ASYMMETRIC LEAST SQUARES ESTIMATION AND TESTING

Whitney K. Newey
James L. Powell

Econometric Research Program
Research Memorandum No. 325

Revised June 1986

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, NJ 08544

ASYMMETRIC LEAST SQUARES ESTIMATION AND TESTING[1]

by

Whitney K. Newey and James L. Powell

1. Introduction

In the estimation of a linear regression function for a dependent variable y_i in terms of a vector of explanatory variables x_i, choice of an estimation method (say, least squares versus a more "robust" method, such as least absolute deviations) is often based solely on the criterion of relative efficiency. When y_i is in fact generated by a linear function of the regressors plus an i.i.d. error term which is independent of the regressors, relative efficiency is the appropriate criterion; only the interpretation of the intercept term depends on the choice of estimation method, provided the latter is well-behaved. However, under weaker conditions on the error terms, interpretation of the estimated coefficients depends crucially on the method used. If, conditional on the regressors, the dependent variable is symmetrically distributed about a linear function of x_i, this function is a "natural" estimand, and estimation methods based on minimization of symmetric empirical loss functions should consistently estimate its coefficients; however, if the conditional distribution of y_i given x_i is both heteroskedastic and asymmetric, different loss functions correspond to fundamentally different estimands. In this setting, "choice" of estimator amounts to choice of estimand, and relative efficiency of estimation is a much less compelling consideration (although, as Bickel and Lehmann (1975) argue, efficiency may be a useful criterion when choosing among possible estimands). Indeed, an adequate characterization of the conditional distribution of the dependent variable may require calculation of several regression coefficient vectors, each corresponding to different notions of "location" of this conditional distribution.

In view of the importance of homoskedasticity and/or conditional symmetry to the interpretation of regression coefficient estimates, it is worthwhile to test whether either of these conditions are applicable. For the linear regression model, several tests of the null hypothesis of homoskedasticity have recently been investigated. The majority of such tests uses residuals from a preliminary fit of the regression equation of interest; this group includes the tests proposed and studied by Anscombe (1961), Glejser (1969), Goldfeld and Quandt (1972), Harvey (1976), Godfrey (1978), Breusch and Pagan (1979), and White (1980). In their simplest form, these tests of homoskedasticity are tests that the coefficients of a second-stage regression of the squared values (or more general even functions) of the residuals on transformations of the regressors are zero. Testing for symmetry of the error distribution has received somewhat less attention in the statistical literature (presumably since it is not directly related to the question of the relative efficiency of weighted to classical least squares, which motivates much of the literature on heteroskedasticity); research has focussed on i.i.d. observations, rather than observations generated from a linear model (see, for example, Antille, Kersting, and Zucchini (1982), and Boos (1982)). Nonetheless, tests for symmetry analogous to those for heteroskedasticity can be constructed using odd rather than even functions of the residuals in a second-stage regression (as discussed in Section 4.2 below).

An alternative approach to testing homoskedasticity has been studied by Koenker and Bassett (1982); the test they propose is based upon the regression analogues of order statistics, termed "regression quantiles," introduced by Koenker and Bassett (1978). For their test, the null hypothesis of homoskedasticity is rejected if the slope coefficients of the regression equation, estimated at different quantiles of the conditional distribution of the dependent variable, are significantly different from one another. Comparing the asymptotic efficiency of this test relative to a corresponding "squared residual regression" test, the authors found some inefficiency of the regression quantiles test when the error distribution is Gaussian, but this conclusion was reversed for contaminated Gaussian error distributions, and the efficiency gains of the regression quantile test appeared to be substantial even for low levels of contamination.

Koenker and Bassett's approach to heteroskedasticity testing exploits the previously-discussed interdependence of estimand and estimator when the error terms are not i.i.d.; by comparing regression coefficients estimated at different quantiles, the question of heteroskedasticity is recast as a question concerning differences in alternative measures of "location" of the conditional distribution of the dependent variable. And, while the authors did not consider testing for conditional symmetry, their approach can easily be extended to this setting. However, there are certain drawbacks to the regression quantile approach. First, because the minimand defining the quantile estimators is not continuously differentiable, the estimators themselves are somewhat difficult to compute (though Koenker and Bassett do point out that the minimization problem can be restated as a linear programming problem, for which efficient algortihms are available). Also, efficiency of the estimators and corresponding tests depends on the precision with which percentiles, rather than moments, of the error distribution can be estimated; while the quantile estimators are "robust" against heavy-tailed error distributions (because they are based on absolute rather than squared error loss minimization), they are relatively inefficient for error distributions which are close to Gaussian or which have low densities at the corresponding percentiles. Finally, and perhaps most importantly, the asymptotic covariance matrix of quantile estimators depends on the values of the density function of the errors at those quantiles; such density function values are notoriously difficult to estimate, and the resulting test statistics for heteroskedasticity and asymmetry will depend on the degree of "smoothing" of the empirical distribution of the residuals, as chosen by the researcher.

The present paper proposes a least squares analogue of regression quantile estimation, as well as associated tests of homoskedasticity and symmetry. The regression coefficent estimators considered here were also investigated by Aigner, Amemiya, and Poirier (1976) but only in the context of a correctly-specified, i.i.d. error distribution. As shown in the following section, the estimands for this approach in a more general, non-i.i.d. setting characterize the conditional distribution of y_i given x_i in much the same way that the regression quantiles do; however, the estimators are simpler to compute (using iteratively-reweighted least squares), and their asymptotic covariance matrix can be estimated without estimation of the density function of the errors. The efficiency of this class of estimators, which includes least squares estimation as a special case, is governed by the efficiency of estimation of the first moment of the error distribution, so comparison of the resulting tests to tests using residuals from a first-stage regression do not depend on the "robustness" of the respective procedures.

In the following section, these least squares analogues of regression quantiles, termed "asymmetric least squares" estimators, are defined, and properties of the corresponding estimands for i.i.d. observations and for the linear regression model are presented. The large sample distributions of the estimators are derived in Section 3, both for the general case when y_i and x_i are jointly i.i.d. and for the special case of locally linear heteroskedasticity and/or asymmetry; in the latter case, test statistics are derived which have limiting noncentral chi-squared distributions under the local alternatives. In Section 4, the local power of the tests of homoskedasticity and symmetry are compared to that of tests based on regression quantiles, and of tests which use first-stage residuals, under the assumption of contaminated Gaussian errors. As the results of this section show, the asymmetric least squares tests have local power functions which are strikingly similar to certain other tests of heteroskedasticity and conditional asymmetry, and they perform quite well relative to regression quantile tests for the range of error distributions considered. The paper concludes with some qualitative observations concerning the general efficacy of the procedures considered. Proofs of the main theorems are given in a technical appendix.

2. Definition of the Asymmetric Least Squares Estimators

The observable data $\{(y_i, x_i'), i = 1,...,n\}$ are assumed to be generated by the linear model

$$(2.1) \qquad y_i = x_i'\beta_0 + u_i,$$

where $\{x_i\}$ is a sequence of regression vectors of dimension p with first component $x_{i1} \equiv 1$, β_0 is a conformable vector of unknown parameters, and $\{u_i\}$ is a sequence of scalar error terms.

The regression quantile (RQ) estimators, proposed by Koenker and Bassett (1978), are defined as those vectors $\hat{b}(\theta)$ which minimize the function

$$(2.2) \qquad Q_n(\beta;\ \theta) \equiv \Sigma_{i=1}^n r_\theta(y_i - x_i'\beta)$$

over β in $\mathbb{R}^p$ for fixed values of θ in $(0, 1)$, where $r_\theta(\cdot)$ is a convex loss function of the form

$$(2.3) \qquad r_\theta(\lambda) \equiv |\theta - 1(\lambda < 0)|\cdot|\lambda|,$$

with $1(A)$ denoting the indicator function for the event A. Under homoskedasticity the probability limits of the regression quantile estimators $\{b(\theta)\}$ for different choices of θ deviate from β_0 only in their intercept terms. That is, under homoskedasticity,

$$(2.4) \qquad \text{plim}\ \hat{b}(\theta) = \beta_0 + \eta(\theta)e_1,$$

where e_ℓ denotes the ℓ^{th} unit vector and $\eta(\theta) \equiv F^{-1}(\theta)$, the quantile function for the error term u_i. Under heteroskedasticity the probability limits for the slope coefficients will in general also vary with θ, with differences depending on the joint distribution of u_i and x_i.

The regression quantile estimators are thus a class of empirical "location" measures for the dependent variable whose sampling behavior involves the true regression coefficients and the stochastic behavior of the error terms. To obtain a similar class of location measures that are more convenient than regression quantiles we consider replacing the "check function" criterion of (2.3) with the following "asymmetric least squares" loss function:

$$(2.5) \qquad \rho_\tau(\lambda) \equiv |\tau - 1(\lambda < 0)| \cdot \lambda^2, \quad \text{for} \quad \tau \quad \text{in} \quad (0, 1).$$

The corresponding class of asymmetric least squares (ALS) estimators $\{\hat{\beta}(\tau)\}$ are defined to minimize

$$(2.6) \qquad R_n(\beta; \tau) \equiv \Sigma_{i=1}^n \rho_\tau(y_i - x_i'\beta)$$

over β, for $\rho_\tau(\cdot)$ given in (2.5). Aigner, Amemiya, and Poirier (1976) show that this estimator can be interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances.

To determine the class of location parameters that are estimated by $\{\hat{\beta}(\tau)\}$, consider the scalar parameter $\mu(\tau)$ which minimizes the function $E[\rho_\tau(Y - m) - \rho_\tau(Y)]$ over m, where the expectation is taken with respect to the distribution of the random variable Y, which is assumed to have finite mean. The parameter $\mu(\tau)$ is easily shown to be the solution of the equation

$$(2.7) \qquad \mu(\tau) - E(Y) = [(2\tau-1)/(1-\tau)] \cdot \int_{[\mu(\tau),\infty)} (y-\mu(\tau))dF(y),$$

where $F(y)$ is the c.d.f. of Y.[2] When $E(Y) = 0$ this equation for $\mu(\tau)$ is identical to the equation for the reservation wage for sequential, costless search from a fixed distribution when the search period interest rate is $(1-\tau)/(2\tau-1)$. Also, as discussed by DeGroot (1970, pp. 244-247), the integral on the right-hand side of equation (2.7) is proportional to the Bayes' risk for the problem of deciding whether a parameter is smaller or larger than a specified value, when f(y) is interpreted as the density function of the mean of the posterior distribution. It is evident from these interpretations that $\mu(\tau)$ is determined by the properties of the expectation of the random variable Y conditional on Y being in a tail of the distribution. Motivated by this fact we will henceforth refer to $\mu(\tau)$ as the τ^{th} <u>expectile</u>.[3]

The expectile function $\mu(\tau)$ summarizes the distribution function in much the same way that the quantile function $\eta(\theta) \equiv F^{-1}(\theta)$ does. Let I_F denote the set $\{y | 0 < F(y) < 1\}$.

Theorem 1: Suppose that $E(Y) = m$ exists. Then for each $0 < \tau < 1$, a unique solution $\mu(\tau)$ to equation (2.7) exists and has the following properties:

(i) As a function $\mu(\tau): (0, 1) \rightarrow \mathbb{R}$, $\mu(\tau)$ is strictly monotonic increasing.

(ii) The range of $\mu(\tau)$ is I_F and $\mu(\tau)$ maps $(0, 1)$ onto I_F.

(iii) For $\tilde{Y} = sY + t$, where $s > 0$, the τ^{th} expectile $\tilde{\mu}(\tau)$ of $\tilde{Y}$ satisfies $\tilde{\mu}(\tau) = s\mu(\tau) + t$.

(iv) If $F(y)$ is continuously differentiable then $\mu(\tau)$ is continuously differentiable, and for $y \neq m$ in I_F and τ_y such that $y = \mu(\tau_y)$,

$$F(y) = -[y - m + \tau_y\mu'(\tau_y)(1 - 2\tau_y)]/[\mu'(\tau_y)(1 - 2\tau_y)^2], \quad (2.8)$$

where this equation holds in the limit for $y = m$ (and $\tau_y = 1/2$).

Property (iii) states that, like the τ^{th} quantile, the τ^{th} expectile is location and scale equivariant. Most important, (ii) and (iv) together imply that the function $\mu(\tau)$ characterizes the distribution of Y. The range of $\mu(\tau)$ is I_F by (ii), and for any y in I_F equation (2.8) gives an expression for $F(y)$ in terms of $\mu(\tau)$ and its derivative.

We see from Theorem 1 that expectiles have properties that are similar to quantiles. It might also be useful to have some idea of how expectiles behave for some common distributions. In Figure 1 we plot the quantile and the expectile functions for the standard normal distribution. We see that the expectile function has a smaller slope than the quantile function near $\tau = .5$ and a larger slope than the quantile function near $\tau = 0$ or $\tau = 1$. The expectile function for the uniform distribution on the unit interval, which can be shown to be $\mu(\tau) = [\tau - \sqrt{\tau(1-\tau)}]/(2\tau - 1)$ by using equation (2.7), also exhibits similar behavior.

In the regression case the vector $\beta(\tau)$ that minimizes $R(\beta,\tau) \equiv E[\rho_\tau(y_i - x_i'\beta) - \rho_\tau(y_i)]$, which is a population version of $R_n(\beta,\tau)$, will be determined by the conditional distribution of y_i given x_i. The first order conditions for this minimization problem can be shown to imply that $\beta(\tau)$ is a solution of the equation

$$(2.9) \qquad \beta(\tau) = \{E[|\tau - 1(y_i < x_i'\beta(\tau))|x_i x_i']\}^{-1} E[|\tau - 1(y_i < x_i'\beta(\tau))|x_i y_i].$$

In general, the conditional τ^{th} expectile of y_i will be a function $\mu(\tau,x_i)$ that minimizes $E[\rho_\tau(y_i - m) - \rho_\tau(y_i)|x_i]$ over m for almost all x_i, and $x_i'\beta(\tau)$ will be a linear (in x_i) approximation to $\mu(\tau,x_i)$. As in Theorem 1 and the accompanying discussion it can be shown that $\mu(\tau,x_i)$ characterizes the conditional distribution of y_i given x_i, although the linear approximation $x_i'\beta(\tau)$ need not.

In some cases $x_i'\beta(\tau)$ has a simple relationship to $\mu(\tau,x_i)$, and this relationship can provide useful information about the conditional distribution of y_i. If u_i is independent of x_i in equation (2.1), so that only the location of y_i depends on x_i, then by property (iii) we have $\mu(\tau,x_i) = x_i'\beta_0 + \mu(\tau)$, where $\mu(\tau)$ is the τ^{th} expectile of u_i. Since the conditional τ^{th} expectile is linear in this case, it follows that

$$(2.10) \qquad \beta(\tau) = \beta_0 + \mu(\tau)e_1, \qquad e_1 \equiv (1, 0, \ldots, 0)',$$

so that changing τ only changes the intercept term in $\beta(\tau)$.

When the scale of y_i also depends linearly on x_i, say $u_i = (x_i'\gamma_0)\varepsilon_i$ and ε_i and x_i independent, then $\mu(\tau,x_i) = x_i'\beta_0 + \mu(\tau)x_i'\gamma_0 = x_i'[\beta_0 + \mu(\tau)\gamma_0]$, where $\mu(\tau)$ is the τ^{th} expectile of ε_i. It follows that

$$(2.11) \qquad \beta(\tau) = \beta_0 + \mu(\tau)\gamma_0.$$

When the scale of y_i varies with x_i, so that heteroskedasticity is present in the regression equation, it follows from equation (2.11) that the slope coefficients in $\beta(\tau)$ also vary with τ. As with regression quantiles, heteroskedasticity can be detected by checking whether or not the slope coefficients in a set of ALS estimators vary with τ. Note that this specification does not restrict the way in which the exogenous variables affect the scale, except that functions of the exogenous variables must enter linearly. We can always redefine the original regression vector to include, say, nonlinear functions of a set of "original" regressors.

Asymmetric least squares coefficients also provide information about symmetry of the conditional distribution of y_i given x_i. Symmetry is an important property of the conditional distribution of y_i because, in the absence of symmetry or homoskedasticity, conclusions about how the location of the distribution of y_i varies with x_i may depend on the choice of a location measure (e.g. mean versus median). Also, it is possible to obtain efficient adaptive estimators of β_0 if symmetry holds (Manski (1984), Newey (1986)). One can check for asymmetry by using the type of specification test considered by Hausman (1978), with a comparison of two regression estimators obtained using different location measures, such as least squares and least absolute deviations. One can also use the following result on the pattern of the asymmetric least squares coefficients to detect asymmetry.

Theorem 2: If the distribution of y_i conditional on x_i is symmetric around $x_i'\beta_0$ with probability one, then

(2.12) $$[\beta(\tau) + \beta(1-\tau)]/2 = \beta_0.$$

In other words, if y_i is symmetrically distributed around $x_i'\beta_0$ then, as a function of τ, $\beta(\tau)$ will be symmetric around β_0, which is equal to $\beta(1/2)$. Note that this result holds even if $\mu(\tau,x_i)$ is nonlinear in x_i for $\tau \neq 1/2$. An analogous result can be shown to hold for regression quantiles.

Misspecification of the regression function may also affect the asymmetric least squares coefficients $\beta(\tau)$, since in general $x_i'\beta(\tau)$ is an approximation to the actual conditional weighted mean function $\mu(\tau,x_i)$. For example, it is well known that misspecification of the regression function can induce heteroskedasticity, so that the slope coefficients of $\beta(\tau)$ might vary with τ due to, for example, the presence of an omitted variable.

An advantage of ALS estimators relative to regression quantiles is that the loss function $\rho_\tau(\lambda)$ is continuously differentiable in λ, so that the estimators $\hat{\beta}(\tau)$ can be computed as iterated weighted least squares estimators, i.e., as the solution to the equation

(2.13) $$\hat{\beta}(\tau) = [\Sigma_{i=1}^{n}|\tau-1(y_i<x_i'\hat{\beta}(\tau))|x_ix_i']^{-1}\Sigma_{i=1}^{n}|\tau-1(y_i<x_i'\hat{\beta}(\tau))|x_iy_i.$$

Furthermore, and perhaps more importantly, consistent estimation of the joint asymptotic covariance matrix of several ALS estimators does not require estimation of the density function of the error terms, as discussed below. Unlike regression quantiles, the estimated covariance matrix will involve no "smoothing" of the empirical distribution or quantile function of the estimated residuals. These convenient properties of asymmetric least squares, along with its relatively favorable performance in the efficiency comparisons of Section 4, suggest that asymmetric least squares merits consideration for use in practice.

The disadvantage of asymmetric least squares relative to regression quantiles is that expectiles may be more difficult to interpret than quantiles. This should not be considered to be a serious disadvantage however, since one can obtain a rough idea concerning the location of a particular ALS estimator $\hat{\beta}(\tau)$ in the conditional distribution of y_i given x_i by calculating the proportion of observations for which $y_i < x_i'\hat{\beta}(\tau)$. In the case where x_i is independent of u_i this proportion will be a consistent estimator of $F(\mu(\tau))$, where $F(u)$ is the c.d.f. of u_i and $\mu(\tau)$ is the τ^{th} expectile of u_i.

3. Large Sample Properties of Asymmetric Least Squares Estimators

The asymptotic theory for the asymmetric least squares estimators and test statistics will be developed under the following assumptions. Let ℓ denote the Lebesgue measure on the real line and let $z \equiv (y, x')$, where x is a $p \times 1$ vector.

Assumption 1: For each sample size n, $z_i = (y_i, x_i')$, $(i=1,\ldots,n)$, is i.i.d. and for γ_n in $\mathbb{R}^q$, z_i has a probability density function $f(y_i|x_i,\gamma_n)g(x_i)$ with respect to a measure $\mu_z = \ell \times \mu_x$ such that $\gamma_n = \gamma_0 + \delta/\sqrt{n}$. Also, the conditional density $f(y|x,\gamma_0)$ is continuous in y for almost all x.

Let $E[\cdot|\gamma]$ denote the expectation taken at $f(y|x,\gamma)g(x)$, and let $E[\cdot] \equiv E[\cdot|\gamma_0]$. Also, let $\psi_\tau(\lambda) \equiv |\tau - 1(\lambda < 0)|\cdot\lambda$.

Assumption 2: There is an open set Γ containing γ_0 such that for almost all z, the conditional density $f(y|x,\gamma)$ is continuous in γ on Γ. Also, $E[x_i\psi_\tau(y_i - x_i'\beta(\tau))|\gamma]$ is continuously differentiable in γ on Γ.

For a matrix $A = [a_{ij}]$, let $|A| \equiv \max_{i,j}|a_{ij}|$.

Assumption 3: There is a constant $d > 0$ and a measurable function $\alpha(z)$ that satisfy $\sup_\Gamma f(y|x,\gamma) \leq \alpha(z)$ and

$$(3.1) \qquad \int |z|^{4+d}\alpha(z)g(x)d\mu_z < +\infty, \quad \int \alpha(z)g(x)d\mu_z < +\infty.$$

Assumption 4: $E[x_i x_i']$ is nonsingular.

Assumption 1 specifies that the data are i.i.d.. We make the identically distributed assumption for ease of interpretation of $\beta(\tau)$, since without this assumption the $\beta(\tau)$ that minimizes $R(\beta,\tau)$ would depend on i. It should also be noted that without identically distributed observations the estimator of the asymptotic covariance matrix of $\hat{\beta}(\tau)$ given below need not be consistent (see White (1983)). Of course, the i.i.d. assumption does not restrict the way that the conditional distribution of y_i depends on x_i, so that, for example, conditional heteroskedasticity is allowed.

Assumption 1 also specifies that the data are generated by a sequence of local alternatives to $f(y|x,\gamma_0)g(x)$. We make this assumption in order to discuss the asymptotic efficiency of various test statistics based on ALS estimators. Of course, the case where the data are generated by a fixed distribution is included as a special case when $\delta = 0$.

Assumption 3 requires that slightly higher than fourth moments of y_i and x_i are bounded uniformly in γ.

For a vector of weights $(\tau_1, \ldots, \tau_m)'$, let $\hat{\xi} \equiv \mathrm{vec}[\hat{\beta}(\tau_1),\ldots,\hat{\beta}(\tau_m)]$ denote the vector of ALS estimators and let $\xi \equiv \mathrm{vec}[\beta(\tau_1),\ldots,\beta(\tau_m)]$ be the population counterpart. For $u_i(\tau) \equiv y_i - x_i'\beta(\tau)$ and $w_i(\tau) \equiv |\tau - 1(u_i(\tau)<0))|$, let

$$W_j \equiv E[w_i(\tau_j)x_ix_i'], \quad W \equiv \mathrm{diag}[W_1, \ldots, W_m],$$

$$V_{jk} \equiv E[w_i(\tau_j)w_i(\tau_k)u_i(\tau_j)u_i(\tau_k)x_ix_i'], \quad V \equiv [V_{jk}], \ (j,k=1,\ldots,m),$$

$$G_j \equiv \partial E[w_i(\tau_j)u_i(\tau_j)x_i|\gamma_0]/\partial\gamma, \quad G \equiv [G_1', \ldots, G_m']',$$

where V is partitioned conformably with ξ. The asymptotic distribution of $\hat{\xi}$ is given in the following result.

Theorem 3: If Assumptions 1 - 4 are satisfied then for each τ in $(0,1)$ a unique solution $\beta(\tau)$ to equation (2.9) exists. Also,

$$(3.2) \qquad \sqrt{n}(\hat{\xi} - \xi) \xrightarrow{d} N(W^{-1}G\delta, \ W^{-1}VW^{-1}).$$

In order to use Theorem 3 to construct large sample confidence intervals or hypothesis tests, a consistent estimator of the asymptotic covariance matrix of $\hat{\xi}$ must be constructed. Unlike the regression quantile estimators, natural, sample moment estimators of the components of this asymptotic covariance matrix can be constructed using the asymmetric least squares residuals $\hat{u}_i(\tau) \equiv y_i - x_i'\hat{\beta}(\tau)$ and the estimated weights $\hat{w}_i(\tau) \equiv |\tau - 1(\hat{u}_i(\tau) < 0)|$. Let

$$\hat{W}_j \equiv \Sigma_{i=1}^n \hat{w}_i(\tau_j) x_i x_i'/n, \quad \hat{W} = \mathrm{diag}[\hat{W}_1, \ldots, \hat{W}_m],$$
$$\hat{V}_{jk} \equiv \Sigma_{i=1}^n \hat{w}_i(\tau_j)\hat{w}_i(\tau_k)\hat{u}_i(\tau_j)\hat{u}_i(\tau_k) x_i x_i'/n, \quad \hat{V} = [\hat{V}_{jk}], \ (j,k=1,\ldots,m).$$

The sample moment estimator of the asymptotic covariance matrix $W^{-1}VW^{-1}$ of $\hat{\xi}$ is $\hat{W}^{-1}\hat{V}\hat{W}^{-1}$.

Theorem 4: If Assumptions 1 - 4 are satisfied then

$$\text{(3.3)} \qquad \hat{W}^{-1}\hat{V}\hat{W}^{-1} \xrightarrow{P} W^{-1}VW^{-1}.$$

This covariance matrix estimator is straightforward to compute. If $\hat{\beta}(\tau)$ is interpreted as a weighted least squares estimator, as in equation (2.13), then $\hat{W}^{-1}\hat{V}\hat{W}^{-1}$ is simply the generalization of the Eicker (1967), White (1980) heteroskedasticity consistent covariance matrix for the vector $\hat{\xi}$ of weighted least squares estimators.

Theorems 3 and 4 allow one to form large sample confidence intervals or hypothesis tests concerning the population values of various asymmetric least squares parameters $\beta(\tau)$. It would be useful to know how inference procedures based on ALS estimators compare with other inference procedures. For example, we would like to know how tests for heteroskedasticity based on ALS estimators compare with other tests for heteroskedasticity, such as the regression quantiles test presented by Koenker and Bassett (1982). One way to make such a comparison is to compare the local power of tests that utilize ALS estimators with the local power of other test statistics. To make this comparison we need to specify the form of tests based on ALS estimates and obtain their distribution under a sequence of local alternatives.

Consider the general linear hypothesis

$$(3.4) \qquad H_0: \; H\xi = h.$$

Two particular hypotheses of interest are homoskedasticity and symmetry. As discussed in Section 2, heteroskedasticity can be detected by checking for differences in the vector of slope coefficients across different weights. Suppose that $(\tau_1, \ldots, \tau_m)$ is ordered so that $0 < \tau_1 < \ldots < \tau_m < 1$. As pointed out by Koenker and Bassett (1982), for this case we have $h = 0$, and the matrix H can be written as

$$(3.5) \qquad H = \Delta^h \otimes \Psi,$$

where Δ^h is an $(m-1) \times m$ matrix with typical element $\Delta^h_{ij} = \delta_{ij} - \delta_{i(j-1)}$, δ_{ij} is the Kronecker delta, and $\Psi = [0, I_{p-1}]$. Also, as discussed in Section 2, nonsymmetry can be detected by checking whether symmetrically placed ALS estimators average up to the least squares estimator. For this case suppose that m is odd and that for j^* equal to the median of $(1, \ldots, m)$,

$$\tau_{j^*} = 1/2, \quad \tau_j = 1 - \tau_{2j^*-j}, \quad 0 < j < j^*.$$

Then $h = 0$ and H can be written as

$$(3.6) \qquad H = \Delta^s \otimes \Psi, \quad \Delta^s = [I_{(m-1)/2}, -2e_{(m-1)/2}, I_{(m-1)/2}],$$

where Ψ is a selection matrix. For the symmetry test we usually have $\Psi = I_p$, but occasionally another choice of Ψ might be appropriate. For example, if it is known a priori that the disturbances are i.i.d. but possibly not symmetrically distributed, then Ψ can be chosen to pick out only the intercept term.

To compare the local power of asymmetric least squares tests of homoskedasticity and symmetry with other tests we need to be more specific about the form of the data generating process. We will restrict attention to local heteroskedasticity and asymmetry that is linear in x_i.

Assumption 5: The observations satisfy $y_i = x_i'\beta_0 + u_i$, where

(3.7) $$u_i = \sigma_i \varepsilon_i, \quad \sigma_i = 1 + x_i'\gamma_{nh} + 1(\varepsilon_i > 0)x_i'\gamma_{ns},$$

where $\gamma_{nh} = \delta_h/\sqrt{n}$, $\gamma_{ns} = \delta_s/\sqrt{n}$, and ε_i is i.i.d., independent of x_i, and symmetrically distributed around zero. Also, ε_i has the c.d.f. $F(\varepsilon)$, which has a continuous density $f(\varepsilon)$.

This assumption specifies that the data is generated by a sequence of local alternatives to a model with i.i.d., symmetric disturbances. If $\delta_h \neq 0$ and $\delta_s = 0$ then we have the local heteroskedastic alternative considered by Koenker and Bassett (1982). If $\delta_h = 0$ and $\delta_s \neq 0$ then we have a non-symmetry alternative like that considered by Antille, Kersting, and Zucchini (1982) and Boos (1982), where the effect of x_i on the distribution of y_i is confined to the upper half of the distribution of u_i.

In order to guarantee that the regularity conditions given above are satisfied for the particular data generating process in Assumption 5, it is useful to make the following assumption.

Assumption 6: x_i has compact support. Also there exist finite constants

D, d > 0 such that

$$(3.8) \qquad f(\varepsilon) \leq D/(1 + |\varepsilon|^{5+d}).$$

The assumption that x_i has compact support is difficult to dispense with in the presence of linear heteroskedasticity.

The vector of ALS estimators and the estimated asymptotic covariance matrix can be used to form a test statistic T for the general linear hypothesis in the usual fashion, with

$$(3.9) \qquad T = n(H\hat{\xi} - h)'[H\hat{W}^{-1}\hat{V}\hat{W}^{-1}H']^{-1}(H\hat{\xi} - h).$$

Under a sequence of local alternatives to H_0, T will have a noncentral chi-square distribution. Let

$$(3.10) \qquad c(\tau, \theta) \equiv E[w_i(\tau)u_i(\tau)w_i(\theta)u_i(\theta)], \quad d(\tau) \equiv E[w_i(\tau)],$$

$$\sigma_{jk} \equiv c(\tau_j, \tau_k)/[d(\tau_j)d(\tau_k)], \ (j,k = 1,\ldots,m),$$

and let Σ be the matrix with typical element σ_{ij}. Also, let $D \equiv E[x_i x_i']$, $\mu = (\mu(\tau_1), \ldots, \mu(\tau_m))'$, where $\mu(\tau)$ is the τ^{th} weighted mean of ε_i, and $\nu = (\nu(\tau_1), \ldots, \nu(\tau_m))'$, where

$$(3.11) \qquad \nu(\tau) = [\tau\int_0^\infty \varepsilon f(\varepsilon)d\varepsilon + (1-2\tau)\int_0^{\max\{0,\mu(\tau)\}}\varepsilon f(\varepsilon)d\varepsilon] \times \{\tau[1 - F(\mu(\tau))] + (1-\tau)F(\mu(\tau))\}^{-1}.$$

Theorem 6: Suppose that Assumptions 1, 4, 5, and 6 are satisfied. Also suppose that H_0 is satisfied when $\gamma = \gamma_0$, Σ is nonsingular, and H has full row rank. Then T converges in distribution to a noncentral chi-squared with rank(H) degrees of freedom and noncentrality parameter

$$(3.12) \qquad (\mu \otimes \delta_h + \nu \otimes \delta_s)'H'[H(\Sigma \otimes D^{-1})H']^{-1}H(\mu \otimes \delta_h + \nu \otimes \delta_s).$$

The local power of asymmetric least squares tests can be compared with that of other tests that have the same degrees of freedom by comparing the respective noncentrality parameters. For the local power comparisons to be considered in the next section it will be useful to have available expressions for the noncentrality parameter of the asymmetric least squares test of homoskedasticity in the absence of local asymmetry and that of the symmetry test in the absence of the local heteroskedasticity. The following result gives these expressions.

Corollary 1: If $H = \Delta^h \otimes \Psi$ and $\delta_s = 0$, then the noncentrality parameter for T is $\kappa^h_{LS} \cdot (\Psi\delta_h)'(\Psi D^{-1}\Psi')^{-1}(\Psi\delta_h)$, where

(3.13) $$\kappa^h_{LS} = (\Delta^h\mu)'(\Delta^h\Sigma\Delta^{h\prime})^{-1}(\Delta^h\mu)$$

Also, if $H = \Delta^s \otimes \Psi$ and $\delta_h = 0$, then the noncentrality parameter for T is $\kappa^s_{LS} \cdot (\Psi\delta_s)'(\Psi D^{-1}\Psi')^{-1}(\Psi\delta_s)$, where

(3.14) $$\kappa^s_{LS} = (\Delta^s\nu)'(\Delta^s\Sigma\Delta^{s\prime})^{-1}(\Delta^s\nu).$$

The noncentrality parameter for the asymmetric least squares test of homoskedasticity under a heteroskedastic alternative that is given in Corollary 1 has a similar form to the noncentrality parameter for the regression quantiles test of homoskedasticity. The matrix Σ is the covariance matrix for a vector of weighted mean estimators. Also, as shown by Koenker and Bassett (1982), the regression quantiles test of homoskedasticity has a limiting noncentrality parameter that is equal to $\kappa^h_{RQ} \cdot (\Psi\delta_h)'(\Psi D^{-1}\Psi')^{-1}(\Psi\delta_h)$, with

(3.15) $$\kappa^h_{RQ} = (\Delta^h\eta)'(\Delta^h\Omega\Delta^{h\prime})^{-1}(\Delta^h\eta),$$

where $\eta = (\eta(\theta_1), \ldots, \eta(\theta_m))'$ is the vector of quantiles, and Ω the asymptotic covariance matrix of the vector of quantile estimators. Thus, κ^h_{RQ}

involves the differences $[\eta(\theta_j)-\eta(\theta_{j-1})]$ of quantiles and the precision with which these differences are estimated, so that by comparing κ^h_{LS} and κ^h_{RQ}, we see that the relative efficiency of the ALS test and RQ test is governed by the sensitivity of quantiles and weighted means to the choice of τ and θ and the precision with which the respective location measures are estimated.

One can also form regression quantile tests of symmetry by using regression quantile estimators and their estimated covariance matrix to form a test statistic, as discussed by Koenker and Bassett (1982), with the H matrix given in equation (3.6). It is straightforward to show that for the asymmetric alternative the regression quantiles test of symmetry has a noncentrality parameter that is equal to $\kappa^s_{RQ}\cdot(\Psi\delta_s)'(\Psi D^{-1}\Psi')^{-1}(\Psi\delta_s)$, with

$$(3.16) \qquad \kappa^s_{RQ} = (\Delta^s\eta^+)'(\Delta^s\Omega\Delta^{s\prime})^{-1}(\Delta^s\eta^+),$$

where $\eta^+ = (\eta(\theta_1)^+, \ldots, \eta(\theta_m)^+)'$ and $\eta(\theta)^+ \equiv \max\{0,\eta(\theta)\}$. As with the homoskedasticity tests, comparison of κ^s_{RQ} and κ^h_{LS} indicates that the relative power of asymmetric least squares and regression quantile tests of symmetry will depend on the sensitivity of the respective location parameters and the precision with which they are estimated.

Finally, it is worth noting that, because of the special form of the Δ^h and Δ^s matrices and the joint covariance matrix Σ of the expectile estimators for values of τ symmetric about $\tau = 1/2$, it can be shown that the ALS test statistics for either homoskedasticity or conditional symmetry are asymptotically independent under either of the null hypotheses (with an analogous result holding for the regression quantile tests.) Thus, significance levels for joint tests of these hypotheses are particularly easy to calculate.

4. Asymptotic Relative Efficiencies of Alternative Tests

In the two subsections below, the efficiencies of tests based on the asymmetric least squares estimators relative to other tests of heteroskedasticity and asymmetry are calculated for the class of contaminated Gaussian error distributions. Section 4.1 compares the ALS test for heteroskedasticity to tests which use the absolute or squared residuals of a preliminary fit of (2.1), and to the regression quantile test of heteroskedasticity. In this context, Koenker and Bassett's (1982) original calculations concerning the latter two tests are revised; due to an algebraic error (described below), their Figures 1 and 2 give a misleading depiction of the relative performance of the tests for this class of error distributions. Section 4.2 discusses how odd functions of residuals can be used to construct tests of symmetry, and compares the performance of such tests to the corresponding tests using asymmetric least squares and regression quantile estimators. A surprising finding is that the ALS test of heteroskedasticity performs virtually identically to Glejser's (1969) absolute residual regression test in terms of local power; another surprising result is that the ALS test for symmetry behaves much like a test based on a comparison of least squares (mean) and least absolute deviations (median) regression coefficients. A general conclusion is that the ALS tests dominate the corresponding RQ tests over a range of error distributions which does not depend on whether homoskedasticity or conditional symmetry is being tested.

4.1 Tests of Homoskedasticity

Here we investigate the local power of tests for heteroskedasticity, assuming conditional symmetry, i.e., $\delta_s = 0$. Following Koenker and Bassett's

(1982) setup, we consider the two-parameter class of contaminated Gaussian distributions, with cumulative distributions of the form

$$(4.1) \qquad F(\lambda \mid \alpha, \sigma) = (1 - \alpha)\cdot\Phi(\lambda) + \alpha\cdot\Phi(\lambda/\sigma) ,$$

for $\Phi(\cdot)$ denoting the standard normal cumulative and for α in the interval (0, 1). For this class of distributions, the τ^{th} weighted mean satisfies

$$(4.2) \qquad \mu(\tau) = \frac{(2\tau - 1)[(1 - \alpha)\varphi(\mu(\tau)) + (\alpha/\sigma)\varphi(\mu(\tau)/\sigma)]}{\tau + (1 - 2\tau)[(1 - \alpha)\Phi(\mu(\tau)) + \alpha\Phi(\mu(\tau)/\sigma)]} ,$$

for $\varphi(\cdot)$ the standard normal density function. To conform to Koenker and Bassett's framework, we consider only the efficiency of the asymmetric least squares test using a single difference of symmetrically chosen weights, i.e., a test based upon $\hat{\beta}(\tau) - \hat{\beta}(1 - \tau)$, for $\frac{1}{2} < \tau < 1$. The corresponding regression quantile test uses $\hat{b}(\theta) - \hat{b}(1 - \theta)$, the difference in symmetric regression quantile estimators, where the same normalization $\frac{1}{2} < \theta < 1$ is imposed. For these tests, the scalars κ^h_{LS} and κ^h_{RQ}, defined in the discussion following Corollary 1 above, can easily be calculated as functions of α and σ, using the special form of the distribution function in (4.1). For example, the term $c(\theta, \tau)$ appearing in the expression for the asymptotic covariance matrix of the ALS estimators is

$$(4.3) \qquad c(\theta, \tau) = \{(1 - \theta)(1 - \tau)G(\mu(\theta)) + \theta(1 - \tau)[G(\mu(\tau)) - G(\mu(\theta))]$$

$$+ \theta\tau[1 - G(\mu(\tau))]\} - \theta(1 - \tau)\cdot\mu(\theta)\cdot\mu(\tau) ,$$

where

$$G(\lambda) \equiv (1 - \alpha)\Phi(\lambda) + \alpha\sigma^2\Phi(\lambda/\sigma) . \tag{4.4}$$

Expressions for the remaining components of κ^h_{LS} and κ^h_{RQ} can be obtained in a similar manner.

Koenker and Bassett compared the scalar κ^h_{RQ} to the corresponding term κ^h_{SR} for a heteroskedasticity test using squared residuals from a preliminary least squares fit of equation (2.1), a test closely related to those investigated by Breusch and Pagan (1979) and White (1980). More generally, tests for heteroskedasticity can be based on the sample correlation of $\ell(\hat{u}_i)$ with the regressors x_i, where $\hat{u}_i \equiv y_i - x_i'\hat{\beta}(\frac{1}{2})$ is the least squares residual and $\ell(\cdot)$ is an even function. To obtain a test with more asymptotic power than the squared residual regression test for (heavy-tailed) nonnormal disturbances, the ℓ function could be chosen to penalize large errors less heavily, e.g., $\ell(u) = |u|^p$ for $1 \leq p < 2$ rather than $p = 2$.

The test statistic for this type of test is

$$T_\ell = nR^2_\ell , \tag{4.5}$$

the sample size n times the constant-adjusted R^2 of the regression of $\ell(\hat{u}_i)$ on x_i. Bickel (1978) has obtained the asymptotic properties of this class of tests when it is assumed that $\delta_h = \beta_o$, but his results can be extended to the more general linear scale model considered in the previous section. With some additional regularity conditions (such as the boundedness of $E[\ell(\varepsilon_1)]^2$) which can be verified for the cases we consider here, the test statistic T_ℓ of (4.5) can be shown to have a limiting noncentral chi-square distribution with $(p - 1)$ degrees of freedom and noncentrality parameter

(4.6) $$[E(\ell'(\varepsilon_1)e_1)]^2[\mathrm{Var}(\ell(\varepsilon_1))]^{-1}(\Psi\delta_h)'(\Psi D^{-1}\Psi')(\Psi\delta_h)$$

$$\equiv \kappa_\ell^h \cdot (\Psi\delta_h)'(\Psi D^{-1}\Psi')(\Psi\delta_h)$$

under the conditions given in section 3.2.

In our application we focus attention on the squared residual regression test ($\ell(u) \equiv u^2$) and the more "robust" test which uses absolute residuals (i.e., $\ell(u) \equiv |u|$). For the former test, the scalar $\kappa_\ell^h \equiv \kappa_{SR}^h$ is

(4.7) $$\kappa_{SR}^h = 4[\{3(1 + \alpha(\sigma^4 - 1))/(1 + \alpha(\sigma^2 - 1))^2\} - 1]^{-1}$$

when the errors are contaminated Gaussian, while for the absolute residual regression test, $\kappa_\ell^h \equiv \kappa_{AR}^h$ is given as

(4.8) $$\kappa_{AR}^h = [\{\pi(1 + \alpha(\sigma^2 - 1))/2(1 + \alpha(\sigma - 1))^2\} - 1]^{-1} .$$

The local power of thehe squared residual regression, absolute residual regression, regression quantile, and asymmetric least squares tests may be compared by computing their Pitman asymptotic relative efficiencies (AREs); since the limiting degrees of freedom for all of these test statistics are equal, the AREs are just the ratios of the respective noncentrality parameters, which in turn reduce to the ratios of the respective κ coefficients. However, the noncentrality parameters of the regression quantiles and asymmetric least squares tests depend upon the particular weights (θ and τ, respectively) chosen. Rather than considering the AREs for these tests for a range of weights, we consider only the weights $[1 - \theta, \theta] = [.13, .87]$ for the regression quantiles test and $[1 - \tau, \tau] = [.46, .54]$ for the asymmetric least squares test. These values of θ and τ were selected after a preliminary calculation of the weights which maximized the respective noncentrality parameters in a grid search for a selection of α and σ combinations; the results of this optimization for tests of homoskedasticity are given in Table 1 below. As the table shows, the optimal θ values for the regression quantiles test are typically between .75 and .90, and decrease as α and σ increase (although there is a sharp reversal in this pattern for values of α near .50). The optimal values of τ for the asymmetric least squares test are usually between .51 and .75, and also typically decrease with increasing α and σ. A simple arithmetic average of the optimal values in Table 1 and Table 3 below (which contains the corresponding optimal weights for tests of conditional symmetry) suggests $\theta = .87$ and $\tau = .54$ are reasonable choices for the respective weights, so these average values were used in this and the following subsection.

It is important to note that the value of the noncentrality parameter is usually quite insensitive to moderate perturbation of the weights from their optimal values. For example, for the regression quantiles test, when $\alpha = .05$ and $\sigma = 5$, use of $\theta = .87$ rather than the optimal $\theta = .89$ results in an efficiency loss of only two percent (although for $\alpha = 0$, the efficiency loss rises to 10 percent, with optimal $\theta = .93$).

Table 2 gives the AREs of the regression quantile, asymmetric least squares, and absolute residual regression tests, all relative to the squared residual regression test. One striking feature of this table is the nearly identical performance of the absolute residual regression test and the asymmetric least squares test. While it is not surprising that these two tests should have qualitatively similar performance (since their respective κ functions are both determined by the efficiency with which the first moment of the residuals can be estimated), the fact that the ARE of the asymmetric least squares test never differs from the ARE of the absolute residual regression test by more than one percent was unexpected. Of course, this outcome is more than coincidental; it can be shown (with some tedious algebra) that the limiting value of the asymmetric least squares noncentrality parameter as $\tau \to 1/2$ is identical to the absolute residual regression noncentrality parameter. Moreover, this result is not special to the family of contaminated normal error distributions considered here, but holds for any symmetric error distribution satisfying the regularity conditions of the previous section. Hence, the large number of values of $\tau \equiv .51$ in Table 1 suggests that for most (but not all) of the distributions considered here, the highest possible efficiency of the expectile-based test is attained by the absolute residual regression test.

Table 2 shows that both the asymmetric least squares and absolute residual regression tests are more efficient than the squared residual regression test except when α and σ are large (or when $\alpha = 0$, in which case the squared residual regression test is locally most powerful). The ARE of the asymmetric least squares test is small for $\sigma = 2$, but increases substantially as σ increases.

Another interesting feature of Table 2 is the behavior of the AREs of the regression quantile test. For $\sigma = 2$ the squared residual regression test is always more efficient than the regression quantile test, and for $\sigma = 3$ the asymmetric least squares (or absolute residual regression) test is efficient relative to the regression quantile test. For $\sigma = 4$ and 5, the regression quantile test is the most efficient of all tests considered when α is between

5 and 20 percent; for $\sigma = 4$, however, its efficiency gain over the asymmetric least squares test is not particularly large, amounting, for example, to 29 percent at $\alpha = .10$.

These results on the ARE of the regression quantile test relative to the squared residual regression test are quite different from those reported in Koenker and Bassett (1982). For example, when $\theta = .75$, the relative scale $\sigma = 5$, and there is 20 percent contamination, we find the ARE of the regression quantile test to be 1.64, rather than the "40+" figure reported previously. This difference is explained by an error in equations (4.12) and (4.14) of Koenker and Bassett (1982);[4] the term corresponding to κ^h_{SR} in these expressions is "$4[\mathrm{Var}(\varepsilon_i^2)]^{-1}$" instead of the correct $\kappa^h_{SR} = 4[E(e_i)^2]^2[\mathrm{Var}(\varepsilon_i^2)]^{-1}$. The omitted term overstates the ARE of the regression quantile test for $\sigma > 1$, particularly when the contamination percentage α is large; hence the "iso-efficiency" contours of Figures 1 and 2 of Koenker and Bassett (1982) should actually be shifted upward and "U"-shaped, with the ARE of the regression quantile test sharply declining as the distinction between the "contaminating" and "contaminated" distributions of the error vanishes.

It should be noted, however, that for sufficiently large σ and sufficiently small α, dramatic efficiency gains of the regression quantile test to the other procedures are attainable. For example, for $\alpha = .0125$ and $\sigma = 10$, the ARE of the regression quantile test is 21.33, over twice as large as that for the asymmetric least squares and absolute residual regression tests; this improvement, though, drops off quite rapidly as α increases. Thus the regression quantile test should perform very well for large data sets which contain a few sizable outliers.

4.2 Tests of Conditional Symmetry

Turning now to the null hypothesis of conditional symmetry of the distribution of u_i about zero, we consider for simplicity only the case with $\delta_s \neq 0$ but $\delta_h = 0$ (that is, the potential heteroskedasticity is confined to the "positive half" of the error distribution). Again, we restrict attention to the family of contaminated Gaussian distributions given in (4.1) above, and evaluate the relative local powers of the tests for the same range of α and σ.

Using the result of Theorem 2, a test of conditional symmetry using asymmetric least squares (or regression quantile) estimators can be based on the "symmetric second difference" $\beta(\tau) + \beta(1-\tau) - 2\cdot\beta(\frac{1}{2})$, which is zero under the null hypothesis of symmetry. In the notation of the discussion following Corollary 1, the Δ matrix would represent the single contrast $[1, -2, 1]$ applied to the matrix $[\hat{\beta}(1-\tau), \hat{\beta}(\frac{1}{2}), \hat{\beta}(\tau)]$, while Ψ can be an arbitrary selection matrix, whose rank determines the degrees of freedom of the corresponding test statistic (thus, if it is known a priori that δ_s is proportional to e^1 -- that is, that the error terms are i.i.d. but possibly asymmetrically distributed -- then Ψ can be chosen to pick out only the intercept term of the contrast). Calculation of the scalars κ_{LS}^{s} and κ_{RQ}^{s} proceeds as in the previous section; the coefficient $\nu(\tau)$, for example, can be computed using the formula

$$(4.9) \qquad \int_a^b \lambda \, dF(\lambda|\alpha, \sigma) = (1-\alpha)[\varphi(a) - \varphi(b)] + \alpha\sigma[\varphi(a/\sigma) - \varphi(b/\sigma)].$$

Table 3 gives results analogous to those in Table 1 of the previous subsection; that is, values of θ and τ in the interval $(\frac{1}{2}, 1)$ which maximize the respective noncentrality parameters are computed for a range of contaminated Gaussian distributions. The optimal values of the weights for tests of symmetry behave quite similarly to those for tests of heteroskedasticity; in comparison to Table 1, values of τ for the ALS test in Table 3 tend to be somewhat closer to $\frac{1}{2}$, and values of θ for the RQ test tend to be slightly closer to 1, but the differences are typically minor. In calculating the relative efficiencies of the RQ and ALS test, the values $\tau = .54$ and $\theta = .87$ were chosen by averaging the values in Tables 1 and 3, as discussed above.

In the previous subsection, the ALS and RQ tests of homoskedasticity were compared to tests based on a regression of an even function $\ell(\hat{u}_i)$ of preliminary least squares residuals on the regressors. Similar tests of conditional symmetry can be constructed by regressing odd functions of residuals -- that is, functions of the form $\ell(\hat{u}_i)\text{sgn}(\hat{u}_i)$, where $\ell(\cdot)$ is an even function as above -- on x_i. However, unlike the "residual regression" tests for heteroskedasticity, the distribution theory for such tests for asymmetry depends on the method of estimation in the "first-stage" regression. For example, taking $\ell(\lambda) = |\lambda|$, regression of $\ell(\hat{u}_i)\text{sgn}(\hat{u}_i) \equiv \hat{u}_i$ on x_i cannot detect conditional asymmetry, since the least squares residuals are orthogonal to the regressors by construction. Nevertheless, for other odd functions of the least squares residuals "residual regression" test statistics for asymmetry can be constructed, which have limiting noncentral chi-square distributions under the sequence of local alternatives. Letting $\hat{\Psi}_\ell$ denote the second-stage coefficient estimates of $\ell(\hat{u}_i)\text{sgn}(\hat{u}_i)$ on x_i, it can be shown that $\hat{\Psi}_\ell$ will satisfy the "asymptotic linearity" relationship

$$(4.10) \qquad \sqrt{n}\,\hat{\Psi}_{\ell} = D^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \operatorname{sgn}(\varepsilon_i)\cdot\{\ell(\varepsilon_i) - [E\,\ell'(\varepsilon_i)]\cdot|\varepsilon_i|\}\right.$$

$$\left. + \delta_s\cdot\operatorname{Cov}[|\ell'(\varepsilon_i)|,\ \max\{0,\ \varepsilon_i\}] + o_p(1)\right]$$

under the local alternatives considered here (and appropriate regularity conditions). For the special case $\ell(\lambda) = \lambda^2$, the corresponding vector $\hat{\Psi}_{SR}$ of regression coefficients of the "signed squared residuals" on x_i will have the asymptotic distribution

$$(4.11) \qquad \sqrt{n}\,\hat{\Psi}_{SR} \xrightarrow{d} N\left(\delta_s\cdot\operatorname{Var}(|\varepsilon_i|),\ E[\varepsilon_i(|\varepsilon_i| - 2E|\varepsilon_i|)]^2\cdot D^{-1}\right) .$$

Though the R^2 from the second-stage regression is not asymptotically chi-squared (the denominator overestimates the variability of $\hat{\Psi}_{SR}$ under the null and local alternative hypotheses), a Wald test of $\delta_s = 0$ (or $\Psi\delta_s = 0$, as discussed above) can be constructed in a straigtforward fashion using $\hat{\Psi}_{SR}$. For this test, the scalar κ^s_{SR} governing the local power is given explicitly as

$$(4.12) \qquad \kappa^s_{SR} = [\,3m_4 - (16/\pi)m_3m_1 + (8/\pi)m_2(m_1)^2\,]^{-1}\cdot[\,m_2 - (2/\pi)(m_1)^2\,]^2,$$

for the contaminated Gaussian distributions, where

$$(4.13) \qquad m_j = m_j(\alpha,\ \sigma) \equiv 1 - \alpha(1 - \sigma^j) \quad \text{for any integer } j.$$

While a symmetry test analogous to the "absolute residual regression" test for heteroskedasticity is not available (for the reasons given above), another test for conditional asymmetry can be based on the difference between least squares and least absolute deviations estimates of β_o. Under the null hypothesis, the asymptotic distribution of this difference should be centered at zero (since both the conditional mean and median of y_i will be $x_i'\beta_o$ if the errors are conditionally symmetric about zero), but the difference can be expected to differ from zero under the alternative of asymmetry. The asymptotic distribution of $\hat{\beta}(\frac{1}{2}) - \hat{b}(\frac{1}{2})$ is

$$(4.14) \qquad \sqrt{n}\left[\hat{\beta}(\tfrac{1}{2}) - \hat{b}(\tfrac{1}{2})\right] \xrightarrow{d} N\left(\tfrac{1}{2}\,\delta_s \cdot E[|\varepsilon_i|],\ E(\varepsilon_i - [2f(0)]^{-1}\mathrm{sgn}(\varepsilon_i))^2 \cdot D^{-1}\right)$$

under the local alternative hypothesis, where $f(\cdot)$ denotes the density function of ε_i; this can be obtained using the "asymptotic linearity" relationship

$$(4.15) \qquad \sqrt{n}\left[\hat{b}(\tfrac{1}{2}) - \beta_o\right] = D^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [2f(0)]^{-1}\mathrm{sgn}(u_i)x_i + o_p(1)$$

(see, for example, Koenker and Bassett (1982)). Except for estimation of the density function of ε_i at zero, construction of a Wald-type test statistic for the null hypothesis $\hat{\beta}(\frac{1}{2}) - \hat{b}(\frac{1}{2}) = 0$ is straightforward, and $f(0)$ might be consistently estimated using, say, kernel-type estimation methods applied to the residuals. The noncentrality parameter for this "median versus mean" regression test of symmetry will be of the same general form as for the foregoing tests, with

$$(4.16) \qquad \kappa_{MM}^{s} = (2\pi)^{-1} m_1^2 \cdot \left[m_2 - 2m_1 m_{-1} + (\pi/2)(m_{-1})^2\right]$$

for the contaminated Gaussian family, where m_j is defined in (4.13) above.

Table 4 gives the relative efficiencies of the regression quantile, asymmetric least squares, and "median versus mean" tests of conditional symmmetry of the error distribution, all relative to the "signed squared residual" test, for the same range of error distributions investigated in Table 2. In comparison with Table 2, it is clear that the performance of the residual regression test of asymmetry is inferior to its heteroskedasticity counterpart; it only dominates the ALS test in the two extreme cases ($\alpha = 0$ and $\alpha = .5$, $\sigma = 5$), and then by very little. The relative performance of the RQ test is also improved relative to the SRR test. For tests of symmetry, it appears that the dependence of residual regression tests on the asymptotic distribution of the first-stage estimator makes such tests much less attractive, particularly to the extent that the loss function $\ell(\cdot)$ is well approximated by an absolute value function.

Comparison of the ALS to the RQ test of symmetry reveals the same pattern of relative performance as for the respective tests of homoskedasticity. This similarity can be more easily seen in Table 5, which gives the relative efficiency of the ALS to the RQ test directly for tests of both hypotheses. This table shows that, overall, the relative performance of the tests depends only on the nature of the underlying error distribution, and not on the particular null hypothesis (homoskedasticity or conditional symmetry) of interest. The general conclusions of section 4.2 -- that the ALS test is preferred except for distributions with small probabilities of large contamination -- thus applies in this circumstance.

The most striking feature of Table 4 is the similarity of the last two columns, which correspond to the ALS and "median versus mean" tests. While the "median versus mean" test is typically more efficient than the ALS test, the magnitude of this difference is no more than five percent throughout. As in the previous section, this result suggests that the local power of these tests is governed by the precision to which the mean of ε_i can be estimated; we conjecture that, as in the previous subsection, the "median versus mean" noncentrality parameter is the limiting value of the ALS noncentrality parameter as $\tau \to 1/2$, though we have not verified it algebraically (which would require a four-fold application of L'Hôpital's rule). In practical terms, the slight efficiency advantage of the "median versus mean" test would be outweighed by the computational burden of least absolute deviations estimation, and, more importantly, by the need to estimate the density function of the residuals to construct the Wald test statistic.

5. Conclusions

From the results of the previous section, we conclude that tests of homoskedasticity and conditional symmetry based upon asymmetric least squares coefficient estimates are reasonably efficient over a wide range of error distributions, relative to the other test procedures considered. Furthermore, the ALS coefficient estimators are of interest in their own right, as useful summary statistics of the conditional distribution of y_i given x_i. Rejection of the null hypotheses of homoskedasticity and symmetry using the ALS estimators with "optimal" weights indicates that the least squares coefficient estimates do not adequately characterize the relationship of the dependent variable to the regressors; computation of $\hat{\beta}(\tau)$ for other values of τ, say, τ = .25 and .75 or τ = .15 and .85 (roughly corresponding to the 33rd/66th or 25th/75th percentiles for Gaussian errors), would give a more complete picture of this relationship.

Mathematical Appendix

Proof of Theorem 1: Let $T_F(\mu) \equiv \int_{[\mu,\infty)}(y-\mu)dF(y)$ and $\alpha(\tau) \equiv (2\tau-1)/(1-\tau)$. As discussed by DeGroot (1970, p. 246), $T_F(\mu)$ is a convex function of μ (and is therefore continuous in μ) and satisfies

$$\text{(A.1)} \quad T_F(\mu) \geq m-\mu, \quad \lim_{\mu\to\infty} T_F(\mu) = 0, \quad \lim_{\mu\to-\infty}[T_F(\mu)-(m-\mu)] = 0.$$

Also, for τ in $(0,1)$, $\alpha(\tau)$ satisfies

$$\alpha(\tau) > -1, \quad d\alpha(\tau)/d\tau = 1/(1-\tau)^2 > 0.$$

It follows that $\mu-m$ is greater (smaller) than $\alpha(\tau)T_F(\mu)$ for μ large (small) enough, so that a solution to equation (2.7) exists by the intermediate value theorem. Also, any such solution must be unique because the convexity of $T_F(\mu)$ and (A.1) imply that for $\mu' > \mu$, $0 \geq T_F(\mu')-T_F(\mu) \geq -(\mu'-\mu)$ (i.e. $T_F(\mu)$ is monotonic decreasing and has a "slope" of at least -1).

The fact that $\mu(\tau)$ is strictly monotonic increasing follows from $\alpha(\tau)$ strictly monotonic increasing in τ and $T_F(\mu)$ monontonic decreasing in μ.

The fact that $\mu(\tau)$ must lie in I_F follows from m an element of I_F, and from $T_F(\mu) = 0$ for μ greater than any element of I_F, while if μ is less than any element of I_F then $\alpha(\tau)T_F(\mu) = \alpha(\tau)(m-\mu) > (-1)(m-\mu) = \mu-m$. To see that $\mu(\tau)$ is onto I_F, note first that $\mu(1/2) = m$. Also, if μ is an element of I_F and $\mu > m$, then $T_F(\mu) > 0$, so that by $\lim_{\tau\to 1}\alpha(\tau) = +\infty$, $\mu(\tau) = \mu$ for some τ in $(0,1)$, while if $\mu < m$ then $T_F(\mu) > m-\mu$, so that by $\alpha(0) = -1$, $\mu(\tau) = \mu$ for some τ in $(0,1)$.

To show the location and scale equivariance of $\mu(\tau)$, note that the mean of Y is $sm+t$, and that by a change of variables $T_F(\mu) =$

$T_F^{\sim}(s\mu+t)/s$. It follows that $s\mu(\tau)+t$ satisfies

$$(A.2) \qquad (s\mu(\tau)+t)-(sm+t) = s\alpha(\tau)T_F(\mu(\tau)) = \alpha(\tau)T_F^{\sim}(s\mu(\tau)+t),$$

which is the defining equation for $\tilde{\mu}(\tau)$.

When $F(y)$ is continuously differentiable $T_F(\mu)$ is a continuously differentiable function of μ. Continuous differentiability of $\mu(\tau)$ then follows from the implicit function theorem. Differentiating both sides of equation (2.7) gives

$$(A.3) \qquad \mu'(\tau) = \alpha'(\tau)T_F(\mu(\tau)) - \alpha(\tau)[1-F(\mu(\tau))]\mu'(\tau)$$
$$= [\alpha'(\tau)/\alpha(\tau)][\mu(\tau)-m] - \alpha(\tau)[1-F(\mu(\tau))]\mu'(\tau).$$

Equation (2.8) is obtained by solving for $F(\mu(\tau))$.

Proof of Theorem 2: It follows from y_i symmetrically distributed around $x_i'\beta_0$ that $u_i = y_i - x_i'\beta_0$ is symmetrically distributed around zero. Let $\delta(\tau) \equiv \beta(\tau) - \beta_0$. Then from equation (2.9) we have

$$(A.4) \qquad \beta(\tau) - \beta_0 = \{E[|\tau-1(u_i<x_i'\delta(\tau))|x_ix_i']\}^{-1}E[|\tau-1(u_i<x_i'\delta(\tau))|x_iu_i]$$
$$= \{E[|1-\tau-1(u_i\geq x_i'\delta(\tau))|x_ix_i'\}^{-1}E[|1-\tau-1(u_i\geq x_i'\delta(\tau))|x_iu_i]$$
$$= \{E[|1-\tau-1(u_i>x_i'\delta(\tau))|x_ix_i']\}^{-1}E[|1-\tau-1(u_i>x_i'\delta(\tau))|x_iu_i]$$
$$= -\{E[|1-\tau-1(-u_i<x_i'[-\delta(\tau))]|x_ix_i']\}^{-1} \times$$
$$E[|1-\tau-1(-u_i<x_i'[-\delta(\tau))]|x_i(-u_i)]$$
$$= -E[|1-\tau-1(u_i<x_i'[-\delta(\tau))]|x_ix_i']\}^{-1}E[|1-\tau-1(u_i<x_i'[-\delta(\tau))]|x_iu_i],$$

where the third equality follows by continuity of the conditional distribution of y_i and the fifth by symmetry of the condtional distribution of u_i about zero. From this equation and equation (2.9) we see that $-[\beta(\tau) - \beta_0]$ solves the equation for $\beta(1-\tau) - \beta_0$ from which it follows that

$$(A.5) \qquad \beta(1-\tau) - \beta_0 = -[\beta(\tau) - \beta_0].$$

The following lemma is useful for proving consistency of asymmetric least squares estimators without requiring the parameter space to be compact.

Lemma A: Let θ_0 be a point in $\mathbb{R}^q$ and Θ an open set containing θ_0. If

(A) $Q_n(\theta)$ converges to $Q(\theta)$ in probability uniformly on Θ,

(B) $Q(\theta)$ has a unique minimum on Θ at θ_0,

(C) $Q_n(\theta)$ is convex in θ;

then for $\theta = \text{argmin}_{\mathbb{R}^q} Q_n(\theta)$,

(i) $\hat{\theta}$ exists with probability approaching one;

(ii) $\hat{\theta}$ converges in probability to θ_0.

Let C be a closed ball of finite radius that is a subset of Θ and that contains θ_0 in its interior. Let B be the boundary of C. By convergence in probability of $Q_n(\theta)$ to $Q(\theta)$ pointwise on Θ, $Q(\theta)$ is convex, and therefore continuous, on Θ. It follows that $\min_B Q(\theta)$ exists and by assumption B that

(A.6) $\quad \delta = \min_B Q(\theta) - Q(\theta_0) > 0.$

By uniform convergence in probability of $Q_n(\theta)$ to $Q(\theta)$ on Θ it follows that $\text{plim}[\max_C |Q_n(\theta)-Q(\theta)|] = 0$. Therefore, by equation (A.6), with probability approaching one

(A.7) $\quad \min_B Q_n(\theta) > Q_n(\theta_0).$

By construction of C and B, for any $\tilde{\theta}$ not in C there exists θ in B and $0 < \lambda < 1$ such that $\tilde{\theta} = \lambda\theta + (1-\lambda)\theta_0$, so that by convexity of $Q_n(\theta)$, equation (A.7) implies

(A.8) $\quad Q_n(\theta) \geq [Q_n(\tilde{\theta}) - Q_n(\theta_0)]/\lambda + Q_n(\theta_0) > Q_n(\theta_0).$

It follows that the minimum of $Q_n(\theta)$ on $\mathbb{R}^q$ exists and lies in C with probability approaching one. Convergence of $\hat{\theta}$ to θ_0 follows from the arbitrary choice of C as a closed ball containing θ_0 in its interior.

Proof of Theorem 3: Note that $\rho_\tau(\lambda)$ is differentiable and convex in λ, so that $\rho_\tau(y_i-x_i'\beta)$ is differentiable and convex in β, with $g_i(\beta) \equiv \partial\rho_\tau(y_i-x_i'\beta)/\partial\beta = -2x_i\psi_\tau(y_i-x_i'\beta)$. Note that for some constants d and d',

(A.9) $$|g_i(\beta)| \le |z_i|^2(d + d'|\beta|),$$

so that by Assumption 3 $g_i(\beta)$ is uniformly dominated by an integrable function on a neighborhood of any β. It follows that $R(\beta,\tau)$ is differentiable in β with

(A.10) $$\begin{aligned}\partial R(\beta,\tau)/\partial\beta &= E[g_i(\beta)] \\ &= -2E[x_i\{\tau\textstyle\int_{x_i'\beta}^{\infty}(y-x_i'\beta)f(y|x_i,\gamma_0)dy \\ &\quad + (1-\tau)\textstyle\int_{-\infty}^{x_i'\beta}(y-x_i'\beta)f(y|x_i,\gamma_0)dy\}].\end{aligned}$$

Note that $\int_{-\infty}^{\alpha}(y-\alpha)f(y|x,\gamma_0)dy$ is continuously differentiable in α, with derivative $-\int_{-\infty}^{\alpha}f(y|x,\gamma_0)dy$, which is uniformly dominated by 1. It follows that $\partial R(\beta,\tau)/\partial\beta$ is continuously differentiable, with

(A.11) $$\begin{aligned}&\partial^2R(\beta,\tau)/\partial\beta\partial\beta' \\ &= 2E[x_ix_i'\{\tau\textstyle\int_{x_i'\beta}^{\infty}f(y|x_i,\gamma_0)dy + (1-\tau)\textstyle\int_{-\infty}^{x_i'\beta}f(y|x_i,\gamma_0)dy\}]. \\ &= 2E[x_ix_i'|\tau-1(y_i<x_i'\beta)|].\end{aligned}$$

Let $\delta = \min\{\tau,1-\tau\}$, and note that $|\tau-1(y_i<x_i'\beta)| > \delta$. It follows that $\partial^2R(\beta,\tau)/\partial\beta\partial\beta'-\delta E[x_ix_i']$ is positive semi-definite. For any β in $\mathbb{R}^p$ a second order mean value expansion of $R(\beta,\tau)$ around $\tilde{\beta}$ gives

$$\text{(A.12)}\quad R(\beta,\tau) - R(\tilde{\beta},\tau) = [\partial R(\tilde{\beta},\tau)/\partial\beta]'(\beta-\tilde{\beta}) + (\beta-\tilde{\beta})'[\partial^2 R(\dot{\beta},\tau)/\partial\beta\partial\beta'](\beta-\tilde{\beta})$$
$$\geq [\partial R(\tilde{\beta},\tau)/\partial\beta]'(\beta-\tilde{\beta}) + \delta m_x p|\beta-\tilde{\beta}|^2,$$

where $\dot{\beta}$ is the mean value and m_x is the minimum eigenvalue of $E[x_i x_i']$, which is positive by Assumption 4. By dividing through equation (A.12) by $|\beta-\tilde{\beta}|^2$ we see that for fixed $\tilde{\beta}$, $R(\beta,\tau) > R(\tilde{\beta},\tau)$ for $|\beta-\tilde{\beta}|$ big enough. That is, $R(\beta,\tau) > R(\tilde{\beta},\tau)$ outside some closed ball centered at $\tilde{\beta}$. It follows from continuity that $R(\beta,\tau)$ has a minimum $\beta(\tau)$ inside this ball, and since $R(\beta(\tau),\tau) \leq R(\tilde{\beta},\tau)$, that $\beta(\tau)$ is a global minimum. By differentiability, $\beta(\tau)$ satisfies $\partial R(\beta(\tau),\tau)/\partial\beta = E[g_i(\beta(\tau))] = 0$, which can be solved to obtain equation (2.9). The fact that $\beta(\tau)$ is a unique global minimum of $R(\beta,\tau)$ follows from (A.12) with $\tilde{\beta} = \beta(\tau)$, so that $\beta(\tau)$ is the unique solution of $E[g_i(\beta)] = 0$ by convexity of $R(\beta,\tau)$.

To obtain the asymptotic distribution result we consider first the case with $m = 1$, and for notational convenience the j subscript is suppressed. Note that there exist constants d and d' such that

$$\text{(A.13)}\quad |\rho_\tau(y_i-x_i'\beta)| \leq |y_i-x_i'\beta|^2 \leq |z_i|^2(d' + d|\beta|^2).$$

Uniform convergence in probability of $R_n(\beta,\tau)/n$ to $R(\beta,\tau)$ on any bounded open set containing $\beta(\tau)$ follows from Assumptions 2 and 3 by Lemma A1 of Newey (1985). Then $\text{plim}\,\hat{\beta}(\tau) = \beta(\tau)$ follows by Lemma A and $\beta(\tau)$ the unique minimum of $R(\beta,\tau)$, which was shown above.

Let $E_n[\cdot] \equiv E[\cdot|\mathcal{F}_n]$. By arguments like those above for $R(\beta,\tau)$, it follows that $E_n[\rho_\tau(y_i-x_i'\beta)]$ is twice continuously differentiable in β for large enough n, with

$$\text{(A.14)}\quad \lambda_n(\beta) \equiv \partial E_n[\rho_\tau(y_i-x_i'\beta)]/\partial\beta = E_n[g_i(\beta)],$$
$$\partial\lambda_n(\beta)/\partial\beta = \partial^2 E_n[\rho_\tau(y_i-x_i'\beta)]/\partial\beta\partial\beta' = 2E_n[x_i x_i'|\tau-1(y_i<x_i'\beta)|].$$

By continuity of $\rho_\tau(y_i - x_i'\beta)$ in β, continuity of $f(y|x,\gamma)$ in γ, Assumption 3, and the dominated convergence theorem $E_n[\rho_\tau(y_i - x_i'\beta)]$ converges uniformly to $E[\rho_\tau(y_i - x_i'\beta)]$ on any compact neighborhood N of $\beta(\tau)$. It follows that there is a sequence $\beta_n(\tau)$ that minimizes $E_n[\rho_\tau(y_i - x_i'\beta)]$ on N such that $\lim_{n\to\infty}\beta_n(\tau) = \beta(\tau)$, and that for large enough n,

$$\text{(A.15)} \qquad 0 = \lambda_n(\beta_n(\tau)) = E_n[g_i(\beta_n(\tau))].$$

Also, by continuity of $f(y|x,\gamma)$ in γ, Assumption 3, and the dominated convergence theorem, $\partial\lambda_n(\beta)/\partial\beta$ converges uniformly on N to $\partial^2 R(\beta,\tau)/\partial\beta\partial\beta'$. By $\lim\beta_n(\tau) = \beta(\tau)$ and $\partial^2 R(\beta(\tau),\tau)/\partial\beta\partial\beta'$ nonsingular, there are positive constants d and d' such that for n large enough,

$$\text{(A.16)} \qquad |\beta - \beta_n(\tau)| < d \Rightarrow |\partial\lambda_n(\beta)/\partial\beta| > d'|\beta - \beta_n(\tau)|.$$

Now let $u(z_i,\beta,d) \equiv \sup_{|\tilde{\beta}-\beta|\le d}|g_i(\tilde{\beta}) - g_i(\beta)|$. By $|\psi_\tau(\tilde{\lambda}) - \psi_\tau(\lambda)| \le |\tilde{\lambda} - \lambda|$,

$$\text{(A.17)} \qquad u(z_i,\beta,d) \le \sup_{|\tilde{\beta}-\beta|\le d} 2|x_i||x_i'(\tilde{\beta}-\beta)| \le 2p|x_i|^2 d.$$

Then by Assumption 3,

$$\text{(A.18)} \qquad E_n[u(z_i,\beta,d)] \le pdM,$$
$$E_n[u(z_i,\beta,d)^2] \le p^2 dM',$$

where $M \equiv \int|x|^2\alpha(z)g(x)d\mu_z$ and $M' \equiv \int|x|^4\alpha(z)g(x)d\mu_z$. From equations (A.15), (A.16), and (A.18) it follows that Assumptions (N-1) - (N-4) of Huber (1965) are satisfied uniformly in n. Also, by $\text{plim}\hat{\beta}(\tau) = \beta(\tau)$ and $\lim\beta_n(\tau) = \beta(\tau)$, $\text{plim}(\hat{\beta}(\tau) - \beta_n(\tau)) = 0$. Then Theorem 3 of Huber (1965) gives

$$\text{(A.19)} \qquad \Sigma_{i=1}^n g_i(\beta_n(\tau))/\sqrt{n} + \sqrt{n}\lambda_n(\hat{\beta}(\tau)) = o_p(1).$$

A mean value expansion of $\lambda_n(\hat{\beta}(\tau))$ around $\beta(\tau)$ gives

(A.20) $[\partial\lambda_n(\dot{\beta}(\tau))/\partial\beta]\sqrt{n}[\hat{\beta}(\tau)-\beta(\tau)] = -\sqrt{n}\lambda_n(\beta(\tau)) - \Sigma_{i=1}^n g_i(\beta_n(\tau))/\sqrt{n} + o_p(1)$,

where $\dot{\beta}(\tau)$ is the mean value. Uniform convergence of $\partial\lambda_n(\beta)/\partial\beta$ and continuity of $\partial^2R(\beta,\tau)/\partial\beta\partial\beta'$ imply $\text{plim}\,\partial\lambda_n(\dot{\beta}(\tau))/\partial\beta = 2W$, which is nonsingular. Also, it can be shown as in the proof of Lemma 2.1 of Newey (1985) that $\Sigma_{i=1}^n g_i(\beta_n(\tau))/\sqrt{n}$ converges in distribution to $N(0,4V)$. The conclusion then follows from the fact that

(A.21) $\lim_{n\to\infty}[-\sqrt{n}\lambda_n(\beta(\tau))] = 2G\delta$,

which can be shown from a mean value expansion argument as in the proof of Lemma 2.1 of Newey (1985). The case with $m \geq 1$ follows similarly.

Proof of Theorem 4: Note that for any positive ϵ and $I(x,\epsilon) \equiv [x'\beta(\tau)-\epsilon|x|,\ x'\beta(\tau)+\epsilon|x|]$.

(A.22) $E_n[1(|u_i(\tau)|\leq\epsilon|x|)|x] = \int_{I(x,\epsilon)} f(y|x,\gamma_n)dy \leq \int_{I(x,\epsilon)}\alpha(z)dy \equiv \alpha_\epsilon(x)$.

By the monotone convergence theorem and the fact that with probability one $\alpha(z)$ is integrable in y with respect to Lebesgue measure, $\alpha_\epsilon(x)$ converges to zero with ϵ, and this convergence is monotone by construction. Note that $\hat{w}_i(\tau)$ differs from $w_i(\tau)$ only if $x_i'\hat{\beta}(\tau) \leq y_i \leq x_i'\beta(\tau)$ or $x_i'\beta(\tau) \leq y_i \leq x_i'\hat{\beta}(\tau)$, so that

(A.23) $|\hat{w}_i(\tau)-w_i(\tau)| \leq |2\tau-1|\cdot 1(|u_i(\tau)| \leq |x_i'[\hat{\beta}(\tau)-\beta(\tau)]|)$
$\leq |2\tau-1|\cdot 1(|u_i(\tau)| \leq p|x_i||\hat{\beta}(\tau)-\beta(\tau)|)$.

By $\text{plim}\,\hat{\beta}(\tau) = \beta(\tau)$, equation (A.30) implies that for any $\epsilon > 0$, with probability approaching one,

(A.24) $|\Sigma_{i=1}^n\hat{w}_i(\tau)x_ix_i'/n - \Sigma_{i=1}^n w_i(\tau)x_ix_i'/n|$

$$\leq \Sigma_{i=1}^{n}|x_i|^2\cdot 1(|u_i(\tau)|\leq\epsilon|x_i|)/n \leq E[|x_i|^2\alpha_\epsilon(x_i)] + \epsilon,$$

where the second inequality follows from Markov's inequality for large n with probability approaching one. Note that $E[|x_i|^2\alpha_\epsilon(x_i)] + \epsilon$ converges to zero with ϵ by the monotone convergence theorem, so that the term in equation (A.24) that precedes the inequalities converges in probability to zero. Suppressing the j subscript for notational convenience, the triangle inequality gives

$$\text{(A.25)}\quad |\hat{W} - W| \leq |\Sigma_{i=1}^{n}\hat{w}_i(\tau)x_ix_i'/n - \Sigma_{i=1}^{n}w_i(\tau)x_ix_i'/n| + |\Sigma_{i=1}^{n}w_i(\tau)x_ix_i'/n - W|.$$

Consistency of $\hat{W}$ then follows from Lemma A1 of Newey (1985) applied to the second term in (A.25). To show consistency of $\hat{V}_{jk}$, note that there exist constants d, d' and d'' such that

$$\text{(A.26)}\quad |x_ix_i'\psi_\tau(y_i-x_i'\beta)\psi_\theta(y_i-x_i'b)| \leq |z_i|^4(d + d'|\beta|^2 + d''|b|^2).$$

The result then follows as in the proof of Theorem 2.2 of Newey (1985).

Proof of Theorem 5: The noncentral chi-square limiting distribution of T will follow immmediately from the asymptotic normality of $\hat{\xi}$ and the consistency of the covariance matrix estimator. Let $f(\varepsilon)$ and $F(\varepsilon)$ be the p.d.f. and c.d.f. of ε_i. Also let $S = (-1/2, 1/2)$. Then for σ in S,

$$\text{(A.27)}\quad [1/(1+\sigma)]f(u/(1+\sigma)) \leq 2f(u/(1+\sigma)) \leq 2D/[1+|u/(1+\sigma)|^{5+d}]$$
$$\leq 2D/[1 + |2u/3|^{5+d}].$$

Furthermore, by x_i having compact support there is an open set Γ in $\mathbb{R}^{2p}$ containing zero such that for $(\gamma_h', \gamma_s')'$ in Γ, $x_i'\gamma_h + 1(\varepsilon_i>0)x_i'\gamma_s$ is an element of S with probability one. It follows that the domination condition of Assumption 3 is satisfied with $\alpha(z) = 2D/[1 + |2(y-x'\beta_0)/3|^{5+d}]$. Continuity of $f(y|x,\gamma)$ in y and γ follows from continuity of $f(\varepsilon)$.

It remains to check the remainder of Assumption 4 and verify the particular formula for the noncentrality parameter that is given. Note that in general the noncentrality parameter would be equal to

$$(A.28)\qquad (HW^{-1}G\delta)'[HW^{-1}VW^{-1}H']^{-1}HW^{-1}G\delta.$$

In this case u_i is independent of x_i when $\gamma = \gamma_0 = 0$, so that $x_i'\beta(\tau) = \mu(\tau) + x_i'\beta_0$, $W_j = d(\tau_j)D$, and $V_{jk} = \sigma_{jk}D$, where $\mu(\tau)$ is the τ^{th} weighted mean of ε_i. It follows that $W^{-1}VW^{-1} = \Sigma \otimes D$. Also note that

$$(A.29)\qquad d(\tau) = (1-\tau)F(\mu(\tau)) + \tau[1-F(\mu(\tau))],$$

and that $E[x_i\psi_\tau(u_i(\tau))|\gamma] = E[x_iE[w_i(\tau)u_i(\tau)|x_i,\gamma]]$, where for $\mu(\tau) > 0$,

$$\begin{aligned}(A.30)\qquad E[w_i(\tau)u_i(\tau)|x_i,\gamma] &= (1-\tau)\int_{-\infty}^{0}[(u-\mu(\tau))/\sigma_{in}]f(u/\sigma_{in})du \\ &+ (1-\tau)\int_0^{\mu(\tau)}[(u-\mu(\tau))/\sigma_{ip}]f(u/\sigma_{ip})du \\ &+ \tau\int_{\mu(\tau)}^{\infty}[(u-\mu(\tau))/\sigma_{ip}]f(u/\sigma_{ip})du \\ &= (1-\tau)\sigma_{in}\int_{-\infty}^{0}\varepsilon f(\varepsilon)d\varepsilon + (1-\tau)\sigma_{ip}\int_0^{\mu(\tau)/\sigma_{ip}}\varepsilon f(\varepsilon)d\varepsilon \\ &+ \tau\sigma_{ip}\int_{\mu(\tau)/\sigma_{ip}}^{\infty}\varepsilon f(\varepsilon)d\varepsilon - (1-\tau)\mu(\tau)\int_{-\infty}^{0}f(\varepsilon)d\varepsilon \\ &- (1-\tau)\mu(\tau)\int_0^{\mu(\tau)/\sigma_{ip}}f(\varepsilon)d\varepsilon - \tau\mu(\tau)\int_{\mu(\tau)/\sigma_{ip}}^{\infty}f(\varepsilon)d\varepsilon,\end{aligned}$$

where $\sigma_{in} = 1 + x_i'\gamma_h$ and $\sigma_{ip} = 1 + x_i'\gamma_h + x_i'\gamma_s$. We now find that $E[w_i(\tau)u_i(\tau)|x_i,\gamma]$ is differentiable in γ on Γ, with

$$(A.31)\qquad \partial E[w_i(\tau)u_i(\tau)|x_i,0]/\partial\gamma = d(\tau)(\mu(\tau),\nu(\tau))' \otimes x_i,$$

which is obviously dominated by an integrable function, so that

$$\begin{aligned}(A.32)\qquad \partial E[x_i\psi_\tau(u_i(\tau))|0]/\partial\gamma &= E[x_i\partial E[w_i(\tau)u_i(\tau)|x_i,0]/\gamma'] \\ &= d(\tau)(\mu(\tau),\nu(\tau)) \otimes D.\end{aligned}$$

A similar calculation shows that equation (A.32) also holds for the case with

$\mu(\tau) < 0$. Consequently, the differentiability hypothesis of Assumption 2 is satisfied with

$$(W_j)^{-1}G_j = (\mu(\tau_j),\nu(\tau_j)) \otimes I_p, \tag{A.33}$$

so that the result follows from (A.28).

Proof of Corollary 1: The result follows upon application of the usual results for matrix inversion and arithmetic involving Kronecker products.

FOOTNOTES

1. This research was supported by NSF grants SES-8309292 at M.I.T. and SES-8410249 at Princeton University. We are grateful for helpful comments provided by A. Bera, G. Chamberlain, A. Goldberger, J. Hausman, J. Heckman, R. Koenker, C. Manski, R. Quandt, two anonymous referees, and participants at several workshops. A. Goldberger provided the calculations used in Figure 1 below. An earlier version of this paper was presented at the Winter 1983 North American meetings of the Econometric Society.

2. An alternative equation for $\mu(\tau)$, which was suggested to us by A. Goldberger, is

$$\tau/(1-\tau) = \left[\int_{(-\infty,\ \mu(\tau))} (\mu(\tau)-y)dF(y) \right] \cdot \left[\int_{(\mu(\tau),\ \infty)} (y-\mu(\tau))dF(y) \right]^{-1}.$$

By comparing this equation to the analogous relation for quantiles, $\theta/(1-\theta) = F(\eta(\theta))/(1-F(\eta(\theta)))$, it is easy to see that expectiles are determined by tail expectations in the same way that quantiles are determined by the distribution function.

3. Alternative terminology for $\mu(\tau)$ has been suggested, including "gravile," "heftile," and "loadile" (by A. Goldberger, motivated by the interpretation of expectation as a center of gravity), as well as "projectile" (by G. Chamberlain, motivated by the fact that $\mu(\tau)$ solves a least squares problem).

4. Roger Koenker has informed us that this error was also pointed out to him by Alistair Hall of Warwick University.

REFERENCES

Aigner, D.J., T. Amemiya, and D. Poirier: "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," International Economic Review, 17 (1976), 377-396.

Anscombe, F.J.: "Examination of Residuals," Proceedings of the Fifth Berkeley Symposium, 1 (1961), 1-36.

Antille, A., G. Kersting, and W. Zuccini: "Testing Symmetry," Journal of the American Statistical Association, 77 (1982), 639-646.

Bickel, P.J.: "Using Residuals Robustly I: Tests for Heteroscedasticity and Nonlinearity," Annals of Statistics, 6 (1978), 266-291.

Bickel, P.J. and E.L. Lehmann: "Descriptive Statistics for Nonparametric Models I: Introduction," Annals of Statistics, 3 (1975), 1038-1044.

Boos, D.D.: "A Test for Asymmetry Associated With the Hodges-Lehmann Estimator," Journal of the American Statistical Association," 77 (1982), 647-651.

Breusch, T.S. and A.R. Pagan: "A Simple Test for Heteroscedasticity and Random Coefficient Variation," Econometrica, 47 (1979), 1287-1294.

DeGroot, M.H.: Optimal Statistical Decisions, New York: McGraw-Hill, 1970.

Eicker, F.: "Limit Theorems for Regressions with Unequal and Dependent Errors," Proceedings of the Fifth Berkeley Symposium, 1 (1967), 59-82.

Glesjer, H.: "A New Test for Heteroscedasticity," Journal of the American Statistical Association, 64 (1969), 316-323.

Godfrey, L.G.: "Testing for Multiplicative Heteroskedasticity," Journal of Econometrics, 8 (1978), 227-236.

Goldfeld, S.M. and R.E. Quandt: Nonlinear Methods in Econometrics, Amsterdam: North-Holland, 1972.

Harvey, A.C.: "Estimating Regression Models with Multiplicative Heteroscedasticity," Econometrica, 44 (1976), 461-466.

Hausman, J.A.: "Specification Tests in Econometrics," Econometrica, 46 (1978), 1251-1271.

Koenker, R. and G.S. Bassett Jr.: "Regression Quantiles," Econometrica, 4 (1978), 33-50.

Koenker, R. and G.S. Bassett Jr.: "Robust Tests for Heteroscedasticity Based on Regression Quantiles," Econometrica, 50 (1982), 43-61.

Manski, C.F.: "Adaptive Estimation of Nonlinear Regression Models," Econometric Reviews, (1983),: 145-194.

Newey, W.K.: "Adaptive Estimation of Regression Models Via Moment Restrictions," manuscript, Princeton University, 1986.

Newey, W.K.: "Maximum Likelihood Specification Testing and Conditional Moment Tests," Econometrica, 53 (1985), 1047-1070.

White, H.: "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48 (1980), 817-838.

White, H.: "Corrigendum," Econometrica, 51 (1983), 513.

Figure 1

Plot of Quantile ($\eta(\theta)$) and Expectile ($\mu(\tau)$) Functions

for the Standard Normal Distribution

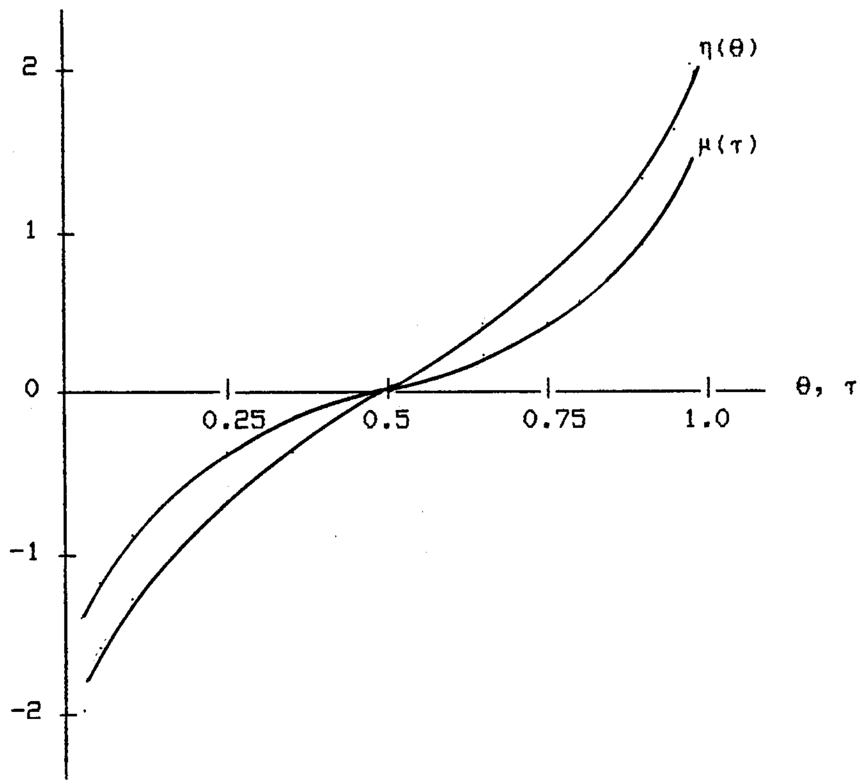

Table 1

Optimal Values of Regression Quantile (Expectile) Weights for Tests of Homoskedasticity

Contamination Percentage α \ Relative Scale σ	2	3	4	5
.05	.92 (.78)	.90 (.61)	.90 (.51)	.89 (.51)
.10	.91 (.73)	.88 (.52)	.87 (.51)	.87 (.51)
.15	.90 (.70)	.87 (.51)	.85 (.51)	.84 (.51)
.20	.89 (.67)	.85 (.51)	.83 (.51)	.82 (.51)
.25	.88 (.66)	.84 (.51)	.82 (.51)	.80 (.51)
.30	.88 (.66)	.83 (.51)	.80 (.51)	.79 (.51)
.35	.87 (.66)	.82 (.51)	.79 (.51)	.77 (.51)
.40	.87 (.66)	.81 (.51)	.77 (.51)	.76 (.51)
.45	.87 (.66)	.80 (.51)	.97 (.51)	.97 (.51)
.50	.87 (.66)	.97 (.52)	.97 (.51)	.97 (.97)*

* The last entry in the table (.97) is correct. For large values of σ, the noncentrality scalar κ_{LS}^{h} is bimodal as a function of τ, with relative maxima at $\tau = .51$ and $\tau = .97$. For the most extreme values of α and σ tabulated, the global maximum occurs at the latter value.

Table 2

Local Efficiencies of Tests for Heteroskedasticity, Relative to

Squared Residual Regression Test

Relative Scale σ	Contamination Proportion α	Regression Quantile	Asymmetric Least Squares	Absolute Residual Regression
1	-	.59	.88	.88
2	.0125	.68	.98	.98
	.025	.74	1.06	1.06
	.05	.84	1.17	1.16
	.10	.94	1.25	1.25
	.15	.96	1.26	1.26
	.20	.93	1.24	1.24
	.25	.89	1.20	1.20
	.30	.83	1.17	1.17
	.40	.74	1.10	1.09
	.50	.66	1.03	1.03
3	.0125	1.15	1.54	1.54
	.025	1.50	1.85	1.85
	.05	1.83	2.04	2.04
	.10	1.87	1.90	1.90
	.15	1.65	1.69	1.69
	.20	1.39	1.51	1.51
	.25	1.15	1.37	1.37
	.30	.94	1.26	1.26
	.40	.64	1.11	1.11
	.50	.49	1.02	1.02
4	.0125	2.28	2.70	2.70
	.025	3.02	3.08	3.08
	.05	3.36	2.89	2.89
	.10	2.84	2.21	2.21
	.15	2.17	1.77	1.78
	.20	1.61	1.50	1.50
	.25	1.18	1.32	1.32
	.30	.85	1.20	1.20
	.40	.45	1.04	1.04
	.50	.34	.95	.95
5	.0125	4.21	4.29	4.29
	.025	5.18	4.28	4.28
	.05	5.04	3.38	3.39
	.10	3.62	2.23	2.24
	.15	2.49	1.70	1.70
	.20	1.69	1.41	1.41
	.25	1.12	1.23	1.23
	.30	.71	1.11	1.11
	.40	.29	.97	.97
	.50	.28	.90	.90

Table 3

Optimal Values of Regression Quantile (Expectile)

Weights for Tests of Symmetry

Contamination Percentage α / Relative Scale σ	2	3	4	5
.05	.93 (.62)	.92 (.51)	.91 (.51)	.91 (.51)
.10	.92 (.57)	.90 (.51)	.89 (.51)	.88 (.51)
.15	.91 (.54)	.89 (.51)	.87 (.51)	.86 (.51)
.20	.91 (.52)	.87 (.51)	.85 (.51)	.84 (.51)
.25	.90 (.52)	.86 (.51)	.83 (.51)	.82 (.51)
.30	.90 (.52)	.84 (.51)	.82 (.51)	.80 (.51)
.35	.89 (.52)	.83 (.51)	.80 (.51)	.78 (.51)
.40	.89 (.52)	.82 (.51)	.79 (.51)	.77 (.51)
.45	.89 (.52)	.81 (.51)	.97 (.51)	.97 (.51)
.50	.89 (.52)	.97 (.51)	.97 (.51)	.97 (.51)

Table 4

Local Efficiencies of Tests for Asymmetry, Relative to

Squared Residual Regression Test

Relative Scale σ	Contamination Proportion α	Regression Quantile	Asymmetric Least Squares	"Median vs. Mean"
1	-	.73	.97	.96
2	.0125	.96	1.25	1.24
	.025	1.12	1.45	1.44
	.05	1.33	1.71	1.70
	.10	1.48	1.88	1.88
	.15	1.47	1.87	1.87
	.20	1.40	1.80	1.80
	.25	1.31	1.71	1.71
	.30	1.21	1.62	1.62
	.40	1.04	1.45	1.46
	.50	.91	1.32	1.32
3	.0125	2.32	2.91	2.89
	.025	3.03	3.64	3.64
	.05	3.40	3.85	3.88
	.10	3.04	3.25	3.30
	.15	2.49	2.67	2.73
	.20	2.02	2.26	2.31
	.25	1.64	1.96	2.00
	.30	1.32	1.74	1.78
	.40	.88	1.45	1.49
	.50	.65	1.28	1.31
4	.0125	5.37	6.23	6.24
	.025	6.25	6.59	6.66
	.05	5.77	5.40	5.52
	.10	4.13	3.53	3.65
	.15	2.98	2.58	2.69
	.20	2.17	2.05	2.14
	.25	1.57	1.73	1.81
	.30	1.12	1.51	1.58
	.40	.56	1.25	1.30
	.50	.40	1.11	1.15
5	.0125	9.75	10.20	10.30
	.025	9.74	8.77	8.97
	.05	7.64	5.80	6.01
	.10	4.74	3.25	3.40
	.15	3.16	2.26	2.38
	.20	2.14	1.77	1.86
	.25	1.40	1.48	1.56
	.30	.87	1.29	1.36
	.40	.32	1.07	1.13
	.50	.29	.96	1.01

Table 5

Relative Efficiency of the Asymmetric Least Squares to Regression

Quantiles Test of Homoskedasticity and Symmetry

Relative Scale σ	Contamination Proportion α	Homoskedasticity	Symmetry
1	-	1.49	1.32
2	.0125	1.45	1.31
	.025	1.42	1.30
	.05	1.38	1.29
	.10	1.33	1.27
	.15	1.32	1.27
	.20	1.33	1.29
	.25	1.35	1.31
	.30	1.39	1.34
	.40	1.48	1.40
	.50	1.57	1.46
3	.0125	1.34	1.25
	.025	1.24	1.20
	.05	1.11	1.13
	.10	1.02	1.07
	.15	1.03	1.07
	.20	1.09	1.11
	.25	1.19	1.31
	.30	1.34	1.31
	.40	1.72	1.65
	.50	2.06	1.96
4	.0125	1.18	1.16
	.025	1.02	1.06
	.05	.86	.94
	.10	.78	.85
	.15	.82	.87
	.20	.93	.95
	.25	1.12	1.10
	.30	1.41	1.36
	.40	2.31	2.24
	.50	2.83	2.81
5	.0125	1.02	1.05
	.025	.83	.90
	.05	.67	.76
	.10	.62	.68
	.15	.68	.72
	.20	.83	.83
	.25	1.10	1.05
	.30	1.57	1.49
	.40	3.32	3.33
	.50	3.24	3.29