ROBUST M-TESTS

by

Franco Peracchi*

Department of Economics
University of California, Los Angeles

November 1987

UCLA Working Paper No. 459

*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 Trudy Cameron for helpful comments.

Abstract

This paper investigates the 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 local misspecification of the distribution of the observations may 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.

JEL No. 211

Keywords: M-tests, Robust tests.

1. Introduction

The problem of the robustness a test, i.e. the stability of its level and power under small changes in the distribution of the observations, has received considerable attention in the statistical literature [see e.g. Rieder (1978), Schrader and Hettmansperger (1980), Kent (1982), Ronchetti (1982), Wang (1982), and Hampel et al. (1986)], but has been largely ignored by econometricians.

This paper investigates the robustness properties of a class of multi-dimensional tests, called M-tests because they are based on M-estimators. This class of tests is defined more generally than in Newey (1985a), because we include Wald tests along with score and Hausman tests. We do not consider likelihood ratio-type tests, such as the robust tests proposed by Ronchetti (1982) for the linear model [see also Hampel et al. (1986)]. The asymptotic distribution of tests of this type may be quite complicated, since it may involve mixtures of χ^2 variates [see e.g. Holly (1986)]. On the other hand, all M-test statistics have a simple asymptotic χ^2 distribution for which tail probabilities are readily available.

We study the asymptotic properties of M-tests under sequences of contaminated local alternatives. Our approach builds on earlier work of Rieder (1978), Ronchetti (1982) and Wang (1982) for one-dimensional, one-sided tests, and on results of Hampel et al. (1986) for the linear model. It is also related to the work of Burguete, Gallant and Souza (1982), because they also consider sequences of misspecified alternatives, but we get sharper results by focusing on a particular type of model misspecification, namely contamination by gross-errors. We show that contamination may have very serious effects, leading to tests that are biased and have no power against some alternatives. However, we also obtain the result that contamination may 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 robust if a small amount of contamination has only a small effect on its asymptotic level and power. We show that robustness is guaranteed if the estimator on which the test statistic is based has a bounded influence function. For the one-dimensional case, Ronchetti (1982) showed that optimal bounded influence estimators lead to tests that are asymptotically locally most powerful invariant in the class of robust tests. For the more general case of multiple restriction, tests based on optimal bounded influence estimators are not invariant, and we can therefore only prove an asymptotic 'admissibility' result.

This paper is organized as follows. Section 2 derives the influence function and the asymptotic distribution of a number of statistics that may be used for testing purposes. Section 3 examines the asymptotic properties of M-tests under sequences of contaminated local alternatives and provides an example based on the linear regression model. Section 4 contains the asymptotic 'admissibility' result.

2. Restricted M-estimators

Let $z_1, \ldots, z_N$ be a sequence of independently and identically distributed (iid) random variables, with values in a known subset $\mathfrak{Z}$ of $\mathbb{R}^m$, and common but unknown distribution function (df) F_0. It is typically assumed that: (i) F_0 belongs to a known set $\mathcal{F}$ of df's over $\mathfrak{Z}$. (ii) The assumed model $\mathcal{F}$ is either a regular parametric model indexed by a p-dimensional parameter θ, i.e. $\mathcal{F} = \{F_\theta: \theta \in \Theta \subset \mathbb{R}^p\}$, or a family of distributions satisfying a number $s \geq p$ of moment restrictions, i.e. $\mathcal{F} = \{F: E_F\ \psi(z,\theta) = 0,\ \theta \in \Theta \subset \mathbb{R}^p\}$. (iii) The structure F_0 is identifiable, i.e. there is a unique $\theta_0 \in \Theta$ such that either $F = F_{\theta_0}$ or $E_0\ \psi(z,\theta) = 0$, where E_0 denotes expectations taken with respect to the df F_0. The assumed model $\mathcal{F}$ may be misspecified in the sense that it may not contain F_0. In this paper we focus on two types of model misspecification. The first, called the 'contamination model', arises when $F_0 = (1 - \varepsilon)\ F + \varepsilon\ H$, where $F \in \mathcal{F}$, H is an unknown df not belonging to $\mathcal{F}$ and $\varepsilon \in [0,1]$. When ε is small, this formalizes the notion that the assumed model $\mathcal{F}$ may be adequate for the majority but not all the observations. The second, called the 'gross-error model', arises when H gives unit mass to a point z in $\mathfrak{Z}$. This is a convenient way of modelling the occurrence of outliers or gross errors.

We consider a general class of estimators of the unknown parameter θ_0, namely the class of M-estimators defined as solutions to the problem

$$(1) \qquad \text{Max}_{\theta \in \Theta_0}\ N^{-1} \Sigma_{n=1}^{N}\ \rho_N(z_n,\theta)$$

where Θ_0 is a subset of $\mathbb{R}^p$ assumed to contain θ_0 in its interior, and the function $\rho_N(z,\theta)$, defined on $\mathfrak{Z} \times \Theta_0$, is continuous and almost everywhere twice-differentiable with respect to θ. The class of M-estimators is very large and includes most common econometric estimators, such as ML, least squares and generalized method of moments estimators.

Under regularity conditions that guarantee the existence of an interior solution, an M-estimator $\hat{\theta}_N$ is a root of the equation system

$$(2) \qquad N^{-1} \Sigma_{n=1}^{N} \eta_N(z_n,\hat{\theta}_N) = 0$$

corresponding to a global maximum. The function $\eta_N(z,\theta) = (\partial/\partial\theta)\ \rho_N(z,\theta)$ is defined almost everywhere and is called the score function associated with $\hat{\theta}_N$. Sometimes M-estimators are defined in terms of (2) rather than (1). The two definitions are essentially equivalent, for solving the equation (2) is the same as minimizing the norm of the sample average of $\eta(z,\theta)$.

Because of the independence assumption, the order of the observations in the sample does not matter. One may therefore replace functions of the observations by statistical functionals, i.e. functionals defined over a set of df's. For example, when η_N has an almost sure limit η, a functional $\hat{\theta}$: $F \longrightarrow \hat{\theta}(F)$ is implicitly defined by the asymptotic first- order conditions

$$\int_{\mathfrak{Z}} \eta(z,\hat{\theta}(F))\ dF(z) = 0$$

where F is some df in a convex set $\mathcal{F}'$ containing F_0, the assumed model $\mathcal{F}$

and all empirical df's over $\mathfrak{Z}$. From now on, $\hat{\theta}$ will denote a statistical functional, and $\hat{\theta}_N = \hat{\theta}(F_N)$ its value at the empirical df F_N.

In the case of statistical functionals it is more reasonable to adopt a definition of consistency that differs slightly from the usual definition of weak consistency [see e.g. Cox and Hinkley (1974)].

DEFINITION 1: Suppose that the model is correctly specified. Then $\hat{\theta}$ is called Fisher-consistent for θ_0 if $\hat{\theta}(F_0) = \theta_0$.

In the remainder of this Section we state rather informally a number of results that will be used repeatedly in what follows. For proofs and regularity conditions see e.g. Serfling (1980) and Fernholz (1983).

A.1: Under appropriate regularity conditions, $\hat{\theta}$ possesses the von Mises expansion

$$(3) \qquad \hat{\theta}(F_N) - \hat{\theta}(F_0) = \int_{\mathfrak{Z}} IF(z,\hat{\theta},F_0)\, dF_N(z) + o_p(N^{-1/2}).$$

for some function $IF(\cdot,\hat{\theta},F_0)$ that depends on $\hat{\theta}$ and F_0.

Thus, under appropriate regularity conditions, the possibly non-linear functional $\hat{\theta}$ can be approximated by a linear functional, with an approximation error that is asymptotically negligible.

The integrand in the linear term of (3) plays a key role in robust statistics, where it is called the influence function (IF) of $\hat{\theta}$ at F_0 [see e.g. Hampel (1974)]. Sometimes the linear term in (3) vanishes, in which case higher-order terms are included and the definition of the IF is based, more generally, on the first non-vanishing term.

A more direct definition of the IF is the following:

DEFINITION 2: Let $\Delta_{(z)}$ be a df with mass concentrated at the point z and let $F_{\varepsilon,z} = (1-\varepsilon)\ F + \varepsilon\ \Delta_{(z)}$. Then the IF of $\hat{\theta}$ at F is defined by

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

provided that the limit exists.

When $\hat{\theta}$ is Fisher consistent, its IF may be interpreted as a measure of the asymptotic bias arising from an arbitrarily small contamination of the df F by a point-mass at z. A natural quantitative measure of the robustness of a functional $\hat{\theta}$ under small deviations from the df F is given by the sup-norm of its IF at F in the metric of some pd matrix B. This is denoted by $\gamma^*(\hat{\theta},F,B) = \sup_{z\in\mathfrak{Z}} \|IF(z,\hat{\theta},F)\|_B$ and called the gross-error-sensitivity of $\hat{\theta}$ at F. When $\gamma^*(\hat{\theta},F,B)$ is finite, we say that $\hat{\theta}$ is B-robust or has a bounded IF.

When $\hat{\theta}$ is an M-estimator, its IF (if it exists) is a non-singular linear transformation of the score function that defines $\hat{\theta}$.

A.2: Assume regularity conditions sufficient to allow interchanging differentiation and integration, and to ensure that the matrix $P(\hat{\theta},F) = E_F\ [-(\partial/\partial\theta')\ \eta(z,\hat{\theta}(F))]$ is pd. Then the IF of an M-estimator $\hat{\theta}$, defined by an equation of the form (5), at a df F, is equal to

$$IF(z,\hat{\theta},F) = P(\hat{\theta},F)^{-1}\ \eta(z,\hat{\theta}(F)).$$

If $\hat{\theta}$ is the ML estimator based on a parametric model with likelihood score function $s(z,\theta)$ and a pd Fisher information matrix $J(\theta)$, then clearly $IF(z,\hat{\theta},F_\theta) = J(\theta)^{-1}\ s(z,\theta)$.

The next result gives a simple expression for the asymptotic bias of $\hat{\theta}$, as an estimator of $\hat{\theta}(F)$, when F_0 is in a small neighborhood of F.

A.3: Assume that $E_0\ IF(z,\hat{\theta},F)$ exists. If F_0 is near F in a suitable metric, the value of $\hat{\theta}$ at F_0 can be approximated by $\hat{\theta}(F) + E_0\ IF(z,\hat{\theta},F)$, with an approximation error that is of smaller order than the distance between F_0 and F.

The von Mises expansion (3) also provides a way of establishing the asymptotic distribution of a functional $\hat{\theta}$ which possesses an IF.

A.4: Assume that $\hat{\theta}$ satisfies (3) and $E_0\ IF(z,\hat{\theta},F_0)\ IF(z,\hat{\theta},F_0)'$ exists. Then

$$N^{1/2}\ [\hat{\theta}_N - \hat{\theta}(F_0)] \xrightarrow{d} \mathcal{N}(0,AV(\hat{\theta},F_0)),$$

where $AV(\hat{\theta},F_0) = E_0\ IF(z,\hat{\theta},F_0)\ IF(z,\hat{\theta},F_0)'$.

In particular, when $\hat{\theta}$ is an M-estimator we obtain that $AV(\hat{\theta},F_0) = P(\hat{\theta},F_0)^{-1}\ Q(\hat{\theta},F_0)\ P(\hat{\theta},F_0)'^{-1}$, where $Q(\hat{\theta},F_0) = E_0\ \eta(z,\hat{\theta}(F_0))\ \eta(z,\hat{\theta}(F_0))'$.

Now suppose that the unknown parameter θ_0 belongs to a proper subspace of Θ, defined by a set of $q \le p$ implicit constraints. Thus $\theta_0 \subset$

Θ, where $\omega = \{\theta \in \Theta: h(\theta) = 0\}$ and the (possibly non-linear) function h: $\Theta \longrightarrow \mathbb{R}^q$ satisfies certain regularity conditions. In particular, we assume that the q×p matrix $H_0 = H(\theta_0) = (\partial/\partial\theta')\ h(\theta_0)$ has rank q.

Often attention focuses on a subset of θ_0. Thus θ_0 may be partitioned as $(\beta_0{}',\gamma_0{}')'$, where $\beta_0 \in B \subset \mathbb{R}^k$ is the parameter of primary interest and $\gamma_0 \in \Gamma \subset \mathbb{R}^{p-k}$ is a nuisance parameter. Sometimes, only the parameter of primary interest is subject to restrictions. Thus an interesting special case is when $h(\theta_0) = h^*(\beta_0)$ with $q \le k$ [see e.g. Holly (1986)]. Results for this case are easily derived from our results, using the fact that $H_0 = [(\partial/\partial\beta')\ h^*(\beta_0)\,\vdots\,0]$ when $h(\theta) = h^*(\beta)$.

A restricted M-estimate $\hat{\theta}_{0N}$ of θ_0 is a solution to

$$\text{Max}_{\theta\in\Theta_0}\ N^{-1}\ \Sigma_{n=1}^{N}\ \rho_N(z_N,\theta) \quad \text{subject to} \quad h(\theta) = 0.$$

Under regularity conditions that guarantee the existence of an internal solution, $\hat{\theta}_{0N}$ satisfies the system of p+q equations

$$N^{-1}\ \Sigma_{n=1}^{N}\ \eta_N(z_n,\hat{\theta}_{0N}) - H(\hat{\theta}_{0N})'\ \hat{\lambda}_{0N} = 0$$

$$h(\hat{\theta}_{0N}) = 0,$$

where $\hat{\lambda}_{0N}$ is a q-vector of Lagrange multipliers associated with the constraints. When η_N has an almost sure limit η, functionals $\hat{\theta}_0$ and $\hat{\lambda}_0$ are implictly defined by the asymptotic first-order conditions

$$(4) \qquad \int \eta(z,\hat{\theta}_0(F))\ dF(z) - H(\hat{\theta}_0(F))'\ \hat{\lambda}_0(F)) = 0$$

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

for all probability distributions in the set $\mathcal{F}'$.

In what follows we assume that the unrestricted M-estimator $\hat{\theta}$ is Fisher-consistent, i.e. $\hat{\theta}(F_0) = \theta_0$. This implies that the restricted estimator $\hat{\theta}_0$ is also Fisher-consistent when the model is correctly specified. Generalizations to the case when the model is misspecified or $\hat{\theta}$ is not Fisher-consistent are straightforward. We also assume that $\hat{\theta}$ satisfies A.1-A.4, and so $\hat{\theta}$ is asymptotically normal with asymptotic variance matrix (AVM) equal to $P_0^{-1} Q_0 P_0'^{-1}$, where $P_0 = E_0 (\partial/\partial\theta') \eta(z,\theta_0)$ and $Q_0 = E_0 \eta(z,\theta_0) \eta(z,\theta_0)'$.

We first derive a relationship between the IF of $\hat{\theta}_0$ and $\hat{\lambda}_0$ and the IF of $\hat{\theta}$. This result enables us to establish the asymptotic distribution and the robustness properties of several statistics that may be used for testing purposes. All proofs are gathered in the Appendix.

PROPOSITION 1: If the model is correctly specified and $\hat{\theta}_0$ is Fisher-consistent and satisfies A.2, then

$$IF(z,\hat{\theta}_0,F_0) = [I_p - P_0^{-1} H_0' R_0 H_0] \; IF(z,\hat{\theta},F_0)$$

$$IF(z,\hat{\lambda}_0,F_0) = R_0 H_0 \; IF(z,\hat{\theta},F_0)$$

where $R_0 = (H_0 P_0^{-1} H_0')^{-1}$ is a q×q matrix of full rank q.

Proposition 1 asserts that the IF's of $\hat{\theta}_0$ and $\hat{\lambda}_0$ are linear transformations of the IF of the unrestricted estimator $\hat{\theta}$, and therefore of the score function $\eta(z,\theta)$. Notice that the linear transformation from $IF(z,\hat{\theta},F_0)$ to $IF(z,\hat{\theta}_0,F_0)$ always has less than full rank.

The IF's of $\hat{\theta}_0$ and $\hat{\lambda}_0$ have a nice projection interpretation. In the metric defined by the matrix P_0^{-1}, the IF of $\hat{\lambda}_0$ is equal to the vector of coefficients of the linear projection of the score $\eta(z,\theta_0)$ onto the q-dimensional subspace spanned by the rows of H_0. The IF of $\hat{\theta}_0$ is equal to the vector obtained by applying P_0^{-1} to the orthogonal complement of this projection, and is therefore orthogonal to H_0.

Proposition 1 enables us to derive the IF of other statistics that will be used in the next Section to construct tests of the hypothesis that $h(\theta_0) = 0$. Let

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

$$\hat{s}_{0N} = S \cdot N^{-1} \Sigma_{n=1}^{N} \eta(z_n,\hat{\theta}_{0N}) = S \cdot H(\hat{\theta}_{0N})' \hat{\lambda}_{0N}$$

$$\hat{\Delta}_N = S \cdot (\hat{\theta}_N - \hat{\theta}_{0N})$$

where $\hat{\theta}_N = \hat{\theta}(F_N)$ denotes the unrestricted estimate and S is an r×p matrix with rank $r \le p$ (S need not be the same for $\hat{s}_{0N}$ and $\hat{\Delta}_N$). Let $\hat{h}$, $\hat{s}_0$ and $\hat{\Delta}$ denote the corresponding functionals. The first statistic gives the value of the constraints at the unrestricted estimate $\hat{\theta}_N$ and forms the basis for Wald-type tests. The other two statistics form the basis for score- and Hausman-type tests respectively.

The following Corollary shows that the IF of $\hat{h}$, $\hat{s}_0$ and $\hat{\Delta}$ are also linear transformations of the IF of the unrestricted estimator $\hat{\theta}$.

COROLLARY: Under the assumptions of Proposition 1

$$IF(z,\hat{h},F_0) = H_0 \; IF(z,\hat{\theta},F_0)$$

$$IF(z,\hat{s}_0,F_0) = [S \; H_0' \; R_0 \; H_0] \; IF(z,\hat{\theta},F_0).$$

$$IF(z,\hat{\Delta},F_0) = [S \; P_0^{-1} \; H_0' \; R_0 \; H_0] \; IF(z,\hat{\theta},F_0).$$

Let $\hat{t}_N$ be any of the statistics $\hat{h}_N$, $\hat{\lambda}_{0N}$, $\hat{s}_{0N}$ and $\hat{\Delta}_N$. It is easily seen that $\hat{t}_N = t(\hat{\theta}_N)$, where the function t is continuously differentiable. Let $T_0 = (\partial/\partial\theta') \; t(\theta_0)$ be the gradient matrix of t evaluated at θ_0. In the case of $\hat{h}_N$ and $\hat{\lambda}_{0N}$, respectively $T_0 = H_0$ and $T_0 = R_0 \; H_0$, and the rank of T_0 is equal to q. In the case of $\hat{s}_{0N}$ and $\hat{\Delta}_N$, respectively $T_0 = S \; H_0' \; R_0 \; H_0$ and $T_0 = S \; P_0^{-1} \; H_0' \; R_0 \; H_0$, and the rank of T_0 is equal to min (r, q). Let $\hat{t}$ denote the functional mapping F into $t(\hat{\theta}(F))$. The IF of $\hat{t}$ at F_0 is of the form $IF(z,\hat{t},F_0) = T_0 \; IF(z,\hat{\theta},F_0)$, and so $\hat{t}$ has a bounded IF provided that $\hat{\theta}$ does. Further, if the model is correctly specified and $\hat{\theta}$ is Fisher-consistent and satisfies A.1-A.4 then

$$N^{1/2} \; [\hat{t}_N - t(\theta_0)] \xrightarrow{d} \mathcal{N} \; (0, \; T_0 \; P_0^{-1} \; Q_0 \; P_0'^{-1} \; T_0').$$

In particular, for the difference $\hat{\Delta} = S \; (\hat{\theta} - \hat{\theta}_0)$ one obtains the well known result that if $\hat{\theta}$ is the ML estimator or the efficient generalized method of moments estimator based on a given set of moment restrictions, then $AV(\hat{\Delta},F_0) = S \; [AV(\hat{\theta},F_0) - AV(\hat{\theta}_0,F_0)] \; S'$ [see e.g. Newey (1985a)]. Moreover, since rank $H_0' \; R_0 \; H_0 = q \le p$, $\hat{s}_0$ and $\hat{\Delta}$ have a singular AVM whenever $q < r$.

3. Asymptotic properties of M-tests under model contamination

This section introduces a number of test statistics for testing the null hypothesis that $h(\theta_0) = 0$ against the alternative that $h(\theta_0) \neq 0$, and examines the asymptotic level and power of the associated tests under sequence of contaminated local alternatives.

Consider the following test statistics :

$$\xi_N^W = N\, \hat{h}_N'\, AV(\hat{h},F_0)^{-1}\, \hat{h}_N$$

$$\xi_N^{LM} = N\, \hat{\lambda}_{0N}'\, AV(\hat{\lambda}_0,F_0)^{-1}\, \hat{\lambda}_{0N}$$

$$\xi_N^S = N\, \hat{s}_{0N}'\, AV(\hat{s}_0,F_0)^{-}\, \hat{s}_{0N}$$

$$\xi_N^H = N\, \hat{\Delta}_N'\, AV(\hat{\Delta},F_0)^{-}\, \hat{\Delta}_N,$$

where the use of a g-inverse, denoted by $-$, is necessary because the AVM of $\hat{s}_0$ and $\hat{\Delta}$ may be singular. The statistics ξ_N^W and ξ_N^{LM} are generalizations of the classical Wald and Lagrange multiplier test statistics. ξ_N^H is the test statistic proposed by Hausman (1978). It may be based on the whole parameter vector, in which case $S = I_p$, or on a subset of it. For example, when $h(\theta) = h^*(\beta)$ one may choose $S = [0 \vdots I_r]$ [see e.g. Holly (1986)]. If $P_{\beta\gamma}$ and $P_{\gamma\gamma}$ denote respectively the top-right and the bottom-right submatrices of P_0, when P_0 is partitioned according to $\theta = (\beta',\gamma')'$, and if $S = [I_k \vdots -P_{\beta\gamma}\, P_{\gamma\gamma}^{-1}]$, then the score test statistic ξ_N^S is equal to Neyman's $C(\alpha)$-test statistic [Neyman (1958)] for testing restrictions on the parameters of interest in the presence of nuisance parameters. Notice that the $C(\alpha)$-test statistic is asymptotically equivalent to the one based on $S = [I_k \vdots 0]$, i.e. on the subset of the score corresponding to the parameters of primary interest.

Since all statistics are based on M-estimates of θ, the resulting tests will be called M-tests. We define this class of tests more generally than in Newey (1985a), for we also include Wald tests. 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 regression model [see also Hampel et al. (1986)]. The asymptotic distribution of this kind of tests may be quite complicated, since it may involve mixtures of χ^2 variates [see e.g. Holly (1986)], in which case computation of tail probabilities is not easy. On the other hand, as we will see, all M-test statistics have a simple asymptotic χ^2 distribution, for which tail probabilities are readily available.

By the results of the previous Section, all M-test statistics are quadratic forms in asymptotically normal random variables $\hat{t}(F_N) = t(\hat{\theta}(F_N))$, where the statistical functional $\hat{t}$ has a bounded IF provided that the unrestricted M-estimator $\hat{\theta}$ does.

Let F_{0N} be the df of the observations under a sequence of local alternatives indexed by the parameter

(5) $$\theta_N = \theta_0 + N^{-1/2}\,\delta,$$

where δ does not lie in the null space of H_0. Consider sequences of contaminated local alternatives of the form

(6) $$F_{\varepsilon,N,z} = (1 - \varepsilon_N)\,F_{0N} + \varepsilon_N\,\Delta_{(z)}$$

where $\varepsilon_N = N^{-1/2}\,\varepsilon$, $\varepsilon \in [0,1]$, and $\Delta_{(z)}$ is the df with mass concentrated at the point $z \in \mathfrak{X}$. When $\delta = 0$ the contaminated distribution is in a shrinking neighborhood of the distribution specified by the null hypothesis. The device of letting $\varepsilon_N \longrightarrow 0$ and $\theta_N \longrightarrow \theta_0$ at the same rate as $N \longrightarrow \infty$ is a standard way of ensuring that the effects of contamination will not vanish nor dominate as the sample size increases [see e.g. Rieder (1978), Wang (1981) and Ronchetti (1982)].

First we establish the asymptotic distribution of all test statistics under the sequence of contaminated local alternatives (5)-(6).

PROPOSITION 2: Assume that the unrestricted estimator $\hat{\theta}$ satisfies A.1-A.4, and let ξ_N denote any of the M-test statistics ξ_N^W, ξ_N^{LM}, ξ_N^S or ξ_N^H. Then, under the sequence of contaminated local alternatives (5)-(6), the asymptotic distribution of ξ_N is a non-central χ^2 with number of degrees of freedom equal to the rank of T_0 and non-centrality parameter equal to

$$\nu(\delta,\varepsilon,z) = \varphi'\ T_0'\ [T_0\ P_0^{-1}\ Q_0\ P_0'^{-1}\ T_0']^-\ T_0\ \varphi$$

where $\varphi = \delta + \varepsilon\ IF(z,\hat{\theta},F_0)$. Moreover, the non-centrality parameter is invariant for any choice of the g-inverse.

The extension to more general models of contamination is straightforward. For example, when the sequence of contaminated local alternatives is given by $F_{\varepsilon,N,G} = (1 - \varepsilon_N)\ F_{0N} + \varepsilon_N\ G$ for some df G, the conclusions of Proposition 2 still hold with $\varphi = \delta + \varepsilon\ E_G\ IF(z,\hat{\theta},F_0)$.

Proposition 2 implies that under the sequence of contaminated local alternatives (5)-(6), the Wald and Lagrange multiplier test statistics ξ_N^W and ξ_N^{LM} have asymptotically the same non-central χ^2 distribution with q degrees of freedom and non-centrality parameter $\varphi' H_0' (H_0 P_0^{-1} Q_0 P_0'^{-1} H_0')^{-1} H_0 \varphi$. The score test statistic ξ_N^S has an asymptotic χ^2 distribution with min (r, q) degrees of freedom and non-centrality parameter $\varphi' H_0' R_0 H_0 S' (S H_0' R_0 H_0 P_0^{-1} Q_0 P_0'^{-1} H_0' R_0 H_0' S')^- S H_0' R_0 H_0 \varphi$, which is invariant for any choice of the g-inverse. Finally, the Hausman test statistic ξ_N^H has an asymptotic χ^2 distribution with min (r, q) degrees of freedom and non-centrality parameter $\varphi' H_0' R_0 H_0 P_0'^{-1} S' (S P_0^{-1} H_0' R_0 H_0 P_0^{-1} Q_0 P_0'^{-1} H_0' R_0 H_0 P_0'^{-1} S')^- S P_0^{-1} H_0' R_0 H_0 \varphi$, which is invariant for any choice of the g-inverse.

The Hausman and score test statistics are asymptotically equivalent when $S = I_p$, i.e. when the Hausman test is based on the whole parameter vector and the score test on the whole score vector. If, in addition, $q = p_0$ (i.e. the number of constraints is equal to the dimension of θ_0) or $P_0 = Q_0$ (i.e. $\hat{\theta}$ is the ML or the efficient GMM estimator), then all tests are asymptotically equivalent in the sense that they all have the same non-central χ^2 distribution, with q degrees of freedom and non-centrality parameter given by $\varphi' H_0' (H_0 P_0^{-1} H_0')^{-1} H_0 \varphi$ when $P_0 = Q_0$, by $\varphi' P_0' Q_0^{-1} P_0 \varphi$ when $q = p$, and by $\varphi' P_0 \varphi$ when $P_0 = Q_0$ and $q = p$.

When $\varepsilon = 0$ (no contamination), Proposition 2 contains as special cases a number of well known results [see e.g. Newey (1985a, 1985b), Gourieroux and Monfort (1985) for the case when $\rho_N(z,\theta) = \rho_N^*(z,\beta)$, and Holly (1986) for the case when $h(\theta) = h^*(\beta)$]. When $\delta = 0$ we obtain the asymptotic distribution of M-test statistics under contamination of the distribution specified by the null hypothesis.

Let τ be an M-test with nominal asymptotic level equal to α. The asymptotic level of τ under ε-contamination at z will be called the actual (asymptotic) level and will be denoted by $\alpha(\varepsilon,z)$. The asymptotic local power of τ against a specific alternative δ under ε-contamination at z will be called the actual (asymptotic) power and will be denoted by $\pi(\delta,\varepsilon,z)$. Clearly $\alpha = \alpha(0,z)$ and $\alpha(\varepsilon,z) = \pi(\delta,0,z)$. Let $\chi^2_{\mu,\nu}$ denote the non-central χ^2 df with μ degrees of freedom and non-centrality parameter ν. Then $\pi(\delta,\varepsilon,z) = 1 - \chi^2_{\mu,\nu(\delta,\varepsilon,z)}(\lambda)$, where the number of degrees of freedom μ and the non-centrality parameter $\nu(\delta,\varepsilon,z)$ are given in Proposition 2, and λ denotes the $(1 - \alpha)$-th quantile of the central χ^2 distribution with μ degrees of freedom. Recall that an α-level test is called unbiased when its power against any fixed alternative is greater than or equal to α [see e.g. Lehmann (1986)]. This property is desirable, for otherwise a test may reject the null hypothesis more frequently when it is true than when it is false.

We now explore a little further the dependence of the actual asymptotic level and power of the M-test based on a statistic $\hat{t}$ on the amount of contamination ε and the IF of $\hat{t}$.

PROPOSITION 3: Assume the regularity conditions of Proposition 2, and let τ be the M-test based on a statistic $\hat{t}$. Then, under the sequence of contaminated local alternatives (5)-(6),

(a) $\alpha(\varepsilon,z) > \alpha$, and $\sup_{z\in\mathfrak{X}} \alpha(\varepsilon,z) = 1$ if and only if $\hat{t}$ does not have a bounded IF.

(b) $\text{Inf}_{z\in\mathfrak{X}} \pi(\delta,\varepsilon,z) = \alpha$.

The first part of Proposition 3 asserts that the actual asymptotic level of the M-test based on a statistic $\hat{t}$ is strictly greater than the nominal level, and also that B-robustness of $\hat{t}$ ensures that the maximum asymptotic level under ε-contamination is bounded away from one. The second part asserts that there exist departures from the null hypothesis against which the test has essentially no power.

The next result establishes a relationship between the IF of $\hat{t}$ and the directions of departure from the null hypothesis against which the M-test based on $\hat{t}$ retains good power properties under ε-contamination.

PROPOSITION 4: Assume the regularity conditions of Proposition 2, and let τ be the M-test based on a statistic $\hat{t}$. Then, under the sequence of contaminated local alternatives (5)-(6),

(a) The actual asymptotic power of τ is smaller than the nominal power against all alternatives δ such that $\|T_0\delta\| \geq -\,[\varepsilon/(2\cos\rho)]\,\|IF(z,\hat{t},F_0)\|$, where ρ denotes the angle between the vectors $T_0\delta$ and $IF(z,\hat{t},F_0)$ in the metric of $AV(\hat{t},F_0)^-$;

(b) τ is asymptotically biased against all alternatives δ such that $\|T_0\delta\| \leq -\,2\,\varepsilon\cos\rho\,\|IF(z,\hat{t},F_0)\|$;

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

where the norm $\|\cdot\|$ is defined in the metric of $AV(\hat{t},F_0)^-$.

Thus, under ε-contamination at z, τ loses power against alternatives that are in the opposite direction to the asymptotic bias of $\hat{t}$. Moreover, τ is biased against alternatives that are in the opposite direction to the asymptotic bias of $\hat{t}$ and not too far from the null hypothesis. For departures from the null hypothesis in the same direction as the asymptotic bias of $\hat{t}$, the test is unbiased and its actual power is greater than the nominal one.

Proposition 4 is illustrated in Figure 1 for given values of ε and $IF(z,\hat{t},F_0)$. The region $\mathcal{L}$ is the set of alternatives against which τ loses power under ε-contamination at z, $\mathcal{B}$ is the set of alternatives against which τ is biased, and δ_0 is the alternative against which τ has no power. Increasing the asymptotic bias of $\hat{t}$ widens the set $\mathcal{B}$, and pushes $\mathcal{L}$ down somewhat. An increase in ε has the same effect.

As an example, consider testing the hypothesis that $\theta_0 = 0$ in the linear regression model $y = x'\theta_0 + r$, where $E_H\, xx' = I_p$, and the distribution of the disturbance r is assumed to be $N(0,1)$. We will compare tests based on two different estimators, namely the LS estimator $\hat{\theta}_{LS}$, with score function $\eta(z,\theta) = x\,(y - x'\theta)$, and the Huber estimator of regression $\hat{\theta}_H$ [see e.g. Huber (1981)], with score function $x\,\psi_c(y - x'\theta)$ where $\psi_c(r) = \min\,(c, \max(-c, r))$ and $c > 0$. The M-test statistics based on $\hat{\theta}_{LS}$ are all numerically the same, and their non-centrality parameter under the sequence of contaminated local alternatives (5)-(6) is equal to

$$(7) \qquad \nu(\delta,\varepsilon,z) = \delta'\delta + 2\,\varepsilon\, r\, \delta'x + \varepsilon^2\, r^2\, x'x.$$

The M-test statistics based on $\hat{\theta}_H$ are not the same numerically, but they all have the same asymptotic distribution with non-centrality parameter of the same form as (7), with r replaced by $\tilde{r} = \psi_c(r)\ /\ [2\,\Phi(c) - 1]$.

Let τ and $\tilde{\tau}$ denote, respectively, a test based on $\hat{\theta}_{LS}$ and a test based on $\hat{\theta}_H$. The different asymptotic behavior of τ and $\tilde{\tau}$ under

ε-contamination at (y, x) depends only on the difference between r and $\tilde{r}$. This behavior is summarized in Table 1 and illustrated in Figure 2.

Consider first the case when contamination occurs at a point (y_1, x) such that $r_1 = y_1 - x'\theta_0 = c$. Let $\mathcal{L}_1$ and $\mathcal{B}_1$ denote the set of alternatives against which τ, based on the LS estimator, respectively loses power and is biased. Now consider the case when contamination occurs at a point (y_2, x) such that $r_2 = y_2 - x'\theta = 2c$, i.e. r_2 is 'large' and positive. In this case, let $\mathcal{L}_2$ and $\mathcal{B}_2$ denote the set of alternatives against which τ respectively loses power and is biased. For $\tilde{\tau}$, based on the Huber estimator, $\tilde{\mathcal{L}}_1$, $\tilde{\mathcal{B}}_1$, $\tilde{\mathcal{L}}_2$ and $\tilde{\mathcal{B}}_2$ are similarly defined. In the case of r_1 there are few difference between τ and $\tilde{\tau}$, but $\tilde{\mathcal{L}}_1$ is slightly smaller than $\mathcal{L}_1$ and $\tilde{\mathcal{B}}_1$is slightly larger than $\mathcal{B}_1$. Moving the contamination to r_2 does not change the behavior of $\tilde{\tau}$, because $\tilde{r}_2 = \tilde{r}_1$. The behavior of τ, however, is altered dramatically. In particular, the set of alternatives against which τ is biased is now much broader than for $\tilde{\tau}$, because $r_2 = 2c$ is much bigger than $\tilde{r}_2 = c/[2\ \Phi(c) - 1]$. The different performance of the two tests reflects the fact that while the Huber estimator is robust under this particular form of contamination, the LS estimator is not.

As we have seen, 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 robustness of a test in terms of the effects of contamination on the asymptotic level and power. 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 3: Let τ be an M-test. Then, the IF of the asymptotic level and asymptotic local power of τ are defined by

$$IF_L(z,\tau,F_0) = (\partial^2/\partial\varepsilon^2)\ \pi(0,\varepsilon,z)|_{\varepsilon=0}.$$

$$IF_P(z,\tau,F_0,\delta) = (\partial/\partial\varepsilon)\ \pi(\delta,\varepsilon,z)|_{\varepsilon=0}.$$

The definition of IF_P is the same as the one given by Ronchetti (1982). The definition of IF_L is slightly different, for it involves second derivatives. This is necessary because, for M-tests, $(\partial/\partial\varepsilon)\ \pi(0,\varepsilon,z)|_{\varepsilon=0} = 0$. IF_L and IF_P can be used to approximate locally the actual level and power of an M-test τ. Thus, $\alpha(\varepsilon,z) = \alpha + (1/2)\ \varepsilon^2\ IF_L(z,\tau,F_0) + o(\varepsilon^2)$ and $\pi(\delta,\varepsilon,z) = \pi(\delta) + \varepsilon\ IF_P(z,\tau,F_0,\delta) + o(\varepsilon)$, where $\pi(\delta)$ denotes the nominal power of the test against the alternative δ. The gross-error sensitivity of the level and power of τ are defined respectively by $\gamma_L^*(\tau,F_0) = \sup_{z\in\mathfrak{X}} |IF_L(z,\tau,F_0)|$ and $\gamma_P^*(\tau,F_0) = \sup_{z\in\mathfrak{X},\delta\in\mathbb{R}^p} |IF_P(z,\tau,F_0,\delta)|$.

DEFINITION 4: An M-test τ is called robust at a df F_0 if $\gamma_L^*(\tau,F_0)$ and $\gamma_P^*(\tau,F_0)$ are both finite.

Thus, when a test is robust its asymptotic level and power change little under a small amount of contamination. The following result provides expressions for IF_L and IF_P in terms of the IF of the statistic on which the test is based.

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

$$IF_L(z,\tau,F_0) = \vartheta(z,F_0)\ IF(z,\hat{t},F_0)'\ AV(\hat{t},F_0)^{-}\ IF(z,\hat{t},F_0)$$

$$IF_P(z,\tau,F_0,\delta) = \varkappa(z,F_0,\delta)\ IF(z,\hat{t},F_0)'\ AV(\hat{t},F_0)^{-}\ IF(z,\hat{t},F_0)$$

where $\vartheta(z,F_0) = -(\partial/\partial\nu)\ \chi^2_{\mu,\nu(0,\varepsilon,z)}(\lambda)|_{\varepsilon=0}$, $\varkappa(z,F_0,\delta) = -(\partial/\partial\nu)\ \chi^2_{\mu,\nu(\delta,\varepsilon,z)}(\lambda)|_{\varepsilon=0}\cdot\cos\rho$, and ρ denotes the angle between $T\delta$ and $IF(z,\hat{t},F_0)$ in the metric of $AV(\hat{t},F_0)$.

Hampel et al. (1986) obtain a similar result for the case of one-sided tests of a single restriction. Since $\vartheta(\cdot,F_0)$ and $\varkappa(\cdot,F_0,\delta)$ are both bounded functions, Proposition 5 has two implications. First, for a small ε the divergence of the actual asymptotic level and power of τ from the nominal is proportional to the squared norm of the IF of $\hat{t}$ in the metric of $AV(\hat{t},F_0)^{-}$. Second, the gross-error sensitivity of the level and power of τ is proportional to the square of the self-standardized sensitivity of $\hat{t}$. Thus we have:

COROLLARY: The M-test τ based on a statistic $\hat{t}$ is robust at the df F_0 if and only if $\hat{\theta}$ has a bounded IF.

This Corollary represents the formal justification for using M-tests based on bounded-influence estimators.

4. Optimal robust M-tests

In general, we are interested in optimality as well as robustness of a statistical procedure. In the absence of a robustness constraint, tests based on the ML estimator $\hat{\theta}_{ML}$ are known to be asymptotically (locally) most powerful invariant if the assumed parametric model F_θ is correctly specified. However, these tests will not generally be robust, particularly if the assumed model is Gaussian. We therefore consider the following problem: Given a family of M-tests, such as the family of Wald tests, can we find a test $\tilde{\tau}$ that is 'optimally robust', i.e. has maximum asymptotic local power at the assumed model among all tests with bounded IF_L and IF_P?

Since M-tests inherit the efficiency or robustness properties of the estimators on which they are based, it seems reasonable to consider tests based on those bounded-influence estimators that have maximum asymptotic precision. Following Hampel et al. (1986), such estimators will be called optimal B-robust estimators. In the one-dimensional case, Ronchetti (1982) shows that optimal B-robust estimators do lead to optimal robust tests. In the more general case of multiple restrictions, the problem is essentially one of finding an unrestricted M-estimator $\tilde{\theta}$, such that (a) the non-centrality parameter $\delta' \; T_0' \; [T_0 \; AV(\tilde{\theta},F_\theta) \; T_0']^- \; T_0 \; \delta$ attains a maximum for all directions δ, and (b) the robustness constraint $\sup_{z\in\mathfrak{Z}} IF(z,\tilde{\theta},F_\theta)' \; T_0' \; [T_0 \; AV(\tilde{\theta},F_\theta) \; T_0']^- \; T_0 \; IF(z,\tilde{\theta},F_\theta) \le c^2$ is satisfied. Maximizing the non-centrality parameter with respect to $\tilde{\theta}$ for all directions δ is equivalent to minimizing the trace of $Q \; AV(\tilde{\theta},F_\theta)$ for all psd matrices Q. By A.4, this is equivalent to minimizing $E_0 \; IF(z,\tilde{\theta},F_\theta) \; Q \; IF(z,\tilde{\theta},F_\theta)$ for all psd matrices Q. Therefore, when the robustness constraint is imposed, the existence of an optimal test depends on the existence of an M-estimator $\tilde{\theta}$ based on a score function that is the solution to the following minimization problem

$$\text{Min}_{\eta(\cdot,\theta)\in\tilde{H}} E_\theta\, \eta(z,\theta)'\, Q\, \eta(z,\theta)$$

(8) $$\text{s.t.} \quad E_\theta\, \eta(z,\theta) = 0$$

(9) $$E_\theta\, \eta(z,\theta)\, s(z,\theta)' = I_p$$

(10) $$\sup_{z\in\mathfrak{Z}} \|\eta(z,\theta)\|_B \leq c$$

for all psd matrices Q and all θ in some set Θ_0 that contains θ_0. The set $\tilde{H}$ is the set of all functions that may be score functions for some regular M-estimator. Constraint (8) requires the M-estimator to be Fisher-consistent. Constraint (9), where $s(z,\theta)$ is the likelihood score of the assumed parametric model, is a normalization condition under which the score and the IF of $\tilde{\theta}$ are identical. Constraint (10), with $B = T_0' [T_0\, AV(\tilde{\theta},F_\theta)\, T_0']^- T_0$, is the robustness condition.

This problem is the same as the minimum norm problem in Peracchi (1987), except that the solution must hold for all psd matrices Q. It is easily seen that when $c = \infty$ a solution always exists and is equal to the likelihood score for the assumed model. When c is finite and sufficiently large a solution exists in the one-dimensional case, but not in the multi-dimensional case. In the latter case, the minimization problem generates instead a whole family of 'weakly optimal' solutions, one for each pair of matrices Q and B. This corresponds to the family of 'optimal B-robust' estimators for a given parametric model discussed in Peracchi

(1987). The form of the score function for these estimators involves shrinking and recentering the likelihood score and is fairly complicated, but can be simplified considerably when Q is a scalar multiple of B [see also Hampel et al. (1986)]. In any case, the optimal estimator depends on the choice of the matrix Q. Therefore, the resulting M-tests are not invariant because they are not equally powerful in all directions.

As an example, consider the case of a Wald test of the linear restriction $\theta_0 = 0$. In this case $B = AV(\tilde{\theta},F_\theta)^{-1}$, and a Wald test based on a Krasker-Welsch type estimator [Krasker and Welsch (1982)] is optimally robust and has highest power against the alternatives implicitly defined by the matrix Q = B.

Let $\mathcal{T}_{\alpha,c}$ be a class of tests (e.g. Wald tests) with the same nominal asymptotic level α and robustness properties, and let $\tilde{\tau}$ denote the test based on a given optimal B-robust estimator $\tilde{\theta}$. When the assumed model is correct, $\tilde{\tau}$ has maximal asymptotic power in $\mathcal{T}_{\alpha,c}$ against some alternatives but not against all. Therefore, there may exist other tests in $\mathcal{T}_{\alpha,c}$ with greater power against specific alternatives. However, as we are going to show, no test in $\mathcal{T}_{\alpha,c}$ dominates $\tilde{\tau}$ in the sense of being at least as powerful as $\tilde{\tau}$ against all alternatives, and more powerful than $\tilde{\tau}$ against some alternatives. This result generalizes the asymptotic 'admissibility' result obtained by Hampel et al. (1986) [Proposition 5, Section 7.3] for Wald tests of linear restrictions in the linear model.

PROPOSITION 6: Let $\mathcal{T}_{\alpha,c}$ be a family of asymptotic α-level robust M-tests, let $\tilde{\tau}$ be the member based on the optimal B-robust estimator $\tilde{\theta}$, and let $\tilde{\pi}(\delta)$ be its asymptotic (local) power function at the assumed model. Then there is no other test τ in $\mathcal{T}_{\alpha,c}$, with asymptotic power function $\pi(\delta)$ at the assumed model, such that $\pi(\delta) \geq \tilde{\pi}(\delta)$ for all $\delta \in \Theta$, and $\pi(\delta) > \tilde{\pi}(\delta)$ for some δ.

The proof does not depend on the metric Q for which $\tilde{\theta}$ is optimally B-robust, and so this Proposition provides a formal justification for using M-tests based on some optimal B-robust estimator. Since optimal robust M- tests are not invariant, choosing a particular estimator implicitly corresponds to choosing a set of alternatives against which the test has optimal power properties.

Appendix

PROOF OF PROPOSITION 1: Let $F_{\varepsilon,z} = (1 - \varepsilon) F_0 + \varepsilon \Delta_{(z)}$, where $\Delta_{(z)}$ is the df with mass concentrated at the point z. Replacing F_N by $F_{\varepsilon,z}$ in (4) -(5), differentiating with respect to ε and evaluating at $\varepsilon = 0$ gives

$$(*) \qquad 0 = - P_0 \, IF(z,\hat{\theta}_0,F_0) + \eta(z,\hat{\theta}_0) - H_0' \, IF(z,\hat{\lambda}_0,F_0)$$

$$0 = H_0 \, IF(z,\hat{\theta},F_0),$$

where $P_0 = - E_0 \, (\partial/\partial\theta') \, \eta(z,\theta_0)$, $H_0 = (\partial/\partial\theta') \, h(\theta_0)$, and we use the fact that $\hat{\theta}_0(F_0) = \theta_0$ and $\hat{\lambda}_0(F_0) = 0$. Premultiplying (*) by P_0^{-1} and rearranging gives the equation system

$$\begin{bmatrix} I_k & P_0^{-1} H' \\ H & 0 \end{bmatrix} \begin{bmatrix} IF(z,\hat{\theta}_0,F_0) \\ IF(z,\hat{\lambda}_0,F_0) \end{bmatrix} = \begin{bmatrix} IF(z,\hat{\theta},F_0) \\ 0 \end{bmatrix}.$$

The desired result then follows from the formulae for the inverse of a partitioned matrix. ■

PROOF OF PROPOSITION 2: Let observations be distributed according to $F_{\varepsilon,N,z} = (1-\varepsilon_N) F_{0N} + \varepsilon_N \Delta_{(z)}$. From A.1 and A.3, for N sufficiently large

$$\hat{\theta}(F_{\varepsilon,N,z}) - \theta_0 \simeq \int IF(x,\hat{\theta},F_0) \, dF_{\varepsilon,N,z}(x)$$

$$\simeq (1-\varepsilon_N) \int IF(x,\hat{\theta},F_0) dF_{0N} + \varepsilon_N \, IF(z,\hat{\theta},F_0)$$

$$\simeq (1-\varepsilon_N)\, N^{-1/2}\, \delta + \varepsilon_N\, IF(z,\hat{\theta},F_0).$$

Thus $\text{plim}_{N\to\infty} N^{1/2} [\hat{\theta}(F_{\varepsilon,N,z}) - \theta_0] = \delta + \varepsilon\, IF(z,\hat{\theta},F_0) \equiv \varphi$. Therefore $N^{1/2} [\hat{\theta}(F_N) - \theta_0] \xrightarrow{d} N(\varphi,\, P_0^{-1} Q_0 P_0'^{-1})$ by A.2 and A.4, and so $N^{1/2}\, \hat{t}_N \simeq N^{1/2}\, T_0' [\hat{\theta}(F_N) - \theta_0] \xrightarrow{d} N(T_0' \varphi,\, T_0' P_0^{-1} Q_0 P_0'^{-1} T_0)$.

From the definition of non-central χ^2 distribution, $N\, \hat{t}_N'\ AV(\hat{t},F_0)^- \hat{t}_N \xrightarrow{d} \chi^2_{\mu,\nu}$, where μ = rank $AV(\hat{t},F_0)$ = rank T_0 since rank $T_0 \le p$, and $\nu = \varphi'\, T_0\, [T_0'\, AV(\hat{\theta},F_0)\, T_0']^-\, T_0'\, \varphi$.

Invariance of the non-centrality parameter to the choice of g-inverse is easily proved by using the argument in Holly (1986). ■

PROOF OF PROPOSITION 3: The asymptotic local power function of τ is given by $\pi(\delta,\varepsilon,z) = 1 - \chi^2_{\mu,\nu(\delta,\varepsilon,z)}(\lambda)$ and is a strictly increasing function of the non-centrality parameter $\nu(\delta,\varepsilon,z)$. From Proposition 2, $\nu(\delta,\varepsilon,z)$ is equal to

$$[T_0\, \delta + \varepsilon\, IF(z,\hat{t},F_0)]'\, AV(\hat{t},F_0)^-\, [T_0\, \delta + \varepsilon\, IF(z,\hat{t},F_0)].$$

The first part of (i) then follows from the fact that $\alpha(\varepsilon,z) = \pi(0,\varepsilon,z)$, and the fact that $\nu(0,\varepsilon,z) = \varepsilon^2\, IF(z,\hat{t},F_0)'\, AV(\hat{t},F_0)^-\, IF(z,\hat{t},F_0)$ is strictly positive.

The second part of (i) follows from the fact that $\sup_{z\in\mathfrak{Z}} \nu(0,\varepsilon,z) = [\varepsilon\, \gamma^*(t,F_0,AV(\hat{t},F_0)^-)]^2$ and $\gamma^*(t,F_0,AV(\hat{t},F_0)^-)$ is unbounded if and only if $\hat{t}$ is not B-robust.

(ii) follows from the fact that $\nu(\delta,\varepsilon,z) = 0$ if $\delta = -\varepsilon\, IF(z,\hat{\theta},F_0)$.■

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

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

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

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

$$\Leftrightarrow \quad \|T_0\delta\| \le -\,[\varepsilon\,/(2\cos\rho)]\,\|IF(z,\hat{t},F_0)\|,$$

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

(ii) τ is asymptotically biased whenever

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

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

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

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

(iii) τ has no power whenever

$$\varphi' T_0' \, AV(\hat{t},F_0)^- \, T_0\varphi = 0 \quad \Leftrightarrow \quad T_0\delta = -\,\varepsilon\, IF(z,\hat{t},F_0). \qquad \blacksquare$$

PROOF OF PROPOSITION 5: Immediate by applying Definition 3 to $\pi(\delta,\varepsilon,z) = 1 - \chi^2_{\mu,\nu(\delta,\varepsilon,z)}(\lambda)$ and $\alpha(\varepsilon,z) = 1 - \chi^2_{\mu,\nu(0,\varepsilon,z)}(\lambda)$, with $\nu(\delta,\varepsilon,z)$ given by Proposition 2. ∎

PROOF OF PROPOSITION 6: The proof is a simple consequence of the following:

Lemma: Let A and B be (p×p) matrices, and suppose that

$$\text{trace } A\, Q < \text{trace } B\, Q$$

for a pd matrix Q. Then A - B is not a psd matrix.

Proof: Suppose on the contrary that $x'(A - B)\,x = \text{trace }(A - B)\,xx' \geq 0$ for all x. Since Q is pd, there exist a set of p linearly independent vectors $(x_1, \ldots, x_p)$ such that $Q = \Sigma_{i=1}^{p} x_i x_i'$. Therefore trace (A - B) Q = trace $(A - B)\,\Sigma_{i=1}^{p} x_i x_i' \geq 0$, a contradiction. ∎

References

Burguete, J., Gallant, A.R., and Souza, G. (1982), "On unification of asymptotic theory of non-linear econometrics", *Econometric Reviews*, 1, 151-190.

Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics*, Chapman and Hall, London.

Fernholz, L.T. (1983), *Von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics 19, Springer, New York.

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

Gourieroux, C. and Monfort, A. (1985), "A general theory for asymptotic tests", CEPREMAP, Paris, mimeo.

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.

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

Holly, A. (1986), "Specification tests: An overview", Cahier No. 8603, Department d'Econometrie et d'Economie Politique, Universite de Lausanne.

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.

Lehmann, E.L. (1986), *Testing Statistical Hypothesis*, 2nd ed., Wiley, New York.

Newey, W.K. (1985a), "Generalized method of moments specification testing", *Journal of Econometrics*, 29, 229-256.

Newey, W.K. (1985b), "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. (1987), Bounded Influence Methods in Econometrics with an Application to the Censored Regression Model, Ph.D. Thesis, Princeton University.

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, Ph.D. Thesis, ETH, Zurich.

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

Sen, P.K. (1982), "On M-tests in linear models", *Biometrika*, 69, 245-248.

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

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

Table 1

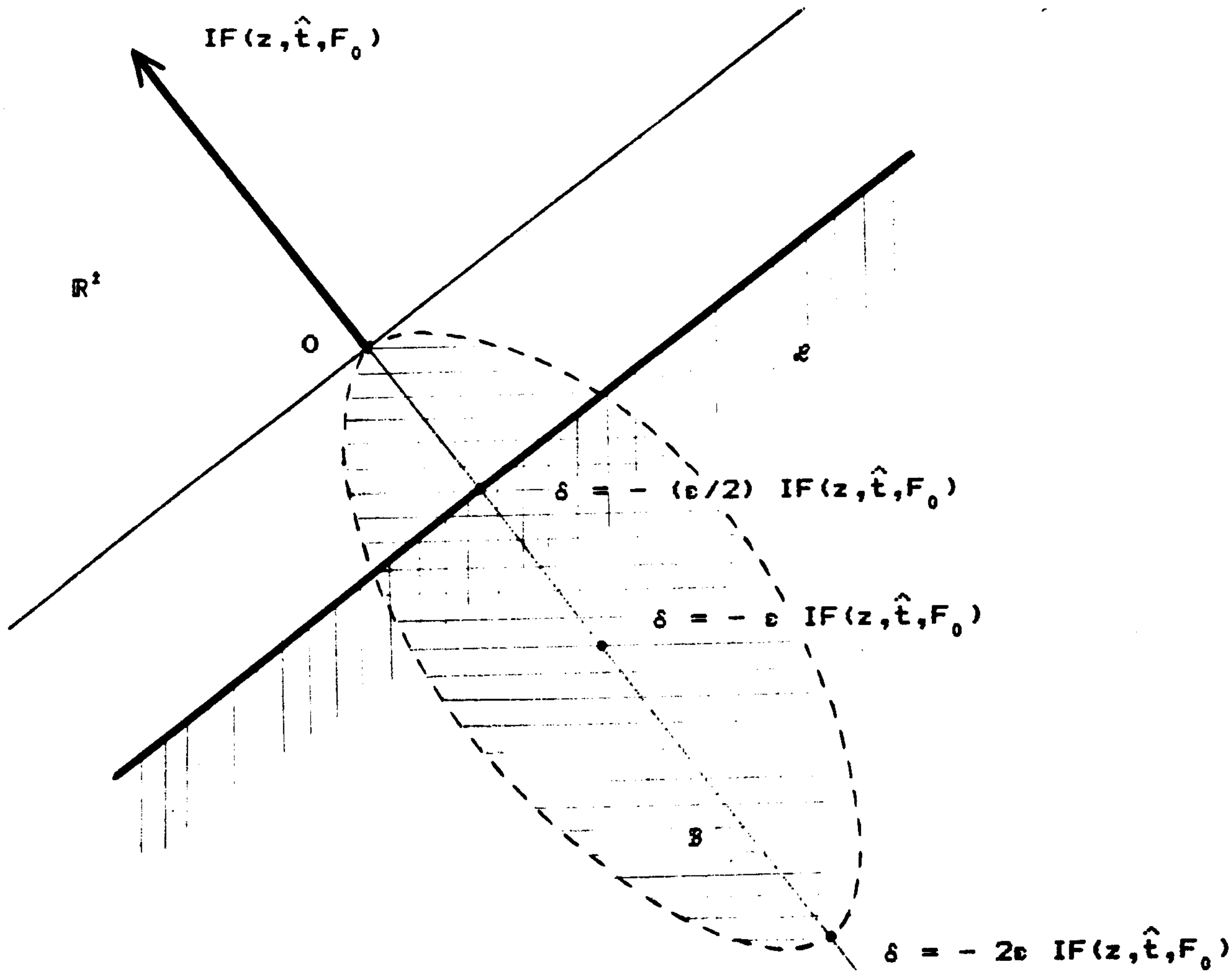
Asymptotic behavior of M-tests based on the LS and Huber estimators of regression under sequences of ε-contaminated local alternatives, with contamination at (y, x).

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

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

Figure 1

Asymptotic behavior of M-tests under sequences of ε-contaminated local alternatives.

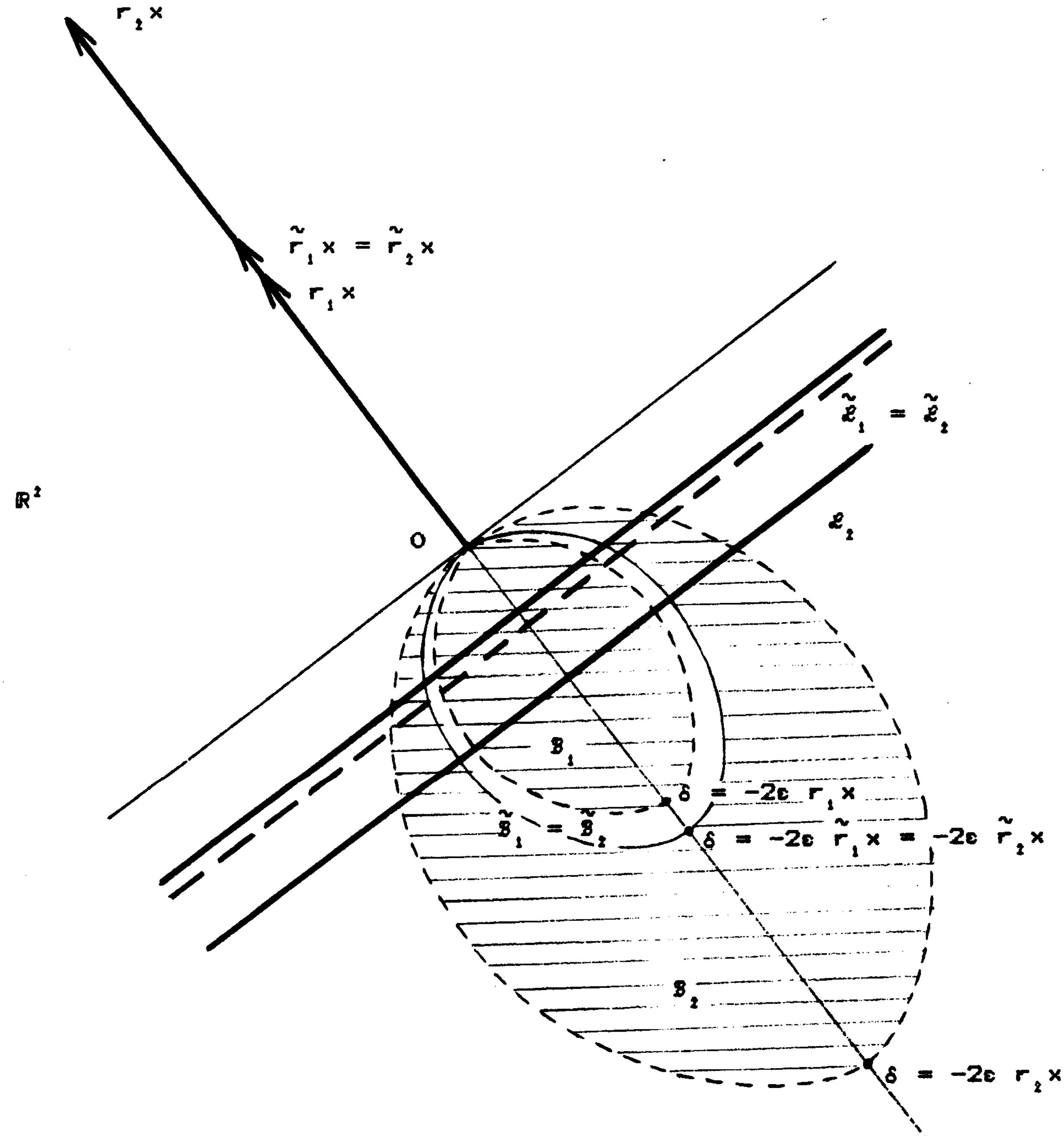


$\mathcal{B} = \{\delta : \|\delta + \varepsilon\, IF(z,\hat{t},F_0)\| \le \varepsilon\, \|IF(z,\hat{t},F_0)\|\}$

$\mathcal{E} = \{\delta : \|\delta + \varepsilon\, IF(z,\hat{t},F_0)\| \le \|\delta\|\}$

Figure 2

Asymptotic behavior of M-tests based on the LS and Huber regression estimators under sequences of ε-contaminated local alternatives.



Note: $r_1 = c$, $r_2 = 2c$, $\tilde{r} = \psi_c(r)\ [2\Phi(c) - 1]$.