THE AUSTRALIAN JOURNAL OF AGRICULTURAL ECONOMICS

VOL 15 AUGUST 1971 NO. 2

A REVIEW OF THE ESTIMATION OF TRANSITION PROBABILITIES IN MARKOV CHAINS

WARREN DENT and RICHARD BALLINTINE*

A chronological review of the development of estimation procedures for unknown constant Markovian transition probabilities is presented with emphasis on applications involving the availability of macrodata, as opposed to microdata. Monte Carlo results comparing various estimation methods are analysed and several suggestions for estimating non-stationary probabilities are made.

The estimation of Markovian transition probabilities and inferences based on these probabilities have been examined in a variety of applied fields, and under a variety of assumptions. The basic concept of the Markov chain was introduced by A. A. Markov in 1912 [42], with subsequent relevant studies by Kolmogorov, Doeblin, and Doob in the early 1940's [14, 15, 16, 17, 18, 32]. Miller's 1952 work in the study of learning processes in psychology [44] seems to offer the first serious application involved with *estimation* of a Markov process. This initial attempt has provided the impetus for a widening band of estimation techniques in instances where applied data is in the form of 'market shares'.

To estimate transition probabilities, ideally the researcher would like to have information over each period on actual movements of individuals from one state of the chain to another. This might be considered the case of 'full information' or of the availability of 'microdata'. Here the maximum likelihood estimators of the probabilities p_{ij} of the transition from state i to state j are easily derived as the ratios of the numbers of individuals moving from state i to state j to the numbers originally in state i. Where observations are available over several time periods and the probabilities are assumed constant, the maximum likelihood estimators are found by a straightforward averaging process. Billingsley [5, 6, 7] and Anderson and Goodman [3, 22] have examined in detail the statistical properties of these estimators and developed tests for time-dependency and the order of chains providing the best fit to given data. Lee, Judge, and Cain [35] have shown in their simulation study that such microdata estimators provide excellent estimates of the true transition probabilities.

* Respectively of the Department of Economics, University of Iowa, and the School of Business Administration, University of Minnesota.

Unfortunately, in most applications researchers do not have available data on movements between states. More likely they have data reflecting the end result of such movements, this data being in the form of proportions or market shares, and called 'macrodata'. This case of 'limited information' is also the more challenging theoretically, in that estimators with strong properties are desired, yet the assumptions postulated must be necessarily weaker.

This review is mainly concerned with estimation in the case of limited data information. It proposes to examine chronologically the development of estimation techniques, and point to possible future avenues of research. A comprehensive selection of published applications is included. The areas from which these are drawn involve agricultural economics [27, 31, 33, 61], international trade [12], fertility [56, 57, 58, 59], industrial structure [1, 11, 51], sociology [20, 71], psychology [29, 44], medicine [9], actuarial science [10], marketing [38, 48], demography [30, 45, 46, 47, 54], stock market analysis [19], capital theory [53], ethnic studies [55], behaviour [2, 52, 63], wage theory [60], social mobility [8, 49, 50] and textile production [26]. A further, excellent review of the state-of-the-art of estimation in micro and macro Markov models is to be found in Lee, Judge and Zellner [37], where detailed mathematical and Monte Carlo analyses of theoretical models are undertaken and reported.

The common model for most applied studies may be written as

$$(1) \qquad m_{jt} = \sum_{i=1}^{r} p_{ij} m_{i\,t-1} + v_{jt} \quad j = 1, \ldots, r; \; t = 2, \ldots, T$$

where m_{jt} is the proportion of the system in state j at time t, there being r states and $T-1$ time periods (time ranging from $t = 1$ to $t = T$), and v_{jt} is a random error component. The transition probabilities p_{ij} are assumed constant over the time span analysed. In matrix notation (1) may be rewritten as

$$(2) \qquad \underset{\sim}{m}_t = P' \underset{\sim}{m}_{t-1} + \underset{\sim}{v}_t \quad t = 2, \ldots, T$$

where $\underset{\sim}{m}_t' = (m_{1t}, m_{2t}, \ldots, m_{rt})$ is the row vector of market shares at time t, $P = (p_{ij})$ is the matrix of transition probabilities, and v_t the error term at time t. The only given relations on the data are

$$(3) \qquad \left.\begin{array}{ll} \underset{\sim}{m}_t' \cdot \underset{\sim}{1} & = 1 \\ \underset{\sim}{m}_t & \geqslant \underset{\sim}{0} \end{array}\right\} \quad t = 1, \ldots, T$$

where $\underset{\sim}{1}$ is an r component column vector of ones. The desired constraints on the transition probabilities are

$$(4) \qquad 1 \geqq P \geqq 0$$

$$(5) \qquad p \cdot 1 = \underset{\sim}{1}$$

A set of probabilities or transitions matrix satisfying (4) and (5) is called 'admissible'. Equation (5) together with the right hand side of (4) implies the left hand inequality of (4).

Unrestricted Least Squares

Denoting the first term on the right of (1) by

$$(6) \qquad \hat{m}_{jt} = \sum_{i=1}^{r} p_{ij} m_{i\,t-1} \quad j = 1, \ldots, r$$

or the predicted share of state j at time t, Miller [44] estimated the unknown probabilities by minimizing the sum of the squares of the differences between the predicted proportions and the proportions actually observed. His objective function was

$$\sum_{t=2}^{T} \underset{\sim}{v}_t' \underset{\sim}{v}_t = \sum_{t=2}^{T} \sum_{j=1}^{T} v_{jt}^2 \tag{7}$$

Although Miller erred in his derivation procedure, Goodman [21] obtained the same result on correction. This was that the least squares estimate of the matrix of transition probabilities is given by the following:

$$P = NM' (MM')^{-1}$$

where M is the matrix of observed proportions for observations one to $n-1$ (n is the total number of observations or time periods), and N is the matrix of observed proportions one observation beyond M (i.e. observations 2 through n).

Inherent in the form of P was the satisfaction of condition (5), or the exhaustion of possible movements from any given state. Unhappily Miller's technique permitted the occurrence of inadmissible estimates, that is negative probabilities, or probabilities greater than unity. Further, the correction terms for each state did not have the same variance so that unweighted estimators were not asymptotically efficient [41]. The independence of the correction terms and the observations, however, did make the estimates consistent with respect to the number of observations and asymptotically consistent with respect to the sample size. Madansky [41], who pointed out the differences in the variances, in the same article proposed a way to overcome the problem. After obtaining the estimates as Miller proposed, Madansky suggested modifying them by forcing the variances of the correction terms to be unity through the use of a weighting technique in the objective function, to be discussed later.

The problem of inadmissible estimates was first attacked by Goodman [21]. His solution was simply a modification of Miller's method. Summarily dismissing the problem, Goodman suggested setting all probabilities beyond the allowable range equal to the closest boundary value (i.e. 0 or 1) and adjusting the remaining estimates so that the sum of the squares of the correction terms was still a minimum. Telser carried this line of reasoning one step further by actually giving the iterative procedure necessary to incorporate such a modification into the solution [64]. His procedure became very complicated when the inadmissible estimates were greater than four in number and naturally, if all the alternative possibilities of assigning extreme values were not examined, it was possible that the combination of estimates which best fitted the least squares criterion would not be found.

Telser also modified the procedure in two other ways. He transformed the relationship (1) by a simple adjustment of the coefficients to yield

$$m_{jt} = \alpha_j + \sum_{\substack{i=1 \\ i \neq j}}^{r} \alpha_{ij} m_{i\,t-1} + v_{jt} \quad j = 1, \ldots, r \tag{8}$$

where

$$\alpha_j = p_{jj} \; \alpha_{ij} = p_{ij} - p_{jj} \tag{9}$$

a formulation which was more consistent with the then current regression expressions. Although this was nothing really new, it did present the estimation technique in a different light. Telser also attempted to modify Madansky's weighting procedure so that the requirement that the transition probabilities for a state would sum to unity (a quality which was lost when Madansky modified Miller's work) would be satisfied. Lagrangian multipliers were used to include this restriction explicitly. Telser found the worth of this modification to be somewhat doubtful since the resulting estimates were unreliable. This was due to the approximately singular nature of the matrices, causing extreme sensitivity to rounding errors. He concluded that the procedure should only be used with great caution, preferably when experience provided some idea of the value of the technique for the process being studied.

Restricted Least Squares

Although attempts were made to adjust the basic least squares technique for admissability, the modifications were somewhat arbitrary and the resulting probabilities no longer were guaranteed to satisfy the least squares criteria. Work such as Madansky's which aimed at solving problems in the technique only resulted in fewer of the restrictions on the probabilities being met. It seemed that only by explicitly including all the restrictions could the problem be solved. The resulting procedure was called restricted least squares.

To enable both the equality restriction (5) and the inequalities (4) to be explicitly included in the formulation, the problem was specified by Lee, Judge and Takayama [28, 34] and by Theil and Rey [68] as a typical quadratic problem. The objective was to minimize the error sum of squares

$$\sum_{t=2}^{T} \underset{\sim}{v}_t' \, \underset{\sim}{v}_t$$

(quadratic in the unknown probabilities) with respect to the linear constraints (4) and (5). The version of Lee, Judge, and Takayama made use of the standard simplex version of the quadratic programming algorithm. Theil and Rey's version modified the unrestricted least squares results in a finite number of successive steps, in each of which certain constraints that were violated in previous steps were imposed in binding (equational) form.

Zellner [74] investigated the statistical properties of these inequality restricted estimators. He found the estimates had distributions of the truncated normal form (partly continuous and partly discrete). However, he found it difficult to evaluate the moments analytically and obtain the sampling properties of the restricted estimators when more than two variables were involved. The difficulty stemmed from the fact that the distributions which were binding were not constant from one observation to the next and should really have been considered as a vector of random variables. Properties of the restricted least squares estimates have been discussed by a number of others. Although some definite conclusions have resulted, such as the derivation of the exact distribution, the general consensus [25] is that other properties are analytically indeterminate.

Weighted Restricted Least Squares

The explicit consideration of the restrictions on the probabilities by the restricted least squares method resulted in the estimates being admissible. The efficiency of the restricted least squares estimator was questioned however, and the weighted restricted least squares estimator was developed.

The problem of heteroscedasticity noted earlier was found to be partly removed when a matrix weight function was introduced into the least squares objective equation. The objective function became

$$\sum_{t=2}^{T} \underset{\sim}{v}' A_t \underset{\sim}{v}_t \tag{10}$$

where A_t is a matrix of weighting factors applicable to time t. The admissability constraints were not explicitly included. In a previously mentioned study, Madansky put forth the first type of weighting structure [41]. His procedure required that the transition probability estimates be determined first using the unrestricted least squares technique. Using these estimates, the estimates of the inverses of the variances of the deviation terms for each period were used as the weighting factors to recalculate the estimates. Madansky found little improvement was gained by repetition of the procedure.

Theil and Rey proposed the use of the inverse of the average proportion in a state over the period of observation as the weighting factor [68]. By inserting this appropriately in their quadratic programming scheme, the unrestricted least squares procedure was by-passed and both of the restrictions on the transition probabilities were met.

Other weighting factors have been suggested by Lee, Judge and Cain [35]. They proposed using the inverses of the estimate of the disturbance variance for each of the equations in the unweighted least squares procedure as weighting factors, yielding an unbiased version of Madansky's procedure. The properties of the estimators so derived and tests to determine their usefulness are presented by Zellner [73], who found that this procedure often resulted in large gains in efficiency when compared to the regular application of least squares. Tests were presented which examined the hypothesis that the transition probabilities were constant. Another possibility mentioned by Lee, Judge, and Cain was to employ the inverse of the product of the average proportion in a state and the average proportion not in that state as a weight for that state in the given period. Comparisons of the accuracy of different weighting factors in simulated use will be made later.

McGuire [39] investigated a number of the weighting proposals. Giving special attention to Madansky's procedure, McGuire showed his estimator was not inefficient. Although no general recommendations were made, McGuire did propose a different statistic to use as a weighting factor, in which the whole covariance matrix of the disturbance terms was used.

Other Estimators

The previous techniques have been concerned with the gradual evolution of the use of least squares criteria for estimating the transition probabilities of the postulated Markov chain underlying the process being investigated. Other published techniques differ in the criteria used to derive optimal estimates. Lee, Judge and Zellner [36] present a method of estimation based on maximizing a likelihood function of the transition probabilities. On the hypothesis that the observed proportions are generated by independent trials with the transition probabilities constant for each trial, the multinomial probability of any given sequence of states at time t is given by

$$\frac{N(t)!}{\prod_{i=1}^{r} n_i(t)!} \prod_{j=1}^{r} q_j(t)^{n_j(t)} \tag{11}$$

where $N(t)$ is the number of trials, $n_i(t)$ is the number of times state i is observed in the N trials, r is the number of states, and $q_j(t)$ is the probability that the j-th state will occur for an individual.

It is possible to maximize the likelihood function of the joint density function if the covariance matrix of the proportions in the various states is known. Without this knowledge, an estimate of the covariance matrix must be made. The estimate is found by replacing the observed proportions in the analysis by the expected proportions, whose covariance matrix can be found by existing techniques [36]. The procedure then is straightforward when the restrictions on the permissible values of the probabilities are ignored. Inclusion of these restrictions, however, requires the use of the reducibility theorem of non-linear programming, and after some manipulation, the problem can be converted into a form that is solvable by use of the standard simplex version of the quadratic programming algorithm developed by Wolfe [72].

Dent [13] and McGuire et al. [40] have developed maximum likelihood procedures which directly account for the probability restrictions. Their methods are based on normality of the usual regression disturbance terms v_{jt} in (1). The errors are assumed to have a special symmetric covariance structure reflecting the singularity of the interdependence of the regression equations, identical marginal distribution variances and identical pair correlations. It is shown that the maximum likelihood estimators are equivalent to the restricted least squares estimators under these conditions.

Taking their cue from the findings of others [43, 67, 69, 70, 75], Lee, Judge and Zellner [36] have shown how Bayesian techniques may be used to estimate transition probabilities. The Bayesian approach may be used in cases of microdata availability when the prior mean and variance may be found from an actual frequency count of the movements of the microdata. The prior probability density function for a given state is assumed to be a basic multivariate beta probability with the given mean and variance. Using the multinomial likelihood, a posterior density may be derived. If a mean and variance are not available *a priori* they may be derived from a predicted moving frequency of past transition probability estimates [36].

Lee, Judge and Zellner initially ignored the restrictions on the probabilities in arriving at an estimating equation for the probabilities, and made various approximations using the prior variance and prior covariance for current unknowns. The initial probability estimates were then used to replace the prior probabilities and the procedure was repeated. Repetition continued until no change in the probabilities was seen. The posterior probabilities used in the procedure were at the mode of the posterior probability distribution.

Since the admissibility restrictions were not included, some of the probability estimates so found were not admissible. It is possible, although cumbersome, to include these restrictions in a manner very similar to that used in the maximum likelihood estimation technique.

A third type of estimation technique that has been proposed is the minimum absolute deviation (MAD) procedure. Ashar and Wallace [4] derived transition probability estimates which minimized

$$\sum_{t=2}^{T} \sum_{j=1}^{r} |v_{jt}|$$

subject to admissibility constraints. Others have pointed out and Ashar and Wallace themselves realized that a MAD estimator is not as efficient as a minimum variance estimator. The advantage of this technique, however, is its computational simplicity. Ashar and Wallace showed how appropriate transformations of the original variables permitted the problem to be formulated as a linear program. Investigation has shown that the estimates resulting from their procedure are normally distributed except for those estimates close to the limits of the allowed probability values [35]. The quality of the results was found to be highly dependent on the sample size, improving as the sample size grew.

Time-dependent Estimators

All the previous estimation techniques have had one assumption in common, that the transition probabilities are constant over time. When this assumption is relaxed, the task immediately grows from finding one set of probabilities to finding nearly as many different sets as there are time periods. Such a problem would be very difficult to solve unless some relation between probabilities in successive periods were assumed. If more than just the passage of time is felt to influence the changes in probabilities, the next question concerns determination of the relevant independent variables in the relation which are responsible for these changes, and the exact form of the relation hypothesized.

Miller investigated the assumption of time independence in his groundbreaking article in the early 50's [44]. Since his work involved psychological phenomena, he considered any time dependence would be due to learning processes. This led him to believe that the dependence would be transitional in nature as the organism advanced from a position of no knowledge to a position of complete knowledge. If the probabilities for both these positions were known, the change per period could be determined (assuming the change to be uniform).

Telser [64] presented a method that looked for possible relations of the transition probabilities with factors other than just the passage of time. Assuming such relations to be linear, he replaced the probabilities with their equivalent value in terms of the other factors in the equation relating past proportions to present proportions. Least squares estimators were used to determine the coefficient of these other factors in their relation to the probabilities. Telser was particularly concerned with the problem of multicollinearity and presented approximation techniques for estimation when such a situation arose. Two examples of the effect of external factors on probability determination were presented. In the first, the transition probabilities between brands of cigarettes were assumed to depend on prices and in the second example, the independent variable was advertising expenditures [65, 66].

Recently, there has been more interest in the determination of the time dependency of the transition probabilities. Initially, the concentration in this research has been at the expense of the generality of the procedure. Thus, the new studies have again started assuming the availability of microdata. Hallberg [23] examined the prediction of future transition probabilities by regression analysis. The assumption on which he based his study was that there existed a time series of known transition probability matrices. With these matrices, it was possible to fit a least squares regression equation of various exogenous factors explaining the transition probability values. The result was used to predict probabilities in future periods and although the probabilities for each state did sum to one as required, some of the estimates were inadmissible. Using an admittedly arbitrary rule, Hallberg suggested setting the inadmissible estimates equal to the closest permitted value (0 or 1). The remaining estimates were then adjusted so they still summed to one.

In the same field of interest, although not concerned with the estimation of the transition probabilities *per se,* is the work of Lipstein [38] who attempted to find the effect of advertising expenditures on brand attitudes. Lipstein dismissed the problem of finding the transition probabilities for a period. As he put it, 'These entries [probabilities] are easily derived from consumer purchase panel data'. He postulated that the probabilities were dependent on three factors: availability, price, and a reaction factor. Since price and availability were supposedly known, the object was to find the reaction factors, which were the probabilities of a brand preference change. Since the transition probabilities (brand preferences), availability, and prices were all known for all periods, the reaction factors could be readily determined for each period. Lipstein hypothesized that the change in the reaction factor matrix from one period to the next could be explained by a causative matrix. In other words, the reaction factor matrix for a period could be found by multiplying the reaction factor matrix of the previous period by the causative matrix. Since all the reaction factors were known, the causative matrix could be found by regression analysis. Assuming the elements of the causative matrix to be dependent on advertising expenditures, a linear regression analysis was run to determine the coefficient relating the two. Lipstein concluded that the resulting coefficients were measures of how advertising expenditures could bring about changes in attitudes. The work has been followed by a statistical analysis of this causative model with special attention on the two state case [24].

Dent [13] suggested a technique for estimating non-stationary probabilities from macrodata restrictions. In essence the method depended on heuristic rules for bounding the movements of individual transition probabilities between successive periods, the rules being determined by the application. The usual objective function was expanded to include the effects resulting from the time dependence. This allowed the model to maintain the regression structure and permitted consideration of external influences at the same time. Prices were the main external factor considered. Quadratic programming was found to be required for solving most problems. An example of the use of the technique in the case of two states over three periods was presented.

Comparisons

It is difficult to compare analytically the various techniques of transition probability estimation that have been described above. Lee, Judge, Cain and Zellner [35, 36] undertook two simulation analyses to compare some of the methods. Using a given set of transition probabilities, 1000 individuals were led through a number of steps of a random walk. The comparison of different estimation techniques (applicable to micro and macrodata cases) was made by examining how well each method could estimate the known transition probabilities from the generated data. The criterion of 'how well' was the size of the sum of the root mean square errors for each of the estimators, the smaller the sum the better the estimation procedure. Unfortunately, in the two analyses performed different numbers of steps were used to find estimates so that direct comparison of some pairs of techniques is impossible.

The first study [35] resulted in the following ranking of methods with the best procedure given first: weighted restricted least squares, unweighted least squares, restricted minimum absolute deviations, and unrestricted least squares. Although the different weighting procedures used in the weighted restricted least squares produced slightly different results, the differences were not significant. Hence, no differentiation was made between the various weighting schemes. Furthermore, the MAD and unweighted least squares techniques were found to be approximately equal. The most significant difference was found between the unrestricted least squares and the other techniques. That is, little difference was found among all restricted estimators, but they were all much better than the unrestricted least squares. Of course, if microdata were assumed to be available, the maximum likelihood techniques which used the microdata were far superior to any other procedure.

The second independent study [36] resulted in the following rank ordering of methods: Bayesian, maximum likelihood, weighted least squares, least squares. This order was found to hold whether the techniques were restricted or not (however, when the restrictions were not applied, least squares did not differ from weighted least squares and the maximum likelihood estimator was not significantly better). All techniques improved with larger sample sizes, and the accuracy of the Bayesian estimators was found to be highly dependent on the accuracy of the prior probabilities.

Future Studies

Looking at past studies, it is evident that an evolutionary trend has been followed. Each study improved on the work in its particular family of studies. The families of studies are differentiated by the different fundamental bases (least squares, maximum likelihood, minimum absolute deviations, etc.) used in their estimating technique. Extension of the trend of the past studies into the future gives a definite suggestion of what the next step will be. Only one existing study has been able to estimate nonconstant transition probabilities without using microdata. Unfortunately, cases where the probabilities are not constant and only proportion data exist are the most common.

Analogous to the restricted least squares formulation in the constant probability case, a possible representation of the non-stationary estimation problem is: find matrices P_t, $t = 1, \ldots, T-1$ which minimize

(12) $$\sum_{t=2}^{T} v_t' v_t$$

subject to

(13) $\underset{\sim}{v}_t = \underset{\sim}{m}_t - P'_{t-1} \underset{\sim}{m}_{t-1} \quad t = 2, \ldots, T$

(14) $P_t \cdot \underset{\sim}{1} = \underset{\sim}{1} \quad t = 1, \ldots, T-1$

(15) $P_t \geqslant 0 \quad t = 1, \ldots, T-1$

(16) $P_t = f(P_{t-1}, \ldots, P_1) \quad t = 2, \ldots, T-1.$

P_t is the matrix of transition probabilities applicable between periods $t - 1$ and t, while $P_t = f(P_{t-1}, \ldots, P_1)$ is some relation, as yet unspecified, relating probabilities in the current period with those in previous periods. Ignoring this relation reduces the problem to the equivalent of $T-1$ restricted least squares problems each using only two data periods.

Dent [13] suggested (16) take the form of a series of inequalities determining bounds on the extent of possible changes in the probabilities from period to period. Another possibility is to assume, as did Lipstein [38] that the P_t are related by a causative matrix C such that

(17) $P_t = C P_{t-1}$ or $P_t = P_{t-1} C \quad t = 2, \ldots, T-1$

where the C and P_t matrices also obey certain other constraints [24]. Essentially such relations imply that the probabilities are changing in a constant manner. The existence and form of the nonlinear terms in $P_t = P_{t-1} C$ in this formulation, however, prohibits the use of quadratic or convex programming methods in determining an explicit solution. At this time an appropriate programming technique which will result in an explicit solution is not known.

Alternately, the P_t could be looked upon as being dependent on previous period probabilities and/or some other observable variable, such as price or advertising expenditures or a combination of such variables. Since the P_t are not known, the exact nature of such relationships cannot be found without further assumptions. Another possibility involves assuming the probabilities constant over a number of distinct time spans and drawing inferences from sets of constant probabilities. For example, suppose proportion data for twenty periods were available. The periods could be separated into groups of five. Constant probabilities over each group of five periods could be estimated by existing procedures. There would then exist four sets of transition probabilities. The relations between them could be found by the techniques discussed under 'Time-dependent Estimators' and the pattern of change could then be applied within the groups to find different probabilities for each year. These estimates could be used again to estimate the yearly variance, from which a final set of transition probabilities would emerge. Work would be required to test the mechanics and the validity of the procedure.

Indications are that the best method of estimation of non-stationary probabilities may depend heavily on the particular application. With limited data availability in the macro case and a high number of independent probabilities to be estimated, the power of any procedure will be limited except under relatively strong assumptions. While various system formulations may be readily constructed, the actual derivation of solutions may be extremely complicated. The use of non-stationary probabilities and their value (e.g. in prediction) appear to be the most pertinent field for immediate analysis.

References

[1] Adelman, I. G. 'A Stochastic Analysis of the Size Distribution of Firms'. *Journal of the American Statistical Association,* LIII (December, 1958), pp. 893-904.

[2] Anderson, T. W. 'Probability Models for Analyzing Time Changes and Attudes'. *Mathematical Thinking in the Social Sciences.* Edited by P. F. Lazarsfeld. Glencoe: The Free Press, 1955.

[3] ————, and Goodman, L. A. 'Statistical Inference about Markov Chains'. *The Annals of Mathematical Statistics,* XXVIII (1957), pp. 89-110.

[4] Ashar, V. G., and Wallace, T. D. 'A Sampling Study for Minimum Absolute Deviation Estimators'. *Operations Research* (1963), pp. 747-758.

[5] Billingsley, Patrick. 'Statistical Methods in Markov Chains'. *The Annals of Mathematical Statistics,* XXXII (March, 1961), pp. 12-40.

[6] ————, 'On testing Markov chains'. Presented to the Institute of Mathematical Statistics, September 10, 1957 (unpublished paper).

[7] ————, *Statistical Inference for Markov Processes.* Institute of Mathematical Statistics, University of Chicago Statistical Research Monographs, University of Chicago Press, Chicago, 1961.

[8] Blumen, Isadore; Kogan, Marvin; and McCarthy, Phillip J. *The Industrial Mobility of Labor as a Probability Process.* Ithaca: Cornell University Press, 1955.

[9] Chiang, Chin Lang. 'A Stochastic Model of Competing Risks of Illness and Competing Risks of Death'. *Stochastic Models in Medicine and Biology.* Edited by John Gurland. Madison: University of Wisconsin Press, 1964.

[10] ————, 'A Stochastic Study of the Life Table and Its Applications: III, The Follow-Up Study with the Consideration of Competing Risks'. *Biometrics,* March (1961), pp. 57-78.

[11] Collins, N. R., and Preston, L. E. 'The Size Structure of Industrial Firms'. *American Economic Review* (December, 1961), pp. 986-1003.

[12] Dent, Warren T. 'Application of Markov Analysis to International Wool Flows'. *Review of Economics and Statistics,* XLIX (1967), pp. 613-616.

[13] ————, 'The Estimation of Constant and Non-Constant Transition Probabilities from Market Shares'. Unpublished paper, University of Minnesota.

[14] Doeblin, W., 'Sur les Proprietes Asymptotique de Mouvement Regis pour Certains Types de Chaines Simples'. *Bull. Math. Soc. Roum. Sci.,* 39, No. 1, pp. 57-115, No. 2, pp. 3-61 (1937).

[15] ————, 'Sur L'equation Matricelle $A^{(t+s)} = A^{(t)} A^{(s)}$ et ses Applications aux Probabilities in Chaines'. *Bull. Sci. Math.* (2), *62,* pp. 21-32 (1938), *64,* pp. 35-37 (1940).

[16] ————, 'Sur deux Problemes de M. Kolmogoroff Concernant les Chaines Denombrables'. *Bull. Soc. Math.* France *66,* pp. 210-220 (1938).

[17] Doob, J. L., 'Topics in the Theory of Markov Chains'. *Annals of Mathematical Statistics,* 52, 455-473, (1942).

[18] ————, 'Markov Chains-denumerable case'. *Annals of Mathematical Statistics,* 58, pp. 455-473 (1945).

[19] Dryden, M. M., 'Share Price Movements: A Markovian Approach'. *The Journal of Finance,* XXIV (March, 1969), pp. 49-60.

[20] Fuguitt, Glenn V., 'The Growth and Decline of Small Towns as a Probability Process'. *American Sociological Review* (June, 1965), pp. 403-411.

[21] Goodman, Leo A., 'A Further Note on Miller's "Finite Markov Processes in Psychology" '. *Psychometrika* (September, 1953), pp. 245-248.

[22] ————, 'Statistical Methods for the Mover—Stayer Model'. *Journal of the American Statistical Association,* LVI (December, 1961), pp. 841-868.

[23] Hallberg, M. C., 'Projecting the Size Distribution of Agricultural Firms—An Application of a Markov Process with Non-Stationary Transition Probabilities'. *Journal of Farm Economics,* 51 (May, 1969).

[24] Harary, Frank; Lipstein, Benjamin; and Styan, George. *A Matrix Approach to Nonstationary Chains.* University of Minnesota Technical Report No. 120, May 20, 1969.

[25] Hocking, R. R., 'The Distribution of a Projected Least Squares Estimator'. *Annals of the Institute of Statistical Mathematics,* XVII (1965), pp. 357-362.
[26] Jarrett, F. G. and Dent, W., 'Fibre Substitution—A Markov Process Analysis'. *Australian Economic Papers,* V. 1960.
[27] Judge, C. G., and Swanson, E., 'Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics'. *Australian Journal of Agricultural Economics,* VI (1962).
[28] ————, and Takayama, T., 'Inequality Restrictions in Regression Analysis'. *Journal of the American Statistical Association,* LXI (March, 1966), pp. 166-181.
[29] Kao, Richard C. W., 'Note on Miller's "Finite Markov Processes in Psychology" '. *Psychometrika* (September, 1953), pp. 241-243.
[30] Kelley, Allen C., and Weiss, Leonard W., 'Markov Processes and Economic Analysis: The Case of Migration'. *Econometrica,* XXXVII (April, 1969), pp. 280-297.
[31] Kislev, Y., and Amaid, A., 'Linear and Dynamic Programming in Markov Chains'. *American Journal of Agricultural Economics,* L (1968), pp. 111-129.
[32] Kolmogorov, A. N., 'Anfongsgrunde der Theorie der Markoffschen Kettes met Unendlichen Vielen Moglichen Zustanden'. *Mat. Sbornik N.S. Ser.,* pp. 607-610 (1936); *Bull. Univ. Moscou* 1 (1937) Russian.
[33] Krenz, Ronald D., 'The Projection of Farm Numbers for North Dakota with Markov Chains'. *Agricultural Economics Research,* XVI (July, 1964), pp. 77-83.
[34] Lee, T. C.; Judge, G. G.; and Takayama, T. 'On Estimating the Transitional Probabilities of a Markov Process'. *Journal of Farm Economics,* XLVII (1965), pp. 742-760.
[35] ————; Judge, G. G.; and Cain, R. L. 'A Sampling Study of the Properties of Estimators of Transition Probabilities'. *Management Science* (March, 1969), pp. 374-398.
[36] ————; Judge, G. G.; and Zellner, A. 'Maximum Likelihood and Bayesian Estimation of Transition Probabilities'. *Journal of the American Statistical Association* LXIII (December, 1968), pp. 1162-1179.
[37] ————; Judge, G. G.; and Zellner, A. *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data,* North-Holland, Amsterdam, 1970.
[38] Lipstein, B. 'A Mathematical Model of Consumer Behaviour'. *Journal of Marketing Research,* II (August, 1965), pp. 259-265.
[39] McGuire, Timothy W. 'More on Least Squares Estimation of the Transition Matrix in a Stationary First-Order Markov Process from Sample Proportion Data'. *Psychometrika* (September, 1969), pp. 335-345.
[40] ————, *et al.* 'Estimation and Inference for Linear Models in which Subsets of the Dependent Variable Are Constrained'. *Journal of the American Statistical Association,* LXIII (December, 1968), pp. 1201-1213.
[41] Madansky, Albert. 'Least Squares Estimation in Finite Markov Processes'. *Psychometrika* (June, 1959), pp. 137-144.
[42] Markov, A. A., *Wahrscheinlichkeitsrechnung,* translated from Russian (1912).
[43] Martin, J. J. *Bayesian Decision Problems and Markov Chains.* New York: John Wiley, 1967.
[44] Miller, George A. 'Finite Markov Processes in Psychology'. *Psychometrika,* (June, 1952), pp. 149-167.
[45] Musham, Helmut V. 'Discussion of Internal Migration as a Stochastic Process'. *Proceedings of the 34th Session,* Internation Statistical Institute, Ottawa, Canada, 1963.
[46] ————, 'Toward a Formal Theory of Internal Migration'. *International Population Union Conference,* 1961, paper 19.
[47] Myers, George C.; McGinnis, Robert; and Masnick, George. 'Preliminary Assessment of a Stochastic Process Model of Internal Migration'. *Population Index,* XXXI (July, 1965), pp. 256-257.
[48] Padberg, Daniel I. 'The Use of Markov Processes in Measuring Changes in Marketing Structure'. *Journal of Farm Economics,* XLIV (February, 1962), pp. 188-199.
[49] Prais, S. J.' 'The Formal Theory of Social Mobility'. *Population Studies,* IX (July, 1955), pp. 72-81.

[50] ———, 'Measuring Social Mobility'. *Journal of the Royal Statistical Society,* CXVIII (Series A, 1955), pp. 56-66.

[51] Preston, L. E., and Bell, E. J., 'The Statistical Analysis of Industry Structure: An Application to Food Industries'. *Journal of the American Statistical Association* (December, 1961).

[52] Regier, Mary H., 'A Two-State Markov Model for Behavioural Change', *Journal of the Americal Statistical Association,* LXIII (September, 1968), p. 993.

[53] Rey, G. 'A Markov Chain Prediction of Value Added and Expenditure Shares, Italy, 1861-1956'. Publication No. 16 of the Research Center of CIRET, 1966.

[54] Rogers, Andrei. "A Markovian Policy Model of Interregional Migration'. *The Regional Science Association Papers,* XVII (November, 1965), pp. 205-244.

[55] Scott, A. J. 'A Procedure for the Estimation of Markov Transition Probabilities Exemplified by a Model of Urban Ethnic Structure'. Unpublished paper, Department of Regional Science, University of Pennsylvania.

[56] Shelps, Mindel C., and Perrin, Edward B. 'The Distribution of Birth Intervals Under a Class of Stochastic Fertility Models'. *Population Studies,* XXVII (March, 1964), pp. 321-331.

[57] ———, 'Human Reproduction as a Stochastic Process', *Biometrics* (March, 1964), pp. 28-45.

[58] ———, 'On the Time Required for Conception'. *Population Studies* XXVIII (July, 1964), pp. 85-97.

[59] ———. 'Pregnancy Wastage as a Factor in the Analysis of Fertility Data'. *Demography,* I, pp. 111-118.

[60] Solow, R. 'Some Long-Run Aspects of the Distribution of Wage Incomes'. *Econometrica,* XIX (1951), pp. 333-334.

[61] Stanton, B. F., and Kettunen, L. 'Potential Entrants and Projections in Markov Process Analysis'. *Journal of Farm Economics,* XLIX (1967), p. 633.

[62] Summers, Robert. 'A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators'. *Econometrica,* XXXIII (1965), pp. 1-41.

[63] Suppes, Patrick, and Atkinson, Richard C. *Markov Learning Models for Multiperson Interactions,* Stanford Univ. Press, Stanford, 1961.

[64] Telser, Lester G. 'Least Squares Estimation of Transition Probabilities'. *Measurement of Economics.* Edited by Carl F. Christ. Stanford: Stanford University Press, 1963.

[65] ———, 'The Demand for Branded Goods as Estimated from Consumer Panel Data'. *The Review of Economics and Statistics,* XLIV (1962), pp. 300-324.

[66] ———, 'Advertising and Cigarettes'. *Journal of Political Economy,* 70 (1962) pp. 471-99.

[67] Theil, H. 'On the Use of Incomplete Prior Information in Regression Analysis'. *Journal of the American Statistical Association,* LVIII (1963), pp. 401-414.

[68] ———, and Rey, Guido. 'A Quadratic Programming Approach to the Estimation of Transition Probabilities'. *Management Science* (May, 1966), pp. 714-721.

[69] Tiao, George, and Zellner, Arnold. 'Bayes Theorem and the Use of Prior Knowledge in Regression Analysis'. *Biometrika,* LI (1964), pp. 219-230.

[70] ———. 'On the Bayesian Estimation of Multivariate Regression'. *Journal of the Royal Statistical Society,* Series B, XXVI (1964), pp. 277-285.

[71] Williamson, J. G., and Swanson, J. A. 'The Growth of Cities in the American Northeast, 1820-1870'. *Explorations in Entrepreneurial History* (Fall, 1966).

[72] Wolfe, P., 'The Simplex Method for Quadratic Programming'. *Econometrica,* Vol. 27 (1959), pp. 382-398.

[73] Zellner, A. 'An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias'. *Journal of the American Statistical Association,* LVII (1962), pp. 348-368.

[74] ———, 'Linear Regression with Inequality Constraints on the Coefficients'. Report 6109 of the International Center for Management Science, 1961.

[75] ———, and Tiao, George C. 'Bayesian Analysis of the Regression Model with Autocorrelated Errors'. *Journal of the American Statistical Association,* LIX (1964), pp. 763-768.