High Correlations or Multicollinearity, and What to Do About Either: Reply to Light*

ROBERT NASH PARKER, *Rutgers University*
M. DWAYNE SMITH, *Tulane University*

Stephen Light has raised an issue which is frequently and not always successfully confronted by many sociologists and others engaged in research involving the relationship between phenomena such as homicide and the social structure. We wish to express our appreciation to Light for his careful and thoughtful consideration of not only our work but the important issues involved in the problem of multicollinearity. Light has given us the opportunity for a fuller discussion of our views on this problem and how it has affected our study of regional homicide rates.

Light's examination of our intercorrelations among independent variables revealed, as it did to us (1980: 146, n5), a number of unusually high values, a finding that sometimes reflects multicollinearity. As multicollinearity can have serious negative impacts on the signs, sizes, and significances of estimated net regression parameters, Light is correct in pointing out the potentially damaging impact of these high intercorrelations on our results and, more importantly, on our evaluation of competing hypotheses concerning the predictors of homicide in the South and elsewhere. However, Light's assumption that high correlation is automatically equal to a multicollinearity problem, and his use of the Haitovsky test to confirm this equality, point to a conceptual issue concerning the nature of multicollinearity which is not generally understood. High correlation, in and of itself, does not necessarily mean that an ordinary least-squares analysis which includes such correlation will suffer from multicollinearity. Conversely, low correlation does not guarantee that a similar analysis will not suffer from multicollinearity. In order for multicollinearity to do its worst, two factors must be simultaneously present: first, one variable or several in combination must exhibit unusually high intercorrelations with another or several other variables in combination; second, those variables or combinations must share substantial covariance with the dependent

*Address correspondence to Robert Nash Parker, Department of Sociology, Rutgers University, New Brunswick, NJ 08903.

variable. This issue points to the major flaw with the Haitovsky test and others like it (e.g., Farrar and Glauber), which is that such tests do not take into account the pattern of relationships between the dependent variable and the variables suspected of being collinear. In addition, the Haitovsky test in particular has other problems, such as insensitivity and inferential misdirectedness (see Campbell and Parker), that make it unhelpful.

This point is clearly illustrated by an examination of the intercorrelations from our study (143). The four variables mentioned by Light as having unusually high intercorrelation, structural poverty, percent nonwhite, the Gini index of income inequality, and Non-South, have an average intercorrelation with primary homicide of .71; the average intercorrelation of these same four variables with non-primary homicide is .25. The Haitovsky test would require that the null hypothesis of singularity not be rejected in both cases, yet a key factor necessary for a multicollinearity problem is absent in the case of non-primary homicide.

The fact that multicollinearity does not appear to be a problem for non-primary homicide does not concern the impact of multicollinearity on the equation we originally presented to predict primary homicide (144). As we have argued for the rejection of typically utilized tests for multicollinearity, we are still left with the question of deciding whether we actually have multicollinearity, and, if we do, what to do about it. It is our contention (see also Campbell and Parker) that ridge regression (Hoerl and Kennard) provides a reasonably good answer to both questions. Ridge regression can be used as a diagnostic tool when one suspects multicollinearity, because coefficients behave in an identifiable and distinct manner in ridge regression. In addition, ridge regression can provide additional information concerning the impact of multicollinearity on results and therefore on inferences from results.

Table 1 illustrates these points. It presents a ridge regression analysis of our original equations for both types of homicide; for convenience of the reader, the left column presents a replication of the original OLS results.[1] We used a recently developed variant of ridge regression (Bulcock and Lee), which allows precise determination of the amount of bias to be introduced so as to achieve the most appropriate solution to multicollinearity. The results for primary homicide in Table 1 show that (1) multicollinearity does not affect the conclusion that Non-South has no independent net effect, (2) structural poverty remains a significant and important net predictor, and (3) although multicollinearity between poverty and percent nonwhite has some important effects, for example, poverty's effect is probably overestimated, and the effect of percent nonwhite is probably underestimated, our basic conclusion that structural poverty is a more stable and important predictor of primary homicide, net of these variables, remains intact. The ridge regression results we referred to in our paper (146) yielded identical conclusions. The results for non-primary homicide further illus-

Table 1. NORMALIZED RIDGE REGRESSION WITH BIASING PARAMETER (k) EQUAL TO .163

	Primary Homicide		Non-Primary Homicide	
	OLS	NRR	OLS	NRR
Structural poverty	.330* (.100)	.177* (.079)	.068 (.054)	.036 (.036)
Percent nonwhite	.029 (.051)	.070 (.043)	.012 (.028)	.018 (.020)
Percent urban	.017 (.015)	.012 (.018)	.023* (.008)	.019* (.008)
Age 20-34	.351* (.153)	.344 (.188)	.097 (.083)	.100 (.086)
Gini Index	-.017 (.012)	-.001 (,013)	-.004 (.007)	-.001 (.006)
Hospital beds	.006 (.006)	007 (.008)	.007 (.004)	.007 (.004)
Non-South	-1.069 (.667)	-.894 (.728)	.145 (.363)	.082 (.334)
R^2	.784	.622	.522	.406

*Unstandardized coefficient is twice its standard error, which is in parentheses.

trate the point made previously: multicollinearity does not affect the estimates in this equation, as neither the coefficients nor their standard errors demonstrate significant instabilities or fluctuations.

Conclusions

Although we share Light's concern about multicollinearity, and the need for more precise operationalization, measurement, and attention to issues such as the appropriate unit(s) of analysis for studies of homicide (see Parker and Loftin), we do not share Light's view that the problems of multicollinearity detract from the value of our typology of homicide. Regardless of the impact of multicollinearity on equations for primary homicide, variables such as poverty, racial composition, and region, long the

mainstays of research on homicide, have very little impact on non-primary homicide. The only variable that seems to have an impact on the latter type of homicide is percent urban, a proxy variable for a large number of potentially interesting effects. Others have begun to apply this typology to studies of related issues such as sentencing and the death penalty (Radelet), and have found it to be of import. We feel that our original conclusion can be made with even greater confidence: "The importance of qualitative distinctions in homicide has been shown to have implications for explanations of criminal homicide . . . The potentials of this distinction should be considered in the development and expansion of perspective regarding its causation and possible prevention" (145).

Note

1. The results presented in the left column of Table 1 are not an exact replication of the original results due to considerations such as different computer hardware and software (the current results were produced on a VAX 11/730 with version 9.1 of SPSS and version V of LISREL) and the fact that a matrix with three digits (as opposed to five or six digits) was used in the replication. The results are similar enough to allay concern, as are the ridge regression results when compared to those originally obtained.

References

Bulcock, J. W., and W. F. Lee. 1983. "Normalization Ridge Regression in Practice." *Sociological Methods and Research* 11(February):259–304.

Campbell, R. T., and R. N. Parker. 1983. "Substantive and Statistical Considerations in the Interpretation of Multiple Measures of SES." *Social Forces* 62(December):450–66.

Farrar, D. E., and R. R. Glauber. 1967. "Multicollinearity in Regression Analysis: The Problem Revisited." *Review of Economics and Statistics* 49(January):92–107.

Haitovsky, Y. 1969. "Multicollinearity in Regression Analysis: A Comment." *Review of Economics and Statistics* 51(April):486–9.

Hoerl, A. E.N and R. W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics 12(January):55–67.

Light, S. 1983. "Multicollinearity in a Study of Regional Variation in Homicide Rates: Comment on Smith and Parker." *Social Forces* 62(March):800–3.

Parker, R. N., and C. K. Loftin. 1983. "Poverty, Inequality, and Type of Homicide: A Reconsideration of Unexpected Results." Paper presented at the Annual Meeting of the American Sociological Association.

Radelet, M. L. 1981. "Racial Characteristics and the Imposition of the Death Penalty." *American Sociological Review* 46(December):918–27.

Smith, M. D., and R. N. Parker. 1980. "Type of Homicide and Variation in Regional Rates." *Social Forces* 59(September):136–47.