Testing the Information Matrix Equality with Robust Estimators

by

Christophe CROUX
Geert DHAENE
Dirk HOORELBEKE

Econometrics

DISCUSSION PAPER

Testing the Information Matrix Equality
with Robust Estimators

Christophe Croux   Geert Dhaene   Dirk Hoorelbeke *
K.U.Leuven   K.U.Leuven   K.U.Leuven

March 18, 2003

Abstract

We study the behaviour of the information matrix (IM) test when
maximum likelihood estimators are replaced with robust estimators.
The latter may unmask outliers and hence improve the power of the
test. We investigate in detail the local asymptotic power of the IM
test in the normal model, for various estimators and under a range
of local alternatives. These local alternatives include contamination
neighbourhoods, Student's t (with degrees of freedom approaching in-
finity), skewness, and a tilted normal. Simulation studies for fixed
alternatives confirm that in many cases the use of robust estimators
substantially increases the power of the IM test.

JEL classification: C12, C15
Key-words: information matrix test, robustness

---

*Corresponding author. Address: K.U.Leuven, Department of Economics, Naamses-
traat 69, 3000 Leuven, Belgium. Tel. +32 16 326652. Fax +32 16 326796. Email:
dirk.hoorelbeke@econ.kuleuven.ac.be

1   Introduction

White (1982) introduced the information matrix (IM) test as an omnibus test for misspecification of a parametric model. The test exploits the well-known property that, at the model, the sum of the Hessian of the log-likelihood and the outer product of the score has zero expectation. So if, at parameter estimates, the sample average of the sum of the Hessian and the outer product of the score differs significantly from zero, this is evidence against the model. The IM test is typically implemented using maximum likelihood (ML) estimates of the parameters. In this paper we explore the potential of replacing the ML estimator with robust estimators. Specific attention is given to the effect on power, conjecturing that unmasking outliers will lead to an increased power of the IM test. In most cases considered, using robust estimators effectively increases the power of the test.

Past research on the IM test has mainly focussed on the poor behaviour of the test under the null hypothesis, and on remedies to overcome this problem (see e.g. Orme (1990) and Chesher and Spady (1991), among many others). Considering that the use of bootstrap critical values largely solves this problem (Horowitz, 1994), we shift our attention to the power of the IM test.

The standard approach in the literature on the IM test is to substitute the ML estimator for the unknown parameter in the IM equality. As an alternative, one can use any estimator which is consistent under the model specification. When the IM test is used in conjunction with the ML estimator, the test may suffer from the masking effect. Outlying observations will not be recognised as such (or less so, compared to robust estimators), and hence the test will have low power against distributions with fatter tails or when outliers are present. We show that, when robust estimators are used, the IM test statistic still has a limiting χ^2 distribution under the null hypothesis. An explicit expression for the asymptotic covariance matrix (V) of the indicator vector, to be used in the construction of the test statistic, is derived.

We analyse the normal location-scale model in detail. It is shown that V does not depend on the estimator of location. For the ML estimator, as is well-known, the IM test is the Jarque-Bera (1980) test for skewness and non-normal kurtosis, and the rank of V equals 2. We show that, for any

1

other M-estimator of scale, the rank of V equals 3, indicating that in this case the IM test is sensitive to 'more' specification error. Explicit expressions for V are derived for two robust M-estimators of scale, namely the M-estimator based on Tukey's biweight (TB) function and the Median Absolute Deviation (MAD). Next, we examine the performance of the IM test under four different sequences of local alternatives: a contaminated normal, Student's t, a skewed normal, and a tilted normal. The asymptotic distribution of the IM statistics under a sequence of local alternatives is non-central χ^2. For the contaminated normal alternative the non-centrality parameter is shown to be minimal, over all M-estimators, at the ML estimator. It turns out that, when outlying observations are present, the IM test using the 25% breakdown point TB estimator (which is a compromise between high robustness and high efficiency) is much more powerful than the IM test using the ML estimator. On the other hand, when the local alternative is Student's t (with degrees of freedom approaching infinity), skewed normal, or tilted normal, the non-centrality parameter is identical for all M-estimators. We derive closed-form expressions of the non-centrality parameter, under any local alternative considered. We also compare, whenever relevant, the local power of the IM tests with the local power of the score test. The latter test is known to be optimal and so it provides a natural benchmark.

In the normal regression model the IM test is a combined test for heteroskedasticity, skewness and non-normal kurtosis (Hall, 1987). We use an S-estimator (Rousseeuw and Yohai, 1984) or an MM-estimator (Yohai, 1987) as robust estimators of regression, and an M-estimator based on Tukey's biweight function as a robust estimator of residual scale. It is well-known that the ML estimator tends to mask outlying observations, and this danger is more severe in the regression model than in the location-scale model. It is therefore expected that the use of robust estimators holds more promise in the regression case. Simulation results indeed indicate that using robust estimators increases the power of the IM test in the case of a thick-tailed alternative like the Cauchy distribution, or in the presence of outliers.

Note that in the presence of outliers, the misspecification test will reject the IM equality. The aim of testing the IM equality with robust estimators is to increase the power of the statistical test, and not to attain level-robustness (Hampel et al., 1986). So when the null hypothesis holds, aside from some outliers, we would like the test to detect this deviation from the specified

model.

In Section 2 we present the IM test with robust estimators and the general form of the IM statistic. The IM test in the normal model, without and with covariates, is considered in Sections 3 and 5, respectively. In Section 4 we study the local asymptotic power of the IM test in the normal model without covariates. Monte Carlo results are presented in Section 6. Section 7 concludes. Technical derivations are gathered in the Appendix.

2 The IM test with robust estimators

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n independent copies of the random variables (X, Y), where X has distribution K (which is left unspecified) and the distribution of Y, given X, is specified by the parametric model $\{F_\theta \mid \theta \in \Theta\}$. Let f_θ be the conditional density corresponding to F_θ, and let $s_\theta = -\frac{\partial}{\partial \theta} \log f_\theta$. The conditional IM equality can be stated as

$$E_{F_\theta}[m(X, Y; \theta)] = 0 \qquad \text{for almost all } X,$$

where E_{F_θ} is the conditional expectation with respect to F_θ and m is the vectorised lower triangular part of

$$s_\theta s_\theta' - \frac{\partial}{\partial \theta'} s_\theta.$$

Integrating with respect to K yields the IM equality

$$E[m(X, Y; \theta)] = 0,$$

where $E[\cdot] = E_K E_{F_\theta}[\cdot]$. Now let $\hat{\theta}$ be an estimator of θ, sufficiently regular to have

$$\hat{\theta} - \theta = \frac{1}{n} \sum_{i=1}^{n} \text{IF}(X_i, Y_i; \hat{\theta}; K, F_\theta) + o_p(n^{-1/2}), \tag{1}$$

where IF is the influence function of the estimator $\hat{\theta}$, and, for

$$\hat{M} = \frac{1}{n} \sum_{i=1}^{n} m(X_i, Y_i; \hat{\theta}),$$

3

to admit the expansion

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}\left[m(X_i, Y_i; \theta) + \left(\frac{\partial}{\partial\theta}m(X_i, Y_i; \theta)\right)'(\hat{\theta} - \theta)\right] + o_p(n^{-1/2}). \quad (2)$$

The estimator $\hat{\theta}$ is B-robust (Hampel et al., 1986) when $\text{IF}(\cdot, \cdot; \hat{\theta}; K; F_\theta)$ is bounded. Assuming the existence of

$$D(\theta) = E[\frac{\partial}{\partial\theta'}m(X, Y; \theta)],$$

we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}m(X_i, Y_i; \theta) \overset{\text{p}}{\to} D(\theta). \quad (3)$$

Now, let

$$\xi(X, Y; \theta) = m(X, Y; \theta) + D(\theta)\text{IF}(X, Y; \hat{\theta}; K, F_\theta). \quad (4)$$

Then, combining (1)-(4),

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}\xi(X_i, Y_i; \theta) + o_p(n^{-1/2}). \quad (5)$$

So we obtain

$$\sqrt{n}\hat{M} \overset{\text{d}}{\to} \text{N}(0, V),$$

with

$$V = E[\xi(X, Y; \theta)\xi(X, Y; \theta)'].$$

Let $\hat{V}^+$ be a consistent estimator of V^+, the Moore-Penrose inverse of V, and define the test statistic

$$T = n\hat{M}'\hat{V}^+\hat{M}.$$

Then, if the parametric model is correctly specified,

$$T \overset{\text{d}}{\to} \chi_q^2,$$

where q is the rank of V. Remark that T can also be written as

$$T = n \left(W\hat{M} \right)' \left(W\hat{V}W' \right)^+ \left(W\hat{M} \right),$$

where W is any non-singular $q \times q$ matrix. Choosing W so as to make WVW' diagonal simplifies T, and this will be exploited below.

Note that if, say, the j-th column of $D(\theta)$ is zero, then V is independent of the estimator $\hat{\theta}_j$ (the j-th element of $\hat{\theta}$) that is used. When V has full rank, an obvious estimator is

$$\hat{V}^+ = \left[\frac{1}{n} \sum_{i=1}^{n} \xi(X_i, Y_i; \hat{\theta}) \xi(X_i, Y_i; \hat{\theta})' \right]^{-1}.$$

For the ML estimator we have

$$\text{IF}(X, Y; \hat{\theta}_{ML}; K, F_\theta) = - \left(E \left[\frac{\partial}{\partial \theta'} s_\theta(X, Y) \right] \right)^{-1} s_\theta(X, Y).$$

It may occur that (some elements of) $D(\theta)$ need to be estimated by empirical counterparts.

3   The normal model

3.1   The IM test

For the normal model without covariates, $F_\theta(y) = \frac{1}{\sigma}\Phi(\frac{y-\beta}{\sigma})$ with $\theta = (\beta, \sigma)'$ and Φ the standard normal cdf. Letting $u = (Y - \beta)/\sigma$, we have (White, 1994, p. 332-333)

$$m(Y; \theta) = \frac{1}{\sigma^2} \begin{pmatrix} u^2 - 1 \\ u^3 - 3u \\ u^4 - 5u^2 + 2 \end{pmatrix},$$

$$\frac{\partial}{\partial \theta'} m(Y; \theta) = -\frac{1}{\sigma^3} \begin{pmatrix} 2u & 4u^2 - 2 \\ 3u^2 - 3 & 5u^3 - 9u \\ 4u^3 - 10u & 6u^4 - 20u^2 + 4 \end{pmatrix},$$

$$D(\theta) = -\frac{1}{\sigma^3} \begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 2 \end{pmatrix},$$

wherefrom, using (4),

$$\xi(Y;\theta) = \frac{1}{\sigma^2} \begin{pmatrix} u^2 - 1 \\ u^3 - 3u \\ u^4 - 5u^2 + 2 \end{pmatrix} - \frac{2}{\sigma^3} \begin{pmatrix} \text{IF}(Y;\hat{\sigma};F_\theta) \\ 0 \\ \text{IF}(Y;\hat{\sigma};F_\theta) \end{pmatrix}.$$

Note that $\xi(Y;\theta)$ does not depend on $\text{IF}(Y;\hat{\beta};F_\theta)$. Take $\hat{\sigma}$ to be equivariant, i.e. $\sigma(aY_1+b, \cdots, aY_n+b) = |a|\sigma(Y_1, \cdots, Y_n)$, so $\text{IF}(Y;\hat{\sigma};F_\theta) = \sigma\text{IF}(u;\hat{\sigma};\Phi)$ and

$$\xi(Y;\theta) = \frac{1}{\sigma^2} \begin{pmatrix} u^2 - 1 - 2\text{IF}(u;\hat{\sigma};\Phi) \\ u^3 - 3u \\ u^4 - 5u^2 + 2 - 2\text{IF}(u;\hat{\sigma};\Phi) \end{pmatrix}. \tag{6}$$

A straightforward calculation shows that $V = \sigma^{-4}B$, where B is a 3×3 matrix with elements B_{ij} given by

$$
\begin{aligned}
B_{11} &= -2 + 4\text{ASV}(\hat{\sigma}), \\
B_{22} &= 6, \\
B_{33} &= 46 + 4\text{ASV}(\hat{\sigma}) - 4E_\Phi(u^4\text{IF}), \\
B_{13} &= 10 + 4\text{ASV}(\hat{\sigma}) - 2E_\Phi(u^4\text{IF}) = B_{31}, \\
B_{12} &= B_{21} = B_{23} = B_{32} = 0,
\end{aligned}
$$

with $\text{IF} = \text{IF}(u;\hat{\sigma};\Phi)$ and $\text{ASV}(\hat{\sigma}) = E_\Phi(\text{IF}^2)$, the asymptotic variance of $\hat{\sigma}$ when $\sigma = 1$. Note that V does not depend on the estimator $\hat{\beta}$ that is chosen. For a given estimator $\hat{\theta} = (\hat{\beta}, \hat{\sigma})'$, let

$$\hat{u}_i = (Y_i - \hat{\beta})/\hat{\sigma}, \qquad \hat{\mu}_j = n^{-1}\sum_{i=1}^n \hat{u}_i^j,$$

and write $\hat{M}$ as $\hat{\sigma}^{-2}\hat{N}$ with

$$\hat{N} = \begin{pmatrix} \hat{\mu}_2 - 1 \\ \hat{\mu}_3 - 3\hat{\mu}_1 \\ \hat{\mu}_4 - 5\hat{\mu}_2 + 2 \end{pmatrix}.$$

Taking $\hat{V}^+$ equal to V^+ with σ replaced by $\hat{\sigma}$ yields the test statistic

$$T = n\hat{M}'\hat{V}^+\hat{M} = n\hat{N}'B^+\hat{N}. \tag{7}$$

6

To simplify the notations and calculations that follow we transform $\hat{M}$ in order to diagonalise V. Let

$$W = \begin{pmatrix} 1+d & 0 & -d \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \tag{8}$$

where

$$d = \frac{E_\Phi\left[(u^4 - 6u^2 + 3)(u^2 - 1 - 2\text{IF})\right]}{E_\Phi\left((u^4 - 6u^2 + 3)^2\right)} = \frac{1}{2} - \frac{1}{12} E_\Phi\left(u^4 \text{IF}\right). \tag{9}$$

Then

$$W\hat{M} = \frac{1}{\hat{\sigma}^2} \begin{pmatrix} \hat{\mu}_2 - 1 - d(\hat{\mu}_4 - 6\hat{\mu}_2 + 3) \\ \hat{\mu}_3 - 3\hat{\mu}_1 \\ \hat{\mu}_4 - 6\hat{\mu}_2 + 3 \end{pmatrix},$$

$$W\xi(Y;\theta) = \frac{1}{\sigma^2} \begin{pmatrix} u^2 - 1 - 2\text{IF}\left(u;\hat{\sigma};\Phi\right) - d\left(u^4 - 6u^2 + 3\right) \\ u^3 - 3u \\ u^4 - 6u^2 + 3 \end{pmatrix},$$

and $\tilde{V} = WVW'$ is given by

$$\tilde{V} = \frac{1}{\sigma^4}\text{diag}[A, 6, 24],$$

where

$$A = -8 + 4\text{ASV}(\hat{\sigma}) + 2E\left(u^4\text{IF}\right) - \frac{1}{6}\left[E\left(u^4\text{IF}\right)\right]^2. \tag{10}$$

We conclude that the IM test statistic in the normal model can be written as the sum of three (asymptotically independent) statistics,

$$T = n\bigg[A^+\left(\hat{\mu}_2 - 1 - d(\hat{\mu}_4 - 6\hat{\mu}_2 + 3)\right)^2 + \frac{1}{6}(\hat{\mu}_3 - 3\hat{\mu}_1)^2$$
$$+ \frac{1}{24}(\hat{\mu}_4 - 6\hat{\mu}_2 + 3)^2\bigg]. \tag{11}$$

3.2 Different parameter estimators

Maximum Likelihood Estimator

If $\hat{\sigma}$ is the ML estimator, then IF $= (u^2 - 1)/2$, $d = A = 0$ (implying $\tilde{V}$ has rank 2), and

$$T = n \left[\frac{(\hat{\mu}_3 - 3\hat{\mu}_1)^2}{6} + \frac{(\hat{\mu}_4 - 6\hat{\mu}_2 + 3)^2}{24} \right],$$

with limiting χ_2^2 distribution. If, moreover, $\hat{\beta}$ is the ML estimator, then $\hat{\mu}_1 = 0$, $\hat{\mu}_2 = 1$, and T reduces to

$$T = n \left[\frac{\hat{\mu}_3^2}{6} + \frac{(\hat{\mu}_4 - 3)^2}{24} \right], \tag{12}$$

the well-known Jarque-Bera (1980) statistic for testing for skewness and non-normal kurtosis.

Robust Estimators

Some straightforward algebra shows that V has rank 2 only if IF $= (u^2 - 1)/2$. Thus, if $\hat{\sigma}$ is not the ML estimator, then V has full rank and T, given by (11), has a limiting χ_3^2 distribution. So, the IM test with robust estimators is sensitive to 'more' specification error than the IM test with ML estimators [1].

Throughout we use the median as a robust estimator of the location parameter β. The asymptotic distribution of T, however, does not depend on this choice. Neither does the local asymptotic power in the examples we will consider. Alternative M-estimators of location could be used, but we use the median since it is minimax robust (Huber, 1964).

As robust estimators of scale we consider two robust M-estimators. An M-estimator of scale, $\hat{\sigma}$, solves, for some chosen $c > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} \rho_c \left(\frac{Y_i - \hat{\beta}}{\hat{\sigma}} \right) = b_c,$$

where $b_c = E_\Phi \left[\rho_c(u) \right]$, $\hat{\beta}$ is an equivariant estimator of β, and ρ_c is an even function, not identically zero, non-decreasing on $[0, \infty[$, differentiable a.e.

---

[1]This property is not unique relative to robust estimators.

and for which $\rho_c(0) = 0$. This estimator is consistent for σ and has influence function (Huber, 1981)

$$\text{IF}(u; \hat{\sigma}; \Phi) = \frac{\rho_c(u) - b_c}{E_\Phi[\rho_c'(u)u]}.$$

For $\hat{\sigma}$ to be robust, ρ_c has to be bounded and $\hat{\beta}$ has to be a robust estimator of location (e.g. the median). The ML estimator of σ, for example, is an M-estimator of scale for which $\rho_c(u) = u^2$ (unbounded), $b_c = 1$, and $\hat{\beta}$ is the sample average, and hence is non-robust.

Consider the robust M-estimator of scale defined by Tukey's biweight (TB) function,

$$\rho_c(u) = \begin{cases} \frac{u^2}{2} - \frac{u^4}{2c^2} + \frac{u^6}{6c^4}, & \text{if } |u| \leq c; \\[2mm] \frac{c^2}{6}, & \text{if } |u| > c. \end{cases}$$

The choice of c is governed by the desired breakdown point of $\hat{\sigma}$. The details of how to compute T, for any choice of c, are given in Appendix A. Table 1 gives the numerical results for 10%, 25% and 50% breakdown points that are needed to compute T using the TB estimator.

Another, simpler, robust M-estimator of scale is the (standardized) Median Absolute Deviation (MAD),

$$\hat{\sigma} = \frac{1}{\Phi^{-1}(3/4)} \text{med}_i\{|Y_i - \text{med}_j Y_j|\},$$

for which $\rho_c(u) = I(|u| \geq c)$ with $c = \Phi^{-1}(3/4) = 0.6745$ and $b_c = 1/2$. The breakdown point of the MAD is 50%. Table 1 gives the constants, derived in Appendix A, that are needed to compute T using the MAD.

4 Local asymptotic power

Let F_n be a sequence of local alternatives tending to the null distribution, i.e. $F_n \to F_\theta$. Under a sequence of local alternatives,

$$H_n : Y \sim F_n,$$

the IM indicator $\hat{M}$ is, given some regularity conditions, asymptotically normally distributed,

$$\sqrt{n}(\hat{M} - m_n) \xrightarrow{\text{d}} \text{N}(0, V),$$

9

Table 1: Constants needed to compute the IM statistic T, as formulated in (11), at the normal model for several robust estimators of scale

estimator of scale	breakdown point	c	b_c	d	A
TB	10%	5.182	0.4476	0.03800	2.417×10^{-4}
	25%	2.937	0.3594	0.1059	0.03163
	50%	1.548	0.1996	0.1925	0.8224
MAD	50%	0.6745	0.5	0.2121	2.362

where $m_n = E_{F_n}[m(Y; \theta_n)]$ and θ_n such that $\hat{\theta} - \theta_n = \frac{1}{n} \sum_{i=1}^{n} \mathrm{IF}(Y_i; \hat{\theta}; F_n) + o_p(n^{-1/2})$. Assuming the existence of

$$b = \lim_{n \to \infty} \sqrt{n} \, m_n,$$

it follows that under H_n the IM test statistic $T = n\hat{M}'\hat{V}^+\hat{M}$ is asymptotically non-central χ^2,

$$T \xrightarrow{\mathrm{d}} \chi_q^2(\delta),$$

with non-centrality parameter

$$\delta = b'V^+b.$$

In the following subsections, we derive explicit expressions for δ when F_θ is the normal location-scale model, and F_n is a specific sequence of local alternatives. As local alternatives we consider a contaminated normal, Student's t, a skewed normal, and a tilted normal. By the results of Section 3.1, the non-centrality parameter takes the form

$$\delta = A^+ \tilde{b}_1^2 + \frac{1}{6}\tilde{b}_2^2 + \frac{1}{24}\tilde{b}_3^2,$$

where

$$\begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \end{pmatrix} = \tilde{b} = Wb,$$

with A and W defined in (10) and (8). Thus, for each alternative considered, we need to find $\tilde{b}$.

The IM test is defined by the parametric model and the estimator employed (also by the estimator of V, but this is of no concern here). It is not oriented towards a specific alternative, hence it is called an 'omnibus' test for misspecification. In contrast, the score test is defined by the parametric model and the alternative, and is usually implemented using the ML estimator. The fact that it is specifically designed as a test against a given alternative, and the way the test is constructed ensure that the score test is most powerful against any given local alternative (Godfrey, 1990). Because of this optimality property, it is natural to use the score test as a benchmark for evaluating the power of the IM test. Thus, we also carry out a local power analysis for the score test, except in the case of contamination because it is unlikely that, in practice, the precise form of potential contamination is known here.

In the remainder of this section, we take $F_\theta(y) = \frac{1}{\sigma}\Phi\left(\frac{y-\beta}{\sigma}\right)$, unless otherwise stated.

4.1   Contaminated normal alternative

Consider the sequence of local alternatives

$$H_n : Y \sim F_n = (1 - \frac{e}{\sqrt{n}})F_\theta + \frac{e}{\sqrt{n}}G \qquad (0 < e < 1),$$

where G is any distribution and e is the level of contamination. We have

$$
\begin{aligned}
b = \lim_{n\to\infty} \sqrt{n}\, m_n &= e \lim_{n\to\infty} \frac{\sqrt{n}}{e} M\left(\left(1 - \frac{e}{\sqrt{n}}\right)F_\theta + \frac{e}{\sqrt{n}}G\right) \\
&= e \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} M\left((1 - \epsilon)F_\theta + \epsilon G\right)
\end{aligned}
$$

with M the functional representation of $\hat{M}$. So, b is the Gâteaux derivative of M at F_θ in the direction of G. Hence (Hampel et al., 1986)

$$b = e \int \mathrm{IF}(y; M; F_\theta)dG(y).$$

By (5), $\mathrm{IF}(y; M; F_\theta) = \xi(y; \theta)$. Therefore, under H_n,

$$T \xrightarrow{\mathrm{d}} \chi_q^2(\delta_G) \tag{13}$$

11

where

$$\delta_G = e^2 E_G[\xi(Y;\theta)]' V^+ E_G[\xi(Y;\theta)].$$

The expression for the non-centrality parameter is valid for any F_θ and G, provided that $E_G[\xi(y;\theta)]$ exists. Note that, for $G = F_\theta$, $\delta_G = 0$ as it should be.

When F_θ is the normal distribution, then

$$\delta_G = e^2 \left[A^+ \left(E_G \tilde{\xi}_1 \right)^2 + \frac{1}{6} \left(E_G \tilde{\xi}_2 \right)^2 + \frac{1}{24} \left(E_G \tilde{\xi}_3 \right)^2 \right], \qquad (14)$$

where

$$
\begin{aligned}
\tilde{\xi}_1 &= u^2 - 1 - 2\text{IF} - d\tilde{\xi}_3, \\
\tilde{\xi}_2 &= u^3 - 3u, \\
\tilde{\xi}_3 &= u^4 - 6u^2 + 3,
\end{aligned}
$$

$\text{IF} = \text{IF}(u;\hat{\sigma};\Phi)$ and $u = \frac{Y-\beta}{\sigma}$. We see from (14) that, for any G, δ_G is minimal when $\hat{\sigma}$ is the ML estimator, because then $A = d = \tilde{\xi}_1 = 0$, which makes the first term in brackets vanish, and because the second and third term are independent of $\hat{\sigma}$. When $\hat{\sigma}$ is not the ML estimator, the first term vanishes only if $E_G \tilde{\xi}_1 = 0$. Thus, replacing the ML estimator with another estimator has two opposite effects on power: it increases (actually, non-decreases) the non-centrality parameter, but also increases the degrees of freedom from 2 to 3.

Taking $G = \Delta_x$, where Δ_x is the Dirac measure with all mass at x (representing outliers in the data at x), the non-centrality parameter is

$$\delta_x = e^2 \big(\xi(x;\theta)\big)' V^+ \xi(x;\theta),$$

while for $G = \frac{1}{2}(\Delta_x + \Delta_{-x})$ (representing symmetric outliers), it is

$$\delta_{-x,x} = \frac{e^2}{4} \big(\xi(x;\theta) + \xi(-x;\theta)\big)' V^+ \big(\xi(x;\theta) + \xi(-x;\theta)\big).$$

Figure 1 gives the non-centrality parameter for $G = \Delta_x$ and the three estimators discussed above (ML, MAD, and TB with 25% breakdown point). The non-centrality parameter for $G = \frac{1}{2}(\Delta_x + \Delta_{-x})$ is plotted in Figure 2.

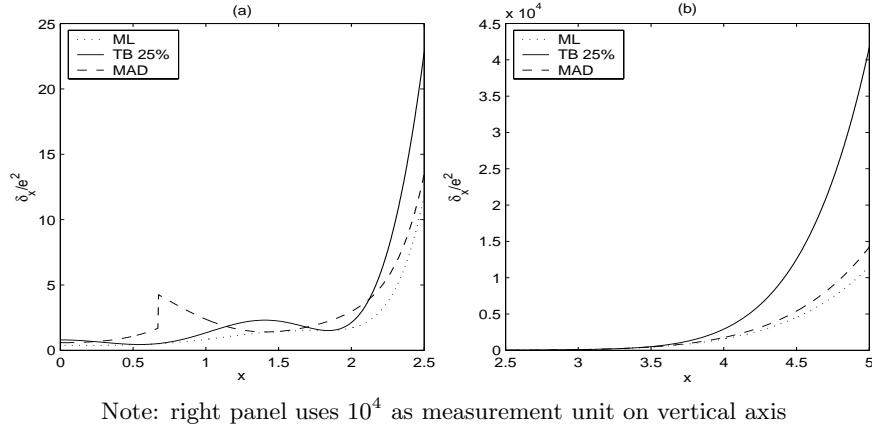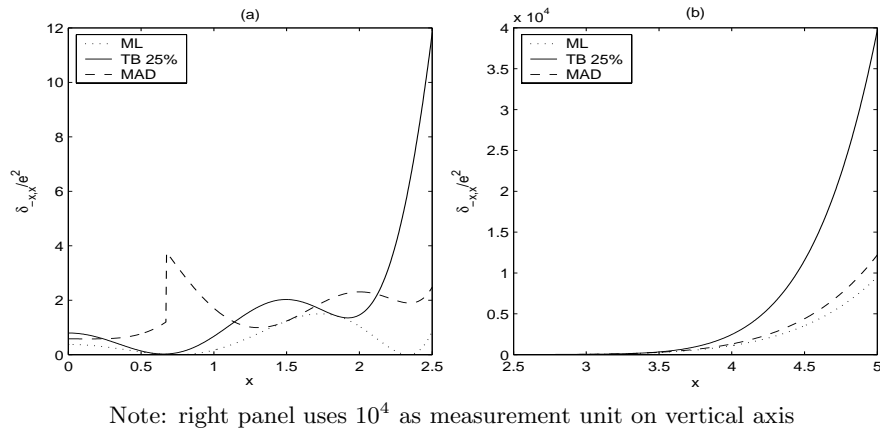Figure 1: Asymmetric contamination: non-centrality parameter

Note: right panel uses 10^4 as measurement unit on vertical axis

Figure 2: Symmetric contamination: non-centrality parameter

Note: right panel uses 10^4 as measurement unit on vertical axis

We see in both figures that the non-centrality parameter corresponding to the ML estimator is uniformly smaller than those corresponding to the other estimators, as shown. The non-centrality parameter associated with the MAD estimator is discontinuous at $x = \Phi^{-1}(3/4) = 0.6745$, where also $\rho_c(\cdot)$ is discontinuous. Figure 3 gives the power curves of 5%-level IM tests with level e contamination at $x = 10$, a clear outlier relative to the $N(0,1)$ distri-

Figure 3: POWER CURVES: (a) ASYMMETRIC CONTAMINATION AT $x = 10$ AND (b) SYMMETRIC CONTAMINATION AT $x = -10, 10$
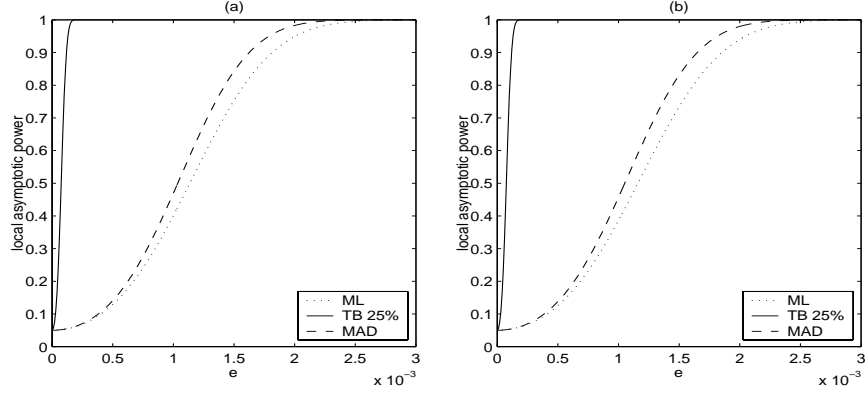
bution. The ordering is unambiguous: the IM test with the TB estimator is far more powerful than with the MAD or ML estimator. Using the MAD is slightly more powerful than using ML. The TB estimator with 25% breakdown point is a compromise between the ML estimator which is efficient but has 0% breakdown point, and the MAD estimator, which has 50% breakdown point but is very inefficient. This compromise yields a more powerful IM test.

4.2 Student's t alternative

Consider the sequence of local alternatives

$$H_n : Y \sim F_n = F_{t(\sqrt{n}/e)} \qquad (e > 0)$$

where $F_{t(p)}$ is the distribution function of a Student's t variate with p degrees of freedom. In Appendix B.1 it is shown that, under H_n,
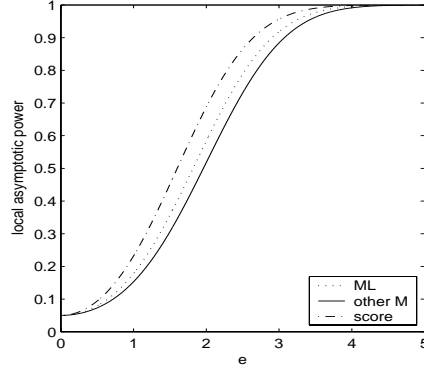
$$T \xrightarrow{\mathrm{d}} \chi_q^2(\delta)$$

with non-centrality parameter

$$\delta = \frac{3}{2}e^2,$$

for all M-estimators of scale.

14

Figure 4: POWER CURVES: STUDENT'S t

Let S be the score test statistic, defined in the usual way. Under the null hypothesis ($p = \infty$), S has a limiting χ_1^2 distribution. Under H_n, as we show in Appendix C.1,

$$S \xrightarrow{d} \chi_1^2(\delta),$$

with non-centrality parameter

$$\delta = \frac{3}{2}e^2.$$

In conclusion, the score test and the IM tests have the same non-centrality parameter in their limiting distribution. The power curves (as a function of e) of 5%-level tests are given in Figure 4. The differences in power are entirely due to differences in degrees of freedom: 1 for the score test, 2 for the IM test with the ML estimator, and 3 for the IM test using any other M-estimator. The difference in power between the IM tests is small.

4.3 Skewed normal alternative

Let $Z \sim N(0,1)$ and denote the distribution of $ZI(Z \leq 0) + (1+\gamma)ZI(Z > 0)$ as F_γ^{sk} (Fernández and Steel, 1998). Under the sequence of local alternatives

$$H_n : Y \sim F_n = F_{e/\sqrt{n}}^{sk},$$

15

Figure 5: POWER CURVES: SKEWED NORMAL
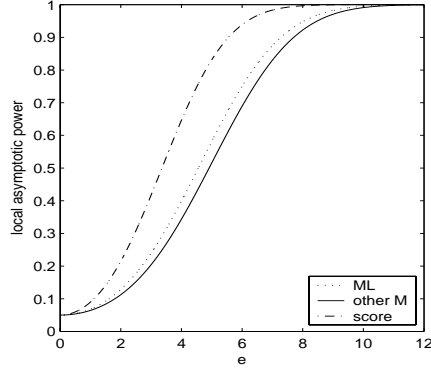
we have, as shown in Appendix B.2,

$$T \overset{\mathrm{d}}{\to} \chi_q^2(\delta)$$

with non-centrality parameter

$$\delta = \frac{3}{4\pi}e^2,$$

which is the same for all M-estimators of scale.

Regarding the score test, we show in Appendix C.2 that, under H_n,

$$S \overset{\mathrm{d}}{\to} \chi_1^2(\delta),$$

with non-centrality parameter

$$\delta = \frac{\pi - 1}{2\pi}e^2.$$

The power curves of 5%-level tests are given in Figure 5. The score test is now considerably more powerful than the IM test, primarily due to the larger non-centrality parameter. The difference between the different IM tests is, again, rather small.

4.4 Tilted normal alternative

Let $h(x)$ be a positive valued, scalar function satisfying $h(0) = 1$ and $\frac{dh(0)}{dx} = 1$, and consider the tilted normal density

$$f(y; \beta, \sigma, \kappa, \lambda) = \frac{1}{\sigma q(\kappa, \lambda)} \phi(u) h\left[\frac{\kappa}{6}(u^3 - 3u) + \frac{\lambda}{24}(u^4 - 6u^2 + 3)\right],$$

where $u = (Y - \beta)/\sigma$ and

$$q(\kappa, \lambda) = E_\Phi \left\{ h\left[\frac{\kappa}{6}(u^3 - 3u) + \frac{\lambda}{24}(u^4 - 6u^2 + 3)\right]\right\},$$

which is assumed to exist. Taking $h(x) = |x + 1|$ yields a density comparable to the first two terms of an Edgeworth expansion. Let $F_{\kappa,\lambda}$ be the distribution corresponding to $f(y; 0, 1, \kappa, \lambda)$. Then, as $\kappa, \lambda \to 0$, the first four moments of $F_{\kappa,\lambda}$ are (see Appendix C.3)

$$\begin{aligned}
E_{F_{\kappa,\lambda}}(Y) &= 0 + o(\kappa, \lambda), \\
E_{F_{\kappa,\lambda}}(Y^2) &= 1 + o(\kappa, \lambda), \\
E_{F_{\kappa,\lambda}}(Y^3) &= \kappa + o(\kappa, \lambda), \\
E_{F_{\kappa,\lambda}}(Y^4) &= 3 + \lambda + o(\kappa, \lambda),
\end{aligned} \tag{15}$$

from which κ and λ have an interpretation as skewness and (excess-)kurtosis parameters. Our interest in this distribution lies in the fact that the score test for $\kappa = \lambda = 0$ is in fact the Jarque-Bera test. Thus, under a sequence of local alternatives

$$H_n : Y \sim F_n = F_{\kappa,\lambda},$$

with $\kappa = k/\sqrt{n}$ and $\lambda = l/\sqrt{n}$, the IM test with ML estimator is optimal. We show in Appendix C.3 that, under H_n,
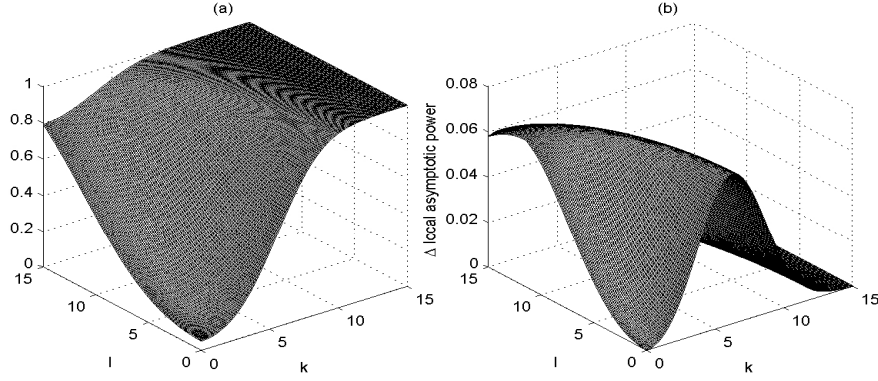
$$T \xrightarrow{\text{d}} \chi_q^2(\delta),$$

with non-centrality parameter

$$\delta = \frac{k^2}{6} + \frac{l^2}{24}$$

for all M-estimators of scale.

17

Figure 6: (a) POWER SURFACE ML ESTIMATOR; (b) DIFFERENCE ML ESTIMATOR-OTHER

The power surface for the IM test with ML estimator is plotted in Figure 6 (a). The difference in power between the IM test with ML estimator and the IM test with other M-estimators is plotted in Figure 6 (b). The maximal difference in power is 0.06623. It is again observed that the loss in local power when using robust estimators for the IM equality test is rather limited.

5 The normal regression model

For the normal model with covariates, $F_\theta(y|x) = \frac{1}{\sigma}\Phi(\frac{y-x'\beta}{\sigma})$ with $\theta = (\beta', \sigma)'$. We obtain, with $u = (Y - X'\beta)/\sigma$,

$$m(X, Y; \theta) = \frac{1}{\sigma^2}\begin{pmatrix} (u^2 - 1)\text{vech}(XX') \\ (u^3 - 3u)X \\ u^4 - 5u^2 + 2 \end{pmatrix},$$

where $\text{vech}(\cdot)$ is the lower triangular stack operator, and $\xi = \xi(X, Y; \theta)$,

$$\xi = \frac{1}{\sigma^2}\begin{pmatrix} (u^2 - 1)\text{vech}(XX') - \frac{2}{\sigma}\text{IF}(X, Y; \hat{\sigma}; K, F_\theta)E_K[\text{vech}(XX')] \\ (u^3 - 3u)X \\ u^4 - 5u^2 + 2 - \frac{2}{\sigma}\text{IF}(X, Y; \hat{\sigma}; K, F_\theta) \end{pmatrix}.$$

18

When $\hat{\sigma}$ is equivariant, $\mathrm{IF}(X, Y; \hat{\sigma}; F_\theta) = \sigma\mathrm{IF}(u; \hat{\sigma}; \Phi)$. Hence $V = \sigma^{-4}C$ with C a partitioned matrix with blocks

$$
\begin{aligned}
C_{11} &= 2E_K[\mathrm{vech}(XX')(\mathrm{vech}(XX'))'] \\
&\quad + (B_{11} - 2)E_K[\mathrm{vech}(XX')]E_K[\mathrm{vech}(XX')]', \\
C_{22} &= B_{22}E_K(XX'), \\
C_{33} &= B_{33}, \\
C_{13} &= B_{13}E_K[\mathrm{vech}(XX')] = C_{31}', \\
C_{12} &= C_{21}' = 0, \qquad C_{23} = C_{32}' = 0,
\end{aligned}
$$

and all B_{ij} as in Section 3. Again, V does not depend on the choice of the estimator of the location parameter β. Replacing E_K with sample averages yields an estimate $\hat{C}$ of C. For a given $\hat{\theta} = (\hat{\beta}', \hat{\sigma})'$, let $\hat{u}_i = (Y_i - X_i'\hat{\beta})/\hat{\sigma}$ and

$$
\hat{N} = \frac{1}{n}\sum_{i=1}^{n}
\begin{pmatrix}
(\hat{u}_i^2 - 1)\mathrm{vech}(X_iX_i') \\
(\hat{u}_i^3 - 3\hat{u}_i)X_i \\
\hat{u}_i^4 - 5\hat{u}_i^2 + 2
\end{pmatrix}
=
\begin{pmatrix}
\hat{N}_1 \\
\hat{N}_2 \\
\hat{N}_3
\end{pmatrix}.
$$

Then T takes the form

$$
T = n\left(\hat{N}_1'\hat{C}^{11}\hat{N}_1 + \hat{N}_2'\hat{C}^{22}\hat{N}_2 + \hat{N}_3'\hat{C}^{33}\hat{N}_3 + 2\hat{N}_1'\hat{C}^{13}\hat{N}_3\right),
$$

where $\hat{C}^{ij}$ is the (i, j)-th block of $\hat{C}^+$.

If $\hat{\theta}$ is the ML estimator and the first column of X is a vector of ones, then the first element of C_{11} is zero, $B_{13} = 0$ and $B_{33} = 24$. Let $p = k(k + 1)/2$, with $k = \dim(\beta)$, and

$$
\begin{aligned}
L &= \left(\mathbf{0}_{p-1\times 1} \;\; I_{p-1}\right), \\
L_i &= L\mathrm{vech}(X_iX_i'), \\
\bar{L} &= \frac{1}{n}\sum_{i=1}^{n}L_i.
\end{aligned}
$$

Then the IM test statistic with ML estimator can be written as (Hall, 1987)

$$
\begin{aligned}
T &= \frac{1}{2}\sum_{i=1}^{n}(\hat{u}_i^2 - 1)L_i'\left(\sum_{i=1}^{n}(L_i - \bar{L})(L_i - \bar{L})'\right)^{-1}\sum_{i=1}^{n}(\hat{u}_i^2 - 1)L_i \\
&\quad + \frac{1}{6}\sum_{i=1}^{n}\hat{u}_i^3 X_i'\left(\sum_{i=1}^{n}X_iX_i'\right)^{-1}\sum_{i=1}^{n}\hat{u}_i^3 X_i + \frac{1}{24}\sum_{i=1}^{n}(\hat{u}_i - 3)^2.
\end{aligned}
$$

19

For the computation of the IM test statistic with robust estimators (in particular, for computing B), we refer to Section 3.2, since all robust scale estimators considered here are M-estimators based on Tukey's biweight function. We consider two robust estimators of β. The first one is the S-estimator (Rousseeuw and Yohai, 1984),

$$\hat{\beta}_1 = \arg\min_\beta s(\beta),$$

where $s(\beta)$ is a robust M-estimator of scale, i.e. it solves

$$\frac{1}{n} \sum_{i=1}^n \rho_{c_1} \left(\frac{Y_i - X_i'\beta}{s(\beta)} \right) = b_{c_1},$$

where ρ_c and b_c are as in Section 3.2. The second one is the MM-estimator (Yohai, 1987), which solves

$$\hat{\beta}_2 = \arg\min_\beta \sum_{i=1}^n \rho_{c_2} \left(\frac{Y_i - X_i'\beta}{\hat{\sigma}} \right),$$

where $\hat{\sigma}$ is an initial residual scale estimator based on a very robust S-estimator, i.e. $\hat{\sigma} = s(\hat{\beta}_1)$. The constant c_2 is chosen large enough to obtain an increase in efficiency upon the S-estimator. By selecting c_1 and c_2 appropriately, this MM-estimator combines the high breakdown property (25%) with a higher statistical efficiency (95% at Gaussian models) than the S-estimator.

6   Monte Carlo results

6.1   The normal model

In Section 4 we studied the local behaviour of the IM test under various alternatives. Here the finite sample power (against fixed alternatives) is investigated by means of Monte Carlo experiments.

We look at three alternative hypotheses: a normal distribution contaminated with outliers, the Cauchy distribution, and the χ^2 distribution. To estimate the location robustly, we use the median, which has a breakdown point of 50%. As robust scale estimators we use the MAD and the TB M-estimator with 25% breakdown point. We carry out 10000 simulations. To

20

Figure 7: RP-POWER CURVES: (a) $n = 20$, 2 OUTLIERS AT 4 (b) $n = 20$, ONE OUTLIER AT -4, ONE OUTLIER AT 4
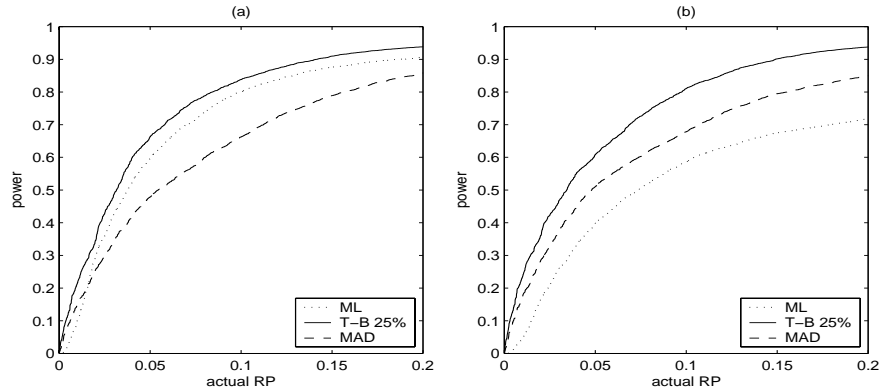
correct power for errors in rejection probability (ERP) [2] we plot power as function of (actual) rejection probability under the null of correct specification, as in Davidson and MacKinnon (1998). For pivotal statistics, such an RP-power curve [3] is obtained as follows. Run R Monte Carlo replications under the null and under the alternative hypothesis. Order the R test statistics obtained under the null from high to low to obtain $T_1^0 \geq \ldots \geq T_R^0$. The power at actual RP $k/(R+1)$ is then estimated as the fraction of test statistics generated under the alternative that are larger than T_k^0. Figure 7 plots the RP-power curves for $n = 20$, the alternative hypothesis being the normal distribution contaminated with (a) two outliers at 4 and (b) one outlier at -4 and one at 4. From Figure 7 (a) it is clear that the IM test with robust estimators may, but need not be, more powerful in the presence of outliers than the IM test with ML estimator. As a second alternative hypothesis we consider the Cauchy distribution. The RP-power curves for the Cauchy distribution are plotted in Figure 8. As conjectured, the IM tests with robust estimators have more power. A χ_3^2 alternative is considered in

---

[2]The ERP of a test is the actual minus the nominal (i.e. chosen) probability of rejecting the null when it is true.

[3]Davidson and MacKinnon (1998) call this a size-power curve. Because the size of a test, defined as the supremum, over the null, of the RP, often differs from the RP, we prefer the term RP-power curve. In this particular model, however, the statistic is pivotal and hence size equals RP.

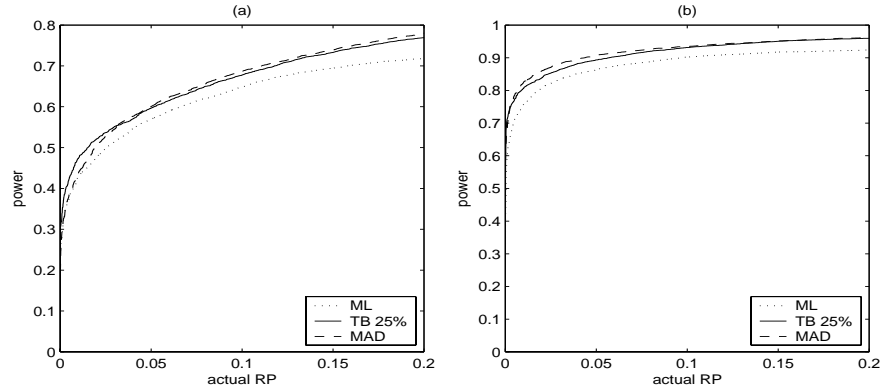Figure 8: RP-Power curves: Cauchy (a) $n = 10$ (b) $n = 20$

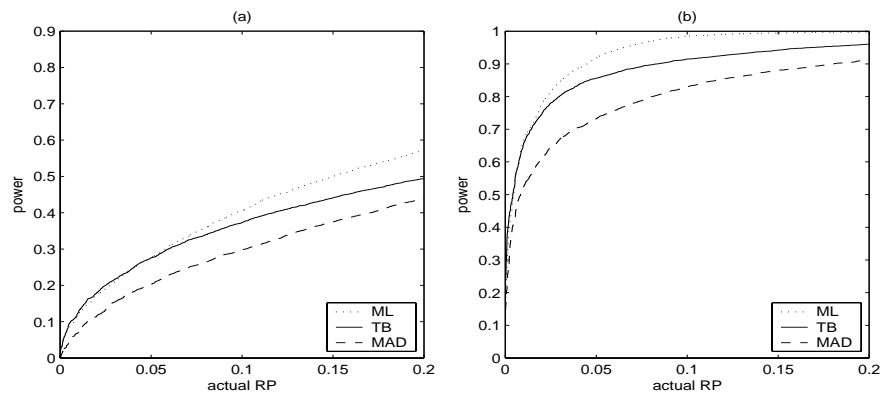Figure 9: RP-Power curves: χ_3^2 (a) $n = 10$ (b) $n = 50$

Figure 9. Here the IM test with ML estimator has better power properties.

6.2 The normal regression model

The following design is used in the Monte Carlo experiments. Throughout $k = 2$, $\beta = 0$ and $\sigma = 1$, with one of the regressors a standard normal variate and the other one a constant. The regressors are kept fixed across

22

Figure 10: REGRESSION: RP-POWER CURVES: CAUCHY (a) $n = 10$ (b) $n = 20$
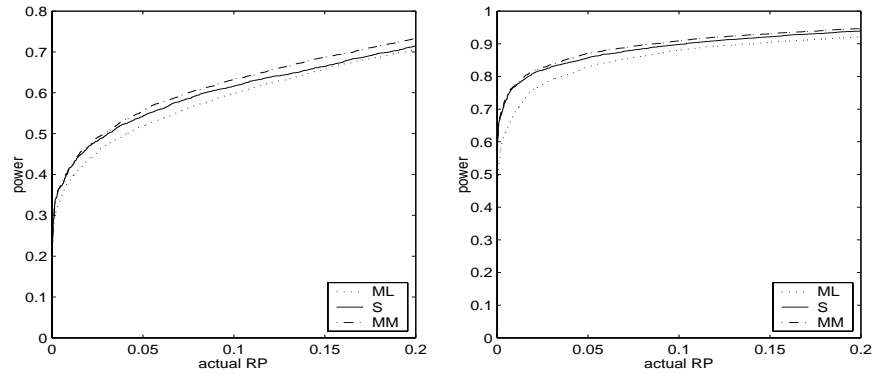
all replications. The sample size equals 10 or 20 for the Cauchy alternative, and 50 and 100 in the case of vertical outliers. The RP-power curves are constructed as explained in Section 6.1. In Figure 10 the RP-power curves are plotted for the Cauchy alternative. As in the normal model, the IM tests with robust estimators perform better than the IM test with ML estimator, although the difference in power is minor. In the presence of 20% vertical outliers lying at (a) $Y = 5$ when $n = 50$ and (b) $Y = 5$ and $Y = -5$ when $n = 100$, the TB and MAD estimators yield much more powerful IM tests than the ML estimator (see Figure 11). In this case, the 5%-level IM test with the ML estimator is biased, meaning that the actual RP under the null exceeds power. Figure 12 plots the RP-power curves for the case of bad leverage points, with sample sizes equal to 50 and 100, and 10% and 5% bad leverage points, respectively. In both cases the IM test with robust estimators has much more power, as expected. These simulation results convincingly demonstrate the usefulness of robust estimators in a regression context in connection with the IM test.

Figure 11: REGRESSION: RP-POWER CURVES: 20% VERTICAL OUTLIERS (a) AT $Y = 5$ ($n = 50$) (b) AT $Y = 5$ AND $Y = -5$ ($n = 100$)
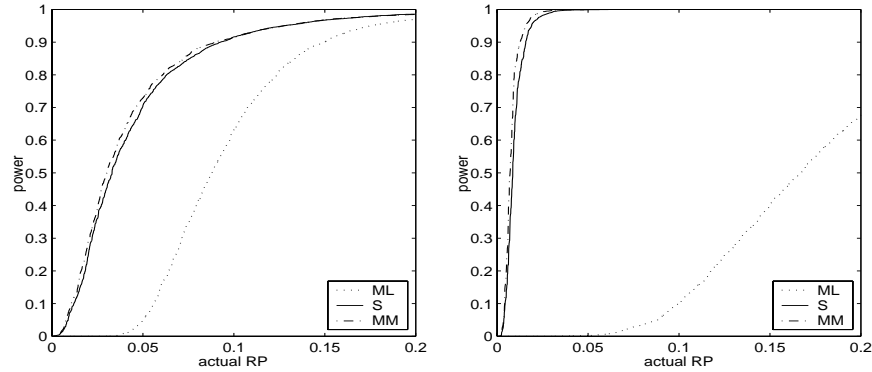
Figure 12: REGRESSION: RP-POWER CURVES: BAD LEVERAGE POINTS: (a) 5 POINTS AT $X = (1\ 6), Y = 6$ ($n = 50$); (b) 5 POINTS AT $X = (1\ 6), Y = 6$ ($n = 100$)
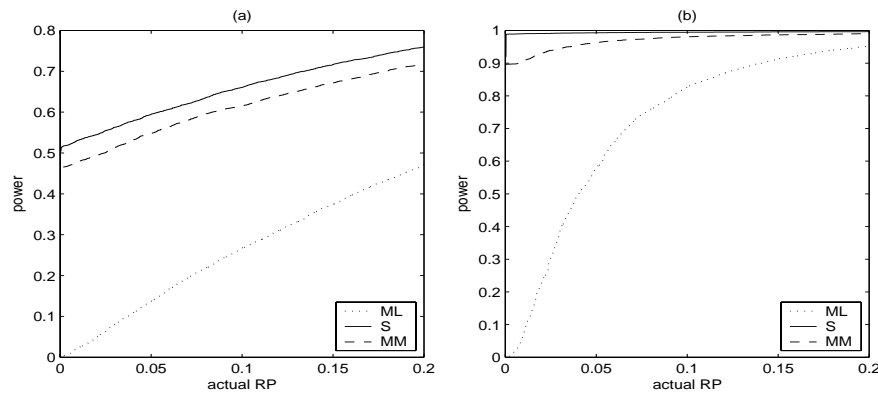
7   Conclusion

We have studied the behaviour of the IM test when robust estimators replace the ML estimator in the construction of the test. Particular attention has been given to the simplest of models, the normal location-scale model without covariates, where the IM test with ML estimator reduces to the

Jarque-Bera test. We have shown that, under local contamination, the ML estimator in fact minimizes the non-centrality parameter that appears in the limiting χ^2 distribution of the test statistic. Under some other alternatives (Student's t, skewed normal, tilted normal), the non-centrality parameter was found to be the same, whether using the ML estimator or robust estimators. In the regression model, the IM test with robust estimators clearly dominates the IM test with ML estimator in the presence of vertical outliers. Somewhat to our surprise, only minor differences between the tests were found under a Cauchy alternative.

The use of robust estimators makes the parameter estimates much less sensitive to outlying observations than when the ML estimator is used. As a result, such observations are more easily recognised as outliers, and outliers are in this context considered as evidence against the model. This intuition is supported by the analytical results concerning local contamination, for an arbitrary contaminating distribution, and by simulation results in the regression case.

We have focussed on the normal location-scale model and the regression model. The potential of using robust estimators in connection with the IM test in other models remains to be explored.

Appendix A
Computation of T using TB or MAD estimator

To compute T using the TB estimator, let, for k even,

$$\nu_c(k) = \int_{-c}^{c} u^k d\Phi(u) = \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) P\left(\frac{k+1}{2}, \frac{c^2}{2}\right),$$

with Γ the gamma and P the incomplete gamma function. Now,

$$
\begin{aligned}
b_c &= \frac{\nu_c(2)}{2} - \frac{\nu_c(4)}{2c^2} + \frac{\nu_c(6)}{6c^4} + \frac{c^2}{3}\left(1 - \Phi(c)\right), \\
E_\Phi\left[\rho_c'(u)u\right] &= \nu_c(2) - \frac{2\nu_c(4)}{c^2} + \frac{\nu_c(6)}{c^4}, \\
E_\Phi\left[\rho_c^2(u)\right] &= \frac{\nu_c(4)}{4} - \frac{\nu_c(6)}{2c^2} + \frac{5\nu_c(8)}{12c^4} - \frac{\nu_c(10)}{6c^6} + \frac{\nu_c(12)}{36c^8} \\
&\quad + \frac{c^4}{18}\left(1 - \Phi(c)\right),
\end{aligned}
$$

and ASV$(\hat{\sigma})$ follows from

$$\text{ASV}(\hat{\sigma}) = \frac{E_\Phi\left[\rho_c^2(u)\right] - b_c^2}{\left\{E_\Phi\left[\rho_c'(u)u\right]\right\}^2}.$$

Furthermore,

$$E_\Phi(u^4\text{IF}) = \frac{1}{E_\Phi\left[\rho_c'(u)u\right]} \left\{\frac{c^2}{6}\left(1 - \nu_c(4)\right) + \frac{\nu_c(6)}{2} - \frac{\nu_c(8)}{2c^2} + \frac{\nu_c(10)}{6c^4} - 3b_c\right\}.$$

The computation of T, from (11), is now straightforward.

For the MAD, $E_\Phi\left[\rho_c'(u)u\right] = 2c\phi(c)$ (with ϕ the standard normal pdf) and $E_\Phi\left[\rho_c^2(u)\right] = \frac{1}{2}$, resulting in

$$ASV(\hat{\sigma}) = \frac{1}{16c^2\left(\phi(c)\right)^2} = 1.361$$

and

$$E_\Phi(u^4\text{IF}) = \frac{1}{E_\Phi\left[\rho_c'(u)u\right]} \left\{\frac{3}{2} - \nu_c(4)\right\}.$$

Appendix B
Local asymptotic power of IM test

B.1 Student's t alternative

Under H_n, $Y \sim F_n = F_{t(p_n)}$ with $p_n = \sqrt{n}/e$. So

$$Y \sim \frac{Z}{\sqrt{U/p_n}},$$

where $U \sim \chi_{p_n}^2$ and $Z \sim N(0,1)$, with U and Z independent. By the Central Limit Theorem, as $p_n \to \infty$,

$$\frac{U}{p_n} = 1 + \sqrt{\frac{2}{p_n}}W + R,$$

where $W \sim N(0,1)$, W and Z are independent, the remainder term R is $O_p\left(p_n^{-1}\right)$, and $E(R) = 0$. Therefore σ_n solves

$$E\left[\rho_c\left(\frac{Z}{\sqrt{1 + \sqrt{2/p_n}W + R}\,\sigma_n}\right)\right] = b_c, \qquad (16)$$

Clearly, $\sigma_n \to 1$, since Student's t tends to $N(0,1)$ as $p_n \to \infty$. Rewrite (16) as

$$E\left[\rho_c\left(\frac{Z}{\sqrt{1+\varepsilon}}\right)\right] = b_c,$$

where

$$\varepsilon = \sigma_n^2 - 1 + \sigma_n^2\sqrt{2/p_n}W + \sigma_n^2 R.$$

Expanding $\rho_c(Z/\sqrt{1+\varepsilon})$ around $\varepsilon = 0$ gives

$$
\begin{aligned}
b_c &= E[\rho_c(Z)] - \frac{1}{2}E\left[(Z\rho_c'(Z)\varepsilon\right] + \frac{1}{8}E\left[(Z^2\rho_c''(Z) + 3Z\rho_c'(Z))\varepsilon^2\right] \\
&\quad + o(E(\varepsilon^2)).
\end{aligned}
$$

Since $E[\rho_c(Z)] = b_c$, $E(\varepsilon) = \sigma_n^2 - 1$ and $E(\varepsilon^2) = (\sigma_n^2 - 1)^2 + \frac{2}{p_n}\sigma_n^4 + o(p_n^{-1})$, we obtain,

$$
\begin{aligned}
b_c &= b_c - \frac{1}{2}(\sigma_n^2 - 1)E\left[Z\rho_c'(Z)\right] + \frac{1}{4p_n}\sigma_n^4 E\left[Z^2\rho_c''(Z) + 3Z\rho_c'(Z)\right] \\
&\quad + o\left(\sigma_n^2 - 1, p_n^{-1}\right).
\end{aligned}
$$

So

$$\sigma_n^2 = 1 + \frac{1}{2p_n}\Sigma_1 + o(p_n^{-1}), \tag{17}$$

where

$$\Sigma_1 = \frac{E\left[Z^2\rho_c''(Z) + 3Z\rho_c'(Z)\right]}{E\left[Z\rho_c'(Z)\right]}.$$

Let $(\tilde{m}_1, \tilde{m}_2, \tilde{m}_3)' = Wm(Y; \theta_n)$ with

$$Wm(Y; \theta) = \begin{pmatrix} u^2 - 1 - d(u^4 - 6u^2 + 3) \\ u^3 - 3u \\ u^4 - 6u^2 + 3 \end{pmatrix}.$$

By the symmetry of Student's t distribution, $E_{F_n}(\tilde{m}_2) = 0$. Furthermore,

$$
\begin{aligned}
\frac{E_{F_n}(Y^2)}{\sigma_n^2} - 1 &= E_{F_n}(Y^2) - \sigma_n^2 + o(p_n^{-1}) \\
&= \frac{p_n}{p_n - 2} - \frac{1}{2p_n}\Sigma_1 - 1 + o(p_n^{-1}) \\
&= \frac{4 - \Sigma_1}{2p_n} + o(p_n^{-1}), \tag{18}
\end{aligned}
$$

27

where the first line uses $(1 + x)/(1 + y) \approx 1 + x - y$ for x and y small, with $x = E_{F_n}(Y^2) - 1$ and $y = \sigma_n^2 - 1$, and the second line uses the second moment of Student's t, $p_n/(p_n - 2)$. We also have that

$$
\begin{aligned}
\frac{E_{F_n}(Y^4)}{\sigma_n^4} - 3 &= E_{F_n}(Y^4)(1 - 2(\sigma_n^2 - 1)) - 3 + o(p_n^{-1}) \\
&= \frac{3p_n^2}{(p_n - 2)(p_n - 4)}\left(1 - \frac{\Sigma_1}{p_n}\right) - 3 + o(p_n^{-1}) \\
&= \frac{18 - 3\Sigma_1}{p_n} + o(p_n^{-1}), \qquad (19)
\end{aligned}
$$

where the first line uses $(1 + x)^{-2} \approx 1 - 2x$ and the second line uses the fourth moment of Student's t, $3p_n^2(p_n - 2)^{-1}(p_n - 4)^{-1}$. Using (18) and (19),

$$
\begin{aligned}
E_{F_n}(\tilde{m}_3) &= \left(\frac{E_{F_n}(Y^4)}{\sigma_n^4} - 3\right) - 6\left(\frac{E_{F_n}(Y^2)}{\sigma_n^2} - 1\right) \\
&= \frac{6}{p_n} + o(p_n^{-1})
\end{aligned}
$$

and

$$
\begin{aligned}
E_{F_n}(\tilde{m}_1) &= \left(\frac{E_{F_n}(Y^2)}{\sigma_n^2} - 1\right) - dE_{F_n}(\tilde{m}_3) \\
&= \frac{1}{2p_n}(4 - \Sigma_1 - 12d) + o(p_n^{-1}).
\end{aligned}
$$

To compute Σ_1, note that, by partial integration and using $\phi'(z) = -z\phi(z)$,

$$
\begin{aligned}
E\left[Z\rho_c'(Z)\right] &= \int_{-\infty}^{\infty} z\phi(z)d\rho_c(z) \\
&= -\int_{-\infty}^{\infty} \rho_c(z)\big(\phi(z) - z^2\phi(z)\big)dz \\
&= E\left[Z^2\rho_c(Z)\right] - b_c.
\end{aligned}
$$

Along the same lines we get

$$
E\left[Z^2\rho_c''(Z)\right] = E\left[Z^3\rho_c'(Z)\right] - 2E\left[Z\rho_c'(Z)\right],
$$

and

$$
E\left[Z^3\rho_c'(Z)\right] = -3E\left[Z^2\rho_c(Z)\right] + E\left[Z^4\rho_c(Z)\right],
$$

28

from which an alternative expression for Σ_1 follows,

$$\Sigma_1 = \frac{E\left[\left(Z^4 - 2Z^2\right)\rho_c(Z)\right] - b_c}{E\left[Z^2\rho_c(Z)\right] - b_c}.$$

From (9),

$$
\begin{aligned}
d &= \frac{E_\Phi\left[(Z^4 - 6Z^2 + 3)(Z^2 - 1 - 2\mathrm{IF})\right]}{E_\Phi\left((Z^4 - 6Z^2 + 3)^2\right)} \\
&= \frac{1}{24} E_\Phi\left[(Z^4 - 6Z^2 + 3)(Z^2 - 1 - 2\mathrm{IF})\right] \\
&= -\frac{1}{12} \frac{E_\Phi\left[\left(Z^4 - 6Z^2\right)\rho_c(Z)\right] + 3b_c}{E\left[Z^2\rho_c(Z)\right] - b_c},
\end{aligned}
\tag{20}
$$

from which it is straightforward that $4 - 12d = \Sigma_1$, and thus $E_{F_n}(\tilde{m}_1) = o(p_n^{-1})$. Replacing p_n with $\sqrt{n}/e$, we obtain

$$\tilde{b} = \lim_{n\to\infty} \sqrt{n} E_{F_n}\left[Wm(Y;\theta_n)\right] = e \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$$

and

$$\delta = \tilde{b}'\tilde{V}^+\tilde{b} = \frac{3}{2}e^2.$$

B.2 Skewed normal alternative

Let $\gamma_n = e/\sqrt{n}$ and $F_n = F_{\gamma_n}^{sk}$. Let β be estimated by an M-estimator of location (e.g. the median). Then β_n and σ_n are the solutions of

$$E\left[\psi\left(\frac{Y - \beta_n}{\sigma_n}\right)\right] = 0,
\tag{21}$$

where ψ is an odd function, non-decreasing, not identically zero, and differentiable a.e., and

$$E\left[\rho_c\left(\frac{Y - \beta_n}{\sigma_n}\right)\right] = b_c,
\tag{22}$$

where expectations are with respect to $F_{\gamma_n}^{sk}$. From (21),

$$0 = \int_{-\infty}^{0} \psi\left(\frac{z - \beta_n}{\sigma_n}\right) d\Phi(z) + \int_{0}^{\infty} \psi\left(\frac{z(1 + \gamma_n) - \beta_n}{\sigma_n}\right) d\Phi(z). \qquad (23)$$

Now, expanding the integrands around $\beta_n = 0$, $\sigma_n = 1$, and $\gamma_n = 0$ gives

$$\psi\left(\frac{z - \beta_n}{\sigma_n}\right) = \psi(z) + \psi'(z)\left(-\beta_n - z(\sigma_n - 1)\right) + o\left(\beta_n, \sigma_n - 1\right) \qquad (24)$$

and

$$\psi\left(\frac{z(1 + \gamma_n) - \beta_n}{\sigma_n}\right) = \psi(z) + \psi'(z)\left(z\gamma_n - \beta_n - z(\sigma_n - 1)\right)$$
$$+ o\left(\beta_n, \sigma_n - 1, \gamma_n\right), \qquad (25)$$

and so (23) becomes

$$0 = E_\Phi\left[\psi(Z)\right] - \beta_n E_\Phi\left[\psi'(Z)\right] - (\sigma_n - 1)E_\Phi\left[Z\psi'(Z)\right]$$
$$+ \gamma_n \int_0^\infty z\psi'(z)d\Phi(z) + o\left(\beta_n, \sigma_n - 1, \gamma_n\right).$$

Since $E_\Phi\left[\psi(Z)\right] = E_\Phi\left[Z\psi'(Z)\right] = 0$, it follows that $\beta_n = \gamma_n \Sigma_2 + o(\gamma_n)$, with

$$\Sigma_2 = \frac{\int_0^\infty z\psi'(z)d\Phi(z)}{E\left[\psi'(Z)\right]}.$$

For σ_n it holds that

$$b_c = \int_{-\infty}^{0} \rho_c\left(\frac{z - \beta_n}{\sigma_n}\right) d\Phi(z) + \int_{0}^{\infty} \rho_c\left(\frac{z(1 + \gamma_n) - \beta_n}{\sigma_n}\right) d\Phi(z). \qquad (26)$$

Since (24) and (25) also hold with ρ_c replacing ψ, (26) becomes

$$b_c = E_\Phi\left[\rho_c(Z)\right] - \beta_n E_\Phi\left[\rho'_c(Z)\right] - (\sigma_n - 1)E_\Phi\left[Z\rho'_c(Z)\right]$$
$$+ \gamma_n \int_0^\infty z\rho'_c(z)d\Phi(z) + o\left(\beta_n, \sigma_n - 1, \gamma_n\right).$$

Now $E_\Phi\left[\rho_c(Z)\right] = b_c$, $E_\Phi\left[\rho'_c(Z)\right] = 0$ and $\int_0^\infty z\rho'_c(z)d\Phi(z) = \frac{1}{2}E_\Phi\left[Z\rho'_c(Z)\right]$, so we get

$$\sigma_n = 1 + \frac{\gamma_n}{2} + o(\gamma_n).$$

30

Further,

$$
\begin{aligned}
E_{F_n}(Y) &= \gamma_n \phi(0), \\
E_{F_n}(Y^2) &= 1 + \gamma_n + o(\gamma_n), \\
E_{F_n}(Y^3) &= 6\gamma_n \phi(0) + o(\gamma_n), \\
E_{F_n}(Y^4) &= 3 + 6\gamma_n + o(\gamma_n).
\end{aligned}
$$

Therefore, letting $(\tilde{m}_1, \tilde{m}_2, \tilde{m}_3)' = W m(Y; \theta_n)$,

$$
\begin{aligned}
E_{F_n}(\tilde{m}_3) &= E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^4\right] - 6 E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^2\right] + 3 = o(\gamma_n), \\
E_{F_n}(\tilde{m}_2) &= E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^3\right] - 3 E_{F_n}\left[\frac{Y - \beta_n}{\sigma_n}\right] = 3\gamma_n \phi(0) + o(\gamma_n), \\
E_{F_n}(\tilde{m}_1) &= E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_1}\right)^2\right] - 1 - d E_{F_n}(\tilde{m}_3) = o(\gamma_n),
\end{aligned}
$$

and, upon replacing γ_n with $e/\sqrt{n}$,

$$
\tilde{b} = \lim_{n \to \infty} \sqrt{n} E_{F_n}[W m(Y; \theta_n)] = e \begin{pmatrix} 0 \\ 3\phi(0) \\ 0 \end{pmatrix}.
$$

Hence $\tilde{b}' \tilde{V}^+ \tilde{b} = \frac{3}{4\pi} e^2$.

Appendix C
Local asymptotic power of score test

First, we review briefly how the local asymptotic power of the score test against specified alternatives can be defined. By an appropriate extension of $f(\cdot; \cdot)$, let the density under the alternative be $f(y; \omega)$, depending on an extended parameter ω, and let $s(y; \omega) = -\frac{\partial}{\partial \omega} \log f(y; \omega)$. Write the null hypothesis as $H_0 : \omega \in \Omega_0$, where Ω_0 is a restricted parameter space (essentially, Θ). Let $\hat{\omega}$ be the restricted ML estimator (essentially, $\hat{\theta}$), i.e. $\hat{\omega}$ solves

$$
\max_{\omega \in \Omega_0} \sum_{i=1}^{n} \log f(Y_i; \omega).
$$

31

Then, the score test statistic of H_0 is

$$S = n\bar{s}'\hat{\mathcal{J}}^{-1}\bar{s},$$

where

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s(Y_i; \hat{\omega})$$

and $\hat{\mathcal{J}}$ is, under H_0, a consistent estimator of

$$\mathcal{J} = E\left[s(Y; \omega_0)s(Y; \omega_0)'\right],$$

where $\omega_0 \in \Omega_0$ is the true value. Consider now a sequence of alternatives

$$H_n : Y \sim F_n,$$

where $F_n = F_{\omega_n}$ and $\omega_n = \omega_0 + e/\sqrt{n}$. Then, under H_n,

$$S \xrightarrow{\text{d}} \chi_q^2(\delta),$$

where $\delta = b'V^{-1}b$, with

$$b = \lim_{n\to\infty} \sqrt{n}E_{F_n} s(Y; \omega_{0,n})$$

and $\omega_{0,n}$ solves

$$\max_{\omega \in \Omega_0} E_{F_n} \log f(Y; \omega).$$

C.1 Student's t alternative

The log-density is

$$\log f(y; \beta, \sigma, \eta) = -\log \sigma - \log t_\eta(u),$$

where $u = (y - \beta)/\sigma$ and $t_\eta(u)$ is the density of a Student's t variable with $1/\eta$ degrees of freedom. The score function, evaluated at $\eta = 0$, is

$$s(y; \beta, \sigma, 0) = -\begin{pmatrix} \frac{u}{\sigma} \\ \frac{u^2-1}{\sigma} \\ \frac{u^4-2u^2-1}{4} \end{pmatrix},$$

32

where the last element follows from

$$t_\eta(u) = \phi(u) \left[1 + \frac{\eta}{4}(u^4 - 2u^2 - 1)\right] + o(\eta).$$

See e.g. Johnson et al. (1995, p. 375). The information matrix, evaluated at $\sigma = 1$ and $\eta = 0$, is

$$
\begin{aligned}
\mathcal{J} &= E\left[s(y;\beta,1,0)s(y;\beta,1,0)'\right] \\
&= E_\Phi \begin{pmatrix} u^2 & u^3 - u & u^5 - 2u^3 - u \\ \cdot & (u^2-1)^2 & \frac{u^6 - 3u^4 + u^2 + 1}{4} \\ \cdot & \cdot & \frac{(u^4 - 2u^2 - 1)^2}{16} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & \frac{7}{2} \end{pmatrix}
\end{aligned}
\tag{27}
$$

Since the first two elements of $\bar{s}$ equal zero at the restricted ML estimator, and since $\mathcal{J}^{33} = 2/3$, the score test statistic equals

$$S = n\frac{(\hat\mu^4 - 3)^2}{24},$$

which is the 'kurtosis part' of the Jarque-Bera statistic (12).

Let F_n be $F_{t(p_n)}$ with $p_n = \eta_n^{-1} = \sqrt{n}/e$. We then have that $\eta_{0,n} = 0$, and, from (17),

$$\sigma_{0,n}^2 = 1 + 2\eta_n + o(\eta_n),$$

since $\Sigma_1 = 4$ for the ML estimator. Using (18) and (19), with $\Sigma_1 = 4$, it follows that

$$E_{F_n}\left[s(Y;\beta_{0,n},\sigma_{0,n},0)\right] = \begin{pmatrix} 0 \\ o(\eta_n) \\ \frac{3}{2}\eta_n + o(\eta_n) \end{pmatrix}.$$

Hence, replacing η_n with $e/\sqrt{n}$,

$$b = \lim_{n\to\infty} \sqrt{n} E_{F_n}\left[s(Y;\beta_{0,n},\sigma_{0,n},0)\right] = e\begin{pmatrix} 0 \\ 0 \\ \frac{3}{2} \end{pmatrix},$$

and, using (27),

$$\delta = b'\mathcal{J}^{-1}b = \frac{3}{2}e^2.$$

33

C.2 Skewed normal alternative

The skewed-normal log-density is

$$\log f(y; \beta, \sigma, \gamma) = \begin{cases} -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{u^2}{2}, & \text{if } y \le \beta; \\ \\ -\frac{1}{2}\log(2\pi) - \log\sigma + \log(1+\gamma) - \frac{u_\gamma^2}{2}, & \text{if } y > \beta; \end{cases}$$

where $u = (y - \beta)/\sigma$ and $u_\gamma = u(1+\gamma)$. The score function, evaluated at $\gamma = 0$, is

$$s(y; \beta, \sigma, 0) = -\begin{pmatrix} u/\sigma \\ (u^2 - 1)/\sigma \\ I(y > \beta)(u^2 - 1) \end{pmatrix}.$$

Now let F_n be skewed normal with $\beta = 0$, $\sigma = 1$, and $\gamma_n = e/\sqrt{n} > 0$. Then, $\gamma_{0,n} = 0$, and by the results of Appendix B.2,

$$\begin{aligned} \beta_{0,n} &= E_{F_n}(Y) = \gamma_n \phi(0), \\ \sigma_{0,n} &= 1 + \frac{\gamma_n}{2} + o(\gamma_n). \end{aligned}$$

It follows that

$$E_{F_n}[s(Y; \beta_{0,n}, \sigma_{0,n}, 0)] = -\begin{pmatrix} 0 \\ o(\gamma_n) \\ E_{F_n}\left\{\left[\left(\frac{Y - \beta_{0,n}}{\sigma_{0,n}}\right)^2 - 1\right] I(Y > \beta_{0,n})\right\} \end{pmatrix}.$$

The third element in parentheses is

$$\begin{aligned} &\int_{\beta_{0,n}}^{\infty} \left(\frac{(1+\gamma_n)z - \beta_{0,n}}{\sigma_{0,n}}\right)^2 d\Phi(z) - 1 + \Phi(\beta_{0,n}) \\ &= \left(\frac{1+\gamma_n}{\sigma_{0,n}}\right)^2 \left(\beta_{0,n}\phi(\beta_{0,n}) + 1 - \Phi(\beta_{0,n})\right) - \frac{2(1+\gamma_n)\beta_{0,n}}{\sigma_{0,n}^2}\phi(\beta_{0,n}) - 1 \\ &\quad + \Phi(\beta_{0,n}) + o(\gamma_n) \\ &= (1+\gamma_n)\left(\gamma_n\left(\phi(0)\right)^2 + 1 - \Phi(\beta_{0,n})\right) - 2\gamma_n\left(\phi(0)\right)^2 - 1 + \Phi(\beta_{0,n}) \\ &\quad + o(\gamma_n) \\ &= \gamma_n\left(\frac{1}{2} - \left(\phi(0)\right)^2\right) + o(\gamma_n) \\ &= \gamma_n\left(\frac{\pi - 1}{2\pi}\right) + o(\gamma_n). \end{aligned}$$

Hence

$$b = \lim_{n \to \infty} \sqrt{n} E_{F_n}[s(Y; \beta_{0,n}, \sigma_{0,n}, 0)] = e \begin{pmatrix} 0 \\ 0 \\ \frac{1-\pi}{2\pi} \end{pmatrix}.$$

The information matrix is

$$J = E\left[s(y; 0, 1, 0)s(y; 0, 1, 0)'\right] = \begin{pmatrix} 1 & 0 & E_\Phi[u(u^2-1)I(u>0)] \\ \cdot & 2 & E_\Phi[(u^2-1)^2 I(u>0)] \\ \cdot & \cdot & E_\Phi[(u^2-1)^2 I(u>0)] \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \phi(0) \\ 0 & 2 & 1 \\ \phi(0) & 1 & 1 \end{pmatrix},$$

wherefrom $\delta = b' J^{-1} b = \frac{\pi - 1}{2\pi} e^2$.

C.3 Tilted normal alternative

Expanding $f(y; \beta, \sigma, \kappa, \lambda)$ around $\kappa = 0$ and $\lambda = 0$ gives

$$f(y; \beta, \sigma, \kappa, \lambda) = \frac{1}{\sigma} \phi(u) \left[1 + \frac{\kappa}{6}(u^3 - 3u) + \frac{\lambda}{24}(u^4 - 6u^2 + 3)\right] + o(\kappa, \lambda),$$

from which the moments given in (15) follow. The score function, evaluated at $\kappa = \lambda = 0$, is

$$s(y; \beta, \sigma, 0, 0) = - \begin{pmatrix} \frac{u}{\sigma} \\ \frac{u^2 - 1}{\sigma} \\ \frac{u^3 - 3u}{6} \\ \frac{u^4 - 6u^2 + 3}{24} \end{pmatrix}.$$

The information matrix is

$$\begin{aligned} J &= E\left[s(y; \beta, 1, 0, 0)s(y; \beta, 1, 0, 0)'\right] \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & \frac{1}{24} \end{pmatrix}. \end{aligned}$$

35

It follows that the score test statistic for testing $\kappa = \lambda = 0$ is

$$S = n \left[\frac{\hat{\mu}_3^2}{6} + \frac{(\hat{\mu}_4 - 3)^2}{24} \right],$$

which is the Jarque-Bera statistic (see (12)).

Let $\kappa_n = k/\sqrt{n}$ and $\lambda_n = l/\sqrt{n}$ and $F_n = F_{\kappa_n, \lambda_n}$. As before, β_n and σ_n are the solutions of (21) and (22) where expectations are now with respect to $F_{\kappa, \lambda}$. From (21) we have, using (24),

$$
\begin{aligned}
0 &= E_\Phi \bigg[\Big\{ \psi(Z) + \psi'(Z)(-\beta_n - Z(\sigma_n - 1)) \Big\} \Big\{ 1 + \frac{\kappa_n}{6}(Z^3 - 3Z) \\
&\quad + \frac{\lambda_n}{24}(Z^4 - 6Z^2 + 3) \Big\} \bigg] + o\left(\beta_n, \sigma_n - 1, \kappa_n, \lambda_n\right). \tag{28}
\end{aligned}
$$

Since $E_\Phi \left[\psi(Z) \right] = 0$ and ψ is odd, it follows that

$$\beta_n = \Sigma_3 \kappa_n + o\left(\beta_n, \sigma_n - 1, \kappa_n, \lambda_n\right), \tag{29}$$

where

$$\Sigma_3 = \frac{1}{6} \frac{E\left[(Z^3 - 3Z)\psi(Z) \right]}{E\left[\psi'(Z) \right]}.$$

Similarly, we have, from (22),

$$
\begin{aligned}
b_c &= E_\Phi \bigg[\Big\{ \rho_c(Z) + \rho_c'(Z)(-\beta_n - Z(\sigma_n - 1)) \Big\} \Big\{ 1 + \frac{\kappa_n}{6}(Z^3 - 3Z) \\
&\quad + \frac{\lambda_n}{24}(Z^4 - 6Z^2 + 3) \Big\} \bigg] + o\left(\beta_n, \sigma_n - 1, \kappa_n, \lambda_n\right). \tag{30}
\end{aligned}
$$

Now $E_\Phi \left[\rho_c(Z) \right] = b_c$ and ρ_c is even, so

$$\sigma_n = 1 + \Sigma_4 \lambda_n + \left(\beta_n, \sigma_n - 1, \kappa_n, \lambda_n\right), \tag{31}$$

where

$$\Sigma_4 = \frac{1}{24} \frac{E\left[(Z^4 - 6Z^2 + 3)\rho_c(Z) \right]}{E\left[Z\rho_c'(Z) \right]}.$$

Since

$$
\begin{aligned}
\frac{Y - \beta_n}{\sigma_n} &= Y(1 - (\sigma_n - 1)) - \beta_n + o\left(\kappa_n, \lambda_n\right) \\
&= Y(1 - \Sigma_4 \lambda_n) - \Sigma_3 \kappa_n + o\left(\kappa_n, \lambda_n\right),
\end{aligned}
$$

36

we obtain

$$E_{F_n}\left[\frac{Y - \beta_n}{\sigma_n}\right] = -\Sigma_3\kappa_n + o(\kappa_n, \lambda_n),$$

$$E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^2\right] = 1 - 2\Sigma_4\lambda_n + o(\kappa_n, \lambda_n),$$

$$E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^3\right] = \kappa_n - 3\Sigma_3\kappa_n + o(\kappa_n, \lambda_n),$$

$$E_{F_n}\left[\left(\frac{Y - \beta_n}{\sigma_n}\right)^4\right] = 3 + \lambda_n - 12\Sigma_4\lambda_n + o(\kappa_n, \lambda_n).$$

Letting $(\tilde{m}_1, \tilde{m}_2, \tilde{m}_3) = Wm(Y; \theta_n)$, this results in

$$E_{F_n}(\tilde{m}_1) = -\lambda_n(2\Sigma_4 + d) + o(\kappa_n, \lambda_n)$$
$$E_{F_n}(\tilde{m}_2) = \kappa_n + o(\kappa_n, \lambda_n)$$
$$E_{F_n}(\tilde{m}_3) = \lambda_n + o(\kappa_n, \lambda_n)$$

Note that, from (20),

$$\begin{aligned} d &= -\frac{1}{12}\frac{E\left[(Z^4 - 6Z^2 + 3)\rho_c(Z)\right]}{E\left[Z\rho_c'(Z)\right]} \\ &= -2\Sigma_4, \end{aligned}$$

and thus $E_{F_n}[\tilde{m}_1] = o(\kappa_n, \lambda_n)$. Since $\kappa_n = k/\sqrt{n}$ and $\lambda_n = l/\sqrt{n}$, we obtain

$$\tilde{b} = \lim_{n \to \infty} \sqrt{n}E_{F_n}\left[Wm(Y; \theta_n)\right] = \begin{pmatrix} 0 \\ k \\ l \end{pmatrix},$$

from which it is straightforward that

$$\delta = \tilde{b}\tilde{V}^+\tilde{b} = \frac{k^2}{6} + \frac{l^2}{24}.$$

Acknowledgements

References

[1] Chesher A. and R. Spady (1991), Asymptotic Expansions of the Information Matrix Test Statistic, *Econometrica* 59, 787-815.

[2] Davidson, R. and J.G. MacKinnon (1998), Graphical methods for investigating the size and power of hypothesis tests, *The Manchester School* 66, 1-26.

[3] Fernández, C. and M. Steel (1998), On Bayesian Modelling of Fat Tails and Skewness, *Journal of the American Statistical Association* 93, 359-371.

[4] Godfrey, L.G. (1990), *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Cambridge University Press,

[5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and W.A. Stahel (1986), *Robust statistics: the approach based on influence functions*, Wiley, New York.

[6] Hall, A. (1987), The information matrix test for the linear model, *The Review of Economic Studies*, 54, 257-263.

[7] Horowitz, J.L. (1994), Bootstrap-based Critical Values for the Information Matrix Test, *Journal of Econometrics* 61, 395-411.

[8] Huber, P.J. (1964), Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35,73-101.

[9] Huber, P.J. (1981), *Robust statistics*, Wiley, New York.

[10] Jarque, C.M. and A.K. Bera (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters* 6, 255-259.

[11] Johnson N.L., Kotz, S. and N. Balakrishnan (1995), *Continuous univariate distributions*, Volume 2, Wiley, New York.

[12] Orme C. (1990), The Small-Sample Performance of the Information-Matrix Test, *Journal of Econometrics* 46, 309-331.

[13] Rousseeuw, P.J. and V.J. Yohai (1984), Robust regression by means of S-estimators, in Robust and nonlinear time series analysis, edited by J. Franke, W. Härdle and R.D. Martin (eds), *Lecture notes in Statistics* 26, Springer, New York.

[14] White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-26.

[15] White, H. (1994), *Estimation, inference and specification analysis*, Cambridge University press, Cambridge.

[16] Yohai, V.J (1987), High breakdown point and high efficiency robust estimates for regression, *Annals of Statistics* 15, 642-656.