Choosing Between Two Income Distribution Models With Contaminated Data

by
Maria-Pia Victoria-Feser

London School of Economics
Houghton Street
London WC2A 2AE, UK

March 1996

Abstract

Choosing between two income distribution models typically involves testing two non-nested hypotheses, that is hypotheses such that one cannot be obtained as a special or limiting case of the other. Cox (1961, 1962) proposed a classical testing procedure based on the comparison of the maximized likelihood functions for the two models. In this paper it is shown that such a procedure is not robust in that a single observation can reverse the decision. A robust version of Cox-type test statistics is proposed which can be used for the comparison of any parametric model. Its robustness properties as well as other properties are shown in simulated examples.

Key words: M-estimators, model choice, robust tests, Income distribution, linear regression.
JEL classification: C12, D31.

1 Introduction

We study the problem of choosing a model which, according to certain criteria, best represents the data. We focus on tests between two non-nested hypotheses as defined by Cox (1961, 1962).

The areas of application are wide and the subject has attracted a lot of researchers; see e.g. Atkinson (1970), Williams (1970), Breusch and Pagan (1980), Davidson and MacKinnon (1981), Fisher and McAleer (1981), Godfrey and Pesaran (1983), MacKinnon, White, and Davidson (1983), Gourieroux, Monfort, and Trognon (1983), Loh 1985 and Dastoor (1985).

Although it is widely accepted that the problem is very important, the test statistics proposed in the literature have been often criticized for several reasons. The most studied one is the lack of accuracy of the approximation of the sample distribution of the statistic by its asymptotic distribution. Another (less studied) reason but at least as important is the lack of robustness of the testing procedure. Hall (1985), in the context of regression models, noting that the full information maximum likelihood estimator under normality is extremely sensitive to misspecifications of the error distribution, stated that *' ... the development of non-nested selection techniques based on more robust estimators would appear extremely desirable...'*. Aguirre-Torres and Gallant (1983) propose a generalization of the Cox statistic based on M-estimators for the parameters. The same idea can be found in Hampel, Ronchetti, Rousseeuw, and Stahel (1986), chap 7. However, not only do they leave open the question of the choice of the ρ-function defining the M-estimators, but also, as we will show, it is not sufficient to use robust estimators for the parameters to obtain a robust test.

The aim of this paper is to propose a robust procedure based on optimal bounded influence parametric tests developed recently by Heritier and Ronchetti (1994). Indeed, it is very important to develop robust procedures for tests, since if we estimate the parameters of a model robustly and then use a classical procedure to test hypotheses about the parameters or about the estimated model, the tests are not necessarily robust.

We begin by showing numerically how much a Cox-type test can be badly influenced by a very small amount of contamination in the data. We then show theoretically this effect by computing the *level influence function* (LIF) (Rousseeuw and Ronchetti 1979, Ronchetti 1982) of the Cox test. Finally, we show that we can bound this LIF with a new test obtained by considering the Cox test as a Lagrange multiplier or score test and applying the results on robust parametric tests. We end with the application of this new procedure to well known models. In particular, we will see that the new procedure is not only robust to small model deviations or contaminations but also that for at least the chosen particular cases, the asymptotic distribution of the robust test statistic is a better approximation of its sample distribution than in the classical case.

2 Robustness properties of Cox-type statistics

2.1 The test statistics

In general, it is assumed under H_0 (the hypothesis under test) that the model is F^0_α (with density $f^0(\cdot;\alpha)$) and that under H_1 (the alternative hypothesis) the model is F^1_β (with density $f^1(\cdot;\beta)$), where α and β are parameter vectors. The observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$, $\mathbf{x}_i \in \mathcal{R}^p$ are then supposed to be generated by F^0_α under H_0. Let $L_0(\mathbf{x};\hat\alpha) = \log f^0(\mathbf{x};\hat\alpha)$ and $L_1(\mathbf{x};\hat\beta) = \log f^1(\mathbf{x};\hat\beta)$ be the (maximum values of the) log-likelihood functions, where $\hat\alpha$ and $\hat\beta$ are the corresponding maximum likelihood estimators and define $L(\mathbf{x};\hat\alpha,\hat\beta) = L_0(\mathbf{x};\hat\alpha) - L_1(\mathbf{x};\hat\beta)$. Cox (1961, 1962) proposed the following test statistic

$$U_{Cox} = n^{-1}\sum L(\mathbf{x}_i;\hat\alpha,\hat\beta) - \int L(\mathbf{x};\hat\alpha,\beta_{\hat\alpha}) f^0(\mathbf{x};\hat\alpha)\mathbf{dx} \tag{1}$$

where $\sum$ stands for $\sum_{i=1}^n$ and $\beta_{\hat\alpha}$ is the pseudo maximum likelihood estimator defined as the solution in β of $\int \partial/\partial\beta \log f^1(\mathbf{x};\beta) f^0(\mathbf{x};\hat\alpha)\mathbf{dx} = 0$. There are two straightforward modifications of U_{Cox} statistics. One was proposed by Atkinson (1970) who suggested to use $\beta_{\hat\alpha}$ instead of $\hat\beta$, because $\beta_{\hat\alpha}$ is the maximum likelihood estimator of β under H_0 and (1) can be viewed as a score test statistic (see below). The second was proposed by White (1982) who suggested to use $\hat\beta$ instead of $\beta_{\hat\alpha}$. In these three cases, the asymptotic null distribution of $n^{1/2}U_{Cox}$ is the normal distribution with mean 0 and variance

$$var(F^0_\alpha) = E[L^2] - E[L]^2 - E[(s^0)^T L]\{E[(s^0)^T s^0]\}^{-1} E[s^0 L]$$

where $L = L(\mathbf{x};\alpha,\beta_\alpha)$, β_α is the limiting value of $\hat\beta$ or $\beta_{\hat\alpha}$ as $n \to \infty$, $s^0 = s^0(\mathbf{x};\alpha) = \partial/\partial\alpha \log f^0(\mathbf{x};\alpha)$ and $E[\cdot]$ is the expectation with respect to F^0_α, the argument of var. In practice one needs a consistent estimator of $var(F^0_\alpha)$, e.g. when α is replaced by $\hat\alpha$.

Atkinson (1970) and others showed that the Cox-type test statistics given in (1) can be interpreted as a Lagrange multiplier or score test. Indeed, if we construct the comprehensive model

$$f^c(\mathbf{x};\theta) = \{f^0(\mathbf{x};\alpha)\}^\lambda \{f^1(\mathbf{x};\beta)\}^{(1-\lambda)} \left[\int \{f^0(\mathbf{y};\alpha)\}^\lambda \{f^1(\mathbf{y};\beta)\}^{(1-\lambda)} \mathbf{dy} \right]^{-1} \tag{2}$$

where $\theta = (\alpha^T, \lambda)^T$, then the Lagrange multiplier or score test statistic corresponding to the hypothesis $\mathrm{H}_0 : \lambda = 1$ against the alternative $\mathrm{H}_1 : \lambda \neq 1$ leads to the Cox, Atkinson or White statistic, depending on the choice of the estimator $\tilde{\beta}$ of β.

2.2 Example

In this subsection we illustrate numerically, through a simulated example, the non robustness properties of Cox-type statistics. We consider here the problem of choosing between two competing regressors in a simple linear regression model (without intercept). Namely, we test $\mathrm{H}_0 : Y = \gamma_0 X + \epsilon_0$ against $\mathrm{H}_1 : Y = \gamma_1 Z + \epsilon_1$, where $\epsilon_0 \sim N(0, \sigma_0^2)$, $\epsilon_1 \sim N(0, \sigma_1^2)$ and the univariate variables X and Z are fixed. The simulated data are represented in the plots in Figures 1 and 2. One can see that X is the best choice for the regressor, except for an outlying point. If we rely on a Cox-type testing procedure, for example a White test, the value of the (standardized) test statistic we get is -2.16 (with corresponding p-value of 0.03). On the other hand, if we test $\mathrm{H}_0 : Y = \gamma_1 Z + \epsilon_1$ against $\mathrm{H}_1 : Y = \gamma_0 X + \epsilon_0$, the value of the (standardized) test statistic is -0.71 (with corresponding p-value of 0.48). Therefore, a classical test procedure would choose the regressor Z.

The classical test is clearly not robust. Intuitively, by looking at (1), one can see that this is not only due to the fact that the maximum likelihood of the parameters is not robust, but also because $\sum L(\mathbf{x}_i; \alpha, \beta)$ can be heavily influenced by a single observation. In our example, it is the outlying observation that influences heavily the value of the test statistic. A robust procedure would give less weight to this observation and lead to the choice of the regressor X.

We will see later that it is possible to build robust versions of Cox-type statistics that are not badly influenced by outlying observations and moreover can automatically detect them.

2.3 Level influence function

In this subsection, we compute the *LIF* of the Cox test. A similar result can be obtained for the Atkinson and White tests. We follow Hampel (1968) infinitesimal approach based on M-estimators. Ronchetti (1982), Rousseeuw and Ronchetti (1979, 1981) were the first to adapt Hampel's optimality problem for estimators to testing procedures, in the case of testing a null hypothesis about a one-dimensional parameter. Hampel's optimality problem for testing procedures can be stated as: Under a bound on the influence of small contamination on the test's level and power (robustness requirement), the power of the test at the ideal model is maximized (efficiency requirement). More recently, Heritier and Ronchetti (1994) have extended the existing theory on robust parametric tests to general parametric models.

To compute the LIF, one assumes a distribution in the neighborhood of the model under the null hypothesis and then studies the effect on the asymptotic level of the test. The neighborhood of the model at H_0 is given by

$$F^0_{\varepsilon_n} = (1-\varepsilon_n)F^0_\alpha + \varepsilon_n \Delta_{\mathbf{z}} \tag{3}$$

where $\Delta_{\mathbf{z}}$ is the distribution which gives a probability of 1 to an arbitrary point $\mathbf{z} \in \mathcal{R}^p$. Note that a more general neighbourhood is defined when instead of $\Delta_{\mathbf{z}}$ we choose any distribution. However, to determine the bias on the level of the test, by choosing (3), one actually has the worst situation, i.e. the LIF determines the worst bias on the level (Hampel et al. 1986, p. 175). To study the effect of this model deviation, it is necessary to choose a contamination ε_n that tends to zero at the rate $n^{-\frac{1}{2}}$ (Hampel et al. 1986, chap 3 and Heritier and Ronchetti 1994). Therefore, $\varepsilon_n = n^{-1/2}\varepsilon$.

The Cox statistic can be written as a functional of any distribution F as

$$n^{1/2}U(F) = n^{1/2}\left(\int L(\mathbf{x};\hat\alpha,\hat\beta)\mathbf{d}F(\mathbf{x}) - \int L(\mathbf{x};\hat\alpha,\beta_{\hat\alpha})f(\mathbf{x};\hat\alpha)\mathbf{dx}\right)$$

At $F = F^{(n)}$ where $F^{(n)}$ denotes the empirical distribution, and under H_0 it has an asymptotic normal distribution with zero mean and variance $var(F^0_\alpha)$. Note that $\hat\alpha$ and $\hat\beta$ can be written as functionals of the empirical distribution, i.e. $\hat\alpha(F^{(n)})$ and $\hat\beta(F^{(n)})$.

The test decision will be to reject H_0 if $|n^{1/2}U(F^{(n)})| > \kappa^*_\omega$, ($\kappa^*_\omega = var(F^0_\alpha)^{1/2}\ \Phi^{-1}(1-\omega/2)$, ω is the nominal level and Φ is the cumulative standard normal distribution). We define the true asymptotic level $\omega(0)$ (the argument 0 stands for no model contamination) at the model F^0_α of the test as the probability that $|n^{1/2}U(F^{(n)})|$ exceeds the critical value κ^*_ω. Let $\kappa_\omega = \kappa^*_\omega / var(F^0_\alpha)^{1/2}$, then we have

$$2^{-1}\omega(0) = 1 - \Phi(\kappa_\omega)\ .$$

Since the model is not always exactly true, the actual asymptotic level will in general be biased. This can lead the test to a rejection of H_0 only because of an infinitesimal proportion of outliers. As a neighborhood distribution, we consider the ε_n-contamination distribution given by (3), with $\varepsilon_n = \varepsilon n^{-1/2}$.

Assuming $F^0_{\varepsilon_n}$ under H_0 we then have that

$$\begin{aligned} n^{1/2}U(F^0_{\varepsilon_n}) &= n^{1/2}\left(\int L(\mathbf{x};\hat{\alpha}(F^0_{\varepsilon_n}),\hat{\beta}(F^0_{\varepsilon_n}))\mathbf{d}F^0_{\varepsilon_n}(\mathbf{x}) - \right. \\ &\left. \int L(\mathbf{x};\alpha,\beta_\alpha)f(\mathbf{x};\alpha)\mathbf{dx}\Big|_{\alpha=\hat{\alpha}(F^0_{\varepsilon_n})}\right) . \end{aligned}$$

If we assume Fréchet differentiability (Heritier and Ronchetti 1994, appendix 2, Clarke 1983 and Clarke 1986), we have that

$$n^{1/2}\left(U(F^{(n)}) - U(F^0_{\varepsilon_n})\right) \xrightarrow{\sim} N\left[0, var(F^0_\alpha)\right]$$

where $\xrightarrow{\sim}$ stands for convergence in distribution as $n \to \infty$. This result enables us to derive the LIF of the test.

Up to $O(1/n)$, the asymptotic level under the hypothetical model $F^0_{\varepsilon_n}$ is given by $2^{-1}\omega(\varepsilon) = 1 - \Phi\left(\{\kappa^*_\omega - \delta(\varepsilon)\}/var(F^0_\alpha)\right) = 1 - \Phi(k(\varepsilon))$, where $\delta(\varepsilon) = n^{1/2}U(F^0_{\varepsilon_n})$. The bias on the asymptotic level can be approximated by a first order Taylor expansion of $\omega(\varepsilon)$ around $\omega(0)$, i.e.

$$\omega(\varepsilon) - \omega(0) = \varepsilon \cdot \frac{\partial\omega(\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0} + O(\varepsilon^2).$$

where $\frac{\partial\omega(\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0} = LIF(\mathbf{z};\omega,F^0_\alpha)$. Therefore, the LIF carries all the information about the first order approximation of the bias on the asymptotic level of the test. If the LIF can take arbitrarily large values, it means the test is not robust, or in other words the decision can be determined by only a few observations.

It can be shown (see Victoria-Feser 1995), that for consistent estimators of β_α, i.e. $\hat{\beta}(F^0_\alpha) = \beta_\alpha$, the bias on the asymptotic level is then given by

$$\begin{aligned} \omega(\varepsilon) - \omega(0) &= -2\cdot\varepsilon\cdot var(F^0_\alpha)^{-1/2}\int_{-\infty}^{\kappa_\omega} y d\Phi(y)\cdot\Big\{\int L(\mathbf{x};\alpha,\beta_\alpha)\mathbf{d}F^0_\alpha(\mathbf{x}) - \\ &\quad L(\mathbf{z};\alpha,\beta_\alpha) + \int L(\mathbf{x};\alpha,\beta_\alpha)s^0(\mathbf{x};\alpha)\mathbf{d}F^0_\alpha(\mathbf{x})\cdot IF^*(\mathbf{z},\hat{\alpha},F^0_\alpha)\Big\} \\ &\quad + O(\varepsilon^2) \ . \end{aligned} \tag{4}$$

where

$$\begin{aligned} IF^*(\mathbf{z},\hat{\alpha},F^0_\alpha) &= \lim_{n\to\infty}\left[\{\partial/\partial\varepsilon_n\}\hat{\alpha}(F^0_{\varepsilon_n})\big|_{\varepsilon=0}\right] \\ &= -\left[\int(\partial/\partial\alpha^T)s^0(\mathbf{x};\alpha)\mathbf{d}F^0_\alpha(\mathbf{x})\right]^{-1} s^0(\mathbf{z};\alpha) \end{aligned}$$

is the *influence function* (IF) of the estimator $\hat{\alpha}$. That means that a single observation $\mathbf{z}$ such that $L(\mathbf{z};\alpha,\beta_\alpha) = \log f^0(\mathbf{z};\alpha) - \log f^1(\mathbf{z};\beta_\alpha)$ or $s^0(\mathbf{z};\alpha)$ is large, can make the bias on the asymptotic level very large. Indeed, the non-robustness of the test, i.e. the bias on the asymptotic level, is due simultaneously to

- the non-robustness of the estimator of the parameter α ($IF^*(\mathbf{z},\hat{\alpha},F^0_\alpha)$)
- the non-robustness of the test statistic ($L(\mathbf{z};\alpha,\beta_\alpha)$)

While $s^0(\mathbf{z};\alpha)$ is up to a multiplicative constant the IF of the maximum likelihood estimator of the parameter under the null hypothesis, $L(\mathbf{z};\alpha,\beta_\alpha)$ is directly related to the influence on the test statistic. Therefore, it is not sufficient to base a test on robust estimators for the parameters only. Indeed, a robust estimator for α guarantees a bounded value for $IF^*(\mathbf{z},\hat{\alpha},F^0_\alpha)$ but not for $L(\mathbf{z};\alpha,\beta_\alpha)$. For example, if we want to test the Gamma (F_{α_1,α_2}, α_1 is the shape parameter and α_2 is the scale parameter) against the Lognormal (F_{β_1,β_2}, $\beta_1 = \mu$ and $\beta_2 = \sigma^2$), the difference between the log-likelihood functions evaluated at any point $\mathbf{z}$ is given by (for a suitable function Q)

$$\left(\alpha_1 - \frac{\beta_1}{\beta_2}\right)\log(\mathbf{z}) + \frac{1}{2\beta_2^2}\log(\mathbf{z})^2 - \alpha_2\mathbf{z} + Q(\alpha_1,\alpha_2,\beta_1,\beta_2)$$

which can be large when $\mathbf{z}$ is large or small. Finally, note that (4) doesn't depend on the IF of $\hat{\beta}$, such that it is not necessary to choose a robust estimator for β to build a robust test statistic.

3 Robust model choice tests

3.1 Robust Cox-type statistics

The robust statistic we propose here is based on the results of Heritier and Ronchetti (1994). They propose a robust score test statistic based on M-estimators for testing subsets of parameter vectors. They show that an optimal test (i.e. one which maximizes the power at the model given a bound on the LIF) is obtained by considering optimal robust self-standardized estimators for the parameters under test, with partitioned parameters given in Hampel et al. (1986), p. 252-257. Let $\theta = (\theta_{(1)}^T, \theta_{(2)}^T)^T$ (of dimension $(1\times r_1, 1\times r_2)$) be the vector of parameters to be tested, such that $\mathrm{H}_0 : \theta_{(2)} = 0$, then for a given bound $c \geq r_2^{1/2}$ (which regulates the degree of robustness), the robust test statistic is based on the following ψ-function

$$\begin{aligned}\psi_c(\mathbf{x};\theta)_{(1)} &= A_{(11)}s(\mathbf{x};\theta)_{(1)}\\ \psi_c(\mathbf{x};\theta)_{(2)} &= h_c\left([A[s(\mathbf{x};\theta)-a]]_{(2)}\right)\end{aligned}$$

where $h_c(x) = x\min(1; c/\|x\|)$ is the Huber function, $a_{(1)} = 0$, the r_2-dimensional vector $a_{(2)}$ and the lower triangular matrix A are determined by the equations

$$\int \psi_c(\mathbf{x};\theta)_{(2)}\mathbf{d}F_\theta(\mathbf{x}) = 0 \tag{5}$$

$$\int \psi_c(\mathbf{x};\theta)\psi_c^T(\mathbf{x};\theta)\mathbf{d}F_\theta(\mathbf{x}) = I\ . \tag{6}$$

The robust score statistic is given by $R^2 = \{n^{-1}\sum\psi_c(\mathbf{x}_i;(\hat\theta_{(1)},0))_{(2)}\}^T\cdot\{n^{-1}\sum\psi_c(\mathbf{x}_i;(\hat\theta_{(1)},0))_{(2)}\}$ where $\hat\theta_{(1)}$ is the maximum likelihood of $\theta_{(1)}$. R^2 has asymptotically a $\chi^2_{r_2}$ distribution. We note that depending on the choice of the bound c, one can have a more or less robust test statistic. The lower the bound c the more robust the test statistic. At $c = \infty$, we have the classical non-robust statistic.

If we consider the compound model (equation (2)), the score function under H_0 ($\lambda = 1$), becomes

$$s^c(\mathbf{x};\theta) = \frac{\partial}{\partial\theta}\log f^c(\mathbf{x};\theta)\bigg|_{\lambda=1} = \begin{bmatrix} s^c(\mathbf{x};\theta)_{(1)}\\ s^c(\mathbf{x};\theta)_{(2)}\end{bmatrix}$$

where the vector $s^c(\mathbf{x};\theta)_{(1)}$ of dimension $\dim(\alpha)\times 1$ is given by

$$s^c(\mathbf{x};\theta)_{(1)} = \frac{\partial}{\partial\alpha}\log f^c(\mathbf{x};\theta)\bigg|_{\lambda=1} = \frac{\partial}{\partial\alpha}\log f^0(\mathbf{x};\alpha) = s^0(\mathbf{x};\alpha)$$

and the scalar $s^c(\mathbf{x};\theta)_{(2)}$ is given by

$$\begin{aligned}s^c(\mathbf{x};\theta)_{(2)} &= s_{Cox}(\mathbf{x};\alpha,\beta) = \frac{\partial}{\partial\lambda}\log f^c(\mathbf{x};\theta)\bigg|_{\lambda=1}\\ &= L(\mathbf{x};\alpha,\beta) - \int L(\mathbf{y};\alpha,\beta)f^0(\mathbf{y};\alpha)\mathbf{dy}\ .\end{aligned}$$

We then have that under H_0

$$\psi_c(\mathbf{x};\theta) = \begin{bmatrix} A_{(11)}s^0(\mathbf{x};\alpha)\\ \left[A_{(21)}s^0(\mathbf{x};\alpha) + A_{(22)}[s_{Cox}(\mathbf{x};\alpha,\beta) - a_{(2)}]\right]W_c(\mathbf{x};\alpha,\beta)\end{bmatrix}$$

where

$$W_c(\mathbf{x};\alpha,\beta) = \min\left\{1; c\cdot\left|A_{(21)}s^0(\mathbf{x};\alpha) + A_{(22)}[s_{Cox}(\mathbf{x};\alpha,\beta) - a_{(2)}]\right|^{-1}\right\}\ . \tag{7}$$

The robust Cox-type or generalized Lagrange multiplier (GLM) test statistic is finally given by

$$U_{GLM} = \frac{1}{n}\sum_{i=1}^{n}\left[A_{(21)}s^0(\mathbf{x}_i;\hat{\alpha}) + A_{(22)}[s_{Cox}(\mathbf{x}_i;\hat{\alpha},\beta) - a_{(2)}]\right] W_c(\mathbf{x}_i;\hat{\alpha},\beta) \tag{8}$$

where $\hat{\alpha}$ is the maximum likelihood estimator of α, the vector $A_{(21)}$ $(1\times\dim(\alpha))$ and the scalars $A_{(22)}$ and $a_{(2)}$ are determined implicitly by (5) and (6), where $F_\theta = F_\theta^c$ is the comprehensive model at $\hat{\theta} = (\hat{\alpha}^T, 1)^T$.

The asymptotic normality of $n^{1/2}U_{GLM}$ under H_0 is proven in Heritier and Ronchetti (1994) (see proof of proposition 2). In practice however, β needs to be estimated. If $\tilde{\beta}$ is a consistent estimator, then the asymptotic normality still holds. Indeed, to prove it, one would follow the proof given by Heritier and Ronchetti (1994) and add a development of the statistic around $\tilde{\beta}$. The added term would disappear for consistent estimators of β.

Knowing that the bias on the asymptotic level (4) is proportional to $s^0(\mathbf{z};\alpha)$ and to $L(\mathbf{z};\alpha,\beta)$, we see that by using the robust version of the Lagrange multiplier test with the comprehensive model (2), we bound exactly the right quantity. Therefore, the use of (8) avoids that the decision is influenced by a small amount of outliers.

3.2 Computation of the robust Cox-type test statistics

The computation of the test statistics is not straightforward. However, we propose an algorithm based on algorithm for robust estimators (see Victoria-Feser and Ronchetti 1994) that works well in practice. For a given bound c, it is given by the following 4 steps.

Step 1: Compute the maximum likelihood estimator for α and let $\hat{\theta} = (\hat{\alpha}^T, 1)^T$.

Step 2: Solve for $A_{(21)}$, $A_{(22)}$ and $a_{(2)}$, the following implicit equations

$$\int \psi_c(\mathbf{x};\hat{\theta})_{(2)} f^0(\mathbf{x};\hat{\alpha})\mathbf{dx} = 0 \tag{9}$$

$$\int \psi_c(\mathbf{x};\hat{\theta})_{(2)}\psi_c(\mathbf{x};\hat{\theta})^T_{(1)} f^0(\mathbf{x};\hat{\alpha})\mathbf{dx} = \mathbf{0} \tag{10}$$

$$\int \psi_c(\mathbf{x};\hat{\theta})^2_{(2)} f^0(\mathbf{x};\hat{\alpha})\mathbf{dx} = 1 \tag{11}$$

Step 3: Compute U_{GLM} given in (8).

Step 4: At the nominal level ω, accept H_0 if $|n^{1/2}U_{GLM}| < \kappa_\omega$, where $\kappa_\omega = \Phi^{-1}(1-\omega/2)$.

The second step is not straightforward since one has to solve a complicated nonlinear system in $a_{(2)}$, $A_{(21)}$ and $A_{(22)}$. We propose to use an iterative process combined with a classical routine to find the zero roots of a system of nonlinear equations. Equations (9), (10) and (11) can be written as

$$\begin{aligned} A_{(21)}E\left[s^0(\mathbf{x};\hat\alpha)W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &+ \\ A_{(22)}E\left[s_{Cox}(\mathbf{x};\hat\alpha,\tilde\beta)W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &- \\ A_{(22)}a_{(2)}E\left[W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &= 0 \end{aligned} \tag{12}$$

$$\begin{aligned} A_{(21)}E\left[s^0(\mathbf{x};\hat\alpha)s^0(\mathbf{x};\hat\alpha)^T W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &+ \\ A_{(22)}E\left[s^0(\mathbf{x};\hat\alpha)^T s_{Cox}(\mathbf{x};\hat\alpha,\tilde\beta)W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &- \\ A_{(22)}a_{(2)}E\left[s^0(\mathbf{x};\hat\alpha)^T W_c(\mathbf{x};\hat\alpha,\tilde\beta)\right] &= \mathbf{0} \end{aligned} \tag{13}$$

$$\begin{aligned} A_{(21)}E\left[s^0(\mathbf{x};\hat\alpha)s^0(\mathbf{x};\hat\alpha)^T W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right]A_{(21)}^T &+ \\ 2A_{(21)}A_{(22)}E\left[s^0(\mathbf{x};\hat\alpha)s_{Cox}(\mathbf{x};\hat\alpha,\tilde\beta)W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right] &- \\ 2A_{(21)}A_{(22)}a_{(2)}E\left[s^0(\mathbf{x};\hat\alpha)W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right] &+ \\ A_{(22)}^2E\left[s_{Cox}(\mathbf{x};\hat\alpha,\tilde\beta)^2W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right] &- \\ 2A_{(22)}^2a_{(2)}E\left[s_{Cox}(\mathbf{x};\hat\alpha,\tilde\beta)W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right] &+ \\ A_{(22)}^2a_{(2)}^2E\left[W_c(\mathbf{x};\hat\alpha,\tilde\beta)^2\right] &= 1 \end{aligned} \tag{14}$$

where the expectations are taken at $F_{\hat\alpha}$. The procedure is then the following: given values for $a_{(2)}$, $A_{(21)}$ and $A_{(22)}$, compute the weights and the expectations and then, given these expectations solve for $a_{(2)}$, $A_{(21)}$ and $A_{(22)}$ equations (12), (13) and (14). As starting values, one could choose $a = 0$ and $A_{(21)}$, $A_{(22)}$ such that $A^{-1}A^{-T} = \int s^c(\mathbf{x};\hat\theta)s^c(\mathbf{x};\hat\theta)^T f^0(\mathbf{x};\hat\alpha)\mathbf{dx}$, since they are the solutions when $c = \infty$.

3.3 Simulation study

In order to study the robustness properties of U_{GLM} when compared to the classical Cox-type statistics, we used the test when the samples are contaminated and non-contaminated. We choose to simulate Pareto samples and test the Pareto distribution against the Exponential distribution by means of the Atkinson statistic. Note that we obtain similar results with other Cox-type statistics.

The Pareto density is given by $f^0(x;\alpha) = \alpha x^{-(\alpha+1)} x_0^\alpha$ with $0 < x_0 \leq x < \infty$, so that as the alternative we considered the truncated exponential distribution given by $f^1(x;\beta) = \beta e^{-\beta(x-x_0)}$. We simulated 1000 samples of 200 observations from a Pareto distribution with parameter $\alpha = 3.0$ ($x_0 = 0.5$), samples that we contaminated by means of $\left(1 - \varepsilon \cdot 200^{-1/2}\right) F_{\alpha,x_0} + \varepsilon \cdot 200^{-1/2} F_{\alpha,10\cdot x_0}$. For amounts of contaminations from $\varepsilon = 0\%$ to $\varepsilon = 20\%$, Table 1 gives the actual levels of the classical and robust Atkinson statistic when testing the Pareto against the Exponential distribution. We chose a very robust version by putting $c = 2.0$.

The actual levels are the probabilities that ($n^{1/2}$ times) the test statistic (absolute value) computed from the simulated samples exceed the critical value κ_ω, where ω is the nominal level. We can observe that the classical statistic has a very strange behaviour since when there is no contamination the null hypothesis is underrejected and with only small amounts of contamination, the null hypothesis is overrejected. The first phenomenon is probably due to the fact that the approximation of the actual distribution of the Cox-type statistics by means of their asymptotic distribution is not accurate (Williams 1970, Atkinson 1970 and Loh 1985). The second phenomenon is the lack of robustness as shown in subsection 2.3. On the other hand, we find that with the robustified Atkinson statistic not only the asymptotic distribution is a good approximation of its sample distribution, but also that the small departures from the model under the null hypothesis do not influence the level of the test at least for amounts of contamination up to about $\varepsilon = 10\%$. With more contamination (15% and 20%), the null hypothesis tends to be slightly overrejected at the 5% and 10% levels, but this is not too drastic compared to the classical case. In other words, the robust test is very stable.

The fact that the level of the robust test is not very much influenced by contamination is due to the structure of the test itself (8). However, that the asymptotic distribution of the robust test statistic is a good approximation of its sample distribution (compared to the classical test) can seem at first rather surprising. This can be understood intuitively by remembering the probable causes of the problems in the classic test: Atkinson (1970) remarked that some rather small (legitimate) observations have a large influence on the value of the test statistic because one often takes their logarithm. With robust techniques the influence of such 'extreme but legitimate' observations is bounded, such that the null hypothesis is not under- or overrejected.

3.4 The regression example

As a second example, we compute the robust White test statistic for the regression example presented in subsection 2.2. The robust test statistic is given by

$$\begin{aligned} U_{GLM} &= \frac{1}{n}\sum_{i=1}^{n}\big[A_{(21)}s^0(y_i|x_i,\hat{\gamma}_0,\hat{\sigma}_0^2)+A_{(22)}[s_{Cox}(y_i|x_i,z_i,\hat{\gamma}_0,\hat{\sigma}_0^2,\hat{\gamma}_1,\hat{\sigma}_1^2)- \\ & a_{(2)}]\big]W_c(y_i|x_i,z_i,\hat{\gamma}_0,\hat{\sigma}_0^2,\hat{\gamma}_1,\hat{\sigma}_1^2) \end{aligned}$$

where

$$s^0(y|x_i,\gamma_0,\sigma_0^2)=\begin{bmatrix} \frac{1}{\sigma_0^2}x_i(y-x_i\gamma_0) \\ -\frac{1}{2\sigma_0^2}+\frac{1}{2\sigma_0^4}(y-x_i\gamma_0)^2 \end{bmatrix}$$

and

$$\begin{aligned} s_{Cox}(y|x_i,z_i,\gamma_0,\sigma_0^2,\gamma_1,\sigma_1^2) &= \frac{1}{2}\Bigg(\frac{1}{\sigma_1^2}(y-z_i\gamma_1)^2- \\ & \frac{1}{\sigma_0^2}(y-x_i\gamma_0)^2-\frac{\sigma_0^2}{\sigma_1^2}-\frac{(x_i\gamma_0-z_i\gamma_1)^2}{\sigma_1^2}+1\Bigg) \end{aligned}$$

We found for the robust White test statistic ($c=1.5$) a value of -1.47 with corresponding p-value of 0.14. On the other hand, if we test the regressor Z (under H_0) against the regressor X, we get for the test statistic a value of -2.06 with corresponding p-value of 0.02. A robust approach would then choose the regressor X.

This is a more comforting result, since common sense (in this case) suggests the same choice. In general, one cannot always rely on common sense, especially in complicated problems where there can be masking effects. A robust procedure that takes automatically into account the possibility of having very influential outliers is then necessary. The robust Cox-type test we propose, computes automatically weights associated to each observation. In our example, the weights are represented by the numbers next to the points on the plot in Figure 3. As one would expect, the 'nice' data receive a weight equal to one, whereas the outlier receives a very small weight.

4 Acknowledgments

The author was partially supported by the 'Fond National Suisse pour la Recherche Scientifique' and by an ESRC research Grant #R000 23 5725.

She would like to thank Prof. A. C. Atkinson for his helpful comments during the preparation of this paper, Prof. E. Ronchetti, Dr. M. Knott and an anonymous referee for their valuable comments on earlier versions of the paper.

References

Aguirre-Torres, V. and A. R. Gallant (1983). The null and non-null asymptotic distribution of the Cox test for multivariatenonlinear regression: Alternatives and a new distribution-free Cox test. *Journal of Econometrics 21*, 5–33.

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society, Series B 32*, 323–353.

Breusch, T. S. and A. R. Pagan (1980). The Lagrange multiplier test and its application to model specificationin econometrics. *Review of Economic Studies 47*, 239–253.

Clarke, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximumlikelihood-type equations. *The Annals of Statistics 11*, 1196–1205.

Clarke, B. R. (1986). Nonsmooth analysis and Fréchet differentiability of M functionals. *Probality Theory and Related Fields 73*, 197–209.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statisticsand Probability 1*, Berkeley, pp. 105–123. University of California Press.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B 24*, 406–424.

Dastoor, N. K. (1985). A classical approach to Cox's test for non-nested hypotheses. *Journal of Econometrics 27*, 363–370.

Davidson, R. and J. G. MacKinnon (1981). Several tests for model specification in presence of alternative hypotheses. *Econometrica 49*, 781–793.

Fisher, G. R. and M. McAleer (1981). Alternative procedures and associated tests of significance for non-nestedhypotheses. *Journal of Econometrics 16*, 103–119.

Godfrey, L. G. and M. H. Pesaran (1983). Tests of non-nested regression models: Small sample adjustments and MonteCarlo evidence. *Journal of Econometrics 21*, 133–154.

Gourieroux, C., A. Monfort, and A. Trognon (1983). Testing nested or non-nested hypotheses. *Journal of Econometrics 21*, 83–115.

Hall, A. (1985). A simplified method of calculating the distribution free Cox test. *Economics Letters 18*, 149–151.

Hampel, F. R. (1968). *Contribution to the Theory of Robust Estimation.* Ph. D. thesis, University of California, Berkeley.

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

Heritier, S. and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association 89*(427), 897–904.

Loh, W.-Y. (1985). A new method for testing separate families of hypotheses. *Journal of the American Statistical Association 80*, 362–368.

MacKinnon, J. G., H. White, and R. Davidson (1983). Tests for model specification in the presence of alternative hypotheses:some further results. *Journal of Econometrics 21*, 53–70.

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

Rousseeuw, P. J. and E. Ronchetti (1979). The influence curve for tests. Research Report 21, ETH Zürich, Switzerland.

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

Victoria-Feser, M.-P. (1995). Choosing between two parametric models robustly. Statistics Research Report LSERR21, London School of Economics, London WC2A 2AE.

Victoria-Feser, M.-P. and R. Ronchetti (1994). Robust methods for personal income distribution models. *The Canadian Journal of Statistics 22*, 247–258.

White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses. *Journal of Econometrics 19*, 301–318.

Williams, D. A. (1970). Discrimination between regression models to determine the pattern of enzymesynthesis in synchronous cell cultures. *Journal of Biometrics 28*, 23–32.

Amount of contamination	Classical statistic Nominal levels (in %) 1%	3%	5%	10%	Robust statistic Nominal levels (in %) 1%	3%	5%	10%
0%	2.1	3.1	3.5	5.2	1.3	3.5	5.5	10.2
3%	6.3	8.7	10.3	14.7	1.2	3.3	5.1	10.3
6%	13.1	18.5	22.5	27.6	1.4	3.6	5.4	10.7
10%	24.4	31.3	35.2	43.9	1.3	3.0	5.6	11.4
15%	35.6	44.6	49.9	58.1	1.4	4.1	7.9	14.5
20%	46.3	54.2	58.6	67.1	0.9	4.1	7.6	14.5

Table 1: **Actual levels (in %) of the classical and robust Atkinson statistic ($c = 2.0$) with contamination (Pareto against Exponential)**

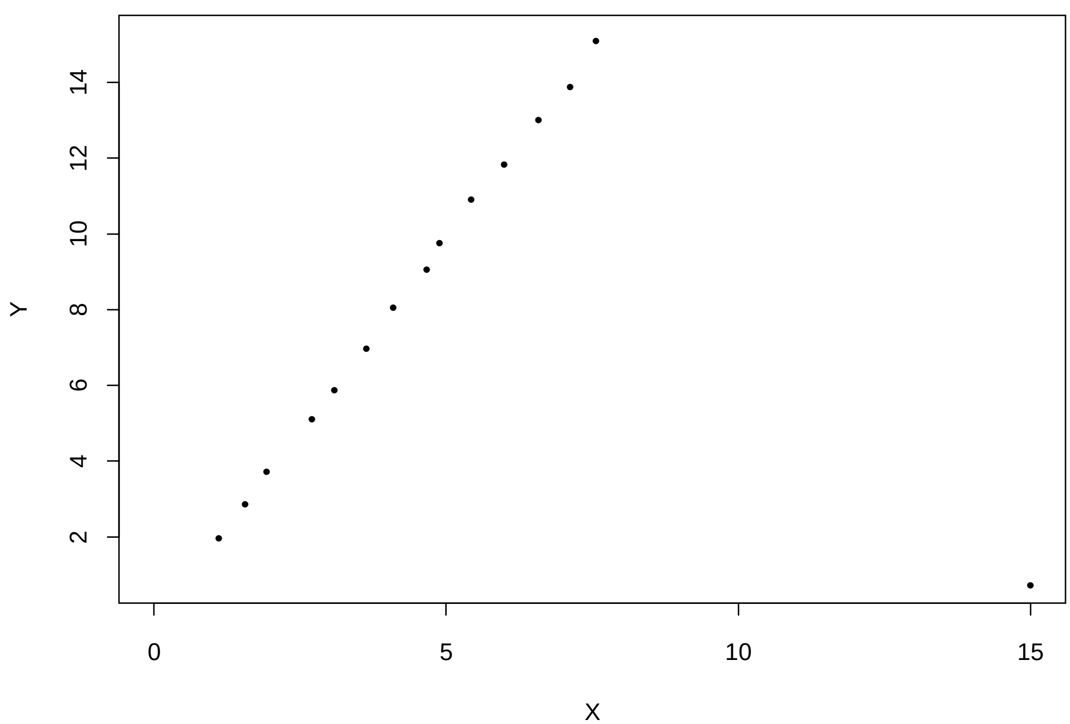

Figure 1: Plot of the response Y against regressor Z

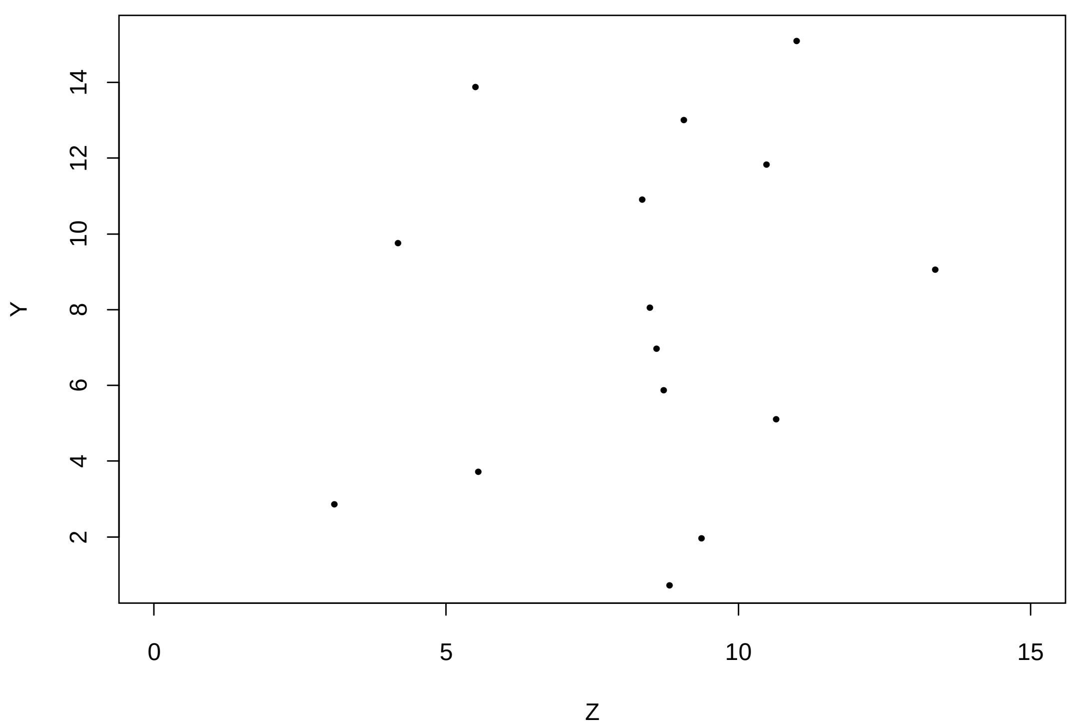

Figure 2: Plot of the response Y against regressor Z

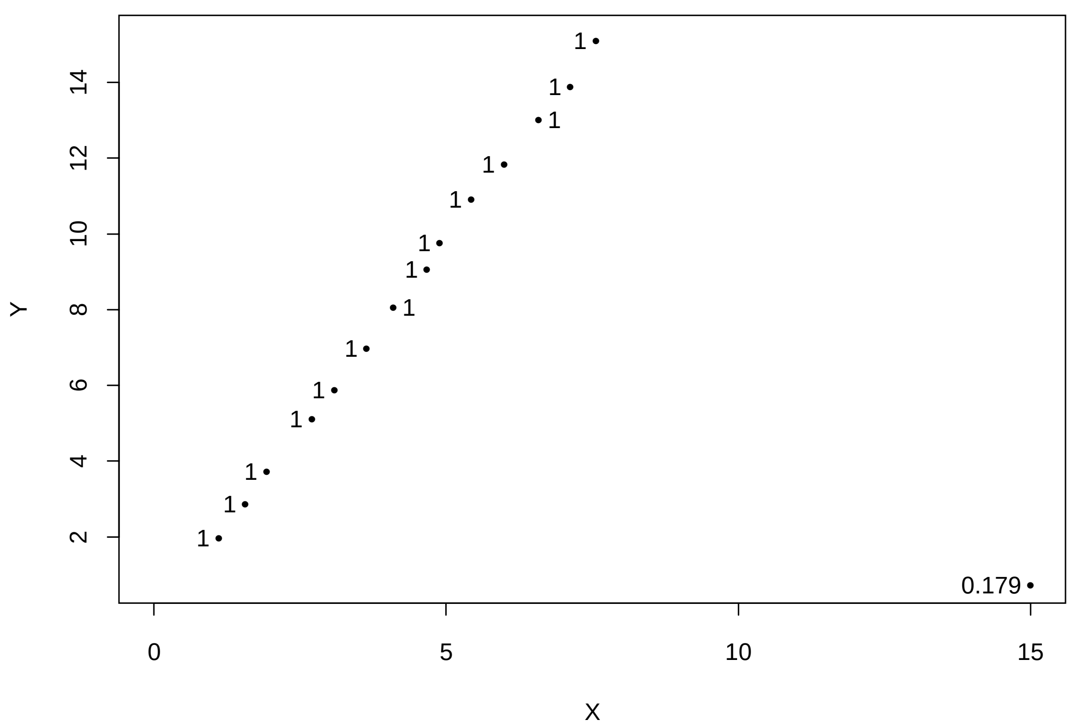

Figure 3: Plot of Y against X with weights computed by the robust test