number of clusters. Note that in the case where there is only one household per cluster, (22) is the standard formula for the heteroskedasticity-consistent variance covariance matrix. Note too that (22) does not require that there be homoskedasticity, either within clusters or between them, nor that there be a common intracluster correlation coefficient. It is therefore robust against quite general forms of intracluster correlations. The equation is implemented in the software package STATA as part of the huber command, and the corresponding procedure for panel data is described by Arellano (1987).

How much does all this matter? The answer seems to be a great deal, certainly more than is the case for the more familiar heteroskedasticity correction. In many applications, the correction is not much less than the design effect, and in my own work, I have frequently found that the usual formulas give standard errors that are understated by a factor of two to three, a much more serious matter than the 30 percent that seems to be common for the heteroskedasticity correction. The problem is exacerbated by the fact that in so many development applications, the explanatory variables are constant within the clusters, the wage, price, and access variables listed above. It would be invidious to list papers that use clustered data without correction, although [see Deaton (1988, 1990a)} for two selected examples, but there are hundreds of papers in development economics looking at labor market questions, at the demand for commodities and nutrition as a function of prices, and at access to education and health services where the true significance levels for t-values should probably be closer to 6 than to 2. Many of these studies will have to be redone, and I suspect that there will have to be a good deal of revision of conclusions. Of course, these problems are not confined to studies of economic development, and similar considerations apply for example, in labor economics. Indeed, Moulton (1990) has provided a particularly dramatic example using American state level data, where a small intrastate correlation coefficient is combined with large numbers of observations in each state to yield a design effect of nearly 10.

2.1.3. Quantile regressions

The method of quantile regression is not one that has been much used in economics to date, perhaps because of computational considerations. These have now been solved the qreg command in STATA is an example — so that this extremely useful tool is readily available without the need for special coding. The basic idea was first introduced into economics by Koenker and Bassett (1978) and can be described as follows.

Quantile regression, like linear regression, is concerned with the distribution of a scalar random variable y conditional on a vector of covariates x. In linear regression, it is assumed that one characteristic of this distribution, its mean, is