Econometrica, Vol. 50 No. 1 (January, 1982)

ROBUST TESTS FOR HETEROSCEDASTICITY BASED ON REGRESSION QUANTILES

BY ROGER KOENKER AND GILBERT BASSETT JR.[1]

A new class of tests for heteroscedasticity in linear models based on the regression quantile statistics of Koenker and Bassett [17] is introduced. In contrast to classical methods based on least-squares residuals, the new tests are robust to departures from Gaussian hypotheses on the underlying error process of the model.

1. INTRODUCTION

THE CLASSICAL THEORY of linear statistical models is fundamentally a theory of conditional expectations.[2] In their influential text Mosteller and Tukey [19] note that,

> "What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of x's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions."

Means and other measures of central tendency rightfully occupy a distinguished place in the theory and practice of statistical data analysis. But we are entitled to ask, "Does the conditional expectation or any other measure of conditional central tendency adequately characterize a statistical relationship among variables?" An affirmative answer seems possible only within the confines of extremely restrictive parametric models. In principle, we would like to know the entire conditional distribution function, or equivalently, but perhaps preferably (see Parzen [20]), the conditional quantile function.

In this paper we suggest a natural approach to the estimation of the conditional quantile function based on the analogues of the sample quantiles for linear models introduced in Koenker and Bassett [17]. The asymptotic theory of these regression quantiles is extended to linear models with a family of linear scale processes. The problem of estimating the precision of these "regression quantile" estimates is addressed, and a new robust approach to problems of testing homoscedasticity is developed. Several examples of the proposed techniques are discussed in a final section. The emphasis here is rather different than our

[1] The authors wish to thank Robert Hogg and Colin Mallows for helpful comments on a previous draft. We are also deeply indebted to an anonymous referee whose "rough calculations" on an example stimulated an extensive revision of Sections 3 and 4. Complete responsibility for remaining errors and malapropisms rests with the authors.

[2] An exposition of least squares theory from this point of view may be found in a valuable monograph by Goldberger [14].

previous work on regression quantiles which emphasized robustness considerations within linear models with iid errors.

2. NOTATION AND PRELIMINARIES

Given a random variable Y with right continuous distribution function $F(y) = \Pr[Y \leqq y]$, it is useful to define the *quantile function*, on $u \in [0, 1]$,

$$Q_Y(u) = F^{-1}(u) = \inf\{y \mid F(y) \geqq u\}. \tag{2.1}$$

Similarly, given a random sample $Y_1, Y_2, \ldots, Y_n$, with empirical distribution function, $\hat{F}(y) = (1/n)\#\{Y_i \leqq y\}$ we may define the *empirical quantile function*,

$$\hat{Q}_Y(u) = \hat{F}^{-1}(u) = \inf\{y \mid \hat{F}(y) \geqq u\}. \tag{2.2}$$

Equivalently, the empirical quantile function may be defined in terms of solutions to a simple optimization problem. Explicitly,

$$\hat{Q}_Y(u) = \inf\left\{y \mid \sum_{i=1}^{n} \rho_u(Y_i - y) = \min!\right\} \tag{2.3}$$

where ρ is the "check function,"

$$\rho_u(z) = \begin{cases} u|z|, & z \geqq 0, \\ (1-u)|z| & \text{for } z < 0. \end{cases} \tag{2.4}$$

The latter definition suggests a natural analogue of the empirical quantile function for the linear model.

Consider the random variable Y having conditional distribution function $F(y \mid x)$ depending upon a row vector of exogenous variables x. We will assume henceforth that the first component of x is an intercept so that $x_{i1} = 1$ for all i. Suppose that the *conditional quantile function* of Y,

$$Q_Y(u \mid x) = \inf\{y \mid F(y \mid x) \geqq u\} \tag{2.5}$$

is a linear function of x, that is,

$$Q_Y(u \mid x) = \sum_{k=1}^{K} x_k \beta_k(u) = x\beta(u) \tag{2.6}$$

for some vector of constants $\beta(u)$ depending upon u. Now by analogy with (2.3) we are led to define an *empirical conditional quantile function*, for the sample $Y_1, \ldots, Y_n$,

$$\hat{Q}_Y(u \mid x) = \inf\left\{x\hat{\beta}(u) \mid \sum_{i=1}^{n} \rho_u(Y_i - x_i\hat{\beta}(u)) = \min!\right\}. \tag{2.7}$$

In Koenker and Bassett [17] we have studied in some detail the behavior of the statistics $\hat{\beta}(u)$ which we call "regression quantiles." l_1 regression is an important (median) special case with $u = 1/2$ (see Bassett and Koenker [4]). In previous work we have emphasized linear models with independent and identically distributed errors. Then the conditional quantile function may be written as

$$Q_Y(u \mid x) = x\beta + Q_\varepsilon(u) \tag{2.8}$$

where β is some fixed vector of parameters and $Q_\varepsilon(u)$ is the quantile function of the error distribution. In this case,

$$\beta(u) = \beta + (Q_\varepsilon(u), 0, \ldots, 0)'. \tag{2.9}$$

The exogenous variables influence only the *location* of the conditional distribution $F(y \mid x)$. Note that due to the intercept we are free to choose the first element of β so that $Q_\varepsilon(\frac{1}{2}) = 0$. In the iid case, the conditional quantile functions constitute a family of parallel hyperplanes.

Departing from the assumption of identically distributed errors, we consider the following rather general model of systematic heteroscedasticity,

$$Y = \mu(x) + \sigma(x)\varepsilon \tag{2.10}$$

where $\mu(x)$ may be thought of as the conditional mean of the regression process, $\sigma(x)$ as the conditional scale, and ε as an error term independent of x from a distribution with quantile function $Q_\varepsilon(u)$. The conditional quantile functions of Y are then simply,

$$Q_Y(u \mid x) = \mu(x) + \sigma(x)\, Q_\varepsilon(u). \tag{2.11}$$

We will assume that both μ and σ are linear functions of x, and we will write (2.11) as

$$Q_Y(u \mid x) = x\beta + (1 + x\gamma)\, Q_\varepsilon(u) \tag{2.12}$$

for some $(\beta, \gamma) \in \mathbf{R}^{2K}$. We are free, of course, to set some elements of β and/or γ to zero *a priori* so the fact that μ and σ depend upon the same vector of exogenous variables reflects only notational convenience.

This linear scale model of heteroscedasticity (2.12) is an important special case of the general class of models with linear conditional quantile functions. It subsumes many models of systematic heteroscedasticity which have appeared in the econometrics literature. The iid error case is simply $\gamma = \sigma e_1 = \sigma(1, 0, \ldots, 0)'$ for $\sigma \geqq 0$. The model of Goldfeld and Quandt [14] with $\sigma(x) = \sigma x_k$ makes $\gamma = \sigma e_k$. Anscombe's [1] model in which scale is linearly related to the mean of the regression process is also a special case with $\gamma = \sigma e_1 + \lambda\beta$. The so-called multiplicative heteroscedasticity model recently studied by Harvey [16] and Godfrey [13] which sets

$$\sigma(x) = \exp(x\gamma) \tag{2.13}$$

can be approximated for small $\|\gamma\|$ by the linear expansion

$$\sigma(x) = 1 + x\gamma + o(\|\gamma\|) \tag{2.14}$$

which takes the form (2.12). We will see presently that calculations of asymptotic power restrict us to local alternative models in which we consider sequences of alternatives $\{\gamma_n\}$ such that $\|\gamma_n\| = O(1/\sqrt{n})$. For such sequences the multiplicative heteroscedasticity model is well approximated by our linear scale model. This is also true of the model recently suggested by Breusch and Pagan **[10]** of which (2.13) is a special case. Bickel **[6]** treats the Anscombe model in this way, suggesting robustified versions of Anscombe's tests. Very much in the spirit of Bickel's work, the present paper proposes an alternative approach to robustified tests for heteroscedasticity.

In the next section we sketch the asymptotic theory of "regression quantile" statistics under slightly weaker conditions than we have employed in previous work. This theory leads directly to new tests for heteroscedasticity in linear models.

3. THE ASYMPTOTIC THEORY OF REGRESSION QUANTILES

The asymptotic theory of regression quantiles in linear models with independent and identically distributed errors is treated in Koenker and Bassett **[17]** and with somewhat different methods in Ruppert and Carroll **[25]**. In this section we weaken the iid assumption slightly to consider the asymptotic behavior of $\sqrt{n}(\hat{\beta}(u) - \beta(u))$ under a particular form of asymptotically vanishing heteroscedasticity. We adopt the linear scale model (2.12), but rather than a fixed vector γ determining the scale parameter, we consider a sequence, $\{\gamma_n\}$, depending on sample size. We make the following additional assumptions:

ASSUMPTION A1 (Density): The error distribution, F, has continuous and strictly positive density, f, for all z such that $0 < F(z) < 1$.

ASSUMPTION A2 (Design): The sequence of design points $\{x_i\}$, satisfies $n^{-1}\sum x_i x_i' \to D$, a positive definite matrix.

ASSUMPTION A3 (Scale): The sequence of scaling functions takes the form $\sigma_n(x) = 1 + x\gamma_n$, where $\gamma_n = \gamma_o/\sqrt{n}$, for some fixed $\gamma_o \in \mathbf{R}^K$.

Let $\zeta = (\beta(u_1), \ldots, \beta(u_M))$ and $\hat{\zeta} = (\hat{\beta}(u_i))$ with $\beta(u)$ as defined in (2.9), and let $Q(u) = (Q_\varepsilon(u_1), \ldots, Q_\varepsilon(u_M))$. We may now state our main result.

THEOREM 3.1: *Under Assumptions* A1–A3, $\sqrt{n}(\hat{\zeta} - \zeta)$ *converges in distribution to an MK-variate Gaussian vector with mean vector* $Q(u) \otimes \gamma_o$ *and covariance matrix* $\Omega(u, F) \otimes D^{-1}$ *where* Ω *has typical element* $\omega_{ij} = [\min(u_i, u_j) - u_i u_j] / f(Q_\varepsilon(u_i)) f(Q_\varepsilon(u_j))$.

PROOF: We adopt the approach and notation of Ruppert and Carroll **[25]**. Let

$$V_n(\Delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \psi_u\left(\sigma(x_i)\varepsilon_i - x_i\Delta/\sqrt{n} - Q_\varepsilon(u)\right) \tag{3.1}$$

where

$$\psi_u(z) = \begin{cases} u & \text{as } z \geqq 0, \\ u - 1 & \text{as } z < 0. \end{cases}$$

Set $\hat{\Delta} = \sqrt{n}\,(\hat{\beta}(u) - \beta(u))$. Clearly, $V_n(\hat{\Delta}) = o_p(1)$; this is simply a statement that a particular element of the subgradient of the defining minimization problem is $o_p(\sqrt{n})$ when evaluated at elements of a solution sequence. Now define,

$$U_n(\Delta) = V_n(\Delta) - V_n(0) - E(V_n(\Delta) - V_n(0)). \tag{3.2}$$

From Ruppert and Carroll [25] and Bickel [6] we have for any $L > 0$

$$\sup_{0 \leqq \|\Delta\| < L} \|U_n(\Delta)\| = O_p(1) \tag{3.3}$$

and for any $\varepsilon > 0$, there exists η, L, and n_o such that

$$\Pr\left\{ \inf_{\|\Delta\| > L} |V_n(\Delta)| < \eta \right\} < \varepsilon \qquad \text{for } n > n_o. \tag{3.4}$$

By (3.4) and $V_n(\hat{\Delta}) = o_p(1)$ we have $\hat{\Delta} = O_p(1)$ so $U_n(\hat{\Delta}) = o_p(1)$ follows from (3.3). Thus,

$$V_n(0) + E(V_n(\hat{\Delta}) - V_n(0)) = o_p(1). \tag{3.5}$$

Now,

$$V_n(\Delta) = \frac{1}{\sqrt{n}} \sum x_i \psi_u\big((1 + x_i\gamma_o/\sqrt{n})\varepsilon_i - x_i\Delta/\sqrt{n} - Q_\varepsilon(u)\big)$$

so for n sufficiently large we may write

$$V_n(\Delta) = \frac{1}{\sqrt{n}} \sum x_i \psi_u(\varepsilon_i - h(\nu_{1i}, \nu_{2i})) \tag{3.6}$$

where

$$h(\nu_1, \nu_2) = (Q_\varepsilon(u) + \nu_1)/(1 + \nu_2)$$

so

$$EV_n(\Delta) = \frac{1}{\sqrt{n}} \sum x_i\big[(u - 1)F(h_i) + u(1 - F(h_i))\big]. \tag{3.7}$$

Expanding $F(h_i)$ to two terms we have

$$F(h(\nu_{1i}, \nu_{2i})) = u + \frac{1}{\sqrt{n}} f(Q_\varepsilon(u)) x_i(\Delta - Q_\varepsilon(u)\gamma_o); \tag{3.8}$$

thus,

$$EV_n(\hat{\Delta}) = -f(Q_\varepsilon(u))\left[\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right](\hat{\Delta} - Q_\varepsilon(u)\gamma_o) + o_p(1). \tag{3.9}$$

The term in square brackets converges to D, so

$$E(V_n(\hat{\Delta}) - V_n(0)) \to -f(Q_\varepsilon(u))D\hat{\Delta} \tag{3.10}$$

and thus (3.5) implies the following asymptotic linearity result:

$$\sqrt{n}\,(\hat{\beta}(u) - \beta(u)) = \frac{1}{\sqrt{n}} f(Q_\varepsilon(u))^{-1}D^{-1} \times \sum_{i=1}^{n} x_i\psi_u(\sigma(x_i)\varepsilon_i - Q_\varepsilon(u)) + o_p(1). \tag{3.11}$$

The design hypothesis (A2) implies $\max_{i,k}|x_{ik}| = o(\sqrt{n}\,)$ so $\sqrt{n}\,(\hat{\beta}(u_1) - \beta(u_1), \ldots, \hat{\beta}(u_M) - \beta(u_M))$ satisfies the standard multivariate Lindeberg condition and therefore is asymptotically Gaussian. The expectation of the right hand side of (3.11) converges to $Q_\varepsilon(u)\gamma_o$ using (3.9) and the covariance matrix follows from routine but somewhat tedious calculations. See, for example, the proof of Corollary 1 in Ruppert and Carroll [25]. *Q.E.D.*

Under iid conditions Theorem 3.1 reduces to Theorem 3.1 of Koenker and Bassett [17]. Conditional quantile functions are parallel hyperplanes in K space. And the slope coefficients of all regression quantiles converge in probability to the same vector. When the errors are not identically distributed the situation is quite different. In the linear scale model of heteroscedasticity described above, $\hat{\beta}(u)$ converges in probability to $\beta(u) + Q_\varepsilon(u)\gamma_o$, so slope coefficients depend in a nontrivial way on u. A convenient aspect of making $\gamma_n = O(1/\sqrt{n}\,)$ is that the limiting covariance structure of $\hat{\beta}(u)$ is independent of γ. This is essential for the hypothesis testing discussion to follow, but by no means necessary for consistency results for example. Consistency obtains as long as $Q_Y(u\,|\,x)$ is strictly linear in x as for example in (2.12).

In Koenker and Bassett [17] we emphasized that robust estimators of the parameters of iid linear models can be constructed from linear combinations of a few regression quantile estimates.[3] In addition, analogues of trimmed means were suggested for iid linear models based on regression quantiles. These "trimmed-least-squares estimators" have been studied intensively by Ruppert and Carroll [25]. Methods of estimation and tests of hypotheses based on regression quantiles in linear models with iid errors can substantially improve upon traditional least-squares methods in non-Gaussian situations. This is especially true when error distributions are longer-tailed than the Gaussian distribution. But all of this

[3] Robert Hogg has recently suggested to us (personal communication) that rather than forming estimates from linear combinations of regression quantiles, one might form linear combinations of several regression quantile objective functions and then minimize, effectively constraining the slope estimates of several regression quantile solutions to be identical. This suggestion yields estimators with similar, but not identical, asymptotic behavior to those we originally suggested. We hope to report in greater detail on such estimators in future work.

earlier work relies heavily on the iid-errors assumption which implies that $Q_Y(u \mid x)$ depends upon x only in *location*. In many applications it may be plausible that exogenous variables influence the scale, tail behavior, or other characteristics of the conditional distribution of Y. It is to tests of one such hypothesis that we now turn.

4. TESTS FOR HETEROSCEDASTICITY

Theorem 3.1 of the preceding section provides a foundation for tests of hypotheses. Having estimated the parameters of several conditional quantile functions and noted discrepancies among the estimated slope parameters, the question naturally arises: "Are these discrepancies 'significant'?" Tests with asymptotic validity may be readily constructed. No parametric assumptions on the shape of the error distribution are required. This nonparametric feature of the tests stands in marked contrast to the dominant approach found in the literature which rests, shakily in our view, on the hypothesis of Gaussianity. As Bickel [6] and others have recently pointed out, tests based on least squares residuals from a preliminary fit of the model are highly sensitive to Gaussian assumptions. Slight perturbations from Gaussianity can wreak havoc with the behavior of least-squares based test statistics. We will illustrate this phenomenon in Section 4.2.

4.1 *A Test Statistic*

Now consider the general linear hypothesis

(4.1) $$H_o : H\xi = h.$$

Under the conditions of Theorem 3.1 we have the following result.

THEOREM 4.1: *Under the null hypothesis H_o, the test statistic,*

(4.2) $$T = n(H\hat{\xi} - h)'\left[H(\Omega \otimes D^{-1})H'\right]^{-1}(H\hat{\xi} - h)$$

is asymptotically non-central chi-square with rank (H) degrees of freedom and noncentrality

(4.3) $$\eta = (H(Q_\varepsilon(u) \otimes \gamma_o))'\left[H(\Omega \otimes D^{-1})H'\right]^{-1}(H(Q_\varepsilon(u) \otimes \gamma_o)).$$

In the homoscedastic case the slope parameters are identical at every quantile. So partitioning $\beta = (\beta_1, \beta_2)'$ and $X = [1 \vdots X_2]$, set $h = 0$, and

(4.4) $$H_\Delta = \Delta \otimes \Phi$$

where Δ is an $(m - 1) \times m$ matrix with typical element $\Delta_{ij} = \delta_{ij} - \delta_{i(j-1)}$, δ_{ij} is the Kronecker delta, and $\Phi = [O \vdots I_{k-1}]$. Thus, for example, when $m = 2$, $\Delta = [1, -1]$ and when $m = 3$

$$\Delta = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}. \tag{4.5}$$

Then $H\hat{\zeta} = [\Delta \otimes \Phi]\hat{\zeta} = (\hat{\beta}_2(u_1) - \hat{\beta}_2(u_2), \ldots, \hat{\beta}_2(u_{M-1}) - \hat{\beta}_2(u_M))$,

$$H(\Omega \otimes D^{-1})H' = \Delta\Omega\Delta' \otimes \Phi D^{-1}\Phi', \tag{4.6}$$

and

$$\begin{aligned} \eta &= (q_\Delta \otimes \Phi\gamma_o)'\left[\Delta\Omega\Delta' \otimes \Phi D^{-1}\Phi'\right]^{-1}(q_\Delta \otimes \Phi\gamma_o) \\ &= q_\Delta'(\Delta\Omega\Delta')^{-1}q_\Delta \cdot (\Phi\gamma_o)'(\Phi D^{-1}\Phi')^{-1}\Phi\gamma_o \end{aligned} \tag{4.7}$$

where $q_\Delta = (Q_\varepsilon(u_1) - Q_\varepsilon(u_2), \ldots, Q_\varepsilon(u_{M-1}) - Q_\varepsilon(u_M))$. Thus, in the homoscedastic case, $\gamma_0 = 0$, T is asymptotically central χ^2 with $(M-1)(K-1)$ degrees of freedom. The power of the test depends as expected on the design, γ_o, q_Δ, and, via Ω, on the density of ε valuated at chosen quantiles. In practice we replace D^{-1} with $n(X'X)^{-1}$, so

$$n\left[\Phi D^{-1}\Phi'\right]^{-1} = (X_2'X_2 - n\bar{x}_2'\bar{x}_2). \tag{4.8}$$

We must also replace Ω with a consistent estimate $\hat{\Omega}$. This problem is addressed in Section 4.3.

4.2 *Asymptotic Relative Efficiency with Respect to a Least-Squares Test*

It is interesting at this point to compare the asymptotic behavior of the test suggested above with a test of H_o recently proposed by Breusch and Pagan [10]. Their test is based on the vector of least-squares residuals,

$$\hat{u} = \left[I - X(X'X)^{-1}X'\right]y. \tag{4.9}$$

Let $\hat{w} = (\hat{u}_i^2)$, $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$, and

$$\hat{\gamma} = (X'X)^{-1}X'\hat{w}. \tag{4.10}$$

Then Breusch and Pagan suggest a test of $\Phi\gamma_o = 0$ based on the test statistic

$$\bar{\xi} = \tfrac{1}{2}(\Phi\hat{\gamma})'\left(\Phi(X'X)^{-1}\Phi'\right)^{-1}\Phi\hat{\gamma}/\hat{\sigma}^4. \tag{4.11}$$

They demonstrate that if ε is a vector of independent and identically distributed Gaussian random variables then under the null hypothesis, $H_o: \gamma = 0$, $\bar{\xi}$ will be asymptotically central χ^2_{K-1}.

Strengthening Assumptions (A1) and (A2) slightly so that $V\varepsilon^2 < \infty$ and

$\max_{ij}|x_{ij}| < \infty$, one can easily show that

$$(4.12) \qquad \xi^* = \bar{\xi}/\lambda \to \chi^2_{K-1}\left(4(\Phi\gamma_o)'(\Phi D^{-1}\Phi')^{-1}\Phi\gamma_o / V\varepsilon^2\right)$$

where $\lambda = \frac{1}{2} V\varepsilon^2/\sigma_\varepsilon^4$. When the error distribution F is Gaussian then $V\varepsilon^2 = 2\sigma_\varepsilon^4$, so $\lambda = 1$, and the asymptotic size of Breusch and Pagan's test based on $\bar{\xi}$ is correct. However, deviations from Gaussianity, or more explicitly, deviations from Gaussian kurtosis imply $\lambda \neq 1$ and the significance levels suggested by Breusch and Pagan will then be incorrect asymptotically. This point is hardly new. It was made by Box [9] and has been recently reemphasized by Bickel [6] in the context of Anscombe's model of heteroscedasticity. The obvious solution to the size problem is to "Studentize" the quadratic form suggested by Breusch and Pagan by an estimate of V_ε^2, say $n^{-1}\sum(\hat{u}_i^2 - \hat{\sigma}^2)^2$, instead of $2\hat{\sigma}^4$, as they suggest. Note that this approach is adopted by White [30] in proposing a similar test for heteroscedasticity based on least squares residuals.

The local power of our test, $\hat{\xi}$, and the appropriately Studentized test, ξ^*, may be compared by computing Pitman's asymptotic relative efficiency (ARE) of the two tests. For any two tests of the same hypothesis, the same size, and same power with respect to the same alternative: if the first test requires n_1 observations and the second requires n_2 observations, then the relative efficiency of the second test with respect to the first is n_1/n_2. The limit of this ratio as both numerator and denominator tend to infinity is the asymptotic relative efficiency of the two tests. See Pitman [22] or Rao [24] for additional details.

In an effort to obtain some quantitative feel for plausible ARE's in applications we have calculated some for members of the family of contaminated Gaussian distributions. This family is often suggested as affording plausible longer-tailed alternatives to strictly Gaussian errors, see Tukey [26]. In its simplest form, the family is indexed by two parameters: π, the proportion of contamination, and ϕ, the scale of the contaminating Gaussian distribution, i.e.,

$$(4.13) \qquad F_{\pi,\phi}(x) = (1-\pi)G_{0,1}(x) + \pi G_{0,\phi}(x)$$

where $G_{0,\phi}$ denotes the Gaussian distribution function with mean 0 and scale ϕ.

To facilitate the comparison we use the simplest possible version of $\hat{\xi}$ based on only two symmetrically placed quantiles. We assume, of course, that a consistent estimate of Ω is available. See the next section for a discussion of how such estimates may be constructed. Thus in the notation of the previous sections $u = [p, 1-p]$, $\Delta = [1, -1]$ and $h = 0$. We will denote by $\hat{\xi}_p$ the test based on the pth and $(1-p)$th quantiles. Since both $\hat{\xi}_p$ and ξ^* are asymptotically χ^2 with the same degrees of freedom, $K-1$, their ARE is simply the ratio of the noncentrality parameters of their limiting distributions, which in our case yields

$$(4.14) \qquad \text{ARE} = \hat{\eta}_p/\eta^* = q_\Delta^2(p)V\varepsilon^2/4(\Delta\Omega\Delta').$$

A few contours of asymptotic relative efficiency of the two tests are illustrated in Figures 1 and 2 for $p = .25$ and $p = .05$. The striking feature of the figures is the rapid deterioration of the least-squares test as modest amounts of contamination are introduced. For a fixed proportion of contamination the asymptotic relative efficiency of ξ^* to $\hat{\xi}(p)$ goes to zero as ϕ the relative scale of the contaminating distribution increases without bound. The test based on $\hat{\xi}_{.25}(\hat{\xi}_{.05})$ needs about 2.72 (1.56) the number of observations needed by ξ^* at the strictly Gaussian case, but with 10 per cent contamination from a Gaussian distribution with relative scale of three, the asymptotic efficiency comparison is dramatically reversed. The least-squares test ξ^* now requires more than three times the number of observations needed by either $\hat{\xi}_{.25}$ or $\hat{\xi}_{.05}$. When there is 20 percent contamination from a distribution with relative scale of five, ξ^* needs 40 times the number of observations that $\hat{\xi}_{.25}$ needs to achieve the same asymptotic power. It should be noted that $p = .25$ is quite pessimistic, in the sense that smaller p's can rather dramatically improve the efficiency of the $\hat{\xi}_p$ test near the strictly Gaussian case. This is evident in the figures, and may suggest that small values of p, say .05–.10, may be desirable unless severe contamination is expected.

Expanding the test statistic $\hat{\xi}_p$ to include additional quantiles has a mixed payoff. The noncentrality parameter of the limiting χ^2 cannot decrease and may increase which would increase the power of the test. But degrees of freedom also

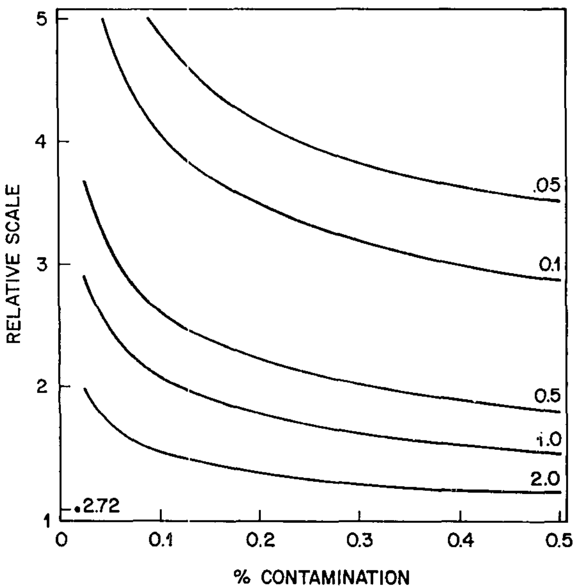

FIGURE 1 —Asymptotic relative efficiency contours $p = .25$.

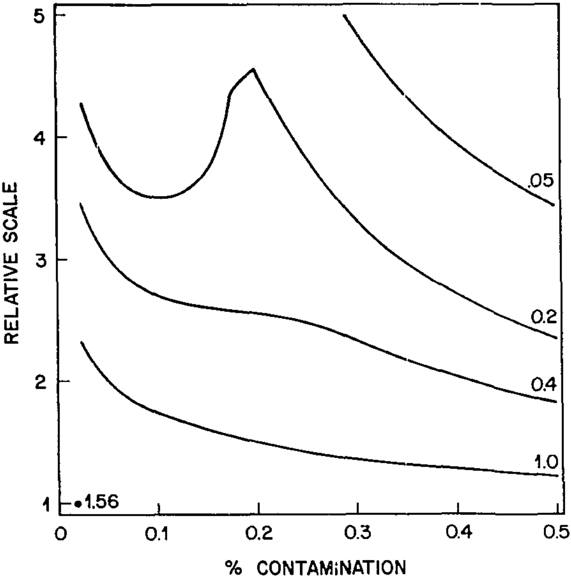

FIGURE 2 —Asymptotic relative efficiency contours $p = .05$.

increase, decreasing power, and counterbalancing the noncentrality effect. Another possibility (which we have not explored) would be to compute weighted averages of several upper and lower quantiles and then test for equality the weighted averages. This would capture the noncentrality effect to some extent without sacrificing the degrees of freedom effect.

4.3 *Estimating* Ω

Tests of hypotheses based upon regression quantile statistics typically require the estimation of certain nuisance parameters of the matrix Ω. In the case of the heteroscedasticity test proposed above the reciprocals of the density function of errors evaluated at each of the quartiles are required. The reciprocal of a density is the derivative of a quantile function. Tukey [27] has aptly called such functions "sparsity functions." See also Bloch and Gastwirth [7].

In the one-sample (location) model the problem of estimating elements of Ω is essentially a problem of smoothing the empirical quantile function. For an important recent discussion of such problems, see Parzen [20]. In the linear model setting a natural approach would be to replace $\hat{F}$ or $\hat{Q}$ based on a single sample with an $\hat{F}$ or $\hat{Q}$ based on residuals from a preliminary fit of the model.

Recent work by Ruppert and Carroll [25], Loynes [18], and Pierce and Kopecky [23] provide asymptotic support for such an approach.

However, here we adopt an alternative approach. In (2.7) we defined an empirical conditional quantile function $Q_Y(u \mid x)$ at an arbitrary design point x. We have studied these functions in some detail in Bassett and Koenker [5]. We restate two important results here without proof.

THEOREM 4.2: *The sample paths of $\hat{Q}_Y(u \mid \bar{x})$ are non-decreasing, left-continuous, jump functions of u.*

THEOREM 4.3: *In linear models with independent and identically distributed errors satisfying Assumptions A1–A2 the random function of u, $\sqrt{n}\,[\hat{Q}_Y(u \mid \bar{x}) - Q_Y(u \mid \bar{x})]$, has finite dimensional distributions which are asymptotically Gaussian with mean vector zero and covariance matrix Ω.*

The first result assures that $\hat{Q}_Y(u \mid \bar{x})$ looks, at least superficially, like an empirical quantile function, i.e., a monotone staircase. However, it is important to note that unlike the one-sample case the steps are not of equal width on the u-axis, but depend upon the design configuration as well as the realization of Y. We also should note that $Q_Y(u \mid x)$ is not necessarily monotone in u for $x \neq \bar{x}$.

The second result establishes that not only is $\hat{Q}_Y(u \mid \bar{x})$ a weakly consistent estimator of $Q_Y(u \mid \bar{x})$, but the *normalized* process $\sqrt{n}\,(\hat{Q} - Q)$ behaves in large samples as if it arose from the location model and is asymptotically independent of design. Thus consistent estimates of the sparsity function may be easily constructed from $\hat{Q}$.

Let $\hat{R}(u)$ denote the minimum value achieved by the objective function at each regression quantile, i.e.,

$$\hat{R}(u) = \min_{b \in \mathbf{R}^K} \left[\sum_{i=1}^{n} \rho_u(y_i - x_i b) \right]. \tag{4.15}$$

In Bassett and Koenker [5] we show that $Q_Y(u \mid x)$ is simply a translated version of the left derivative of $\hat{R}(u)$ with respect to u,

$$\hat{Q}_Y(u \mid \bar{x}) = \bar{y} - \hat{R}'(u - 0) \tag{4.16}$$

where $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$. The computation of $\hat{R}(u)$ may at first appear prohibitive since it requires the solution of an entire one-dimensional family of linear programs indexed by $u \in (0, 1)$. However, the parametric programming problem posed in Koenker and Bassett [17] is readily solved for $\hat{R}(\cdot)$. Even more efficient methods are undoubtedly possible using the techniques developed by Barrodale and Roberts [2] for l_1 estimation. See the Appendix of Bassett and Koenker [5] for further details on computation. The computation of $\hat{R}(u)$ for the examples discussed in Section 5 was carried out by the IBM-MPSX linear programming package, and checked with an adaptation of the Bartels and Conn [3] l_1 regression algorithm.

Let $\tilde{R}(u)$ denote a smoothed version of $\hat{R}(u)$; then estimates of the sparsity function of errors are immediately available by differentiating twice:

$$\tilde{s}(u) = \tilde{R}''(u). \tag{4.17}$$

In our examples below we have chosen to use a variant of the histospline methods of Boneva, Kendall, and Stefanov [8], and Wahba [28]. The smoothed estimates $\hat{R}(u)$ are cubic splines—piecewise cubic polynomials. The smoothness of $\hat{R}$ is governed by the number of knots chosen at abscissa points. At knots the third derivative of the spline has a jump discontinuity In the examples reported in the next section knots were equally spaced on $(0, 1)$ with "natural" end conditions. Computation of the smoothing splines and their derivatives was carried out with the aid of B-spline routines in the PORT Library described in Fox [12].

There is an element of inevitable arbitrariness in any smoothing or derivative estimation technique since many quite different techniques all yield consistent estimates. The obvious element here is the choice of the degree of smoothness governed by the knot selection. While several formal methods of choosing a degree of smoothing exist, notably the cross-validation methods of Wahba [29], we have chosen in the examples below to compare results for several different degrees of smoothness.

5. SOME EXAMPLES

We now illustrate the methods introduced above with several examples. We begin with a particularly simple form of design. The next two empirical examples are simple bivariate models which lend themselves to visual analysis.

5.1 *Two Sample Problems*

We have already noted that in the one-sample (location) model our methods specialize to consideration of the ordinary sample quantiles. In the two sample problem, $y = (y_1, y_2)'$, with design

$$X = \begin{bmatrix} 1_{n_1} & 0 \\ 0 & 1_{n_2} \end{bmatrix}, \tag{5.1}$$

it is readily shown that any uth regression quantile estimate, $\hat{\beta}(u) = (\hat{\beta}_1(u), \hat{\beta}_2(u))$, has the property that $\hat{\beta}_1(u)$ is a uth sample quantile from sample 1 consisting of the first n_1 observations and $\hat{\beta}_2(u)$ is a uth sample quantile from sample 2 consisting of the last n_2 observations. Thus, denoting the empirical quantile functions of the two samples by $\hat{Q}_i(u)$ we have for the two sample problem,

$$\hat\beta(u) = (\hat Q_1(u), \hat Q_2(u)). \tag{5.2}$$

The conditional empirical quantile function for the two sample problems is thus simply a weighted average of the two one-sample empirical quantile functions. In particular, we have

$$\hat Q_Y(u \mid \bar x) = \frac{n_1}{n} \hat Q_1(u) + \frac{n_2}{n} \hat Q_2(u). \tag{5.3}$$

It proves convenient to transform the design to accommodate an intercept,

$$\tilde X = XA = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} & \vdots & \mathbf{0} \\ \mathbf{1} & \cdots & \\ & \vdots & \mathbf{1} \end{bmatrix}. \tag{5.4}$$

By the equivariance result (Theorem 3.2(iv)) of Koenker and Bassett **[17]** our estimates for the transformed model are simply

$$\tilde\beta(u) = A^{-1}\hat\beta(u) = \begin{pmatrix} \hat Q_1(u) \\ \hat Q_2(u) - \hat Q_1(u) \end{pmatrix}. \tag{5.5}$$

If we now consider tests for departures from homoscedasticity like those suggested above, we see that they reduce to tests of hypotheses of the form,

$$(Q_2(u_1) - Q_1(u_1)) - (Q_2(u_2) - Q_1(u_2)) = 0. \tag{5.6}$$

5.2 *Engel's Food Expenditure Data*

In this subsection we investigate the data originally presented by Ernst Engel **[11]** to support the proposition that food expenditure constitutes a declining share of personal income. The data consists of 235 budget surveys of 19th century European working class households. An interesting discussion of the data sources and Engel's analysis of the data may be found in Perthel **[21]**. Following established custom we transform the data on annual household income and food expenditure to natural logarithms. In Figure 3 we present the scatter of original data on double log axes. Fitted regression quartile lines have been superimposed on the same figure. The fitted quartile lines suggest a weak increasing conditional scale effect. The estimated slope parameters (Engel elasticities) for the three quartiles are .85, .88, .92 respectively. To test the "significance" of this effect we begin by estimating the reciprocal of the error density at the quartiles.

Our method is described in Section 4.2. The estimate $\hat R(u)$ was computed, and several smoothed histopline estimates $\tilde R(u)$ were made. The degree of smoothness of $\tilde R(u)$ is governed by the number of knots: estimates with 5, 10, and 15 equally-spaced interior knots in $(0, 1)$ were computed. Our sparsity function

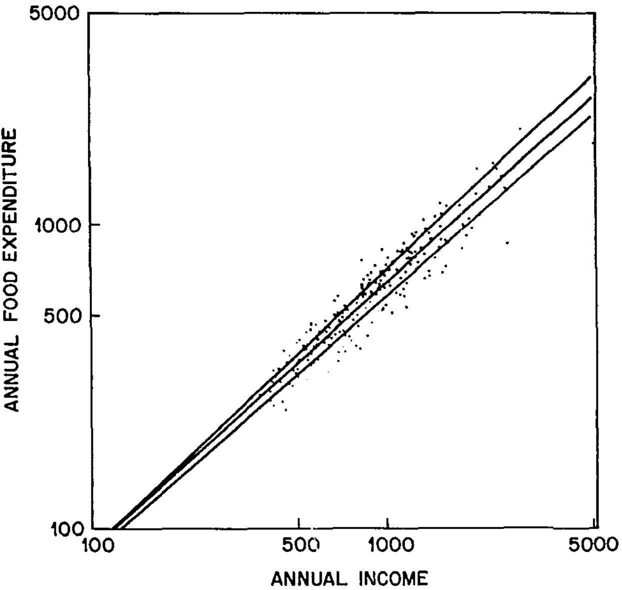

FIGURE 3 —Quartile Engel curves for food.

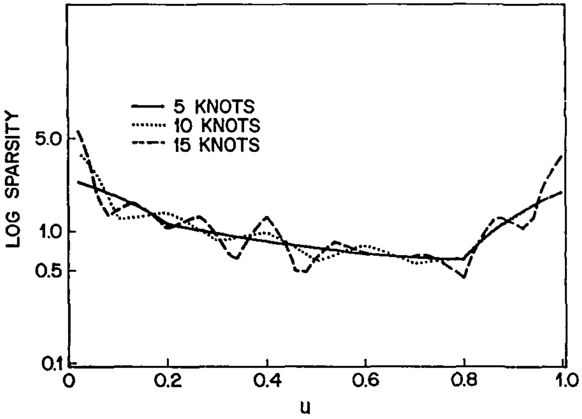

FIGURE 4 —Sparsity functions for the Engel data.

TABLE I

SPARSITY ESTIMATES AND χ^2 TESTS FOR HETEROSCEDASTICITY

Smoothness	knots	$\tilde{s}(1/4)$	$\tilde{s}(1/2)$	$\tilde{s}(3/4)$	T
smooth	5	.5446	.3754	.3087	4.42
moderate	10	.5586	.3046	.3009	5.00
rough	15	.6256	.3154	.3139	4.41

estimates $\tilde{s}(u)$ are second derivatives of $\tilde{R}(u)$. They are linear splines, i.e., continuous, piecewise linear functions. The three resulting estimates are illustrated in Figure 4. Sparsity estimates at the quartiles for the three degrees of smoothness are given in Table I.

We illustrate the calculations of the test statistic for the case of moderate smoothing with 10 knots. The estimates of the sparsity function at the quartiles are

$$\left(\tilde{s}\left(\tfrac{1}{4}\right), \tilde{s}\left(\tfrac{1}{2}\right), \tilde{s}\left(\tfrac{3}{4}\right)\right) = (.5586, .3046, .3009)$$

so we have

$$\hat{\Omega} = \begin{bmatrix} .0585 & .0213 & .0105 \\ .0213 & .0232 & .0115 \\ .0105 & .0115 & .0170 \end{bmatrix}.$$

With Δ as in (4.5) we have

$$\left[(x'x - n\bar{x}^2)^{-1}\Delta\Omega\Delta'\right]^{-1} = \begin{bmatrix} 1168.50 & 65.45 \\ 65.45 & 2629.00 \end{bmatrix}.$$

So with

$$\hat{\zeta} = \begin{pmatrix} \hat{\beta}_2\left(\tfrac{1}{4}\right) - \hat{\beta}_2\left(\tfrac{1}{2}\right) \\ \hat{\beta}_2\left(\tfrac{1}{2}\right) - \hat{\beta}_2\left(\tfrac{3}{4}\right) \end{pmatrix} = \begin{pmatrix} -.027 \\ -.039 \end{pmatrix}$$

we have $\hat{\zeta}'\hat{\Sigma}^{-1}\hat{\zeta} = 5.0$. In Table I we present sparsity estimates and test statistics for varying degrees of smoothness. A χ^2 variable on 2 degrees of freedom exceeds 6.0 with probability .05 so the null hypothesis of homoscedasticity cannot be rejected on the basis of the tests reported in Table I.

5.3 *Demand for Admen*

In this subsection we investigate a simple model of labor demand by advertising agencies. In Figure 5 we have plotted data on the number of employees and annual U.S. billings in 1967 of 108 of the largest American advertising firms. The data is taken from *Advertising Age*. As in the previous example we have superimposed the three regression quartile lines on the raw data. The quartiles have slopes of 6.25, 6.67, 7.20 employees per million dollars of annual billing in 1967 dollars. It is interesting to note in this example that the least squares

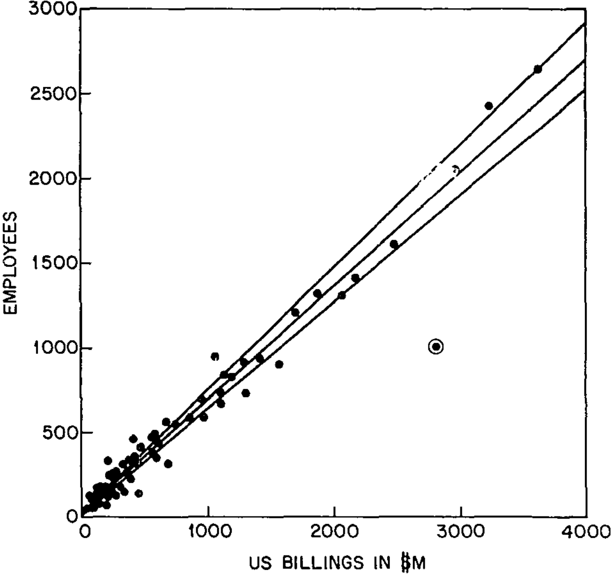

FIGURE 5—Quartile labor demand by advertising firms.

regression line with $\hat{\beta}_2 = 6.4$ persons/million dollars looks roughly like the first quartile fit due to the strong influence of the anomalous behavior of McCann-Erickson Inc. which is circled in the figure and has roughly half as many employees as one would predict from the remaining data. Again we have a conical configuration of quantile lines possibly reflecting an increasing conditional scale effect.

Again we have estimated several smoothed estimates $\tilde{R}(u)$ for this example. Proceeding as in the previous example using the moderate smoothing estimate (with 10 equally spaced knots) we obtain a test statistic of 46.4 for equality of slopes of the quartile models. The smoother estimate with only 5 knots yields a test statistic of 83.9. Since a χ^2 variable on 2 degrees of freedom exceeds 9.2 with only 1 per cent probability, the hypothesis of homoscedasticity is firmly rejected. Dispersion in labor demand measured by interquartile distances significantly increases with scale. Note that this inference is extremely robust with respect to the position of the anomalous McCann-Erickson point. Moving McCann-Erickson up as far as the first quartile line or down exerts no effect on the quartile estimates presented above (see Theorem 3.4 of Koenker and Bassett [17]), but such movements would exert a strong effect on least-squares estimates and an even stronger effect on tests for heteroscedasticity based on least-squares residuals like those of Goldfeld-Quandt [15] and Breusch-Pagan [10].

6. CONCLUSION

We have proposed a method for estimating linear models for conditional quantile functions. The method makes no parametric assumptions on the shape of the error distribution, and employs the "regression quantile" statistics previously introduced in Koenker and Bassett [17]. Tests for heteroscedasticity are proposed based on the large sample theory of the proposed estimators. We believe these methods should prove to be useful diagnostic tools in a wide range of linear model applications.

Bell Laboratories
and
University of Illinois-Chicago Circle

Manuscript received September, 1979; revision received February, 1981.

REFERENCES

[1] Anscombe, F. J.: "Examination of Residuals," *Proceedings of the 5th Berkeley Symposium*, 1 (1961), 1-36.
[2] Barrodale, I., and F. D. K. Roberts: "Solution of an Overdetermined System of Equations in the $L1$ Norm," *Communications of the ACM*, 17(1974), 319-320.
[3] Bartels, R. H., and A. R. Conn: "Linearly Constrained Discrete l_1 Problems," *ACM Transactions on Mathematical Software*, 6(1980), 594-608.
[4] Bassett, G., and R. Koenker: "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association*, 73(1978), 618-622.
[5] ———: "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, to appear.
[6] Bickel, P. J.: "Using Residuals Robustly I: Tests for Heteroscedasticity and Nonlinearity," *Annals of Statistics*, 6(1978). 266-291.
[7] Bloch, D. A., and J. L. Gastwirth: "On a Simple Estimate of the Reciprocal of the Density Function," *Annals of Mathematical Statistics*, 39(1968), 1083-1085.
[8] Boneva, L. I., D. Kendall, and I. Stefanov: "Spline Transformations: Three New Diagnostic Aids for the Statistical Data—Analyst," *Journal of the Royal Statistical Society*, Series (B), 33(1971), 1-70 (with discussion).
[9] Box, G. E. D.: "Non-normality and Tests on Variances," *Biometrika*, 40(1953), 318-335.
[10] Breusch, T. S., and A. R. Pagan: "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47(1980), 1287-1294.
[11] Engel, E.: "Die Productions-und Consumtionsver-hältnisse des Konigreichs Sachsen," in *Zeitschrift des Statistischen Bureaus des Koniglich Sachsischen Ministeriums des Innern*, 8(1857), 1-54.
[12] Fox, P.: *The Post Mathematical Subroutine Library*. Murray Hill, New Jersey: Bell Laboratories, 1976.
[13] Godfrey, L. G.: "Testing for Multiplicative Heteroscedasticity," *Journal of Econometrics*, 8(1978), 227-236.
[14] Goldberger, A. S.: *Topics in Regression Analysis*. Toronto: MacMillan, 1968.
[15] Goldfeld, S. M., and R. E. Quandt: "Some Tests for Heteroscedasticity," *Journal of the American Statistical Association*, 60(1965), 539-547.
[16] Harvey, A. C.: "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, 44(1976), 461-466.
[17] Koenker, R., and G. Bassett: "Regression Quantiles," *Econometrica*, 46(1978), 33-50.
[18] Loynes, R. M.: "The Empirical Distribution Function of Residuals from Generalized Regression," *Annals of Statistics*, 8(1980), 285-298.
[19] Mosteller, F., and J. W. Tukey: *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley, 1977.

[20] PARZEN, E.: "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74(1979), 105–131 (with discussion).
[21] PERTHEL, D.: "Engel's Law Revisited," *International Statistical Review*, 43(1975), 211–218.
[22] PITMAN, E. J. G.: *Some Basic Theory for Statistical Inference*. London: Halsted, 1979.
[23] PIERCE, D. A., AND K. J. KOPECKY: "Testing Goodness of Fit for the Distribution of Errors in Regression Models," *Biometrika*, 66(1979), 1–5.
[24] RAO, C. R.: *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons, 1973.
[25] RUPPERT, D., AND R. J. CARROLL: "Robust Regression by Trimmed Least Squares," *Journal of the American Statistical Association*, 75(1980), 828–838.
[26] TUKEY, J. W.: "A Survey of Sampling from Contaminated Distributions," in *Contributions to Probability and Statistics*, ed. by I. Olkin. Stanford: Stanford University Press, 1960.
[27] ———: "Which Part of the Sample Contains the Information," *Proceedings of the National Academy of Sciences*, 53(1965), 127–134.
[28] WAHBA, G.: "Histosplines with Knots which are Order Statistics," *Journal of the Royal Statistical Society*, Series B, 38(1976), 140–151.
[29] ———: "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society*, Series B, 40(1978), 364–372.
[30] WHITE, H.: "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48(1980), 817–838.