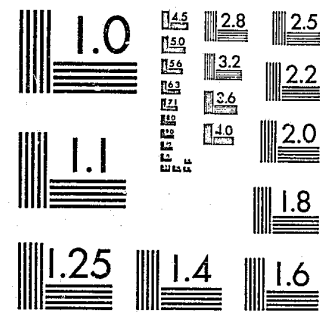


National Criminal Justice Reference Service



This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D. C. 20531

7/25/83

FINAL REPORT

LEAA Grant # 79-NI-AX-0058

Inferring Individual Behavior From
Grouped Data

John Wanat

86079-
86081

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by

Public Domain/LEAA
U.S. Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

FINAL REPORT

LEAA Grant # 79-NI-AX-0058

Inferring Individual Behavior From Grouped Data

John Wanat
Department of Political Science
University of Illinois at Chicago Circle
Box 4348
Chicago, Illinois 60680

INDEX

I. Introduction to the Research Approach	I-1
II. Applications	
A) Estimating Juvenile Recidivism by Cross-Level Inference 86080	II-A-1
B) Estimating Cell Entries in Contingency Tables: Distributional Assumptions vs. Marginal Constraints 86081	II-B-1
C) Estimating Individual Level Change in Public Opinion from Aggregate Data	II-C-1
D) Estimating the Degree of Mobilization and Conversion in the 1890's: The Nature of Political Change in One Critical Election (with Karen Burke)	II-D-1
E) Evaluation With Sparse Nominal Data: The Case of Differential Compliance with the 55 mph Limit	II-E-1
III. Illustrative FORTRAN Program Embodying the Research Technique Used in the Research	III-1

ACQUISITIONS

This research project attacked the classic problem of inferring cell entries in contingency tables from marginals. The traditional approaches to cell estimation take two forms. One traditional approach used the marginals to set upper and lower limits that each cell could assume. Since no cell can be larger than the smaller of its row or column marginal, extrema are reasonably easy to identify. Subtracting the maxima of all cells save the one of interest from the marginals in either the row or column generates the minimal value of the cell. One problem with this approach is that multiple constraints may be operative and extreme values of cells may be theoretically reached under the basically unidimensional constraints just indicated, but may not be attainable because constraints on other cells may influence the extrema of the cell of interest. A second problem is that this approach may yield very wide ranges of the values a cell may take.

The other traditional approach to cross-level inference assumes that the cell entries are distributed according to a given and known parametric distribution. Assuming that the distribution of cell entries follows a multinomial probability distribution allows the investigator to employ a maximum likelihood criterion to pick out that one combination of cell values which satisfies the marginal constraints and maximizes likelihood. Although this traditional approach yields point estimates, that precision is purchased at the expense of assuming the researcher knows the distributional form underlying the cells.

Yet another problem with the traditional approaches is their inability to incorporate inequality relations among cells into the estimation. the

class of problems this research focuses on allows the researchers to postulate that certain cells are related in specific ways to other cells. Traditional approaches are inapplicable to those problems.

In particular this research centered on problems in which a given population is partitioned into three subgroups at two points in time and the aggregate size of those subgroups is known at both times but the movement of the elements is not known. We know, for example, how many drivers exceeded 70 mph, how many were under 55 mph, and how many drove between 55 and 70 in the years prior to and just after the imposition of the 55 mph limit. We wish to estimate how many in each of those speed categories shifted to other categories. We know the marginals and can postulate that fewer 1973 drivers increased speed than decreased or held steady in speed. How can this situation yield estimates of precision without making distributional assumptions?

The approach employed in the research is based on the premise that to know what did happen in the quasi-experimental design expressible as a 3x3 table, the investigator must know what could have happened. My research extends the Davis and Duncan approach of finding extreme cell values from 2x2 tables to 3x3 tables and gets a handle on multi-operative constraints by using a program that enumerates all possible solutions and examines each to identify the smallest and largest values for each cell.

The enumeration of all possible solutions also allows bypassing the assumption of a distribution underlying the cell entries. By tallying all cell values an 'empirically' derived probability distribution underlying the cell that may incorporate inequality relations is mapped. The research-

cher is not assuming any parametric form but is allowing the constraints to set up limits and so to define the possibility of each cell taking on any given value. These distributions can then be used to choose a "best" solution from among the myriad solutions.

The tallying of all solutions goes beyond the advantages just listed because it allows exploring the relations among cells not related by assumption. Probabilistic statements about cells' relationships are thus possible. By such tallying the investigator can make statements of the form, "In 98.7% of all solutions, cell A exceeds cell B," which gives important ordinal information along with a measure of confidence in the statement.

In the next five sections the development of the techniques suggested above is illustrated. In each paper some aspect of the estimation approach is developed and applied to a substantive problem. The first paper (Section II-A) attempts to estimate the extent of recidivism among juveniles from aggregate data. In that paper the 'empirically' derived possibility distributions were used as probabilities and the one solution that maximized the joint probability was chosen as the best estimate. While that generated a nice estimate, it was based on the assumption of independence of the cell probabilities, which is not always a reasonable assumption.

The second paper (Section II-B) surmounts the independence assumption by choosing a solution by a least squares criterion. The best solution in this paper is the one that minimizes the 'distance' from it to all possible other solutions. That paper also shows that this approach yields estimates that approximate maximum likelihood estimates but without the usual distributional assumptions required by the ML approach.

The third paper (Section II-C) applies the basic technique to the problem of estimating changes in public opinion among various groups from only aggregate data collected at two time points. One again the basic strategy of enumerating all solutions is employed. Point estimates of the cells are derived using the least squares criterion on the frequencies empirically identified as possible. This paper, however, modifies the technique by establishing confidence intervals about the point estimates. It also examines the ordinal relations among selected cells. Examining the number of solutions in which various cells exceed others, it appears, uncovers powerful ordinal relationships.

In the fourth paper (Section II-D) the mapping of solutions is applied to archival voting data from the 1890's. The variant of the technique applied here is to avoid point estimates but to examine the ordinal relations among cells and cell groups. The program used in all this research was modified to compare each cell in every solution to every other one and to record the proportion of time any cell exceeded any other. This kind of information gives the researcher a measure of confidence in the ordinal relations among cells. In a pleasing number of comparisons, all solutions had certain cells exceeding others all the time so that certainty is obtained about some cell relationships.

The last paper tried to assess differential compliance with the 55 mph speed limit imposed in 1974 by tallying all solutions to the 3x3 cross-tabulation representing 1973 and 1974 aggregate data on the distribution of passenger car speeds. In this paper the ordinal relations among cells in solutions were sought as were the limits each cell could assume. This

paper made heavy use of cell groupings in its comparisons because frequently collections of cells are of interest. The limits on selected cell groups and the ordinal relations among the cell groups were also identified and analyzed.

The five papers demonstrate the refinement of the basic notion that examining all possible solutions to 3x3 tables representing basically the same population at two points in time will yield information not automatically evident but which can be important. The use of computers allows quick enumeration and analysis so that what would have been impossible is now easy. The last section of this report consists of the FORTRAN program used in the speeding paper. That program incorporates virtually all the refinements used in the previous four papers. It is presented for interactive use at a terminal.

86080

ESTIMATING JUVENILE RECIDIVISM BY CROSS-LEVEL INFERENCE

John Wanat

with the assistance of Karen Burke

Department of Political Science
University of Illinois - Chicago Circle
Box 4348, Chicago, Illinois 60680

A Paper Prepared for Presentation at the Special National Workshop
on Research Methodology and Criminal Justice Program Evaluation,
March 17-19, 1980, Baltimore, Maryland.

ESTIMATING JUVENILE RECIDIVISM BY CROSS-LEVEL INFERENCE

II-A-1

John Wanat, with the assistance of Karen Burke
Department of Political Science
University of Illinois - Chicago Circle

This paper applies a non-analytic technique currently under development to the problem of estimating juvenile recidivism from aggregate data. Where only aggregate level data on juvenile crime for two successive years are available no precise estimate of recidivism is possible with present techniques. Yet by using only those data, usually ignored relations among the variables, and examining the map of solutions to the 3x3 table representing the recidivism situation it is possible, this paper argues, to identify a solution that is "most possible."

The best or "most possible" solution is estimated by charting the frequency distributions of solutions to each cell entry in the 3x3 table. That information constitutes a probability density function for each cell. By making the perhaps cavalier but intuitively justifiable assumption of independence, the joint probability of each solution is computed and the solution with the highest joint probability is nominated the most possible estimate.

Sensitivity testing and a Monte Carlo approach allow assessing the confidence the investigator should put on the estimate. While this approach lacks elegant intellectual theory, it is intuitively pleasing and is superior to alternatives for estimating recidivism in terms of time and cost.

II-A-2

Problems in Estimating Juvenile Recidivism

Participants in the criminal justice system and scholars studying the system occasionally need to estimate the extent of recidivism among juvenile delinquents. Decisions on the effectiveness of police, counseling, custodial, and court policy as well as assessment of the impact of various forces on the criminal justice system often require some estimate of a repeater population.

No matter what definition of recidivism is used (multiple apprehensions, arrests, court action, custody, incarceration), it is often difficult to gauge the extent of recidivism. Estimating recidivism among juveniles is complicated by legal constraints, the expense and shortcomings of cohort and survey research, and time required for most estimating studies. The legal constraints are the most obvious. Because society feels that it would be unfair to stigmatize persons throughout their lives for actions committed before maturity, arrest and court records are usually sealed for juveniles. By promising confidentiality, bona fide researchers can occasionally gain access to sealed records. However, even when this is allowed, the records are not usually organized to facilitate research. Where access to individual records is denied, researchers are left with aggregate data indicating how many were apprehended, charged, convicted, sentenced, etc. In many cases breakdowns by offense, age, sex, race and other characteristics may be available. But problems with present methods of making cross-level inferences from such grouped data to individual behavior make such data less than useful.

The most common approach to estimating recidivism is through a cohort study. Wolfgang's work (1972) is an excellent example of this approach, but it also illustrates the shortcomings. For one, it is expensive to identify and track large numbers of subjects. Second, the practitioner is not generally interested in the cohort but in a series of cohorts since the original cohort may be unique. Third, cohort studies do not concern usually themselves with migration questions, though some researchers in this area are sensitive to this problem (Shaw, 1929). Juveniles move temporally by aging and becoming subject to adults laws, and juveniles move spatially. Cohort studies can examine the aging but are poor in assessing immigration to the jurisdiction under study. Many cohort studies also take a long time perspective while practitioners in the criminal justice system are interested in a shorter time perspective, both for completion of a study and for recidivism.

To cover a jurisdiction surveys can be used to measure various attributes of the target population. Unfortunately it takes a fair degree of sophistication to elicit sensitive information, such as that about illegal activity. Once again, it takes time to collect and analyze such data.

The most commonly available data are the monthly, weekly, or yearly reports of the police and courts. Data in those reports are aggregated, however, and so are not very useful although their potential is strong. Since the records are aggregate confidentiality is preserved. Since the reports are periodic short term changes in the population might be assessed. And since the records are collected for managerial and accountability reasons anyway, their use for research would be virtually costless.

In what follows aggregate data from public records will be used to estimate juvenile recidivism.

Problem Formulation and Illustration

To illustrate how aggregate public data can be utilized to estimate recidivism probabilities by a technique being developed by the author, consider the problem of estimating the rate of recidivism in arrests among juveniles in Chicago for all crimes from 1970 to 1971. In 1970 juvenile male arrests numbered 46,583. In the next year the count was 57,727. The problem is to estimate how many of the 46,583 were among the 53,727.

To make our recidivism estimate, the total population of eligible male juveniles must be first appraised. The 1970 census yields 423,576 males aged 11 through 17. Since the literature indicates few arrests for those under 11, we used that as the lower cutoff. Age 18 marks majority in Illinois and so 17 is the higher cutoff. The 423,576 constitute the 1970 base.

To make the 1971 base we must add the 66,319 who were 10 year olds in 1970 to the 1970 base and subtract the 56,493 17 year olds in 1970 who became 18 year olds in 1971. This exercise accounts for the age curve variation from one year to the next.

Migration into and out of Chicago must now be taken into account. The 1970 census asked about residence in 1965 as well as 1970. By using responses to those questions it is possible to estimate in-and out-migration to the Chicago metropolitan area at 1.46% and 1.99% per year, respectively. Although we are interested only in Chicago we use

those figures for illustrative purposes, although they should be reasonably accurate. Obviously, other data sources, such as school records and utility company records, could be used to assist in refining migration data.

Ideally our problem could be case into a 2x2 contingency table with arrests and non-arrests in 1970 on the rows and with arrests and non-arrests in 1971 on the columns. However there is slippage in the population from one year to the next. Ten year olds are relevant in 1971 and not in 1970, while 17 year olds are relevant in 1970 but not in 1971. And migration into and out of Chicago must be taken into account. Therefore a pool or residual category must be created. We must consider all those who were 11 to 17 in 1970 as well as all those who became 11 to 17 in 1971 by either aging or migration. To handle the pool or residual category for both years, the problem must be cast into a 3x3 table. See Figure 1.

The first Figure indicates the arithmetic and assumptions used to derive the marginals for the 3x3 table whose cell entries we will estimate. In terms of that table, $N_1/46,583$ is the recidivism rate from 1970 to 1971; $N_4/376,993$ is the rate of entry to criminal arrest.

Certain constraints are clear. Cell entries must sum to marginals, Hence,

$$N_1 + N_2 + N_3 = M_{11} \quad (1)$$

$$N_4 + N_5 + N_6 = M_{12} \quad (2)$$

$$N_7 + N_8 + N_9 = M_{13} \quad (3)$$

$$N_1 + N_4 + N_7 = M_{21} \quad (4)$$

$$N_2 + N_5 + N_8 = M_{22} \quad (5)$$

$$N_3 + N_6 + N_9 = M_{23} \quad (6)$$

Figure 1

Set-Up of Recidivism Problem

		1971			
1970		Arrests	Non Arrests	Pool	
Arrests	N_1	N_2	N_3	46,583	M_{11}
Non Arrests	N_4	N_5	N_6	376,993	M_{12}
Pool	N_7	N_8	N_9	73,472	M_{13}
		53,729	377,430	65,880	$N = 497,048$
		M_{21}	M_{22}	M_{23}	

Total Population:

1970 population	423,576
"new" 10 year olds	66,319
1.46% inmigration	<u>7,153</u>
Total	497,048

$$1970 \text{ Pool} = \text{Total} - 1970 \text{ population} = 497,048 - 423,513 = 73,472$$

1971 Non-Pool:

1970 Population	423,576
+ 10 year olds	66,319
- 17 year olds	-56,493
1.46% inmigration	6,184
1.99% out migration	<u>-8,429</u>
	431,159

$$1970 \text{ Pool} = \text{Total} - 1971 \text{ Non Pool} = 497,048 - 431,159 = 65,889$$

For simplicity's sake it is assumed that juveniles are arrested no more than once a year.

We now have six equations and nine unknowns, an algebraically insoluble problem. The infinite number of solutions to equations (1) through (6) can be reduced to a finite though large number of solutions by converting the marginal data into integer percentages. If we further require that all N_i be non-negative integers, we have a more tractable problem.

The problem can be made even more tractable by using ancillary information to reject certain solutions. Use of ancillary information gives the researcher leverage with which he can go beyond standard techniques. For example, because the 56,493 seventeen year olds in 1970 must be put into the pool category for 1971, we know that, in terms of percentages,

$$N_3 + N_6 > 56493/497048 \quad (7)$$

Similarly, because 66,319 ten year olds in 1970 move from the pool category, it is clear that in terms of percentages,

$$N_7 + N_8 > 66,319/497,048 \quad (8)$$

Furthermore, previous research, folklore, and common sense can be incorporated into restrictions on possible solutions. It is reasonable to expect that most male juveniles who have not been arrested until now, will still be arrest free a year from now, or

$$N_5 > .5(373,993/497,048) \quad (9)$$

Similarly, although we do not know the recidivism rate it might be reasonable to assume it does not exceed 50%, or

$$N_1 < .5(46,583/497,048) \quad (10)$$

Yet other constraints can be imposed. Since both those arrested and those not arrested age, it is reasonable to require that in 1970 those leaving the population of interest should be composed of delinquents and non-delinquents in roughly the same proportion as the 1970 marginals. A comparable requirement for those entering the population of interest should apply. Hence we require that:

$$\frac{M_{11}}{M_{11} + M_{12}} - .1 < \frac{N_3}{N_3 + N_6} < \frac{M_{11}}{M_{11} + M_{12}} + .1 \quad (11)$$

$$\frac{M_{21}}{M_{21} + M_{22}} - .1 < \frac{N_7}{N_7 + N_8} < \frac{M_{21}}{M_{21} + M_{22}} + .1 \quad (12)$$

Even with equations (1) - (6) and inequalities (7) - (12), there will be a number, possibly a large number, of solutions. In order to see what the solutions look like, a computer program was written that identifies all non-negative integer solutions to conditions (1) - (12). Table 1 displays the frequencies with which each cell entry appears. Note that some cell entries occur far more frequently than others. For N_5 , for instance, a value of 57 or 58 is more frequent a solution than any other. Intuitively then we would want one solution that had $N_5 = 57$ more than one in which $N_5 = 55$. The solution we choose should maximize the frequency of cell entries.

To be more formal, if $(N_1', N_2', N_3', \dots, N_9')$ is a solution satisfying conditions (1) through (12), and if N_1' occurs F_1 times, if N_2' occurs F_2 times, etc, then define $W = \prod_{i=1}^N F_i$. We define the most possible solution as the one that maximizes W . To do this our program first searches out all solutions and maps their cell-wise frequencies. Then it re-examines all solutions, and using the frequencies, computes the W

for each. Finally it chooses the solution(s) that maximizes W. That we define as the best estimate, the most possible solution.

Recidivism Problem Estimates

Table 1 displays the first pass cell-wise frequency distributions of solutions to our problem and the optimal solutions. Even though some of the distributions are peaked, a number are uniform. Hence it is not surprising that as many as 12 solutions generate the same high value of W, although this rarely happens. Now the problem lies in choosing from among the 12 optimal solutions. The approach is the same as before. Frequency distributions of the values of each cell among the 12 solutions are compiled and, using that information, the W statistic for each solution is compiled. This second pass generates two "best" solutions, which, expressed as 9-tuples, are (2, 6, 1, 7, 57, 12, 2, 13, 0), (2, 5, 2, 7, 58, 11, 2, 13, 0). In both cases, $N_1 = 2$, which gives us recidivism rate of 2/9.

The value of 2/9 is rather coarse, which is to be expected since the problem we have attacked is the original problem cast into integer percentage terms. To gain a more refined estimate the marginals are expressed in thousandths instead of percentages or hundredths. To reduce computer time the values searched are restricted to those around the two optimal solutions listed above. Because our problem has four degrees of freedom, I vary N_1 , N_3 , N_5 , and N_9 to cover all possible solutions. Going from hundredths to thousandths, bracketing the best solutions by ten thousandths, and incrementing by thousandths means N_1 ranges from 10 to 30; N_3 runs from 1 to 30; N_5 from 560 to 590; and N_9 run from 0 to 10. The 1970 marginals, in thousandths, for

Table 1

First Pass Solutions and Frequencies

FREQUENCIES OF SOLUTIONS

VALUE	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9
0	6	0	0	0	0	0	0	0	30
1	6	0	15	0	0	0	10	0	0
2	6	0	15	0	0	0	10	0	0
3	6	3	0	0	0	0	10	0	0
4	6	6	0	0	0	0	0	0	0
5	0	6	0	4	0	0	0	0	0
6	0	6	0	6	0	0	0	0	0
7	0	6	0	6	0	0	0	0	0
8	0	3	0	6	0	0	0	0	0
9	0	0	0	4	0	0	0	0	0
10	0	0	0	2	0	0	0	0	0
11	0	0	0	0	0	15	0	0	0
12	0	0	0	0	0	15	0	10	0
13	0	0	0	0	0	0	0	10	0
14	0	0	0	0	0	0	0	10	0
54	0	0	0	0	0	0	0	10	0
55	0	0	0	0	1	0	0	0	0
56	0	0	0	0	3	0	0	0	0
57	0	0	0	0	5	0	0	0	0
58	0	0	0	0	6	0	0	0	0
59	0	0	0	0	6	0	0	0	0
60	0	0	0	0	5	0	0	0	0
61	0	0	0	0	3	0	0	0	0
	0	0	0	0	1	0	0	0	0

* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
OPTIMAL ESTIMATES									
Solution #	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9
1	0	7	2	8	57	11	3	12	0
2	1	7	1	7	57	12	3	12	0
3	1	6	2	8	57	11	2	13	0
4	1	6	2	7	58	11	3	12	0
5	2	6	1	7	57	12	2	13	0
6	2	5	2	8	57	11	1	14	0
7	2	6	1	6	58	12	3	12	0
8	2	5	2	7	58	11	2	13	0
9	3	5	1	7	57	12	1	14	0
10	3	5	1	6	58	12	2	13	0
11	3	4	2	7	58	11	1	14	0
12	4	4	1	6	58	12	1	14	0

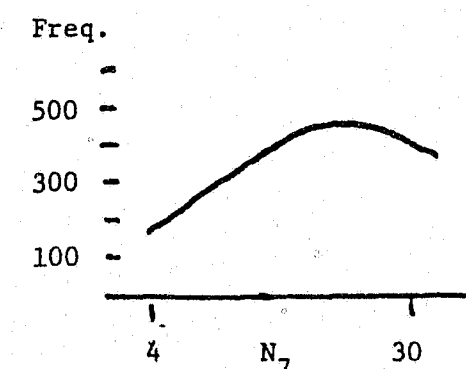
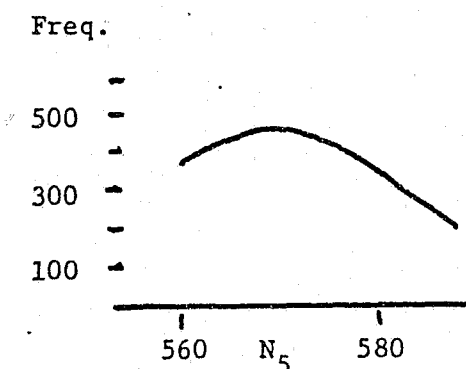
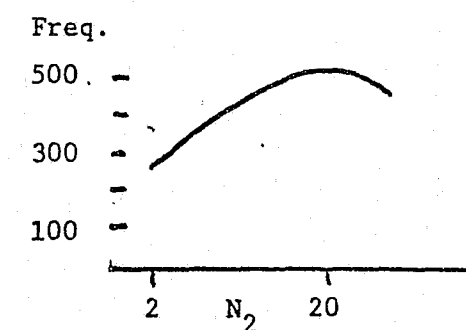
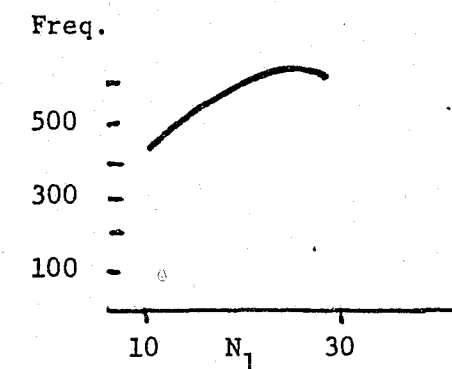
the arrested, non-delinquent, and pool categories are 94, 758, and 148. Comparable marginals for 1971 are 108, 759, and 133. The frequency distributions of cell solutions from the more refined problem formulation is found in Figure 2.

Note that the distributions of solutions in Figure 2 are not uniform. Moreover they tend to be asymmetric. And they are numerous. Also note that only one "best" or most possible solution is found: (20, 57, 17, 67, 575, 116, 21, 127, 0). That solution yields the transition matrix:

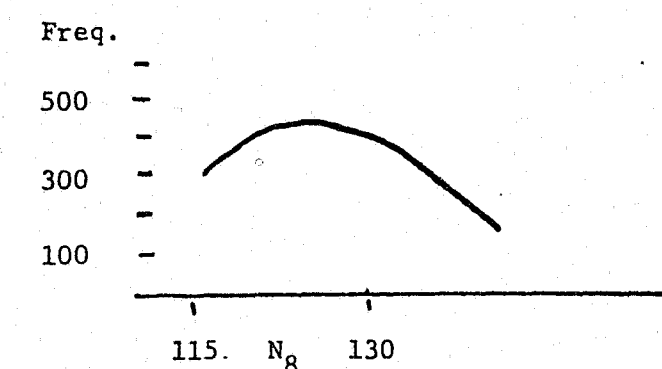
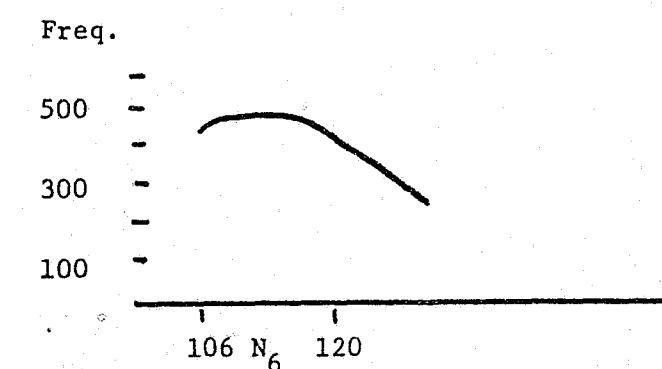
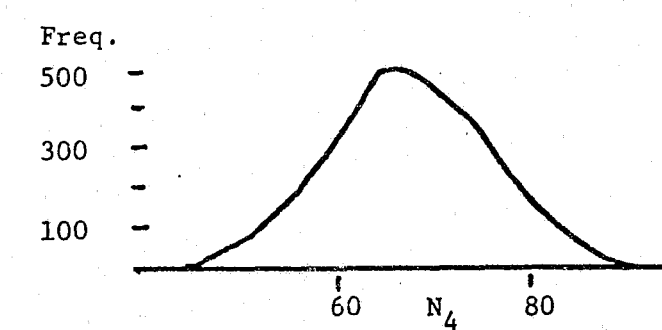
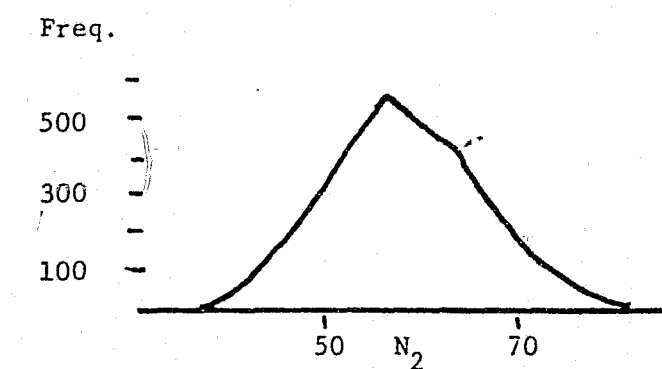
	1971		
1970	Arrest	No Arrest	Pool
Arrest	.213	.606	.181
No Arrest	.088	.759	.153
Pool	.142	.858	0

The problem just analyzed by the most possible approach was presented to illustrate the technique. Since the technique is still under development it is not possible to ascribe specific characteristics to the estimates. Yet it is worth mentioning that a small Monte Carlo experiment that used the Most Possible Estimate (MPE) approach to estimate cell entries from marginals with no inequalities constraints yielded very good results. The MPE estimates approximated the Maximum Likelihood Estimate very closely, considering that the MPE is restricted to integers and the usual MLE is not. A further Monte Carlo experiment where inequality constraints were imposed, yielded pleasingly accurate estimates. All 197 tables in that experiment had a total N of 100; in 75% of the time a confidence interval of 3 about the estimate contained the true value.

Cell Frequencies for Solutions on Third Pass
(N_i 's all expressed in thousandths)



For all 11,062 solutions, $N_9 = 0$



The accuracy of the estimates in general at this stage of development will be a function of how closely the constraints represent the phenomenon under analysis and how binding those constraints are. The advance this approach makes over other techniques lies in its ability to incorporate the relations, usually inequality relations, among the variables. There is literature that attempts to estimate cell entries from marginals. But the approaches in the literature require more than the sparse data MPE uses (see Davis and Duncan, 1953; Robinson, 1950; Goodman, 1959; Lee, Judge, and Zellner, 1970; Shively, 1974). The use of the inequality relations gives leverage to the analyst the other approaches lack.

By varying the inequality relations and observing whether the estimates are affected the analyst can determine the influence of particular assumptions. Sensitivity analysis of this type is particularly easy since the costs of analysis are reasonably cheap. In a similar vein if the assumptions are settled, random data conforming to those assumptions can be repeatedly generated, MPE analysis can be undertaken on those data, and confidence intervals about the original problem's estimates can be drawn. Confidence bands thus can assist the analyst in evaluating the utility of the approach in a particular instance.

Justification of Choice Criterion

Because of its central position in the technique being described, some discussion of the choice criterion is in order. There is obvious value in mapping the solution space for our problem because this defines the universe of possibilities, just as counting the faces on a die or the

number of hearts in a partial deck is important in assessing the probabilities in a dice or card game.

In a classical probability sense one assumes that each solution in the solution space is equi-probable. But even if the n-tuples representing solutions are equally probable (or if this convention merely represents our ignorance of the likelihood of each solution), their characteristics, i.e., their elements, are not. For illustrative purposes consider the list of first pass "best" solutions to our problem. See Table 1. There are twelve solutions we wish to choose from, each presumably as likely as any other. But note that for N_1 some values are more common than others. While the meso-level, classical approach suggests equi-probability, it would be foolish to reject the ancillary information that $N_1 = 2$ occurs more often than any other value for N_1 .

The operational problem lies in incorporating this ancillary information. If we use conditional probabilities, then everything "cancels" out and each solution has probability of 1/12. That approach throws away information. We do not follow that avenue but proceed in a straight forward if slightly cavalier fashion. By acting as if the cell entries were independent of each other the probability of each solution is simply the product of the frequencies of each cell entry's value. This uses the ancillary information in what appears to be an effective manner.

This approach also avoids some complications arising from other approaches by explicitly incorporating the inequality relations among cells and marginals without bias. One alternative approach would be

parametric but that is not reasonable inasmuch as there is no easy way to presume the distributional model. A second and, on the face of it, more reasonable approach would seek the solution with the greatest entropy. But that approach is biased toward the "flatter" distributions. It does that because, as an extension of the Fisher exact test, it presumes no interaction (independence by rows), a condition explicitly inappropriate given the inequality relations characterizing our problem.

One other problem with the entropy maximization and some parametric approaches lies in the assumption of distinguishable elements. That is unrealistic for two reasons. First, we are not dealing with individual behavior but with groups of individuals. Our focus is on meso, not micro level phenomena. Secondly, unlike dealing with red and black balls in urns which can truly be chosen by chance and are undistinguishable, we are enumerating humans who have acted. Either the juvenile delinquent committed a second offense or not. If the meso-level solution says that 2 of 9 are recidivists, it does not make sense to ask how many ways nine can be taken two at a time because two repeated and they can be taken from the nine only one way. Thus the meso-level data are all we can rely upon. Only at the meso-level can we preserve the determinate action of the individual level and also incorporate the effects of the constraints.

Consequently, because of no strong justification for the alternatives and because of the desire to incorporate the ancillary information, we proceed as if the cell-wise frequency distributions are independent. Ongoing research will help assess the validity of this approach.

SELECTED REFERENCES

- Duncan, O. Dudley and Beverly Davis (1953), "An Alternative to Ecological Correlation," 18 American Sociological Review, 665-666.
- Goodman, Leo A., (1959), "Some Alternatives to Ecological Correlation," 64 American Journal of Sociology, 610-625.
- Lee, T.C., G.G. Judge, and A. Zellner, (1970), Estimating the Parameters of the Markov Probability Model From Aggregate Time Series Data, Amsterdam and London: North-Holland Publishing Company.
- Robinson, W.S., (1950), "Ecological Correlations and the Behavior of Individuals," 15 American Sociological Review, 351-357.
- Shaw, Clifford R., Zorbaugh, Frederick M., McKay, Henry D., and Cottrell, Leonard S., (1929), Delinquency Areas, Chicago: University of Chicago Press.
- Shively, W. Phillips, (1975), "Utilizing External Evidence in Cross-Level Inference," 1 Political Methodology, 61-73.
- Wolfgang, Marvin E., Figlio, Robert M., Sellin, Thorsten, (1972), Delinquency in a Birth Cohort, Chicago: University of Chicago Press.

parametric but that is not reasonable inasmuch as there is no easy way to presume the distributional model. A second and, on the face of it, more reasonable approach would seek the solution with the greatest entropy. But that approach is biased toward the "flatter" distributions. It does that because, as an extension of the Fisher exact test, it presumes no interaction (independence by rows), a condition explicitly inappropriate given the inequality relations characterizing our problem.

One other problem with the entropy maximization and some parametric approaches lies in the assumption of distinguishable elements. That is unrealistic for two reasons. First, we are not dealing with individual behavior but with groups of individuals. Our focus is on meso, not micro level phenomena. Secondly, unlike dealing with red and black balls in urns which can truly be chosen by chance and are undistinguishable, we are enumerating humans who have acted. Either the juvenile delinquent committed a second offense or not. If the meso-level solution says that 2 of 9 are recidivists, it does not make sense to ask how many ways nine can be taken two at a time because two repeated and they can be taken from the nine only one way. Thus the meso-level data are all we can rely upon. Only at the meso-level can we preserve the determinate action of the individual level and also incorporate the effects of the constraints.

Consequently, because of no strong justification for the alternatives and because of the desire to incorporate the ancillary information, we proceed as if the cell-wise frequency distributions are independent. Ongoing research will help assess the validity of this approach.

SELECTED REFERENCES

- Duncan, O. Dudley and Beverly Davis (1953), "An Alternative to Ecological Correlation," 18 American Sociological Review, 665-666.
- Goodman, Leo A., (1959), "Some Alternatives to Ecological Correlation," 64 American Journal of Sociology, 610-625.
- Lee, T.C., G.G. Judge, and A. Zellner, (1970), Estimating the Parameters of the Markov Probability Model From Aggregate Time Series Data, Amsterdam and London: North-Holland Publishing Company.
- Robinson, W.S., (1950), "Ecological Correlations and the Behavior of Individuals," 15 American Sociological Review, 351-357.
- Shaw, Clifford R., Zorbaugh, Frederick M., McKay, Henry D., and Cottrell, Leonard S., (1929), Delinquency Areas, Chicago: University of Chicago Press.
- Shively, W. Phillips, (1975), "Utilizing External Evidence in Cross-Level Inference," 1 Political Methodology, 61-73.
- Wolfgang, Marvin E., Figlio, Robert M., Sellin, Thorsten, (1972), Delinquency in a Birth Cohort, Chicago: University of Chicago Press.

86081

ESTIMATING CELL ENTRIES IN CONTINGENCY TABLES:
DISTRIBUTIONAL ASSUMPTIONS VS. MARGINAL CONSTRAINTS

John Wanat
Department of Political Science
University of Illinois at Chicago Circle
Box 4348
Chicago, Illinois 60680

THIS RESEARCH WAS SUPPORTED BY THE LAW ENFORCEMENT ASSISTANCE
ADMINISTRATION UNDER LEAA GRANT # 79-NI-AX-0058.

II-B-1

ABSTRACT

ESTIMATING CELL ENTRIES IN CONTINGENCY TABLES:
DISTRIBUTIONAL ASSUMPTIONS VS. MARGINAL CONSTRAINTS

Testing for independence in two-way contingency tables commonly involves using a chi-square test that measures the goodness of fit between the observed data and the expected values. This paper examines the expected values usually employed in such analyses. It first notes that the usually derived expected values are based on assumptions about the distributional form and indistinguishability of the counted items, assumptions that may not always be reasonable. Then the paper presents an alternative method of estimating expected values, the most possible estimate approach, that does not rely on a priori assumptions about the distributional form or distinguishability of items being counted. Finally, by establishing that the most possible estimates approximate maximum likelihood estimates, the robustness of the standard ML estimator is demonstrated and the researcher is thereby assured of safety in using it without attending to distributional assumptions. This analysis applies to 3x3 tables.

ESTIMATING CELL ENTRIES IN CONTINGENCY TABLES:
DISTRIBUTIONAL ASSUMPTIONS VS. MARGINAL CONSTRAINTS

INTRODUCTION

Testing for independence in two way contingency tables commonly involves using a chi-square test that measures the goodness of fit between the observed data and the expected values. This paper examines the expected values as usually employed in such analyses. It first notes that the expected values usually derived are based on assumptions about the distributional form and indistinguishability of the counted items, assumptions that may not always be reasonable. Then the paper presents an alternative method of estimating expected values, the most possible estimate approach, that does not rely on a priori assumptions about the distributional forms. Finally, by establishing that the most possible estimates approximates maximum likelihood estimates, the robustness of the ML estimator is demonstrated. This analysis is restricted to 3x3 tables.

STANDARD ESTIMATION TECHNIQUES

The standard algorithm for estimating the expected value for a cell entry in a two way table is RC/N , where R is the row marginal and C is the column marginal for the cell being estimated and N is the total table count. This algorithm is derived by assuming the cell entries are distributed according to a given probability distribution function and by finding the cell value that maximizes the function. Regardless of whether one assumes the underlying PDF to be Poisson, multinomial, or product multinomial, the expected cell value will be RC/N (Bishop, Fienberg, and Holland, 1975, Chapter 3).

Unfortunately, it is not always possible to assume the form of the PDF underlying the cells. For example, the multinomial distribution so commonly employed assumes indistinguishable counted items. In many social systems it is clearly not appropriate to assume that all persons under scrutiny are interchangeable. And in particle physics, in fact, the Maxwell-Boltzmann approach assumed distinguishable particles and has been shown to be inapplicable to particle phenomena. The use of an a priori model here was obviously incorrect (Feller, p. 41).

An alternative approach to finding expected cell entries which makes fewer assumptions about the PDF is therefore desirable.

Estimating Cell Entries....

THE MOST POSSIBLE ESTIMATE APPROACH

The logic behind the most possible estimate (MPE) approach is simply to make exhaustive use of the constraints expressed in the row and column marginals. It is suggested that these marginal constraints are so overwhelming that they generally make assumptions about cell entry distributions unnecessary. It will be demonstrated that estimates based on the marginally imposed constraints closely approximate the estimates made by the more profligate use of assumptions.

Let $n(i,j)$ represent the cell in the i -th row and the j -th column of the 3×3 table with row marginals $n(i,.)$ and column marginals $n(.,j)$. Hence there are six equality constraints on any set of cell entries or solutions:

$$\sum_{i=1}^3 n(i,j) = n(.,j) \text{ for } j=1,2,3 \quad (1)$$

$$\sum_{j=1}^3 n(i,j) = n(i,.) \text{ for } i=1,2,3 \quad (2)$$

Estimating Cell Entries....

We seek a set $(n'(1,1), n'(1,2), n'(1,3), n'(2,1), n'(2,2), n'(2,3), n'(3,1), n'(3,2), n'(3,3))$ that satisfies equations (1) and (2) "better" than all other sets of $n(i,j)$. Because the $n(i,j)$ must be non-negative integers for most substantive interpretations, as well as for ease of computation and exposition, there are only a finite, though conceivably very large, set of $n(i,j)$'s that satisfy the marginal constraints. Hence, it is logical to explore just what lies in the realm of possibility. We want to know what the set of solutions sets looks like. Mapping the solutions space is therefore in order.

Consider the contingency table at the top of Display 1. To explore the solution space for that table a computer program was written to identify all possible, non-negative integer solutions and to indicate the frequencies with which each cell takes on values. The Display, for example, notes that cell $n(1,1)$ is 0 in 79 of the 170 solutions, is 1 in 59 solutions, and takes a values of 2 in the remaining 32 solutions.

The argument now moves from the level of the solutions sets to cell entries. It can be argued that since there were 170 distinct solutions, each is equally likely and no reasonable inferences can be made about a best or most likely solution set. Inspection of the frequency distributions in the Display, however, makes it clear that some

Estimating Cell Entries....

values of the cells are more common than others. It would be foolish, therefore, to ignore the information that, for example, $n(1,3)=6$ occurs in more than three times as many solutions as $n(1,3)=2$. The unequal frequency distributions of possible cell entries is consequently considered as evidence that even if the solutions are equally likely, the cell entries are not. We use that information to choose one solution that is better than the rest.

The frequency distributions of the cells constitute an envelope of possible distributions of the cell solutions. No matter what probability distribution may govern $n(1,1)$, data in the Display say that the probability of $n(1,1)=0$ is less than or equal to $79/170$; that the probability of $n(1,1)=1$ is less than or equal to $59/170$; and that the probability of $n(1,1)=2$ is less than or equal to $32/170$. Data in Display 1 show that without any a priori assumptions about the nature of the PDF governing solutions to the table, about the parameters of the PDF's, or about the interchangeability of counted items, upper limits on the likelihoods, probabilities, or perhaps most accurately, the possibilities of cell entries can be derived.

The envelope distribution for each cell can be used to generate an optimal estimate. If n' represents an optimal solution for a given cell, we seek to minimize the deviation represented as the sum of the squared differences between

Estimating Cell Entries....

the optimal cell value and each possible cell value found in the search for all solutions. In essence we seek the least squares estimate given the data found in the map of all possible cell solutions. By differentiating the sum of squared deviations and setting it equal to zero, it becomes clear that the best estimate is the expected value or the mean.

$$I = \sum_T (n_t - n')^2 \quad (3)$$

$$dD/dn = -2 \sum_T (n_t - n') = 0 \quad (4)$$

$$n' = (1/T) \sum_T n_t \quad (5)$$

Or, in terms of grouped data, where $f(n_t)$ is the frequency of n_t ,

$$n' = \sum_T f(n_t) n_t \quad (6)$$

In other words, the expected value of the cell, given the envelope distribution, is our optimal value.

Estimating Cell Entries....

We next show that the expected values of the cells satisfy all the row and marginal constraints. Consider the first row.

$$n(1,1) + n(1,2) + n(1,3) = n(1,-) \quad (7)$$

Taking the sum of the cell values for the first row for each solution,

$$\sum_T n(1,1) + \sum_T n(1,2) + \sum_T n(1,3) = T n(1,-) \quad (8)$$

where T = number of solutions

$$(1/T) \sum_T n(1,1) + (1/T) \sum_T n(1,2) + (1/T) \sum_T n(1,3) = n(1,-) \quad (9)$$

Equation (9) states that the expected values, our optimal estimates, for the first row sum to the first row marginal. The same holds true for all rows and columns. The estimates therefore observe the marginal constraints.

MCST POSSIBLE ESTIMATES AND MAXIMUM LIKELIHOOD

Under the assumption that cell entries are distributed according to the multinomial, the Poisson, or the product multinomial distribution and there is independence between row and column categories, the maximum likelihood estimate of a cell is

Estimating Cell Entries....

$$n'(i,j) = n(i,-) n(-,j) / n(-,-) \quad (10)$$

To assess the relationship between the maximum likelihood estimate (MLE) and the most possible estimate (MPE), the marginals for 3x3 tables were generated, both the MLE and MPE estimates were calculated, and the estimates were compared. Three sets of one tables each were generated. One hundred tables whose total count ranged from 50 to 100, then one hundred tables whose count ranged 100 to 200, and finally one hundred whose count ranged from 200 to 300 were created.

To make sure that the ML estimates would be appropriately made, tables in the first set were randomly generated by choosing a uniformly generated random number between 50 and 100 to be the total table count, T. Next, uniformly drawn random numbers between 0 and 1, designated $r(1)$, $r(2)$, and $r(3)$, were used to generate the row marginals defined as

$$n(1,-) = T r(1) / (r(1) + r(2) + r(3))$$

$$n(2,-) = T r(2) / (r(1) + r(2) + r(3))$$

$$n(3,-) = T - n(1,-) - n(2,-)$$

Then the order of the row marginals were randomly interchanged. The same process was employed to generate the column marginals. Finally, the row and column marginals were used in equation (10) to find the ML estimates. If any

Estimating Cell Entries....

of the estimates were less than 5, the table was discarded because small expected cell estimates cannot be accurately derived with the standard MLE algorithm. The entire process was repeated until 100 tables were generated. Then the process was repeated for the other two ranges of total table count.

Marginals from the tables were employed to make both ML and MP estimates. They were compared according to two criteria: numerical closeness and statistical likelihood. First, the root mean square error was estimated for all cells. For the 900 cells tallied from the 100 tables whose count ranged from 50 to 100, the RMSE was 0.543. The mean absolute percentage error was 3.5%. Thus the "average" error was about one half of a cell count between the maximum likelihood and the most possible estimate. For the 100 to 200 count tables the error measures were 8.31% mean absolute % error and 1.82 RMSE. Comparable figures for the 200 to 300 count tables were 10.51% and 3.22 RMSE. These measures indicate a close agreement between the MLE and the MPE.

Secondly, to see if the two approaches generate estimates whose differences could have occurred by chance, a standard chi square test was employed. The ML estimates were taken as the expected values and the MP estimates were considered the observed. For each of the tables, then, the chi square statistic was computed to assess the chances that

Estimating Cell Entries....

the tables were statistically indistinguishable. As Table 1 indicates, the likelihood that there are significant differences between the two sets of estimates is very small in the 50 to 100 count tables. In 93% of the cases, the differences are "significant" at the 95% level. In other word, in 93 cases out of 100 the probability that the two estimated tables were not distinguishable was 95%. And in no case was the probability that the estimates were different less than 80%.

In the tests where total counts were larger, the data are not as immediately impressive, though they are convincing. Out of 100 chi-square goodness of fit tests where there is no relation, one expects, on the average, that 10 would be significant at the 10% level, 30 at the 30% level, 60 at the 60% level, and so on. It is clear in the center column of the table that the 100 through 200 count tables do better than expected. The entries in the column are always less than the significance levels, which indicates that the differences between the MP and ML estimates are not even as large as chance would allow. In the set of largest tables, those in the last column, the same degree is apparent for the overwhelming majority of tests. Only the first four entries in that column are over expectations, but not overwhelmingly so.

Estimating Cell Entries....

Overall, it is clear that the MP estimates approximate the MLE. The closeness is especially good when dealing with small count tables. When the count is large, asymptotic properties of most distributions assure the researcher that distributional assumptions are not important. When the table count is in the middle range, this paper indicates the assumptions are not important.

The MP approach may be particularly useful in analysis of tables where expected cell entries by the usual algorithm are less than 5. The standard algorithm is not appropriate then, but the MPE technique has no constraints on the expected values. Hence it could be used to derive the expected values when they might be small and so prevent the usual expedient of collapsing cells to bring expected cell values up over 5.

The data thus suggest that MP estimates approximate ML estimates and do so without assuming the nature of an underlying distribution. The close equivalence between the MLE and MPE possibly offers an explanation of why the same maximum likelihood algorithm for expected cell values applies for at least three PDF's. It is conjectured that the marginal constraints so dominate the functional forms that the standard algorithm reflects more of the marginal constraints than of the functional form assumptions.

Estimating Cell Entries....

Perhaps most importantly this paper shows that researchers can employ the standard algorithm for the expected values in chi-square tests without worrying about assumptions of how the data are distributed.

Estimating Cell Entries....

DISPLAY 1

Frequency of Cell Solutions to Table

n(1,1)	n(1,2)	n(1,3)	9
n(2,1)	n(2,2)	n(2,3)	6
n(3,1)	n(3,2)	n(3,3)	5
2	7	11	

Frequency Distributions

Value	n(1,1)	n(1,2)	n(1,3)	n(2,1)	n(2,2)	n(2,3)	n(3,1)	n(3,2)	n(3,3)
0	79	22	1	86	32	16	89	38	22
1	59	28	4	57	32	22	56	38	28
2	32	31	10	27	32	28	25	35	34
3	0	29	16	0	29	31	0	30	37
4	0	24	22	0	24	32	0	20	31
5	0	18	28	0	15	26	0	9	18
6	0	12	31	0	6	15	0	0	0
7	0	6	29	0	0	0	0	0	0
8	0	0	20	0	0	0	0	0	0
9	0	0	9	0	0	0	0	0	0

Estimating Cell Entries....

TABLE 1

Differences Between ML and MP Estimates

Sig. Level	Cumulative # of Cases		
	N=50 to 100	N=100 to 200	N=200 to 300
.025%	0	1	1
1	0	2	8
5	0	3	11
10	0	4	13
20	0	4	17
30	0	7	25
40	0	10	30
50	0	16	42
60	0	21	52
70	0	30	58
80	3	45	70
90	7	52	77
95	14	63	84
97.5	28	70	90
99	39	81	95
99.5%	100	100	100
Cell RMSE	.54	1.82	3.22
Mean Absolute % Error/Cell	3.5%	8.3%	10.5%

Estimating Cell Entries....

BIBLIOGRAPHY

Bishop, Yvonne M., Stephen E. Fienberg, and Paul W. Holland,
DISCRETE MULTIVARIATE ANALYSIS: THEORY AND PRACTICE,
Cambridge, Massachusetts: The MIT Press, 1975.

Feller, William, AN INTRODUCTION TO PROBABILITY AND ITS
APPLICATIONS, 3rd Edition, Volume 1, NY: Wiley, 1968.

ESTIMATING INDIVIDUAL LEVEL CHANGE IN
PUBLIC OPINION FROM AGGREGATE DATA

John Wanat
Department of Political Science
University of Illinois at Chicago Circle
Box 4348
Chicago, Illinois 60680

This paper was presented at the 1980 Annual Meeting of the Midwest Association for Public Opinion Research, December 5, 1980, Chicago, Illinois. I am grateful to Lee Dutter, Doris Graber, and Gerald Strom for their comments on an earlier draft. Support for this research came in part from the Law Enforcement Assistance Administration under Grant 79-NI-AX-0058. IEAA bears no responsibility for the judgments and analysis in this paper.

ESTIMATING INDIVIDUAL LEVEL CHANGE IN
PUBLIC OPINION FROM AGGREGATE DATA

This paper estimates shifts in public opinion on presidential approval surrounding certain salient international events. It estimates the proportion of approvers, disapprovers, and neutral respondents in Gallup polls taken before the events who shift to other responses in polls after the events. This is done by analyzing all possible non-negative, integer solutions to the 3x3 tables representing the pre- to post-event change. By means of a most possible criterion to make a point estimate and by examination of the relations among cell values in solutions it is shown that most of the increase in support accruing to the president comes from those who had disapproved of his performance before and not from those who had no opinion.

ESTIMATING INDIVIDUAL LEVEL CHANGE IN PUBLIC
OPINION FROM AGGREGATE DATA

John Wanat

Department of Political Science
University of Illinois at Chicago Circle

INTRODUCTION

The presidential support literature documents at least two major patterns. First, the proportion of those surveyed responding favorably to the Gallup question about how well the president is doing his job generally declines as the president's term wears on (Mueller, 1970, 1971, 1973; Stimson, 1976). Second, in times of certain presidential action on the international scene, the public will register a decided increase in approving how the incumbent is carrying out his duties (e.g., Polsby, 1964:25).

The explanations advanced for these patterns and other changes in the level of presidential support are well known and reasonable: disillusionment, dissolution of coalitions, war, state of the economy, and rally-round-the-flag emotions, to name but a few (Mueller, 1970:19-25; Stimson, 1976). While not denigrating these explanations, it is noteworthy that none pays

serious attention to subpopulation dynamics and all appear to presume the population to be homogeneous. To be sure, the mechanics and perhaps even the dynamics are alluded to in the explanations, but very little attention is paid to which subpopulations dissolve the coalitions, which become more disillusioned, or which are more susceptible to sabre-rattling and patriotic appeals. Repeated cross-sectional data preclude answering those questions with standard techniques.

Although we lack evidence on the dynamics of public opinion change, we still need it. We would benefit in explaining the macro level trends if we could tie those trends into meso-level and preferably micro-level behavior. For example, are Republican presidential disapprovers more susceptible to rally-round-the-flag appeals than disapprovers who are Democrats? Does support from higher SES respondents drop off less precipitously than support from low SES respondents? Are women who approve of the president's handling of his job more or less likely than men approvers to decrease support for the president in times of war or economic decline? The differential response to changes in the economy, war, or any stimulus by various subpopulations would assist in explaining the macro level trends and patterns so amply described in the literature. It should also allow analysts to move toward prediction. On a more pragmatic note, politicians would like to assess how various events will influence the support the president garners from particular groups in society.

Mueller (1970:19) conjectured that changes in evaluation would follow a trickle through flow. If, for example, an event

is followed by an aggregate increase in presidential approval, Mueller suspected that approvers would stand fast, some of those with no opinion would become approvers, and some disapprovers would shift to the no opinion category.

However, Mueller rejected the conjecture because there was an extremely high negative correlation ($r = -.98$) between approval and disapproval scores and because the no opinion category stayed remarkably stable at about 12-14 percent. He suggested therefore that support shifts from approval to disapproval or vice-versa. While I think he is largely correct, that judgment cannot be made on the basis of the evidence presented because inferences to individual behavior from aggregate data can be dangerous. Moreover, human nature being what it is, there must be some slippage (Stimson, 1976:5). Not all no-opinion holders will sit at the sidelines and certainly some of the approvers and disapprovers will edge into the no-opinion category.

Two questions must therefore be asked: What are the overall dynamics of changes among presidential evaluations over time and what is the degree or degrees of change that takes place? This research proposes to examine the shift in presidential evaluation among the three response categories found in the Gallup questions before and after salient international events associated with large aggregate level shifts in presidential approval. Not only is the use of those categories intrinsically interesting, but because of differing levels of support among various subgroups, it may be possible to plot changes in support in portions of those groups.

To answer the kinds of questions just raised typically requires panel data. Because of the expense of plotting general trends in panel studies there would be little point in carrying out a massive panel study on the chance that an international incident might occur and pre- and post-event data would therefore have been collected. This paper will present a technique that uses the repeated cross-sectional survey data Gallup collects and publishes to estimate shifts of respondents from one opinion category to another. The technique does not provide individual level data; it estimates the possibility of meso-level states representing the shifts among subpopulations. These estimates will allow estimating the probabilities of individual level behavior.

Specifically the paper will examine the four surges in presidential approval scores exceeding ten percentage points associated with presidential action in international matters over the last twenty years. Jumps in evaluation associated with international events are studied because they constitute a clear-cut phenomenon and so are more susceptible to analysis. The events examined are the Cuban Missile Crisis, the Vietnam peace treaty, the Mayaguez incident, and the Camp David Accords. For all four, Gallup reports the national response to the standard question "Do you approve or disapprove of the way (incumbent's name) is handling his job as President?" prior to and after the events. To present the technique and begin analysis, the problem will next be formalized.

FORMALIZATION OF THE PROBLEM

For the purpose of exposition, consider the problem for the Cuban Missile Crisis. National polls asking whether respondents approved, disapproved, or held no opinion about JFK's presidency were conducted just prior to and just after the crisis. The favorable responses rose dramatically after the event. I wish to estimate the proportion of each group in the pre-event survey that shifted to the other two categories. Because the changes in public opinion are reasonably large both in absolute terms and compared to sampling error, the point estimates from the Gallup surveys are used, and sampling error is not considered important for the purposes of this study. Since the two surveys were administered close together, it is presumed that we are dealing with the same population. Attrition and augmentation of the population is taken to be negligible.

To formalize the problem it is cast into a 3x3 table format. See Figure 1 on p. 20. The row marginals represent the approve (A), disapprove (D), and no opinion (N) aggregates at time 1, the pre-event survey. The column marginals represent the same categories at time 2, the post event survey. As Figure 1 illustrates, we want to estimate the cell entries. The relations among the cell entries and marginals are:

$$a + b + c = M11 \quad (1)$$

$$d + e + f = M12 \quad (2)$$

$$g + h + j = M13 \quad (3)$$

$$a + d + g = M21 \quad (4)$$

$$b + e + h = M22 \quad (5)$$

$$c + f + j = M23 \quad (6)$$

The problem of estimating cell given marginals has an extensive literature (for example, see Bishop, Feinberg, and Holland, 1975). Unfortunately, all approaches presume that the nature of the probability density function governing the distribution of the items being counted is known. The existing estimation literature generally presumes a multinomial or Poisson distribution and shows that the standard maximum likelihood estimate of row times column marginals divided by total table count is justified under the assumption of either probability distribution. In the case of changes in public opinion we do not know the probability distribution. Moreover, I will argue that inequality constraints among the cells should be employed in estimation and nothing in the standard literature gives guidance when inequality relations among the cells obtain. Consequently no assistance comes from the standard statistical estimation literature and a different tack is needed. A first step should be to examine the range and characteristics of possible solutions.

Elementary algebra demonstrates that there are an infinite number of solution sets to equations (1) through (6). The number of solutions can be reduced some by restricting the cell

entries to non-negative integers, a reasonable restriction since the actual surveys generated non-negative integer data. But that still can leave hundreds and thousands of solutions, which is clear from the work Davis and Duncan (1953) have done when they used marginal constraints to identify the range of solutions to similar problems.

Another way to reduce the range of possibilities is to bring in ancillary information to further sharpen the possibilities. Shively (1975) was successful in doing this in a different context. The additional information I intend to use is the commonly observed inertia or persistence in human behavior over time. Although people clearly change over time, it is very reasonable to assume that Democrats tend to stay Democrats, conservatives remain conservatives, and bigots continue as bigots -- at least over a period of a few months. Similarly I presume that most approvers will remain approvers, most disapprovers will maintain that posture and most people with no opinion will continue in that state. Some obviously change, but the inertial constraints simply say that most, though it can be as little as one third of the subpopulation, will persist in their opinion in the few months or less between surveys. Although not typical, these assumptions are no different in degree than the kind that social scientists make -- although unwittingly at times -- about linearity and additivity of relationships, independence of variables, and the properties of residuals in the frequently used multiple regression technique.

Expressed in terms of the 3x3 table, the inertial constraints are:

$$a > b \text{ or } a = b \quad (7)$$

$$a > c \text{ or } a = c \quad (8)$$

$$e > d \text{ or } e = d \quad (9)$$

$$e > f \text{ or } e = f \quad (10)$$

$$j > g \text{ or } j = g \quad (11)$$

$$j > h \text{ or } j = h \quad (12)$$

Even with the set of solutions restricted in number by imposing constraints (1) through (12) there will generally be a number of solutions that satisfy every constraint. How is the analyst to choose one "best" solution, particularly if each one of these solutions is a priori just as likely as any other?

It is important at this point to distinguish between the solutions to the tables, a set composed of nine integers, and the individual cell entries, consisting of single integers. An analogy with poker may make sense. If a person is dealt five cards out of a deck, there are an enormous number of equally likely sets of cards that could be dealt. This corresponds to the set of solutions. Yet certain kinds of hands are more likely than others. This corresponds to the individual cell entries. A player is more likely to have two pair than four of a kind because more of the original possibilities are comprised of two pair than four of a kind. I am arguing that all hands (solutions) are equally likely, but the kinds of hands (solution characteristics) are not. No circularity of reasoning is thereby implied later when one estimate is identified as being more possible than all others.

Playing with a standard and untampered with deck guarantees that there are four suits, each with an ace, deuce, three, four, ... jack, queen, and king. The odds of drawing any particular kind of hand are therefore reasonably well known. But if you are forced to play with a deck a group of five year olds have previously played with, it is wise to enumerate all the cards in the deck. If, for example, only 41 cards remain and all the jacks are missing, the likelihood of being dealt a straight is substantially reduced. In other words, the solution set must be delineated to realistically assess the likelihood of any particular outcome.

Similarly, in the set of all possible solutions to our problem, certain values of each cell will crop up more frequently than others because of constraints (1) through (12). In choosing a solution, therefore, the analyst would want a solution set in which the estimated value for every cell crops up frequently in the set of all solutions. Parallel to a careful card-player checking the deck to see that there are four kings, four queens, etc., so that the priori odds are as expected, the most possible estimator approach (Wanat, 1980) examines all solutions to set the a priori odds of a given cell having values 0, 1, 2,...

To illustrate this a computer program was written to identify all non-negative integer solutions and to tabulate the frequency of occurrence. Tables 1 through 4 (pp. 21-24) display the data for each of the problems. Note that the frequency distributions for each cell vary. They do not all come from the

same family of distributions. Even those that do appear to come from the same kind of distribution would be characterized by different parameters. These distributions constitute envelopes of possible distributions. These distributions represent the influence of constraints on solutions. In the absence of any information about the underlying distributions the envelope or limiting distribution will be used in the analysis that follows.

It is obvious from the Tables that some values of each cell are more possible than others. It is intuitively clear that a solution to the whole problem whose individual cell value distributions occur very frequently is desirable. The estimates would be more likely in the sense of being more possible.

Wanat's 1980 research on vote switching searched out all possible solutions and used the relative frequencies of cell values to approximate probabilities. It then sought the joint probability of solutions based on the individual cell frequencies. Unfortunately parallelism with probability theory demands the assumption of independence be made, one that is not justified. The results derived by his approach are good though theoretically inelegant. To improve on that research the present research employs an analytically neater criterion.

Given the empirically derived set of possible integer solutions to constraints (1) through (12), the analyst wants to choose one that minimizes the chances of error. Hence a least squares strategy is employed. For example, if a' is the best estimate for cell a , we seek to minimize the distance, or error, from every possible cell value for a by minimizing

$$\sum (a_i - a')^2.$$

Taking the derivative with respect to a' and setting that equal to zero yields $-2 \sum (a_i - a') = 0$. Solving for a' generates the desired value of $a' = (\sum a_i) / N$, where N is the number of solutions. The optimal solution for any cell is therefore the expected value based on the frequency distribution derived from the set of possible solutions.

It turns out that the expected values of the desired estimates are consistent with each other. In other words, the estimates computed from the possibility distributions separately all sum to the appropriate marginals. It can be seen, for example, that the estimates for the first row all sum to the row marginal by adding the cell values a'_i , b'_i , and c'_i for every solution:

$$\sum a'_i + \sum b'_i + \sum c'_i = \sum M_{11}$$

Since there are N solutions, $\sum M_{11} = (N)(M_{11})$. Dividing every term in the equation by N gives

$$(1/N) \sum a'_i + (1/N) \sum b'_i + (1/N) \sum c'_i = M_{11}$$

which says that the sum of the estimates for each of the first three cells equals the row marginal.

The same logic applies to all the rows and columns. Hence the expected value estimates conform to the marginal constraints. Since the individual solutions consist of cell entries such that one cell is larger than or equal to the other two row entries, the sum of the solutions will also satisfy the inequality relationships.

FINDINGS

The data for the four events (Cuban Missile Crisis, Vietnam Accords, Mayaguez Incident, and Camp David Agreement) are displayed in Table 5 on p. 25. The table displays the marginals and the dates the surveys were taken. Table 6 on p. 26 shows the estimates of transition probabilities among the approving, disapproving, and no opinion subpopulations. Table 6 also displays some measures of confidence in the estimates.

It is clear that in all four cases the same dynamics appear operative. Of those supporting the president prior to the event, the overwhelming majority continue supporting the chief executive after the event with a small and relatively even split of the remainder going to the other two categories.

Among those who disapproved of presidential execution of his duties prior to the international event, roughly 50 percent continued expressing disapproval. But a large proportion, approximately 40 percent, shifted to approval. With only a few percent shifting to the no opinion category. These estimates support Mueller's judgment that no trickle-through process operates in opinion change. One can guess that the no opinion holders are somewhat outside the pale of involvement.

Examination of those who held no opinion on presidential performance before the international event suggests that about half of them stay in that state after the event. Of those who shift, a bit more become supportive than express disapproval. The impact of these shifts, however, is reasonably small because

of the small proportion (12% to 17%) of those surveyed who expressed no opinion.

By taking an interval one standard deviation about the estimate it is possible to indicate the proportion of the possible occupants in the interval. Those high and low estimates provide a range of estimation associated with the proportion of the presumably equally likely possible solutions. For example, in estimating the proportion of approvers who stayed approvers after the Cuban Missile Crisis, the technique makes a point estimate of 95.1%. The mean standard deviation from the expected value for all solutions over all cells is 1.27. Using that as a confidence interval, this approach says that 79% of the solutions fall in the range of 93.1% to 97.1% standing pat in their evaluation of JFK. A range like that can be adequate for many purposes. In a case such as this both the high and low estimates generate the same conclusions as the point estimates. Approvers stay approvers; almost half the disapprovers shift to approval, a few shift to holding no opinion, and the remainder stay disapprovers; and about half of those with no opinion stay in that state with most of the others shifting to approval.

While it is clear that some cell values are more likely than others, the intervals about the point estimates vary in the proportion of the presumably equally likely values included. In the four incidents examined, the proportion of solutions falling within the interval ranged from 44% to 94% with a mean of 67%. Clearly the questions being asked should determine the size of the confidence interval used. For ease of initial investiga-

tions, however, it was thought that the mean standard deviation from the estimates would be appropriate. However, the investigator should choose the interval according to the problem at hand.

The breadth of the distributions of cell values should not be cause for dismay. In picking a point estimate and examining the frequency distribution about it, it seems that although the point estimate may be the best one possible, it is not usually radically different from nearby estimates. Yet social scientists routinely accept such estimates in their work. In the use of the chi square estimate in goodness of fit tests the expected value is a point estimate which does not actually differ from many other estimates by much. In fact, I have enumerated all possible solutions in tables without any inequality constraints and the distributions of cell values about the chi square estimates are the same kind as those seen in Tables 1 through 4. If we routinely accept a point estimate from a wide range of possibilities as a base point in chi square tests, consistency should not allow us to cavil at the parallel use of a point estimate in the problems under study.

The enumeration of solutions also generates data for analysis that is not focused on point estimates. Because the non-supporters' ranks decline after an international event, it is appropriate to concentrate on the relationship between the size of cells d, e, and f. By assumption, cell e is greater than or equal to cells d and f. But that leaves open the relationship between cells d and f, the number of non-supporters who became

supporters or who adopted no opinion. Our point estimate analysis says that the estimate for cell d exceeds the estimate for cell f. But in how many of the individual solutions is that also the case?

The program that searches and enumerates the solutions also counts the number of solutions in which cell d exceeds cell f. For the Cuban missile crisis, the Vietnam accords, and the Camp David agreement, in every one of the solutions more pre-event disapprovers turned into approvers than into no-opinion holders. In the Mayaguez incident, of 2908 solutions 2899, or 99.69%, were such that more one-time disapprovers became approvers than no opinion respondents. Although moving from point estimates suggests a lack of accuracy, the step to analyzing the relationship among cell values in the solutions gives more encompassing information. These data say that given the constraints set by the marginals and behavioral inertia, in all but an infinitesimal number of solutions, more one-time disapprovers become approvers than become neutral. The process of opinion change in our circumstances is clearly, therefore, one of conversion from rather extreme positions rather than gradual shift.

It may be possible by examination of the marginals and extreme values of cells to analytically determine the relations among a few cells. But that approach is not always possible as the Mayaguez case, where dominance is not complete, indicates. Therefore all solutions are examined as they are enumerated.

Given the variety in the marginals in at least three of the four cases, perhaps the most striking aspect of the results is

the similarity of the estimates. This suggests that the dynamic of opinion change is not dependent on the marginals, which vary, but inheres in the reaction to the presidential action. An alternative explanation of the similarity of estimates is that the technique itself is incapable of generating anything else. To investigate that possibility and to address some of the questions about differences in opinion change in demographic subpopulations, estimates of opinion change on Republican and Democratic respondents will be sought. We expect, for instance, that Republican respondents at the time of the Mayaguez incident will react differently than will the Democratic respondents. If the estimates are indeed different for the subpopulations and if they conform to what party loyalty would predict, we would have evidence that: 1) the technique is sensitive to marginals and does not grind out the same estimates regardless of input, 2) more faith can be put into the conclusion that the conversion dynamic is in operation in presidential support change, and 3) subpopulation dynamics are identifiable.

For both the Mayaguez and Camp David events, the Gallup organization collected the approve/disapprove/no opinion responses separately for Democrats and Republicans prior to and just after the events. Those data were analyzed and the results are displayed in Table 7 on p. 27. That table clearly shows that partisan respondents generate differing estimates. The Republicans' estimates for the Mayaguez incident are very close to the Democrats' estimates for the Camp David accords. In both of those cases we are talking about respondents who share the party

affiliation of the President. In the case of respondents with partisan affiliation differing from the President the responses are similar between Democrats with regard to Mayaguez and Republicans with regard to Camp David. The similarity is not as close as in the first case, but we did not postulate any particular pattern; we merely expected that differences would appear and they would be substantively plausible. And it is certainly within the realm of substantive interpretation that cases of congruence between respondent and presidential party affiliation should generate similar responses while cases of partisan difference should yield similar results.

The major difference in estimates between respondents and presidents of opposing parties lies in what the pre-event disapprovers did. Proportionately more disapproving Democrats supported Ford's actions on the Mayaguez than did disapproving Republicans support Carter's Camp David actions. But this is to be expected given that the military nature of the Mayaguez seizure would bring national pride into play more than mediating between two other nations. In any case it is clear that the technique can estimate opinion change in various subpopulations.

To shift from point estimates to the relations between cells, analysis was carried out on the proportion of solutions in which the number of disapprovers turned approvers exceeded the number of disapprovers turned to no-opinion holders. For both Democratic and Republican respondents centering on the Camp David event, in every solution more approvers became disapprovers than took on no opinion. Regarding the Mayaguez incident, in 97.98%

of the solutions involving Republican respondents and in 97.96% of the solutions involving Democratic respondents more disapprovers became approvers than assumed a no-opinion stance. Once again the evidence says that conversion is the dominant dynamic in opinion change in the face of international actions by the president.

IMPLICATIONS

This research offers evidence that increases in the approval of presidential handling of his duties after an international event largely comes from a conversion of those who previously had disapproved of his execution of duties. The shift to and from the no opinion category is small and balances out, leaving the largest net increase coming from the former disapprovers.

This suggests that survey research might spend less time on those who hold no opinion. For one, it is often harder to get responses from them. But more importantly, this research suggests that they are largely irrelevant when changes in aggregate approval is examined. They are not active participants in political life and may be ignored for some purposes. Hence we have provided additional evidence that the standard practice of using the "approve of the president" percentage as a dependent variable in measuring presidential popularity (e.g. Mueller) is justified.

The volatility of respondents in shifting from disapproval to approval suggests the importance of the various rally-round-

the-flag explanations for opinion changes. The national interest apparently overrides previous evaluative predispositions. This betokens a substantial basic support for the country in its presidential embodiment regardless of the respondent's prior positions. In fact, the rally-round-the-flag dynamic is powerful enough to shift a healthy proportion -- about 40% -- of disapprovers.

In the voting arena there is evidence that candidates are better served by holding the waverers and enticing the uncommitted voters than by trying to convert those of the opposing party. The voting act is an example of moderately stable behavior. In presidential evaluation in times of international action, however, support can come from those who were on the opposite side of the political fence. Political activism or awareness is enough, it appears, to augment the president's support when he acts decisively in international matters.

But perhaps most important in this paper is the refinement and application of a technique that allows estimating sub-macro-level changes in public opinion from repeated cross-sectional data. Archival data can now be mined and processed to generate new insights and to offer corroborative data or to suggest invalidating evidence for conjectures. The technique does not promise certitude, but it clearly indicates the odds of estimates being correct. The researcher can thereby judge the utility of the estimates in each application and use the estimates if appropriate.

FIGURE 1: FORMALIZATION OF ESTIMATION PROBLEM

	approve	dis- approve	no opinion			
approve	I I a I	I I b I	I I c I	M11	I a I t I	
disapprove	I I d I	I I e I	I I f I	M12	I t I i I m I e I	
no opinion	I I g I	I I h I	I I j I	M13	I 1	
	M21	M22	M23			
	at time 2					

constraints on solutions:

a + b + c =	M11	(1)
d + e + f =	M12	(2)
g + h + j =	M13	(3)
a + d + g =	M21	(4)
b + e + h =	M22	(5)
c + f + j =	M23	(6)
a > b or a = b		(7)
a > c or a = c		(8)
e > d or e = d		(9)
e > f or e = f		(10)
j > g or j = g		(11)
j > h or j = h		(12)

TABLE 1: FREQUENCIES OF CELL VALUES IN SOLUTIONS AND ESTIMATES FOR CHANGES IN PRESIDENTIAL SUPPORT AT THE CUBAN MISSILE CRISIS

Cell Value	Appr to Appr	Appr to Disap	Appr to No Op	Disap to Appr	Disap to Disap	Disap to No Op	No Op to Appr	No Op to Disap	No Op to No Op
0	0	54	0	0	0	52	0	31	0
1	0	39	67	0	0	37	0	25	0
2	0	26	43	0	0	34	1	33	0
3	0	15	24	0	0	13	5	23	0
4	0	7	7	0	0	6	14	22	0
5	0	1	1	1	0	0	27	7	0
6	0	0	0	5	0	0	42	1	14
7	0	0	0	14	0	0	37	0	32
8	0	0	0	25	0	0	16	0	51
9	0	0	0	38	3	0	0	0	30
10	0	0	0	40	20	0	0	0	15
11	0	0	0	19	47	0	0	0	0
12	0	0	0	0	32	0	0	0	0
13	0	0	0	0	22	0	0	0	0
14	0	0	0	0	12	0	0	0	0
15	0	0	0	0	6	0	0	0	0
55	1	0	0	0	0	0	0	0	0
56	6	0	0	0	0	0	0	0	0
57	16	0	0	0	0	0	0	0	0
58	27	0	0	0	0	0	0	0	0
59	35	0	0	0	0	0	0	0	0
60	34	0	0	0	0	0	0	0	0
61	23	0	0	0	0	0	0	0	0

expected values based on most possible frequency distributions

58.99 1.19 1.82 9.04 11.77 1.18 5.96 2.04 8.00

associated transition probabilities

.951 .019 .629 .411 .535 .054 .373 .127 .500

TABLE 2: FREQUENCIES OF CELL VALUES IN SOLUTIONS AND ESTIMATES
FOR CHANGES IN PRESIDENTIAL SUPPORT SURROUNDING
THE VIETNAM ACCORDS

Cell Value	Appr to Appr	Appr to Disap	Appr to No Op	Disap to Appr	Disap to Disap	Disap to No Op	No Op to Appr	No Op to Disap	No Op to No Op
0	0	18	0	0	0	45	1	7	0
1	0	17	50	0	0	22	2	6	0
2	0	15	18	0	0	4	8	16	0
3	0	11	3	0	0	0	11	13	0
4	0	6	0	0	0	0	25	21	11
5	0	3	0	0	0	0	17	7	32
6	0	1	0	0	0	0	7	1	28
12	0	0	0	1	0	0	0	0	0
13	0	0	0	3	0	0	0	0	0
14	0	0	0	7	0	0	0	0	0
15	0	0	0	11	0	0	0	0	0
16	0	0	0	15	0	0	0	0	0
17	0	0	0	17	0	0	0	0	0
18	0	0	0	17	6	0	0	0	0
19	0	0	0	0	18	0	0	0	0
20	0	0	0	0	17	0	0	0	0
21	0	0	0	0	14	0	0	0	0
22	0	0	0	0	8	0	0	0	0
23	0	0	0	0	5	0	0	0	0
24	0	0	0	0	2	0	0	0	0
25	0	0	0	0	1	0	0	0	0
44	1	0	0	0	0	0	0	0	0
45	4	0	0	0	0	0	0	0	0
46	9	0	0	0	0	0	0	0	0
47	13	0	0	0	0	0	0	0	0
48	16	0	0	0	0	0	0	0	0
49	16	0	0	0	0	0	0	0	0
50	12	0	0	0	0	0	0	0	0

expected values based on most possible frequency distributions
47.90 1.76 1.34 16.18 20.39 0.42 3.92 2.85 5.24
associated transition probabilities
.939 .035 .026 .437 .551 .011 .326 .237 .437

TABLE 3: FREQUENCIES OF CELL VALUES IN SOLUTIONS AND ESTIMATES
FOR CHANGES IN PRESIDENTIAL SUPPORT SURROUNDING
THE MAYAGUEZ INCIDENT

Cell Value	Appr to Appr	Appr to Disap	Appr to No Op	Disap to Appr	Disap to Disap	Disap to No Op	No Op to Appr	No Op to Disap	No Op to No Op
0	0	318	0	0	0	393	262	430	0
1	0	318	516	0	0	382	338	490	0
2	0	318	482	0	0	406	358	428	0
3	0	318	439	0	0	373	433	461	0
4	0	314	389	0	0	369	432	382	0
5	0	302	333	0	0	314	504	394	91
6	0	278	273	2	0	278	375	227	340
7	0	242	210	8	0	201	206	96	553
8	0	196	146	20	0	129	0	0	576
9	0	145	82	38	0	52	0	0	455
10	0	93	32	63	0	11	0	0	342
11	0	45	6	95	0	0	0	0	245
12	0	17	0	135	0	0	0	0	160
13	0	4	0	183	0	0	0	0	93
14	0	0	0	223	0	0	0	0	42
15	0	0	0	255	0	0	0	0	11
16	0	0	0	280	0	0	0	0	0
17	0	0	0	298	0	0	0	0	0
18	0	0	0	310	1	0	0	0	0
19	0	0	0	310	18	0	0	0	0
20	0	0	0	280	65	0	0	0	0
21	2	0	0	220	136	0	0	0	0
22	10	0	0	140	223	0	0	0	0
23	28	0	0	48	318	0	0	0	0
24	58	0	0	0	318	0	0	0	0
25	100	0	0	0	318	0	0	0	0
26	149	0	0	0	318	0	0	0	0
27	201	0	0	0	301	0	0	0	0
28	252	0	0	0	267	0	0	0	0
29	288	0	0	0	216	0	0	0	0
30	303	0	0	0	171	0	0	0	0
31	299	0	0	0	122	0	0	0	0
32	280	0	0	0	80	0	0	0	0
33	253	0	0	0	36	0	0	0	0
34	220	0	0	0	0	0	0	0	0
35	182	0	0	0	0	0	0	0	0
36	140	0	0	0	0	0	0	0	0
37	95	0	0	0	0	0	0	0	0
38	48	0	0	0	0	0	0	0	0

expected values based on most possible frequency distributions
30.74 4.33 3.93 16.72 25.83 3.45 3.54 2.84 8.62
associated transition probabilities
.788 .111 .101 .363 .562 .075 .236 .189 .575

TABLE 4: FREQUENCIES OF CELL VALUES IN SOLUTIONS AND ESTIMATES FOR CHANGES IN PRESIDENTIAL SUPPORT SURROUNDING THE CAMP DAVID ACCORDS

Cell Value	Appr to Appr	Appr to Disap	Appr to No Op	Disap to Appr	Disap to Disap	Disap to No Op	No Op to Appr	No Op to Disap	No Op to No Op
0	0	214	0	0	0	268	42	142	0
1	0	210	338	0	0	233	69	181	0
2	0	198	282	0	0	233	90	160	0
3	0	175	225	0	0	184	134	190	0
4	0	145	169	0	0	156	162	162	0
5	0	133	115	0	0	93	216	174	0
6	0	81	65	0	0	48	231	133	80
7	0	53	27	0	0	13	181	65	185
8	0	29	7	0	0	0	103	21	264
9	0	9	0	0	0	0	0	0	261
10	0	1	0	2	0	0	0	0	192
11	0	0	0	8	0	0	0	0	133
12	0	0	0	20	0	0	0	0	78
13	0	0	0	39	0	0	0	0	35
14	0	0	0	65	0	0	0	0	0
15	0	0	0	93	0	0	0	0	0
16	0	0	0	121	0	0	0	0	0
17	0	0	0	149	0	0	0	0	0
18	0	0	0	175	0	0	0	0	0
19	0	0	0	192	9	0	0	0	0
20	0	0	0	182	48	0	0	0	0
21	0	0	0	132	119	0	0	0	0
22	0	0	0	50	214	0	0	0	0
23	0	0	0	0	203	0	0	0	0
24	0	0	0	0	181	0	0	0	0
25	0	0	0	0	148	0	0	0	0
26	2	0	0	0	113	0	0	0	0
27	10	0	0	0	85	0	0	0	0
28	28	0	0	0	57	0	0	0	0
29	55	0	0	0	36	0	0	0	0
30	87	0	0	0	15	0	0	0	0
31	119	0	0	0	0	0	0	0	0
32	147	0	0	0	0	0	0	0	0
33	166	0	0	0	0	0	0	0	0
34	173	0	0	0	0	0	0	0	0
35	162	0	0	0	0	0	0	0	0
36	134	0	0	0	0	0	0	0	0
37	95	0	0	0	0	0	0	0	0
38	50	0	0	0	0	0	0	0	0

expected values based on most possible frequency distributions

33.37 2.81 2.82 17.87 23.92 2.21 4.77 3.27 8.97

associated transition probabilities

.856 .072 .072 .406 .544 .050 .281 .192 .527

TABLE 5: EVENTS AND THEIR AGGREGATE CHARACTERISTICS

Event	Survey Dates	Approve (A)	Disapprove (D)	No Opinion (N)
Cuban Missile Crisis (1)	20-25 Sep 1962 16-21 Nov 1962	62% 74%	22% 15%	16% 11%
Vietnam Peace Treaty (2)	12-15 Jan 1973 26-29 Jan 1973	51% 68%	37% 25%	12% 7%
Mayaguez Incident (3)	18-21 Apr 1975 30 May-2 June	39% 51%	46% 33%	15% 16%
Camp David Accords (4)	21-24 Jul 1978 19 Sept 1978	39% 56%	44% 30%	17% 14%

- (1) The Cuban Missile Crisis began on 16 Oct when photos of Russian missiles were released. JFK announced the presence of the missiles and initiated the naval blockade on 22 Oct. On 28 Oct the USSR agreed to remove the weapons.
- (2) Nixon announced the agreement on 23 Jan. The peace pact was signed in Paris on 27 Jan.
- (3) Cambodian naval vessel took the Mayaguez on 12 May. U.S. forces recaptured the Mayaguez on 15 May.
- (4) The Mideast summit at Camp David ended with the signing of the framework for peace on 17 Sept.

TABLE 6: ESTIMATED TRANSITION PROBABILITIES AND ASSOCIATED STATISTICS FOR OPINION CHANGE IN FOUR EVENTS

ESTIMATED TRANSITION PROBABILITIES

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Cuban....	.951	.019	.029	.411	.535	.054	.373	.127	.500
Vietnam..	.939	.035	.026	.437	.551	.011	.326	.237	.437
Mayaguez..	.788	.111	.101	.363	.562	.075	.236	.189	.575
Camp Dav..	.856	.072	.072	.406	.544	.050	.281	.192	.527

LOW ESTIMATES BASED ON ONE STD DEV ERROR INTERVAL

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Cuban....	.931	.000	.009	.353	.477	.000	.293	.048	.421
Vietnam..	.916	.011	.003	.405	.519	.000	.226	.137	.337
Mayaguez..	.720	.043	.032	.305	.504	.017	.058	.012	.397
Camp Dav..	.801	.018	.018	.358	.495	.002	.155	.067	.402

HIGH ESTIMATES BASED ON ONE STD DEV ERROR INTERVAL

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Cuban....	.972	.040	.050	.469	.593	.112	.452	.207	.579
Vietnam..	.963	.058	.050	.470	.584	.044	.426	.337	.537
Mayaguez..	.857	.179	.169	.421	.620	.133	.414	.367	.753
Camp Dav..	.910	.127	.127	.454	.592	.099	.406	.317	.652

PROPORTION OF POSSIBLE CASES BETWEEN HIGH AND LOW ESTIMATES

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Cuban....	.676	.838	.944	.725	.711	.866	.746	.570	.796
Vietnam..	.634	.451	.958	.606	.437	.944	.746	.704	.845
Mayaguez..	.489	.526	.659	.500	.523	.730	.839	.741	.863
Camp Dav..	.528	.593	.826	.519	.700	.656	.605	.559	.843

Event	Standard Deviation Error Interval
Cuban.....	1.271
Vietnam...	1.201
Mayaguez..	2.668
Camp David	2.126

TABLE 7: ESTIMATED TRANSITION PROBABILITIES AND ASSOCIATED STATISTICS FOR PARTISAN RESPONDENTS IN THE MAYAGUEZ (Mygz) AND CAMP DAVID (CmpD) INCIDENTS

ESTIMATED TRANSITION PROBABILITIES

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Mygz/Rep	.908	.042	.050	.356	.553	.091	.239	.181	.580
Mygz/Dem	.628	.191	.181	.332	.574	.094	.206	.194	.600
CmpD/Rep	.704	.230	.066	.162	.826	.011	.265	.265	.469
CmpD/Dem	.895	.060	.045	.383	.573	.044	.280	.208	.512

LOW ESTIMATES BASED ON ONE STANDARD DEVIATION ERROR INTERVAL

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Mygz/Rep	.877	.011	.019	.289	.486	.024	.068	.010	.409
Mygz/Dem	.518	.082	.071	.276	.519	.038	.000	.000	.373
CmpD/Rep	.611	.137	.000	.125	.790	.000	.072	.072	.275
CmpD/Dem	.858	.022	.008	.329	.518	.000	.144	.072	.375

HIGH ESTIMATES BASED ON ONE STANDARD DEVIATION ERROR INTERVAL

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Mygz/Rep	.939	.072	.081	.423	.620	.158	.410	.352	.751
Mygz/Dem	.737	.301	.291	.388	.630	.150	.433	.421	.828
CmpD/Rep	.797	.323	.160	.199	.863	.048	.459	.459	.663
CmpD/Dem	.933	.097	.083	.438	.629	.098	.417	.345	.648

PROPORTION OF POSSIBLE SOLUTIONS BETWEEN HIGH AND LOW ESTIMATES

	A->A	A->D	A->N	D->A	D->D	D->N	N->A	N->D	N->N
Mygz/Rep	.574	.642	.540	.456	.650	.647	.762	.610	.786
Mygz/Dem	.466	.493	.586	.539	.493	.566	.973	.910	.844
CmpD/Rep	.321	.401	.976	.401	.401	1.000	.860	.860	1.000
CmpD/Dem	.537	.427	.935	.546	.434	.909	.556	.501	.864

Event	Std. Dev. Error	Pre-event A D N	Post-Event A D N
Mayaguez/Republicans	1.881	61% 28% 11%	68% 20% 12%
Mayaguez/Democrats	3.182	29% 57% 14%	40% 41% 19%
Camp David/Republicans	2.326	25% 63% 12%	31% 61% 8%
Camp David/Democrats	1.911	51% 35% 14%	63% 26% 11%

REFERENCES

- Bishop, Yvonne M.M., Stephen E. Feinberg, and Paul W. Holland, DISCRETE MULTIVARIATE ANALYSIS: THEORY AND PRACTICE, Cambridge, Mass.: MIT Press, 1975.
- Davis, Beverly and C. Dudley Duncan, "An Alternative to Ecological Correlation," AMERICAN SOCIOLOGICAL REVIEW, 18(1953) 665-666.
- Mueller, John E., "Presidential Popularity from Truman to Johnson," AMERICAN POLITICAL SCIENCE REVIEW, 64(1970) 18-34.
- _____, "Trends in Popular Support for the Wars in Korea and Vietnam," AMERICAN POLITICAL SCIENCE REVIEW, 65(1971) 358-375.
- _____, WAR, PRESIDENTS, AND PUBLIC OPINION, New York: Wiley, 1973.
- Polsby, Nelson, CONGRESS AND THE PRESIDENCY, Englewood Cliffs, N.J.: Prentice-Hall, 1964.
- Shively, W. Phillips, "Utilizing External Evidence in Cross-Level Inference," POLITICAL METHODOLOGY, 1(1975) 61-73.
- Stimson, James A., "Public Support for American Presidents: A Cyclic Model," PUBLIC OPINION QUARTERLY, 40(1976) 1-21.
- Wanat, John, "The Application of a Non-Analytic, Most Possible Estimation Technique: The Relative Impact of Mobilization and Conversion of Votes in the New Deal," POLITICAL METHODOLOGY, 6(1980) 357-374.

ESTIMATING THE DEGREE OF MOBILIZATION AND CONVERSION IN THE 1890's:
THE NATURE OF POLITICAL CHANGE IN ONE CRITICAL ELECTION

John Wanat and Karen Burke
Political Science Department
University of Illinois at Chicago Circle
Box 4348
Chicago, Illinois 60680

This paper was prepared for presentation at the Annual Meeting of the Midwest Political Science Association, April 16, 1981, Cincinnati. Support in part came from the Law Enforcement Assistance Administration under grant 79-NI-AX-0058. LEAA bears no responsibility for judgments or analysis in the paper. We are indebted to Richard Jensen, Gerald Strom, Charles Williams and Linda Cecsarini for advice given in the course of our research.

ABSTRACT

ESTIMATING THE DEGREE OF MOBILIZATION AND CONVERSION IN THE 1890's:
THE NATURE OF POLITICAL CHANGE IN ONE CRITICAL ELECTION

The primary voting dynamic in the critical presidential election of 1896 is held in the literature to be the conversion of disenchanted Democrats to the Republican banner. Mobilization of new voters is not given much attention. To assess the role of mobilization, the vote shift possibilities from 1892 to 1896 were formalized, and analysis of all possible scenarios conforming to the aggregate data characterizing the electoral shift were carried out by computer. Solutions to the 3x3 tables representing the 1892 to 1896 vote history show that in the Midwest more voters were mobilized in the 1896 election than were converted. In the Northeast the conventional wisdom was untouched.

ESTIMATING THE DEGREE OF MOBILIZATION AND CONVERSION IN THE 1890's:
THE NATURE OF POLITICAL CHANGE IN ONE CRITICAL ELECTION

John Wanat and Karen Burke
Department of Political Science
University of Illinois at Chicago Circle

Though certain elections have been classified with great regularity as critical elections, the definition of what constitutes a critical election has not been clear. V. O. Key (1971: 28) initially classified an election as critical if an enduring realignment was produced by sharp and intense changes in party support. Later Key (1959:198) noted that secular realignment is also compatible with a critical election, but the change in party support is the result of gradual shifts in voter affiliation. Burnham (1970: 4), using Key's formulation, named the sharp change in political behavior a critical realignment. Pomper (1971: 182) modified the concept further by calling elections where the majority party retains pre-eminence through a change in the voter base converting elections; if the majority party loses

its majority status he called the elections realigning. Sundquist (1973: 7), rejecting the distinction between forms of realignment, suggested that the term realignment encompasses both critical and secular realignment if there is an organic change in the party system. Despite such variety in definition, at the core of the concept is a consensus that critical elections are characterized by a decomposition of habitual party loyalty, resulting in a shift of decisive minorities from one political party to another. (Burnham, 1970: 6; Campbell, 1971: 117; Sundquist, 1973: 18)

Some scholars, however, have suggested that the notion of a critical election merits elaboration. For example, Sellers (1971: 159), though accepting the notion that a critical election is partially the result of a shift in voter allegiance, places more emphasis on the contributions of the young and the new voter entering the voting fray. Seagull (1980: 70) attempts to differentiate between a critical election and a secular realignment by noting that secular realignments place less stress upon heated issues and more stress on population shifts and alterations of social bases. Although Seagull is still willing to provide a distinction between the two categories of elections, he suggests that each of these are composed of issues and population shifts, though the contributions of each varies from election to election.

Since a political system at any time is composed of continuing voters as well as new voters resulting from demographic changes we find the Seagull and Sellers positions on the components of a critical election appealing. We do not mean to imply that scholars totally ignore the impact of new voters, but the extent of most treatments is usually a mild reference to the fact that new voters may have some impact. The volume of literature on vote switching in critical elections indicates that prime emphasis lies in the conversion of votes. With the exception of Andersen (1976) and Wanat (1979) virtually no empirical attention has been paid to the relative weight of conversion and mobilization.

The 1896 presidential election is universally considered to be a critical election, not only because of its impact on the politics of succeeding years, but also because of the conversion of voters. Some theorists (e.g., Burnham, 1970; Sundquist, 1973) suggest that the volatile, divisive economic issues in 1896 resulted in an obvious polarization of political choices forcing voters out of the comfort of traditional voting patterns into new territory and new enduring relationships. The result is a conversion of 1892 voters.

However intriguing, when applied to 1896, the grand conversion theory suffers from some methodological problems. Burnham accurately described the severe economic chaos in 1896, presented cogent reasons why the Republican and Democratic positions ap-

pealed to some geographical regions and not others, and equally as painstakingly analyzed the aggregate electoral outcomes. Conversion consequently appears plausible. But problems develop when aggregate data are assumed to represent a compilation of individual conversions. The question thus arises whether any evidence exists that the outcome of the 1896 election is based on factors other than conversion. Although the literature leans in the conversion direction, mobilization of new voters, as Andersen (1976) and Wanat (1979) have shown for the 1932 election, provides an alternative and probably better explanation for some critical elections. We suggest three reasons why mobilization might form the basis for much of what happened in 1896.

First, relying on census data, we note that migration among states could result in an influx of potential voters that might change the relative proportion of party advocates in each state. Within this migratory movement lies the possibility of some states displaying growth in eligible voters and others witnessing a decline in voting population from migration. For example, after accounting for birth and death rates, we note that New York between 1880 and 1890 lost 146,400 native whites over ten years old to migration and Illinois lost 170,000 native whites to migration while Massachusetts gained 31,900 native white migrants in that same period. If the foreign-born and black populations are included in this calculation the enormous impact of migratory patterns becomes even more noticeable. Illinois, in this case,

showed a loss from migration of 59,000 between 1870 to 1880 and a gain of 170,300 people between 1880 and 1890. Pennsylvania gained 19,100 migrants between 1870 and 1880 and gained 285,100 migrants between 1880 and 1890 (HISTORICAL STATISTICS, 1975: 91-93). The migrant, consequently, was not necessarily converted in 1896, but could have carried his normal political affiliation into the polls in his new state of residence.

Second, if mobilization of new voters was a goal of political parties in 1896, there was a supply of twenty-one year old males that had never experienced voting in a presidential election before. For example, based on the 1890 census, Pennsylvania had 660,000 people in the 15 - 24 year old category. Assuming that approximately 50% of this category are males, and excluding all those that could have voted in 1892, there are approximately 215,000 fifteen to eighteen year olds who would fall into the twenty-one year old male category in 1896. In like manner, Illinois had 158,600 potential new voters and Ohio 150,200 (HISTORICAL STATISTICS, 1975: 24-37).

Finally, the late 1800's may be seen as a period of immigrant influx unmatched in any earlier period. It has been estimated that the growth attributed to immigration during this period was one-half million annually (Jensen, 1971: 187-188). In 1890, the foreign-born population stood at 9,249,560 people, or approximately 14.7% of the total population (Carpenter, 1969: 308). This swell in population from immigration offered a pool

of voters ripe for mobilization. Though the extent of immigrant contribution to electoral outcome would vary by state, we can illustrate the approximate importance of the immigrant to electoral outcome by examining Chic. Of the 1,016,464 males twenty-one and over in 1890, roughly 15.1% were naturalized citizens (HISTORICAL STATISTICS, 1975: 1068). Thus, in Ohio in 1890, approximately 108,486 immigrants fell into the potential voter camp. Though some historians (e.g., Jensen, 1971: 254) feel that immigrants, due to a lack of political integration, were less inclined to vote than native citizens, the economic situation of 1896 and/or the active pursuit of new voters by the political parties may have initiated previously inactive voters into the system.

Certainly in some states (e.g., Indiana and Wisconsin) the immigrant was well received. Merely by meeting normal voter requirements, such as residency, the immigrant could cast a vote without full citizenship. Further, even in states requiring full citizenship prior to obtaining voting rights, immigrants were obtaining citizenship, and even if they opted not to change their status their children were reaching voting age. For example, the native white population born in the United States of foreign parents increased from 16.5% in 1890 to 20.6% in 1900 (Carpenter, 1969: 6). The total impact of this immigrant influx is that a pool of voters existed for which the historical and social cues of party preference did not exist.

This brief review of the components underlying aggregate data suggest that voter conversion is only one explanation among a universe of dynamic possibilities. As suggested earlier the process underlying a critical election may be viewed as a mixture of conversion and mobilization behaviors.

AGGREGATE DATA ANALYSIS

In Table A the votes for each party's presidential candidate in both 1892 and 1896 are presented, as are measures of vote change for the parties. While it is clear that one cannot infer individual level behavior from aggregate data such as these, they can generate conjectures. One striking regularity in these data leads to a conjecture and demands investigation.

Note that with the exception of Wisconsin, all the Midwestern states show an increase in Democratic vote, while all states in the Northeast show a decline in the Democratic vote from 1892 to 1896. The natural conjecture for the Northeast, therefore, is that conversion is likely -- particularly since the increase in the total of those voting is small. But in the Midwest it does not appear as likely to have been a zero sum game. Consequently, the mobilization of new voters, who were relatively numerous in that region, may well explain or at least contribute to the Republican hegemony in 1896.

FORMALIZATION OF THE PROBLEM

We are interested in where the 1896 voters came from and where the 1892 voters went to in 1896. If conversion were the only operative mechanism, the problem could be expressed as a 2x2 crosstabulation. The row marginals would be the 1892 Democratic and Republican vote; the column marginals would be the 1896 Democratic and Republican vote; the problem would be to estimate the cell entries. As clear from his chart, Sundquist's conceptualization of realignment operates on this basis (Sundquist, 1973: 19).

Earlier in the paper, we suggested that conversion alone could not have occurred. The augmentation of the 1892 voting age population by naturalization of immigrants and attainment of voting age by the native born along with the diminution of that population by death and outmigration constitute dynamics outside the realm of conversion. To accommodate these realities of augmentation and attrition, a third category, which we call the pool category, is needed. That category represents those in any year who, though eligible, did not vote, those who voted for third party candidates, and a small adjustment population needed to make the population commensurate between 1892 and 1896. (That will be explained more fully later.)

The problem, therefore, can be expressed as a 3x3 table with the row marginals representing 1892 Democrats, Republicans, and pool members and the column marginals representing the same groups in 1896. Figure 1 lays out the general crosstabulation as well as a particular expression of the problem for Minnesota.

In the Figure, cells DD and RR represent Democratic and Republican voters who stayed faithful to their parties' candidates in both 1892 and 1896. Cell DR holds the Democrats who defected to the Republicans in 1896 and cell RD are those who defected the other way. The size of those two cells indicates the degree of conversion. Mobilization, on the other hand, is seen in cells PD and PR, which indicate how many from the pool in 1892 voted Democratic and Republican, respectively, in 1896. Cells DP, FP, and PP represent those in the 1892 eligible electorate who died, did not vote or voted for third party candidates in 1896.

Voting statistics tell us some of the marginals: D1, F1, D2, and R2. But others, F1 and P2, must be computed. It must be remembered that the entire table is composed of all persons who were eligible to vote in either 1892 or 1896. Although those two groups are largely the same, there are differences because of newly enfranchized voters, newly interested voters, voter dropouts, and death.

We proceed to compute the pool marginals as follows. Burnham's turnout rates (HISTORICAL STATISTICS, 1975: 1071-1072)

are used to calculate the eligible electorate for both 1892 and 1896. Hence, if the 1892 turnout rate is called T092 and the 1896 figure is T096, the available electorates, AE92 and AE96, are computed as:

$$\begin{aligned} \text{AE92} &= \frac{\text{1892 vote}}{(\text{E1} + \text{E2})} / \text{T092} \\ \text{AE96} &= \frac{\text{1896 vote}}{(\text{E2} + \text{E3})} / \text{T096} \end{aligned}$$

In all the states we examined save one, the 1896 electorate exceeds the 1892 electorate, which indicates that the total table population must be at least as large as AE96. It will be larger than the 1896 available electorate by the number in the 1892 available electorate who died between 1892 and 1896. Data indicate that mortality in that period is about 7.56% (see below and HISTORICAL STATISTICS, 1975: 63). Therefore, total table population is:

$$\begin{aligned} \text{TCT} &= \text{AE96} + 7.56\%(\text{AE92}) \\ &= \frac{\text{1896 vote}}{(\text{E2} + \text{E3})} / \text{T096} + .0756 \frac{\text{1892 vote}}{(\text{E1} + \text{E2})} / \text{T092} \end{aligned}$$

Knowing the total available population allows calculation of the pool population as:

$$\text{P1} = \text{TCT} - \text{D1} - \text{R1}$$

$$\text{P2} = \text{TCT} - \text{D2} - \text{R2}$$

We are now left with the classic ecological inference problem: how to estimate cell entries from marginals. The marginals allow us to state the conditions that connect cell entries and marginals:

$$\text{DD} + \text{DR} + \text{DP} = \text{D1} \quad (1)$$

$$\text{RD} + \text{RR} + \text{RP} = \text{R1} \quad (2)$$

$$\text{PD} + \text{PR} + \text{PP} = \text{P1} \quad (3)$$

$$\text{DD} + \text{RD} + \text{PD} = \text{D2} \quad (4)$$

$$\text{DR} + \text{RR} + \text{PR} = \text{R2} \quad (5)$$

$$\text{DP} + \text{RP} + \text{PP} = \text{P2} \quad (6)$$

Knowledge of elementary algebra indicates that no unique solution is possible for conditions (1) through (6) because there are 9 unknowns and only 6 equations. In fact, there are an infinite number of solutions if the cell variables are allowed to be real numbers. But since the problem represents enumerated data, it is fully appropriate to require that the variables be non-negative integers. That stricture reduces the number of solutions to a finite, though potentially very large number.

To make the formalization of the problem more isomorphic to the phenomenon under study and, fortunately, to further reduce the number of solutions, we incorporate the effect of death on the 1892 electorate into the problem. The death rates in Massachusetts in 1893, 1894, 1895, and 1896 were 20.5, 19.1, 19.0 and 19.3 deaths per thousand (HISTORICAL STATISTICS, 1975: 63). Since those data are the only ones available, we assume they are reasonably representative of the Northeast and Midwestern states we investigate. Their compound effect indicates that 7.56% of the 1892 electorate would be dead by 1896. To incorporate mortal attrition into our model we require that at least 7.56% of all

1892 categories move to the pool category in 1896. The inequality is required because there are more reasons for not voting than dying. Expressed more succinctly,

$$LP > .0756D1 \quad (7)$$

$$RP > .0756R1 \quad (8)$$

$$PP > .0756P1 \quad (9)$$

Because infant mortality rates are usually higher than adult rates, the actual attrition rates for the voting age population may be less than 7.56%. Using as low as a 7% attrition rate generated essentially the same results as those to be reported below. Hence we will use the 7.56% mortality rates in the rest of the paper with confidence.

Yet another aspect of reality that must be incorporated into the model is the inertial quality of most human behavior. For instance, given a group of conservatives at one time it is reasonable to assume that most of them will still be conservatively oriented a few years later. In the case of the 1896 election it is reasonable to assume that most of those who voted Republican in 1892 would continue to do so four years later. There is no way that any individual's voting pattern can be identified or predicted, but in a quasi-aggregate sense, it is possible to say that most of those voting in a particular way at one time continued to do so at a second point in time.

Given that the Republicans increased their vote from 1892 to 1896 and given that the qualitative literature argues that vote

defections were mostly or heavily on the part of the Democrats, we feel it is reasonable to require that most 1892 Republicans who voted in 1896 voted for McKinley. To be more specific we will assume that no more than 25% of the 1892 Northeastern and Midwestern Republicans would vote Democratic in 1896. All the narrative and qualitative literature on the politics of the 1890s stresses the cohesiveness and strength of the Republican party both absolutely and relative to the fractionated Democratic party. A 25% latitude for Republican defection is therefore both reasonable and liberal. On the other hand the Democrats are thought to have been in disarray, particularly in the Northeast. We feel therefore that more defection should characterize the 1892 Democratic voters. Too much defection would be unreasonable, but we feel that allowing 50% of the 1892 Democrats to defect would circumscribe what had been going on. More than that is tantamount to saying that there was no Democratic party, an assertion to which the aggregate data gives lie. Setting those limits does not in any way specify a particular solution. The limits merely identify reasonable bounds on what could have happened. Note that these limits are more generous than Burnham's definition of the conversion limits in a critical election which he sets at from one-fifth to one-third of the normal vote shifting (Burnham, 1970: 6). Those conditions are formalized as:

$$RR > .75(RR + RD) \quad (10)$$

$$DD > .50(DD + DR) \quad (11)$$

The final specification of our problem is meant to simplify computation. The center portion of Figure 1 lays out the marginals in thousands of votes for Minnesota. To allow quicker computation those vote figures are converted to percentages. Further analysis is then completed on the portions of the Table at the bottom. In particular, we will examine all non-negative integer solutions to the 3x3 table with the marginals expressed in percentage points that also satisfy conditions (1) through (11).

ANALYTIC APPROACH

Our substantive problem centers on the degree of mobilization and conversion of voters in the 1896 election. Since we want to compare cells RD and ER (conversion) with cells PD and PR (mobilization), what can be decided if there are numerous solutions?

A basic tenet or guiding principle of this research is that to analyze or estimate what did happen, we must examine what could have happened. We contend that valuable information can be garnered by enumerating all possible solutions and examining the relationships among cell entries in the solutions. This approach is simply identifying the logical consequences of the model

formalized in Figure 1. No new empirical information is generated. But since the logical entailments of the model are not apparent by just looking, for instance, at Figure 1, it is essential to work out those consequences. We used a computer to enumerate all solutions. Because of the inequality relationships, an analytic approach to the relations among cells is not possible. But the computer enumeration is fast and provides all the needed information.

Each solution represents a scenario of how 1892 voters could have behaved in the 1896 election, a configuration of how many party voters stood pat, defected, or dropped out. For most states there are literally a few thousand solutions. No one can really know which scenario actually occurred. But if the vast majority of them indicate more people were mobilized than converted, we deem it reasonable to believe that not only is conversion not to be assumed, but mobilization is more likely.

By examining all possible solutions or scenarios we are setting the groundwork for making some probabilistic statements. In this paper we will be making statements of the type: "In X% of all possible solutions, relationship Y exists." The approach we employ gives the researcher a specific measure of the certainty for each assertion. To say that 97% of all scenarios for state Z have more people mobilized than converted tells the researcher a specific number, a significance level. In conformity with standard usage for significance when more than 95% of

all solutions are congruent with "statement Y," we consider "statement Y" well supported -- significant at the .05 level as it were. Since social scientists accept probabilistic statements on a regular basis in the context of statistical significance, our approach, although novel, should not be uncomfortable. Happily, in a good number of Midwest and Northeastern states, 100% of the solutions exhibit consistency, which gives us certainty that "relationship Y exists."

DATA

Our general thesis is that mobilization as well as conversion characterizes the Republican ascendancy in 1896. But since regional variations are usually found in every presidential election, we will go below the national level to the state level for two pertinent regions to make our case. Because the South maintained its special relationship to the Democratic party, we will ignore it in the analysis. Because the West was small in electoral college votes and because some of its ties to the populist and bimetallic positions kept it in the Democratic camp, it too will be neglected in what follows.

The most interesting parts of the country in the 1896 election were the Midwest and the Northeast. Jensen, Burnham and others have focused on the Midwest as central to understanding

the election. Moreover, the size of the electoral college votes found in those two regions makes their study important. But perhaps most importantly, conversion of votes was supposed to be most salient in the Midwest and Northeast. Our study therefore analyzes the seven midwestern states of Minnesota, Wisconsin, Iowa, Illinois, Michigan, Indiana and Ohio. It includes the six New England states of Maine, New Hampshire, Vermont, Connecticut, Massachusetts, and Rhode Island. Lastly it takes in the three Mid-Atlantic states of New York, New Jersey, and Pennsylvania. Together they represent 227 electoral votes or 51% of the total.

For each of these 16 states, actual votes (rounded to thousands) for the Democratic and Republican presidential candidates in 1892 and 1896 were taken from the HISTORICAL STATISTICS OF THE UNITED STATES. That same source yielded Burnham's estimates of the voter turnout by state for each of the two elections. Those data were used, as outlined above, to generate the marginals for the 3x3 tables similar to that in Table 1. Using those marginals and conditions (1) through (11) we sought all possible solutions for each state.

FINDINGS

Our general goal is to convince the analyst that mobilization has to be considered more seriously and included more explicitly than in any studies to date. More proximately we will show that it is likely that in many states mobilization was very prevalent. Most immediately we will demonstrate that in some states mobilization was more prevalent than conversion.

To measure mobilization we have taken the sum of cells PD and PR to represent the segment of the 1892 electorate that came into the two party fray in 1896. These voters may be persons never voting before, persons who had voted Populist or other minor party in 1892, or persons who had dropped out from the voting booth in 1892 though they could have voted for one of the major parties prior to 1892. The extent of conversion is measured by the sum of cells DR and RD. These cells indicate the size of the defector population in the 1896 election. Our first pass was to search all possible solutions and identify those wherein $PR + PD > RD + DR$.

Table B displays various characteristics of the solutions to our problem by state. The first set of entries tally the proportion of solutions or voting scenarios in each state wherein more people are mobilized than are converted. Note that in all states but New Hampshire more than 84% of the solutions are mobilization dominant. In the Midwest in particular there is overwhelming

evidence that in four states (Minnesota, Wisconsin, Illinois, and Michigan) every possible voting scenario has more people mobilized than converted. In the remaining states in that region the data say that the probability exceeds 98% that the electoral dynamics were mobilization dominant. In the Northeast the data are not generally as strong but are still reasonably thought provoking. In Vermont it is certain that mobilization exceeded conversion. Save New Hampshire, the odds range from a low of 5 to 1 to a high of 10 to 1 that mobilization dominated. All this, of course, is based on the reasonable assumption that every solution or scenario is as likely as any other.

These data say that in 5 of 16 states we know incontrovertably that more mobilization than conversion occurred. In another 4, the odds are 50 to 1 that the extent of mobilization exceeded conversion. Mobilization, consequently, must be accorded more importance in the 1896 election than is found in the literature to date.

The second column of data in Table B counts the proportion of solutions or voting scenarios in which the mobilized go to the Democrats rather than the Republicans. In Wisconsin 96% of the scenarios had more mobilized voters going Republican than Democratic, but in the rest of the Midwest the data are not conclusive. In the New England states, on the other hand, in all but Massachusetts the odds are greater than 95% that the Republicans got the bulk of the newly mobilized voters. In Vermont we have certainty that that is the case.

CONTINUED

1 OF 2

The source of the victorious Republican vote is touched on in the third column of the table where the percentage of solutions in which more of the Republican vote came from the pool than from defecting Democrats is laid out. Only in Minnesota, Wisconsin, Vermont, and Pennsylvania are the data strong enough to say that Republicans got more support from the mobilized than the converted. In the rest of the states converted voters could well have and probably did play a major part in explaining the Republican success. But in those four states more than 95% of the scenarios have the mobilized more important than the converted, which is contrary to the conventional wisdom.

The last two columns in the Table indicate the vitality of the parties. Each shows the proportion of solutions (voting scenarios) in which more voters are mobilized to the party candidate than are lost by defection to the other party. With the exception of Indiana, Ohio, and Massachusetts it is clear that over 97% of all scenarios have the Republicans gaining more than they lose. In fact, in 8 of the 16 states in every possible scenario the Republicans gained more from the pool category than they lost from defection. This comports well with the emergence of the Republicans as the dominant party.

The case for Democratic party vitality is not as convincing. In four of the Midwestern states (Minnesota, Michigan, Indiana, and Ohio) all scenarios have fewer Democrats lost to the Republicans than are gained from the pool, indicating active and

reasonably successful state parties. The remaining Midwestern states do not display enough consistency in their set of solutions to say one way or the other. In the New England states, however, most of the states (excepting Massachusetts) show the Democrats losing more than they gain. What is interesting here is that the Northeast fits what the literature says about conversion, but some of the Midwest does not.

To this point, then, our enumeration and analysis of solutions indicates that in general mobilization must be counted as a serious component of the vote dynamic in 1896. In five states it is certain that mobilized voters outnumber converted voters while in another four states the weight of probability points in the same direction. Moreover we have presented evidence that although the Republican state parties seem full of vitality in both the Northeast and Midwest, the Democrats appear comparably strong only in the Midwest.

To understand mobilization's role more fully requires us to estimate its magnitude in cardinal rather than ordinal terms. Just how many people were mobilized? Or, lacking that, what is the smallest number mobilized? The answer is highly dependent on turnout. Table C displays the turnout in 1892, 1896, and the difference between those two. Clearly the turnout was higher in the Midwest than in the Northeast in absolute terms in 1896. Moreover the Midwest experienced an increase in turnout from 1892 while the Northeastern states, except Pennsylvania, all expe-

rienced a decline in turnout. Since the Republican party seems to have experienced continued vitality in that election, the only possible explanation for the turnout is that the Democrats failed to mobilize voters that were potentially theirs in the Northeast. It is very possible, then, that the criticality of the 1896 election comes not from inordinate numbers of people changing their party allegiance, but from large numbers of people who were not urged out by the Democratic party in some states while the Republicans were out cracking the whip to get to the polls those who were likely to be theirs.

In an effort to estimate the magnitude of the "new" voter, we proceed as follows. We seek to identify the minimum number of people in cells PD and PR , the cells representing the newly mobilized. Seeking the minimum of $PD+PR$ is tantamount to seeking the maximum value of PP since we know the marginal $P1$. But the maximum value of PP is equivalent to knowing the minimum values of DP and RP since we know the marginal $P2$. The minimum values of DP and RP , however, fall out directly from the attrition by death data which says that $PD+PR$ must minimally exceed 7.56% of $D1+R1$.

The fourth column of Table C lists the estimates of minimal new voters in thousands. The New England states have very small numbers of new voters compared to the comparable figures in the Midwest. New York and Pennsylvania are the only two Northeastern states which have a large number of newly mobilized voters. The

fifth column expresses the new voters as a percentage of the total Democratic and Republican voters in 1896. Once again, it is clear that in the Midwest generally about 25% of the electorate had not voted in 1892 or had voted for a third party. Since the third party vote in 1892 was small in comparison with the absolute numbers of the minimal estimates, we can be confident that most of the "new" voters were indeed new. (In the Midwest the Populist vote in 1892 as a percentage of Democratic and Republican votes ranged from a lonely extreme of 13% in Minnesota to a low of 1.8% in Ohio. For the region as a whole, the mean Populist vote was 4% of the Democratic and Republican vote.) In the Northeast the proportion of the voters that were mobilized is substantially lower. The reader must remember that these estimates of new voters are in no way dependent on conditions (10) and (11), i.e. defection outer limits are irrelevant in these estimates.

In the Midwest, then, mobilization was indeed an important force in the 1896 election. If a minimum of one quarter of the electorate is new to the polls, it is hard to justify the silence the literature showers on mobilization. Conversion is a more dramatic explanation than mobilization and so may have drawn greater attention. But we have raised questions about the importance of mobilization, especially in the Midwest.

Although we cannot argue with certainty that mobilization made the difference in the electoral outcome, we close this

section by noting that in the Midwest we cannot ignore the possibility. If the Republican plurality in 1896 is taken from Table A and compared with the minimum number of new voters from Table C, it is clear that the minimum number of new voters exceeds the Republican plurality in every Midwestern state. In Wisconsin the new voters only slightly outnumber the Republican plurality, but in the other six states new voters are two to five times more numerous than the plurality. Since we know relatively little about the voting preferences of the newly mobilized voter, all we can say is that mobilized voters had the potential for a substantial impact, a fact that must be acknowledged in the same breath as the one arguing for the impact of conversion.

SUMMARY AND CONCLUSIONS

This paper has shown that in the 1896 presidential election the role of new voters was important. While the literature has emphasized the conversion of 1892 Democratic voters to the Republican party, we have shown that in addition to the converted, the mobilized were very numerous.

In the Midwest mobilization was very common and in most states provided more voters than those lost by defection. In the

Northeast conversion presumably was still very important, although mobilization was not unknown. Part of the reason mobilization was low in the Northeast probably lies in the lack of effort on the part of the Democratic party except in Irish dominated cities (Jones, 1964: 347) and in the commonly felt opinion in both parties that the major battlefield was to be the Midwest. Midwestern Democratic party enthusiasm for Bryan got the new vote out, a voting group that outnumbered those defecting from the party. In the Northeast, however, suspicion of the firebrand Bryan was manifested by the relatively low turnout and the lack of evidence that mobilization was substantial.

In the Midwest, then, the critical nature of the election marking Republican ascendancy lies not so much in the shift of allegiance on the part of one-time Democrats but more from the mobilization of new voters. Political change in this case is not the wrenching, psychologically radical shift ascribed to most critical elections. Instead, much of the Republican support comes from the entrance of new electoral participants. We have shown, we believe, that major and enduring change can come through the incremental addition of new voters to the electoral arena.

We find the notion of Republican victory through mobilization in the Midwest attractive because of the notion that in that time political affiliation was near to a primary allegiance, one that would not be easy to change. We suggest that in general it

is far easier to bring in new voters whose political allegiance is perhaps unformed and certainly is relatively malleable than to change the allegiance of a voter confirmed in his or her belief through repeated voting acts.

FIGURE 1: Formulation of the Problem and an Illustration for Minnesota

1892	Dem		1896 Rep		Pool		
Democrat	I	II	I	DR	I	DP	I D1
Republican	I	RI	I	RR	I	RP	I R1
Pool	I	PI	I	PR	I	PP	I P1
	L2		R2		P2		

VOTES (in thousands)

1892	Dem		1896 Rep		Pool		
Democrat	I	LD	I	DR	I	DP	I 101
Republican	I	RD	I	RR	I	RP	I 123
Pool	I	PD	I	PR	I	PP	I 262
	140		194		152		486

PERCENTAGES

1892	Dem		1896 Rep		Pool		
Democrat	I	LD	I	DR	I	DP	I 21%
Republican	I	RD	I	RR	I	RP	I 25%
Pool	I	PD	I	PR	I	PP	I 54%
	29%		40%		31%		100%

Table A: Party Vote and Vote Change in 1892 and 1896
(in thousands of votes)

State	Democratic Vote			Republican Vote			% Increase In Vote
	1892	1896	1896 -1892	1892	1896	1896 -1892	
Minnesota	101	140	+39	123	194	+71	27.61%
Wisconsin	177	166	-11	171	268	+97	20.49
Illinois	426	465	+39	399	609	+210	24.49
Michigan	202	237	+35	223	293	+70	16.92
Indiana	263	306	+43	254	324	+70	15.40
Iowa	196	224	+28	220	289	+69	17.61
Ohio	405	477	+72	405	526	+121	19.04
Vermont	16	10	-6	38	51	+13	14.29
New Hampshire	42	21	-21	46	57	+11	-5.60
Connecticut	82	57	-25	77	110	+33	5.45
Maine	48	35	-13	63	80	+17	1.72
Massachusetts	177	106	-71	203	279	+76	2.81
Rhode Island	24	14	-10	27	37	+10	3.77
New Jersey	171	134	-37	156	221	+65	10.42
New York	655	551	-104	609	820	+211	6.51
Pennsylvania	452	427	-25	516	728	+212	19.04

Source: HISTORICAL STATISTICS, (1975: 1079).

Table B: Characteristics of Solutions Allowing Maximum of
25% Republican Defection and 50% Democratic
Defection from 1892 to 1896

State	Percentage of Solutions Wherein:				
	PR+PD>RD+DR	PD>PR	PR>DR	PR>RD	PD>DR
Minnesota	100.0%	19.2%	99.8%	100.0%	100.0%
Wisconsin	100.0	4.0	97.0	100.0	60.8
Illinois	100.0	31.7	76.5	100.0	87.2
Michigan	100.0	51.6	70.9	99.5	100.0
Indiana	98.7	62.0	54.2	83.1	100.0
Iowa	99.6	44.7	65.2	98.7	87.8
Ohio	99.1	60.1	55.9	85.2	100.0
Vermont	100.0	0.0	100.0	100.0	6.7
New Hampshire	49.9	3.8	49.6	96.0	0.0
Connecticut	85.1	4.3	76.6	100.0	9.5
Maine	90.2	2.0	88.4	100.0	17.3
Massachusetts	87.2	80.2	45.6	71.3	91.0
Rhode Island	84.4	0.2	84.2	100.0	3.2
New Jersey	85.2	14.9	66.9	99.8	13.8
New York	92.4	22.8	72.5	98.7	41.1
Pennsylvania	99.9	5.8	96.0	100.0	69.3

Table C: Turnout and Minimum New Vote by State

State	Turnout			Minimum New Vote	
	1892	1896	1896-1892	In 1000s	As % of 1896 2 Party Vote
Minnesota	66.6%	75.2%	+8.6%	127	38%
Wisconsin	76.8	84.9	+8.1	112	26
Illinois	86.0	95.7	+9.7	309	29
Michigan	73.2	95.3	+22.1	137	26
Indiana	89.0	95.1	+6.1	152	24
Iowa	88.5	96.1	+7.6	128	25
Ohio	86.2	95.5	+9.3	254	25
Vermont	60.4	67.5	+7.1	11	18
New Hampshire	85.8	78.1	-7.7	0	0
Connecticut	85.4	83.3	-2.1	20	12
Maine	63.5	63.0	-0.5	12	11
Massachusetts	74.6	70.6	-4.0	34	9
Rhode Island	63.0	59.2	-3.8	4	7
New Jersey	90.3	88.4	-1.9	53	15
New York	86.3	84.3	-2.0	203	15
Pennsylvania	75.7	81.8	+6.1	260	22

Source: HISTORICAL STATISTICS, (1975: 1071-1072).

REFERENCES

- Andersen, Kristi (1976). "Generation, Partisan Shift, and Realignment: A Glance Back to the New Deal." In Norman H. Nie, Sidney Verba, and John R. Petrocik (eds.), *THE CHANGING AMERICAN VOTER*. Cambridge, Mass.: Harvard University Press.
- Eurnham, Walter Dean (1970). *CRITICAL ELECTIONS AND THE MAINSPRINGS OF AMERICAN POLITICS*. New York: W. W. Norton and Company, Inc.
- Campbell, Angus (1971). "The Development of Party Identification." In Jerome M. Clubb and Howard W. Allen (eds.), *ELECTORAL CHANGE AND STABILITY*. New York: The Free Press.
- Carpenter, Niles (1969). *IMMIGRANTS AND THEIR CHILDREN*. New York: Arno Press and The New York Times.
- Jensen, Richard (1971). *THE WINNING OF THE MIDWEST*. Chicago: The University of Chicago Press.
- Jones, Stanley L. (1964). *THE PRESIDENTIAL ELECTION OF 1896*. Madison, Wisconsin: The University of Wisconsin Press.
- Key, V. C. Jr. (1959). "Secular Realignment and the Party System." *XXI, JOURNAL OF POLITICS*, 198-210.
- Key, V. C. Jr. (1971). "A Theory of Critical Elections." In Jerome M. Clubb and Howard W. Allen (eds.), *ELECTORAL CHANGE AND STABILITY*. New York: The Free Press.
- Pomper, Gerald (1971). "Classification of Presidential Elections." In Jerome M. Clubb and Howard W. Allen (eds.), *ELECTORAL CHANGE AND STABILITY*. New York: The Free Press.
- Seagull, Louis M. (1980). "Secular Realignment: The Concept and its Utility." In Bruce A. Campbell and Richard J. Trilling (eds.), *REALIGNMENT IN AMERICAN POLITICS*. Austin, Texas: University of Texas Press.
- Sellers, Charles (1971). "The Equilibrium Cycle in Two-Party Politics." In Jerome M. Clubb and Howard W. Allen (eds.), *ELECTORAL CHANGE AND STABILITY*. New York: The Free Press.
- Sundquist, James L. (1973). *DYNAMICS OF THE PARTY SYSTEM*. Washington, D.C.: The Brookings Institution.

U.S., Bureau of the Census (1975). HISTORICAL STATISTICS OF THE UNITED STATES, COLONIAL TIMES TO 1970, Bicentennial Edition. Washington, D.C.: Government Printing Office.

Wanat, John (1979). "The Application of a Non-Analytic, Most Possible Estimation Technique: The Relative Impact of Mobilization and Conversion of Votes in the New Deal," VI, POLITICAL METHODOLOGY, 357-374.

EVALUATION WITH SPARSE NOMINAL DATA:
The Case of Differential Compliance with the 55 mph Limit

John Wanat
Department of Political Science
University of Illinois at Chicago Circle
Box 4348
Chicago, Illinois 60680

This research was supported in part by LEAA grant 79-WI-AX-0059.
LEAA is not responsible for any analysis in this paper.

EVALUATION WITH SPARSE NOMINAL DATA:
The Case of Differential Compliance With the 55 mph Limit

Evaluating programs under a quasi-experimental design with scanty data is unfortunately not rare. Changes in policy often are followed by a call for assessing the policy's impact. Yet not enough foresight is usually employed to collect enough data before the policy change to make adequate comparisons possible. The analyst is therefore forced to use data collected for other purposes to estimate the impact of a policy.

One example of this situation is the attempt to identify differential compliance with the drop in speed limits to 55 mph in 1974. Panel studies were not planned. Recall data on possibly illegal activity is suspect. The policy intervention is long past. All that remains are routinely collected cross-sectional data on the proportion of the traffic exceeding various speeds. How can one use aggregate data collected just before and just after the drop in speed limit to infer changes in driver behavior?

Existing approaches to the ecological inference problem this case represents are generally not adequate to answering whether the high speed drivers comply more fully than those in the middle ranges, or whether the drop in average speed comes about because everyone reduced speed equally. These and other questions about compliance behavior have not been answered so far.

In this paper the problem will first be specified and formalized. Data sources will be discussed. An estimation technique based on examining all solutions to the 3x3 table representing the policy change impact in aggregate terms will then be described. Next, techniques will be developed to identify the bounds of certain kinds of driving behavior and to identify the probabilities of certain changes in driving behavior on Michigan freeways. Finally those techniques will be applied.

PROBLEM SPECIFICATION AND FORMALIZATION

The problem we attack is to estimate the proportion of people driving in a given speed range who shift into other speed ranges and to do this using only nominal level data collected prior to and just after the 1974 drop in speed limits to 55 mph. Display 1 shows the proportion of passenger cars traveling on Michigan freeways in 5 mile per hour intervals in both 1973 and 1974. The legislation clearly had an impact because the curve has shifted to the left. But exactly who shifted downward by how much is not known. A computerized bibliographic search failed to identify any literature on who complied or how driving habits were affected by speed law changes.

Since speed limits in 1973 were 70 mph on Michigan freeways and they dropped to 55 mph in 1974, those two speed limits form natural cut points. Display 2 in the upper left shows the 3x3 crosstabulation representing those who fell into the low speed range (0-55 mph), the middle range (56-70 mph), and the high speed range (71+ mph) in both 1973 and 1974. The row marginals indicate the percentage of 1973 drivers who drove in the three speed ranges while the column marginals indicate the percentages of those surveyed in the various categories in 1974.

The cells in the table represent differential compliance. Cell G, for example, indicates the percentage of those who in 1973 were exceeding 70 mph and hence violating the law but who came into compliance when the law mandated the 55 mph limit. Cell A, E, and J represent those who drove in the same speed ranges both before and after the speed limit change. Cell H corresponds to those who were speeders in 1973 but who dropped speed in 1974 though not enough to comply with the new law. We are interested in the cells and their relative sizes because they indicate, within the ranges specified, how various groups responded to the lowered speed limit.

(Although the Arab oil embargo occurred at about the same time as the change in speed limit, we made the assumption that the aggregate drop in speed is the result of the change in speed limit. We feel the embargo's impact is minimal in comparison to the law change. Some people reduced driving speeds because of

greater fuel economy, but that factor is negligible since in 1975 the distribution of speeds was virtually the same as in 1974 while the number of passenger miles driven was back up to the 1973 levels.)

DATA

As indicated, data on driving habits are not available for the 1973-74 period on the individual level. The best source of data is the annual surveys the Federal Highway Administration publishes. Each state surveys the distribution of driving speeds on various kinds of roads for various kinds of vehicles. The surveys are taken on level roads under good driving conditions during daylight hours.

This research assumes that the surveys represent with reasonable accuracy how people drive in a given year. It also assumes that two successive surveys are sampling the same driving population. These assumptions allow using the univariate distributions from two successive years as the marginals for a 3x3 table that represents the shifts among the driving categories. These data are not ideal, but they are the only ones available and, as we will show, they are sufficiently good to enable

researchers to approximate panel studies with repeated diachronic data and the estimation approach to be explained next.

RESEARCH APPROACH

To answer our questions about changes in driving behavior we must estimate the cells found in Display 2. The 3x3 table in Display 2 can be expressed by the following six equations representing the marginal constraints:

$$A + B + C = L1 \quad (1)$$

$$D + E + F = M1 \quad (2)$$

$$G + H + J = H1 \quad (3)$$

$$A + D + G = L2 \quad (4)$$

$$B + E + H = M2 \quad (5)$$

$$C + F + J = H2 \quad (6)$$

Since there are only 6 equations but 9 unknowns, there are an infinite number of solutions. Because the cell entries represent enumerated data, restricting the cells to non-negative integers is appropriate and will reduce the number of solutions from an infinite number to a potentially large but finite number. The number of solutions can be reasonably reduced even further because it is unlikely that large numbers of drivers increased their speed. Individual drivers might have done so, but we

assume that fewer drivers moved into a higher speed range than stayed in the same range. Therefore our search for solutions will be restricted to those for which equations (1) through (6) obtain as well as the following three inequalities.

$$A > B \quad (7)$$

$$B > C \quad (8)$$

$$E > F \quad (9)$$

Even though the potentially large number of solutions to our problem may have been substantially reduced by our characterization of the problem, many solutions remain. Attempting to choose one solution from among the many would be a fruitless task. This research, therefore, will not try to find the one solution or scenario of driver speed changes that actually occurred. Instead we proceed on the assumption that by examining all possible solutions (or scenarios, in terms of the substantive problem) certain patterns and limits common to all the solutions may be uncovered. Those limits and patterns should tell the researcher something about changes in driving habits.

This research was carried out by writing a computer program that enumerated every solution to the marginals of the 3x3 contingency table representing the speed change that satisfied conditions (1) through (9). For each solution every cell and particular groups of cells were examined to identify maxima and minima as well as the relationships among the cells and cell groups. Because of the inequality constraints and the fact that

it is not possible to analytically identify extreme values of the cell groups of interest, enumeration was the only possible approach. Because we wish to map the inequalities among the cells (e.g., is cell G always greater than Cell H?), enumeration is also called for.

Seeking extreme values of cells and cell groups follows in the tradition developed by Davis and Duncan (1953). They dealt with a different kind of problem, but they showed the utility of locating maxima and minima of cells when point estimates cannot be made with any confidence. My approach goes beyond that of Davis and Duncan and their intellectual successors (e.g., Goodman, 1959) because the enumeration allows statements like "cell H exceeds cell G in 97% of all solutions." Statements like that make possible generalizations about particular behaviors (e.g., of the 1973 traffic violators who dropped speed in 1974, fewer complied with the new law than merely dropped speed) with a confidence level specified (e.g., in 97% of all possible cases). By identifying limits and by noting the frequency of inequality relationships and patterns among cells and groups of cells, it will be possible in analysis of some 3x3 tables to generate information that is not obvious and which is illuminating.

FINDINGS

At the top left of Display 2 the 3x3 table for passenger car vehicles on Michigan Freeways for 1973 and 1974 is displayed. The top right lays out the 3x3 table with cells grouped to indicate drivers who were law abiders under both speed limits (AA), who were abiders but became violatos (AV), who were violators but became abiders (VA), and who speed law violators in both 1973 and 1974 (VV). Maxima, minima, and inequality relations among those four groups were tallied when all solutions were enumerated.

In the center portion of Display 2 the maximal and minimal values of the important cells and cell groupings are shown, while the bottom part of that figure lists the proportion of solutions in which inequalities among cells and cell groups obtain. These data present the kind of information needed to make some judgments about who changed their driving habits in what way. Because we cannot identify any one solution as representing what really happened, we cannot expect to make pinpoint statements. But these data at hand will allow us to state what happened within certain ranges and with particular likelihoods.

Consider first cells D, G, H, and cell group G and H. Those cells represent drivers who have slowed down into lower speed ranges. Cell D represents those in the 56 to 70 mph range in 1973 who dropped below 56 in 1974. Cell G indicates those exceeding 70 mph in 1973 who complied with the new limit in 1974.

Cell H stands for those exceeding 70 mph who dropped into the 56 to 70 range the next year. And cell group G and H represents the 1973 violators who drop below 70 mph after the speed limit shifted from 70 mph to 55 mph.

We are particularly interested in whether those in the high range differed from those in the mid-range in slowing down. The maximum of both cells D and G is 18, which says that at maximum of about 30% (18/60) of the midrange drivers comply with the new law while a maximum of about 50% (18/37) of the high range drivers comply. Since the minimum in both cells is 0, comparison of cells D and G only tells the analyst that more of the high range drivers than mid-range drivers in 1973 could have complied with the new law.

Although analysis of compliance is so far indeterminant, study of speed drops is more satisfying. Compare the extreme values of cell D (1973 law abiders who dropped speed) to those of cells G and H (1973 speeders who dropped speed). Remember that a maximum of 30% of the mid-range drivers dropped into the next lowest speed range. But Display 2 says that a minimum of 33/37 (89%) of those in the high range in 1973 dropped below 70 mph. In other words a minimum of 89% of those exceeding 70 mph dropped into a lower speed range, but a maximum of 30% of those in the 56 to 70 mph range slowed into a lower speed range. The new law apparently had a greater impact on those violating the old law than those obeying the old laws.

Although we know that a large proportion of the 1973 speed law violators reduced speed, it is unclear whether the drop in speed actually brought them into compliance. The analysis therefore now focuses on the relationship between cell G (those dropping to below 55 mph) and cell H (those dropping to the 56-70 mph range). Display 2 shows that a maximum of 48.6% (18/37) of high range drivers move to below 56 mph and also that a close minimum of 43.25 (16/37) moved to the 56 to 70 range. Unfortunately, there is some overlap. It seems reasonable to suspect, however, that more dropped speed into an illegal speed range than into the lower legal range.

To gain a better grasp of the likelihood of more 1973 speeders reducing to a still illegal speed than fully complying, more attention is paid to the relative size of cells G and H in the set of all solutions. As the bottom portion of Display 2 shows, in 97.6% of all the solutions more 71+ drivers moved into the 56-70 mph range than into the less than 55 mph range. Clearly the investigator here is not given certainty about what the drivers did, but if we assume, as we are forced to lacking any other evidence, that each solution or scenario of driving changes is as likely as any other, then the probability of stating that most high speed drivers reduced speed but did not comply with the new law is 97.6%. Since social science researchers typically accept a 95% confidence level in hypothesis testing, and since our probability exceeds that figure, I hold that the statement is worth accepting.

The second set of cell groupings categorized driving behavior according to whether drivers were in compliance or in violation of the speed limits in 1973 and 1974. The computer program, while enumerating all solutions, added the appropriate cells into the groups of interest, compared their magnitude, and tallied the maxima and minima. The range of the highest and lowest values clearly shows that the group of 1973 law abiders who became 1974 violators is the largest, ranging from minimum of 44% to a maximum of 62% of all drivers. The proportion of drivers who violated both speed limit laws could range from 19% to 37%. Those who abided by both laws could range from 1% to 19% and those who were violators of the 1973 limit but complied with the 55 mph limit could range from 0% to 18%. Analysis of the extreme values of the compliance categories shows that changing the law had the effect of reducing speeds but putting most law abiding drivers into violation of the new law. The other three categories have overlapping extreme points and so another form of analysis must be sought.

Solution by solution comparison of the four compliance groups yields the unequivocal evidence that abiders turned violators are the most numerous group, followed by violators staying violators, and the continual violators outnumber both violators turned abiders and permanent law abiders. While this inequality analysis yielded certainty because in 100% of the solutions the inequalities held, it lacks the numerical precision

the extrema analysis provided. Their joint use can, however, give the investigator a good idea of the changes in law abider and violator status before and after the new speed law was introduced.

So far our analysis has been restricted to a 3x3 table with cut points or category markers tied to the 1973 and 1974 speed limits on Michigan's freeways. The FHA data, however, are collected in 5 mph increments. Once analysis has been done for the law abider/violator cut points, we proceed to apply our enumeration of all solutions to finer breakdowns of the given data. Because the enumeration program is only feasible for 3x3 or smaller tables, our analysis uses the FHA data broken into all possible combinations of 3 categories from 55 through 75 mph in 5 mph increments. Those breakdowns are displayed in the range categories of Table 1.

Two cell comparisons are particularly interesting in all tables: cells D and E and cells G and H. Cells D and E are those who were in the midrange and who dropped speed into the low range (D) or who stayed in the same speed range (E). Cells G and H represent those high speed range drivers who reduced speed to the low range (G) or the mid range (H). Table 1 lays out the proportion of solutions in which cell D exceeds cell E as well as the proportion of solutions in which cell G exceeds cell H.

It is clear from Table 1 that in a reasonably large proportion of tables analyzed the percentage of solutions when cell E > cell D or when cell G > cell H more than 95% or less than 5% of the times is reasonably large. This means that some generalizations about speed change behavior are possible.

In the first column of Table 1 there are three entries of 0%. This means that in 100% of the solutions cell D > cell E or cell D = cell E under three sets of circumstances. In particular, when the 1973 driver was going 66 to 70 mph, 66 to 75 mph, or 61 to 75 mph, more drivers dropped below those ranges than stayed in them. In other words, for drivers exceeding 66 mph, more reduced speed than stayed in the same speed range. But for drivers going 56 to 75 mph or 56 to 70 mph, 100 % of all solutions have cell E > cell D. More drivers in those speed ranges stay in those ranges than reduce speed. Lower speed drivers, it appears, are less likely to reduce speed than those driving at a higher speed.

Considering all possible combinations of cell cut points also permits generating some information about how much speeds are reduced. The second column in Table 1 says that in 100% of the possible solutions, of those drivers exceeding 75 mph, more moved into the 0 to 70 mph range than dropped into the 71 to 75 mph range. Similarly for those exceeding 70 mph, in every possible scenario, more drivers reduced speed by more than 5 mph than dropped by less than 5 mph. In 98.3% of the solutions

representing drivers exceeding 65 mph, it is true that more drop more than 5 mph than less than 5 mph in speed. When the drivers are considered who were exceeding 60 mph, in all cases more dropped less than 5 mph than more than 5 mph. To generalize, high speed drivers slowed more than lower speed drivers.

Viewing the same phenomenon from a slightly different perspective occurs when the proportion of solutions wherein drivers moved into compliance with the 55 mph limit is examined. It is 100% sure that where drivers exceeded 60 mph in 1973, more reduced speed less than 5 mph than reduced speed more than that to comply with the 55 mph limit. Likewise it is certain that for drivers exceeding 65 mph, more dropped into the 56 to 65 range than complied with the new limit. It is 98.8% (100%-1.2%) sure that for those exceeding 70 mph in 1973, fewer complied than dropped up to 15 mph. Our analysis tells us nothing about those exceeding 75 mph in 1973.

Maxima and minima for speed reductions in various ranges are displayed in Table 2. From that display there is further corroboration of what we either had suggested or demonstrated with the likelihood technique. Enumeration of all solutions and recording the extreme values shows that larger proportions of drivers in the higher speed ranges drop speeds from 1973 to 1974 than do drivers in the relatively lower speed ranges.

CONCLUSIONS

This paper began by noting the inability of analysts to evaluate the impact of certain policy interventions because of sparse nominal level data. Using the premise that knowing the limits of what could have happened will assist the investigator to know what did happen, this research enumerated all possible solutions or, in real world terms, scenarios that represent changes in driving behavior before and after the imposition of the 55 mph limit in 1974 on Michigan freeways. That enumeration tallied extreme values of all cells and cell groups in the 3x3 table representing the speed changes from 1973 to 1974. It also sought the probabilities that various cells and cell groups exceeded other cells and cell groups.

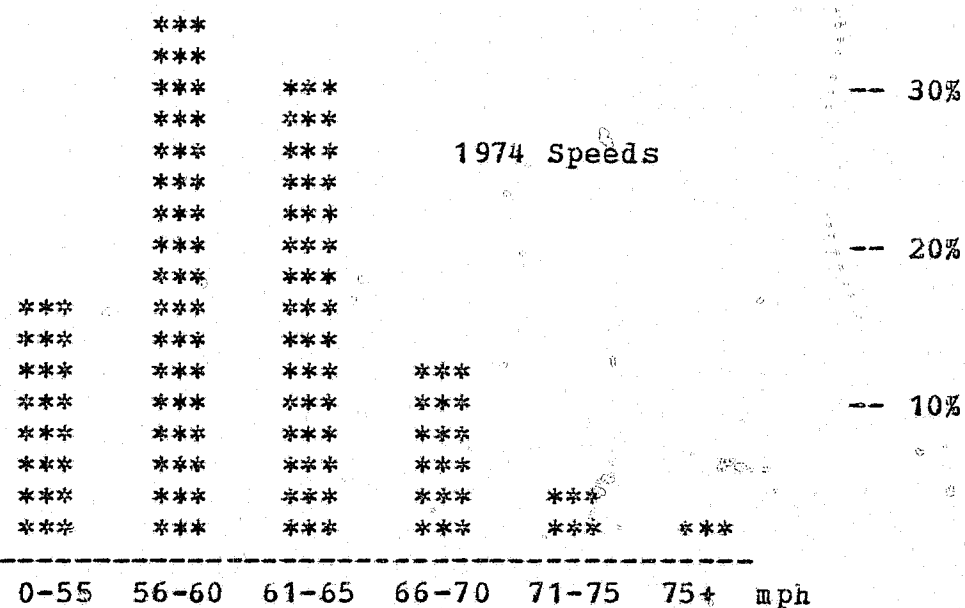
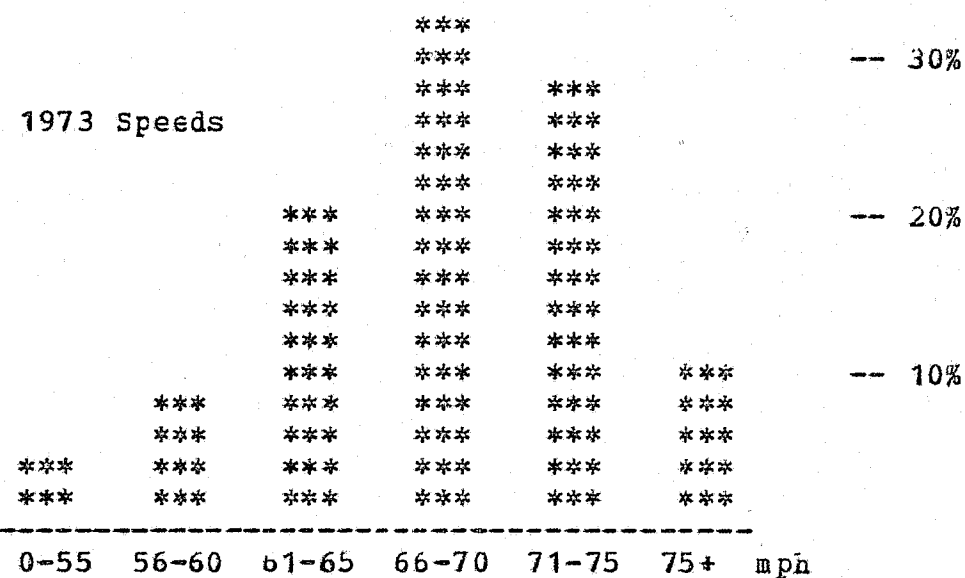
Substantively this investigation demonstrated that those 1973 drivers who travelled relatively fast modified their driving behavior more radically than those who were closer to the new 1974 speed limit. The imposition of the 55 mph limit brought about changes in driving behavior more dramatically on faster drivers than slower ones. But it showed that it also created a large group of law violators out of previously law abiding drivers.

While the technique developed here did not offer pinpoint estimates of changes in driving behavior, it did advance our knowledge of how various driving groups complied with a reduced speed limit in one state.

REFERENCES CITED

- Duncan, O. Dudley and Beverly Davis, (1953), "An Alternative to Ecological Correlation," 18 AMERICAN SOCIOLOGICAL REVIEW 665-666.
- Goodman, Leo A., (1959), "Some Alternatives to Ecological Correlation," 64 AMERICAN JOURNAL OF SOCIOLOGY 610-625.

Display 1: Passenger Car Speed Distributions on
Michigan Freeways in 1973 and 1974



Display 2: Formalization and Solution Limits to Passenger Car Speeding Behavior on Michigan Freeways from 1973 to 1974

1974										1973									
0-55			56-70			71+			1973										
0-55mph	I	A	I	B	I	C	I	L1	I	I								I	
	I						I		I	AA	I		AV				I		
56-70	I	D	I	E	I	F	I	M1	I	I							I		
	I		I		I		I		I		I						I		
71+	I	G	I	H	I	J	I	H1	I	VA	I		VV				I		
	L2			M2			H2												

Cell Display

Law Abider/Violator
Display

* * *

Extreme Values of Selected Cells and Cell Groupings

	Cells and Cell Groupings							
	D	G	H	G+H	AA	VA	AV	VV
Max:	18	18	37	37	19	18	62	37
Min:	0	0	16	34	1	0	44	19

* * * * *

Likelihood of Solutions Wherein:

Cell E > Cell D	=	100.0%
Cell H > Cell G	=	97.6
Cell D > Cell G	=	48.2
Cell H > Cell J	=	100.0

Table 1: Inequality Relationships Showing Likelihood
Of Reducing Speed to Other Speed Categories

Range Categories	Proportion of Solutions Wherein:	
	# midrange maintainers > # midrange reducers (Cell E > Cell D)	# high range reducing to low range > those reducing to mid- range (Cell G > Cell H)
0-70, 61-75, 76+	0.0%	100.0%
0-65, 66-75, 76+	0.0	59.7
0-60, 61-75, 76+	31.5	47.0
0-55, 56-75, 76+	100.0	47.2
0-65, 66-70, 71+	0.0	100.0
0-60, 61-70, 71+	39.4	49.9
0-55, 56-70, 71+	100.0	1.2
0-60, 61-65, 66+	63.6	98.3
0-55, 56-65, 66+	78.4	0.0
0-55, 56-60, 61+	60.0	0.0

Table 2: Ranges of Speed Reducing Behavior

Initial Range	Lower Range	% in Initial Range	Reducing to Lower Range		% Reducing to Lower Range	
			Max	Min	Max	Min
71-75	0-70	27	27	23	100.0%	85.2%
66-75	0-65	58	58	42	100.0	72.4
61-75	0-60	79	49	33	62.0	41.8
56-75	0-55	89	18	6	20.7	6.9
76+	0-75	10	10	9	100.0	90.0
71+	0-70	37	37	33	100.0	91.8
66+	0-65	68	67	52	98.5	76.5
61+	0-59	89	48	43	53.9	48.3

Illustrative FORTRAN Program Embodying the
Research Technique Used in the Research

```

00000010 C
00000020 C TALLY.ROAD TALLY.ROAD TALLY.ROAD TALLY.ROAD TALLY.ROAD TALLY.ROA
00000030 C
00000040 C
00000050 C ESTIMATES ARE GENERATED BY USING THE EXPECTATIONS
00000060 C BASED ON THE EMPIRICALLY DERIVED MOST POSSIBLE DISTRIBUTIONS
00000070 C OF CELL ENTRIES TO A 3X3 TABLE CONSTRAINED ONLY BY THE MARGINALS.
00000080 C EXTREME VALUES OF CELLS AND CELL GROUPS ARE DERIVED AS ARE
00000090 C THE PROPORTION OF SOLUTIONS IN WHICH EACH CELL AND CELL GROUP
00000100 C EXCEEDS ALL OTHERS.
00000110 C
00000120 C IMPLICIT INTEGER(8-Y)
00000130 C REAL FLOAT,SQRT
00000140 C DIMENSION N(9),MX(9),MN(9),AXP(9),AXP2(9),FRQ(100,9)
00000150 C DIMENSION T(9,9),ZT(9,9),ATITLE(18),ZM(9,9),ZSD(9,9)
00000160 C DIMENSION ZP(9),S(9,9),ZS(9,9)
00000170 C DIMENSION VMX(9),VMN(9)
00000180 C DIMENSION V(9),W(9,9),ZW(9,9)
00000190 C
00000200 C N HOLDS THE SOLUTIONS AS THEY ARE ENUMERATED
00000210 C V HOLDS GROUPS OF SOLUTION CELLS
00000220 C MN, MX, VMX, & VMN ARE MAX AND MIN VALUES OF N AND V
00000230 C FRQ TALLIES THE FREQUENCY WITH WHICH EACH CELL VALUE OCCURS
00000240 C
00000250 C ENTER THE RUN TITLE AND THE TIME 1 AND TIME 2 MARGINALS
00000260 C 11 CONTINUE
00000270 C WRITE(6,50)
00000280 C 50 FORMAT(// ' ENTER TITLE ON FIRST LINE AND TIME1 AND TIME2 MARGINAL
00000290 C &S ON SECOND IN FREE FORMAT')
00000300 C READ(5,51) (ATITLE(I),I=1,18)
00000310 C 51 FORMAT(' ',18A4)
00000320 C WRITE(6,51) (ATITLE(I),I=1,18)
00000330 C READ *,D1,E1,F1,D2,E2,F2
00000340 C
00000350 C
00000360 C INITIALIZE VARIABLES, GENERALLY TO 0, AND FIND MAX VALUES THAT
00000370 C CELLS 1, 3, 5, AND 9 CAN TAKE ON IN THE SEARCH FOR ALL SOLUTIONS.
00000380 C
00000390 C DMIN=MIN0(D1,D2)+1
00000400 C EMIN=MIN0(E1,E2)+1
00000410 C FMIN=MIN0(F1,F2)+1
00000420 C GMIN=MIN0(D1,F2)+1
00000430 C NSOLN=0
00000440 C AA=0.0
00000450 C BA=0
00000460 C AB=0.0
00000470 C NTRY=0
00000480 C DO 210 J=1,9
00000490 C AXP2(J)=0.0
00000500 C MN(J)=100
00000510 C ZP(J)=0.0
00000520 C V(J)=0
00000530 C VMX(J)=0
00000540 C VMN(J)=100
00000550 C MX(J)=0
00000560 C AXP(J)=0.0
00000570 C DO 220 I=1,9
00000580 C T(I,J)=0
00000590 C ZM(I,J)=0.
00000600 C ZSD(I,J)=0.
00000610 C ZT(I,J)=0.
00000620 C W(I,J)=0
00000630 C ZS(I,J)=0.0
00000640 C ZW(I,J)=0.0
00000650 C S(I,J)=0
00000660 C 220 CONTINUE

```

```

00000530 VMX(J)=0
00000540 VMN(J)=100
00000550 MX(J)=0
00000560 AXP(J)=0.0
00000570 DC 220 I=1.9
00000580 T(I,J)=0
00000590 ZM(I,J)=0.
00000600 ZSD(I,J)=0.
00000610 ZT(I,J)=0.
00000620 W(I,J)=0
00000630 ZS(I,J)=0.0
00000640 ZW(I,J)=0.0
00000650 S(I,J)=0
00000660 220 CONTINUE
00000670 DO 210 I=1.100
00000680 FRQ(I,J)=0
00000690 210 CONTINUE
00000700 C
00000710 C SEARCH THROUGH ALL VALUES THAT CELLS 1, 3, 5, & 9 CAN ASSUME TO
00000720 C IDENTIFY THOSE THAT MEET MARGINAL AND ANY OTHER REQUIREMENTS
00000730 C THE RESEARCHER PUTS ON THE SOLUTIONS.
00000740 C DO LOOP ENDED BY STATEMENT 300 WILL TALLY ALL THE CHARACTERISTICS
00000750 C OF THE SOLUTION SET.
00000760 C
00000770 DO 300 I1=1,DMIN
00000780 DO 300 I5=1,EMIN
00000790 DO 300 I9=1,FMIN
00000800 DO 300 I3=1,GMIN
00000810 N(1)=I1-1
00000820 N(5)=I5-1
00000830 N(9)=I9-1
00000840 N(3)=I3-1
00000850 NTRY=NTRY+1
00000860 N(2)=D1-N(1)-N(3)
00000870 IF(N(2).LT.0) GO TO 300
00000880 N(6)=F2-N(9)-N(3)
00000890 IF(N(6).LT.0) GO TO 300
00000900 N(4)=E1-N(5)-N(6)
00000910 IF(N(4).LT.0) GO TO 300
00000920 N(8)=E2-N(2)-N(5)
00000930 IF(N(8).LT.0) GO TO 300
00000940 N(7)=F1-N(8)-N(9)
00000950 IF(N(7).LT.0) GO TO 300
00000960 C
00000970 C NOW CREATE VARIABLES, V, THAT REPRESENT COMBINATIONS OF VARIOUS
00000980 C CELLS OF INTEREST TO THE SPEED REDUCTION PROBLEM.
00000990 C
00001000 V(1)=N(1)+N(4)
00001010 V(2)=N(2)+N(3)+N(5)+N(6)
00001020 V(3)=N(7)
00001030 V(4)=N(8)+N(9)
00001040 V(5)=N(4)+N(7)+N(8)
00001050 V(6)=N(2)+N(3)+N(6)
00001060 V(7)=N(1)+N(5)+N(9)
00001070 V(8)=N(7)+N(8)
00001080 V(9)=N(4)+N(8)
00001090 C
00001100 C CREATE VARIABLES ASKING IF % OF MIDRANGE DRIVERS EXCEED % OF
00001110 C HIGH RANGE DRIVERS REDUCE SPEED.
00001120 C
00001130 AA1=FLOAT(N(4))/FLOAT(E1)
00001140 AA2=FLOAT(N(7))/FLOAT(F1)
00001150 IF(AA1.GT.AA2) BA=BA+1
00001160 C
00001170 C TALLY WHETHER EACH CELL GROUP EXCEEDS ANY OF THE OTHERS AND
00001180 C PUT THE TALLY INTO MATRIX S.
00001190 C
00001200 DO 280 I=1.8
00001210 J1=I+1
00001220 DO 280 J=J1.9
00001230 IF(V(I).GT.V(J)) S(I,J)=S(I,J)+1
00001240 IF(V(I).LT.V(J)) S(J,I)=S(J,I)+1
00001250 280 CONTINUE
00001260 299 CONTINUE
00001270 C
00001280 C TALLY WHETHER INDIVIDUAL CELLS EXCEED THE OTHERS AND PUT THE TALLY
00001290 C INTO MATRIX T. ALSO PREPARE FOR COMPUTING THE MEANS AND STD DEVS
00001300 C OF THE FREQUENCY DISTRIBUTIONS OF THE CELL SOLUTION VALUES.
00001310 C
00001320 DO 400 I=1.8
00001330 J1=I+1
00001340 DO 400 J=J1.9
00001350 IF(N(I).GT.N(J)) T(I,J)=T(I,J)+1
00001360 ZM(I,J)=ZM(I,J)+FLOAT(N(I)-N(J))
00001370 IF(N(I).LT.N(J)) T(J,I)=T(J,I)+1
00001380 ZSD(I,J)=ZSD(I,J)+FLOAT((N(I)-N(J))*(N(I)-N(J)))
00001390 400 CONTINUE
00001400 C
00001410 C SEARCH FOR EXTREMA OF CELL GROUPS, V
00001420 C
00001430 DO 305 I=1.9
00001440 IF(V(I).GT.VMX(I)) VMX(I)=V(I)
00001450 IF(V(I).LT.VMN(I)) VMN(I)=V(I)
00001460 305 CONTINUE

```

```

00001190 C      DO 280 I=1,8
00001200      J1=I+1
00001210      DO 280 J=J1,9
00001220      IF(V(I).GT.V(J)) S(I,J)=S(I,J)+1
00001230      IF(V(I).LT.V(J)) S(J,I)=S(J,I)+1
00001240      280 CONTINUE
00001250      299 CONTINUE
00001260 C
00001270 C      TALLY WHETHER INDIVIDUAL CELLS EXCEED THE OTHERS AND PUT THE TALLY
00001280 C      INTO MATRIX T. ALSO PREPARE FOR COMPUTING THE MEANS AND STD DEVS
00001290 C      OF THE FREQUENCY DISTRIBUTIONS OF THE CELL SOLUTION VALUES.
00001300 C
00001310 C      DO 400 I=1,8
00001320      J1=I+1
00001330      DO 400 J=J1,9
00001340      IF(N(I).GT.N(J)) T(I,J)=T(I,J)+1
00001350      ZM(I,J)=ZM(I,J)+FLOAT(N(I)-N(J))
00001360      IF(N(I).LT.N(J)) T(J,I)=T(J,I)+1
00001370      ZSD(I,J)=ZSD(I,J)+FLOAT((N(I)-N(J))*N(I)-N(J))
00001380      400 CONTINUE
00001390 C
00001400 C      SEARCH FOR EXTREMA OF CELL GROUPS, V
00001410 C
00001420 C      DO 305 I=1,9
00001430      IF(V(I).GT.VMX(I)) VMX(I)=V(I)
00001440      IF(V(I).LT.VMN(I)) VMN(I)=V(I)
00001450      305 CONTINUE
00001460 C
00001470 C      SEARCH FOR EXTREMA OF CELL VALUES, N
00001480 C
00001490 C      DO 310 I=1,9
00001500      IF(N(I).GT.MX(I)) MX(I)=N(I)
00001510      IF(N(I).LT.MN(I)) MN(I)=N(I)
00001520      310 CONTINUE
00001530 C
00001540 C      TALLY THE FREQUENCY OF CELL VALUES AND NUMBER OF SOLUTIONS.
00001550 C
00001560 C      DO 330 L=1,9
00001570      AXP(L)=AXP(L)+FLOAT(N(L))
00001580      AXP2(L)=AXP2(L)+FLOAT(N(L)*N(L))
00001590      FRQ(N(L)+1,L)=FRQ(N(L)+1,L)+1
00001600      330 CONTINUE
00001610      NSOLN=NSOLN+1
00001620      300 CONTINUE
00001630 C
00001640 C      PRINT OUT TITLE AND EXTREME VALUES OF CELLS
00001650 C
00001660 C      WRITE(6,515) (ATITLE(I),I=1,18),D1,E1,F1,D2,E2,F2
00001670 C      515 FORMAT(/' ',18A4/12X,'T1 MARGINALS',6X,3I6/12X,'T2 MARGINALS',6X,3
00001680      I6/)
00001690      WRITE(6,520) (MX(K),K=1,9),(MN(L),L=1,9)
00001700      520 FORMAT(' ',10X,'MAX(I)='',3X,9I6/11X,'MIN(I)='',3X,9I6/)
00001710      IF(NSOLN.EQ.0) GO TO 2222
00001720 C
00001730 C      WRITE OUT THE DISTRIBUTION OF CELL SOLUTION VALUES
00001740 C
00001750 C      WRITE(6,800)
00001760 C      800 FORMAT(' ',5X,'CELL FREQUENCIES IN SOLUTION SETS '/5X,'VALUE',5X,'
00001770      &N(I) = '/')
00001780      RSUM=0
00001790      RSUM1=-1
00001800      DO 850 I=1,200
00001810      K=I-1
00001820      DO 851 JT=1,9
00001830      RSUM=RSUM+FRQ(I,JT)
00001840      851 CONTINUE
00001850      IF(RSUM1.EQ.RSUM) GO TO 850
00001860      WRITE(6,860) K,(FRQ(I,J),J=1,9)
00001870      860 FORMAT(5X,I5,12X,9I6)
00001880      RSUM1=RSUM
00001890      IF((9*NSOLN-RSUM).EQ.0) GO TO 855
00001900      850 CONTINUE
00001910      855 CONTINUE
00001920 C
00001930 C      WRITE(6,900) NSOLN,NTRY
00001940 C      900 FORMAT(10X,'# OF SOLUTIONS = ',I8,5X,'# OF TRIES = ',I8/)
00001950 C
00001960 C      ASOLN=FLOAT(NSOLN)
00001970 C
00001980 C
00001990 C
00002000 C      COMPUTE THE EXPECTED VALUES OF THE CELLS AND THEN COMPUTE THE
00002010 C      ESTIMATED TRANSITION PROBABILITIES OF GOING FROM ONE STATE
00002020 C

```

```

00001960 900 FORMAT(10X, '# OF SOLUTIONS = ', I8, '5X, '# OF TRIES = ', I8, '/')
00001970 C
00001980 C
00001990 ASOLN=FLOAT(NSOLN)
00002000 C
00002010 C COMPUTE THE EXPECTED VALUES OF THE CELLS AND THEN COMPUTE THE
00002020 C ESTIMATED TRANSITION PROBABILITIES OF GOING FROM ONE STATE
00002030 C TO ANOTHER.
00002040 DO 1800 J=1,9
00002050 AXP(J)=AXP(J)/ASOLN
00002060 AXP2(J)=SQRT(AXP2(J)/ASOLN-AXP(J)*AXP(J))
00002070 1800 CONTINUE
00002080 ZP(1)=AXP(1)/FLOAT(D1)
00002090 ZP(2)=AXP(2)/FLOAT(D1)
00002100 ZP(3)=AXP(3)/FLOAT(D1)
00002110 ZP(4)=AXP(4)/FLOAT(E1)
00002120 ZP(5)=AXP(5)/FLOAT(E1)
00002130 ZP(6)=AXP(6)/FLOAT(E1)
00002140 ZP(7)=AXP(7)/FLOAT(F1)
00002150 ZP(8)=AXP(8)/FLOAT(F1)
00002160 ZP(9)=AXP(9)/FLOAT(F1)
00002170 WRITE(6,1820) (AXP(J),J=1,9),(AXP2(K),K=1,9),(ZP(L),L=1,9)
00002180 1820 FORMAT(1X, ' EXPECTED VALUES BASED ON MOST POSSIBLE FREQUENCY DISTR
00002190 &IBUTIONS' /5X,9F7.2/1X, ' STD DEV ' /5X,9F7.2/1X, ' TRANSITION PROBABILI
00002200 &LITIES' /5X,9F7.4)
00002210 C
00002220 C COMPUTE PROPORTION OF CELLS THAT EXCEED OTHER CELLS AND RELATED
00002230 C STATISTICS.
00002240 C
00002250 IF(ASOLN.EQ.0.0) GO TO 1870
00002260 DO 1860 I=1,9
00002270 DO 1860 J=1,9
00002280 ZT(I,J)=FLOAT(T(I,J))/ASOLN
00002290 ZM(I,J)=ZM(I,J)/ASOLN
00002300 ZSD(I,J)=SQRT(ZSD(I,J)/ASOLN-ZM(I,J)*ZM(I,J))
00002310 1860 CONTINUE
00002320 C
00002330 C COMPUTE PROPORTION OF CELL GROUPS THAT EXCEED OTHER CELL GROUPS.
00002340 C
00002350 DO 1870 I=1,9
00002360 DO 1870 J=1,9
00002370 ZS(I,J)=FLOAT(S(I,J))/ASOLN
00002380 1870 CONTINUE
00002390 C
00002400 C PRINT OUT REMAINDER OF OUTPUT
00002410 C
00002420 WRITE(6,1830)
00002430 1830 FORMAT(/' PROPORTION OF SOLUTIONS WHEREIN GIVEN CELL EXCEEDS OTHER
00002440 &S: ')
00002450 DO 1850 I=1,9
00002460 WRITE(6,1840) (ZT(I,J),J=1,9)
00002470 1840 FORMAT(5X,9F8.3)
00002480 1850 CONTINUE
00002490 C
00002500 WRITE(6,1871)
00002510 1871 FORMAT(/' MEAN DIFFERENCES AND STD. DEV. BETWEEN CELLS: ')
00002520 DO 1880 I=1,9
00002530 WRITE(6,1882) (ZM(I,J),J=1,9),(ZSD(I,J),J=1,9)
00002540 1882 FORMAT(5X,9F8.3/7X,9F8.3)
00002550 1880 CONTINUE
00002560 C
00002570 C
00002580 C
00002590 WRITE(6,1930)
00002600 1930 FORMAT(/' PROPORTION OF SOLUTIONS WHEREIN GIVEN CELL GROUPINGS EXC
00002610 &EED OTHERS: ')
00002620 &' VARIABLE 1 = ABIDER STAYING ABIDER' /
00002630 &' VARIABLE 2 = ABIDER TURNED VIOLATOR' /
00002640 &' VARIABLE 3 = VIOLATOR TURNED ABIDER' /
00002650 &' VARIABLE 4 = VIOLATOR STAYING VIOLATOR' /
00002660 &' VARIABLE 5 = THOSE DECREASING SPEED' /
00002670 &' VARIABLE 6 = THOSE INCREASING SPEED' /
00002680 &' VARIABLE 7 = THOSE STAYING IN THE SAME SPEED RANGE' /
00002690 &' VARIABLE 8 = THOSE IN HIGH RANGE WHO REDUCE SPEED' /
00002700 &' VARIABLE 9 = THOSE DECREASING SPEED BY ONE CATEGORY' /
00002710 DO 1950 I=1,9
00002720 WRITE(6,1940) (ZS(I,J),J=1,9)
00002730 1940 FORMAT(5X,9F8.3)
00002740 1950 CONTINUE
00002750 C
00002760 WRITE(6,1970) (VMX(K),K=1,9),(VMN(L),L=1,9)
00002770 1970 FORMAT(/' .10X, ' MAX V(I)=', 3X,9I6/11X, ' MIN V(I)=', 3X,9I6/)
00002780 C
00002790 AA=FLOAT(BA)/ASOLN
00002800 WRITE(6,2000) AA
00002810 2000 FORMAT(' PROPORTION OF SOLUTIONS WHEREIN N(4)/E1 > N(7)/F1 = ', F8.
00002820 &3/)
00002830 C
00002840 2222 CONTINUE
00002850 WRITE(6,1700)
00002860 1700 FORMAT(/' IF YOU WANT TO RUN ANOTHER DATA SET, TYPE 1')
00002870 READ *,NOTHER
00002880 IF(NOTHER.EQ.1) GO TO 11
00002890 STOP
00002900 END

```

END