ERIC M. USLANER
University of Maryland-College Park

*Straight Lines and Straight Thinking: Can All of Those Econometricians Be Wrong?**

Lyons (1977) challenges my critique of the inappropriate uses of the per capita transformation (Uslaner, 1976). In rejoinder, I maintain that one cannot simply employ such measures without prior theoretical justification. In particular, Lyons' arguments do not adequately consider the large body of econometric/mathematical statistical literature on such transformations. The mathematical criticisms he makes are shown to be without foundation. In short, I stand by what I originally wrote in total (i.e., not discounted by any per capita factor).

Constructing a defense for per capita indices *in general* is only slightly less difficult than trying to raise a defense fund for Al Fatah in Queens. Yet, William Lyons (1977) has undertaken such a task in an attempt to counter some arguments I made in a recent issue of this journal (Uslaner, 1976). At first glance, I thought that he had simply overreacted to my criticism of the use of per capita indices when they are not theoretically justified. Others may interpret his article as a "useful corrective" to mine. Upon further consideration, however, it is clear to me (at least) that there are some fundamental disagreements in our positions, some of which have led Lyons astray in his attempt to salvage the per capita index as a simple algebraic transformation which need not distort the results of one's data analysis. Here, I shall concentrate on the differences between us and the misinterpretations of my article by Lyons (both substantively and mathematically). The simplest way in which to proceed is to outline briefly what I maintained in my original article.

"The Pitfalls of Per Capita" Revisited

The basic point I was stressing in my article, "The Pitfalls of Per Capita" is: Be careful. I maintained that much of aggregate data analysis in political

*I am grateful to my colleague Kathleen Peroff for helpful suggestions.

science has employed per capita measures of various policy, political, and socioeconomic variables without adequate concern for whether the theoretically relevant measures should be so transformed (Uslaner, 1976, pp. 125–126). I did not say "never use" such transformations, but rather only employ them with caution (Uslaner, 1976, pp. 131–132). Lyons does indeed note that I so argued, but his own recommendation is rather different: He suggests that one should *always* employ such a transformation "*unless* your 'original' variable is *not* a function of population" (Lyons, 1977, p. 181, emphasis in original). I cannot think of any situation in which the cautionary note Lyons adds would hold and the researcher would attempt a per capita transformation. What I was arguing against is the blind use of such per capita measures if they just happen to be in the data set you have acquired from someone else—or because you think that it is the "thing to do." (I have seen too many people spend countless hours in creating per capita indices only to realize that the raw data was what they really wanted. Why did they go to such trouble? Because the literature is filled with per capita measures, I am told!) Whether we need such measures is, I argued, a theoretical question, not a statistical one. However, once a theoretical mistake has been committed, then statistical problems will surely arise. The problem is commonly known as "specification error."

I began my argument with the rather explicit assumption that the correctly specified formulations in each of the three "cases" I considered involved using the untransformed data. This approach is sufficient, but not necessary to demonstrate my point, as two recently published critiques of similar problems in data analysis in political science (Hazlewood, 1976; Cortés and Przeworski, 1977)[1] clearly establish. If the correctly specified models involve the untransformed data, and the transformation being employed is nonlinear, then I maintain that we simply have no clear-cut way of demonstrating what the pattern of interrelationships in the transformed data will be (Uslaner, 1976, pp. 126, 131). Alas, this is not a terribly original contribution: I relied heavily upon similar arguments made by Madansky (1964) and Schuessler (1973, 1974) in presenting my case.

So, we cannot make any conclusive inferences about the overall results of our analyses, although we can state that whatever parameters we derive will be biased, inefficient and inconsistent. The first two statistical properties are

[1] Cortés and Przeworski (1977) is a revised version of the 1971 paper cited in Lyons (1977) and Uslaner (1976). Both this paper and the Hazlewood one (presented at the 1974 Annual Meeting of the International Studies Association) were written before mine, but I was aware only of the former at the time.

not helped by increasing the sample size arbitrarily. However, a reasonably alert observer should be able to make at least some educated guesses as to the nature of the biases in extreme situations. Thus, I proposed to investigate two such cases. In the first, the variables chosen for analysis would, in their untransformed values, be as close to being independent of each other as possible. In the second, we would try to get variables which are as strongly related to each other as possible (without regressing a variable upon itself). Thus, in the first instance, I generated eleven (not twelve, as Lyons states) simulated random variables, each normally distributed with identical means and variances. A factor analysis of the untransformed data, as well as the zero-order correlation matrix (cf. Uslaner, 1976, p. 127), demonstrated that I had succeeded in producing a very random situation indeed. When each case of the first ten (not eleven, as Lyons argues) variates was divided by its respective case on the eleventh, the new "transformed" variables produced a very coherent pattern of covariation. This is not surprising. If you know *a priori* that: (1) your original variables are uncorrelated; and (2) the transformed variables will yield biased, inefficient, and inconsistent estimates of the coefficients in a linear model, then one would expect some pattern of intercorrelation to emerge. As it happened, there was a very strong set of relationships in the "data set" I produced. This is not conclusive (I never claimed that it was), but it did indicate that it is *possible* to create order out of chaos.

Similarly, in the case of two original variables posited to be virtually perfectly related (state expenditures and state revenue), we would expect at least some reduction in the correlation coefficient (since you cannot improve upon unity). Indeed, the transformed data outperformed my expectations, indicating that it is *possible* to produce chaos out of order. Neither example made any significant progress in advancing social science theory, but that was not the purpose of my article.

Finally, I examined a case of "middle-range" bias, in which there was a fairly high correlation observed in another scholar's analysis, but where there also appeared to be no apparent reason for employing the per capita transformation. The study I reexamined was Sharkansky's analysis of budgetary incrementalism in the American states (1968). Since the nature of incremental change, as well as what we know of budgeting in general, implies that cross-state comparisons are not relevant to decision-makers at the state level, I saw no theoretical reason to employ per capita measures. While comparisons of correlation coefficients among different dependent variables are at a minimum very risky (see Rao and Miller, 1971, p. 16), the reanalysis did raise the coefficient of determination from .66 to .94 for the Sharkansky data set

(Uslaner, 1976, p. 130). It could have just as easily lowered the overall relationship if it were not for some identifiable factors (Uslaner, 1976, pp. 130–131). Lyons argues that these results are "inconclusive" (p. 178). That is precisely the point! You cannot always predict what will happen with nonlinear transformations. This would seem like an appropriate point at which to terminate the argument, but Lyons has made so many erroneous inferences which might lead well-meaning followers astray that I believe further discussion is warranted.

The Sources of Disagreement

Hopefully, the above recapitulation of what I originally maintained will clarify any remaining misconceptions of skeptical readers. In particular, the arguments I made follow in a long line of research on the topic by econometricians, mathematical statisticians, and even some sociologists and political scientists (see the references in Uslaner, 1976, pp. 132–133), only one of which is cited by Lyons–and, as it turns out, it is the one which has been refuted many times since. The reader may now say, "All right, he has made his point about what he did and did not say. What about the counterarguments of Lyons?". It is to these that I now turn.

Two minor points will demonstrate some of the more fundamental flaws in the Lyons critique. First, I am accused (Lyons, 1977, p. 177) of "misapplying" proofs. I not only did not present any proofs in my article, but also stated (Uslaner, 1976, p. 126) that *in general* ". . . no direct proof of the existence or the magnitude of the distortions arising from per capita transformations is available." Secondly, Lyons maintains (p. 178) that my factor analysis of random numbers would lead one to "infer . . . that any observed correlation between two per capita variables was spuriously inflated. . . ." My analysis of revenue-expenditure data for 1970 as well as the reanalysis of the Sharkansky data suggests just the contrary. It is indeed the case that this argument was made by Pearson (1897), but it has been since refuted by Neifeld (1927) and Madansky (1964).

It is in the section of his article, "Correct Specification of Per Capita," that the fundamental disagreements between us emerge. He states that I assume a "real meaning" (Lyons, 1977, p. 178) to the original variables. Putting aside the possibility that anything has some inherent "real meaning" to it, I do maintain that different specifications of a hypothesis can–and generally do–impose different sorts of order on a data set. For example, the preference of the "median" voter is a very different concept from the preference of the "modal" voter. The way in which we specify our variables has theoretical importance–and, not coincidentally, ramifications for the statistical techniques we employ. Even Lyons (1977, p. 179, emphasis added) comes close to admitting this when he describes regression analysis as "simply fitting a line through a *meaningful* set of points."

The critical question, however, is this: What happens when we employ the per capita transformation? Lyons maintains: Not much. My response is: We don't know with any reasonable degree of security. Of course, if my original statement of the transformation was wrong, then the entire logic of the "The Pitfalls of Per Capita" must be questioned. And, indeed, Lyons makes this argument. He defines algebraically (Lyons, 1977, p. 179) the per capita transformation as: (AB)/A, where A = the population of a state and B = the relative occurrence of a social phenomenon. In contrast, my definition of the transformation was simply B/A. Which is correct? Given the capacity for mathematical manipulation, the answer of course is: Both are correct. But, let us follow Lyons' logic to an Aristotelian conclusion and see where we wind up. First, let B = the proportion of people in a state who are doctors and assume that that quantity is equal to .01. Let there be 1000 residents of the state. Thus, by the Lyons formula, there will be a total of .01 (1000) = 10 doctors in the state, or 10/1000 = .01 doctors per capita. This answer is correct. But isn't is simpler to state that the proportion of doctors in a state is .01—the result you get either by starting with B or at the end of the computation. If the relative occurrence of a social phenomenon is *not* a proportion, then the Lyons formula is even less useful. Suppose that we want to compute per capita income. Then, what measure will be consistent with the Lyons formula? The only answer is: mean income. But mean income is the *same thing* as per capita income. Thus, the Lyons formula would require that we employ per capita income as one of the "elements" of a mathematical construct which yields per capita income!

It is not difficult to see why this is the case. The quantity (AB)/A simply reduces to B by rather straightforward algebra. The B term can only be a per capita term or a proportion which equals the per capita term. (Now, both transformations are subject to the same problems in data analysis. I did not specifically criticize the use of proportions in my original article because I did not believe they lead to the types of theoretical confusions wrought by per capita measures. Besides, I often use proportions in my own research—disagreeing with those who maintain that such transformations are invalid under any conditions whatsoever.) Even if we have a proportion, my example above also indicates that it can be interpreted as a per capita measure. But we are still left with the question of how to derive the original per capita measure. To derive such a formula, one does not *multiply* the numerator by anything (much less population); instead, one *sums* across the individuals corresponding to the case in aggregate analysis and then divides by the population figure. If one already has the summated data, as is typical in practice, then the appropriate per capita transformation is indeed B/A, where B is the aggregate measure and A is population.

Lyons next criticizes the specific examples I employ. First, he attacks my use of random numbers by noting that Pearson (1897) showed that random variables with an expected correlation of zero will have an expected correlation of .500 when the data are transformed into per capita measurements (Lyons, 1977, p. 180). In reply, I note that Pearson's approximation to the expected value of the correlation coefficient is not very accurate (Kuh and Meyer, 1955; Madansky, 1964). Specifically, Kunreuther (1966) maintained that Pearson incorrectly argued that higher order (i.e., greater than squared) terms can safely be ignored in computing the correlation. My "per capitized" set of random variates were indeed all positive (cf. Schuessler, 1974, p. 368): 33 of the 45 off-diagonal correlations are greater than *.600* and only four fall below Pearson's value of .500. The Lyons critique simply does not consider the literature since Pearson's original article.

Lyons next argues (p. 180) that my r^2 of .22 for per capita expenditures and per capita revenues "is simply too low." He then offers another example employing a similar design, but based upon a sample of 285 cities. The fact that his data anlysis is only marginally deflated by the per capita transformation is interesting. It only serves, however, to reinforce the point that I stressed in my article: There is no simple, catch-all rule of correspondence between raw data and per capita variables. My second extreme example showed that one *can* produce chaos out of order, not that one invariably does. In any event, the question remains: Why per capitize? What does this transformed data set tell us about the budgetary process—particularly in contrast to the raw data? In using per capita data, one is at least implicitly assuming that the decision-maker in at least one city employs the expenditure level of some other cities as a benchmark for one's own budget.[2] The blind

[2] On the problem of poorly specified variables, my 1970 data analysis in Uslaner (1976) employed total *state* expenditures and total *state* revenues, not state and local data as one might naively expect. How misleading was my original argument? To check this, I asked my research assistant, Phillip Rourk, a graduate student in economics, to read my original article and then to check the codebook for the data he thought I employed. The only short-cut provided was to inform him which ten decks (of a total of 134) the 1970 data of all types was on. Within five minutes, he had selected the appropriate variables—despite his total unfamiliarity with the literature in the field. Data on state and local combined expenditures and revenue were in the relevant decks, but he maintained that the original statement of the problem never led him to consider these variables. There was no help or other prodding involved.

use of such transformations can often lead one into trouble, as Hazlewood's critique (1976) of the Dimensionality of Nations study clearly demonstrates. Somehow, I find it strange for Lyons (1977, p. 177) to argue that employing operationalizations and concepts which are related to the theoretical concerns of the research is "deleterious to the research efforts of those studying aggregate data."

Summing Up

Even if one accepts the arguments I have presented, there is still a nagging question left to answer: What can be done about the unit of analysis problem? I suggested (Uslaner 1976, p. 132) that the researcher would be better off employing order-preserving transformations such as standard scores as original variables or beta weights (standardized regression coefficients). Now, these do not "solve the problem." We are not dealing with a situation that admits of a nice and easy "solution." It is not surprising that Lyons would find this advice unsatisfactory and seek other means of obtaining greater comparability across variables. His advice, of course, is to employ per capita measures in the first place and he sees the danger (Lyons, 1977, p. 181) that researchers following my advice will resort to "residualization"–i.e., regressing the original variable upon population and treating the residuals as new dependent variables independent of population size. I do not quite see how readers of my article will rush to commit this mistake, but in case any of them have done so, I caution: such a procedure, called "stagewise" regression analysis (Draper and Smith, 1966, pp. 173–176), will produce biased estimators of the regression coefficients unless all of the independent variables in the "residualized" regression are uncorrelated with population size.

Finally, it is somewhat interesting to note the source of Lyons' objection to residualization. He does not approve of this procedure because (Lyons, 1977, p. 181) "population size is, in itself, a component of much social theory." Yet, not too many pages earlier, he stated (p. 179): "We perform the per capita correction only when the unadjusted total occurrences for each unit are largely a function of the fact that the unit contains more people. Otherwise comparing raw numbers across states, cities, or countries would be useless. Cities with the most people also have the most Democrats, Republicans, diseases, doctors, etc. Higher amounts of these 'social phenomena' covary. Are we to build social theory from this co-occurrence? Surely not!"

I make no pretense to have achieved a "breakthrough" in the analysis of data in the social sciences in my article. Indeed, the arguments I made are not at all new, but have generally not been readily accessible to political scientists, particularly in a reasonably nontechnical form.[3] I merely hoped that my article would not only make the arguments more widely available, but would (by the illustrations employed, using data from the discipline) show how the problem can directly affect the researcher. A transformation is not something to be employed whenever the researcher is unsatisfied with his (her) results. Yet Lyons (1977, p. 182) ends his article with the statement, "Without per capita transformations everything would be related to everything." This immediately brings to mind the judgment of Charles O. Jones (1973: p. 34) on what we have learned from comparative state policy research: ". . . lots of things are related to lots of things, other things being equal."

Manuscript submitted 15 July 1976
Final manuscript received 17 September 1976

REFERENCES

Cortés, Fernando and Adam Przeworski. 1977. Comparing partial and ratio regression models. *Political Methodology*.

Draper, N. R. and H. Smith. 1966. *Applied regression analysis*. New York: John Wiley.

Hazlewood, Leo. 1976. An appraisal of the methodological and statistical practices used in the Dimensionality of Nations Project. In Francis Hoole and Dina A. Zinnes, eds., *Quantitative international politics: An appraisal*. New York: Praeger Special Studies.

Jones, Charles O. 1973. State and local policy analysis: A review of progress. In American Political Science Association, *Political Science and State and Local Government*. Washington: American Political Science Association, pp. 27–54.

Kuh, Edwin, and John R. Meyer. 1955. Correlation and regression estimates when the data are ratios. *Econometrica*, 23 (October 1955): 400–416.

Kunreuther, Howard. 1966. The use of the Pearsonian approximation in comparing deflated and undeflated regression estimates. *Econometrica*, 34 (January 1966): 232–234.

Lyons, William. 1977. Per capita index construction: A defense. *American Journal of Political Science*, 21 (February 1977): 177–182.

[3] The question of nontechnical discussions of data analytic problems has long plagued us. To this end, a new series has started publication: *Sage University Papers on Quantitative Applications in the Social Sciences.*

Madansky, Albert. 1964. Spurious correlation due to deflating variables. *Econometrica,* 32 (October 1964): 652–655.

Neifeld, M. R. 1927. A study of spurious correlation. *Journal of the American Statistical Association,* (September 1927): 331–338.

Pearson, Karl R. 1897. Mathematical contributions to the theory of evolution: On a form of spurious correlation when indices are used in the measurement of organs. *Proceedings of the Royal Society of London,* 60 (1897): 489–498.

Rao, Potluri and Roger Lee Miller. 1971. *Applied econometrics.* Belmont, Cal.: Wadsworth.

Schuessler, Karl. 1973. Ratio variables and path models. In Arthur S. Goldberger and Otis Dudley Duncan, eds., *Structural Equation models in the social sciences.* New York: Seminar Press, pp. 201–228.

——. 1974. Analysis of ratio variables: Opportunities and pitfalls. *American Journal of Sociology,* 80 (September 1974): 379–396.

Sharkansky, Ira. 1968. *Spending in the American states.* Chicago: Rand McNally.

Uslaner, Eric M. 1976. The pitfalls of per capita. *American Journal of Political Science,* 20 (February 1976): 125–133.