DISCUSSION

Martin Frankel, University of Michigan

As the authors discovered, work in the area of optimal sample allocation for the estimation of what we might call "analytical" parameters is sorely needed. This lack of theory is compounded by the large gaps between the literature of experimental design and the literature of survey sampling. The experimental design and econometric literature usually assumes that we are taking simple random (with replacement) samples from hypothetical superpopulations. However, when the population of interest is widespread (e.g., families in the U.S.) practical economic probability sample designs are usually stratified and use clusters of sample elements. When regression models are used clustered samples usually induce positive covariance between error terms. Thus additional parametrization of the model would be required.

In the Graduate Work Incentive Experiment the formal population was a small area in New Jersey, and a systematic element sample was used. However, if it is deemed desirable to be able to make formal statistical inference to the entire U.S., a clustered sample would probably be used.

Independent of the type of probability sampling used (either simple random or clustered) there are two assumptions of the stated regression model which are probably always violated to some extent.

The violation of the assumption of homoscedasticity (equal error variances) is probably the less serious, resulting simply in a non-optimal allocation. However, assuming a known parametric form $y = bx$ may lead to serious problems when certain very disproportionate allocation schemes are used.

We are all aware that given enough real multivariate data we can usually reject any specific parametric hypothesis concerning the functional distribution of conditional means. This fact does not deter us from assuming a given form for the regression equation when we feel that the specific violations in various subranges of the design values will tend to average out over the entire range of interest. However, whenever most of the sample is allocated to a specific subrange of design values, the specific anomaly found in that subrange will be overrepresented in our estimate.

Thus in situations where the optimal allocation is extremely disproportionate, we might want to proceed with caution before we actually allocate the real sample.