once standard in econometric texts — [see for example Cramer (1969, pp. 142-143)]— but even so, many regressions using survey data are run in weighted form.

In cases where heterogeneity is suspected, there are several useful strategies. When there are only a few strata rural versus urban would be the most frequent it clearly makes sense to run separate regressions, and to use covariance analysis where the homogeneity hypothesis is of separate interest. When the number of strata is large, with relatively few observations in each, random coefficient specifications would seem more useful, and, as a result, analysts should routinely expect heteroskedasticity in OLS regressions. Standard heteroskedasticity tests can be used, for example that given by Breusch and Pagan (1979), which in this case would involve regressing squared residuals on dummy variables for each stratum and comparing half the resulting explained sum of squares with a χ^2 with degrees of freedom equal to the number of strata. Heteroskedastic consistent variance covariance matrices should also be routinely used, see Section 2.1 below.

I should conclude by noting that there is a school of thought that does not accept the argument against weighted regressions, Kish and Frankel (1974) being perhaps the most eloquent example. They argue that the stratification in many surveys is not of substantive interest in its own right, and that the parameters of a hypothetical census regression are indeed of interest. Others, such as Pfefferman and Smith (1985) take a view similar to that here, arguing (among other things) that a complete population is of no great interest since it is only one of the many possible populations with which we might have been confronted.

1.1.8. Estimation and other design features: clustering

Even if the regression coefficients are homogeneous across strata, standard formulae for standard errors may be incorrect depending on the survey design. Two-stage sampling will induce non-independence between households in the same cluster if households who live in the same village are subject to common unobservables, such as weather, tastes, or prices. Under such circumstances, whether we are estimating means or regressions, standard formulae for variances are incorrect and can be seriously misleading.

Consider first the straightforward use of survey data to estimate a mean. Given a set of n observations *x,* standard procedures call for the estimation of the mean and variance according to

$$\hat{\mu} = n^{-1}\sum_{i} x_i, \quad \hat{\sigma}^2 = (n-1)^{-1}\sum_{i}^{n} (x \tag{12}$$