DEMOGRAPHY© Volume 23, Number 2 May 1986

A COMMENT ON GLENN FIREBAUGH'S "POPULATION DENSITY AND FERTILITY"

Eric Jensen
Department of Economics, College of William and Mary, Williamsburg, Virginia 23185

In a recent issue of this journal, Firebaugh (1982) presents a model in which increasing population density causes decreases in the crude birth rate (CBR). It is a somewhat puzzling theory because it is not clear how, within the structure of the model, villages with initially relatively low birth rates became more densely populated. If low fertility is consistently associated with high-density villages, one should expect eventual convergence of birth rates and population densities, other things constant. No evidence of the process of convergence is apparent in the eleven-year time span of Firebaugh's Indian data; in fact, there appears to be a slight divergence in CBRs over this time span (Firebaugh 1982: graph, 483). It is possible that eleven years is too short a time period to allow one to see any movement toward convergence. This is unlikely, since the mean CBR declined by more than 20 percent over this interval, and in any case, this possibility does not resolve the contradiction inherent in low-fertility villages having more dense initial positions.

I propose that the problem with Firebaugh's results is a statistical one. Firebaugh estimates a coefficient for density (with CBR as the dependent variable) that is significantly different from zero at the .01 significance level, using pooled data and an error-components framework. Other coefficients for density, not statistically significant, are estimated using a set of slightly different models and cross-section data. I argue below that neither the pooled nor the cross-sectional estimators are correct, that both suffer from omitted variable bias. The high level of significance attached to the error-components coefficient is artifactual, the result of underlying structural change that has been omitted from the model.

The error-components model is one of several generalized least squares (GLS) techniques for dealing with pooled data. Such techniques carry two major benefits with them: they are efficient relative to ordinary least squares (OLS), and they yield consistent estimators of variances. They do not, as Firebaugh claims (1982: 482), relieve omitted variable bias. The consistency of error-components coefficient estimators derives from the consistency of the OLS estimators. Therefore, omission of relevant explanatory variables will yield inconsistent estimators, whether or not an error-components model is used.

There are several variables that probably should have been included in the model. Age structure differences, as Firebaugh notes, will yield different CBRs for two populations in which age-specific behavior is identical. Mortality, through the physiological effects of shortened birth intervals, deliberate replacement of children who have died, and "hoarding" in response to parental perceptions of mortality odds in the village, is another important determinant of fertility left out of the Firebaugh model. Including migration rates in the model might have helped resolve the high CBR/low density paradox but would also complicate matters by making density an endogenous variable in a simultaneous equations system.

A final modeling problem is more difficult to solve than the problem above, stemming in part from the use of ratio variables and in part from estimation using an incomplete model. The problem is this: as population grows over time, density increases (with area held constant). Over time the CBR has been falling, in a decline attributed in the literature to a host of socioeconomic factors. This decline has taken the form of a denominator (population size) that increases faster than the numerator (births) does. The time trend variable included in Firebaugh's error-components regression can capture only the linear decline in the CBR. Thus, to the extent that changes in population density better correspond to changes in the CBR than does a linear time trend, the effect of outside-of-model changes in the CBR will be reflected in the coefficient of density. In other words, if both the CBR and population density (through migration, mortality, etc.) respond to a third, omitted variable, causal inferences of the effect of density on the CBR will be ill-founded.

Even if the true relationship, causal or not, between these two variables is weak, pooling the data will create spurious correlation between them. The CBR is the ratio of births to population, and density is the ratio of population to area (lagged one year in the pooled regression equation). Area can take only 22 values; alternatively, it changes only one time for every 12 changes in population size. Conceptually, it is as if we had estimated 22 different time series equations, one per village, and then taken a weighted average of the density coefficients from these equations. In each village-level equation, area is constant. Thus, the coefficient of density reflects the relationship between the within-village CBR and the inverse of the last period's population size. In any given village, because it is so small (between 500 and 5000 people), the variability of births is likely to be large relative to the number of births. Hence, the significance of such within-village relationships between CBR and population density is illusory, a statistical artifact based on the strong negative correlation of population size with its lagged inverse.

Firebaugh's results support this contention. The time trend variable is the least significant one in the pooled regression, and the density variable there has the largest standardized coefficient (Firebaugh 1982: table 3, 489). In contrast, in the cross-sectional regressions, population density is never significant at commonly used levels (Firebaugh 1982: table 5, 490). Therefore, in a cross-sectional data set in which the variability of land area has not been artificially reduced (by pooling) relative to the variability of population size and births, one sees no causal relationship between population density and CBR.[1]

CONCLUSION

Firebaugh has attempted to test an interesting, if counterintuitive, theory. I have argued that his empirical model is incompletely specified, and that one should reserve judgement on his hypothesis until a better means of testing it can be developed. In particular, the use of ratio variables and pooled data create substantial statistical problems.[2] The simultaneous responses of the CBR and village population size to outside-of-model variation yield an inflated estimate of the effect of population density on fertility.

NOTES

[1] Caution should be used in interpreting these *t*-ratios, since they are based on biased estimators.

[2] Firebaugh's note 10 (Firebaugh 1982: 493) is incorrect. If the occurrence of event *A* is inversely correlated with something, the occurrence of "not *A*" must be *positively* correlated with that same thing, *not* inversely correlated. Thus, the "debunking" of "the notion of the inherent negative relationship between density and birth rate" is unconvincing.

REFERENCES

Firebaugh, G. 1982 Population density and fertility in 22 Indian villages. *Demography* 19:481–494.