Estimation of Transition Probabilities Using Median Absolute Deviations

C.S. Kim and Glenn Schaible

Abstract. *The probability-constrained minimum absolute deviations (MAD) estimator appears to be superior to the probability-constrained quadratic programming estimator in estimating transition probabilities with limited aggregate time series data. Furthermore, one can reduce the number of columns in the probability-constrained MAD simplex tableau by adopting the median property.*

Keywords. *Minimum absolute deviations, transition probabilities, median absolute deviations, quadratic programming.*

Markov processes are a special class of mathematical models that are often applied to economic decisionmaking in stochastic dynamic programming *(5)*, structural changes of an industry or changes in size economies *(23)*, or international trade *(6)*.[1] To estimate a meaningful transition matrix, researchers need time-ordered data that reflect intertemporal changes of micro units over states (or classifications). However, time-ordered changes of microeconomic units are generally not available for most economic variables; therefore, researchers must often work with aggregate time series data. In an ingenious article, Lee, Judge, and Takayama *(13)* showed how one can estimate transition probabilities for a Markov process reflecting the behavior of micro units with only aggregate time series data. They concluded from a limited trial, based on the assumption of normality of the error terms, that the probability-constrained quadratic programming (QP) estimator is superior to the probability-constrained minimum absolute deviations (MAD) estimator in estimating transition probabilities. In a subsequent article, Lee, Judge, and Zellner concluded from their sampling experiment that the probability-constrained MAD estimator is inferior to the probability-constrained QP estimator (*14*, p. 135).

We prove here that the probability-constrained MAD estimator is superior to the probability-constrained QP estimator when estimating transition probabilities with limited aggregate time series data. Second, we present an alternative model, minimization of median absolute deviations (MOMAD), based on the assumptions that the error terms are nonnormally distributed and that the researcher has *a priori* information about the dynamic nature of the Markov process. Third, we prove that the MOMAD estimator is identical with the probability-constrained MAD estimator, which Bassett and Koenker *(3)* concluded is a more efficient estimator for any error distribution for which the median is superior to the mean as an estimator of location. Moreover, the constraint matrix associated with the MOMAD model involves fewer columns in the simplex tableau.

Notation and Minimization of Absolute Deviations

The stochastic process of a finite Markov Chain can be expressed as

$$\begin{aligned}\Pr(S_{it}, S_{j,t+1}) &= \Pr(S_{it}) \cdot \Pr(S_{j,t+1} \mid S_{it}, S_{i,t-1}, \ldots, S_{i0}) \\ &= \Pr(S_{it}) \cdot \Pr(S_{j,t+1} \mid S_{it}) \qquad (1)\\ &\text{(for all } i \text{ and } j)\end{aligned}$$

where $\Pr(S_{it})$ represents the probability that state S_i occurs on trial t, $\Pr(S_{it}, S_{j,t+1})$ is the joint probability of S_{it} and $S_{j,t+1}$, and $\Pr(S_{j,t+1} \mid S_{it})$ represents the conditional probability for the state S_j. Equation 1, presented by Kemeny and Snell *(12)*, explains that the probability of going to each of the states depends only on the present state and is independent of how we arrived at that state.

Summing both sides of equation 1 over all possible outcomes of the state S_i may be represented by

$$\begin{aligned}\Pr(S_{j,t+1}) &= \sum_{i}^{r} \Pr(S_{it}) \cdot \Pr(S_{j,t+1} \mid S_{it}) \qquad (2)\\ &= \sum_{i}^{r} \Pr(S_{it}) \cdot P_{ij}\end{aligned}$$

Kim and Schaible are economists with the Commodity Economics Division and the Resources and Technology Division, ERS. They appreciate the helpful comments of V.A. Sposito at Iowa State University, T.C. Lee at the University of Connecticut, and G.G. Judge at the University of California Berkeley on an earlier draft of this paper.

[1] Italicized numbers in parentheses refer to items in the References at the end of this article.

where P_{ij} represents the transition probability and has the following properties

$$P_{ij} \geq 0 \quad \text{for all } i \text{ and } j \tag{3}$$

$$\sum_j P_{ij} = 1 \tag{4}$$

By replacing $Pr(S_{j,t+1})$ and $Pr(S_{it})$ with the observed proportions y_{jt} and $x_{i,t-1}$, respectively, we can write equation 2 in the following conventional notation for regression analysis

$$y_{jt} = \sum_{i}^{r} X_{i,t-1} \cdot P_{ij} + \epsilon_{jt} \; (j = 1, 2, \ldots, r) \tag{5}$$

where y_{jt} reflects the observed proportion in state j in time t, $X_{i,t-1}$ is the observed value of the proportion in state i in time t-1, and ϵ represents a random disturbance

In estimating models of the type described in equation 5, researchers have made extensive use of the methods of minimizing the sum of absolute and/or squared errors Although the method of least squares is superior to the MAD procedure if the random events being considered are normally distributed, Bassett and Koenker *(3)* and Hill and Holland *(9)* demonstrate that the MAD estimator is a superior robust method, especially for nonnormal error distributions Bassett and Koenker show that, for any error distribution for which the median is superior to the mean as an estimator of location, the MAD estimator is preferable to the least squares estimator, in the sense of having strictly smaller asymptotic confidence regions Bassett and Koenker note that this condition holds for an enormous class of distributions that either have peaked density at the median or have long tails

The observed proportions for each time period in equation 5 are multinomially distributed, and the multinomial reduces to the binomial when the individual is considered either to be or not to be in state i The binomial probabilities increase monotonically until they reach a maximum value and then decrease monotonically One can show whether or not the binomial is symmetrically distributed by proving that $\alpha_3 = U_3/\sigma^3$ equals zero where U_3 is the third moment about the mean of the binomial distribution For the binomial distribution, with the probability θ of being in state i, the components of α_3 can be derived as $U_3 = n\theta(1-\theta)(1-2\theta)$ and $\sigma^3 = [n\theta(1-\theta)]^{3/2}$, where n is the sample size Therefore, for the binomial distribution, the measure of skewness can be written as

$$\alpha_3 = \frac{U_3}{\sigma^3} = \frac{1 - 2\theta}{[n\theta(1-\theta)]^{1/2}} \tag{6}$$

From equation 6, the binomial is symmetric if $\theta = 1/2$ and/or the sample size n becomes exceedingly large Because aggregate time series data are used to estimate transition probabilities, it is reasonable to assume that the sample size is not large When there are more than two states, so that the probability of the individual being in state i cannot be 0 5 for each state because of constraint 4, the binomial is asymmetrically distributed and the probability-constrained MAD estimator would be superior to the probability-constrained QP estimator.

Consider the problem of estimating an r^2 dimensional vector of unknown parameters P_{ij} from a sample of independently observed proportions for each time period on the random variables $Y_{11}, \ldots, Y_{rT}$ with the following probability distribution

$$Pr[Y_{jt} < y_{jt}] = F(y_{jt} - \sum_{i=1}^{r} X_{i,t-1} \cdot P_{ij}) \tag{7}$$

where $j = 1, 2, \ldots, r$, and $t = 1, 2, \ldots, T$

The probability-constrained MAD estimator $\tilde{P}$ is a solution to the following problem

$$\underset{P \in \mathbf{R}^{r \times r}}{\text{Minimize}} \; [\sum_{j=1}^{r} \sum_{t=1}^{T} | y_{jt} - \sum_{i=1}^{r} X_{i,t-1} \cdot P_{ij} |] \tag{8}$$

Following Barrodale and Young *(2)*, Lee and others *(14)*, Sposito *(20)*, *(21)*, and Spyropoulos and others *(22)*, the probability-constrained MAD estimator is then a solution to the problem

$$\text{Minimize} \sum_{j=1}^{r} \sum_{t=1}^{T} (U_{jt} + V_{jt}) \tag{9}$$

$$\text{subject to} \sum_{j}^{r} P_{ij} = 1.0 \text{ for } i = 1, 2, \ldots, r \tag{10}$$

$$\sum_{i}^{r} X_{i,t-1} \cdot P_{ij} - U_{jt} + V_{jt} = y_{jt} \tag{11}$$

$$\text{for } j = 1, 2, \ldots, r, \; t = 1, 2, \ldots, T$$

$$U_{jt}, V_{jt}, \text{ and } P_{ij} \geq 0 \tag{12}$$

$$\text{for all } i, j, \text{ and } t$$

Minimization of Median Absolute Deviations

Since Hazell *(8)* introduced the minimization of total absolute deviations (MOTAD) model, several economists have identified the MAD criterion as "minimizing the mean absolute deviations" (see *4, 10, 11, 24*). However, the median property has not received sufficient attention among economists. A number of authors have discussed the concept of the median property. Andrews *(1)*, Bassett and Koenker *(3)*, Harvey *(7)*, and Hill and Holland *(9)* showed that the minimum absolute deviations estimator is superior to the least-squares estimator, when the median is superior to the mean as an estimator of location for nonnormal distributions. Furthermore, Spyropoulos and others *(22)* showed that a median property can be used to improve the rate of convergence of linear programming solutions associated with minimum absolute deviations (see *(16)* for the case of nonconvergence). Finally, Parzen *(18)* and Sposito *(21)* show that, for a random variable e, the quantity $\sum_i | e_i - c |$ achieves its minimum value when c is equal to the median.

Following Bassett and Koenker *(3)*, we assume that P_{ij} for all i and j are located so that the probability distribution function F in equation 7 has median zero. Because the median is the point that divides the area under the probability density function, we have the following equality:

$$\Pr\left(\sum_{i}^{r} X_{i,t-1} \bullet P_{ij} > y_{jt}\right) = \Pr\left(\sum_{i}^{r} X_{i,t-1} \bullet P_{ij} < y_{jt}\right) = 1/2 \tag{13}$$

In several situations, researchers have *a priori* knowledge about the dynamic nature of transition probabilities. As energy costs have risen and irrigation water has become more scarce, for example, irrigation technology adopted by farmers has changed from high-pressure, water-intensive systems to low-pressure, energy- and water-efficient systems. Recent irrigation technology shifts in the Southern High Plains have involved a transition from high-pressure center-pivot systems to low-energy precision application (LEPA) systems, whereas Southwest irrigation of tree crops has been shifting from gravity-fed to drip irrigation systems. The proportion of energy- and water-efficient irrigation systems has been increasing, suggesting positive median deviations. As an example suggesting negative deviations over time, we have observed that the number of smokers among professionals has decreased, and that this trend is likely to continue.

In these cases, researchers may be interested in the positive or negative median deviations in equation 13, depending on whether the dynamic nature of transition probabilities moves toward positive or negative deviations. These cases suggest an alternative specification for the probability-constrained MAD model based on minimizing only the sum of the absolute values of the negative median deviations or the sum of the absolute values of the positive median deviations. We can minimize the sum of the absolute values of the negative median deviations by solving the following linear programming model:

Model I

$$\text{Minimize} \quad \sum_{j=1}^{r} \sum_{t=1}^{T} Z_{jt}^{-} \tag{14}$$

$$\text{subject to} \quad \sum_{j}^{r} P_{ij} = 1.0 \text{ for } i = 1, 2, \ldots, r \tag{15}$$

$$\sum_{i}^{r} X_{i,t-1} \bullet P_{ij} + Z_{jt}^{-} \geq y_{jt} \tag{16}$$

$$\text{for } j = 1, 2, \ldots, r,\ t = 1, 2, \ldots, T \text{ and}$$

$$Z_{jt}^{-}, \text{ and } P_{ij} \geq 0 \tag{17}$$

where $\sum_{j=1}^{r} \sum_{t=1}^{T} Z_{jt}^{-}$ is the sum of the absolute values of the negative median deviations.

An alternative model can be specified that minimizes only the sum of the absolute values of the positive median deviations as follows:

Model II

$$\text{Minimize} \quad \sum_{j=1}^{r} \sum_{t=1}^{T} Z_{jt}^{+} \tag{18}$$

$$\text{subject to} \quad \sum_{j}^{r} P_{ij} = 1.0 \text{ for } i = 1, 2, \ldots, r \tag{19}$$

$$\sum_{i}^{r} X_{i,t-1} \bullet P_{ij} - Z_{jt}^{+} \leq y_{jt} \tag{20}$$

$$\text{for } j = 1, 2, \ldots, r,\ t = 1, 2, \ldots, T \text{ and}$$

$$Z_{jt}^{+}, \text{ and } P_{ij} \geq 0 \tag{21}$$

where $\sum_{j=1}^{r} \sum_{t=1}^{T} Z_{jt}^{+}$ is the sum of the absolute values of the positive median deviations

For any error distribution for which the median is superior to the mean as an estimator of location, the MOMAD estimator for both model I and model II is identical with the probability-constrained MAD estimator We can easily prove the identity by first converting equations 9 through 12 into matrix notation as follows

$$\text{Minimize} \quad (U + V)' e_{rT} \tag{9'}$$

$$\text{subject to} \quad GP = e_r \tag{10'}$$

$$XP - U + V = Y \tag{11'}$$

$$P, U, V \geq 0 \tag{12'}$$

where U and V are $(rT \times 1)$ column vectors of surplus and slack variables, respectively, e_{rT} is an $(rT \times 1)$ column vector with all elements 1, X is an $(rT \times r^2)$ block diagonal matrix, P is an $(r^2 \times 1)$ column vector, Y is an $(rT \times 1)$ column vector, and G is an $(r \times r^2)$ coefficient matrix, such that $G=[I_1, I_2, \ , I_r]$ with each I_i an $(r \times r)$ identity matrix Now define variable Z as follows

$$Z = (U + V) \tag{22}$$

where Z is an $(rT \times 1)$ column vector

Rearranging equation 22, we have the equation

$$V = Z - U \tag{23}$$

or equivalently

$$U = Z - V \tag{24}$$

Inserting equations 22 and 23 into equations 9′ and 11′, respectively, the probability-constrained MAD model can be rewritten as follows

$$\text{Minimize} \quad Z' e_{rT} \tag{25}$$

$$\text{subject to} \quad GP = e_r \tag{26}$$

$$XP + Z - 2U = Y \tag{27}$$

$$P, Z, U \geq O \tag{28}$$

or equivalently as

MOMAD Model I

$$\text{Minimize} \quad Z' e_{rT} \tag{29}$$

$$\text{subject to} \quad GP = e_r \tag{30}$$

$$XP + Z \geq Y \tag{31}$$

$$P, Z \geq O \tag{32}$$

which is identical with the MOMAD Model I given in equations 14 through 17, where $Z = Z^-$

In cases where equations 22 and 24 are inserted into equations 9′ and 11′, respectively, the probability-constrained MAD model can be rewritten as follows

$$\text{Minimize} \quad Z' e_{rT} \tag{33}$$

$$\text{subject to} \quad GP = e_r \tag{34}$$

$$XP - Z + 2V = Y \tag{35}$$

$$P, Z, V \geq O \tag{36}$$

or equivalently as

MOMAD Model II

$$\text{Minimize} \quad Z' e_{rT} \tag{37}$$

$$\text{subject to} \quad GP = e_r \tag{38}$$

$$XP - Z \leq Y \tag{39}$$

$$P, Z, \geq O \tag{40}$$

which is identical with the MOMAD model II given in equations 18 through 21, where $Z = Z^+$

Consequently, the probability-constrained MAD estimators are identical with the probability-constrained MOMAD estimators However, the MOMAD procedure reduces rT variables from the probability-constrained MAD procedure to estimate the transition probabilities of the finite Markov Process

Properties of the MOMAD Estimator

Properties of the QP and MAD estimators associated with the probability constraints in equation 10 are unknown Therefore, we restrict our discussion to the QP and MAD estimators without the probability constraints Since the MOMAD estimator is conceptually identical with the probability-constrained MAD estimator when the median is superior to the mean as an estimator of location, we shall concentrate our discussion on the properties of the MAD estimator only

Let m represent the population median For a continuous random variable e, the sample median is asymptotically normal with mean m and variance $[4rTf^2(m)]^{-1}$, where $f(\bullet)$ is the population density function Under the assumption that P_{ij} is located so that the distribution function F in equation 7 has median zero, $\sqrt{rT}(\bar{P} - P)$ converges in distribution to an r^2 dimensional Gaussian random vector with mean zero and covariance matrix $W^2 \bullet Q^{-1}$ *(3)* Here $\bar{P}$ is a vector of the MAD estimator $\tilde{P}_{ij}$, P is a vector of the parameter

P_{ij}, $W^2 = [4f^2(0)]^{-1}$, and $Q = \lim (rT)^{-1} X'_{rT} X_{rT}$ In other words, the MAD estimator is consistent as well as asymptotically Gaussian for a large sample, with a covariance matrix $[W^2 \cdot Q^{-1}]$ Thus, the MAD estimator has strictly smaller asymptotic confidence regions than the QP estimator for linear models from any distribution function F for which the sample median is a more efficient estimator of location than the sample mean

A Numerical Example

To illustrate the MOMAD procedure as well as to demonstrate that the MAD estimator is superior to the QP estimator, we use the numerical example used by Lee, Judge, and Takayama *(13)* In matrix notation form, the transition probabilities to be estimated are as follows

$$P = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} 0\ 6 & 0\ 4 & 0 & 0 \\ 0\ 1 & 0\ 5 & 0\ 4 & 0 \\ 0 & 0\ 1 & 0\ 7 & 0\ 2 \\ 0 & 0 & 0\ 1 & 0\ 9 \end{bmatrix} \end{array} \tag{41}$$

Table 1 shows the synthetic data relating to the sample proportions in each state As Lee, Judge, and Takayama experimented, we assumed that we do not know the transition probability matrix (equation 41), but have only the information contained in the aggregate data in table 1 Under this assumption, we estimate the transition probabilities by the probability-constrained QP, MAD, and MOMAD procedures (tables 2 and 3) Table 2 con-

Table 1—Synthetic data relating to the sample proportions in each state

Time period	Proportion in state (i)			
	S_1	S_2	S_3	S_4
8	0 0815	0 1890	0 3999	0 3296
9	0678	1671	3885	3766
10	0574	1495	3765	4166
11	0494	1354	3650	4502
12	0431	1239	3546	4784
13	0383	1147	3457	5013
14	0345	1072	3380	5203
15	0314	1012	3315	5359
16	0290	0963	3261	5486
17	0270	0924	3216	5590
18	0254	0892	3180	5674

Table 2—The probability-constrained QP, MAD, and MOMAD estimates of the transition matrix from different ascending portions of the aggregate data for a Markov process

Time period	Estimators		
	QP	MAD	MOMAD[1]
t = 8,9, ,18	$\begin{bmatrix} 0\ 753 & 0\ 237 & 0 & 0\ 010 \\ 0 & 624 & 376 & 0 \\ 016 & 071 & 716 & 197 \\ 0 & 004 & 095 & 901 \end{bmatrix}$	$\begin{bmatrix} 0\ 598 & 0\ 402 & 0 & 0 \\ 101 & 508 & 391 & 0 \\ 0 & 094 & 706 & 200 \\ 0 & 002 & 098 & 900 \end{bmatrix}$	$\begin{bmatrix} 0\ 598 & 0\ 402 & 0 & 0 \\ 101 & 508 & 391 & 0 \\ 0 & 094 & 706 & 200 \\ 0 & 002 & 098 & 900 \end{bmatrix}$
t = 9,10, ,18	$\begin{bmatrix} 755 & 245 & 0 & 0 \\ 0 & 595 & 405 & 0 \\ 016 & 086 & 699 & 199 \\ 0 & 0 & 099 & 901 \end{bmatrix}$	$\begin{bmatrix} 597 & 403 & 0 & 0 \\ 101 & 509 & 390 & 0 \\ 0 & 093 & 707 & 200 \\ 0 & 002 & 098 & 900 \end{bmatrix}$	$\begin{bmatrix} 597 & 403 & 0 & 0 \\ 101 & 509 & 390 & 0 \\ 0 & 093 & 707 & 200 \\ 0 & 002 & 098 & 900 \end{bmatrix}$
t = 10,11, ,18	$\begin{bmatrix} 754 & 246 & 0 & 0 \\ 0 & 589 & 411 & 0 \\ 016 & 088 & 698 & 198 \\ 0 & 0 & 099 & 901 \end{bmatrix}$	$\begin{bmatrix} 0\ 608 & 392 & 0 & 0 \\ 097 & 524 & 379 & 0 \\ 0 & 089 & 712 & 199 \\ 0 & 004 & 096 & 900 \end{bmatrix}$	$\begin{bmatrix} 608 & 392 & 0 & 0 \\ 097 & 524 & 379 & 0 \\ 0 & 089 & 712 & 199 \\ 0 & 004 & 096 & 900 \end{bmatrix}$
t = 11,12, ,18	$\begin{bmatrix} 749 & 0 & 251 & 0 \\ 0 & 728 & 272 & 0 \\ 017 & 068 & 718 & 197 \\ 0 & 0 & 098 & 902 \end{bmatrix}$	$\begin{bmatrix} 765 & 235 & 0 & 0 \\ 0 & 596 & 404 & 0 \\ 015 & 086 & 699 & 200 \\ 0 & 0 & 100 & 900 \end{bmatrix}$	$\begin{bmatrix} 765 & 235 & 0 & 0 \\ 0 & 596 & 404 & 0 \\ 015 & 086 & 699 & 200 \\ 0 & 0 & 100 & 900 \end{bmatrix}$
t = 12,13, ,18	$\begin{bmatrix} 856 & 0 & 0 & 144 \\ 0 & 758 & 242 & 0 \\ 0 & 059 & 776 & 165 \\ 004 & 0 & 083 & 913 \end{bmatrix}$	$\begin{bmatrix} 687 & 0 & 313 & 0 \\ 057 & 736 & 207 & 0 \\ 005 & 066 & 729 & 200 \\ 0 & 0 & 100 & 900 \end{bmatrix}$	$\begin{bmatrix} 0\ 687 & 0 & 313 & 0 \\ 057 & 736 & 207 & 0 \\ 005 & 066 & 729 & 200 \\ 0 & 0 & 100 & 900 \end{bmatrix}$
t = 13,14, ,18	$\begin{bmatrix} 824 & 0 & 0 & 0\ 176 \\ 0 & 856 & 0 & 144 \\ 0 & 0 & 923 & 077 \\ 006 & 018 & 038 & 938 \end{bmatrix}$	$\begin{bmatrix} 608 & 0\ 228 & 164 & 0 \\ 097 & 600 & 303 & 0 \\ 0 & 086 & 713 & 201 \\ 0 & 0 & 101 & 899 \end{bmatrix}$	$\begin{bmatrix} 608 & 228 & 164 & 0 \\ 097 & 600 & 303 & 0 \\ 0 & 086 & 713 & 201 \\ 0 & 0 & 101 & 899 \end{bmatrix}$

[1] Estimators for MOMAD models I and II

Table 3—The probability-constrained QP, MAD, and MOMAD estimates of the transition matrix from different descending portions of the aggregate data for a Markov process

Time period	Estimators: QP	MAD	MOMAD[1]
t = 8,9, ,17	$\begin{bmatrix} 0~752 & 0~248 & 0 & 0 \\ 0 & 615 & 371 & 014 \\ 016 & 073 & 720 & 191 \\ 0 & 004 & 093 & 903 \end{bmatrix}$	$\begin{bmatrix} 0~600 & 0~400 & 0 & 0 \\ 100 & 503 & 390 & 007 \\ 0 & 098 & 706 & 196 \\ 0 & 0 & 098 & 902 \end{bmatrix}$	$\begin{bmatrix} 0~599 & 0~401 & 0 & 0 \\ 101 & 504 & 388 & 007 \\ 0 & 097 & 707 & 196 \\ 0 & 0 & 098 & 902 \end{bmatrix}$
t = 8,9, ,16	$\begin{bmatrix} 750 & 250 & 0 & 0 \\ 0 & 612 & 371 & 017 \\ 017 & 075 & 719 & 189 \\ 0 & 003 & 093 & 904 \end{bmatrix}$	$\begin{bmatrix} 613 & 387 & 0 & 0 \\ 091 & 516 & 388 & 005 \\ 002 & 094 & 708 & 196 \\ 0 & 001 & 097 & 902 \end{bmatrix}$	$\begin{bmatrix} 613 & 387 & 0 & 0 \\ 091 & 516 & 388 & 005 \\ 002 & 094 & 708 & 196 \\ 0 & 001 & 097 & 902 \end{bmatrix}$
t = 8,9, ,15	$\begin{bmatrix} 749 & 228 & 0 & 023 \\ 0 & 631 & 369 & 0 \\ 017 & 070 & 721 & 192 \\ 0 & 004 & 093 & 903 \end{bmatrix}$	$\begin{bmatrix} 599 & 401 & 0 & 0 \\ 101 & 503 & 388 & 008 \\ 0 & 098 & 707 & 195 \\ 0 & 001 & 097 & 902 \end{bmatrix}$	$\begin{bmatrix} 598 & 402 & 0 & 0 \\ 101 & 502 & 389 & 008 \\ 0 & 098 & 707 & 195 \\ 0 & 0 & 098 & 902 \end{bmatrix}$
t = 8,9, ,14	$\begin{bmatrix} 746 & 216 & 0 & 038 \\ 0 & 631 & 369 & 0 \\ 017 & 074 & 721 & 188 \\ 0 & 002 & 093 & 905 \end{bmatrix}$	$\begin{bmatrix} 699 & 301 & 0 & 0 \\ 032 & 574 & 389 & 005 \\ 012 & 085 & 707 & 196 \\ 0 & 001 & 098 & 901 \end{bmatrix}$	$\begin{bmatrix} 699 & 301 & 0 & 0 \\ 032 & 574 & 389 & 005 \\ 012 & 085 & 707 & 196 \\ 0 & 001 & 098 & 901 \end{bmatrix}$
t = 8,9, ,13	$\begin{bmatrix} 772 & 228 & 0 & 0 \\ 0 & 632 & 368 & 0 \\ 008 & 069 & 722 & 201 \\ 005 & 004 & 092 & 899 \end{bmatrix}$	$\begin{bmatrix} 756 & 243 & 001 & 0 \\ 0 & 608 & 385 & 007 \\ 014 & 080 & 710 & 196 \\ 001 & 0 & 097 & 902 \end{bmatrix}$	$\begin{bmatrix} 756 & 243 & 001 & 0 \\ 0 & 608 & 385 & 007 \\ 014 & 080 & 710 & 196 \\ 001 & 0 & 097 & 902 \end{bmatrix}$
t = 8,9, ,12	$\begin{bmatrix} 767 & 0 & 233 & 0 \\ 0 & 784 & 216 & 0 \\ 014 & 047 & 735 & 204 \\ 0 & 0 & 104 & 896 \end{bmatrix}$	$\begin{bmatrix} 598 & 396 & 006 & 0 \\ 101 & 505 & 380 & 014 \\ 0 & 099 & 711 & 190 \\ 0 & 0 & 096 & 904 \end{bmatrix}$	$\begin{bmatrix} 598 & 396 & 006 & 0 \\ 101 & 505 & 380 & 014 \\ 0 & 099 & 711 & 190 \\ 0 & 0 & 096 & 904 \end{bmatrix}$

[1] Estimators for MOMAD models I and II

tains the estimators of the transition matrix from different ascending portions of the aggregate data, while table 3 used different descending portions of the aggregate data The probability-constrained QP estimator, using the trials (t = 8, 9, , 18) in table 2, differs from that presented by Lee, Judge, and Takayama *(13)* Similarly, the probability-constrained QP estimator for the trials (t = 8, 9, , 12) in table 3 differs from that presented by Lee, Judge, and Zellner *(14)* These authors used a simplex algorithm developed by Wolfe *(25)*, whereas we used Minos, developed by Murtagh and Saunders *(17)*, which uses the reduced-gradient algorithm, also developed by Wolfe *(26)*

Tables 2 and 3 show that the probability-constrained MAD and MOMAD estimators are identical Furthermore, the MOMAD estimator is more efficient than the probability-constrained QP estimator However, the efficiency between these two estimators needs further study

Comparison of the QP and MOMAD Estimators

The sample median is asymptotically normal with mean (m) and variance $[4nf^2(m)]^{-1}$, where (m) is the population median and f(•) is the population probability density function Because the probability density function f(m) is unknown, there are no meaningful statistical test procedures based on the sample median Therefore, a nonparametric statistical method (the binomial test) is used to check the significance of the differences in the dispersion of the estimators about the true parameters *(14)*

The null hypothesis to be tested is as follows

$$H_0 \cdot \Pr[\,|\bar{P}_{ij} - P_{ij}| > |\hat{P}_{ij} - P_{ij}|\,] = 1/2$$

relative to the alternative

$$H_A \; \Pr[\,|\bar{P}_{ij} - P_{ij}| > |\hat{P}_{ij} - P_{ij}|\,] > 1/2$$

where $\bar{P}_{ij}$, $\hat{P}_{ij}$, and P_{ij} are the probability-constrained QP estimator, the MOMAD estimator, and the true parameter, respectively

The procedures of the binomial test and its statistical table can be found in Siegel *(19)* We applied the test using only those pairs in which there is no tie (see *15*) The results of the binomial test show that the MOMAD estimator is at least as efficient as, or more efficient than, the probability-constrained QP estimator in estimating the transition probabilities (table 4)

Table 4—The binomial tests for H_0
$\Pr[\,|\bar{P}_{ij} - P_{ij}| > |\hat{P}_{ij} - P_{ij}|\,] = 1/2$
vs
$H_A \; \Pr[\,|\bar{P}_{ij} - P_{ij}| > |\hat{P}_{ij} - P_{ij}|\,] > 1/2$

Time period	Probabilities associated with values in the binomial test	Superior estimator based on the binomial test at α = 0 05
t = 8, 9, , 18	0	MOMAD
t = 9, 10, , 18	194	QP and MOMAD
t = 10, 11, , 18	275	QP and MOMAD
t = 11, 12, , 18	006	MOMAD
t = 12, 13, , 18	046	MOMAD
t = 13, 14, , 18	0	MOMAD
t = 8, 9, , 17	0	MOMAD
t = 8, 9, , 16	0	MOMAD
t = 8, 9, , 15	001	MOMAD
t = 8, 9, , 14	001	MOMAD
t = 8, 9, , 13	212	QP and MOMAD
t = 8, 9, , 12	033	MOMAD

Conclusions

We have proposed the use of the minimization of median absolute deviations (MOMAD) to estimate transition probabilities of a finite Markov chain with limited aggregate time series data The MOMAD model is conceptually identical with the MAD model However, the MOMAD model is simpler to use than the probability-constrained MAD procedure, while using a linear programming algorithm We also showed that the MOMAD estimators are more efficient than the QP estimators by demonstrating that (1) the MOMAD and MAD models are conceptually identical, and (2) the MAD estimators and, therefore, the MOMAD estimators are more efficient than the QP estimators

References

1 Andrews, David F "A Robust Method for Multiple Linear Regression," *Technometrics*, Vol 16, 1974, pp 523-31

2 Barrodale, I , and A Young "Algorithms for Best L_1 and L_∞ Linear Approximations on a Discrete Set," *Numerical Mathematics*, Vol 8, 1966, pp 295-306

3 Bassett, Gilbert, Jr , and R Koenker "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association*, Vol 73, 1978, pp 618-22

4 Buccola, S T "Minimizing Mean Absolute Deviations to Exactly Solve Expected Utility Problems Comment," *American Journal of Agricultural Economics*, Vol 64, 1982, pp 789-91

5 Burt, O R "Economics of Conjunctive Use of Ground and Surface Water," *Hilgardia*, Vol 36, 1964, pp 31-111

6 Dent, W T "Application of Markov Analysis to International Wool Flows," *Review of Economics and Statistics*, Vol 49, 1967, pp 613-16

7 Harvey, A C "A Comparison of Preliminary Estimators for Robust Regression," *Journal of the American Statistical Association*, Vol 72, 1977, pp 910-13

8 Hazell, P B R "A Linear Alternative to Quadratic and Semivariance Programming for Farm Planning Under Uncertainty," *American Journal of Agricultural Economics*, Vol 53, 1971 pp 53-62

9 Hill, R , and P W Holland 'Two Robust Alternatives to Least-Squares Regression," *Journal of the American Statistical Association*, Vol 72, 1977, pp 828-33

10 Johnson, D , and M Boehlje "Minimizing Mean Absolute Deviations to Exactly Solve Expected Utility Problems," *American Journal of Agricultural Economics*, Vol 63, 1981, pp 728-29

11 __________ "Minimizing Mean Absolute Deviations to Exactly Solve Expected Utility Problems Reply," *American Journal of Agricultural Economics*, Vol 64, 1982, pp 792-93

12 Kemeny, J G , and J L Snell *Finite Markov Chains* Princeton, NJ D Van Nostrand Company, Inc , 1960

13 Lee, T C , G G Judge, and T Takayama. "On Estimating the Transition Probabilities of a Markov Process," *Journal of Farm Economics*, Vol 47, 1965, pp 742–62

14 Lee, T C , G G Judge, and A Zellner *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data* 2nd rev ed New York North-Holland Publishing Co , 1977

15 Lindgren, B W *Statistical Theory* 2nd ed New York Macmillan Company, 1968

16 McCormick, G F , and V A. Sposito "A Note on L_1 Estimation Based on the Median Positive Quotient," *Royal Statistical Society*, Series C, Vol 24, 1975, pp 347–350

17 Murtagh, B A , and M Saunders *Minos User's Guide* Tech Rept SOL 83–20 Systems Optimization Laboratory, Dept of Operations Research, Stanford Univ , 1983

18 Parzen, E *Modern Probability Theory and its Application* New York John Wiley and Sons, 1960

19 Siegel, S *Nonparametric Statistics for the Behavioral Sciences*, New York McGraw-Hill, 1956

20 Sposito, V A "On Unbiased L_p Regression Estimators," *Journal of the American Statistical Association*, Vol 77, 1982, pp 652–53

21 ———— *Linear and Nonlinear Programming* Ames· Iowa State Univ Press, 1975

22 Spyropoulos, K , E Kiountouzis, and A Young "Discrete Approximation in the L_1 Norm," *The Computer Journal*, Vol 16, 1973, pp 180–86

23 Telser, L G "Advertising and Cigarettes," *Journal of Political Economy*, Vol 70, 1962, pp 471–99

24 Thomson, K.J , and P B R Hazell "Reliability of Using the Mean Absolute Deviation to Derive Efficient E, V Farm Plans," *American Journal of Agricultural Economics*, Vol 54, 1972, pp 503–06

25 Wolfe, P "The Simplex Method for Quadratic Programming," *Econometrica*, Vol. 27, 1959, pp 382–98

26 ———— "Methods of Nonlinear Programming," in *Recent Advances in Mathematical Programming*, ed R L Graves and P Wolfe New York McGraw-Hill, 1963