970), "A Method for Discriminating Between Models," *Biometrika*, 65, 337–347.

——— (1978), "Posterior Probabilities for Choosing a Regression Model," *Biometrika*, 65, 39–48.

Barry, C. B. and A. R. Wildt (1977), "Statistical Model Comparison in Marketing Research," *Management Science*, 24 (December), 387–392.

Bass, Frank (1969), "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research*, 6 (August), 291–300.

Blattberg, Robert C. and Subrata K. Sen (1973), "An Evaluation of the Application of Minimum Chi-Square Procedures to Stochastic Models of Brand Choice," *Journal of Marketing Research*, 10 (November), 421–427.

——— and ——— (1975), "A Bayesian Technique to Discriminate Between Stochastic Models of Brand Choice," *Management Science*, 21 (February), 682–696.

Box, G. E. P. and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Reading, Mass.: Addison-Wesley.

Bozdogan, H. and S. L. Sclove (1982), "Multi-Sample Cluster Analysis with Varying Parameters Using Akaike's Information Criterion," paper presented at the 142nd Annual Meeting, The American Statistical Association, Cincinnati, Ohio.

Carmone, Frank H. and Paul E. Green (1981), "Model Misspecification in Multiattribute Parameter Estimation," *Journal of Marketing Research*, 18 (February), 87–93.

Carroll, J. Douglas, Paul E. Green and Wayne S. DeSarbo (1979), "Optimizing the Allocation of a Fixed Resource: A Simple Model and Its Experimental Test," *Journal of Marketing*, 43 (January), 51–57.

Churchill, Gilbert A. (1979), *Marketing Research: Methodological Foundations*, 2nd Ed., Hinsdale, Ill.: The Dryden Press.

Cox, D. R. (1961), "Testing of Separate Families of Hypotheses," *Proceedings of the 4th Berkeley Symposium*, 105–123.

——— (1962), "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society B*, 24, 406–424.

Dixon, W. J. (1981), *BMDP Statistical Software 1981*, Berkeley, Cal.: University of California Press.

Gaver, Kenneth M. and Martin S. Geisel (1974), "Discriminating among Alternative Models: Bayesian and Non-Bayesian Methods," in Paul Zarembka (Ed.), *Frontiers in Econometrics*, New York: Academic Press.

Gorman, J. W. and R. J. Toman (1966), "Selection of Variables for Fitting Equations to Data," *Technometrics*, 8, 27–51.

Greene, Jerome D. (1982), *Consumer Behavior Models for Non-Statisticians*, New York: Praeger Publishers.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach," *Operations Research*, 26 (May–June), 406–421.

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

Inagaki, Nobuo (1977), "Two Errors in Statistical Model Fitting," *Annals of the Institute of Statistical Mathematics, Part A*, 29, 131–152.

Johnson, Norman L. and Samuel Kotz (1977), *Urn Models and Their Application*, New York: John Wiley & Sons.

Jones, R. H. (1977), "Multivariate Statistical Problems in Meteorology," in P. R. Krishnaiah (Ed.), *Multivariate Analysis. IV*, New York: North-Holland.

Kitagawa, G. (1979), "On the Use of AIC for the Detection of Outliers," *Technometrics*, 21, 193–199.

Kleinman, Joel C. (1973), "Proportions with Extraneous Variance: Single and Independent Samples," *Journal of the American Statistical Association*, 68, 46–54.

Larreche, J. C. and D. B. Montgomery (1977), "Framework for the Comparison of Marketing Models: A Delphi Study," *Journal of Marketing Research*, (November), 487–498.

Lilien, Gary L. (1970), "Model Relativism: A Situational Approach to Model Building," *Interfaces*, 5 (May), 11–18.

Lindley, D. V. and A. F. M. Smith (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society B*, 34, 1–41.

Little, John D. C. (1970), "Models and Managers: The Concept of a Decision Calculus," *Management Science*, 16 (April), B466–B485.

——— (1979), "Aggregate Advertising Models: The State of the Art," *Operations Research*, 27, 629–667.

Mahajan, Vijay and E. Muller (1979), "Innovation Diffusion and New Product Growth Models in Marketing," *Journal of Marketing*, (Fall), 55–68.

Mosteller, Frederick and John W. Tukey (1968), "Data Analysis, Including Statistics," in *Handbook of Social Psychology*, G. Lindzey and E. Aronson (Eds.), vol. 2, Reading, Mass.: Addison-Wesley.

Narasimhan, C. and S. K. Sen (1983), "New Product Models for Test Market Data," *Journal of Marketing*, (Winter), 11–24.

Neftci, S. N. (1982), "Specification of Economic Time Series Models Using Akaike's Criterion," *Journal of the American Statistical Association*, 77, 537–540.

Raiffa, Howard (1970), *Decision Analysis*, Reading, Mass.: Addison-Wesley.

Rust, Roland T. (1981), "A Probabilistic Measure of Model Superiority," Working Paper 81-23, Graduate School of Business, University of Texas, Austin.

——— and Elizabeth O. Bornman (1982), "Distribution-Free Methods of Approximating Nonlinear Marketing Relationships," *Journal of Marketing Research*, (August), 372–374.

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

Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Sharma, S. and V. Mahajan (1980), "Early Warning Indicators of Business Failure," *Journal of Marketing* 44, 80–89.

Shenton, L. R. and K. O. Bowman (1977), *Maximum Likelihood Estimation in Small Samples*, New York: Macmillan.

Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika*, 63, 117–126.

Shocker, Allan D. and V. Srinivasan (1979), "Multiattribute Approaches for Product Concept Evaluation and Generation: A Critical Review," *Journal of Marketing Research*, (May), 159–180.

Smith, A. F. M. (1973), "A General Bayesian Linear Model," *Journal of the Royal Statistical Socieity B*, 35, 67–75.

——— and D. J. Spiegelhalter (1980), "Bayes Factors and Choice Criteria for Linear Models," *Journal of the Royal Statistical Society B*, 2, 213–220.

Stacey, E. W. (1962), "A Generalization of the Gamma Distribution," *Annals of Mathematical Statistics*, 33, 1187–1192.

Stone, M. (1974), "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society B*, 36, 111–147.

——— (1977), "On Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion," *Journal of the Royal Statistical Society B*, 39, 44–47.

——— (1979), "Comments on Model Selection Criteria of Akaike and Schwarz," *Journal of the Royal Statistical Society B*, 41–278.

Takane, Y. (1981), "Multidimensional Successive Categories Scaling: A Maximum Likelihood Method," *Psychometrika*, 46, 9–27.

——— (1982), "Maximum Likelihood Additivity Analysis," *Psychometrika*, 417, 225–241.

Theil, Henri (1978), *Introduction to Econometrics*, New York: Wiley.

Tukey, John W. (1977), *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley.

Wilson, R. Dale (1979), "The Role of Model Testing in Theory Development: The Case of Mathematical Marketing Models," in *Conceptual and Theoretical Development in Marketing*, O. C. Ferrell, S. W. Brown, and C. W. Lamb, Jr. (Eds.), Chicago: AMA, 86–101.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley.