not so, the fact is in itself important, since it implies that robust inferences are not possible, and that the assumptions of the investigator are as necessary as the data for drawing the conclusions.

Even for standard and well-understood techniques, such as linear regression, inferences can be made more robust, either by moving away from OLS to alternatives such as quantile regression, or Less radically, by calculating standard errors in ways that are robust against the failures of standard assumptions that are common in survey data. I begin this section with these topics.

2.1.1. Heteroskedasticity and linear regression

As is well-known, the presence of heteroskedasticity in linear regression affects neither the unbiasedness nor the consistency of OLS estimation. However, the assumptions of the Gauss-Markov theorem are violated, so that OLS is no longer efficient, and the usual formula for the variance-covariance matrix of the parameter estimates is no longer valid. In particular, if the regression model is, for i = 1, . . n,

$$y, = x,p + ,; E(u,)\ 0; E(1e)= d_i > 0\ , \qquad (20)$$

and the OLS estimator is, as usual, $(X'X)^{-1}X'y$, then the variance-covariance matrix is given by

$$V= (X^1)0\ I'DX(X'X)^{-i} \qquad (21)$$

where D is an n X n diagonal matrix whose diagonal is the d's from (20). Although V in (21) cannot be evaluated without knowledge of the d's, it has been shown by Eicker (1967), Huber (1967), Fuller (1975) and White (1980), that it can be consistently estimated by replacing D by the diagonal matrix whose elements are the squared OLS residuals. Note that the consistency here is of the matrix V, not of D, the number of elements in which increases with the sample size, and which therefore cannot be consistently estimated. Following White and MacKinnon (1985), this relatively straightforward calculation can be modified and extended in a number of ways, some of which are likely to yield improvements in performance. These methods yield estimates of the variance covariance matrix that are asymptotically valid, and do not require the user to know or to specify the specific form of the heteroskedasticity in (20).

As I argued in Section 1, the stratification of surveys is likely to generate heteroskedasticity, and even without it, experience suggests that residuals are more often heteroskedastic than not. There are a number of tests for heteroskedasticity, of which perhaps the most convenient is that suggested by