

COMPARISON OF INFERENCE TECHNIQUES FOR
MARKOV PROCESSES ESTIMATED FROM
MICRO VS. MACRO DATA

Christina M.L. Kelton
Department of Economics
Wayne State University
Detroit, Michigan

W. David Kelton
Department of Industrial & Operations Engineering
The University of Michigan
Ann Arbor, Michigan

Technical Report 85-9

**Comparison of Inference Techniques for Markov Processes
Estimated from Micro vs. Macro Data**

Christina M.L. Kelton
Department of Economics
Wayne State University
Detroit, Michigan

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

Abstract

We estimate the parameters of a Markov chain model using two types of simulated data: micro, or actual interstate transition counts, and macro aggregate frequency. We compare, by means of Monte Carlo experiments, the validity and power for micro likelihood ratio tests with their macro counterparts, developed previously by the authors to complement standard least-squares point estimates. We consider five specific null hypotheses, including parameter stationarity, entity homogeneity, a zero-order process, a specified probability value, and equal diagonal probabilities. The results from these micro-macro comparisons should help to indicate whether micro panel data collection is justified over the use of simpler state frequency counts.

Key words: Markov processes, macro data, micro data, estimation, hypothesis testing

1. Introduction

Despite a rather impressive list of applications of Markov process models to physical and social phenomena (see, for example, Adelman 1958; Collins and Preston 1961; Telser 1962a, 1962b; Collins 1972; Lever 1972; Spilerman 1972; Yakowitz 1973, 1976; Trinkl 1974; Meredith 1976; Shah 1976; Kelton and Kelton 1982; Kalbfleisch, Lawless, and Vollmer 1983; and Kelton 1984), many researchers have been reluctant to employ a stochastic-process methodology. The usefulness of this type of modeling has been generally restricted by the difficulty in performing statistical inference, such as hypothesis tests, from the results. For empirical implementation of Markov models, there are essentially two types of potentially available data. If one has access to *micro* data or actual interstate transition counts, maximum likelihood estimators of the stationary or nonstationary state transition probabilities can be computed in a straightforward manner (see Section 2). These estimators have been shown to be consistent, asymptotically unbiased, and asymptotically normal. Furthermore, likelihood ratio and chi-square hypothesis tests have been developed (Anderson and Goodman 1957; Billingsley 1961a, 1961b; Kullback, Kupperman, and Ku 1962).

On the other hand, micro data are not always available, due to individual privacy or business disclosure rules for social processes or simply due to high collection costs. Instead, we may have to rely on *macro*, aggregate frequency (for example, census) data, where we know only the number of entities occupying a given state at a given time. Actual state-to-state transitions are not observed. For the case of macro data, perhaps the most common method of point estimation is restricted least squares (implemented by a standard quadratic programming procedure and reviewed briefly in Section 3). See Miller (1952), Madansky (1959), Lee, Judge, and Zellner (1977), MacRae (1977), Kelton (1981), Kalbfleisch, Lawless, and Vollmer (1983), and Kalbfleisch and Lawless (1984) for development and refinement of the least squares procedure. Bedall (1978) proposed certain specialized chi-square tests based on analogy to frequency table analysis techniques. Kelton and Kelton (1984a, 1984b, 1984c) proposed a general methodology, following Chow (1960), Fisher (1970), and Theil (1971), for devising hypothesis tests when only macro data are available. The methodology is described in Section 3 below. This general framework was used to develop three tests for the adequacy of the stationary Markov chain model (first-order process, stationarity, and homogeneity) as well as to develop tests of a more specialized nature — constant diagonal elements and a probability equal to a specified value. The distributions of the test statistics were investigated in factorially designed Monte Carlo experiments. In general, it was found that treating the test

statistics as having F distributions with appropriate degrees of freedom under the null hypotheses of interest led to rejection proportions close to the desired levels. Additional Monte Carlo results indicated favorable power of the proposed tests.

Nevertheless, despite their seemingly good performance, the macro estimation and testing procedures are based on inferior, aggregated data, and it is of some interest to learn whether they compare favorably with estimates and tests based on individual transition counts. Lawless and McLeish (1984) examined the loss of information due to aggregation of micro panel data and found that, although the aggregated data were generally less informative than micro data, they still, in some cases, could give rather good estimates and predictions. For the sampling experiments of Lee, Judge, and Zellner (1977), one notes that the root mean square errors for the restricted least squares estimates are consistently larger for various sample sizes than for the micro maximum likelihood estimates, (although both estimation procedures performed better than unrestricted least squares, according to the mean square error criterion).

In our study, we are also interested in this micro-macro comparison, but rather than compare point estimator performance, we compare power and size of hypothesis tests for various null hypotheses of interest. In the next section, we develop five specific tests to complement micro maximum likelihood estimates, which are based on the likelihood ratio criterion. Section 3 reviews the macro-data least squares estimation procedure and the accompanying hypothesis-testing methodology from Kelton and Kelton (1984a, 1984b, 1984c). In Section 4, we present the results from factorially designed simulations which allow a comparison between micro and macro test performance. Finally, a general discussion may be found in Section 5, along with several recommendations for implementation of micro or macro testing procedures.

2. Micro Data

Let N be the number of entities (for example, individuals), and assume that we can observe the sequence of states occupied by each entity at times $t = 0, 1, \dots, T$. Let $n_{ij}(t)$ be the number of entities in state i at time $t-1$ and in state j at time t , and let $n_{ij} = \sum_{t=1}^T n_{ij}(t)$ be the total number of one-step transitions from state i to state j .

Estimation

To obtain maximum likelihood estimates of the stationary probabilities (the p_{ij} 's), the log of the

relevant likelihood function (multiplied by a constant) is maximized, subject to the row-sum constraints that all rows of the probability matrix sum to one:

$$\max_{p_{ij}} \left(\ln \prod_{i,j} p_{ij}^{n_{ij}} \right) - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right), \quad (1)$$

where λ_i is the row-specific Lagrange multiplier. Unless otherwise noted, the indices on all summations and products run from 1 through R , where R is the number of states in the system. The maximum likelihood estimator is simply $\hat{p}_{ij} = n_{ij} / \sum_k n_{ik}$.

Testing

Here, the likelihood ratio criterion of Anderson and Goodman (1957) is used to develop specifically the five tests mentioned in the introduction, whose performance can be directly compared with that of the macro tests of Section 3. Test development proceeds by adding (or loosening, in some cases) constraints as necessary in the constrained maximization problem (1). We discuss briefly estimates obtained under five different null hypotheses. The first two are fairly specialized, while the last three are basic assumptions of the Markov chain model, and may be tested in order to assess the validity of the model for the process of interest.

Constant Diagonal Probabilities

For the null hypothesis $H_0 : p_{ii} = p_{jj} (= p_D)$ for all i, j , (for example, all brands' of a product having the same repeat-purchase probabilities), our problem becomes

$$\max_{p_{ij}} \left(\ln \prod_{i,j} p_{ij}^{n_{ij}} \right) - \sum_i \lambda_i \left(\sum_{j \neq i} p_{ij} + p_D - 1 \right), \quad (2)$$

where p_D has been substituted for p_{ii} in the constraint. When we solve this problem, we obtain

$$\hat{p}_D = \sum_i n_{ii} / (NT)$$

and

$$\hat{p}_{ij} = \left(n_{ij} / \sum_{k \neq i} n_{ik} \right) \left(\sum_h \sum_{k \neq h} n_{hk} / (NT) \right) \quad \text{for } i \neq j,$$

where NT is the total number of transitions in the system, between any state i and any state j at any time t .

A Specified Probability

We now look at the hypothesis $H_0 : p_{i_0 j_0} = c$ for fixed i_0 and j_0 and specified constant c . The new constrained maximization is

$$\max_{p_{ij}} \left(\ln \prod_{i,j} p_{ij}^{n_{ij}} \right) - \sum_{i \neq i_0} \lambda_i \left(\sum_j p_{ij} - 1 \right) - \lambda_{i_0} \left(\sum_{j \neq j_0} p_{i_0 j} + c - 1 \right), \quad (3)$$

replacing $p_{i_0 j_0}$ by c in the i_0 th row's constraint. Again, we obtain rather intuitive estimators, with those for the i_0 th row scaled by a factor of $1 - c$:

$$\hat{p}_{ij} = n_{ij} / \sum_k n_{ik} \quad \text{for } i \neq i_0,$$

and

$$\hat{p}_{i_0 j} = \left(n_{i_0 j} / \sum_{k \neq j_0} n_{i_0 k} \right) (1 - c) \quad \text{for } j \neq j_0.$$

A Zero-Order Process

To test the hypothesis that the process is zero order (current system state independent of past state) against a higher order alternative (with the Markov chain's being a first order process), we examine $H_0 : p_{ij} = p_j$ for all i, j , i.e., all rows of the transition probability matrix are identical. Solving

$$\max_{p_j} \left(\ln \prod_j p_j^{n_j} \right) - \lambda \left(\sum_j p_j - 1 \right), \quad (4)$$

we find that

$$\hat{p}_j = n_j / (NT),$$

where $n_j = \sum_i n_{ij}$. With micro data, it is also possible to test for a second- or higher-order process, whereas macro data do not permit such tests.

Stationarity

Although, following Anderson and Goodman (1957), a more general micro-data stationarity test can be developed, we formulate our null hypothesis in this case to allow direct comparison with the macro-data test, $H_0 : P(e) = P(f)$ where $P(e)$ or $[p_{ij}(e)]_{R \times R}$ is the probability matrix for the

early time records, $t = 1, \dots, T/2$, and $P(f)$ for the later times $t = T/2 + 1, \dots, T$. (Arbitrarily, T is divided in half for this test.) Thus, we can solve for nonstationary estimates:

$$\max_{p_{ij}(t)} \left(\ln \prod_t \prod_{i,j} p_{ij}(t)^{n_{ij}(t)} \right) - \sum_t \sum_i \lambda_{ti} \left(\sum_j p_{ij}(t) - 1 \right), \quad (5)$$

where the range on t is $\{e, f\}$, and where $n_{ij}(e) = \sum_{t=1}^{T/2} n_{ij}(t)$; $n_{ij}(f) = \sum_{t=T/2+1}^T n_{ij}(t)$. When (5) is solved, we have

$$\dot{p}_{ij}(t) = n_{ij}(t) / \sum_k n_{ik}(t), \quad t = e, f.$$

Homogeneity

Our particular hypothesis of homogeneity (in lieu again of a more general option) is very similar to that of stationarity, $H_0 : P(g) = P(h)$, where $P(g)$ is for entities $1, \dots, N/2$, and $P(h)$ is for entities $N/2+1, \dots, N$. We now solve for estimates which can violate the homogeneity assumption:

$$\max_{p_{ij}(s)} \left(\ln \prod_s \prod_{i,j} p_{ij}(s)^{n_{ij}(s)} \right) - \sum_s \sum_i \lambda_{si} \left(\sum_j p_{ij}(s) - 1 \right), \quad (6)$$

where the range on s is $\{g, h\}$. Our estimators are, then,

$$\dot{p}_{ij}(s) = n_{ij}(s) / \sum_k n_{ik}(s), \quad s = g, h.$$

Computation of Test Statistics

Following the likelihood ratio criterion, we note that, for the constant diagonal, specified probability, and zero-order hypotheses,

$$-2 \ln \left[\prod_{i,j} (\dot{p}_{ij}^R / \dot{p}_{ij}^U)^{n_{ij}} \right] \sim \chi_q^2, \quad (7)$$

approximately, where \dot{p}_{ij}^R are the “restricted” probability estimates obtained by imposing the additional restrictions of H_0 — (2), (3), and (4) above; \dot{p}_{ij}^U are the unrestricted estimates obtained from (1) above without the additional restrictions; and q is the number of additional restrictions imposed by H_0 .

For stationarity and homogeneity, respectively,

$$-2 \ln \left[\prod_t \prod_{i,j} (\dot{p}_{ij} / \dot{p}_{ij}(t))^{n_{ij}(t)} \right] \sim \chi_q^2, \quad (8a)$$

and

$$-2 \ln \left[\prod_s \prod_{i,j} (\dot{p}_{ij} / \dot{p}_{ij}(s))^{n_{ij}(s)} \right] \sim \chi_q^2, \quad (8b)$$

approximately, where, in these two cases, the \dot{p}_{ij} are the restricted (stationary or homogeneous) estimates from (1), and $\dot{p}_{ij}(t)$ and $\dot{p}_{ij}(s)$ are the nonstationary and nonhomogeneous estimates from (5) and (6), respectively.

For computational implementation, we do not actually make use of expressions (7), (8a), and (8b), but employ, respectively,

$$\begin{aligned} & -2 \sum_{i,j} n_{ij} \ln(\dot{p}_{ij}^R / \dot{p}_{ij}^U), \\ & -2 \sum_t \sum_{i,j} n_{ij}(t) \ln(\dot{p}_{ij} / \dot{p}_{ij}(t)), \end{aligned}$$

and

$$-2 \sum_s \sum_{i,j} n_{ij}(s) \ln(\dot{p}_{ij} / \dot{p}_{ij}(s)).$$

It is easily shown that when any probability estimate equals zero, so does n_{ij} , $n_{ij}(t)$, or $n_{ij}(s)$. In this way, we can avoid undefined expressions.

3. Macro Data

In this section, we review the restricted least-squares estimation procedure and the general hypothesis testing methodology described in more detail in Kelton and Kelton (1984a, 1984b, 1984c). We make use of the following additional notation:

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

$y_i(t)$ = proportion of entities in state i at time t ;

$\pi_i(t)$ = true unconditional probability of the system's being in state i at time t ;

$\mathbf{y}(t) = [y_1(t), \dots, y_R(t)]$;

$\boldsymbol{\pi}(t) = [\pi_1(t), \dots, \pi_R(t)]$; and

$P = [p_{ij}]_{R \times R}$.

Estimation

The restricted least-squares estimation technique (popular empirically due to the computational

formidability of the macro maximum likelihood estimates), develops from the basic recursion of the Markov chain:

$$\pi(t) = \pi(t-1)P. \quad (9)$$

Since $\pi(t)$ in (9) is not observable, we estimate

$$y(t) = y(t-1)P + \varepsilon(t), \quad t = 1, \dots, T, \quad (10)$$

where $E[y(t)] = \pi(t)$; $\varepsilon(t) = [\varepsilon_1(t), \dots, \varepsilon_R(t)]$; and $E[\varepsilon_i(t)] = 0$. Then, for $1 \leq i \leq R$ and $1 \leq j \leq R-1$ (omitting the last column of P to avoid redundancy), we minimize $\sum_{t=1}^T \sum_{k=1}^{R-1} [\varepsilon_k(t)]^2$ subject to the nonnegativity constraint $p_{ij} \geq 0$ and the row-sum constraint $\sum_{j=1}^{R-1} p_{ij} \leq 1$ — a problem easily solved in standard quadratic programming formulation. We thus obtain least squares estimates \hat{p}_{ij} of the transition probabilities; let $\hat{p}_{iR} = 1 - \sum_{j=1}^{R-1} \hat{p}_{ij}$. We obtain as well the residual sum of squares $SSR = \sum_{t=1}^T \sum_{k=1}^{R-1} [\hat{\varepsilon}_k(t)]^2$ where $\hat{\varepsilon}(t) = y(t) - y(t-1)\hat{P}$, which plays an important role in the hypothesis testing methodology.

Testing

The general testing procedure is to fit two models to the same set of data, one which embeds the restrictions imposed by H_0 , and one which does not. The respective least squares minimization problems are solved. Let SSR_R be the residual sum of squares from the restricted problem; let SSR_U be the sum of squared residuals from the unrestricted problem. We propose (following Chow 1960, Fisher 1970, and Theil 1971) treating

$$F_{q,v} = \frac{(SSR_R - SSR_U)/q}{SSR_U/v}$$

as having an F distribution with (q, v) degrees of freedom, where q is the number of additional restrictions imposed by H_0 , which corresponds to the micro-test degrees of freedom from (7), (8a), and (8b) above, and v refers to the “degrees of freedom” from the unrestricted fit or the number of observations less the number of parameters or transition probabilities to be estimated.

Development of the five specific tests which correspond to the micro-data tests shown above may be found in Kelton and Kelton (1984a, 1984b, 1984c), but we offer one example here — that of constant diagonal probabilities, where $H_0 : p_{ii} = p_{jj} (= p_D)$ for all i and j imposes $q = R - 1$ additional restrictions. As the diagonal elements in this model are not estimated directly, the restricted least squares problem is

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

for $1 \leq j \leq R - 1$, and all t , where $p_D = 1 - \sum_{k=1}^{R-1} p_{Rk}$. We make use of (10) above for the unrestricted fit; and v is $T(R - 1) - R(R - 1)$, the number of observations less the number of transition probabilities estimated in the unrestricted model.

4. Simulation Results

Since the macro-data test statistics proposed in Section 3 need not have the desired F distributions (as (10)'s error term violates the classical least squares assumptions), and since the small-sample properties of the micro-data statistics have not been previously examined, we undertook fairly extensive Monte Carlo simulation studies designed to compare the size and power of the macro and micro tests and to reveal any modeling policies with respect to parameter selection for best test performance. In this section we present our simulation results.

Validity

Of particular interest in hypothesis testing is the probability that χ_q^2 , for micro data, or $F_{q,v}$, for macro data, exceeds a critical value obtained from a standard statistical table; this probability is the actual Type I error incurred in a test with a stated (nominal) significance level. For each of the five null hypotheses developed in Section 2, N independent realizations of the process were simulated for T transitions; from the realizations, either micro data on individual transitions $n_{ij}(t)$ were collected, or macro proportions $y_j(t)$ were obtained; the appropriate restricted and unrestricted models were fit to these data, and an observation on the test statistic computed. These steps were then independently replicated 100 times for the micro tests and 200 times for the macro tests, yielding that many independent observations on the test statistic. As the most direct measure of test robustness, we noted the percentage of the 100 or 200 observations which fell above the upper 10%, 5%, and 1% critical values of the proposed χ_q^2 or $F_{q,v}$ distribution, C_{10} , C_5 , and C_1 . Further, as a measure of test stability, we also tallied the average absolute deviations of these rejection percentages from their target values, $|C_{10} - 10|$, $|C_5 - 5|$, and $|C_1 - 1|$. We used the random number generator of Lewis, Goodman, and Miller (1969).

In conducting our Monte Carlo experiments, we did not set the five parameters of the model — R , N , T , $\pi(0)$, and P — in a purely ad hoc fashion, but rather set them according to a formal, resolution V , 2^{5-1} fractional factorial design constructed by writing a full 2^4 factorial design in the first four factors and letting the level (sign) for P be the positive product of the signs of the levels of

the other four factors (see Box, Hunter, and Hunter 1978). With such a design, we could evaluate the effects of the process factors on the size performance of the tests. Thus, 16 independent sets of 100 or 200 test statistics were generated, yielding 1600 or 3200 independent observations for each null hypothesis. With some appreciation for real-world applications, we let the “-” and “+” levels for R be, respectively, 2 and 4; for T , 25 (or 26 for the stationarity test) and 50; and, for N , 100 and 500. For $\pi(0)$, the “-” level specification was a uniform distribution on the R states, whereas the “+” level put a high probability mass on a particular state. When $R = 2$, the “+” level was (.79,.21), and when $R = 4$, it was (.79,.11,.05,.05), except for the zero-order tests when the “+” level was a distribution degenerate on state 1. Since the five null hypotheses that we consider place different restrictions on the p_{ij} 's, the settings for the levels of P were test-dependent, and are presented in the Appendix, and, for each test, are meant to indicate two different types of entity behavior.

Table 1 presents a summary of validity results for the five tests. All of the estimated Type I error probabilities, averaged across the 16 design points, appear to be fairly close to their target values, implying good overall validity of both testing methodologies; the average rejection percentages over the five micro tests were 9.3, 4.7, and 1.0, while the averages over the macro tests were 10.2, 5.8, and 1.5. Further, for both test groups, the average absolute deviations of the rejection percentages from their respective targets were small, indicating rather stable performance. The macro tests, in fact, show slightly greater stability based on this criterion, with lower average departures from target. Finally, it appears that, for the macro tests, based on our designed experiments, enhanced test validity may be expected if the state space R is kept as small as possible, and there is some evidence that large data records (large N and T) lead to better test performance. The size and stability of the micro tests seem to be generally less sensitive to factor settings. The influences of R , N , and T (factors under direct control of the modeler) were found to be small and inconsistent.

Power

We performed additional designed simulation experiments to investigate the power properties of the tests. The “-” and “+” levels for R , N , T , and $\pi(0)$, as well as the experimental design matrix, were the same as for the validity studies; the values for P for each test are discussed below. Data were generated in violation of H_0 , and 50 (micro) and 100 (macro) replications of the process were made at each design point instead of 100 and 200 since we wanted to consider several alternative

hypotheses.

For the two more specialized tests of constant diagonal elements and a specified probability, we found a simple, intuitive parameterization and developed partial power curves for both micro and macro tests. For the other three tests, we conducted one general design in which the null hypothesis was violated weakly for some design points and strongly for others.

Constant Diagonal Probabilities

For a constant diagonal, transition matrices were chosen such that $p_{11} \neq p_{22} = \dots = p_{RR}$; only p_{11} was allowed to differ from the otherwise constant diagonal. (This should provide a lower bound on power.) Thus, the partial power curves (for the 5% tests) in Figure 1 show power (or the average rejection percentages over the 16 design points) as a function of $d = p_{11} - p_{22}$, constant for all matrices within a given design. The percentages for $d = 0$ correspond to the robustness results from the first and seventh rows in Table 1. As anticipated, the power curves rise as H_0 is violated more severely. However, the micro tests are seen to be much more powerful than the macro tests against slight violations of the null hypothesis. For both sets of tests, and consistent with the validity evidence for the macro tests, a small state space leads to greater power. Further, results from the experimental design suggest that long time records and a large number of entities observed should lead to increased power.

For macro data, an additional design was undertaken which allowed all diagonal elements to differ from each other. The rejection percentages were quite high in this case: 97.25 for a 10% test, 94.13 for a 5% test, and 86.31 for a 1% test.

A Specified Probability

In violating $H_0 : p_{21} = 0.3$, we simply took the actual value of p_{21} to lie over the range $[0,1]$. In other words, the “+” and “-” levels for P were the same as for the validity studies except that row 2 in each case was altered to violate H_0 . Figure 2 shows the 5% partial power curves for the micro and macro tests. Again, power is seen to increase with the extent of departure from H_0 , with the micro test’s being extremely powerful against even very modest departures. Again, the curves pass through the average estimated Type I error probabilities, from the second and eighth rows of Table 1. The policy recommendations are the same as above: small R , large N , and large T seem to have a positive effect on power.

Other Tests

For the zero-order, stationarity, and homogeneity tests, there is not such a natural scalar against which to develop power curves. For these three tests, then, one 16-point design was performed, with the “-” level for P indicating only a small violation of H_0 , and the “+” level a large departure. For first-order dependence, at the “-” level, the rows of P differed only slightly (within 0.1 for each column), whereas, at the “+” level, the difference between p_{ij} and p_{kj} was as much as 0.8. For parameter stationarity and homogeneity, we specified two transition matrices $P(e)$ and $P(f)$ or $P(g)$ and $P(h)$ to generate the realizations of the first and second halves, respectively, of either the T time instants or the N entities. For the “-” level, $P(e)$ and $P(f)$ or $P(g)$ and $P(h)$ were allowed to differ only slightly; $|p_{ij}(e) - p_{ij}(f)| \leq 0.2$ and $|p_{ij}(g) - p_{ij}(h)| \leq 0.2$. For the “+” level, these absolute differences were as much as 0.7.

In Table 2, we give a summary of average rejection percentages, over 16 design points. Although the macro tests are seen to have reasonably high rejection percentages, the micro tests are impressive indeed, with entries of 100% for all three significance levels.

5. Discussion

In this simulation study, we have attempted to assess the degree of micro-data superiority by comparing the size and power performance of micro-data and macro-data tests. Although both the likelihood ratio tests and least squares tests are fairly robust, i.e., have proper Type I error probabilities, and although both are rather powerful against various alternative hypotheses, the micro tests appear to have very steeply rising power curves, “outperforming” the macro tests especially for small degrees of departure from H_0 .

We suggest a sequence of tests, where the basic assumptions of the model would be tested first, followed by tests of a more specialized nature, such as, for example, constant diagonal probabilities or other tests developed from the general frameworks discussed in Sections 2 and 3. Both methodologies are very general and permit any number of specific tests to be developed from them. However, the micro tests are more versatile in allowing, for example, the first-order nature of a process to be tested against either a zero-order alternative or a higher-order alternative. Moreover, the micro-data stationarity test can have a more general character than can its macro counterpart which is based on an arbitrary subperiod division. Finally, it seems that, regardless of testing methodology, a small state space, a long time series, and a large number of entities usually lead to

better performance of the tests.

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.
- 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 and Sons.
- 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.
- FISHER, F.M. (1970), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38, 361-366.
- 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*.
- KELTON, C.M.L. (1981), "Estimation of Time-Independent Markov Processes with Aggregate Data: A Comparison of Techniques," *Econometrica*, 49, 517-518.
- (1984), "Nonstationary Markov Modeling: An Application to Wage-Influenced Industrial Relocation," *International Regional Science Review*, 9, 75-90.
- 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, C.M.L., and KELTON, W.D. (1984a), "Development of Specific Hypothesis Tests for Estimated Markov Chains," The University of Michigan Department of Industrial and Operations Engineering Technical Report 84-6.
- KELTON, C.M.L., and KELTON, W.D. (1984b), "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*, eds. X.J. Avula, et al., Zurich, Switzerland: Pergamon Press, 299-304.
- KELTON, W.D., and KELTON, C.M.L. (1984c), "Hypothesis Tests for Markov Process Models Estimated from Aggregate Frequency Data," *Journal of the American Statistical Association*, 79, 922-928.

- KULLBACK, S., KUPPERMAN, M., and KU, H.H. (1962), "Tests for Contingency Tables and Markov Chains," *Technometrics*, 4, 573-608.
- LAWLESS, J.F., and McLEISH, D.L. (1984), "The Information in Aggregate Data from Markov Chains," *Biometrika*, 71, 419-430.
- 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.
- 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.
- SPILERMAN, S. (1972), "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain," *American Sociological Review*, 37, 277-294.
- 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 and Sons.
- 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
Mean Rejection Percentages under H_0

H_0	C_{10}	C_5	C_1	$ C_{10}-10 $	$ C_5-5 $	$ C_1-1 $
Micro Data						
$p_{ii}=p_{jj}$	10.4	4.9	1.2	2.1	1.4	1.1
$p_{21}=0.3$	10.3	5.9	1.1	2.1	2.0	0.7
$p_{ij}=p_j$	10.8	5.6	1.3	3.1	1.6	0.8
$p_{ij}(e)=p_{ij}(f)$	7.4	3.9	0.7	4.1	2.4	0.8
$p_{ij}(g)=p_{ij}(h)$	7.6	3.4	0.8	3.6	2.2	0.6
Average	9.3	4.7	1.0	3.0	1.9	0.8
Macro Data						
$p_{ii}=p_{jj}$	8.9	4.7	1.2	2.3	1.1	0.7
$p_{21}=0.3$	12.4	7.4	2.1	3.8	2.9	1.4
$p_{ij}=p_j$	7.7	4.2	1.0	2.8	1.3	0.4
$p_{ij}(e)=p_{ij}(f)$	9.9	5.5	1.2	1.9	1.5	0.7
$p_{ij}(g)=p_{ij}(h)$	12.3	7.1	1.8	2.6	2.3	1.0
Average	10.2	5.8	1.5	2.7	1.8	0.8

Table 2

Mean Rejection Percentages Violating H_0

H_0	Level of Test		
	10%	5%	1%
Micro Data			
$p_{ij}=p_j$	100	100	100
$p_{ij}(e)=p_{ij}(f)$	100	100	100
$p_{ij}(g)=p_{ij}(h)$	100	100	100
Macro Data			
$p_{ij}=p_j$	62	55	49
$p_{ij}(e)=p_{ij}(f)$	71	60	43
$p_{ij}(g)=p_{ij}(h)$	77	66	48

Appendix

Levels for P in Factorial Designs for Test Validity

H_0	R=2		R=4	
	"-"	"+"	"-"	"+"
$P_{11}=P_{jj}$	$\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix}$	$\begin{bmatrix} .6 & .4 \\ .4 & .6 \end{bmatrix}$	$\begin{bmatrix} .8 & .2 & .0 & .0 \\ .1 & .8 & .1 & .0 \\ .0 & .1 & .8 & .1 \\ .0 & .0 & .2 & .8 \end{bmatrix}$	$\begin{bmatrix} .6 & .2 & .1 & .1 \\ .2 & .6 & .1 & .1 \\ .1 & .1 & .6 & .2 \\ .1 & .1 & .2 & .6 \end{bmatrix}$
$P_{21}=0.3$	$\begin{bmatrix} .8 & .2 \\ .3 & .7 \end{bmatrix}$	$\begin{bmatrix} .6 & .4 \\ .3 & .7 \end{bmatrix}$	$\begin{bmatrix} .7 & .2 & .1 & .0 \\ .3 & .7 & .0 & .0 \\ .0 & .2 & .7 & .1 \\ .0 & .1 & .2 & .7 \end{bmatrix}$	$\begin{bmatrix} .5 & .2 & .2 & .1 \\ .3 & .6 & .1 & .0 \\ .1 & .1 & .7 & .1 \\ .1 & .2 & .2 & .5 \end{bmatrix}$
$P_{1j}=P_j$	$\begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$	$\begin{bmatrix} .79 & .21 \\ .79 & .21 \end{bmatrix}$	$\begin{bmatrix} .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \end{bmatrix}$	$\begin{bmatrix} .79 & .11 & .05 & .05 \\ .79 & .11 & .05 & .05 \\ .79 & .11 & .05 & .05 \\ .79 & .11 & .05 & .05 \end{bmatrix}$
$P_{1j}(e)=P_{1j}(f);$ $P_{1j}(g)=P_{1j}(h)$	$\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix}$	$\begin{bmatrix} .6 & .4 \\ .5 & .5 \end{bmatrix}$	$\begin{bmatrix} .8 & .2 & .0 & .0 \\ .1 & .8 & .1 & .0 \\ .0 & .1 & .8 & .1 \\ .0 & .0 & .2 & .8 \end{bmatrix}$	$\begin{bmatrix} .5 & .2 & .2 & .1 \\ .2 & .6 & .1 & .1 \\ .1 & .1 & .7 & .1 \\ .1 & .2 & .2 & .5 \end{bmatrix}$

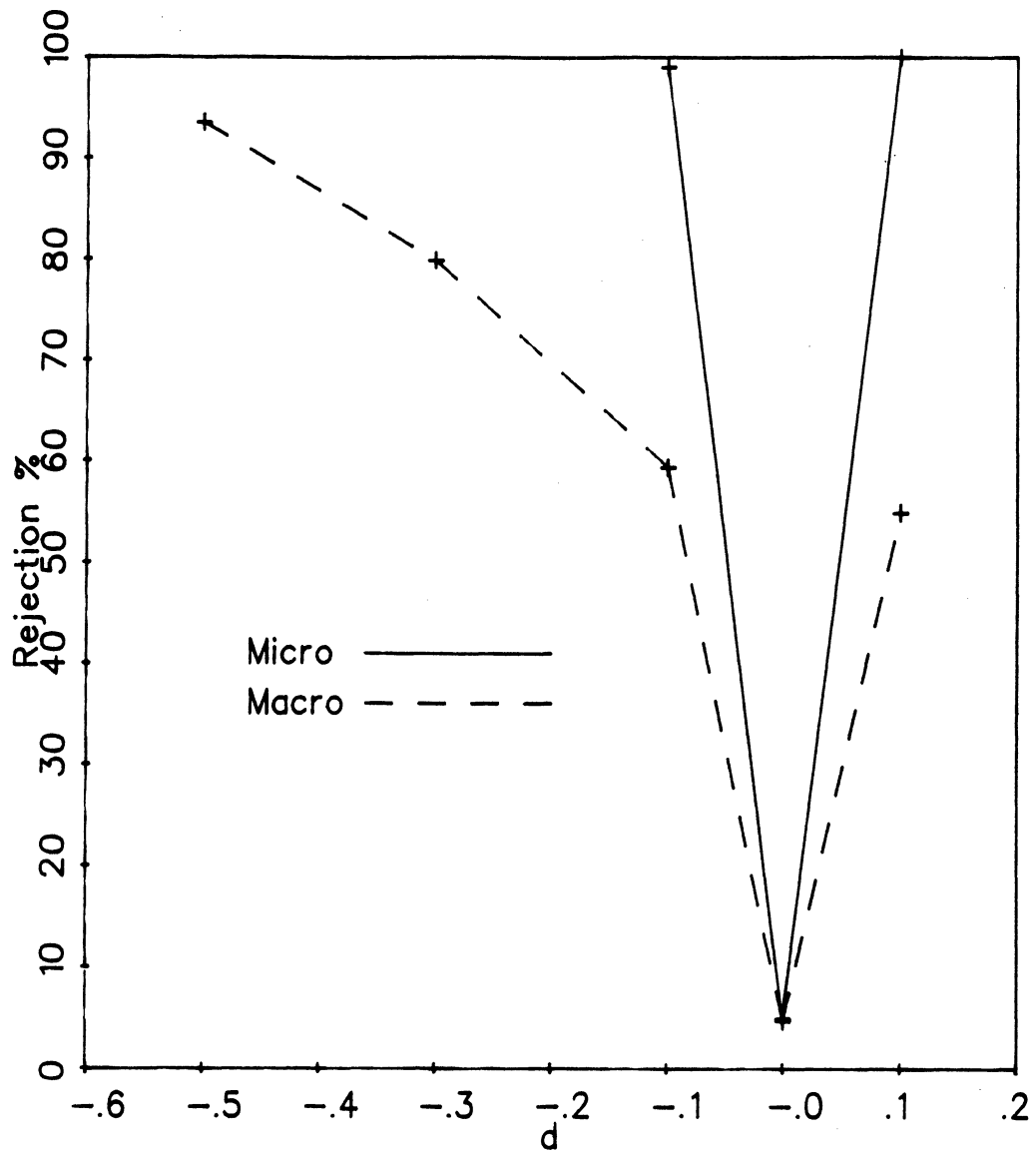


Figure 1. 5% Partial Power Curves for Constant Diagonal Probabilities.
(Means over 16 design points)

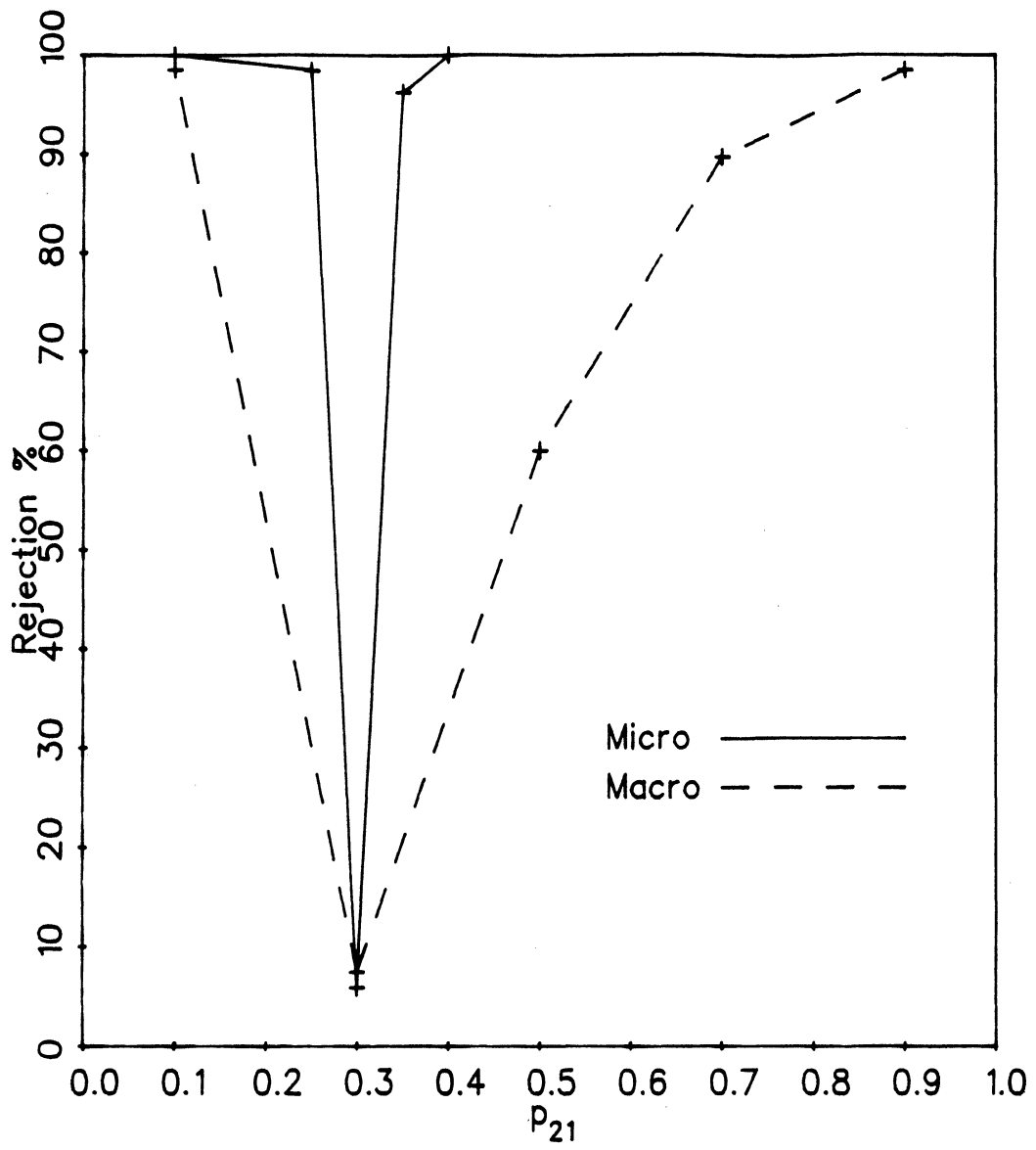


Figure 2. 5% Partial Power Curves for a Specified Probability.
 (Means over 16 design points)