MORE ON LEAST SQUARES ESTIMATION OF THE TRANSITION MATRIX IN A STATIONARY FIRST-ORDER MARKOV PROCESS FROM SAMPLE PROPORTIONS DATA

TIMOTHY W. McGUIRE*
CARNegie-MELLON UNIVERSITY

Miller suggested ordinary least squares estimation of a constant transition matrix; Madansky proposed a relatively more efficient weighted least squares estimator which corrects for heteroscedasticity. In this paper an efficient generalized least squares estimator is derived which utilizes the entire covariance matrix of the disturbances. This estimator satisfies the condition that each row of the transition matrix must sum to unity. Madansky noted that estimates of the variances could be negative; a method for obtaining consistent non-negative estimates of the variances is suggested in this paper. The technique is applied to the hypothetical sample data used by Miller and Madansky.

The notation used follows closely that in Miller [1952], Kao [1953], Goodman [1953], and Madansky [1959]. Let
\[ M_i = [m_{i,1}, \ldots, m_{i,n-1}] \]
and
\[ M = [M_1, \ldots, M_n] \]; where
\[ m_{i,j} \]
is the observed proportion on trial \( j \) in alternative category \( i \) of a multinomial population based on a sample of size \( S \), and let \( \mu = E(M) \). Also, let \( T_i = [t_{i,1}, \ldots, t_{i,n}] \) and
\[ T = [T_1, \ldots, T_a] \]
where \( t_{i,j} \) is the probability of an observation in category \( j \) on one trial occurring in category \( i \) on the subsequent trial. Since the elements of each column of \( T \) are a mutually exclusive and exhaustive set of probabilities,

\[ \sum_{i} T_i = 1 \],

where \( 1 \) is an \( a \)-component vector with each element equal to unity, and

\[ t_{i,j} \geq 0, \quad i, j = 1, \ldots, a. \]

Let \( N_i = [m_{i,2}, \ldots, m_{i,n}] \), \( N = [N_1, \ldots, N_a] \), \( \nu_i = E(N_i) \), and \( \nu = E(N) \); then a stationary first-order Markov process is defined by

\[ \nu = T\mu. \]

The problem is to estimate \( T \) given only \( M \) and \( N \). This problem may be cast as a linear statistical model by writing

\[ N = TM + C, \]

*I am indebted to a referee for his thoughtful suggestions on content and style.
where \( C \) is an \( a \times (n - 1) \) random matrix with typical element \( c_{i,j} \) (in the previously cited works, \( C \) is defined by \( C = TM - N \)). Miller [1952] was the first to apply ordinary least squares to each row of (3) to estimate the transition matrix one row at a time. Kao [1953] noted an error in Miller's derivation, and Goodman [1953] provided a correct derivation. Goodman proved that the estimated transition matrix satisfies condition (1). He also noted that the least squares estimates of the transition probabilities do not necessarily satisfy the non-negativity requirement (2) and suggested an estimating procedure which does satisfy this condition. Madansky [1959] showed that the disturbances \( c_{i,j} (j = 2, \ldots, n) \) associated with the \( i \)th row of (3) are heteroscedastic, and he derived a weighted least squares estimator which he proved to be more efficient than the estimator proposed by Miller. However, even the Madansky estimator is not efficient, since it ignores the covariance among the disturbances on each trial. To demonstrate this deficiency and to derive an efficient estimator of \( T \) it is convenient to modify the notation. First, however, it is necessary to examine the properties of model (3) in somewhat more detail. Since each column of \( M \) and \( N \) contains a full set of sample proportions,

\[
1'J = 1'JN = 1'J_{a-1}.
\]

Thus premultiplication of each side of (3) by \( 1'J \) gives

\[
1'JN = 1'JTM + 1'JC,
\]

or

\[
1'J_{a-1} = 1'JM + 1'JC \quad \text{[by (1) and (4)]}
\]

\[
= 1'J_{a-1} + 1'JC_{a-1} \quad \text{[by (4)]}
\]

whence

\[
1'JC = 0_{a-1},
\]

where \( 0_{a-1} \) is an \( (a - 1) \)-component vector with elements equal to zero. Thus the disturbances associated with each trial are linearly dependent; all of the information provided by any trial is contained in at most any \( (a - 1) \) of the observations for that trial. Consequently, with no loss of generality the \( a \)th observation may be deleted for each trial. Define

\[
N^* = [N'_1, \ldots, N'_{a-1}],
\]

\[
T^* = [T'_1, \ldots, T'_{a-1}],
\]

\[
C_i = [c_{i,2}, \ldots, c_{i,n}],
\]

\[
C^* = [C'_1, \ldots, C'_{a-1}],
\]

and