

Development of Specific Hypothesis Tests  
for Estimated Markov Chains

Christina M.L. Kelton  
Department of Economics  
Wayne State University

W. David Kelton  
Department of Industrial and Operations Engineering  
The University of Michigan

Technical Report 84-6

March 1984

## Development of Specific Hypothesis Tests for Estimated Markov Chains

Christina M.L. Kelton  
Department of Economics  
Wayne State University

W. David Kelton  
Department of Industrial and Operations Engineering  
The University of Michigan

We develop and evaluate the validity and power of two specific tests for the transition probabilities in a Markov chain estimated from aggregate frequency data. The two null hypotheses considered are (1) constancy of the diagonal elements of the one-step transition probability matrix and (2) an arbitrarily chosen transition probability's being equal to a specific value. The formation of tests uses a general framework for statistical inference on estimated Markov processes; we also indicate how this framework can be used to form tests for a variety of other hypotheses. The validity and power performance of the two tests formed in this paper are examined in factorially designed Monte Carlo experiments. The results indicate that the proposed tests lead to Type I error probabilities which are close to the desired levels and to high power against even small deviations from the null hypotheses considered.

**KEYWORDS:** Markov Chains; Estimation and Testing; Aggregate Frequency Data; Monte Carlo Experiment; Test Validity; Test Power

## 1. INTRODUCTION

Stochastic processes often provide useful models of real world phenomena in which randomness is an important feature. Many applications are found in the physical and life sciences. See, for example, Berlin *et al.* (1979) for the use of a Markov illness-death model; Read and Ashford (1968) and Kalbfleisch *et al.* (1983) for modeling an organism's path through consecutive stages of development; and Shah (1976) for consideration of drug transport through different body components. Markovian analyses have also been employed for selection and evaluation of mental health programs (see Trinkl 1974; Meredith 1976). For applications to hydrologic systems, see Denny *et al.* (1974) and Yakowitz (1973, 1976). Further, demographers have employed these models in studying human and industrial migration (Spilerman 1972; Collins 1972; Lever 1972), and brand shift and firm size change have been so modeled by marketing researchers and economists (Telser 1962a, 1962b; Kelton and Kelton 1982; Adelman 1958; Collins and Preston 1961; Hallberg 1969).

Valid modeling of an existing real world process requires that the parameters of the chosen model be estimated from data which are available or can be collected. Furthermore, it is often of interest to perform statistical inference about the true model parameters, in addition to obtaining point estimates.

A very simple stochastic model which has found wide use is a stationary Markov chain. The number of parameters to be estimated may be few relative to more sophisticated models, yet the Markov chain offers considerable predictive ability. When observing the operation of a Markov chain, there are essentially two different kinds of data which might be collected. Micro data are the most desirable, since we observe the number of actual individual transitions from each state  $i$  to each state  $j$ . If one has access to these

data, maximum likelihood estimators of the state transition probabilities can be computed in a straightforward manner; furthermore, hypothesis tests are available (see Anderson and Goodman 1957; Billingsley 1961a, 1961b; Kullback et al. 1962).

On the other hand, micro data are frequently unavailable. Instead, we may have to rely on macro data (also called aggregate frequency data), where we know only  $n_i(t)$ , the number of entities occupying state  $i$  at time  $t$ . Here, the actual state-to-state transitions are not observed. Such data arise, for example, in the Census of Population, market share reports or trade bulletins, and in the Census of Manufactures. Whereas the problem of point estimation has received considerable attention for these data, development of hypothesis tests has been lacking. Perhaps the most common method of point estimation is the method of restricted least squares, described briefly in Section 2; see also Miller (1952), Madansky (1959), Lee et al. (1977), MacRae (1977), Kelton (1981), Kalbfleisch et al. (1983), and Kalbfleisch and Lawless (1984). Bedall (1978) proposed certain specialized chi-square tests based on analogy to frequency table analysis techniques.

In this paper, we also propose specific hypothesis tests from macro data, but these are based on a general framework developed in Kelton and Kelton (1984a). The value of such an approach is that this general framework could be used to develop any number of specific tests of interest in a given study; this paper shows how to use this framework in two cases. The first null hypothesis considered is that all diagonal elements  $p_{ii}$  of the one-step transition probability matrix are the same, and the second hypothesis is that a particular  $p_{ij}$  equals a particular value. We then investigate the validity, robustness, and power of these two tests by carrying out fairly extensive, designed Monte Carlo studies.

In Section 2 we briefly review least squares point estimation from macro

data and the general hypothesis test development framework from Kelton and Kelton (1984a). Section 3 develops tests for the two cases of interest, and gives an example of applying these kinds of tests to a brand shift model for the brewing industry. The validity and power of the tests proposed in this paper are empirically investigated in Section 4, and a general discussion follows in Section 5.

## 2. REVIEW OF LEAST SQUARES ESTIMATION AND HYPOTHESIS TEST FRAMEWORK

In this section, we review the least squares method of obtaining point estimates for the transition probabilities  $p_{ij}$  of moving from state  $i$  to state  $j$  in one step. For a discussion of alternative estimators (such as those obtained by weighted least squares) and their properties, see Lee *et al.* (1977), MacRae (1977), and Kalbfleisch and Lawless (1984). We also briefly review the general hypothesis test methodology from Kelton and Kelton (1984a). We will need the following notation:

$R$  = the number of states

$T$  = the number of transitions

$N$  = the number of entities (i.e., individuals) observed

$n_i(t)$  = the number of the  $N$  entities in state  $i$  at time  $t$

$y_i(t) = n_i(t)/N$  = the proportion of the  $N$  entities in state  $i$  at time  $t$

$y(t) = [y_1(t), \dots, y_R(t)]$

$P$  = the one-step transition probability matrix with  $(i,j)$ th element

$P_{ij}$

$\pi(t)$  = a  $1 \times R$  row vector with  $i$ th entry's being the (true) probability that an entity occupies state  $i$  at time  $t$ .

## 2.1. Least Squares Point Estimator of P

The idea behind the LS estimators of the  $p_{ij}$ 's is based on noticing that  $y(t)$  is an unbiased estimator of  $\pi(t)$ , and recalling the fundamental property of Markov chains that  $\pi(t) = \pi(t-1)P$ . Thus, we would expect that  $y(t) \cong y(t-1)P$ , and the LS estimators minimize the sum of squared deviations between  $y_j(t)$  and  $\sum_{i=1}^R y_i(t-1) p_{ij}$  for states  $j = 1, \dots, R-1$  and time periods  $t = 1, \dots, T$ , subject to nonnegativity of the  $p_{ij}$ 's and to the row-sum constraints on the estimated P. (State R is omitted to avoid redundancy.)

This problem can be phrased as a standard quadratic programming (QP) problem with linear inequality constraints, as follows. Let

$$p = [p_{11}, \dots, p_{R1}, \dots, p_{1,R-1}, \dots, p_{R,R-1}]',$$

$$y = [y_1(1), \dots, y_1(T), \dots, y_{R-1}(1), \dots, y_{R-1}(T)]',$$

$$X^* = \text{a } T \times R \text{ matrix with } (t+1, i)\text{th element } y_i(t), \text{ for } t = 0, \dots, T-1, \text{ and } i = 1, \dots, R,$$

and

$$X = \text{a } T(R-1) \times R(R-1) \text{ block-diagonal matrix with } R-1 \text{ copies of } X^* \text{ along the diagonal,}$$

where ' denotes transposition. With this notation, the approximations of the previous paragraph may be stated as  $y \cong Xp$ . (Note that p contains all the parameters to be estimated, since  $p_{iR} = 1 - \sum_{j=1}^{R-1} p_{ij}$ .) The LS estimator of p is the solution  $\hat{p}$  to the QP

$$\begin{aligned} & \min_p (y - Xp)' (y - Xp) \\ & \text{subject to } p_{ij} \geq 0 \text{ and } \sum_{j=1}^{R-1} p_{ij} \leq 1, \text{ for } i = 1, \dots, R. \end{aligned}$$

From the solution, we obtain a sum of squared residuals  $SSR = (y - X\hat{p})' (y - X\hat{p})$  which is used in our hypothesis testing framework. This QP problem may be solved by a simplex-like pivotal algorithm, such as Lemke's (1968).

## 2.2. Hypothesis Tests for P

A general hypothesis testing framework proposed in Kelton and Kelton (1984a) is now briefly reviewed. A null hypothesis  $H_0$  generally imposes some additional restrictions on the  $p_{ij}$ 's. Taking into account these restrictions, corresponding  $y$ ,  $X$ , and  $p$  matrices are defined (which may differ from those above), the associated QP is solved, and we obtain a sum of squared residuals  $SSR_R$  for the restricted model. From the same data, an unrestricted model is fit where  $y$ ,  $X$ , and  $p$  are defined ignoring the restrictions imposed by  $H_0$ . A sum of squared residuals  $SSR_U$  is obtained for the unrestricted model. While it will always be the case that  $SSR_R \geq SSR_U$ , the test attempts to ascertain whether the difference is significant.

Let

$q$  = the number of (additional) restrictions imposed by  $H_0$ ,  
 $v$  = the degrees of freedom (number of independent observations minus number of parameters estimated) in the unrestricted model,

and

$$F_{q,v} = [(SSR_R - SSR_U)/q] / (SSR_U/v).$$

Following Chow (1960), Fisher (1970), and Theil (1971), we proposed that  $F_{q,v}$  be treated as having an F distribution with  $(q,v)$  degrees of freedom (d.f.), under  $H_0$ . Many assumptions for the general linear model (e.g.,  $y-Xp$  is a vector of uncorrelated, homoskedastic, normally distributed errors), however, are violated so that the actual distribution of  $F_{q,v}$  under  $H_0$  is unknown; Monte Carlo studies for robustness are thus appropriate.

Using factorially designed Monte Carlo experiments, we have already evaluated three such hypothesis tests formed in the above manner. These tests

were aimed at examining the adequacy and validity of the simple stationary Markov chain model for all entities alike. The Type I error probability estimates were generally close to the desired levels; for example, the overall average observed percentage of rejections at the (desired) 10%, 5%, and 1% levels were 9.97%, 5.61%, and 1.32%, respectively. Moreover, the three tests proved fairly powerful to various considered violations of the null hypotheses. We refer the reader to Kelton and Kelton (1984a, 1984b) for details.

### 3. FORMATION OF TESTS

We now use the general methodology of Section 1 to develop tests for two specific  $H_0$ 's. The discussion should also give an indication of how tests for other hypotheses of interest can be formed.

#### 3.1. Constant Diagonal Probabilities

The null hypothesis is  $H_0: p_{ii} = p_{jj}$  for  $i = 1, \dots, R$  and  $j = 1, \dots, R$ . Under this  $H_0$ , the probability that an entity will remain in its current state for the next time period is the same for all states. Alternatively, we can think of  $H_0$  as stating that the outmovement probabilities  $1-p_{ii}$  are the same for all states  $i$ . This is of interest, for example, in studying geographic migration of population or industry, and in consumer brand selection to detect differences in repeat- and transfer-purchase probabilities across brands.

$H_0$  imposes the following set of  $q = R-1$  (additional) restrictions on  $P$  (recall that the  $p_{iR}$ 's are never explicitly estimated):

$$p_{11} = p_{22} = \dots = p_{R-1,R-1} = 1 - \sum_{j=1}^{R-1} p_{Rj}.$$

As  $p_{R1}, \dots, p_{R,R-1}$  must be estimated anyway, we retain them in the (restricted) model and drop  $p_{11}, p_{22}, \dots, p_{R-1,R-1}$ . Thus, the parameters to



be estimated compose

$$p = [p_{21}, p_{31}, \dots, p_{R1}, p_{12}, p_{32}, \dots, p_{R2}, \dots, p_{1, R-1}, \dots, p_{R-2, R-1}, p_{R, R-1}]'$$

Here  $y(t) \equiv y(t-1) P$  may be expressed equivalently as

$$y_j(t) \equiv \sum_{\substack{i=1 \\ i \neq j}}^R y_i(t-1) p_{ij} + y_j(t-1) (1 - \sum_{k=1}^{R-1} p_{Rk}), \quad (1)$$

for  $j = 1, \dots, R-1$  and  $t = 1, \dots, T$ . Writing (1) in matrix notation leads to  $y \equiv Xp$ , where

$$y = [y_1(1)-y_1(0), \dots, y_1(T)-y_1(T-1), \dots, y_{R-1}(1)-y_{R-1}(0), \dots, y_{R-1}(T)-y_{R-1}(T-1)]'$$

and  $X$  is constructed as follows. For  $t = 0, \dots, T-1$  define

$$X_1^* \text{ to have } (t+1)\text{st row } [y_2(t), \dots, y_{R-1}(t), y_R(t)-y_1(t)],$$

$$X_{R-1}^* \text{ to have } (t+1)\text{st row } [y_1(t), \dots, y_{R-2}(t), y_R(t)-y_{R-1}(t)],$$

and, for  $k = 2, \dots, R-2$ , define

$$X_k^* \text{ to have } (t+1)\text{st row } [y_1(t), \dots, y_{k-1}(t), y_{k+1}(t), \dots, y_{R-1}(t), y_R(t)-y_k(t)].$$

For  $k = 1, \dots, R-1$ , define the  $T \times (R-1)$  matrix  $Z_k$  to have  $(t+1)$ st row  $[0, \dots, 0, -y_k(t)]$ , for  $t = 0, \dots, T-1$ . Finally, let

$$X = \begin{bmatrix} X_1^* & Z_1 & Z_1 & \cdots & Z_1 \\ Z_2 & X_2^* & Z_2 & \cdots & Z_2 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ Z_{R-1} & \cdots & Z_{R-1} & & X_{R-1}^* \end{bmatrix}.$$

The restricted QP is then

$$\begin{aligned} & \min_p (y - Xp)'(y - Xp) \\ & \text{subject to } p_{ij} \geq 0 \text{ for } j = 1, \dots, R-1 \\ & \quad \quad \quad i = 1, \dots, R, i \neq j \\ & \quad \quad \quad \sum_{\substack{j=1 \\ j \neq i}}^{R-1} p_{ij} + (1 - \sum_{k=1}^{R-1} p_{Rk}) \leq 1 \text{ for } i = 1, \dots, R-1 \\ & \quad \quad \quad \sum_{j=1}^{R-1} p_{Rj} \leq 1. \end{aligned}$$

These constraints can be further simplified and collapsed into matrix form for a standard QP statement. Solving this QP leads to  $SSR_R$ .

The unrestricted model in this case is exactly as stated in Section 2.1, and  $v = (T-R)(R-1)$ .  $F_{q,v}$  is computed as in Section 2.2.

### 3.2. A Specified Probability

In some situations there may be special interest in the probability of one-step transition from a specified state  $i_0$  to a specified state  $j_0$ , e.g. population migration from the Northeast to the South in the U.S. The appropriate null hypothesis is thus  $H_0: p_{i_0, j_0} = c$ , where  $c$  is a fixed, specified constant between 0 and 1; this  $H_0$  imposes only one additional restriction on  $P$ , so  $q = 1$ . In this case,  $p$  is as in Section 2.1, except that  $p_{i_0, j_0}$  does not appear (so  $p$  is a row vector containing  $R(R-1)-1$  elements). Now,  $y(t) \equiv y(t-1)P$  is equivalent to

$$y_j(t) \equiv \sum_{i=1}^R y_i(t-1)p_{ij}, \text{ for } j \neq j_0 \quad (2)$$

$$y_{j_0}(t) \equiv \sum_{\substack{i=1 \\ i \neq i_0}}^R y_i(t-1)p_{ij_0} + y_{i_0}(t-1)c, \quad (3)$$

for  $j = 1, \dots, R-1$  and  $t = 1, \dots, T$ . Transforming (2) and (3) to matrix form with all  $j$  and  $t$  included leads to  $y \equiv Xp$ , with  $y$  and  $X$  defined as follows. Let  $y^*$  be the  $y$  in Section 2.1, and let  $z$  be a  $T(R-1) \times 1$  vector with  $y_{i_0}(0), y_{i_0}(1), \dots, y_{i_0}(T-1)$  in locations  $(j_0-1)T + 1, (j_0-1)T + 2, \dots, j_0T$ , respectively, and zeros otherwise. Then we let  $y = y^* - cz$ .  $X$  is defined as in Section 2.1, except that column  $(j_0-1)R + i_0$  is deleted. The restricted QP is

$$\begin{aligned} & \min_P (y - Xp)'(y - Xp) \\ & \text{subject to } p_{ij} \geq 0 \text{ for } i = 1, \dots, R, \\ & \qquad \qquad \qquad j = 1, \dots, R-1 \\ & \qquad \qquad \qquad (i,j) \neq (i_0, j_0) \\ & \qquad \qquad \qquad \sum_{j=1}^{R-1} p_{ij} \leq 1 \text{ for } i = 1, \dots, R \\ & \qquad \qquad \qquad i \neq i_0 \\ & \qquad \qquad \qquad \sum_{\substack{j=1 \\ j \neq j_0}}^{R-1} p_{i_0j} \leq 1 - c. \end{aligned}$$

Again, these constraints can be put into standard QP form, and we solve the problem to obtain  $SSR_R$ . Also, the unrestricted model is as in Section 2.1, with  $v = (T-R)(R-1)$ .

For this test and for that of Section 3.1, the unrestricted fit was made simply by applying the model of Section 2.1. However, that model will not always suffice for the unrestricted fit. We refer the reader to Kelton and Kelton (1984a) for two null hypotheses requiring different treatment for the unrestricted model.

### 3.3. An Example

As one application of the two hypothesis tests proposed above as well as two of three tests from our earlier paper (Kelton and Kelton 1984a), we look at market shares in the brewing industry from 1951 through 1971. The data consist of yearly shares of national production or barrelage for Anheuser-Busch (AB), Miller (M), and all other brewers (O), taken from Keithahn (1978). For a more detailed econometric analysis of the brewing industry and of the rise over time of Anheuser-Busch and Miller in particular, see Kelton and Kelton (1982). The market-share data can be found in Table 1.

The basic unrestricted model of Section 2.1 was fitted to obtain the following estimated one-step transition probability matrix:

$$\begin{array}{c}
 \text{AB} \quad \text{M} \quad \text{O} \\
 \text{AB} \begin{bmatrix} 0.954 & 0.046 & 0.000 \\ 0.349 & 0.651 & 0.000 \\ 0.000 & 0.008 & 0.992 \end{bmatrix}. \\
 \text{M} \\
 \text{O}
 \end{array}$$

Before applying the two tests above, we used two of the three tests developed in Kelton and Kelton (1984a) to assess model adequacy and validity. Note that ignorance of  $N$ , interpreted roughly in this case as the number of beer consumers, does not preclude application of the proposed estimation and testing procedure. Testing for stationary transition probabilities, we obtained an  $F$  statistic of 1.114, which, with (6,28) d.f., yields a  $p$ -value of 0.379 (i.e., under the null hypothesis of stationarity, the probability of obtaining a test statistic in excess of 1.114 is 0.379). Thus, the required assumption of stationary transition probabilities appears to be quite safe for these data. The test for a zero-order process yielded a test statistic of 220.953, which, with (4,34) d.f., is highly significant, indicating strong evidence that the process is autocorrelated.

The null hypothesis of equal diagonal probabilities from Section 3.1 led to a (restricted) transition probability matrix of

$$\begin{array}{c} \text{AB} \quad \text{M} \quad \text{O} \\ \text{AB} \begin{bmatrix} 0.993 & 0.006 & 0.001 \\ \text{M} \begin{bmatrix} 0.007 & 0.993 & 0.000 \\ \text{O} \begin{bmatrix} 0.007 & 0.000 & 0.993 \end{bmatrix} \end{bmatrix} \end{bmatrix}.$$

Comparing this matrix with the basic unrestricted one above, we note that the outmovement probability for Miller has been forced to a considerably different value. The test statistic, with (2,34) d.f., was 3.815 for a p-value of 0.032. This null hypothesis of equal repeat-purchase probabilities thus appears to be suspect.

The test of Section 3.2 was used to investigate whether it is safe to assume that the repeat-purchase probability for Miller in particular is 0.5, i.e., whether Miller retains 50% of its consumers over one year. (If we numbered Miller as state 2, we are testing  $p_{22} = 0.5$ .) The estimated matrix under this restriction is as follows:

$$\begin{array}{c} \text{AB} \quad \text{M} \quad \text{O} \\ \text{AB} \begin{bmatrix} 0.922 & 0.078 & 0.000 \\ \text{M} \begin{bmatrix} 0.455 & 0.500 & 0.045 \\ \text{O} \begin{bmatrix} 0.000 & 0.010 & 0.990 \end{bmatrix} \end{bmatrix} \end{bmatrix}.$$

Compared with the first, unrestricted model, this matrix would not appear to be substantially different. Indeed, the observed F statistic here is 0.677 with a p-value of 0.416, with (1,34) d.f. Thus, there is little evidence to reject a 0.5 repeat-purchase probability for Miller.

#### 4. ASSESSMENT OF TEST VALIDITY AND POWER

As discussed in Section 2, the test statistics proposed above need not have the desired F distributions, under the null hypotheses. In this section we report the results of a fairly extensive Monte Carlo study designed to assess how closely the distribution of  $F_{q,v}$  resembles an F distribution. Of particular interest in a testing context is the probability that  $F_{q,v}$  exceeds a critical value obtained from a standard F table; this probability is the actual Type I error incurred in a test with a stated (nominal) significance level.

Under each null hypothesis of Section 3, N independent realizations of the process were simulated for T transitions, the appropriate models were fit to these data, and an observation on the test statistic was obtained. These steps were independently replicated 200 times, yielding 200 independent observations on the test statistic. These values were used in chi-square ( $\chi^2$ ) and Kolmogorov-Smirnov (KS) goodness-of-fit tests for the desired F distribution. As a more direct measure of test robustness, we noted the percentage of the 200 observations which fell above the upper 10%, 5%, and 1% critical values of the proposed F distribution; the average absolute differences of these percentages from their target values were also tallied. The random number generator used for all experiments is that of Lewis, Goodman, and Miller (1969).

Since the performance of the tests might be affected by various parameters of the data, we employed a formal experimental design to specify their levels and combinations. The five experimental factors in this context, are R, T, N,  $\pi(0)$ , and P. A resolution V,  $2^{5-1}$  fractional factorial design was constructed by writing a full  $2^4$  factorial design in the first four factors, and taking the level (sign) for P to be the positive product of the signs of the levels

of the other four factors (see Box et al. 1978). Thus, 16 independent sets of 200 independent test statistics were generated for each null hypothesis. The numerical levels of the factors, including the selected small values for R, were set with some consideration for our experience with the kinds of data typically available in practice. The "-" and "+" levels for R were, respectively, 2 and 4, for T were 25 and 50, and for N were 100 and 500. For  $\pi(0)$ , the "-" level specification was a uniform distribution on the R states, and the "+" level was (0.79, 0.21) when R = 2 and was (0.79, 0.11, 0.05, 0.05) when R = 4; these levels were judged to be "opposite" in the sense that one gives an equiprobable initial state while the other puts a fairly heavy probability mass on a particular state. As for P, the null hypotheses we consider place different restrictions on the  $p_{ij}$ 's, so the values of P will be given separately with each null hypothesis.

To investigate the power properties of the proposed tests, further designed Monte Carlo studies were undertaken, in which the data were generated in violation of  $H_0$ . The design matrix used was the same as for the validity studies, but, because of the fairly large number of alternative hypotheses that we wanted to consider, 100 (rather than 200) replications were made at each design point. The "-" and "+" levels for R, T, N, and  $\pi(0)$  are the same as for validity; the levels for P are discussed separately for each test.

#### 4.1. Constant Diagonal Probabilities

For this test, the "-" and "+" levels of P were

$$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$$

when  $R = 2$ , and, when  $R = 4$ , the two levels for  $P$  were

$$\begin{bmatrix} 0.8 & 0.2 & 0.0 & 0.0 \\ 0.1 & 0.8 & 0.1 & 0.0 \\ 0.0 & 0.1 & 0.8 & 0.1 \\ 0.0 & 0.0 & 0.2 & 0.8 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.2 & 0.6 \end{bmatrix}.$$

(Note that, in accordance with this null hypothesis, the diagonal elements are always constant within a matrix.) These values of  $P$  were chosen, on the one hand, to induce a reluctance to change state, and, on the other hand, to promote somewhat greater outmovement probability.

The average responses as well as the main effects of the factors are presented in Table 2. The value of the  $\chi^2$  test statistic is denoted by  $\chi^2$ , under which we also give the associated p-value. For the KS test, we report

$$D' = D_n[(n)^{.05} + 0.12 + 0.11/(n)^{0.5}],$$

where  $D_n$  is the usual KS test statistic; the statistic  $D'$  was developed by Stephens (1974) to allow the use of a very compact table of critical values.  $C_{10}$ ,  $C_5$ , and  $C_1$  are the percentages of observed  $F_{q,v}$ 's which exceeded the upper 10%, 5%, and 1% critical values, respectively, of the proposed  $F$  distribution.

The average p-value of  $\chi^2$  is 0.36, indicating a generally good fit, although, for 6 of 16 individual runs, the  $\chi^2$  values were significant at the 0.10 level. The  $C_r$ 's are 8.9, 4.7, and 1.2, being quite close to the desired levels of 10, 5, and 1. (A 90% confidence interval failed to cover the target rejection percentage in 12 out of 48 cases.) To demonstrate that these average rejection percentages' being close to their targets is not the result



of averaging some extremely high values with some extremely low values, the absolute values of  $C_{p-r}$  were averaged to obtain 2.3, 1.1, and 0.7; these results are further indication of good Type I error probabilities.

Looking at the main effects in Table 2, it appears that R is an important factor. (This was further confirmed by probability plots.) In particular, better performance is usually obtained when R is small. Thus, if one has a choice in modeling, it would be advantageous to keep the number of states as small as possible to obtain the desired information from the model. There is as well some evidence to suggest that obtaining longer time records (large T) has a desirable effect on test performance. Note the desirable negative effect of T on both  $|C_{10}-10|$  and  $|C_1-1|$ .

The power studies undertaken for this hypothesis test use the same factor levels (for R, T, N, and  $\pi(0)$ ) as the robustness studies. For four of our five designs, the transition matrices were chosen such that  $p_{11} \neq p_{22} = \dots = p_{RR}$ , i.e., only  $p_{11}$  was allowed to deviate from the otherwise constant diagonal; this should provide a lower bound on power. To parameterize the extent of deviation from  $H_0$ , we let  $d = p_{11} - p_{22}$ , which was constant for all matrices within a given design. For the four designs, d took on the values -0.5, -0.3, -0.1, and +0.1, respectively.

In Figure 1, the average (over the 16 design points) rejection percentages are shown as a function of d for 10%, 5%, and 1% tests. The percentages for  $d = 0$  correspond to the validity results from the second column in Table 2. The three power curves have an anticipated shape; power rises as  $|d|$  increases. The test seems fairly powerful against even modest departures from  $H_0$ .

Table 3 presents the average (over the four designs) main effects of the factors on the rejection percentages. Consistent with the validity results above, it is seen that a small state space leads to higher power. Further,

long time records and a large number of entities observed should enhance power.

Finally, a fifth design was conducted which allowed all the diagonal elements  $p_{ij}$  to be unequal. As expected, the rejection percentages in this case were quite large: 97.25 for a 10% test, 94.13 for a 5% test, and 86.31 for a 1% test.

#### 4.2. A Specified Probability

Here we tested the null hypothesis that  $p_{21} = 0.3$  regardless of the other parameter values. When  $R = 2$ , the "-" and "+" levels for  $P$  were taken to be

$$\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix},$$

and, when  $R = 4$ , the "-" and "+" levels of  $P$  were

$$\begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 \\ 0.3 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.7 & 0.1 \\ 0.0 & 0.1 & 0.2 & 0.7 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.5 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 & 0.0 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.2 & 0.2 & 0.5 \end{bmatrix}.$$

Again, the null hypothesis is satisfied in each case, and the levels of  $P$  were chosen to model different outmovement probabilities.

The average  $p$ -value of  $\chi^2$  in Table 4 again indicates reasonable overall fit. For this test, only 5 of 16  $\chi^2$  values are significant at the 0.10 level. The average values of  $C_r$  are again fairly close to their targets, although

they are slightly large in this case. (90% confidence interval coverage was not achieved in 17 out of 48 cases.) The average absolute deviations of  $C_r$  from  $r$  are somewhat larger than for the other test, but are still only a few percentage points. As before, the most important factor on test performance is  $R$ , and it is clear that small  $R$  is desirable in this test. Furthermore, it again appears that long time records could be expected to yield better results. Although it is not clear what practical implications there are for the user, better results are also obtained if the true  $P$  induces higher outmovement probabilities.

For this hypothesis test, violation of  $H_0$  for power investigation involved simply altering row 2 of the transition probability matrix  $P$ , specified in the above validity studies. In four designs,  $p_{21}$  was set, respectively, at 0.1, 0.5, 0.7, and 0.9. Again partial averaged power curves are generated (see Figure 2), and they show that the test is powerful against even small departures from  $H_0$ . We also note from Figure 2 that power increases with deviation of  $p_{21}$  from its hypothesized value, i.e., with  $|p_{21}-0.3|$ , and that the curves pass through the average estimated Type I error probabilities, from the second column in Table 4, when  $H_0$  is true.

Table 5 shows the average (over the four different designs) power performance effects of the factors considered. The policy recommendations are consistent with those suggested by Table 3 (as well as by our validity studies). The large negative values of the main effect of  $R$  imply that increasing the size of the state space reduces power. On the other hand,  $T$  and  $N$  are seen to have positive main effects on power, again suggesting the benefits of long time records and a large number of entities observed.  $P$  and  $\pi(0)$ , the two factors essentially beyond the control of the modeler, have relatively small impacts on power.

## 5. DISCUSSION

In this paper, we have shown how a general hypothesis testing framework can be used to develop specialized tests which may be of interest for a specific estimation and inference problem involving Markov chains with only aggregate frequency macro data. The results of our Monte Carlo studies lead us to anticipate that the proposed tests should be valid in the sense of producing nearly the proper Type I error probability, and should have good power to detect departures from the null hypothesis of interest. The ability to carry out such tests could allow more meaningful use of stochastic models in physical and social sciences. When considering the use of a Markov chain model in practice, we would recommend first that the three tests developed in Kelton and Kelton (1984a) be applied to assess the adequacy and validity of the model. (In that paper we acknowledge, however, that the test for zero-order dependency does not establish the length of memory of the process if it is not Markovian.) Then, we recommend proceeding with more specialized tests such as those in the present paper, or others of interest in different empirical applications.

As the Monte Carlo experiment was designed, we could observe the effects of various modeling choices on test validity and power. There is clear evidence that a small state space is desirable, and some evidence that lower Type I error probabilities as well as higher power can be expected from long time records of data. These recommendations are consistent with those of our earlier paper.

#### ACKNOWLEDGEMENTS

This research was partially supported by a Faculty Development Grant from the Standard Oil Company (Ohio) to the Graduate School of Management at Kent State University. We appreciate the input of Professor John Geweke, Department of Economics, University of Wisconsin-Madison. We also thank the Madison Academic Computing Center of the University of Wisconsin for supplying the program QUADMP used in this study.

## REFERENCES

- ADELMAN, I.G. (1958), "A Stochastic Analysis of the Size Distribution of Firms," Journal of the American Statistical Association, 53, 893-904.
- ANDERSON, T.W., and GOODMAN, L.A. (1957), "Statistical Inference About Markov Chains," The Annals of Mathematical Statistics, 28, 89-110.
- BEDALL, F.K. (1978), "Test Statistics for Simple Markov Chains. A Monte Carlo Study," Biometrical Journal, 20, 41-49.
- BERLIN, B., BRODSKY, J., and CLIFFORD, P. (1979), "Testing Disease Dependence in Survival Experiments with Serial Sacrifice," Journal of the American Statistical Association, 74, 5-14.
- BILLINGSLEY, P. (1961a), Statistical Inference for Markov Processes, Chicago, Illinois: The University of Chicago Press.
- (1961b), "Statistical Methods in Markov Chains," The Annals of Mathematical Statistics, 32, 12-40.
- BOX, G.E.P., HUNTER, W.G., and HUNTER, J.S. (1978), Statistics for Experimenters, New York: John Wiley & Sons, Inc.
- CHOW, G.C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, 28, 591-605.
- COLLINS, L. (1972), Industrial Migration in Ontario: Forecasting Aspects of Industrial Activity through Markov Chain Analysis, Ottawa: Statistics Canada.
- COLLINS, N.R., and PRESTON, L.E. (1961), "The Size Structure of Industrial Firms," American Economic Review, 51, 986-1003.
- DENNY, J.L., KISIEL, C.C., and YAKOWITZ, S.J. (1974), "Procedures for Determining the Order of Dependence in Streamflow Records," Water Resources Research, 10, 947-954.
- FISHER, F.M. (1970), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrica, 38, 361-366.
- HALLBERG, M.C. (1969), "Projecting the Size Distribution of Agricultural Firms -- an Application of a Markov Process with Non-Stationary Transition Probabilities," American Journal of Agricultural Economics, 51, 289-302.
- KALBFLEISCH, J.D., and LAWLESS, J.F. (1984), "Least Squares Estimation of Transition Probabilities from Aggregate Data," to appear in Canadian Journal of Statistics.
- KALBFLEISCH, J.D., LAWLESS, J.F., and VOLLMER, W.M. (1983), "Estimation in Markov Models from Aggregate Data," to appear in Biometrics.
- KEITHAHN, C.F. (1978), The Brewing Industry, Staff Report of the Bureau of Economics, Federal Trade Commission.

- KELTON, C.M.L. (1981), "Estimation of Time-Independent Markov Processes with Aggregate Data: A Comparison of Techniques," Econometrica, 49, 517-518.
- KELTON, C.M.L., and KELTON, W.D. (1982), "Advertising and Intraindustry Brand Shift in the U.S. Brewing Industry," The Journal of Industrial Economics, 30, 293-303.
- KELTON, W.D., and KELTON, C.M.L. (1984a), "Hypothesis Tests for Markov Process Models Estimated from Aggregate Frequency Data," The University of Michigan Department of Industrial and Operations Engineering Technical Report 84-2.
- KELTON, C.M.L., and KELTON, W.D. (1984b, forthcoming), "Markov Process Models: A General Framework for Estimation and Inference in the Absence of State Transition Data," in Mathematical Modelling in Science and Technology, Proceedings of the 4th International Conference on Mathematical Modelling, Zurich, Switzerland: Pergamon Press, 299-304.
- KULLBACK, S., KUPPERMAN, M., and KU, H.H. (1962), "Tests for Contingency Tables and Markov Chains," Technometrics, 4, 573-608.
- LEE, T.C., JUDGE, G.G., and ZELLNER, A. (1977), Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, 2nd ed., Amsterdam: North-Holland Publishing Company.
- LEMKE, C.E. (1968), "On Complementary Pivot Theory," in Mathematics of the Decision Sciences, eds. G.B. Dantzig and A.F. Veinott, Providence, Rhode Island: American Mathematical Society, 95-114.
- LEVER, W.F. (1972), "The Intra-Urban Movement of Manufacturing: A Markov Approach," Transactions, Institute of British Geographers, 56, 21-38.
- LEWIS, P.A.W., GOODMAN, A.S., and MILLER, J.M. (1969), "A Pseudo-Random Number Generator for the System/360," IBM Systems Journal, 8, 136-146.
- MACRAE, E.C. (1977), "Estimation of Time-Varying Markov Processes with Aggregate Data," Econometrica, 45, 183-198.
- MADANSKY, A. (1959), "Least Squares Estimation in Finite Markov Processes," Psychometrika, 24, 137-144.
- MEREDITH, J. (1976), "Selecting Optimal Training Programs in a Hospital for the Mentally Retarded," Operations Research, 24, 899-915.
- MILLER, G.A. (1952), "Finite Markov Processes in Psychology," Psychometrika, 17, 149-167.
- READ, K.L.Q., and ASHFORD, J.R. (1968), "A System of Models for the Life Cycle of a Biological Organism," Biometrika, 55, 211-221.
- SHAH, B.K. (1976), "Data Analysis Problems in the Area of Pharmacokinetics Research," Biometrics, 32, 145-157.
- SPILERMAN, S. (1972), "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain," American Sociological Review, 37, 277-294.

STEPHENS, M.A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," Journal of the American Statistical Association, 69, 730-737.

TELSER, L.G. (1962a), "Advertising and Cigarettes," The Journal of Political Economy, 70, 471-499.

—— (1962b), "The Demand for Branded Goods as Estimated from Consumer Panel Data," The Review of Economics and Statistics, 44, 300-324.

THEIL, H. (1971), Principles of Econometrics, New York: John Wiley & Sons, Inc.

TRINKL, F.H. (1974), "A Stochastic Analysis of Programs for the Mentally Retarded," Operations Research, 22, 1175-1191.

YAKOWITZ, S.J. (1973), "A Stochastic Model for Daily River Flows in an Arid Region," Water Resources Research, 9, 1271-1285.

—— (1976), "Small-Sample Hypothesis Tests of Markov Order, with Application to Simulated and Hydrologic Chains," Journal of the American Statistical Association, 71, 132-136.



Table 1. Share of National Barrelage by Brewer\*

Year	Anheuser- Busch	Miller	Other
1951	0.0653	0.0312	0.9035
1952	0.0711	0.0359	0.8930
1953	0.0780	0.0248	0.8972
1954	0.0700	0.0252	0.9048
1955	0.0661	0.0258	0.9081
1956	0.0690	0.0264	0.9046
1957	0.0725	0.0275	0.9000
1958	0.0827	0.0263	0.8910
1959	0.0920	0.0269	0.8811
1960	0.0964	0.0270	0.8766
1961	0.0956	0.0303	0.8741
1962	0.0991	0.0308	0.8701
1963	0.1002	0.0311	0.8687
1964	0.1051	0.0333	0.8616
1965	0.1179	0.0365	0.8456
1966	0.1302	0.0398	0.8300
1967	0.1452	0.0428	0.8120
1968	0.1651	0.0435	0.7914
1969	0.1609	0.0446	0.7945
1970	0.1819	0.0422	0.7759
1971	0.1876	0.0401	0.7723

\*Source: 1978 Federal Trade Commission Report, Table VIII.

Table 2. Means and Main Effects for Testing Constant Diagonal Probabilities

Response	Means	Main Effects				
		R	T	N	$\pi(0)$	P
$\chi^2$	24.05	11.50	3.10	7.70	0.55	0.15
p-value of $\chi^2$	0.36	-0.45	-0.07	-0.23	-0.04	0.04
D'	1.31	0.81	0.33	0.22	-0.38	0.10
C <sub>10</sub>	8.91	-1.19	0.44	-1.31	-0.44	-1.94
C <sub>5</sub>	4.69	-0.38	0.63	-1.00	0.38	-1.50
C <sub>1</sub>	1.16	0.06	-0.44	-0.69	0.69	-0.56
C <sub>10</sub> -10	2.28	1.06	-0.31	0.69	0.56	0.31
C <sub>5</sub> -5	1.06	0.63	0.13	-0.00	0.38	0.75
C <sub>1</sub> -1	0.72	-0.06	-0.56	-0.06	0.06	-0.19

Table 3. Averaged Main Effects for Power of Test for  
Constant Diagonal Probabilities

Factor	Level of Test		
	10%	5%	1%
R	-45.91	-54.00	-62.47
T	5.91	6.00	9.53
N	5.09	6.63	9.84
$\pi(0)$	15.28	16.06	16.91
P	-11.09	-11.50	-9.84

Table 4. Means and Main Effects for Testing a Specified Probability

Response	Means	Main Effects				
		R	T	N	$\pi(0)$	P
$\chi^2$	24.75	16.25	-5.30	3.25	3.20	-7.80
p-value of $\chi^2$	0.39	-0.40	-0.08	-0.04	-0.01	-0.03
D'	1.12	0.44	-0.19	0.18	-0.11	-0.26
C <sub>10</sub>	12.38	5.75	-2.00	0.25	-2.13	-2.38
C <sub>5</sub>	7.44	4.75	-1.75	1.25	-0.50	-3.13
C <sub>1</sub>	2.13	2.38	-0.25	1.13	0.13	-1.63
C <sub>10</sub> -10	3.75	4.00	-0.75	0.25	-0.13	-2.63
C <sub>5</sub> -5	2.88	4.38	-1.38	0.88	0.38	-2.75
C <sub>1</sub> -1	1.44	2.25	-0.63	0.50	0.50	-1.25

Table 5. Averaged Main Effects for Power of Test for  
a Specified Probability

Factor	Level of Test		
	10%	5%	1%
R	-7.59	-9.94	-13.88
T	7.47	9.31	13.63
N	4.66	6.88	8.63
$\pi(0)$	-1.22	-1.88	-2.13
P	-3.97	-4.19	-5.06

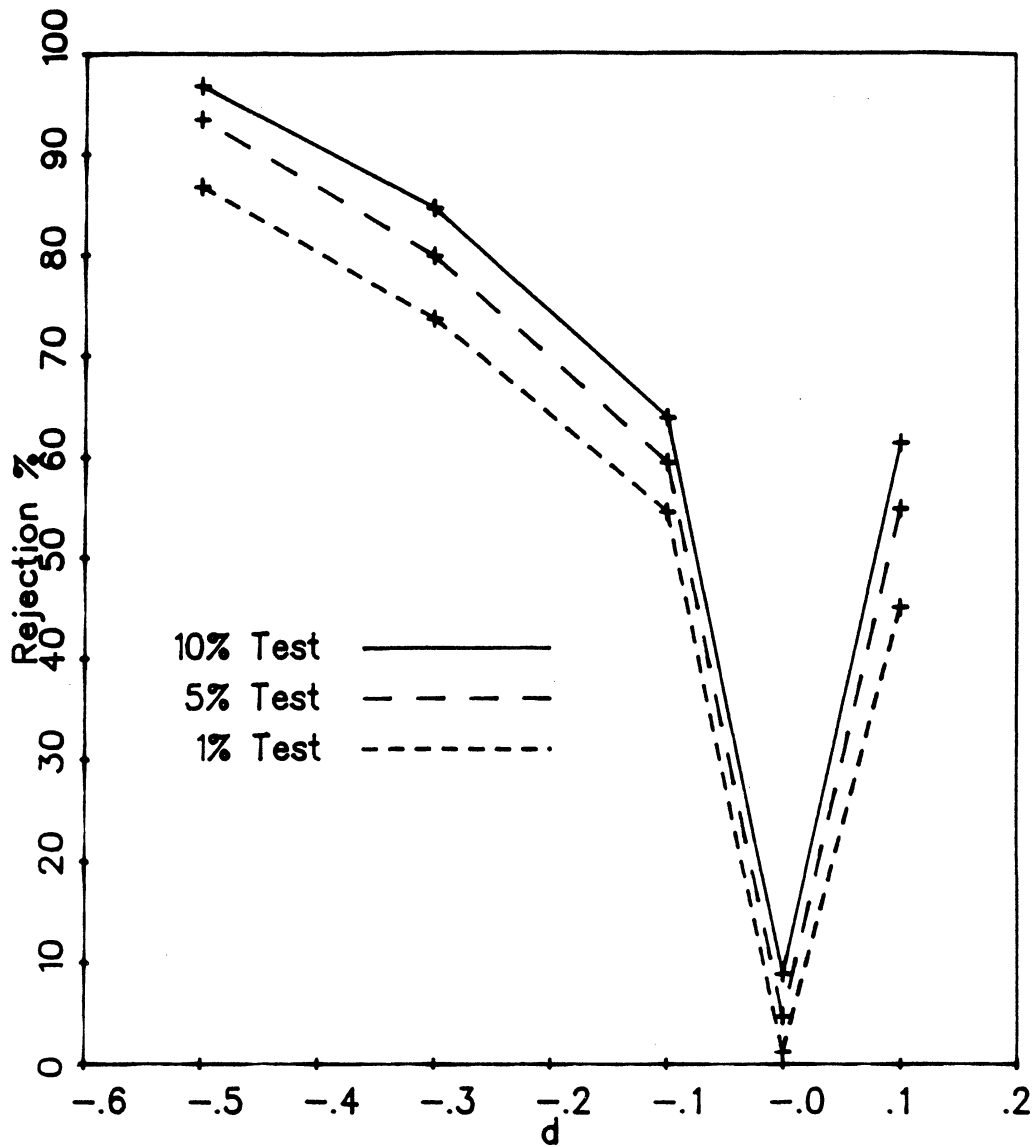


Figure 1. Average Rejection Percentages for Testing  
Constant Diagonal Probabilities

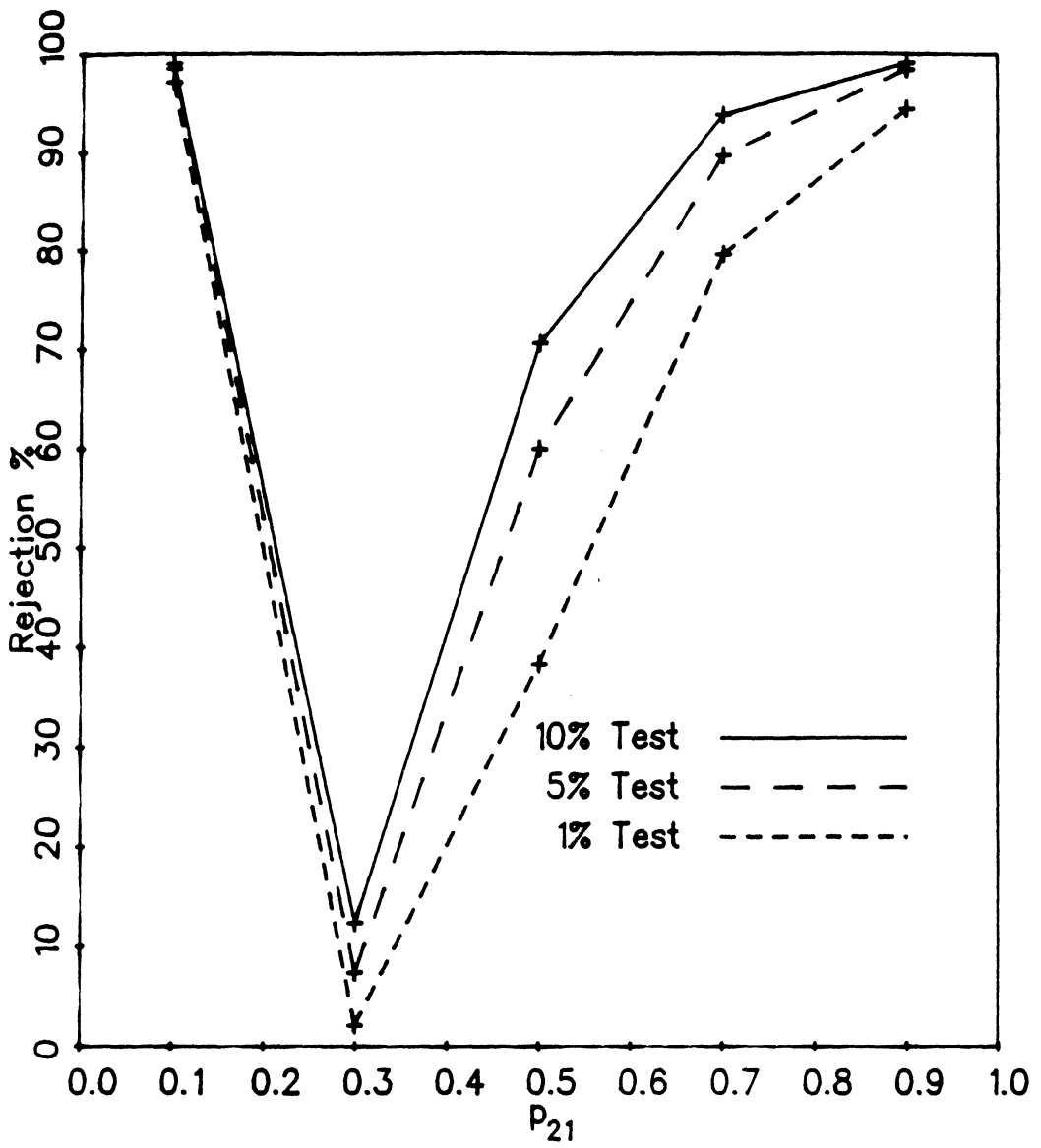


Figure 2. Average Rejection Percentages for Testing  
A Specified Probability

UNIVERSITY OF MICHIGAN



3 9015 03994 8115