CURVE FITTING BY MARKOV TRANSIENTS

A THESIS

Presented to

The Faculty of the Division of Graduate Studies

By

Douglas John Bryant

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

March, 1978

CURVE FITTING BY MARKOV TRANSIENTS

Approved:

Donovan B. Young, Chairman

Robert J. Graves

Douglas C. Montgomery

Date approved by Chairman: 13 Feb '78

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Donovan Young who served as chairman of the thesis committee. His encouragement, enthusiasm and guidance were paramount in the successful completion in the research effort.

I would also like to thank Dr. Bob Graves and Dr. Doug Montgomery, as members of the committee, for their helpful suggestions on the final draft.

Finally, to my wife, Bonnie, and my children Jason and Jessica I am deeply indebted. Their love, understanding and sacrifice made this research possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

Stationary first-order homogeneous Markov processes often satisfactorily model general stochastic systems, but estimation of transition probabilities is difficult in practice. Reliable techniques are available to handle complete micro data, but micro data are usually costly and sometimes not available. Other techniques dealing with complete macro data have been developed. Both approaches are geared at estimating transition probabilities whereas in some applications the use made of them is to construct Markov chains that best replicate the observed proportions. They fail when only incomplete data are available.

The purpose of this research is to develop and demonstrate a method for constructing a stationary first-order homogeneous Markov process that best replicates available state occupancy proportions of an actual stochastic process, using a nonlinear programming approach that adjusts transition probabilities in such a way as to obtain best fit to the observed data.

In applications from the literature it is shown that the procedure gives better curve fits than traditional methods, provides capability to incorporate a priori knowledge about micro data into the estimation procedure, routinely handles incomplete macro data, giving estimates comparable to those

obtained with complete data, and in the presence of information on one state probability provides estimates of its steady state value.

CHAPTER I

INTRODUCTION

Many economically important systems from such fields as queueing, inventory control, cost control, quality control, buying behavior and reliability may be usefully modeled as stationary first-order homogeneous Markov processes. Such systems are considered to exist in a definite state or condition at each observation, with movement from state to state governed completely by a probability law associated with the current state.

Except for games and certain thermodynamic systems, few actual systems completely obey the required stationarity, order, and homogenity assumptions associated with the above process. A system may fail to be stationary, having a transition matrix that varies as a function of some outside influence. It may fail to be first-order, having transition probabilities that depend on prior information. Or, it may fail to be homogeneous, lumping dissimilar entities into a single state. In practice, however, failure to comply strictly with the above assumptions does not exclude a system from being successfully modeled as a stationary, first-order Markov process; rather, the systematic deviations introduced by the failure reduce the robustness of the result.

The relaxation of absolute compliance to these constraints opens up many new areas of application, making estimation of transition probabilities the next major obstacle to the application of a Markov model. When micro data (time ordered transition counts for all states) are available, reliable techniques exist for determination of the transition matrix. Usually, however, such data can only be obtained at high cost, and in some applications based on historical records it cannot be obtained at all.

To handle this situation, estimation techniques which do not depend on the micro data have been developed. These are able to estimate the transition matrix from purely aggregate data. This task is more difficult that estimation using micro data because the information in the macro data is masked by the aggregation. These techniques are aimed at producing statistically efficient, unbiased estimates of the transition probabilities rather than at replicating the observed data. They estimate the observed proportion at time t from the observed proportion at time t-1 in the classical least-squares manner. While this practice is consistent in analysis of transition behavior, forecasting and extrapolation, it is not consistent with respect to obtaining the process that gives the best replication of the observed data as may be desired in design of new systems based on data from existing systems, reconstruction of behavior after an interruption, and interpolation.

Additionally, there are many potential applications where even macro data are difficult to obtain. Such a case exists where the intervals between observations are not constant, as in testing of medical patients. Traditional techniques are completely unable to cope with this, and new techniques must be developed.

Nontraditional techniques are also required when the least squares metric is inappropriate. Such a case occurs when the residuals are found to have a double exponential rather than a normal distribution. Lee, Judge, and Zellner [19] develop a minimum absolute deviation estimator that can be used in such cases. In other cases the error depends on a cost function [32] which may be discontinuous; here special techniques are required.

The purpose of this research is to develop and demonstrate a method for construction of a stationary first-order homogeneous Markov process that best replicates the available state occupancy proportions (aggregage or macro data) of an actual stochastic process. This is based on the idea of choosing a transition matrix such that the resulting transient solution gives the best fit to the observed data.

CHAPTER II

SURVEY OF RELATED LITERATURE

In recent years the applications of Markov chains have extended beyond the physics of Brownian motion into many areas of the social and economic sciences. These have come from such diverse areas as learning theory [24], consumer brand preference behavior [21], and projections of the numbers of firms within industrial classifications [28].

In 1952, George H. Miller attempted to model the response of rate in a T maze as a two-state Markov process. In his research 10 rats were run in a T maze for 20 trails each. Miller suggested that the successive trials constituted a Markov chain having transition probabilities equal to the conditional probability that a rat would make a correct or incorrect choice given the choice in the previous trial. He derived a least-squares estimate of the transition probabilities using the macro data made up of proportions of rats choosing the correct turn on each trial.

Although he presumably had available data on the performance of each rat at each trial, he did not consider this micro data in the determination of the transition matrix. As pointed out in a later paper by Goodman [8] such micro data would have allowed him directly to compute the transition frequencies, which are known [2] to be maximum likelihood estimators of the transition probabilities, asymptotically normally distributed, consistent, and tending to zero bias as the sample size increases.

In many instances, however, micro data on individual transitions are not available, or may be obtained only at high cost. In consumer brand preference testing, for example, it is fairly easy for a market research firm to determine from sales data the proportion of consumers purchasing the product of a particular manufacturer at any given time. It would be much more difficult to determine the number of consumers who switched from one brand to another during any particular time period. This would require questioning individual consumers.

Problem using micro data to estimate the transition probabilities make up the bulk of the applications problems reported to date, probably because of the ease with which micro data can be transformed into these estimates rather than from an abundance of problems for which micro data can be easily obtained.

In the orderly development of techniques for estimating the transition probabilities, Miller's work comes first. He estimated the vector of state probabilities at time t from the observed state frequencies at time t-1. That is,

$$\hat{\pi}(t) = \pi(t-1)\ P \qquad (2.1)$$

where P is the transition matrix, $\hat{\pi}(t)$ is the estimated proportions at time t, and $\pi(t-1)$ represents the observed proportions at time t-1. Miller suggested that the observed proportions differed from the estimated proportions by an error term U(t) and rewrote equation (2.1) as

$$\hat{\pi}(t) = \pi(t-1)\ P + U(t) \tag{2.2}$$

where the row vector U(t) represents the differences between the observed data and the estimated proportions.

Minimizing the sum of the square of these errors, where the sum is taken both over time and over the state space, forms the basis for Miller's estimate.

When rewritten in vector notation the expanded to include all time periods zero through T, equation (2.2) appears as

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_r \end{bmatrix} = \begin{bmatrix} x_1 & & & & \\ & x_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & x_r \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_r \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_3 \end{bmatrix} \tag{2.3}$$

where each of the y_i is a (T x 1) vector of empirical data representing the observed proportions in state i at time 1 through T, and $x_1 = x_2 = x_r$ are (T x r) matrices representing observed proportions in all states 1 through r at times 0 through T-1, p_j is an (r x 1) vector of transitional probabilities, and u_j is a vector of random errors. In more compact form this may be expressed as

$$Y = X P + U \tag{2.4}$$

The classical methods of unweighted least squares give the solution to the minimization of the product UU' as

$$P = (X'X)^{-1} X'Y. \tag{2.5}$$

Although this solution is guaranteed to meet the requirements that the row sum constraints, $\sum_j p_{ij} = 1$ for all i, are satisfied, non-negativity of the estimates is not guaranteed. This estimator, then is referred to as the unweighted, unrestricted least squares estimator.

The approach followed in the development of this estimator is important not just because it sets that stage for most of the later work, but also because it contrasts with the approach to be taken later in this research. It is significant in this regard that Miller notes "We do not have a least-squares fit of the function, $p^{(n)}(R)$ to the observed data; we have a least squares fit for the transformation, T."

It is important, also, to note the underlying assumptions implied in Miller's model. Foremost among these is that the data can be represented by a first-order stationary Markov process. "First order", defines the length of the memory associated with the process; the outcome at time t+1 depends only on the outcome at time t and on the transition matrix, not on the outcomes at earlier times. "Stationary" means that the matrix of transition probabilities does not change through time. Although this is a common assumption, its validity is in question for many applications.

Another important assumption not discussed by Miller, but implied by his work, is that the sample is homogeneous. In the context of Miller's test this means that each rat in the sample was operating under the same set of behavioral rules; that each rat obeyed a Markov behavioral model with the same set of parameters in the transition matrix. Here, two types of non-homogeneity are possible. First, it is possible that not all of the rats operate under Markov behavior rules. Second, it is possible that the transition probability matrix differs among rates. Jones [13] and Harary [9] attempted to deal with some cases of non-homogeneous samples.

Miller's work, then, is important because it introduced modeling by Markov processes into new areas and because the method of estimation and the underlying assumptions appear in most of the later work in the area.

A theoretical fault in Miller's mathematics has been pointed out by Kao [14]. Goodman [8] provides a corrected version of Miller's proof and goes on to address the question of inadmissible estimates of transition probabilities caused by the fact that estimates given by the general least-squares technique are not guaranteed to be greater than or equal to zero. Goodman states that if an inadmissible estimate is encountered, the admissible estimate will lie on the boundary of the admissible subspace. This means that whenever a negative estimate is encountered its value should be taken as zero. Goodman did not provide, however, a method for adjusting the values of the other parameters to maintain the row sums of unity.

The next significant methodological advance occurred when Madansky [20] observed that the least squares estimates developed by Miller do not possess the optimal properties normally associated with least-squares estimates. The estimates do not have minimum variance, because the errors at each time are not uncorrelated. Madansky, however, offered a relatively simple solution to this problem by minimizing an objective function weighted by the covariance matrix. He developed an iterative procedure in which the unweighted estimates are determined first, and then used to determine the weightings for the next determination of the transition matrix. He stated that this procedure produces estimates of the transition probabilities that are asymptotically more efficient than those given by the unweighted methods. These weighted least squares estimates continue to suffer from the problem that some of the estimates may be inadmissible. Additionally, the weighted estimates are no longer guaranteed to meet the row sum constraints.

Further difficulties arose in an example computation of the weighted estimates. In the calculation of the improved estimate associated with Miller's rat problem, Madansky encountered negative covariances. Since negative or zero covariances are not allowed, he arbitrarily replaced these inadmissible estimates of the covariance with 0.001 and proceeded with the determination of the estimates. Upon completion, Madansky observed that the work required to produce the improved matrix was as great as the work required to produce the original unweighted estimates although significant improvement is not guaranteed. Unfortunately, no method is known that will give an a priori determination of the decrease in the variance to be gained through the calculation of the weighted estimates.

At the same time that Kao, Goodman, and Madansky were refining Miller's least squares estimates, others were making advances in the application of Markov process to actual situations. Marshall and Goldhammer [22] postulated three Markov models for the epidemiology of mental disease and attempted to test one of these models with data obtained about mental patient admissions to hospitals in Ohio and Illinois. The models they suggested were all five-state, first-order stationary Markov models. Each of the states in the various models represented some stage in the life of a person suffering from mental illness. The states began with complete freedom from illness and ended in an absorbing state, death.

Knowledge of the structure of their applied problem allowed Marshall and Goldhammer to assign values of zero or one to many of the transition probabilities. The exploitation of the special structure in this manner allowed them to reduce the number of parameters requiring estimation from 25 to only 4.

Richard Maffai [21] introduced the methods of Markov processes into the area of consumer behavior modeling. He developed a two-state model for consumer brand choice behavior. Choices were limited to "our brand" and "all others". This model was used to analyze the effects that occur in the steady state market when it is perturbed by advertising or promotional activities.

Preston and Bell [28] followed the suggestion of Adelman [1] applied a Markov chain model to the size distribution of firms in the food-packing industry. They collected data on the 25 largest firms in 1958 and used records of the Federal Trade Commission to discover past histories of these firms. Although this allowed Preston and Bell to work with a constant group of firms, the use of data gathered in this manner imparted a bias in that they included firms which grew from small to large, but did not include those firms that declined from large to small. Their method for gathering data also made it possible for them to obtain micro data from the time histories of the individual firms. Preston and Bell employed this data in obtaining the standard transition-frequency estimates of the transition probabilities.

Harary and Lipstein [9] extended the applications of Markov processes in the area of market dynamics with their analysis of non-homogeneous populations. They divided the population into two groups based on propensity to switch brands, deriving different transition matrices for the two groups. They point out that advertising and promotional activities have a much higher impact on one group than on the other.

In a work that builds on some of the techniques developed earlier, Styan and Smith [33] developed a four-state model and applied it to brand choice behavior of consumers of soap powders and detergents. Their straightforward application involved the collection of the micro data and the development of the standard transition frequency estimates of the transition probabilities from this data. They also applied tests for the order of the Markov chain and for the stationarity of the transition matrix. Their short paper provides a clear example of an application for which micro data were available.

Another example of an application of Markov processes to the problem of market dynamics is provided by Farris and

Padbert [6]. In consultation with personnel of the Agricultural Marketing Service they established eight size categories for Florida citrus packers. Packers were classified into categories by the amount of produce shipped in the interstate fruit market. Like Preston, Farris and Padberg decided first on the group of firms to be considered and then checked past records of those firms to obtain micro data. Subsets of these data were used to determine several estimates of the transition probabilities for shorter time periods within the overall 1948 to 1960 period. They encountered significant differences (as determined by a Chi square test) in the projected consequences of these estimates; they decided that the Markov process was not stationary. External factors such as the end of World War II and the rapid growth of the frozen food industry in the early 1950s were suggested as possible causes for this non-stationarity.

Krenz [16] presented a similar application dealing with the classification of farms in North Dakota. He attempted to model the size distribution of farms with a six-state Markov chain. Unlike earlier practioners in this area, Krenz did not have direct access to the micro data. The quinquennial U.S. Census of Agriculture records only the numbers of farms in each size category and does not provide information on the numbers of farms moving from one size group to another. Thus, Krenz could not directly use the transition-frequency estimator. Although there were methods available to handle the aggregate data (such as Miller's least squares technique), Krenz approached the problem by assuming away some of the possible transitions. With this method he eventually arrived at a set of "allowable" transitions which he uses to estimate the numbers of farms actually making a transition for each of five time periods. He then applied the standard transition frequency estimator to this manufactured micro data.

The next major advance in the techniques for estimating the Markov transition probabilities from aggregate data was produced by Lester Telser. [34] He built on the work of Miller, Kao, Goodman and Madansky by providing a method for adjusting the transition probabilities when the estimates obtained from the least squares technique are inadmissible. Telser handled the possible violation of the non-negativity constraints with a cyclic procedure that sets each combination of the inadmissible probabilities equal to zero in turn, with an arbitrary adjustment of the other elements of the matrix. The combination that has the least sum of residual squares is chosen to be the best estimate.

Telser also attacked the problem that arises when Madansky's weighted least squares estimate is employed. This is that the resulting estimates may not meet either the row sum constraints or the non-negativity constraints. He suggested that a Lagrangian multiplier approach could be applied to guarantee that the equality constraints are satisfied. Telser suggested that if any of the non-negativity constraints are violated, a quadratic programming (QP) approach could provide the best solution method.

Theil and Rey [36] applied the quadratic programming approach suggested by Telser to solution of Telser's cigarette brand switching problem. They first determined the unrestricted solutions to the weighted and unweighted least squares problems and then applied the QP method of Theil and van de Panne [35] to arrive at solutions to problems under explicit specification of the row sum and non-negativity constraints.

Lee, Judge and Takayama [17] applied the standard QP algorithm of Wolf [38] to arrive at an identical solution to the unweighted problem without intermediate calculation of the unrestricted estimates. They pointed out that the estimates arrived at in this manner have lower residual sum of squares than those arrived at through the use of Telser's adjustment procedure. So, it appears that the use of Telser's procedure is not guaranteed to satisfy Goodman's requirement that the estimates by chosen on the boundary of the subspace in such a way as to minimize the residual sums of squares. In an experimental example, these authors noted that the restricted least squares estimates based on the aggregate data in some cases could provide estimates that are superior to those obtained through use of the transition frequency estimates for micro data. They also developed a minimum-absolute-deviations estimator for the transition probabilities. No sampling characteristics were presented.

Lee, Judge and Zellner [19] and [18] developed maximum likelihood and Bayesian estimators to be used with macro data. If the covariance matrix is known, the calculation of the unrestricted maximum likelihood (ML) estimates are straightforward. They noted, however, that this is usually not the case, as the covariance matrix depends itself on the unknown transition probabilities. Accordingly, they implemented a cyclic procedure in which the covariance matrix is estimated and then used in the estimates of the transition probabilities.

The explicit inclusion of the non-negativity and rowsum constraints complicates this procedure by requiring solution of a QP problem each time the transition probabilities are estimated. Thus, a recursive QP problem must be solved. This procedure is continued until the estimates of the transition probabilities converge.

The additional inclusion of prior information allowed Lee, Judge and Zellner to develop Bayesian estimates of the transition probabilities. They followed a scheme similar to that with which they produced the ML estimator. Again, estimation of the restricted transition probabilities called for solution of a recursive QP problem.

Finally, Lee, Judge, and Zellner made tests on the performance of many of the available estimators. Simulation data was generated using an actual Markov chain and samples were taken of this simulated data. As a means for comparison of the estimators they chose the root-mean-squared-error. They rank the estimators according to their performance in the test as follows: Bayesian (first); restricted ML, restricted-weighted least squares, and restricted-unweighted least squares (second); and unrestricted ML and unrestricted least squares (third). The Bayesian estimators performed better than the others both when leptokurtic (peaked) prior specifications with true means and small variances were given, and when platykurtic (non-peaked) prior specifications with equal means and large variances were given. In the posterior matrix only those estimators, when the true value was actually zero, showed a serious departure from normality.

Ali Ezzati [5] made direct application of the techniques presented above in his analysis of market shares of home heating systems.

In recent papers, Saffer [31] and Saffer and others [30] investigated an alternative approach to estimation of the transition probabilities. Rather than attempting to find the "best estimate" of the transition probabilities they have attempted to find the estimate of the transition probabilities that produces the best fit to the observed data. That is, they have attempted to find the best "least squares fit of the function $p^{(n)}(R)$ to the observed data" that Miller observed he was not getting. Their objective was to provide a Markov chain model for the movement of Rose Bengal through the body. They developed a four-state model in which each state represents a different biological system. Knowledge of the special structure of their problem allowed them to assign values to some of the transition probabilities thereby reducing the number of parameters requiring estimation from 16 to 4. They apply a nonlinear programming algorithm to arrive at estimates of the P_{ij}.

Since the approach taken in these papers is very similar to the approach taken in this research, more details of the development and formulation of the problem may be found in the next chapter.

CHAPTER III

DESCRIPTION OF THE RESEARCH

In the previous chapter several methods for estimating Markov transition probabilities were reviewed. In general, these methods sought "best" estimates in the sense of efficient, unbiased estimates of the true transition probabilities under various data conditions. When micro data were available, the transition frequencies gave maximum likelihood estimates. When only macro data were available several techniques were used: Least squares estimates were improved by adding weighting factors to improve asymptotic efficiency, several methods including direct application of quadratic programming were used to maintain feasible estimates, and maximum likelihood and Bayesian estimation procedures were developed.

The Basis for a New Estimate

The abovementioned methods, as Miller noted, do not necessarily give best fit to the observed data. In addition, they require complete data; if data are lacking for some time periods or for some states, they cannot be applied. These estimates are also limited to least-squares and related error metrics.

The purpose of this research is to provide a method for obtaining an estimate of the transition probabilities that gives a good fit to the observed data. These estimates should be usable when only incomplete macro data exist and should not be limited to the euclidean or squared metrics. Finally an important objective is to apply the method to a wide selection of problems in order to ensure its robustness and generality of application.

Recall that Miller estimated the state space vector at time t from the observed data at time t-1. That is

$$\hat{\pi}(t) = \pi(t-1)P \tag{3.1}$$

for all times, t=1 to T.

In this research a different method of determining the $\hat{\pi}(t)$ is suggested. That is, that the vector at time t be determined not from the observation at time t-1, but from the <u>estimated</u> proportions at time t-1. This gives

$$\hat{\pi}(t) = \hat{\pi}(t-1)P \tag{3.2}$$

This expression holds for all t, except that at time t=1, the initial estimate of the state proportions is based on the observed data at time zero, so that

$$\hat{\pi}(1) = \pi(0)P \tag{3.3}$$

The selection of this method for determining the subsequent state space proportions gives rise to an obvious class of error criteria: sums of functions of the differences (errors) between the estimated and the observed data. The formulation of the problem employing such functions is shown below:

$$\text{Minimize} \sum_{i=1}^{N} \sum_{t=1}^{T} f_{it}(\hat{\pi}_i(t), \pi_i(t)) \tag{3.4}$$

$$\text{subject to} \sum_{j=1}^{N} p_{ij} = 1, \forall i \tag{3.5}$$

$$\text{and} \quad 0 \leq p_{ij} \leq 1 \quad \forall i \, \forall j \tag{3.6}$$

When the familiar unweighted least squares metric is employed, the objective function in (3.4) may be rewritten as

$$\text{Minimize} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{\pi}_i(t) - \pi_i(t))^2. \tag{3.7}$$

This objective function was arrived at independently by Saffer [31] and employed in the Rose-Bengal transport problem.

To write an explicit specification of this function requires that each of the $\hat{\pi}_i(t)$ be broken down in terms of the original state proportions $\pi_i(0)$ and in terms of the transition probabilities, the p_{ij}. For example, $\hat{\pi}_1(2)$ is written in this manner as

$$\hat{\pi}_1(2) = \hat{\pi}_1(1)p_{11} + \hat{\pi}_2(1)p_{21} + \hat{\pi}_3(1)p_{31}. \tag{3.8}$$

Since this expression is not yet written in terms of the original vectors, the $\hat{\pi}_i(0)$; each of the $\hat{\pi}_i(1)$ must be expanded further. This is,

$$\begin{aligned}\hat{\pi}_1(2) = &(\pi_1(0)p_{11} + \pi_2(0)p_{21} + \pi_3(0)p_{31})p_{11} \\ &+ (\pi_1(0)p_{12} + \pi_2(0)p_{22} + \pi_3(0)p_{32})p_{21} \\ &+ (\pi_1(0)p_{13} + \pi_2(0)p_{23} + \pi_3(0)p_{33})p_{31}\end{aligned} \tag{3.9}$$

Collecting terms, this reduces to

$$\begin{aligned}\hat{\pi}_1(2) = &\pi_1(0)\ (p^2_{11} + p_{12}p_{21} + p_{13}p_{31}) \\ &+ \pi_2(0)(p_{21}p_{11} + p_{22}p_{21} + p_{23}p_{31}) \\ &+ \pi_3(0)(p_{31}p_{11} + p_{32}p_{21} + p_{33}p_{31})\end{aligned} \tag{3.10}$$

This term would then be differenced with the observed data $\pi_1(2)$ in the form

$$(\hat{\pi}_1(2) - \pi_1(2))^2 \tag{3.11}$$

to make up a single term of the objective function given in (3.7). The breakdown and recombination of these terms becomes increasingly difficult as the state space and the number of time periods expand, giving an expression with many products and sums of projects of the p_{ij}. In practical problems the expansion of the objective in this manner and the differentiation with respect to the p_{ij}, as would be called for in classical minimization, is prohibited by the magnitude of the resulting polynomials. Additionally, taking the partial derivatives and solving the set of equations that results when these are set to zero does not directly guarantee the satisfaction of the constraints specified in (3.5) and in (3.6).

Returning, then, to the problem stated in (3.4) we have, in essence, an objective function made up of sums of polynomials where the polynomials are of several variables and of varying degree. No specific solution techniques are available for problems of this type. Nonlinear programming procedures (NLP), however, will usually find at least a local optimum. Additionally, many of these techniques have the advantage that the objective function does not require explicit specification, but only the value of the objective at particular points.

After trying several different nonlinear programming procedures (cyclic coordinates, unconstrained gradient, and a Markov pair adjustment procedure) with limited success, the decision was finally made to use a constrained gradient method of solution.

Since gradient information is not directly available, this information is obtained through the use of the two-point difference approximation. Each of the p_{ij} are independently given a small increment, Δ, and the resulting functional values are differenced with the original value and divided by Δ to provide the approximation to the gradient in this direction.

This information is used with the current values of the p_{ij} to obtain new values of the p_{ij} in the following manner:

$$\hat{P}^{(n)} = P^{(n+1)} - \nabla^{(n)} d$$

where $P^{(n)}$ is the nth estimate of the transition matrix, $\nabla^{(n)}$ is the gradient approximation derived from this estimate, and d is a scalar variable determined by the unidimensional line search.

Explicit satisfaction of the constraints in (3.6) is guaranteed by explicitly limiting the p_{ij} to be greater than or equal to zero. Whenever a particular p_{ij} to be tested was found to be less than zero, it was set exactly to zero. The requirement that $\sum_i p_{ij} = 1$, for all i, was maintained by an adjustment procedure that projects the gradient direction onto the associated constraints. This is done by dividing all elements in a particular row by the sum of that vector. This normalizes the estimate of the transition probabilities to be tested.

The testing procedure itself calls for calculation of each of the estimated state proportions at times 1 through T. This is done in the recursive manner already presented in equations (3.2) and (3.3). The differences between the values estimated in this manner and the observed values were calculated and summed. This sum was the value of the value of the objective function associated with this estimate of the transition matrix.

The line search procedure for determining how far the old P-matrix should be moved in the gradient direction called for acceptance of this new point (d), and the associated transition matrix, only if the value of the objective function was superior to all previous values. When this is the case the estimated transition probabilities are accepted and the best value of the objective function is updated. The value of the variable d is recorded.

When the objective value associated with the trial set of transition probabilities is worse than the best objective already available the trial matrix is rejected.

In either case the line search procedure continues by updating the value of d and generating another trial matrix based on the previous base matrix and on the outcome of the last trial. This is continued until a predetermined accuracy has been obtained for the parameter d. At this time the best point found by the line search is accepted as the new base point and the nonlinear procedure is started again with recalculation of a new gradient direction. Cycling of this procedure is prevented by acceptance of only improving points, thereby giving a steadily decreasing objective function.

The recursive nonlinear programming procedure (Figure 1) is continued until the gradient is close enough to zero (as determined by a preset value).

It is possible for a procedure such as this one to terminate without actually finding the optimal values of the transition probabilities. This can happen in several ways. First, the condition that all elements of the gradient be equal to zero does not give a theoretical guarantee that the minimum has been found. The condition is also met by a maximum (although for our application such a case is unlikely) and by a saddle point. Second, it is possible that the method of approximating the gradient can give values of zero when there exist possible directions of improvement. This is caused by the need to evaluate the function at points close to the current point in order to determine the differences. If the nearby points are not close enough, it is possible to miss an improving direction. Finally, even if the method truly converged to a minimum, there is no guarantee that the minimum found is the global minimum. It is possible that many local minima exist. Figure 2 gives Saffer's two-variable interpretations of the multiple variable

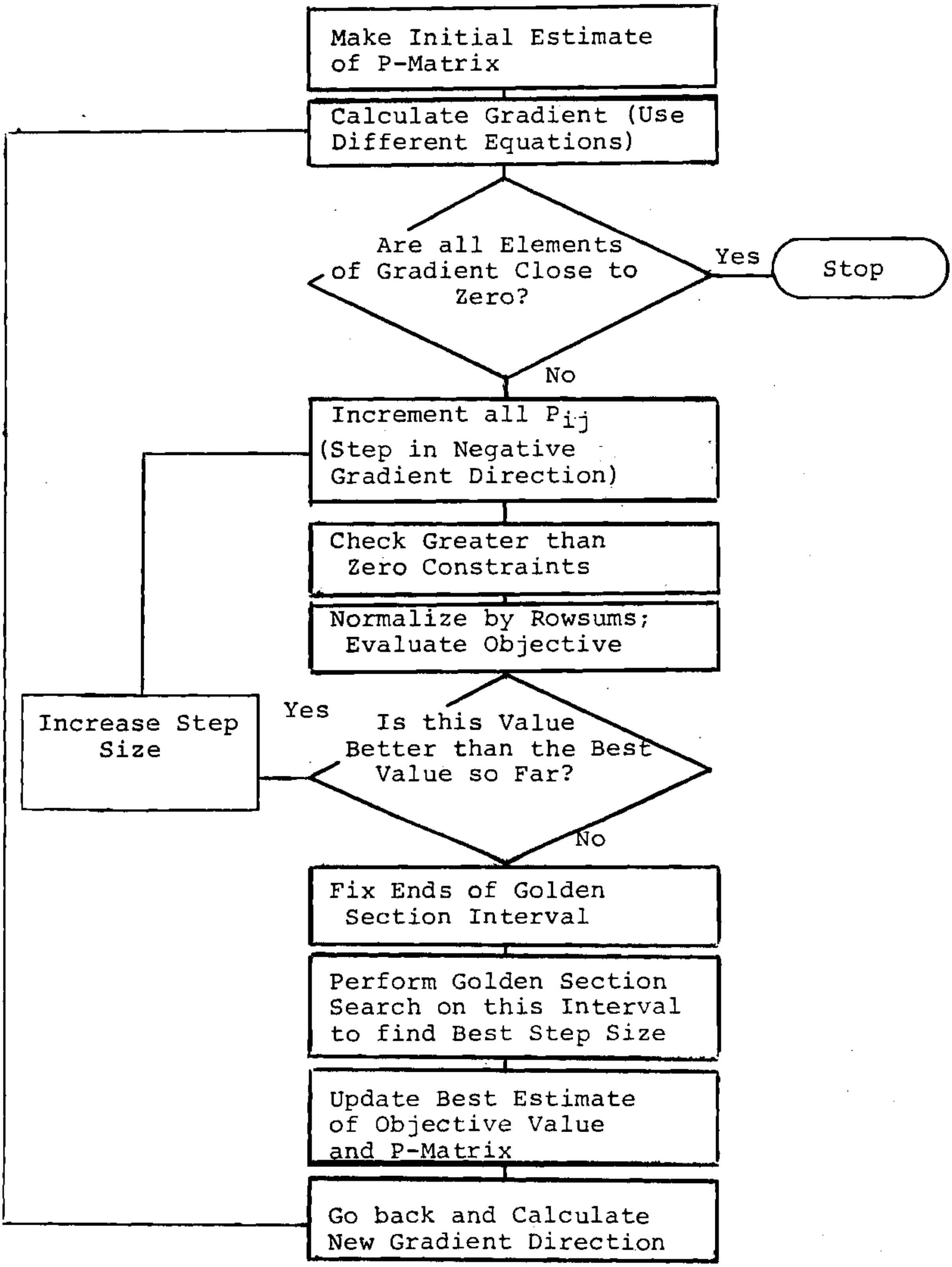

Figure 1. Overview of the Non-linear Programming Method.

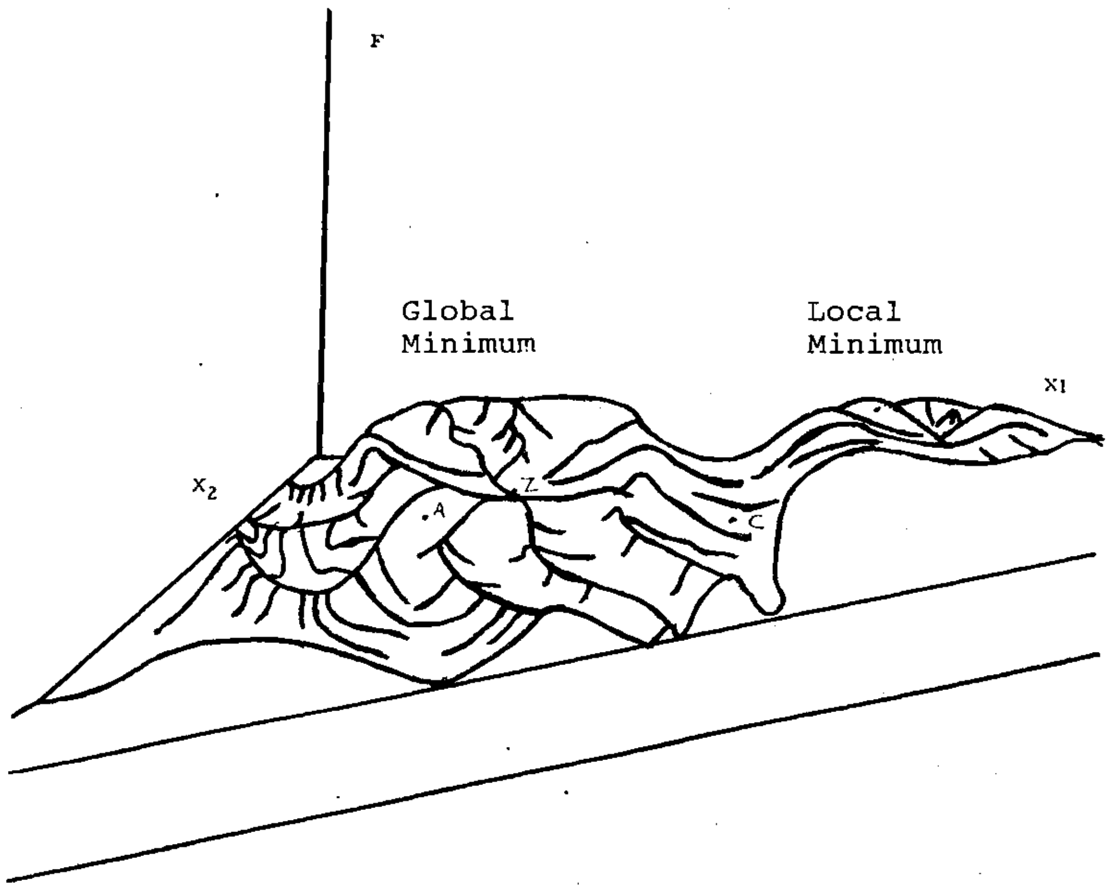

Figure 2. Interpretation of the Objective Surface

objective surface for the Rose-Bengal problem. Point m represents a local minimum.

Procedures are available for handling all of these problems. When it is felt that the procedure has terminated at a point away from which an improving direction exists, the search may be continued with a reduction in the step size that governs the selection of the points required for calculation of the gradient. The reduction of this parameter means that the points are chosen closer to the original point. If an improving direction exists, sufficient reduction of this step size (within the limits of the computer) should find it. For the other problems, saddle points and local minima, the best procedure is to restart the algorithm at several different points. If a sufficient number of these are chosen, there should be little doubt as to the location of the global optimum.

In a similar problem Saffer [31] experimented with several different solution techniques: simplexing, game tree, Gauss least squares, and a gradient approach. For the Rose-Bengal transport problem he found that a combination of pattern search with Gauss least squares provided the most efficient solution method. He reports convergence and execution time difficulties with many of the methods tried.

Testing of the Procedure

Testing of the gradient projection procedure suggested in this research was initiated with the production of several sample problems. Actual Markov chains were generated from arbitrarily selected transition matrices. The transient solutions generated in this manner were used to test the convergence of the algorithm. Next the degree of difficulty was increased by adding normally distributed random errors to the transient solutions already generated. The non-negativity constraints on the πs were maintained by setting any negative elements to zero. The requirement that the sums of the row equal one was maintained by then dividing through by the new row sums to normalize the elements of the $\pi(t)$ vectors. Several different guesses of the transition matrix were used to start the algorithm.

Next, several problems from the literature were attempted. Krenz's farm distribution problem, Telser's cigarette brand switching problem and Ezzati's home heating fuel problem were solved. Then the performance of the estimates given by the procedure was compared to several of the estimates given by Lee, Judge and Zellner's maximum likelihood and Bayesian techniques for their simulated data. Constrained and unconstrained versions of Saffer's Rose-Bengal transport problem, characterized by the intermittent nature of the data, were solved.

The investigation of the performance of the method when complete data were not available was continued with the construction of new problems from the earlier ones of Telser and Ezzati. Certain subsets of the available data were ignored and the performance under these conditions was compared to the performance under conditions of full data.

CHAPTER IV

RESULTS

The nonlinear constrained gradient (NLP) procedure described in the previous chapter was coded in Fortran IV, implemented on the Control Data Cyber 74 computer and applied to several problems. These were of three types. First the procedure was tested for its ability to recover an actual Markov matrix from transient solutions masked with random error. Next it was applied to several problems taken from the literature in which Markov transitions matrices were estimated from macro data; results were compared with those of the authors who had used traditional estimation techniques. Finally, it was applied to problems to which these traditional techniques could not be applied. These were characterized by scarce data and intermittent observations. Where applicable, results from these problems have been compared to results obtainable with complete data.

Markov Test Problems

Informal experiments in the development of the NLP procedure had confirmed the ability accurately to reconstruct Markov matrices from actual transient data. As a further test, random errors were introduced into these transient solutions. Six problems, each of three states and ten time periods, were constructed through the application of two types of random errors to the transient solutions generated by actual matrices. The three problems in the first group were characterized by small amounts of error. Actual transient solutions were generated and then rounded off to four decimal places. This rounding was equivalent to the introduction of a random error uniformly distributed in the range 1×10^{-4} to -1×10^{-4}. Table 1 gives both the original generating Markov matrices and the solution matrices that resulted when these transients were fed into the NLP optimization procedure. The error listed is the summed squared amount by which the transient solution generated by the matrix deviates from the rounded data. In Problem Two, for example, it can be seen that approximately 1/4 of the total error achieved by the NLP technique could be attributed directly to the introduction of the uniform errors. A total squared error of 4×10^{-8} means that the average absolute error by which an estimated $\pi_i(t)$ deviated from the actual value was only 3.8×10^{-5}. This compares to the standard deviation of the random error of 2.9×10^{-5}. The estimates for Problems One and Three are even closer. Table 2 lists actual and estimated state proportions for Problem One.

Although the small amounts of error and the close similarity of the estimated and test matrices indicate that the global optimum has probably been found, this was checked by restarting the problems several times. Problem One was

Table 1. Markov Test Problems with Small Errors

Problem 1	0.4000	0.1000	0.5000	0.4000	0.1000	0.5000
	0.6000	0.2000	0.2000	0.6000	0.2001	0.1999
	0.3000	0.6000	0.1000	0.3000	0.6000	0.1000
	error 1 x 10^{-8}			error 1 x 10^{-8}		
Problem 2	0.2200	0.4600	0.3200	0.2204	0.4599	0.3197
	0.3700	0.4000	0.2300	0.3838	0.3877	0.2285
	0.6800	0.0800	0.2400	0.6616	0.0959	0.2425
	error 1.2 x 10^{-8}			error 4 x 10^{-8}		
Problem 3	0.4000	0.6000	0.0000	0.4001	0.5996	0.0004
	0.2000	0.2000	0.6000	0.1994	0.2004	0.5997
	0.3000	0.1000	0.6000	0.3000	0.1001	0.5999
	error 1 x 10^{-8}			error 1 x 10^{-8}		

Table 2. Actual (π) and Estimated ($\hat{\pi}$) State Proportions for Problem One.

t	$\hat{\pi}_1(t)$	$\pi_1(t)$	$\hat{\pi}_2(t)$	$\pi_2(t)$	$\hat{\pi}_3(t)$	$\pi_3(t)$
1.	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000
2.	0.4000	0.4000	0.1000	0.1000	0.5001	0.5000
3.	0.3700	0.3700	0.3600	0.3600	0.2700	0.2700
4.	0.4450	0.4450	0.2710	0.2710	0.2840	0.2840
5.	0.4258	0.4258	0.2691	0.2691	0.3051	0.3051
6.	0.4233	0.4233	0.2795	0.2795	0.2972	0.2972
7.	0.4262	0.4262	0.2766	0.2766	0.2972	0.2973
8.	0.4256	0.4256	0.2763	0.2763	0.2981	0.2981
9.	0.4255	0.4254	0.2767	0.2767	0.2979	0.2979
10.	0.4256	0.4256	0.2766	0.2766	0.2978	0.2978

Note: Due to rounding, rows may not sum to one.

started from six different initial points. Each time the procedure converged to a matrix that showed good agreement with the test matrix. In no case was the aggregate squared error greater than 3×10^{-8}. For Problems Two and Three the results were similar, with no solution having a square error greater than 3×10^{-6}.

The above tests were repeated with the introduction of normally distributed random errors in addition to the rounding errors. These errors were about 70 times greater than the errors in part one, having a mean of zero and a standard deviation of 0.002.

In each of these three problems the NLP procedure converged to an answer that was slightly different than the original matrix. The average error by which the replicated data differed from the observed data was about 3.3×10^{-5}; in all cases (Table 3) the error was less than that obtained through the use of the original test matrix to replicate the data. This means that the procedure converged to a point that was slightly better than the original matrix, in the sense of giving better replication of the transients with randomized error.

The Distribution of Farms in North Dakota

The next step was to apply the procedure to some problems reported in the literature. The problem of Krenz [16], which dealt with the size distribution of farms, was selected as a test problem. Lack of micro data had forced

Table 3. Summed Squared Errors for Test Problems.

	NLP Solution Matrix	Prior Generating Matrix
Problem 1	4.29×10^{-5}	7.49×10^{-5}
Problem 2	1.61×10^{-5}	2.19×10^{-5}
Problem 3	4.00×10^{-5}	7.49×10^{-5}

Note: Uniform starting guesses were used in the NLP procedure.

Krenz to make simplifying assumptions. First many of the transition probabilities were arbitrarily fixed at zero or one. Then a set of simultaneous equations was solved to arrive at the estimates of the probabilities. When Krenz' matrix (Table 4, Matrix 1) is used to replicate the observed data based on the 1935 size distribution, a squared error of 0.00409 results. Application of the NLP procedure under the same constraints, starting from the uniform matrix, results in the Matrix 2 in the table. The squared error associated with this matrix is 0.00269. Complete relaxation of the Krenz constraints gives further improvement: Matrix 3, with a corresponding error of 0.00231. Comparison of this third matrix with Krenz' original matrix shows many more state-to-state transition possibilities than hypothesized by Krenz.

This difference brings up the question of the validity of Krenz' assumptions. While some of these are possibly made on sound agricultural-economics grounds, dictated by experience in the field, others are made for methodological convenience. The NLP procedure has the capability easily to handle assumptions that involve fixing some of the parameters, while leaving many others free to vary. It may be known from experience, for example, that once a farm reaches the greatest size (class 6) that its probability of remaining there is essentially one. The effects of this assumption may be tested using the NLP procedure. The best

Table 4. Solutions to Farm Distribution Problem

1) Krentz' Farm Distribution Matrix

1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1958	0.8042	0.0000	0.0000	0.0000	0.0000	0.0000
0.2554	0.0000	0.7446	0.0000	0.0000	0.0000	0.0000
0.1307	0.0000	0.0000	0.8301	0.0392	0.0000	0.0000
0.0997	0.0000	0.0000	0.0000	0.8491	0.0512	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.9420	0.0580
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

2) Constrained NLP Best-Fit Solution

1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2060	0.7940	0.0000	0.0000	0.0000	0.0000	0.0000
0.2619	0.0000	0.7381	0.0000	0.0000	0.0000	0.0000
0.0633	0.0000	0.0000	0.8337	0.1029	0.0000	0.0000
0.0949	0.0000	0.0000	0.0000	0.8474	0.0577	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.9415	0.0585
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

3) NLP Best-Fit Solution Matrix (Relaxed Constraints)

0.9806	0.0065	0.0000	0.0000	0.0000	0.0000	0.0129
0.2588	0.6945	0.0194	0.0044	0.0066	0.0000	0.0163
0.3530	0.0000	0.5995	0.0000	0.0475	0.0000	0.0000
0.1053	0.0000	0.0051	0.8229	0.0488	0.0077	0.0102
0.0340	0.0004	0.0313	0.0004	0.8402	0.0896	0.0041
0.0000	0.0086	0.0116	0.0010	0.0000	0.8963	0.0825
0.1206	0.0000	0.0000	0.0000	0.0000	0.0000	0.8794

4) One Constraint, $P_{66} = 1$

0.9428	0.0000	0.0000	0.0000	0.0000	0.0572	0.0000
1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1756	0.0000	0.6075	0.2170	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.6534	0.3466	0.0000
0.0000	0.0265	0.0405	0.0629	0.7610	0.0852	0.0240
0.1183	0.0564	0.0000	0.0000	0.0000	0.8004	0.0249
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

5) NLP Best-Fit Solution Matrix (Uniform Starting Matrix)

0.7091	0.0000	0.0000	0.0000	0.0001	0.1322	0.1586
0.1861	0.0433	0.1433	0.0566	0.2889	0.2240	0.0621
0.0063	0.0678	0.5006	0.0745	0.3508	0.0001	0.0000
0.1261	0.0456	0.1195	0.0632	0.3343	0.2627	0.0487
0.0000	0.0456	0.0424	0.0879	0.6147	0.2094	0.0001
0.2242	0.0000	0.0000	0.0000	0.0823	0.4660	0.2274
0.4629	0.0000	0.0000	0.0000	0.0000	0.3261	0.2109

fit matrix to the observed data with the inclusion of the assumption is shown by Matrix 4 in the Table; the error is 0.002159. With this procedure available, an agricultural economist would be able to test various sets of assumptions to define the minimum set that would give reasonable transition probabilities according to experience. Thus, micro data could be generated from macro data supplemented by "ballpark" knowledge about some of the micro data.

One final test was made on Krenz' farm problem. The NLP procedure was again used to estimate the transition matrix under complete relaxation of the fixed-parameter assumptions. Instead of starting at the solution to the fixed-parameter problem, the procedure was started at the uniform transition matrix. The result is Matrix 5 in the table. It is noted that this matrix differs from the others in several ways. It has several more non-zero elements, indicating more possible state transitions. Also the magnitudes of many elements have changed. This matrix and Matrix 3 were both arrived at under relaxation of the constraints, but they are very dissimilar. It seems unlikely that these matrices could be brought closer together through reduction in the stopping criterion and allowing the NLP procedure to run longer. The error associated with Matrix 5 is 0.00189. This means that the Matrices 1, 2, 3, 4 would have to become more like number five to achieve further reduction. Already a high degree of accuracy has been employed with these problems, and it may be that a local optimum exists in the neighborhood of these matrices. Such local optima were encountered in a similar problem by Saffer [31].

Cigarette Brand Switching Problem

Results for Telser's classical brand switching problem [34] are given in Table 5. The maximum likelihood estimate of Lee, Judge, and Zellner and the restricted least squares estimate of Theil and Rey are also given for comparison. Considerable difference is seen in these two matrices, on one hand, and in the NLP matrix on the other. Although a matrix resembling these two matrices has been arrived at from some starting points of the NLP procedure, the squared error associated with this matrix is more than that associated with the matrix shown. As shown in the table the NLP matrix has a lower squared error than either the maximum likelihood matrix or the restricted least squares matrix. Table 6 makes a comparison of the observed data with the replicated proportions given for each of the maximum liklihood, least squares and NLP solution matrices. A graph of the fit to the "Camels" curve is given in Figure 3 . Here, both the curve generated by the maximum liklihood and the NLP estimates are shown. Although the transition matrices differ considerably, the maximum liklihood and NLP estimates both produce similar fits to the observed data.

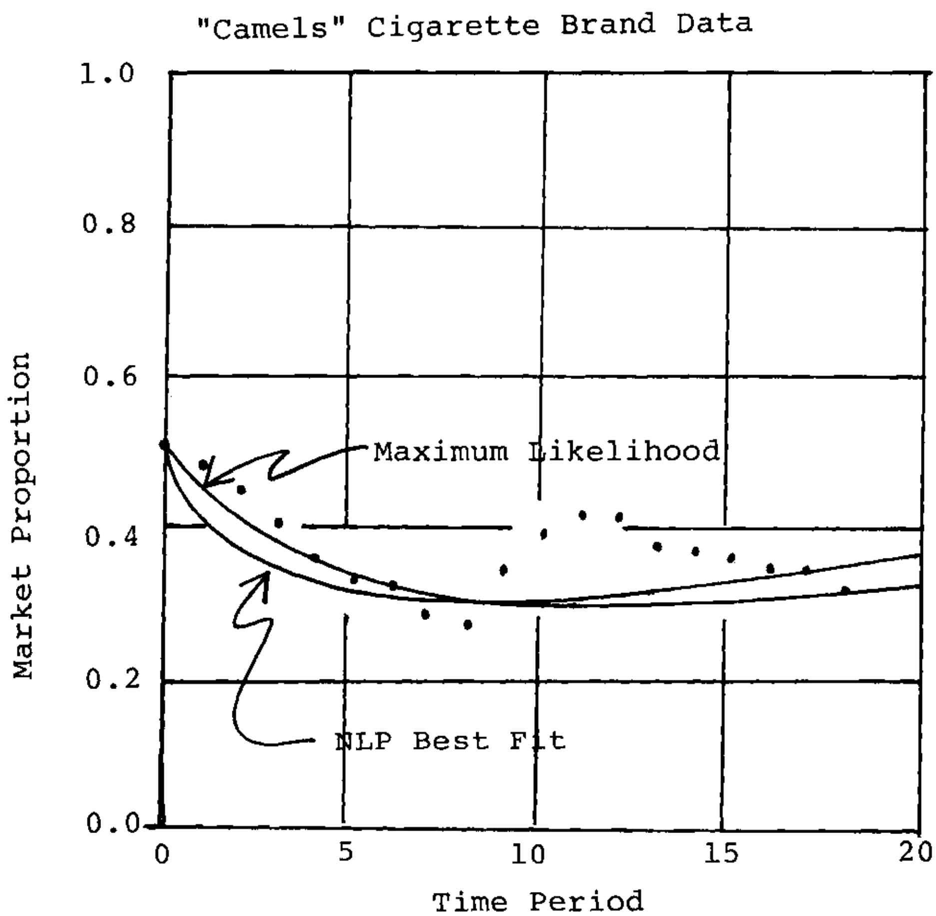

Figure 3. Data Projections from Time Zero for Various Transition Matrices.

Table 5. Solutions to Telser's Brand Switching Problem

Maximum Likelihood Estimate (L.J.Z)

0.6689	0.1425	0.1886
0.0000	0.8707	0.1293
0.4011	0.0000	0.5989

Total Squared Error = 0.09769

Restricted Least Squares

0.6686	0.1423	0.1891
0.0000	0.8683	0.1317
0.4019	0.0000	0.5981

Total Squared Error = 0.09644

NLP Best-Fit Matrix

0.7851	0.0000	0.2149
0.2162	0.7381	0.0457
0.0000	0.3333	0.6667

Total Squared Error = 0.08644

Table 6. Aggregate Market Shares: Cigarette Brands

			Camels	
Year	Actual	NLP	Maximum Likelihood	Least Squares
1925	0.5056	0.5056	0.5056	0.5056
1926	0.4879	0.4409	0.4552	0.4552
1927	0.4504	0.3995	0.4233	0.4236
1928	0.4068	0.3755	0.4016	0.4023
1929	0.3637	0.3630	0.3862	0.3872
1930	0.3365	0.3577	0.3750	0.3762
1931	0.3311	0.3561	0.3666	0.3681
1932	0.2936	0.3563	0.3604	0.3621
1933	0.2794	0.3572	0.3557	0.3576
1934	0.3418	0.3582	0.3522	0.3543
1935	0.3867	0.3591	0.3496	0.3518
1936	0.4074	0.3597	0.3476	0.3499
1937	0.4084	0.3600	0.3461	0.3485
1938	0.3842	0.3603	0.3450	0.3474
1939	0.3746	0.3604	0.3442	0.3466
1940	0.3708	0.3605	0.3435	0.3460
1941	0.3579	0.3605	0.3430	0.3456
1942	0.3527	0.3605	0.3427	0.3452
1943	0.3276	0.3605	0.3424	0.3450

			Lucky Strike	
Year	Actual	NLP	Maximum Likelihood	Least Squares
1925	0.2028	0.2028	0.2028	0.2028
1926	0.1899	0.2469	0.2486	0.2480
1927	0.2236	0.2863	0.2813	0.2802
1928	0.3039	0.3161	0.3053	0.3035
1929	0.3616	0.3361	0.3230	0.3208
1930	0.4118	0.3483	0.3363	0.3337
1931	0.4425	0.3551	0.3462	0.3433
1932	0.4498	0.3583	0.3537	0.3504
1933	0.4008	0.3551	0.3593	0.3558
1934	0.3301	0.3583	0.3636	0.3598
1935	0.3013	0.3596	0.3668	0.3629
1936	0.2906	0.3598	0.3692	0.3651
1937	0.2949	0.3595	0.3710	0.3668
1938	0.3195	0.3592	0.3723	0.3681
1939	0.3358	0.3588	0.3733	0.3691
1940	0.3500	0.3586	0.3741	0.3698
1941	0.3653	0.3584	0.3747	0.3703
1942	0.3851	0.3583	0.3751	0.3707
1943	0.3875	0.3582	0.3755	0.3710

Sales Forecasts for Home Heating Units

An additional problem to which the NLP procedure was applied dealt with the forecasting of market shares of alternative types of home heating devices. In this problem, Ali Ezzati employed the ML and Bayesian estimates to forecast future requirements for various types of heating units. Using his reported aggregate sales data, the NLP best-fit procedure was employed to arrive at several different estimates of the matrix. When the procedure was started at the informationless uniform matrix it produced an estimate having a squared error of 0.0273. This compares to errors of 0.0308 and 0.0589 for his maximum liklihood and Bayesian estimates. When the NLP procedure used these estimates as starting points, an error of 0.0223 resulted for both of the resulting solutions. The three resulting matrices and Ezzati's solution matrices are found in Table 7. Quite good agreement is found among these matrices.

While extrapolation using NLP matrices is not necessarily valid, Table 8 shows close agreement with forecasts obtained through the use of the asymptotically efficient unbiased, asymptotically normal, consistent ML estimates.

One additional fit to Ezzati's data was generated. From the data it appeared that the market may have undergone a change in the mid 1950's. The increased usage of electric units from 1959 onward suggests that the process was not stationary over the time period 1955-1965, but rather that

Table 7. Transition Probability Matrices for Sales of Home Heating Devices

	NLP Procedure			NLP Procedure			NLP Procedure		
	Oil	Gas	Electric	Oil	Gas	Electric	Oil	Gas	Electric
Oil	0.8160	0.1840	0.0000	0.8165	0.1835	0.0000	0.8165	0.1835	0.0000
Gas	0.0517	0.9157	0.0326	0.0514	0.9160	0.0327	0.0513	0.9160	0.0327
Electric	0.1144	0.0000	0.8856	0.1149	0.0000	0.8851	0.1149	0.0000	0.8851

	Ezzati's ML			Ezzati's Bayesian		
	Oil	Gas	Electric	Oil	Gas	Electric
Oil	0.8250	0.1750	0.0000	0.8563	0.1025	0.0412
Gas	0.0600	0.9190	0.0210	0.0513	0.9372	0.0115
Electric	0.0490	0.0000	0.9510	0.0001	0.0536	0.9463

Table 8. Forecasts of Alternative Types of Home Heating Units (Based on 1960 Actual Sales).

	Oil Burners			Gas Burners			Electric Heat		
	NLP	ML	Bayesian	NLP	ML	Bayesian	NLP	ML	Bayesian
1980	0.2609	0.2460	0.2099	0.5783	0.5658	0.5629	0.1608	0.1882	0.2272
1981	0.2612	0.2458	0.2086	0.5775	0.5635	0.5612	0.1613	0.1907	0.2301
1982	0.2614	0.2456	0.2075	0.5769	0.5613	0.5597	0.1617	0.1931	0.2328
1983	0.2617	0.2455	0.2064	0.5764	0.5593	0.5583	0.1620	0.1952	0.2353
1984	0.2618	0.2453	0.2054	0.5759	0.5574	0.5570	0.1622	0.1973	0.2376
1985	0.2620	0.2452	0.2054	0.5756	0.5556	0.5558	0.1624	0.1992	0.2397

the introduction of electric units was influencing the market. Accordingly, another transition matrix was calculated with disregard for the data in the period 1955-1957. The total squared error associated with this matrix was 0.00794, giving an average error per observation of 0.00022. This compares to an average error of 0.00049 when the time period is included.

Sampling Characteristics: Simulated Aggregate Data

In reviewing the traditional estimation techniques available for aggregate data, Lee, Judge and Zellner [19], generate a data base of 1000 individuals from the Monte-Carlo simulation. The aggregate time series that results from sampling this data base forms the basis for the next problem. The NLP procedure was applied to their data for time periods 2-14. The matrix that resulted from a uniform starting point is given in Table 9. The ML and Bayesian (leptokurtic prior) estimates developed by Lee, Judge and Zellner are also shown here. The squared error that results when the NLP matrix is used to replicate the observed data is 0.00244. This compares to errors of 0.00314 and 0.00435 for the Ml and Bayesian matrices.

As a test of their matrices, Lee, Judge and Zellner provide a Chi-square test that depends on the aggregate data. In their application the data at each time period are estimated from the data at the previous time period. Although this is not the case with our method of computing

Table 9. Results for the Simulated Data of Lee, Judge and Zellner.

Maximum Likelihood Estimates

0.5613	0.3540	0.0847	0
0.1212	0.5358	0.3430	0
0	0.0879	0.7209	0.1912
0	0	0.0861	0.9139

Bayesian Estimates (Leptokurtic prior)

0.6001	0.3999	0	0
0.0993	0.4992	0.4015	0
0	0.0981	0.7022	0.1997
0	0	0.0993	0.9006

NLP Best Fit (Uniform start)

0.6195	0.1654	0.2154	0
0.0664	0.7139	0.2196	0
0.0130	0.0436	0.7461	0.1973
0	0	0.0894	0.9106

the projected proportions, the Chi-square test can still be used. Essentially the statistic is used to test the hypothesis that the observed (actual) data could have resulted from the process described by the estimated transition matrix. The form of the statistic is given below.

$$\chi^2_{(r-1)T} = \sum_{t}^{T} \sum_{i}^{r} N(t)\left(y_i(t) - \hat{y}_i(t)\right)^2 / \hat{y}_i(t)$$

where the observed proportions are the $y_i(t)$, the estimated proportions are the $\hat{y}_i(t)$, and $N(t)$ denotes the number of individuals involved. T is the number of time periods for which the prediction is made, and r is the rank of the Markov matrix. Based on the value of the statistic the observed proportions are classified as usual or unusual. If the outcomes are found to be usual it may tentatively be concluded that the observed proportions are not the outcome of the process described by the Markov matrix. This reflects back on the Markov matrix to say that perhaps the estimate is not a good one. This is a weak conclusion since it is quite possible for unusual observations to be obtained from the actual process.

The values for the statistics obtained in this manner are as follows: For the ML estimate of Lee, Judge and Zellner a value of 15.5; for the use of this estimate to project future time periods from period 2, a value of 14.5; and for the best NLP projections from period 2, a value of

13.5. When the NLP estimate is used to make projections in their linear manner, the value is 17.6. All of these estimates are well below the critical value of 47.1 for the Chi-square with 36 degrees of freedom at the 10% level. This is consistent with the hypothesis that the aggregate (observed) data are outcomes of the process represented by the above transition matrices.

It should be noted here that the use of the Chi-square statistic to measure the goodness of fit of the predicted proportions to the observed proportions when the predicted proportions are based on a single earlier time period deviates slightly from the use made by Lee, Judge, and Zellner, where the predicted proportions are measured from the most recent observed proportions.

The Rose-Bengal Transport Mechanism

Saffer's investigation of Rose-Bengal transport through the liver [30] provides a unique type of problem. Because of the intermitent nature and general lack of data, standard ML or least squares techniques could not be applied. To counter this situation Saffer turned to a formulation similar to that suggested in this research. He attempted to find the set of transition probabilities that produced a transient solution giving a good fit to that data that was available. First, he simplified the problem by fixing many of the transition probabilities. He then applied a combination of Hooke and Jeeves pattern search with Gauss least squares to estimate the remaining parameters. The application of the NLP steepest descent gradient method gave identical results to this constrained problem. (Table 10). The objective value was 0.00409.

Relaxation of the fixed parameter constraints makes the problem somewhat more difficult because more parameters require estimation. It also increases the potential for a good fit. When the NLP procedure was restarted at the solution to the constrained problem a further reduction of the objective function to 0.00405 was achieved, while maintaining the same level of accuracy in the termination criterion.

When an attempt was made to solve the unconstrained version of the problem directly (started from the uniform matrix) very slow convergence was encountered. After many iterations the value of the objective had dropped only to 0.00602. This formulation of the problem appeared to be particularly difficult from the standpoint of computational time. First, the large number of time periods (72) combined with the order of the matrix (4) to make functional evaluations very time consuming. Second, the inclusion of the additional variables changes the direction of steepest descent so that some of the zero/one variables may move far from that value for a time. While this improves the direction of steepest descent at each step, it does not make this direction point more accurately toward the optimal solution.

Table 10. Solutions to the Rose-Bengal Problem

Constrained (0,1) Solution Matrix (NLP, Saffer)

0.6373	0.0204	0	0.3423
0	1	0	0
0	0	1	0
0.0377	0	0.0101	0.9532

Relaxed Solution Matrix (NLP)

0.6362	0.0219	0	0.3419
0	0.9959	0.0041	0
0	0	1	0
0.0385	0.0003	0.0095	0.9517

Independent (Uniform Start-unconstrained) Matrix

0.4487	0.4777	0	0.0736
0.1991	0.6482	0	0.1528
0	0.0052	0.9948	0
0	0.0296	0.0127	0.9577

In looking back it can be seen that a priori fixing of some of the parameters at, or close to, their optimal value improved the convergence of the algorithm. Even when the constraints were relaxed, the values of the fixed parameters did not move far from their earlier zero/one values. (Table 10, Matrix 2)

In attempting to improve the convergence of the procedure when this a priori information is not available several alternative non-linear programming procedures were investigated. A cyclic coordinate search method did not work well for large problems since time-consuming line searches were required in each direction. A pairs adjustment procedure also had the same problem. As the rank of the matrix grew, the number of pairs of p_{ij}s in a row grew faster. These were discarded in favor of the gradient method which required only a single line search. Two attempts were made at modification of the basic gradient procedure for faster convergence. The first was the conjugate gradient algorithm of Fletcher and Reeves. This approach generates new search directions based on the current direction of steepest descent and on the past movement and gradient information. When applied to the above problem this method showed no clear advantage over the method of steepest descent. The second modification, a fixed-step approach which eliminated the time-consuming line search in favor of updating a step distance also was tried without success.

Saffer [31] reported a similar wide-ranging investigation of solution techniques before arriving at his combination of simplifying assumptions, pattern search and Gauss least squares. Since Saffer's choice produced identical results to the constrained Rose-Bengal problem the question arises as to whether combination of the Gauss least squares with the gradient search would provide a good method for general solution of the unconstrained problems. Preliminary experimentation with the unconstrained version of Saffer's problem indicated that this was not the case. The unconstrained Gauss least squares procedure requires that the initial point be fairly close to a local minimum. Otherwise the estimates given by the procedure tend to be extremely large, violating feasability. Saffer reports similar difficulties in the application of this technique.

Estimating Parameters with Incomplete Data

In this section the performance of the NLP method for dealing with incomplete data will be investigated. Two types of incompleteness will be considered. In the first type, complete data are available on one or several states and no data exists on the proportions in the remaining states. In the second type, data is available about each state, but not at each time period. Combinations in which some data are available about all states at differing time periods will also be considered.

The approach taken here is to use some of the data that were presented in the earlier sections and to compare the estimates derived after discarding some of this data with the estimates previously derived under consideration of the full data.

The first type of incompleteness occurs when data are available about one state, but where no data are available about other states. It is assumed that the total number of states is known. Such a case could occur in industry. A manufacturer may know the total sales of his firm, and he may also know the total sales of industry as a whole. He may not know the sales of his competitors, and so will be unable to determine their market proportions. He will be able to determine the proportion of the total industry that his sales make up.

Telser's cigarette problem was selected as a test problem for this situation. The NLP procedure was used to derive estimates of the parameters based only on the aggregate proportions in the "Camels" state. Several runs were made. In the first it was assumed that the actual proportions at the time zero were known. This might be the case when census-type data are collected every ten years. Such a year could be used as a starting point for the estimation procedure. In the second and third runs this assumption was not made. In the second run proportions remaining after subtracting the known proportion from one were divided unequally among the other two states. In the third run, one state was assigned a zero value and the other state took up the slack to meet the unity requirement.

The solution matrices to these problems are given in Table 11. The previous NLP estimate, derived under complete data, and the maximum likelihood estimate of Lee, Judge and Zellner are also given for comparison. A great deal of similarities exist between these matrices. In particular it does not appear that accurate knowledge of the $\pi_i(0)$ is necessary. This can be seen from a comparison of matrix 1 with matrices 2 and 3.

It also should be noted that the parameters of particular interest to the manufacturing decision-maker remain surprisingly stable across the five matrices. The elements P_{11} are all fairly close together. Calculation of the steady state conditions can be done easily by the decision-maker since they do not depend on the $\pi_i(0)$. These are given in Table 12.

Notice that, although the steady state proportions associated with states 2 and 3 vary considerably, the parameter of interest, the market share of company 1, "Camels" is fairly constant across the matrices. This means that the NLP technique can be used to project the market share of a company when only the market proportions for that company are known. The decision-maker does not require information on the market proportions of his competitors, or on the

Table 11. Solution Matrices to Telsers Cigarette Problem with Incomplete Data.

Matrix 1 - Initial $\pi_i(0)$ Known

0.6998	0.3002	0
0	0.7483	0.2517
0.4802	0	0.5198

Matrix 2 - $\pi_1(0)$ Known $\pi_2(0)$, $\pi_3(0)$ divided up

0.6914	0.3086	0
0	0.7261	0.2739
0.4467	0	0.5533

Matrix 3 - $\pi_1(0)$ Known $\pi_2(0) = 0$, $\pi_3(0) = 1-\pi_1(1)$

0.6813	0.3187	0
0	0.6104	0.3896
0.3179	0	0.6821

Matrix 4 - Solution to previous problem - all data known

0.7851	0	0.2149
0.2162	0.7381	0.0457
0	0.3333	0.6667

Matrix 5 - ML estimate of Lee, Judge, & Zellner - all data known

0.6689	0.1425	0.1886
0	0.8707	0.1293
0.4011	0.0000	0.5989

Table 12. Steady State Market Projections

Matrix 1 - Initial $\pi_1(0)$ Known		
0.3549	0.4233	0.2219
Matrix 2 - $\pi_1(0)$ Unknown		
0.3549	0.3999	0.2452
Matrix 3 - $\pi_1(0)$ Unknown		
0.3545	0.2900	0.3554
Matrix 4 - All data known		
0.3604	0.3582	0.2815
Matrix 5 - ML estimate of Lee, Judge, and Zellner		
0.3417	0.3763	0.2820

proportions at time zero.

As might be expected, the errors resulting from the fit to the single state were very similar for each of these matrices. They were, for matrices 1, 2, 3, 4, and 5, 0.01860, 0.01840, 0.01823, 0.02458, and 0.02806, respectively. Obviously the requirement that the extra data in curves 2 and 3 be considered has had a greater effect than the absence of information about $\pi_i(0)$.

The above method appears to work well when the manufacturer knows the number of competitors, and so, is able to determine the rank of the Markov matrix. This is not always the case. When a new competitor enters the field, there is some uncertainty about the point at which he should be considered major enough to be given a place in the model. This point was addressed by running the same problem with differing numbers of competitors. Four and five state Markov chains were postulated. The steady state results of three runs are given below:

Run 1 (4 x 4) (0.3550 0.1680 0.2647 0.2123)

Run 2 (4 x 4) (0.3561 0.1765 0.2458 0.2216)

Run 3 (5 x 5) (0.3630 0.1763 0.1487 0.1755 0.1364)

$\pi_i(n)$ again shows good agreement with the values in Table 13. This means that it is not necessary to know the number of market competitors. A reasonable estimate of the rank of the process will suffice.

In the above discussion it has been assumed that complete time-ordered information on the state in question has been available. "Type two" incompleteness results when some data are available on some or all states, but not at all time periods. This was the case in the Rose-Bengal transport problem where complete time-ordered samples could not be obtained.

As a further investigation into this type of error the home-heating-fuel problem [5] will be used for a test. Earlier, the transition parameters for this problem were estimated by the use of the NLP technique and complete data. Successive observations have been deleted from this set of data and the NLP estimate was determined at each stage. Of the original 45 data points (15 time periods by 3 states) 21 were initially deleted, leaving 24. Subsequent deletions left 18, 14, and 7 data points. The resulting NLP estimates are given in Table 13. Matrices 1, 2, 3, and 4 are almost identical; it is not until the data are reduced to only 7 data points that changes begin to take place in the matrix. Errors associated with replication of the complete set of data by these matrices are 0.02230, 0.02341, 0.02326, 0.02221, and 0.07208 for the five matrices, respectively. This demonstrates the robustness of the NLP procedure when data are incomplete.

Table 13. Home Heating Fuel

Matrix 1 - Complete Data (45 Points)

0.8160	0.1840	0
0.0517	0.9157	0.0326
0.1144	0	0.8856

Matrix 2 - 24 Data Points

0.7951	0.2049	0
0.0678	0.9049	0.0273
0.0808	0	0.9191

Matrix 3 - 18 Data Points

0.8269	0.1731	0
0.0516	0.9200	0.0284
0.0919	0	0.9081

Matrix 4 - 14 Data Points

0.8382	0.1618	0
0.0500	0.9235	0.0265
0.0792	0	0.9208

Matrix 5 - 7 Data Points

0.8117	0.1883	0
0.0763	0.8251	0.0987
0	0.4393	0.5607

Computational Experience

In the search for a good method of solving the nonlinear programming problem several methods were tried. Each of these were evaluated on a number of problems with some methods receiving considerable attention and some being rejected almost immediately. One cyclic coordinate method was tried on over a hundred problems before it was rejected. Different starting points and different convergence criteria were tried. Methods of adjustment of these parameters were developed. It was found that the cyclic method worked best when the coordinate directions were ranked according to their potential for improving the objective. After many evaluations the cyclic coordinate approach was discarded in favor of a gradient method. Although the gradient information was not available, it could be approximated through use of the forward and rearward difference equations. The extra time required to make these approximations was more than balanced by the savings associated with a single line search.

Even with the improved method computational times remained quite high with 10 minutes of CPU time and 80,000 functional evaluations being common occurrences when all elements of the gradient were required to be less than 1×10^{-5}. Sharper convergence, of course, required even longer times; conversely, computational savings, but loss of accuracy were associated with convergence to gradients of only 1×10^{-3}.

The performance of the method varied considerably over the problems solved. Problems with more states and more time periods required more time than smaller problems. The number of parameters to be estimated played an important part in the convergence, with the constrained Rose-Bengel problem converging in quite a reasonable amount of time (162 seconds) and only 8,500 functional evaluations (including those required to determine the gradient). The unconstrained Rose-Bengal problem was termined due to slow convergence (at the 1×10^{-5} gradient level) after 2,400 seconds and 80,000 functional evaluations.

Computational time also depended on the choice of starting guess and search parameters. There was no apparent way to pick the best starting guess by a priori analysis of the data. It was found, however, that for fully determined (complete data) problems the uniform starting matrix generally worked well. For severally underdetermined problems, however, it was difficult to pick a worse starting point than the uniform matrix. This was because the gradient evaluation procedure was unable to choose between equally attractive or unattractive states. Unless information was available about more than one state, either at time zero or at later times, the uniform matrix would not work as a starting matrix. For these problems arbitrary starting guesses, with the rows of the matrix as dissimilar as possible, performed best.

In the first few steps the gradient method reduced the objective function by a considerable amount (Figure 4), but then convergence slowed. This could be helped by appropriate choices of the gradient determination step parameter Δ, and of the line search termination criterion. A Δ of 0.0001 worked well in the initial phases, with later reduction to 0.00000001 required to get convergence. In the beginning, the gradient (steepest descent) direction did not necessarily point toward the true optimum, so inaccuracies in the direction were unimportant. With a Δ of 0.0001 only those parameters that could contribute significantly to the reduction of the objective were changed. As the procedure approached the optimum, however, more accurate information was required.

Similarly, early in the procedure a coarse line search sufficed; later, however, greater accuracy was required. Typical values for the line search parameter ranged from 1×10^{-2} to 1×10^{-5}. The computer program was set up so that the user could periodically adjust these parameters.

Finally, in the immediate neighborhood of the optimum, oscillation could occur at the bottom of a steep valley. Here the gradient elements change sign from one iteration to the next. At this point some method that does not move in the direction of steepest descent would be preferred.

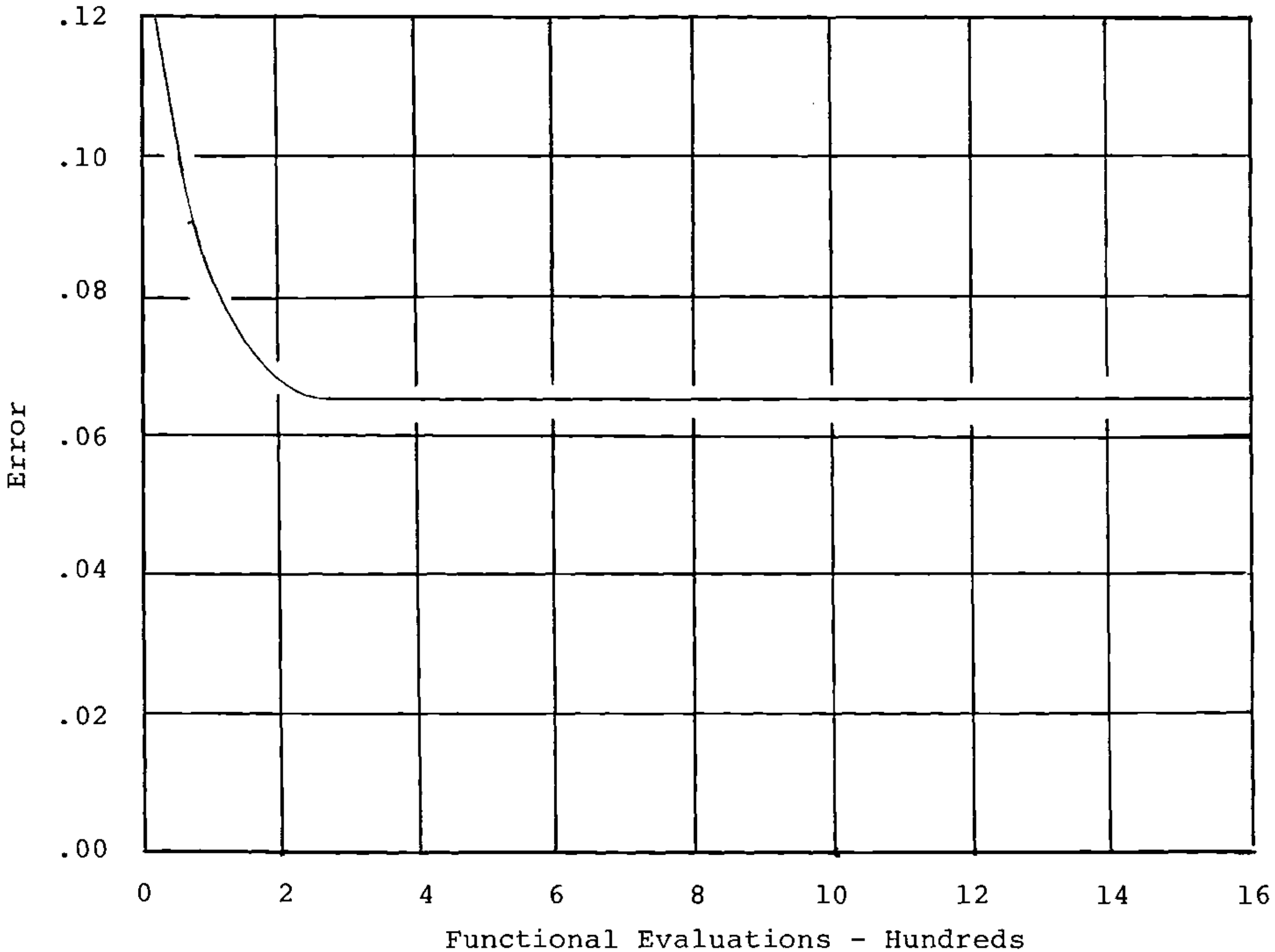

Figure 4. Convergence for the Unconstrained Rose-Bengal Problem.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

In the preceding chapters the NLP method for estimating the parameters of a Markov process from aggregate data was developed and applied to several applications problems, and comparisons were made with estimates arrived at in the traditional manner.

Overall Performance

In the problems of Telser [34] and Ezzati [5] and in the simulated experiment of Lee, Judge and Zellner [18] the procedure reproduced observed data quite well and gave better curve fits than traditional methods. In one case projections were made using both the NLP estimate and the asymptotically efficient, unbiased, asymptotically normal, consistent maximum likelihood estimate. While the NLP estimate does not have these properties, close agreement was found.

The applicability of the NLP procedure was extended by allowing the investigator to fix certain elements in the transition matrix. In this way assumptions or knowledge about the transitions of the micro data may be examined or exploited. The investigator can test varying assumptions based on personal experience and background which could not readily be tested with traditional procedures.

The ability of the NLP procedure to handle incomplete aggregate data was demonstrated for several problems. In the Rose-Bengal Transport problem [31] some data were available for each state, but these observations were arbitrarily distributed along the time axis. Standard techniques could not cope. The NLP procedure was compared to an independently developed non-linear programming procedure; the results were identical.

Investigation of the performance of the NLP in solving problems of this type was made with two additional problems. For the Home Heating Unit problem, previously solved with complete data, several tests were made in which larger and larger numbers of observations were discarded. The results arrived at in this manner were nearly as good as the results obtained with full data. Errors in replicating observed data in all cases were less than obtained by traditional methods. The matrices themselves closely resembled the matrix obtained under the maximum likelihood approach.

In the brand switching problem a second type of incomplete data was investigated. Here, it was assumed that information was available on only one state. It was shown that the steady state proportion for this curve can be accurately derived in the absence of information about the initial state proportions, the $\pi_i(0)$, or the actual size of the state space.

In summary, the NLP technique can provide four things:

1) In the presence of complete information the technique provides as good an estimate as the traditional methods with respect to replicating the observed data. Use of the squared error metric is not required.

2) The technique provides the capability, not available with traditional methods for the modeler to test assumptions about the behavior of the micro data.

3) In the absence of complete information about all states and all time periods the technique is still able to obtain a good estimate of the transition matrix.

4) In the presence of information about only one state the technique is able to provide an estimate of the transition matrix that gives a good fit to the steady state solution for the known state.

Recommendations for Further Research

Two areas for further development have arisen from this research, both aimed at improving the performance of the NLP procedure in solving real problems. As was pointed out in Chapter IV the NLP approach sometimes experiences stalling of the convergence sequence. This occurs when the procedure is moving along the bottom of a narrow valley. Elements of the gradient change sign from positive to negative and back again as new points are accepted. Essentially the steepest descent direction causes the procedure to jump back and forth across the valley rather than proceeding down the valley. This is a typical problem with the method of steepest descent.

Although some other algorithms were tried in this research and also independently by Saffer, no systematic survey has been made of the ability of the non-linear programming algorithms currently available to solve problems of this type. At the cost of increased computer storage and algorithmic complexity the variable metric method of Davidon, Fletcher, and Powell which has been demonstrated to perform well on a number of problems [10] could be employed. This method makes a gradual changeover from the direction of steepest descent to Newton's direction as information on the second derivatives is accumulated. The search methods of Powell [26] and [27] which do not require first derivatives at each stage also show promise.

A second area of potential payoff is in obtaining a good starting guess of the transition matrix. Any non-linear programming method will perform better if started close to the optimal solution. It seems possible that some guess of the optimal solution may be made subjectively by the experimenter. Ideally this would not require the use of a complicated computational scheme, which would involve as much work as obtaining a reasonable estimate from a poor one via the non-linear programming approach, but rather would be based on the general shape of the data curves.

Lore on the analytic geometry of Markov transient solutions gained in the course of this research suggests that, with experience, one could sometimes predict which elements in the transition matrix should be large or small. The high degree of interaction among the elements in the transition matrix in generating curves, however, tends to make this procedure somewhat imprecise. A specific effort could be made to generate a "dictionary" of general curve shapes from which a likely starting guess could be picked.

APPENDIX

FORTRAN LISTINGS FOR MAJOR SECTIONS OF THE NON-LINEAR PROGRAMMING CODE

```
50     CYCLCST=OBJB
       INDEXS(1) = 0
C        ***********************************************************
C        *                                                         *
C        *    BEGIN MAJOR MATRIX-ELEMENT LOOP:  INITIALIZE  VARIABLES *
C        *                                                         *
C        ***********************************************************
C
       NCYCL=NCYCL     +     1
       POINT1 = 0.0
       END1 = 0.0
       STEP = .1
C               SET UP MATRIX OF TRIALP ELEMENTS
       DO 80 I2=1,MATDIM
       DO 80 J2=1,MATDIM
       PBASE(I2,J2) = P(I2,J2)
       TRIALP(I2,J2)=P(I2,J2)
80     CONTINUE
C ***  ***********************************************************
C ***  *                                                         *
C ***  *       SET UP GRADIENT MATRIX                            *
C ***  *                                                         *
C ***  ***********************************************************
```

```
C *** ************************************************************
C *** *                                                          *
C *** *       SET UP GRADIENT MATRIX                             *
C *** *                                                          *
C *** ************************************************************
      DO 90 J2=1,MATDIM
      DO 90 I2=1,MATDIM
      TRIALP(I2,J2) = P(I2,J2) +DELTA
      OBJT = F(TRIALP,NCYCL,D,MATDIM,ICALCT)
      NFUNCE = NFUNCE + 1
      GRAD(I2,J2) = (OBJB - OBJT) / DELTA
      TRIALP(I2,J2) = P(I2,J2)
      IF(GRAD(I2,J2) .GT. 0) GO TO 90
C *** TRY A STEP IN THE OPPOSITE DIRECTION
      TRIALP(I2,J2) = P(I2,J2) - DELTA
      IF( TRIALP(I2,J2) .LT. 0) TRIALP(I2,J2) = 0.0
      OBJT2 = F(TRIALP,NCYCL,D,MATDIM,ICALCT)
      NFUNCE = NFUNCE + 1
      GRAD(I2,J2) = (OBJT2 - OBJB)  /  DELTA
      IF( GRAD(I2,J2) .GE. 0.) GRAD(I2,J2) = 0.0
      TRIALP(I2,J2) = P(I2,J2)
90    CONTINUE
C ***    EVALUATE THE GRADIENT MATRIX TO SEE IF WE SHOULD STOP
      DO 95 I2=1,MATDIM
      DO 95 J2=1,MATDIM
      IF(ABS(GRAD(I2,J2)).GE. GRDSTP) GO TO 100
95    CONTINUE
C ***    ALL GRADIENT ELEMENTS ARE .LE. STOPPING CRITERION
C ***    WRITE OUT INFORMATION IN PREPERATION FOR STOPPING
      DO 96 I2=1,MATDIM
      WRITE(7,104)(GRAD(I2,J2),J2=1,MATDIM)
96    WRITE(6,380)(GRAD(I2,J2),J2=1,MATDIM)
      WRITE(6,97)
97    FORMAT(* THE FINAL P-MATRIX *)
             OBJB=F(P,NCYCL,D,MATDIM,ITIME)
             DO 99 I2=1,MATDIM
             WRITE(6,380)(ADJP(I2,J2),J2=1,MATDIM)
             WRITE(7,380)(ADJP(I2,J2),J2=1,MATDIM)
99           CONTINUE
             WRITE(7,2511)OBJB
             WRITE(6,2511)OBJB
             READ(5,*)ISSTOP
             IF(ISSTOP .LE. 0)GO TO 8000
             GO TO 1850
8000         CONTINUE
             WRITE(7,8010)GRDSTP
8010         FORMAT(* ENTER DIVISION FACTOR FOR GRADIENT-TESTING*
     1       * CRITERION*,/,* CURRENT CRITERION EQUALS*,E12.4)
             READ(5,*)GRADDIV
             IF(GRADDIV .LE. 1)GO TO 1850
             GRDSTP  = GRDSTP  / GRADDIV
100   CONTINUE
105   CONTINUE
      IMPROV = 0
C *** *********************************************************
C *** *                                                       *
C ***    AT LEAST ONE GRADIENT ELEMENT IS NOT CLOSE TO ZERO
C ***    MOVE ALL ELEMENTS OF P-MATRIX ACCORDING TO THE GRADIENT
C *** *                                                       *
C *** *********************************************************
Cca
```

```
109   CONTINUE
C *** ***********************************************************
C *** *                                                         *
C *** *          SET UP THE GOLDEN SECTION INTERVAL             *
C *** *                                                         *
C *** ***********************************************************
C
C
      DO 110 I2=1,MATDIM
      DO 110 J2=1,MATDIM
      TRIALP(I2,J2) = P(I2,J2) + GRAD(I2,J2)*STEP
110   CONTINUE
112   CONTINUE
C *** TEST TO MAKE SURE THAT ALL ELEMENTS OF TRIALP ARE .GE. ZERO
      DO 120 I2=1,MATDIM
      DO 120 J2=1,MATDIM
      IF(TRIALP(I2,J2) .GE. 0)GO TO 120
C
C ***    THIS ELEMENT IS LESS THAN ZERO AND MUST BE ADJUSTED
C ***    TO AVOID COMPUTER ROUND-OFF ERRORS, SET THIS ELEMENT = 0 !
      TRIALP(I2,J2) = 0.0
120   CONTINUE
C ***    NOTE!  ROWS DO NOT NECESSARILY SUM TO 1.0
C ***    THIS CONSTRAINT IS PRESERVED IN THE FUNCTIONAL EVALUATION
C ***    BY NORMALIZING BY THE ACTUAL SUM OF THE ROWS
C
C ***    NOW EVALUATE THE OBJECTIVE AND COMPARE TO PREVIOUS VALUE
      OBJF = F(TRIALP,NCYCL,D,MATDIM,ICALCT)
      NFUNCE = NFUNCE + 1
C
      IF(OBJF .LT. OBJB)GO TO 130
C ***    THE POINT CHECKED WAS A FAILURE,REDUCE THE STEP SIZE
C ***    SINCE THE GRADIENT SHOWS US THE DIRECTION TO MOVE,
C ***    WE DO NOT NEED TO TRY A NEGATIVE STEP SIZE
C ***    THE POINT CHECKED PROVIDES THE SECOND END OF THE GOLDEN
C ***    INTERVAL(ZERO, POINT CHECKED)
      END2=STEP
      GO TO 145
130   CONTINUE
C ***    THE POINT CHECKED WAS A SUCCESS, THE STEP SIZE MUST BE INCREASED
      IMPROV = 1
C ***    SET UP THE BASIC P-MATRIX TO REMEMBER
      DO 135 I2=1,MATDIM
      DO 135 J2=1,MATDIM
      PBASE(I2,J2) = ADJP(I2,J2)
135   CONTINUE
      END1 = POINT1
      POINT1 = STEP
      STEP=STEP*3.
      OBJB = OBJF
C *** GO BACK AND TRY AGAIN--EVALUATE THE NEW POINT UNTILE A FAILURE IS
C *** REACHED
      IF(NPOUT(8) .EQ. 1)GO TO 220
C
      GO TO 105
```

```
      FUNCTION F(Y,NDUMMY,D,MATDIM,ITIME)
      COMMON /BHAT/ QC(50)
C *** PROCEDURE INVOLVED IN MINIMIZATION.
      DIMENSION Y(10,10),D(10,51)
      COMMON /ADJUST/ ADJP(10,10),ROWSUM(10),P(10,10)
      COMMON PI(10,51)
      COMMON /FLAGS/ LOGODS,INIT,IFIRST,ISET,KOPOUT(20),INDEXS(10)
      IF(IFIRST .EQ. 0)GO TO 8
      DO 6 I=1,MATDIM
      DO 6 J=1,ITIME
6     PI(I,J)=0.0
      WRITE(7,600)MATDIM
600   FORMAT(" ENTER THE",I3,"  PI(I,1)")
      READ(5,*)(PI(I,1),I=1,MATDIM)
      IFIRST=0
8     CONTINUE
      DO 10 J=2,ITIME
      DO 10 I=1,MATDIM
      PI(I,J)=0
10    CONTINUE
C *** SET UP INITIAL ADJUSTED P-MATRIX
      DO 14 I2=1,MATDIM
      DO 14 J2=1,MATDIM
      ADJP(I2,J2) = Y(I2,J2)
14    CONTINUE
C **J NOW COMPUTE ADJUSTED P MATRIX:  SUM THE ROWS
      DO 420 II=1,MATDIM
      ROWSUM(II)=0.
      DO 420 JJ=1,MATDIM
420   ROWSUM(II)=Y(II,JJ)+ROWSUM(II)
      DO 430 II=1,MATDIM
      DO 430 JJ=1,MATDIM
430   ADJP(II,JJ)=Y(II,JJ)/ROWSUM(II)
C *** NOW THE PI AT TIME T=1,ITIME WILL BE CALCULATED
      DO 30 J=2,ITIME
      DO 30 I=1,MATDIM
      DO 30 K=1,MATDIM
30    PI(I,J)=ADJP(K,I)*PI(K,J-1)+PI(I,J)
C *** NOW FIND SQUARED ERROR RESULTING FROM PI CALCULATION
      ERROR=0.0
      BETA=1.000000
      DO 50 JK=1,INIT
      DO 50 K=1,ITIME
50    ERROR=ERROR+(PI(JK,K)-D(JK,K) )**2.*BETA**K
      F=ERROR
      RETURN
      END
 -END OF FILE-
?
```

```
      GO TO 105
C ***  ***********************************************************
C ***  *                                                         *
C ***  *          BEGIN GOLDEN SECTION SEARCH ON STEP-SIZE       *
C ***  *                                                         *
C ***  ***********************************************************
145   CONTINUE
      DIFFSAV = END2 - END1
150   CONTINUE
C ***    THE ENDS OF THE GOLDEN SECTION HAVE NOW BEEN SET UP
C ***    PERFORM A GOLDEN SECTION ON THE INTERVAL(END1, END2)
      DIFF = END2 - END1
      GOLDEN =.382
      IF(POINT1 .LE. END1+.5*DIFF) GOLDEN = .618
      TSTEP = END1 + GOLDEN*DIFF
C ***    SET UP A NEW TRIALP-MATRIX USING THIS STEP SIZE
      DO 160 I2=1,MATDIM
      DO 160 J2=1,MATDIM
      TRIALP(I2,J2) = P(I2,J2) + TSTEP*GRAD(I2,J2)
160   CONTINUE
C ***    MAKE SURE ALL ELEMENTS ARE NON-NEGATIVE
      DO 170 I2=1,MATDIM
      DO 170 J2=1,MATDIM
      IF(TRIALP(I2,J2) .GE. 0)GO TO 170
      TRIALP(I2,J2)= 0.0
170   CONTINUE
C
      OBJF = F(TRIALP,NCYCL,D,MATDIM,ICALCT)
      NFUNCE = NFUNCE + 1
      IF(OBJF .LE. OBJB)GO TO 200
C *** FAILURE
      IF(TSTEP .GT. POINT1)END2=TSTEP
      IF(TSTEP .LE. POINT1)END1=TSTEP
C     OBJB = OBJB
      STOPPER = EPSILON
      IF(KOPOUT(7) .EQ. 1) STOPPER = EPSILON * DIFFSAV
      IF(ABS(END2-END1) .LE. STOPPER)GO TO 220
      GO TO 150
200   CONTINUE
C *** SUCCESS
      IMPROV = 1
      IF(TSTEP .GT. POINT1)END1=POINT1
      IF(TSTEP .LE. POINT1)END2=POINT1
      POINT1 = TSTEP
      DO 225 I2=1,MATDIM
      DO 225 J2=1,MATDIM
      PBASE(I2,J2)=ADJP  (I2,J2)
225   CONTINUE
      OBJB = OBJF
      STOPPER = EPSILON
      IF( KOPOUT(7) .EQ. 1) STOPPER = EPSILON * DIFFSAV
      IF(ABS(END2-END1) .LE. STOPPER)GO TO 220
      GO TO 150
220   CONTINUE
C      RESET TIME PERIOD VALUE TO ACTUAL(NOT QUICKIE) VALUE
      ICALCT = ITIME
C       **********************************************************
C       *                                                        *
C       *               END OF THE GOLDEN SECTION LINE SEARCH    *
C       *                                                        *
C       **********************************************************
C
```

BIBLIOGRAPHY

1. Adelman, Irma G.: A Stochastic Analysis of the Size Distribution of Firms, Journal of the American Statistical Association, 53:893(1958).

2. Anderson, T. W. and L. A. Goodman: Statistical Inference About Markov Chains, The Annals of Mathematical Statistics, 28:89(1955).

3. Chung, Kai Lai: Markov Chains with Stationary Transition Probabilities,Springer-Verlag (Berlin, 1960).

4. Cooper, Leon and David Steinberg, Introduction to Methods of Optimization, W. B. Saunders Company (Philadelphia, 1970).

5. Ezzati, Ali: Forecasting Market Shares of Alternative Home-Heating Units by Markov Process Using Transition Probabilities Estimated from Aggregate Time Series Data , Management Science, 21:462(December 1974).

6. Farris, P. L. and S. I. Padberg: Measures of Market Structure Change in the Florida Fresh Citrus Packing Industry , Agricultural Economics Research, 10:93(1964).

7. Goldfield, Stephen M. and Richard E. Quandt: Nonlinear Methods in Economics, North Holland (Amsterdam, 1972).

8. Goodman, L. A.: A Further Note on Miller's 'Finite Markov Processes in Psychology' , Psychometrika, 18:245 (1953).

9. Harary, Frank and Benjamin Lipstein: 'The Dynamics of Brand Loyalty: A Markovian Approach' , Operations Research, 10:19 (1962).

10. Himmelblau, David M.: Applied Nonlinear Programming, McGraw Hill (New York, 1972).

11. Hoel, Paul G., Sidney C. Port and Charles J. Stone: Introduction to Stochastic Processes, Houghton Mifflin; (Boston, 1972).

12. Howard, R. A.: Stochastic Process Models of Consumer Behavior , Journal of Advertising Research, 3:50, (September, 1963).

13. Jones, J. Morgan: A Composite Heterogeneous Model for Brand Choice Behavior , Management Science, 19:499 (January, 1973).

14. Kao, R. C. W.: Note on Miller's 'Finite Markov Processes in Psychology' , Psychometrika, 18:241 (1953).

15. Kemeny, John G. and J. Laurie Snell, and Anthony W. Nnopp: Denumerable Markov Chains, D. Van Nostrand Company, Inc. (Princeton, 1966).

16. Krenz, R. D.: Projection of Farm Numbers for North Dakota with Markov Chains , Agricultural Economics Research, 16:77(1964).

17. Lee, T. C., G. G. Judge and T. Takayama: On Estimating the Transition Probabilities of a Markov Process , Journal of Farm Economics, 47:742(1965).

18. Lee, T. C., G. G. Judge and A. Zellner: Maximum Likelihood and Bayesian Estimation of Transition Probabilities , Journal of the American Statistical Association, 16:1162(1968).

19. Lee, T. C., G. G. Judge and A. Zellner: Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North Holland (Ampsterdam, 1970).

20. Madansky, A.: Least Squares Estimation in Finite Markov Processes , Psychometrika, 24:137(1959).

21. Maffei, Richard B.: Brand Preferences and Simple Markov Processes , Operations Research, 8:210 (1960).

22. Marshall, Andrew W.: and Herbert Goldhamer, An Application of Markov Processes to the Study of the Epidemiology of Mental Disease , Journal of the American Statistical Association, 50:99 (1955).

23. Massey, William F.: David B. Montgomery, and Donald G. Morrison, Stochastic Models of Buying Behavior, the MIT Press (Cambridge, Massachusetts, 1970).

24. Miller, George A.: Finite Markov Processes in Psychology , Psychometrika, 17:149(1952).

25. Powell, M. J. D.: A Survey of Numerical Methods for Unconstrained Optimization, SIAM Review, 12:79 (1970)

26. Powell, M. J. D.: An Efficient Method for Finding the Minimum of Several Variables without Calculating Derivatives, Computer Journal, 7:155(1964).

27. Powell, M. J. D.: A Method for Minimizing a Sum of Squares of Non-linear Functions without Calculating Derivatives, Computer Journal, 7:303(1965).

28. Preston, L. E. and E. J. Bell: The Statistical Analysis of Industry Structure: An Application to the Food Industry , Journal of the American Statistical Association, 56:925(1961).

29. Romanovsky, V. I. Discrete Markov Chains translated by E. Seneta, Wolters-Noordhoff (Groningen, 1970).

30. Saffer, Shelley I., Charles E. Mize, U. Naroyan Bhat and Stephen A. Szygenda: Use of Non-Linear Programming and Stochastic Modeling in the Medical Evaluation of Normal-Abnormal Liver Function, IEEE Transactions on Biomedical Engineering, BME-23:200 (May 1976).

31. Saffer, Shelley I.: A Stochastic Simulation Model of Normal-Abnormal Liver Function, PhD Dissertation, Southern Methodist University (August, 1973).

32. Sleder, Albert J.: Time Series Forecasting Techniques for Scheduling of Multiprocessor Computer Jobs, M.S. Thesis, Georgia Institute of Technology (August, 1974).

33. Styan, George P. H. and Harry Smith, Jr.: Markov Chains Applied to Marketing , Journal of Marketing Research, 1:50 (1964).

34. Telser, L. G.: Least Squares Estimation of Transition Probabilities", Measurement of Economics. Stanford University Press (Stanford, 1963).

35. Theil, H. and C. van de Panne: Quadratic Programming as an Extension of Classical Quadratic Maximization , Management Science, 7:1 (October, 1970).

36. Theil, H. and Guido Rey: A Quadratic Programming Approach to the Estimation of Transition Probabilities , Management Science, 12:714(1966).

37. Walsh, G. R.: Methods of Optimization, John Wiley and Sons (New York, 1975).

38. Wolfe, P.: The Simplex Method for Quadratic Programming , Econometrica, 27:382(1959).