ECONOMIC RESEARCH REPORTS

ROBUST M-TESTS

by

Franco Peracchi

R.R.# 90-25

June, 1990

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

ROBUST M-TESTS

by

Franco Peracchi*

Department of Economics
New York University
New York, NY 10003

August 1989
Revised June 1990

JEL No. 211
Keywords: M-tests, Robust tests.

Abstract

This paper investigates the local robustness properties of a general class of multi-dimensional tests based on M-estimators. These tests are shown to inherit the efficiency and robustness properties of the estimators on which they are based. In particular, it is shown that small perturbations of the distribution of the observations can have arbitrarily large effects on the asymptotic level and power of tests based on estimators that do not possess a bounded influence function. An asymptotic 'admissibility' result is also presented, that provides a justification for tests based on optimal bounded-influence estimators.

*This paper is based on a Chapter of my dissertation at Princeton University. I am very grateful to my advisors, Angus Deaton, Whitney Newey and Elvezio Ronchetti, for their generous guidance. I also thank Roger Koenker and Quang Vuong for helpful comments, and the C.V. Starr Center for financial support. An earlier version of this paper was presented to the 1988 North American Summer Meetings and the 1988 European Meetings of the Econometric Society.

1. INTRODUCTION

The problem of the robustness of a test, i.e. the stability of its level and power under small changes in the underlying probability distribution of the observations, has received considerable attention in the statistical literature, mainly with reference to one-dimensional tests or the linear model [Rieder (1978), Schrader and Hettmansperger (1980), Lambert (1981), Rousseeuw and Ronchetti (1981), Kent (1982), Ronchetti (1982), Wang (1982), and Hampel et al. (1986)], but has been largely ignored by econometricians. This paper investigates the local robustness properties of a broad class of multi-dimensional tests, called M-tests because they are based on M-estimators. This class of tests includes most common tests in econometrics, such as Wald, score and Hausman tests.

We study the asymptotic properties of M-tests under small perturbations of the assumed probability distribution of the observations. The particular kind of perturbations that we consider are 'contamination models' where the assumed distribution is contaminated, with small but positive probability, by some extraneous distribution. This is a convenient way of representing the fact that an econometric model is at best an approximation to the true data generation process, and this approximation may be adequate for the majority but not all the observations. Our approach builds on earlier work of Rieder (1978), Rousseeuw and Ronchetti (1981), Ronchetti (1982) and Wang (1982) for one-dimensional, one-sided tests, and on results of Hampel et al. (1986) for the linear model. We show that contamination of the assumed model can have very serious effects, leading to tests that are biased and inconsistent. However, contamination may also increase the power of a test against alternatives that are in the same direction as the asymptotic bias of the statistic on which the test is based.

An M-test is called locally robust if a small amount of contamination has only a small effect on its asymptotic level and power. We show that local robustness is guaranteed whenever the estimator on which the test statistic is based has a bounded influence function. This provides a further justification for the use of bounded-influence estimators, since not only point estimates but also inference are relatively unaffected by a small amount of contamination.

For the one-dimensional case, Ronchetti (1982) showed that optimal bounded-influence estimators lead to tests that are most powerful in the class of locally robust tests. In the multi-dimensional case, tests based on optimal bounded influence estimators cannot be shown to have maximal power against all directions, but we can prove an asymptotic 'admissibility' result that provides a justification for their use.

The rest of this paper is organized as follows. Section 2 contains definitions and basic assumptions. Section 3 examines the asymptotic properties of M-tests under sequences of contaminated local alternatives. Section 4 illustrates with the example of testing linear restrictions in a linear regression model subject to gross-errors. Section 5 characterizes the class of locally robust M-tests. Section 6 contains the asymptotic 'admissibility' result.

The following notation will be used: F_0 denotes the true distribution function (d.f.) of a single observation and E_0 denotes expectations taken with respect to F_0. Expectations taken with respect to a member F_θ of a parametric family of d.f's are denoted by E_θ. Expectations taken with respect to any other d.f. F are denoted by E_F.

2. THE STATISTICAL MODEL AND THE CLASS OF M-ESTIMATORS

For simplicity, we shall consider the case when the observations $Z_1,\ldots, Z_N$ are randomly drawn from some unknown probability distribution, identified with its cumulative distribution function.

ASSUMPTION A.1: $Z_1,\ldots, Z_N$ *are independently and identically distributed random vectors with values in* $\mathcal{Z} \subseteq \mathbb{R}^m$ *and common d.f.* F_0.

Typically F_0 is unknown, but *a priori* knowledge, convenience or simply custom, may suggest restricting F_0 to some family $\mathcal{F}$ of d.f's on $\mathcal{Z}$.

DEFINITION 1: The *assumed statistical model* is a triple $(\mathcal{F},\Theta,\bar{\theta})$, where $\mathcal{F}$ is a set of d.f.'s on $\mathcal{Z}$, Θ is a non empty subset of $\mathbb{R}^p$, and $\bar{\theta}$ is a functional that maps $\mathcal{F}$ onto Θ.

The pair $(\Theta,\bar{\theta})$ defines the parameterization of the model. The model is parametric if the functional $\bar{\theta}$ is one-to-one, and is semi-parametric otherwise. To allow for model misspecification we do not require that F_0 belong to $\mathcal{F}$. The set consisting of F_0 and $\mathcal{F}$ will be denoted by $\mathcal{F}_0$.

DEFINITION 2: An *M-estimator* associated with the criterion function ρ_N, where ρ_N is a real function defined on $\mathcal{Z} \times \Theta$, is a global maximum $\hat{\theta}_N$ of $\Sigma_{n=1}^{N} \rho_N(Z_n,\theta)$.

The class of M-estimators contains most common econometric estimators, such as least squares, maximum-likelihood (ML) and generalized method of moments estimators [see e.g. Hansen (1982)]. To guarantee that $\hat{\theta}_N$ exists asymptotically and converges to some limit, we make the following assumptions:

ASSUMPTION A.2: *The parameter space* Θ *is an open subset of* $\mathbb{R}^p$.

ASSUMPTION A.3:

(*a*) *There exists a real function* ρ, *defined on* $\mathcal{Z} \times \Theta$ *and integrable with respect to any* $F \in \mathcal{F}_0$, *and an open neighborhood* $\mathcal{N} \subseteq \Theta$ *such that* $N^{-1} \Sigma_{n=1}^{N} \rho_N(Z_n,\theta)$ *converges almost surely-F, uniformly on* $\mathcal{N}$, *to* $E_F\, \rho(Z,\theta)$.

(*b*) $\forall\, F \in \mathcal{F}_0$, $E_F\, \rho(Z,\cdot)$ *has a unique global maximum on* $\mathcal{N}$.

(*c*) $\forall\, z \in \mathcal{Z}$, $\rho(z,\cdot)$ *is continuous on* $\mathcal{N}$ *and its right derivative, denoted by* $\eta(z,\cdot)$, *exists everywhere on* $\mathcal{N}$ *and is square integrable with respect to any* $F \in \mathcal{F}_0$.

(*d*) $\forall\, F \in \mathcal{F}_0$, $E_F\, \rho(Z,\cdot)$ *is differentiable on* $\mathcal{N}$ *and* $(\partial/\partial\theta)\, E_F\, \rho(Z,\cdot) = E_F\, \eta(Z,\cdot)$.

We shall associate with the 'score' $\eta(z,\theta)$ a functional $\hat{\theta}(\cdot)\colon \mathcal{F}_0 \rightarrow \mathcal{N}$, implicitly defined by the 'asymptotic first order condition' $E_F\, \eta(Z,\hat{\theta}(F)) = 0$, where $\hat{\theta}(F)$ is the unique root that corresponds to the global maximum of E_F

$\rho(Z,\theta)$. The functional $\hat{\theta}$ is assumed to be 'regular' in the following sense:

ASSUMPTION A.4:

(*a*) $\hat{\theta}$ *can be extended to some convex set* $\mathcal{F}_*$ *containing* $\mathcal{F}_0$ *and all empirical d.f.'s on* $\mathcal{Z}$.

(*b*) $\hat{\theta}$ *is Fisher-consistent, that is,* $\hat{\theta}(F) = \overline{\theta}(F)$, $\forall\ F \in \mathcal{F}$.

(*c*) $\forall\ F \in \mathcal{F}_*$, $E_F\ \eta(Z,\cdot)$ *is differentiable* on $\mathcal{N}$ *and the matrix* $P(F) = -(\partial/\partial\theta')\ E_F\ \eta(Z,\hat{\theta}(F))$ *exists and is positive definite for all* $F \in \mathcal{F}_0$.

(*d*) *There exists a function* ψ, *defined on* $\mathcal{Z} \times \mathcal{F}_0$ *and integrable with respect to any* $G \in \mathcal{F}_*$, *such that* $\lim_{\varepsilon \to 0} [\hat{\theta}((1 - \varepsilon)\ F + \varepsilon\ G) - \hat{\theta}(F)]/\varepsilon$ *exists and is equal to* $E_G\ \psi(Z,F)$ *for all* $F \in \mathcal{F}_0$.

(*e*) $\hat{\theta}(F_N) - \hat{\theta}(F_0) = N^1\ \Sigma_{n=1}^{N}\ \psi(Z_n,F_0) + o_P(N^{1/2})$.

A.4(a) justifies replacing $\hat{\theta}_N$ by the asymptotically equivalent functional $\hat{\theta}(F_N)$, where F_N denotes the empirical d.f. of the observations. Convexity of $\mathcal{F}_*$ is needed because $\hat{\theta}$ is to be evaluated at 'ε-contamination models' of the form $(1 - \varepsilon)\ F + \varepsilon\ G$, where $\varepsilon \in [0,1]$. A.4(b) and the weak convergence of F_N to F_0 imply that $\hat{\theta}(F_N)$ is consistent for $\overline{\theta}(F_0)$ whenever the model is correctly specified and $\hat{\theta}$ is sufficiently smooth. A.4(c)-(d) provide the necessary smoothness conditions. Finally, the asymptotic linearity assumption A.4(e) implies that $\hat{\theta}(F_N)$ is $\sqrt{N}$-consistent and asymptotically normal. This set of assumptions is satisfied by most M-estimators, including regular ML and GMM estimators [see e.g. Serfling (1980) and Fernholz (1983)].

Even if an estimator is consistent at the assumed model, slight violations of the model assumptions may result in a bias. As a quantitative measure of local robustness, Hampel (1974) and Rousseuw (1981) proposed the supremum of the asymptotic bias of $\hat{\theta}$, as an estimator of $\hat{\theta}(F)$, under an arbitrarily small contamination of the d.f. F by a point mass.

DEFINITION 3: Let $\hat{\theta}$ be a functional defined on $\mathcal{F}_*$ and let $\Delta_{(z)}$ be the d.f. with mass concentrated at $z \in \mathcal{Z}$. Then the *influence function* (IF) of $\hat{\theta}$ at the d.f. $F \in \mathcal{F}_*$ is defined by

$$\mathrm{IF}(z,\hat{\theta},F) = \lim_{\varepsilon \to 0} [\hat{\theta}((1 - \varepsilon)\ F + \varepsilon\ \Delta_{(z)})) - \hat{\theta}(F)] / \varepsilon,$$

provided that the limit exists. If $\mathrm{IF}(z,\hat{\theta},F)$ exists for all $z \in \mathcal{Z}$, then $\gamma^*(\hat{\theta},F) = \sup_{z \in \mathcal{Z}} \| \mathrm{IF}(z,\hat{\theta},F) \|$, where $\|\cdot\|$ denotes some norm on $\mathbb{R}^p$, is called the *(gross-error) sensitivity* of $\hat{\theta}$. If $\gamma^*(\hat{\theta},F) < \infty$, then $\hat{\theta}$ is called *B-(bias-) robust* at F.

Given A.4(d), $\mathrm{IF}(z,\hat{\theta},F)$ exists for all $z \in \mathcal{Z}$ and $F \in \mathcal{F}_0$ and is equal to $\psi(z,F)$. Assumption A.4(c) then implies that $\mathrm{IF}(\cdot,\hat{\theta},F) = P(F)^{-1}\,\eta(\cdot,\hat{\theta}(F))$ [see e.g. Serfling (1980)]. Thus, $\hat{\theta}$ is B-robust at F if and only if its score function is bounded. Further, since

$$\mathrm{E}_F\ \mathrm{IF}(Z,\hat{\theta},F)\ \mathrm{IF}(Z,\hat{\theta},F)' = P(F)^{-1}\ Q(F)\ (P(F)^{-1})',$$

where $Q(F) = \mathrm{E}_F\ \eta(Z,\hat{\theta}(F))\ \eta(Z,\hat{\theta}(F))'$, A.4(e) implies that $N^{1/2}\ [\hat{\theta}(F_N) - \hat{\theta}(F_0)] \xrightarrow{d} N(0, P_0^{-1}\ Q_0\ P_0'^{-1})$, where $P_0 = P(F_0)$, $Q_0 = Q(F_0)$, and $\xrightarrow{d}$ denotes convergence in distribution as $N \longrightarrow \infty$.

Testable hypotheses concerning the data are often defined by a set of smooth but possibly non linear restrictions on the parameter of interest.

HYPOTHESIS $\mathcal{H}_0$: *$F_0 \in \mathcal{F}$ and $h(\bar{\theta}(F_0)) = 0$, where $h\colon \Theta \longrightarrow \mathbb{R}^q$ is a continuously differentiable function with Jacobean matrix of full rank $q \leq p$.*

Often attention focuses on a subset of θ. If θ is partitioned as $(\beta', \gamma')'$, where β is the parameter of primary interest and γ is a nuisance parameter, our set-up covers the case when $\mathcal{H}_0$ only places restrictions on β, as in Holly (1987). It also covers the case when the restrictions are in either 'mixed' or 'explicit' form, as in Gourieroux and Monfort (1985).

Under our set of assumptions, imposing $\mathcal{H}_0$ in estimation gives two functionals, $\hat{\theta}_0$ and $\hat{\lambda}_0$, implicitly defined on $\mathcal{F}_*$ by the 'asymptotic first-order conditions'

$$\mathrm{E}_F\ \eta(Z,\hat{\theta}_0(F))\ -\ [(\partial/\partial\theta)\ h(\hat{\theta}_0(F))]\ \hat{\lambda}_0(F) = 0 \qquad (1)$$

$$h(\hat{\theta}_0(F)) = 0 \qquad (2)$$

where $\hat{\theta}_0(F)$ corresponds to the restricted M-estimator of $\theta(F)$, and $\hat{\lambda}_0(F)$ corresponds to the Lagrange multiplier associated with the constraint. We also consider three other functionals defined on $\mathcal{F}_*$, namely

$$\hat{h}(F) = h(\hat{\theta}(F))$$

$$\hat{s}_0(F) = S \cdot \mathrm{E}_F\, \eta(z,\hat{\theta}_0(F))$$

$$\hat{\Delta}(F) = D \cdot [\hat{\theta}(F) - \hat{\theta}_0(F)],$$

where S is some $r \times p$ matrix with rank $r \leq p$, and D is some $k \times p$ matrix with rank $k \leq p$. These functionals represent the basis for Wald, score and Hausman-type tests respectively. First we derive expressions for the IF of these functionals. All proofs are gathered in the Appendix.

PROPOSITION 1: *Assume that A.1-A.4 hold. If F satisfies $\mathcal{H}_0$, then*

$$\mathrm{IF}(z,\hat{h},F) = H(F) \cdot \mathrm{IF}(z,\hat{\theta},F)$$

$$\mathrm{IF}(z,\hat{s}_0,F) = S\ [H(F)'\ R(F)\ H(F)] \cdot \mathrm{IF}(z,\hat{\theta},F)$$

$$\mathrm{IF}(z,\hat{\Delta},F) = D\ [P(F)^{-1}\ H(F)'\ R(F)\ H(F)] \cdot \mathrm{IF}(z,\hat{\theta},F).$$

where $H(F) = (\partial/\partial\theta')\ h(\hat{\theta}(F))$ and $R(F) = [H(F)\ P(F)^{-1}\ H(F)']^{-1}$.

Thus, if $\hat{t}$ is any of the functionals $\hat{h}$, $\hat{s}_0$ or $\hat{\Delta}_0$, the IF of $\hat{t}$ at a distribution F which satisfies $\mathcal{H}_0$ is a (possibly singular) linear transformation of the IF of the unrestricted estimator $\hat{\theta}$, and so $\hat{t}$ inherits both the asymptotic normality and the robustness properties of $\hat{\theta}$. In particular, if $\hat{\theta}$ is B-robust at F, then $\hat{t}$ is also B-robust at F. Further, if F_0 satisfies $\mathcal{H}_0$, then $N^{1/2}\ \hat{t}(F_N) \xrightarrow{d} N(0,\ \mathrm{AV}(\hat{t},F_0))$, where $\mathrm{AV}(\hat{t},F_0) = T_0\ P_0^{-1}\ Q_0\ P_0'^{-1}\ T_0'$ and T_0 is the matrix defining the linear mapping from $\mathrm{IF}(\cdot,\hat{\theta},F_0)$ to $\mathrm{IF}(\cdot,\hat{t},F_0)$. Notice that $\mathrm{AV}(\hat{t},F_0)$ may be singular even if $P_0^{-1}\ Q_0\ P_0'^{-1}$ is positive definite. In the case of $\hat{s}_0$, this typically occurs when $r > q$, and in the case of $\hat{\Delta}$ when $k > q$.

3. M-TESTS AND THEIR ASYMPTOTIC PROPERTIES

Any of the functionals $\hat{h}$, $\hat{s}_0$ or $\hat{\Delta}$ can be used to construct tests of the hypothesis $\mathscr{H}_0$ against the alternative that $F_0 \in \mathscr{F}$ but $h(\theta(F_0)) \neq 0$.

DEFINITION 4: An *M-test statistic* is a quadratic form

$$\xi_N = N\, \hat{t}_N'\, A_N\, \hat{t}_N$$

where $\hat{t}_N = \hat{t}(F_N)$, $\hat{t}$ is any of the functionals $\hat{h}$, $\hat{s}_0$ or $\hat{\Delta}$, and A_N is an estimate of the generalized (g-) inverse of $\mathrm{AV}(\hat{t},F_0)$. A test that rejects $\mathscr{H}_0$ for large values of ξ_N is called an *M-test.*

The use of a g-inverse of $\mathrm{AV}(\hat{t},F_0)$, denoted by -, is necessary because the asymptotic variance of $\hat{t}$ may be singular. When $\hat{t}$ is equal to $\hat{h}$ one obtains the Wald test statistics. When $\hat{t}$ is equal to $\hat{s}_0$ one obtains a test statistic based on some linear combination of the elements of the score vector, such as the classical score test statistic and Neyman's $C(\alpha)$-test statistic [Neyman (1958)]. Finally, when $\hat{t}$ is equal to $\hat{\Delta}$ one obtains the test statistic proposed, among others, by Hausman (1978). The class of M-tests therefore contains most common tests in econometrics, with the only exception of likelihood ratio-type tests, such as the robust tests proposed by Ronchetti (1982) for the linear model. These tests are excluded because their asymptotic distribution may involve mixtures of χ^2 variates [Foutz and Srivastava (1977), Holly (1987) and Basawa and Koul (1988)], in which case computation of tail probabilities is complicated.

In order to study the local robustness properties of M-tests, we first derive the asymptotic distribution of the statistic ξ_N under a sequence of contaminated local alternatives of the null hypothesis $\mathscr{H}_0$.

DEFINITION 5: Let F_0 satisfy $\mathscr{H}_0$, and let $\{F_{0N}\}$ be a sequence of d.f.'s in $\mathscr{F}$ converging to F_0 and such that $\theta(F_{0N}) = \theta(F_0) + N^{-1/2}\,\delta$. A *sequence of contaminated local alternatives* $\{F_{\varepsilon,N,G}\}$ is a sequence of d.f.'s in $\mathscr{F}_*$ such that $F_{\varepsilon,N,G} = (1 - \varepsilon_N)\, F_{0N} + \varepsilon_N\, G$, where $\varepsilon_N = N^{-1/2}\varepsilon$, $\varepsilon \in [0,1]$.

This formalizes the notion that the assumed model may be adequate for the majority but not all the observations. In particular, putting $G = \Delta_{(z)}$ gives a way of modelling the occurrence of outliers and gross-errors. Notice that contamination is allowed to vanish asymptotically at the same rate with which $\theta(F_{0N})$ converges to $\theta(F_0)$. This is a standard way to prevent its effects from becoming negligible or dominating completely as the sample size increases [see e.g. Rieder (1978), Wang (1981) and Ronchetti (1982)]. When $\varepsilon = 0$ we have a standard sequence of local alternatives. When $\varepsilon > 0$ and $\delta = 0$ the common d.f. of the observations is in a shrinking neighborhood of the set of d.f's specified by $\mathscr{H}_0$.

PROPOSITION 2: *Assume that A.1-A.4 hold. Further assume that* A_N *converges in probability to* $\mathrm{AV}(\hat{t},F_0)^-$ *under the sequence of contaminated local alternatives* $\{F_{\varepsilon,N,G}\}$. *Then, for any choice of g-inverse of* $\mathrm{AV}(\hat{t},F_0)$, *the asymptotic distribution of* ξ_N *under the sequence of contaminated local alternatives is a non-central* χ^2 *with number of degrees of freedom equal to the rank of* $\mathrm{AV}(\hat{t},F_0)$ *and non-centrality parameter*

$$\nu(\delta,\varepsilon,G) = \varphi' T_0' \, \mathrm{AV}(\hat{t},F_0)^- \, T_0\varphi,$$

where $\varphi = \delta + \varepsilon \, \mathrm{E}_G \, \mathrm{IF}(Z,\hat{\theta},F_0)$. *Moreover, the non-centrality parameter is invariant for any choice of g-inverse of* $\mathrm{AV}(\hat{t},F_0)$, *and is equal to zero if and only if* $T_0\varphi = 0$.

When $\varepsilon = 0$ (no contamination) or $\mathrm{E}_G \, \mathrm{IF}(Z,\hat{\theta},F_0) = 0$, Proposition 2 contains as special cases a number of well known results, including the ones of Newey (1985), Gourieroux and Monfort (1985), and Holly (1987). When $\varepsilon > 0$, the effect of asymptotically vanishing contamination is only to modify the non-centrality parameter of the asymptotic χ^2 distribution of ξ_N.

Let τ_α be an asymptotic α-level M-test defined by the rejection region $\{\xi_N \in \mathbb{R}_+ : \xi_N > \lambda_\alpha\}$, where λ_α denotes the upper α-th quantile of a central χ^2 distribution with $\mu = \text{rank } \mathrm{AV}(\hat{t},F_0)$ degrees of freedom. By Proposition 2, the nominal asymptotic local power of τ_α against a specific alternative δ is given by $\pi(\delta) = 1 - \chi^2_{\mu,\nu(\delta,0,G)}(\lambda_\alpha)$, where $\chi^2_{\mu,\nu}$ denotes the non-central χ^2

d.f. with μ degrees of freedom and non-centrality parameter ν. Under the sequence of contaminated local alternatives $\{F_{\varepsilon,N,G}\}$, the actual asymptotic local power of τ_α against a specific alternative δ is given by $\pi(\delta,\varepsilon,G) = 1 - \chi^2_{\mu,\nu(\delta,\varepsilon,G)}(\lambda_\alpha)$, and its actual asymptotic level by $\alpha(\varepsilon,G) = \pi(0,\varepsilon,G)$. In general $\pi(\delta,\varepsilon,G)$ will differ from $\pi(\delta)$, and $\alpha(\varepsilon,G)$ from α. We now consider the relationship between the actual asymptotic level and power of τ_α and the robustness properties of $\hat{t}$.

PROPOSITION 3: *If the condition of Thorem 2 are satisfied, then under the sequence of contaminated local alternatives* $\{F_{\varepsilon,N,G}\}$:

(*a*) $\alpha(\varepsilon,G) \geq \alpha$.

(*b*) $\sup_{G \in \mathscr{F}_*} \alpha(\varepsilon,G) < 1$ *only if* $\hat{t}$ *is B-robust.*

(*c*) $\inf_{G \in \mathscr{F}_*} \pi(\delta,\varepsilon,G) = \alpha$.

Parts (a) and (b) assert, respectively, that the actual asymptotic level of τ_α can be greater than the nominal level α, and that B-robustness of $\hat{t}$ is sufficient to guarantee that the actual probability of Type I error is bounded away from one. Part (c) asserts that there are alternatives against which the test has no power under model contamination.

PROPOSITION 4: *If the conditions of Theorem 2 are satisfied, then under the sequence of contaminated local alternatives* $\{F_{\varepsilon,N,G}\}$:

(*a*) *The actual asymptotic local power of* τ_α *is smaller than the nominal one against any alternative* δ *such that* $(\cos \kappa)\ \|T_0\delta\|_{AV} < -\ (\varepsilon/2)\ \|E_G\ IF(z,\hat{t},F_0)\|$, *where the (pseudo) norm* $\|\cdot\|$ *and the angle* κ *between the vectors* $T_0\delta$ *and* $E_G\ IF(z,\hat{t},F_0)$ *are defined in the metric of* $AV(\hat{t},F_0)^-$.

(*b*) τ_α *is asymptotically biased against any alternative* δ *such that* $\|T_0\delta\| < -2\ \varepsilon \cos \kappa\ \|E_G\ IF(z,\hat{t},F_0)\|$.

(*c*) τ_α *has no asymptotic power against any alternative* δ *such that* $T_0\delta = -\ \varepsilon\ E_G\ IF(z,\hat{t},F_0)$.

This result establishes a relationship between the magnitude and the direction of the asymptotic bias of $\hat{t}$, and the set of departures from the null hypothesis against which τ_α loses power under model contamination. In particular, τ_α may loose power against alternatives that are in the opposite direction to the asymptotic bias of $\hat{t}$, and is biased against alternatives that are in the opposite direction to the asymptotic bias of $\hat{t}$ but not too far from the null hypothesis. For alternatives that are in the same direction as the asymptotic bias of $\hat{t}$, τ_α is unbiased and has greater asymptotic power than nominal.

4. AN EXAMPLE

As an example, let $Z = (Y, X')'$ and consider the problem of testing for $\theta_0 = 0$ in the classical orthonormal linear regression model $Y = X'\theta_0 + U$, where the unobservable random variable U has a $N(0,1)$ distribution and $E\, XX' = I_p$. If the assumed model is correctly specified, optimal tests can be based on the ordinary LS estimator $\hat{\theta}$. The score and the IF associated with $\hat{\theta}$ are both equal to $x(y - x'\theta)$. Clearly, $\hat{\theta}$ does not have a bounded IF and so it is not B-robust. It is easily seen that all M-test statistics based on $\hat{\theta}$ are numerically the same, and their non-centrality parameter under a sequence of contaminated local alternatives $\{F_{\varepsilon,N,\Delta_{(z)}}\}$ is equal to

$$\nu(\delta,\varepsilon,\Delta_{(z)}) = \delta'\delta + 2\,\varepsilon\, u\, \delta'x + \varepsilon^2\, u^2\, x'x,$$

where $z = (y, x')'$ and $u = y - x'\theta_0$.

We shall compare tests based on $\hat{\theta}$ with those based on the Huber estimator of regression $\tilde{\theta}$ [see e.g. Huber (1981)]. The score associated with $\tilde{\theta}$ is equal to $x\, \psi_c(y - x'\theta)$, where $\psi_c(u) = \min\,(c, \max\,(-c, u))$ and c is a finite, positive constant. The IF of $\tilde{\theta}$ at the assumed Gaussian model is therefore equal to $[2\Phi(c) - 1]^{-1}\, x\, \psi_c(y - x'\theta)$, where Φ denotes the $N(0,1)$ d.f. The M-test statistics based on $\tilde{\theta}$ are not the same numerically, but under our sequence of contaminated local alternatives, they all have the same asymptotic χ^2 distribution with non-centrality parameter

$$\tilde{\nu}(\delta,\varepsilon,\Delta_{(z)}) = K(c)\, [\delta'\delta + 2\,\varepsilon\, \tilde{u}\, \delta'x + \varepsilon^2\, \tilde{u}^2\, x'x]$$

where $K(c) = [2\Phi(c) - 1]^2 / [E_\Phi\, \psi_c(U)^2] < 1$ and $\tilde{u} = \psi_c(u) / [2\Phi(c) - 1]$.

Let τ and $\tilde{\tau}$ denote two asymptotic α-level M-tests based, respectively, on $\hat{\theta}$ and $\tilde{\theta}$. The different asymptotic behavior of τ and $\tilde{\tau}$ under ε-contamination depends on the differences between the IF's of the two estimators. This behavior is summarized in Table 1. An illustration is provided in Figure 1. First consider the case when point-mass contamination occurs at a point (y_1, x) such that $u_1 = y_1 - x'\theta_0 = c$. Let $\mathscr{L}_{LS}$ and $\mathscr{B}_{LS}$ denote the set of alternatives against which τ, based on the LS estimator, respectively loses power and is biased, and let analogous sets $\mathscr{L}_H$ and $\mathscr{B}_H$ be defined for $\tilde{\tau}$, based on the Huber estimator. In this case there are only small differences between τ and $\tilde{\tau}$. In particular, since $u_1/\tilde{u}_1 = 2\Phi(c) - 1 < 1$, $\mathscr{L}_H$ is slightly smaller than $\mathscr{L}_{LS}$ and $\mathscr{B}_H$ is slightly larger than $\mathscr{B}_{LS}$.

Now consider the case when contamination occurs at a point (y_2, x) such that $u_2 = y_2 - x'\theta = 2c$. Since $\tilde{u}_2 = \tilde{u}_1$, the behavior of $\tilde{\tau}$ does not change with respect to the first case. The behavior of τ, however, is altered dramatically. In particular, since $u_2/\tilde{u}_2 = 2[2\Phi(c) - 1] > 1$, the set of alternatives against which τ is biased is now much broader than for $\tilde{\tau}$. The difference between the two tests reflects the fact that while the Huber estimator is robust under this particular form of contamination, the LS estimator is not.

5. LOCALLY ROBUST M-TESTS

Clearly, what is crucial for a test is the effect of contamination on the decision of rejecting or not rejecting the null hypothesis. This leads quite naturally to a definition of local robustness in terms of the effects of a small amount of contamination on the asymptotic level and power of a test. Related approaches include Lambert (1981), where robustness is defined in terms of the p-value of a test, and Field and Ronchetti (1985), where robustness is defined in terms of the finite sample tail area of a test.

DEFINITION 6: The *IF of the asymptotic level* and the *IF of the asymptotic local power* of an M-test τ at the d.f. $F \in \mathscr{F}_*$ are defined by

$$\mathrm{IF}_L(z,\tau,F) = (\partial^2/\partial\varepsilon^2)\ \pi(0,\varepsilon,\Delta_{(z)})\,|_{\varepsilon=0}$$

$$IF_P(z,\tau,\delta,F) = (\partial/\partial\varepsilon)\ \pi(\delta,\varepsilon,\Delta_{(z)})\big|_{\varepsilon=0}$$

provided that the appropriate limits exist. If $IF_L(z,\tau,F)$ and $IF_P(z,\tau,\delta,F)$ exist for all $z \in \mathcal{Z}$, then $\gamma^*_L(\tau,F) = \sup_{z\in\mathcal{Z}} |IF_L(z,\tau,F)|$ and $\gamma^*_P(\tau,F) = \sup_{z\in\mathcal{Z},\delta\in\mathbb{R}^p} |IF_P(z,\tau,\delta,F)|$ are called respectively the *level-* and *power-sensitivity* of τ. An M-test τ is called *locally robust* at the d.f. F if $\gamma^*_L(\tau,F)$ and $\gamma^*_P(\tau,F)$ are both finite.

The definition of IF_P is analogous to the one proposed by Ronchetti (1982). The definition of IF_L is different, for it involves second derivatives. This is necessary because, for M-tests, the linear term in the expansion of $\pi(0,\varepsilon,\Delta_{(z)})$ about $\varepsilon = 0$ is equal to zero. $\gamma^*_L(\tau,F)$ and $\gamma^*_P(\tau,F)$ provide a quantitative measure of the local robustness properties of τ. When τ is locally robust its asymptotic level and power change little under a small amount of contamination, and these changes can be approximated using IF_P and IF_L. The next result provides expressions for the IF_L and IF_P of an M-test τ in terms of the IF of the statistic $\hat{t}$ on which the test is based.

PROPOSITION 5: *Let τ be an M-test based on the statistic $\hat{t}$. Then, under the conditions of Theorem 2,*

$$IF_L(z,\tau,F) \propto IF(z,\hat{t},F)'\ AV(\hat{t},F)^-\ IF(z,\hat{t},F)$$

$$IF_P(z,\tau,\delta,F) \propto IF(z,\hat{t},F)'\ AV(\hat{t},F)^-\ IF(z,\hat{t},F).$$

This Proposition, which generalizes the results obtained by Hampel et al. (1986) for one-sided tests of a single restriction, implies that, for a small ε, the divergence between the actual and the nominal asymptotic level and power of τ are both proportional to the squared norm of the IF of $\hat{t}$ in the metric of $AV(\hat{t},F)^-$. Thus, both the level- and power-sensitivity of τ are proportional to the square of the so-called self-standardized sensitivity of $\hat{t}$ [Krasker and Welsch (1982)], and so τ is locally robust at F if and only if $\hat{t}$ is B-robust. Finally, Propositions 1 and 5 together imply that if $\hat{\theta}$ is B-robust at F, then any M-test based on $\hat{\theta}$ is locally robust at F. This provides a formal justification, in terms of robustness, for using M-tests based on some B-robust estimator.

6. OPTIMAL LOCALLY ROBUST M-TESTS

In general, one is interested in optimality as well as robustness of a statistical procedure. We shall discuss optimality with reference to a parametric model $\mathscr{F} = \{F_\theta, \theta \in \Theta\}$, with likelihood score function $s(z,\theta)$ and a finite positive definite Fisher information matrix. If $\mathscr{F}$ is correctly specified, M-tests based on the ML estimator are known to be asymptotically locally most powerful invariant. However, since the likelihood score need not be bounded, these tests need not be robust in general. Thus, consider the following question: Given a class of asymptotic M-tests with the same nominal level and the same local robustness properties, can we find one which is locally most powerful invariant?

Since M-tests inherit the efficiency and local robustness properties of the estimators on which they are based, it seems reasonable to consider tests based on B-robust estimators with maximal asymptotic precision. Given a p.d. matrix M and a finite, positive constant c, an estimator that minimizes the trace of $[\mathrm{AV}(\hat\theta,F_\theta)\ M]$ among those for which $\gamma^*(\hat\theta,F_\theta) < c$, is called optimal B-robust at $\mathscr{F}$ with respect to (M,c). In the one-dimensional case ($p = 1$), Ronchetti (1982) showed that an optimal B-robust estimator does lead to robust tests that are locally most powerful invariant. We now show that his result does not extend to general M-tests.

For concreteness and without any loss of generality, let $\mathscr{T}$ be the class of Wald tests with the same asymptotic level at the assumed model. The various tests in $\mathscr{T}$ differ with respect to the choice of the unrestricted M-estimator $\hat\theta$. Let $\mathscr{R} \subset \mathscr{T}$ be the subset of locally robust Wald tests with the same level- and power-sensitivity at the assumed model. By Proposition 5, a test belongs to $\mathscr{R}$ if it is based on an estimator $\hat\theta$ such that $\gamma^*(\hat\theta,F_\theta) < c$, where c depends on the given sensitivity bound for the test. Let $\hat{t}(\cdot) = h(\hat\theta(\cdot))$ and let T_θ be the matrix defining the linear mapping from $\mathrm{IF}(\cdot,\hat\theta,F_\theta)$ to $\mathrm{IF}(\cdot,\hat{t},F_\theta)$. Then a test in $\mathscr{R}$ is most powerful invariant if $\hat\theta$ maximizes the non-centrality parameter $\delta'\ T_\theta'\ [T_\theta\ \mathrm{AV}(\hat\theta,F_\theta)\ T_\theta']^-\ T_\theta\ \delta$ for all directions

δ and all values of θ, or equivalently, if $\hat{\theta}$ is optimal B-robust with respect to (M,c) for all choices of M. It turns out that, for $p > 1$ and $c < \infty$, an optimal B-robust estimator depends on the choice of the matrix M and the norm $\|\cdot\|$ for the estimator's sensitivity. Thus, one can only prove the following:

PROPOSITION 6: *Let* $\tilde{\tau} \in \mathcal{R}$ *be a locally robust test based on an estimator that is optimal B-robust with respect to* (M,c), *and let* $\tilde{\pi}(\cdot)$ *denote its asymptotic local power function at the assumed model* $\mathcal{F}$. *Then, there exists no other robust test* $\tau \in \mathcal{R}$, *with asymptotic local power function* $\pi(\cdot)$ *at* $\mathcal{F}$, *such that* $\pi(\delta) \geq \tilde{\pi}(\delta)$ *for all* δ, *and* $\pi(\delta) > \tilde{\pi}(\delta)$ *for some* δ.

This Proposition, which generalizes the result obtained by Hampel et al. (1986) [Proposition 5, Section 7.3] for tests of linear restrictions in the linear model, provides a formal justification for using M-tests based on some optimal B-robust estimator. It shows that $\tilde{\tau}$ is asymptotically 'admissible', that is, no other robust test in $\mathcal{R}$ is asymptotically as powerful as $\tilde{\tau}$ against all alternatives, and more powerful against some of them. As the proof makes it clear, $\tilde{\tau}$ has maximal asymptotic local power in $\mathcal{R}$ against some alternatives, but not against all. Thus, choosing a particular optimal B-robust estimator implicitly corresponds to choosing a set of alternatives against which the resulting robust test has maximal power.

7. CONCLUSIONS

Since statistical models are at best an approximation to the true process generating the data, it would be desirable that inference remained stable under small deviations from the model assumptions. Unfortunately, this is not necessarily the case for tests based on estimators whose IF is not bounded. We show that this result holds for a broad class of tests that includes most commonly used tests in econometrics. On the other hand, tests based on B-robust estimators enjoy desirable local robustness properties, for both their level and power remain stable under small, but otherwise quite general departures from the assumed model. In addition, if the underlying estimator attains the best trade-off between robustness and efficiency, these tests also attain the best trade-off between robustness and power.

Appendix

PROOF OF PROPOSITION 1: Replacing F by $(1 - \varepsilon) F + \varepsilon \Delta_{(z)}$ in (1)-(2), differentiating with respect to ε and evaluating at $\varepsilon = 0$ gives

$$0 = - P \cdot \mathrm{IF}(z,\hat{\theta}_0,F) + \eta(z,\hat{\theta}_0(F)) - H' \; \mathrm{IF}(z,\hat{\lambda}_0,F) \tag{3}$$

$$0 = H \cdot \mathrm{IF}(z,\hat{\theta},F_0),$$

where $P = P(F)$, $H = (\partial/\partial\theta') \; h(\hat{\theta}(F))$, and we used the fact that, if F satisfies $\mathscr{H}_0$, then $\hat{\theta}_0(F) = \hat{\theta}_0(F)$ and $\hat{\lambda}_0(F) = 0$. Pre-multiplying (3) by P^{-1} and rearranging gives the equation system

$$\begin{bmatrix} I_p & P^{-1}H' \\ H & 0 \end{bmatrix} \begin{bmatrix} \mathrm{IF}(z,\hat{\theta}_0,F) \\ \mathrm{IF}(z,\lambda_0,F) \end{bmatrix} = \begin{bmatrix} \mathrm{IF}(z,\hat{\theta},F) \\ 0 \end{bmatrix}.$$

By the usual partitioned inverse formulae:

$$\mathrm{IF}(z,\hat{\theta}_0,F) = (I_p - P^{-1}H'R \; H) \cdot \mathrm{IF}(z,\hat{\theta},F)$$

$$\mathrm{IF}(z,\hat{\lambda}_0,F) = R \; H \cdot \mathrm{IF}(z,\hat{\theta},F)$$

where $R = (H \; P^{-1}H')^{-1}$. Therefore

$$\mathrm{IF}(z,\hat{h},F) = H \cdot \mathrm{IF}(z,\hat{\theta},F)$$

$$\mathrm{IF}(z,\hat{\Delta},F) = D \; P^{-1}H'R \; H \cdot \mathrm{IF}(z,\hat{\theta},F)$$

and, since $\hat{s}_0(F) = S \cdot \mathrm{E}_F \; \eta(z,\hat{\theta}_0(F)) = S \cdot H(\hat{\theta}_0(F))' \; \hat{\lambda}_0(F)$,

$$\mathrm{IF}(z,\hat{s}_0,F) = S \; H'R \; H \cdot \mathrm{IF}(z,\hat{\theta},F).$$

PROOF OF PROPOSITION 2: Consider the limiting distribution of the normalized difference $N^{1/2} \; [\hat{\theta}(F_N) - \hat{\theta}(F_0)]$ under the sequence of contaminated local alternatives $\{F_{\varepsilon,N,G}\}$. The distribution of $N^{1/2}\ [\hat{\theta}(F_N) - \hat{\theta}(F_{\varepsilon,N,G})]$ is asymptotically normal with mean zero and variance converging to $P_0^{-1}\ Q_0\ P_0'^{-1}$. By A.4(c),

$$\hat{\theta}(F_{\varepsilon,N,G}) - \hat{\theta}(F_{0N}) - \varepsilon_N\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{\theta},F_{0N}) = o(N^{-1/2}),$$

and by Fisher consistency [A.4(b)], $\hat{\theta}(F_{0N}) - \hat{\theta}(F_0) = N^{-1/2}\ \delta$. Thus $N^{1/2}$ $[\hat{\theta}(F_{\varepsilon,N,G}) - \hat{\theta}(F_0)] = \delta + \varepsilon\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{\theta},F_{0N}) + o(1)$, and so

$$\lim_{N\to\infty} N^{1/2}\ [\hat{\theta}(F_{\varepsilon,N,G}) - \hat{\theta}(F_0)] = \delta + \varepsilon\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{\theta},F_0) \equiv \varphi,$$

since $F_{0N} \to F_0$ as $N \to \infty$. Therefore $N^{1/2}\ [\hat{\theta}(F_N) - \hat{\theta}(F_0)] \xrightarrow{d} N(\varphi,\ P_0^{-1}\ Q_0\ P_0'^{-1})$ under the sequence of contaminated local alternatives. Because $N^{1/2}\ \hat{t}_N$ has the same asymptotic distribution as $N^{1/2}\ T_0\ [\hat{\theta}(F_N) - \hat{\theta}(F_0)]$, it follows that under the sequence of contaminated local alternatives, $N^{1/2}\ \hat{t}_N \xrightarrow{d}$ $N\ (T_0\varphi,\ \mathrm{AV}(\hat{t},F_0))$. Since $\mathrm{AV}(\hat{t},F_0) = T_0\ P_0^{-1}\ Q_0\ P_0'^{-1}\ T_0'$, the column space of T_0 is contained in the column space of $\mathrm{AV}(\hat{t},F_0)$, and so $T_0\varphi$ is contained in the column space of $\mathrm{AV}(\hat{t},F_0)$ for any $\varphi \in \mathbb{R}^P$. The conclusions then follow from Theorem 1 of Vuong (1987).

PROOF OF PROPOSITION 3: By Proposition 2, $\pi(\delta,\varepsilon,G)$ is a strictly increasing function of the non-centrality parameter $\nu(\delta,\varepsilon,G) = [T_0\delta + \varepsilon\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{t},F_0)]'\ \mathrm{AV}(\hat{t},F_0)^-\ [T_0\delta + \varepsilon\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{t},F_0)]$. Then $\alpha(\varepsilon,G) \geq \alpha$, because $\alpha(\varepsilon,G) = \pi(0,\varepsilon,G)$ and $\nu(0,\varepsilon,G)$ is non-negative. The second part of (a) follows from the fact that $\nu(0,\varepsilon,G)$ need not be finite if $\hat{t}$ is not B-robust. Part (b) follows from the fact that $\nu(\delta,\varepsilon,G) = 0$ if $\delta = -\ \varepsilon\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{\theta},F_0)$.

PROOF OF PROPOSITION 4: The actual asymptotic power of τ is smaller than the nominal whenever

$$\varphi' T_0'\ \mathrm{AV}(\hat{t},F_0)^-\ T_0\varphi < \delta' T_0'\ \mathrm{AV}(\hat{t},F_0)^-\ T_0\delta$$

$$\Leftrightarrow\quad 2\ \varepsilon\ \delta' T_0'\ \mathrm{AV}(\hat{t},F_0)^-\ \mathrm{E}_G\ \mathrm{IF}(z,\hat{t},F_0)\ +$$

$$+ \varepsilon^2 \, E_G \, IF(z,\hat{t},F_0)' \, AV(t,F_0)^- \, E_G \, IF(z,\hat{t},F_0) < 0$$

$$\Leftrightarrow \quad 2 \, \varepsilon \cos \kappa \, \| T_0\delta \| \, \| E_G \, IF(z,\hat{t},F_0) \| + \varepsilon^2 \, \| E_G \, IF(z,\hat{t},F_0) \|^2 < 0$$

$$\Leftrightarrow \quad (\cos \kappa) \, \| T_0\delta \| < - (\varepsilon /2) \, \| E_G \, IF(z,\hat{t},F_0) \|,$$

where the (pseudo) norm $\| \cdot \|$ and the angle κ between the two vectors $T_0\delta$ and $E_G \, IF(z,\hat{t},F_0)$ are both defined in the metric of $AV(\hat{t},F_0)^-$.

Finally, τ is asymptotically biased whenever

$$\varphi' T_0' \, AV(\hat{t},F_0)^- \, T_0\varphi < \varepsilon^2 \, E_G \, IF(z,\hat{t},F_0)' \, AV(\hat{t},F_0)^- \, E_G \, IF(z,\hat{t},F_0)$$

$$\Leftrightarrow \quad \delta' T_0' \, AV(\hat{t},F_0)^- \, T_0\delta + $$
$$+ 2 \, \varepsilon \, \delta' T_0' \, AV(\hat{t},F_0)^- \, E_G \, IF(z,\hat{t},F_0) < 0$$

$$\Leftrightarrow \quad \| T_0\delta \|^2 + 2 \, \varepsilon \cos \kappa \, \| T_0\delta \| \, \| E_G \, IF(z,\hat{t},F_0) \| < 0$$

$$\Leftrightarrow \quad \| T_0\delta \| < - 2 \, \varepsilon \cos \kappa \, \| E_G \, IF(z,\hat{t},F_0) \|.$$

τ has no power whenever $\varphi' T_0' \, AV(\hat{t},F_0)^- \, T_0\varphi = 0$, or, equivalently $T_0\delta = - \varepsilon \, E_G \, IF(z,\hat{t},F_0)$.

PROOF OF PROPOSITION 5: Immediate from the fact that the asymptotic local power of an M-test is a strictly increasing, continuously differentiable function of the non-centrality parameter.

PROOF OF PROPOSITION 6: The existence of an optimal test in $\mathcal{R}$ is equivalent to the existence of an M-estimator whose score function solves the following problem

$$\text{Min}_{\eta(\cdot,\theta) \in \mathcal{H}} \; E_\theta \, \eta(Z,\theta)' \, M \, \eta(Z,\theta) \tag{4}$$

s.t. $$E_\theta\ \eta(Z,\theta) = 0 \tag{5}$$

$$E_\theta\ \eta(Z,\theta)\ s(Z,\theta)' = I_p \tag{6}$$

$$\sup_{z \in \mathscr{Z}} \|\eta(z,\theta)\| \leq c \tag{7}$$

for a given $c < \infty$, all $\theta \in \Theta$ and all p.d. matrices M. The set $\mathscr{H}$ is the set of functions that are square integrable with respect to all $F \in \mathscr{F}$. The objective functional (4) is the asymptotic mean squared error (MSE) criterion. Constraint (5) requires an estimator to be Fisher-consistent, (6) is a normalization condition under which the IF and the score function of an estimator are identical, and (7) is the B-robustness condition. Problem (4)-(7) is the same as the minimum norm problem in Peracchi (1990). It follows from his results that when $c = \infty$, that is, the robustness constraint is vacuous, a solution always exists, is invariant to the choice of the matrix M, and is equal to the likelihood score for the assumed parametric model. When c is finite, a solution exists if c is sufficiently large, but is invariant to the choice of M only if $p = 1$. When $p > 1$, the solution depends on M and the norm $\|\cdot\|$ for the estimator's sensitivity. Therefore, the estimator $\tilde{\theta}$ that solves (4)-(7) is optimal only in a weak sense, namely with respect to a given MSE criterion. In turns, this implies that tests based on $\tilde{\theta}$ are locally most powerful only against the set of alternatives implicitly defined by the matrix M. The conclusions of the Proposition then follow from the fact that if A and B are $p \times p$ matrices, and trace $A\,M <$ trace $B\,M$ for some p.d. matrix M, then $A - B$ is not a positive semi-definite matrix.

References

Basawa, I.L. and Koul, H.L. (1988), "Large-sample statistics based on quadratic dispersion", *International Statistical Review*, 56, 199-219.

Fernholz, L.T. (1983), *Von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics No. 19, Springer, New York.

Field, C.A., and Ronchetti, E. (1985), "A tail area influence function and its application to testing", *Communications in Statistics, C*, 4, 19-41.

Foutz, R.V., and Srivastava, R.C. (1977), "The performance of the likelihood ratio test when the model is incorrect", *Annals of Statistics*, 5, 1183-1194.

Gourieroux, C. and Monfort, A. (1985), "A general theory for asymptotic tests", Document de Travail No. 8509, INSEE, Paris.

Hampel, F.R. (1974), "The influence curve and its role in robust estimation", *Journal of the American Statistical Association*, 69, 383- 393.

Hampel, F.R., Ronchetti, E., Rousseeuw, P.J., and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

Hansen, L.P. (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029-1054.

Hausman, J.A. (1978), "Specification tests in econometrics", *Econometrica*, 46, 1251-1272.

Holly, A. (1987), "Specification tests: An overview", in T.F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Vol. 1, Cambridge University Press, New York.

Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

Kent, J.T. (1982), "Robust properties of likelihood ratio tests", *Biometrika*, 69, 19-27.

Krasker, W.S. and Welsch, R.E. (1982), "Efficient bounded-influence regression estimation", *Journal of the American Statistical Association*, 77, 595-604.

Lambert, D. (1981), "Influence function for testing", *Journal of the American*

Statistical Association, 76, 649-657.

Newey, W.K. (1985), "Maximum likelihood specification testing and conditional moment tests", *Econometrica*, 53, 1047-1069.

Neyman, J. (1958), "Optimal asymptotic tests of composite statistical hypotheses", in U. Grenander (ed.), *Probability and Statistics: The Cramer Volume*, Almqvist and Wiksells, Uppsala, Sweden.

Peracchi, F. (1990), "Robust M-estimators", *Econometric Reviews*, 9, 1-30.

Rieder, H. (1978), "A robust asymptotic testing model", *Annals of Statistics*, 6, 1080-1094.

Ronchetti, E. (1982), Robust Testing in Linear Models: The Infinitesimal Approach, unpublished Ph.D. Thesis, ETH, Zurich.

Rousseeuw, P.J. (1981), "A new infinitesimal approach to robust estimation", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56, 127-132.

Rousseeuw, P.J., and Ronchetti, E. (1981), "Influence functions for general statistics", *Journal of Computational and Applied Mathematics*, 7, 161-166.

Schrader, R.M. and Hettmansperger, T.P. (1980), "Robust analysis of variance based upon a likelihood ratio criterion", *Biometrika*, 67, 93-101.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Vuong, Q.H. (1987), "Generalized inverses and asymptotic properties of Wald tests", *Economic Letters*, 24, 343-347.

Wang, P.C.C. (1982), "Robust asymptotic tests of statistical hypotheses involving nuisance parameters", *Annals of Statistics*, 9, 1096-1106.

Table 1

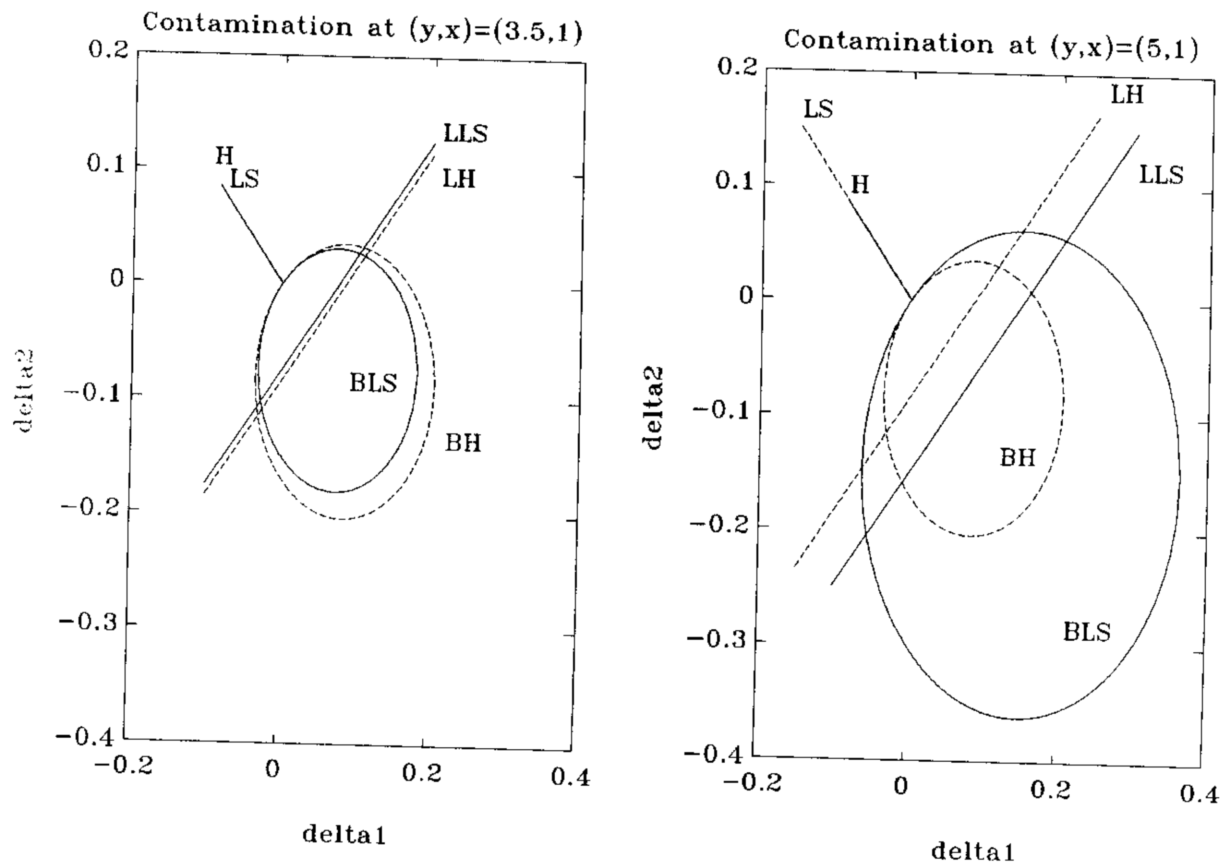
Asymptotic behavior of M-tests based on the LS and Huber regression estimators under sequences of ε-contaminated local alternatives, with contamination at (y, x).

Asymptotic behavior	Alternatives	
	LS	Huber
Loss of power	$\{\delta: \delta' x < -(\varepsilon/2)\, u x' x\}$	$\{\delta: \delta' x < -(\varepsilon/2)\, \tilde{u} x' x\}$
No power	$\delta = -\varepsilon\, u\, x$	$\delta = -\varepsilon\, \tilde{u}\, x$
Biased test	$\{\delta: \delta' x u < -(2\varepsilon)^{-1}\, \delta'\delta\}$	$\{\delta: \delta' x \tilde{u} < -(2\varepsilon)^{-1}\, \delta'\delta\}$

Note: $u = y - x'\theta_0$, $\tilde{u} = \psi_c(u) / [2\, \Phi(c) - 1]$.

Figure 1

Asymptotic behavior of M-tests based on the LS and Huber regression estimators under sequences of 5%-contaminated localalternatives.



Legend: LS: Bias of the LS estimator

H: Bias of the Huber estimator (with $c = 1.5$)

LLS:Loss of power for the LS-test

LH: Loss of power for the Huber-test

BLS:LS-test biased

BH: Huber-test biased