



Department of Statistical Sciences
University of Padua
Italy

Robust inference in composite transformation models

Luca Greco

Department of Statistics
University of Venice
Italy

Laura Ventura

Department of Statistical Sciences
University of Padua
Italy

Abstract: The aim of this paper is to base robust inference about a shape parameter indexing a composite transformation model on a quasi-profile likelihood ratio test statistic. First, a general procedure is presented in order to construct a bounded profile estimating function for shape parameters. This method is based on a standard truncation argument from the theory of robustness. Hence, a quasi-likelihood test is derived. Numerical studies and applications to real data show that its use reveals extremely powerful, leading to improved inferences with respect to classical robust Wald and score-type test statistics.

Keywords: Bootstrap, Generalized profile likelihood, Quasi-profile likelihood, Unbiased estimating equations.

Contents

1	Introduction	1
2	Bounded profile estimating functions for shape parameters	3
3	Quasi-profile likelihood based inference	4

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

Corresponding author:
Luca Greco
tel: +39 041 234 7448
greco@dst.unive.it

tel: +39 049 8274168
fax: +39 049 8274170
<http://www.stat.unipd.it>

Robust inference in composite transformation models

Luca Greco

Department of Statistics
University of Venice
Italy

Laura Ventura

Department of Statistical Sciences
University of Padua
Italy

Abstract: The aim of this paper is to base robust inference about a shape parameter indexing a composite transformation model on a quasi-profile likelihood ratio test statistic. First, a general procedure is presented in order to construct a bounded profile estimating function for shape parameters. This method is based on a standard truncation argument from the theory of robustness. Hence, a quasi-likelihood test is derived. Numerical studies and applications to real data show that its use reveals extremely powerful, leading to improved inferences with respect to classical robust Wald and score-type test statistics.

Keywords: Bootstrap, Generalized profile likelihood, Quasi-profile likelihood, Unbiased estimating equations.

1 Introduction

Composite transformation models include parametric models of the form

$$F(y; \theta) = F(y; \mu, \sigma, \tau) = F_0\left(\frac{y - \mu}{\sigma}; \tau\right), \theta = (\mu, \sigma, \tau), \quad (1)$$

indexed by a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$ and a shape parameter $\tau \in T \subseteq \mathbb{R}$.

For instance, they are encountered in the setting of regression-scale and shape models, in which $\mu = x^\top \beta$, where $\beta \in \mathbb{R}^p$, $p > 1$, is a p -dimensional vector of regression coefficients and x is a vector of covariates. In these models, the error distribution is constructed by enlarging the normal model. Typically, this is achieved by adding a shape parameter τ . An attractive feature of these models is that, allowing τ to vary, they fit continuous variations from normality to non-normality. Hence, they offer the possibility of checking the assumption of normality through formal tests on the shape parameter.

Classical inference about τ in the presence of the nuisance parameter $\lambda = (\mu, \sigma)$, or $\lambda = (\beta, \sigma)$, is based on a pseudo-likelihood function, that is a function of y and

τ having properties similar to those of a genuine likelihood function when there is no nuisance parameter. The most commonly used pseudo-loglikelihood is the profile loglikelihood

$$\ell_p(\tau) = \ell(\tau, \hat{\lambda}_\tau) = \sum_{i=1}^n \ell(\tau, \hat{\lambda}_\tau; y_i) , \quad (2)$$

where $\ell(\theta) = \ell(\tau, \lambda)$ denotes the usual loglikelihood for θ and $\hat{\lambda}_\tau$ is the maximum likelihood estimate (MLE) of λ for fixed τ . Another solution is represented by the marginal likelihood (Severini, 2000).

It is well-known that standard likelihood procedures can be badly affected by data or model inadequacies. Mild deviations from $F(y; \theta)$ and from the related hypotheses can give rise to non negligible changes in inferential results. For instance, the maximum likelihood estimator becomes biased and inefficient under violations of the strict model assumptions.

Suppose that the model $F(y; \theta)$ does not describe exactly the reality, because of the occurrence of anomalous observations in the sample or the approximate character of the theoretical model itself. Robust statistics provides procedures that help to prevent the effects due to departures of the distribution of the data from the specified model and, at the same time, preserve good properties of efficiency and consistency when assumptions underlying the ideal model are relaxed (see Hampel *et al.*, 1986; Carroll and Ruppert, 1988; Markatou and Ronchetti, 1997). To this end, inferential procedures are studied in a neighborhood of the model, aiming at fitting only the majority of the data.

The theory of unbiased estimating equations supplies some effective tools for robust inference. An estimating equation is an equation in θ of the form $\Psi(\theta; y) = 0$, whose solution defines an estimate for θ . The function $\Psi(\theta; y)$ is called an estimating function. The estimating equation is said to be unbiased if $E_\theta(\Psi(\theta; Y)) = 0, \forall \theta \in \Theta$

Let $\tilde{\theta}$ be a robust estimator for θ , solution of the unbiased estimating equation

$$\Psi_\theta = \Psi(\theta; y) = \sum_{i=1}^n \psi(\theta; y) = 0 , \quad (3)$$

where $\psi(\cdot)$ is a suitable known real function. Estimators defined as solutions of such estimating equations are called M-estimators. Under broad conditions, $\tilde{\theta}$ is consistent and asymptotically normal with mean θ and variance

$$V(\theta) = M(\theta)^{-1} \Omega(\theta) (M(\theta)^{-1})^\top$$

, where $M(\theta) = E_\theta(\Psi_{\theta/\theta})$ and $\Omega(\theta) = E_\theta(\Psi_\theta \Psi_\theta^\top)$, and the symbol / as a subscript indicates differentiation. Moreover, $\tilde{\theta}$ is said B-robust at the assumed model F_θ if its influence function $IF(y, \tilde{\theta}, F_\theta) = M(\theta)^{-1} \Psi_\theta$ is bounded. It follows that $\tilde{\theta}$ is B-robust if and only if Ψ_θ is bounded.

Despite all the existing robust literature about inference on the whole parameter θ , the situation with a nuisance parameter has been somewhat neglected. Nuisance parameters, generally, are treated parallelly with the parameters of interest. No inferential techniques are considered to reduce or eliminate their effects. The aim of this paper is to extend the results on the treatment of nuisance parameters in a

natural fashion from the likelihood theory to the field of robust statistics. To this end, the profile approach provides a general way to eliminate nuisance parameters. In particular, the attempt is to make robust inference on the shape parameter τ in (1) in a different spirit from that illustrated in Hampel et al. (1986, section 4.4). Here, actually, the focus is on the use of bounded profile estimating functions and robust quasi-profile likelihood functions.

The remainder of the paper is organized as follows. In the next section a general method to find a bounded profile estimating function for the shape parameter τ is outlined. In section 3 a robust quasi-profile likelihood ratio test for inference on the shape parameter τ is obtained, starting from the proposed estimating functions, and some numerical studies and applications are presented.

2 Bounded profile estimating functions for shape parameters

Assume that, when θ splits up as $\theta = (\tau, \lambda)$, Ψ_θ is similarly partitioned as $(\Psi_\tau, \Psi_\lambda)$, where $\Psi_\tau = \Psi_\tau(\theta; y)$ and $\Psi_\lambda = \Psi_\lambda(\theta; y)$ are the estimating functions corresponding to τ and λ , respectively.

Paralleling classical likelihood inference, a profile-type estimating function for τ is

$$\tilde{\Psi}_\tau = \Psi(\tau, \tilde{\lambda}_\tau; y), \quad (4)$$

where $\tilde{\lambda}_\tau$ is an estimate of the nuisance parameter, evaluated for fixed τ , solution of an unbiased estimating equation $\Psi_\lambda = 0$. The corresponding estimating equation defines a B-robust estimator for τ if and only if both Ψ_τ and Ψ_λ are bounded.

Literature offers a wide variety of solutions for location and scale parameters but it is difficult to dispose of any bounded estimating function for the shape parameter in a composite transformation model. In order to obtain a bounded profile estimating function (4) for the shape parameter τ , when nuisance parameters can be eliminated by means of robust estimates of location and scale, we refer to a general approach described in Greco and Ventura (2006). It consists in bounding the influence of observation on a generalized profile score function by an appropriate downweighting of its components. The generalized profile score function is defined as the gradient with respect to τ of a generalized profile loglikelihood function $\ell(\tau, \tilde{\lambda}_\tau)$, in which the ordinary MLE $\hat{\lambda}_\tau$ is replaced by a consistent robust estimate $\tilde{\lambda}_\tau$ of the nuisance parameter (Severini, 1998).

A bounded profile estimating function for the shape parameter τ , in the presence of the nuisance λ , is given by

$$\tilde{\Psi}_\tau = \sum_{i=1}^n \psi_\tau(y_i; \tau, \tilde{\lambda}_\tau) = \sum_{i=1}^n h_i(\tau; c) \ell_\tau(\tau, \tilde{\lambda}_\tau; y_i), \quad (5)$$

where the weighting function $h(\tau; c)$ depends on y and on a tuning constant c which controls the balance between robustness and efficiency. The simplest choice for $h(\tau; c)$ may be similar to that involved in classical Huber-type estimation.

A robust estimate for τ is defined as the root of the estimating equation $\tilde{\Psi}_\tau = 0$. We call it robust generalized MLE (RGMLE). In general, to find the solution an iteratively algorithm must be used, in which weights are updated on each iterations based on the current value of τ .

Large-sample tests and confidence intervals for τ can be constructed following, for instance, the theory presented in Stefanski and Boos (2002) about M -estimators.

3 Quasi-profile likelihood based inference

Robust inference about the parameter of interest τ may be based on a robust pseudo-likelihood function, whose gradient is a bounded profile estimating function, derived as in (5). In this way, on the one hand, it is possible to obtain a likelihood ratio-type test statistic whose asymptotic distribution is the same as in the classical framework; on the other hand we may improve robust inference on the parameter of interest with respect to the use of Wald-type and score-type tests. Hence, the availability of such a likelihood ratio-type test with the usual limiting behavior is an important achievement. Then, the possibility of representing graphically the robust pseudo-likelihood can give more information about the structure of the bulk of the data.

A possible solution in this direction is provided by the quasi-profile loglikelihood function (Adimari and Ventura, 2002). This quasi-profile loglikelihood function is based on a suitable scale adjustment $\omega(\cdot)$ to a profile estimating function $\Psi_\tau(\tau, \tilde{\lambda}_\tau)$, leading to a rescaled function

$$\tilde{\Psi}_\tau^\dagger(\tau, \tilde{\lambda}_\tau; y) = \omega(\tau, \tilde{\lambda}_\tau) \tilde{\Psi}_\tau(\tau, \tilde{\lambda}_\tau; y) ,$$

such that its variance is equal to minus the expected derivative matrix, when expectations and derivatives are computed at $(\tau, \tilde{\lambda}_\tau)$ (McCullagh and Tibshirani, 1990).

Actually, the scale adjustment $\omega(\cdot)$ is obtained by requiring that

$$\text{Var}_{(\tau, \tilde{\lambda}_\tau)} \tilde{\Psi}_\tau^\dagger(\tau) = -\mathbf{E}_{(\tau, \tilde{\lambda}_\tau)} \left\{ \frac{\partial \tilde{\Psi}_\tau^\dagger}{\partial \tau}(\tau) \right\} .$$

Solving for $\omega(\tau, \tilde{\lambda}_\tau)$, it turns out that

$$\omega(\tau, \tilde{\lambda}_\tau) = \frac{\left(\frac{\partial}{\partial \tau} \mathbf{E}_{(\tau, \tilde{\lambda}_\tau)}(\tilde{\Psi}_\tau) - \mathbf{E}_{(\tau, \tilde{\lambda}_\tau)}(\tilde{\Psi}_{\tau\tau}) \right)}{\text{Var}_{(\tau, \tilde{\lambda}_\tau)}(\tilde{\Psi}_\tau)} . \quad (6)$$

A first-order approximation to (6) can be computed by first-order asymptotic expressions for the moments of the derivatives of the estimating function for τ (Adimari and Ventura, 2002). Since the calculation of the analytical approximation can be burdensome, it may be preferable to resort to a simulation process in which the information bias of $\tilde{\Psi}_\tau$ is estimated by parametric bootstrap. Anyway, even when first order approximations are available, bootstrap provides an exact adjustment which leads to improved inference.

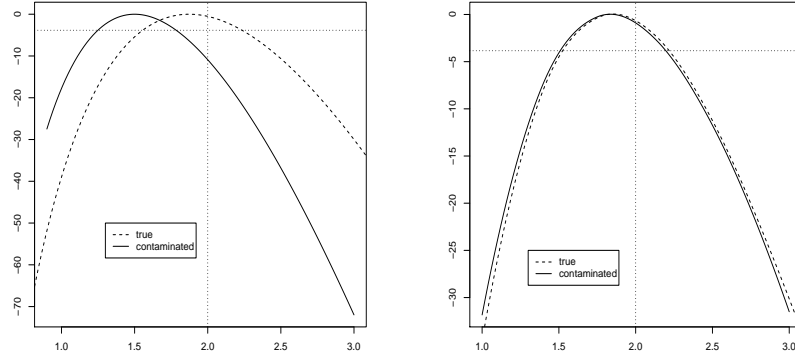


Figure 1: Gamma distribution: simulated sample of size $n = 200$. W_P (left) and W_{QP}^b (right) for the shape parameter (the horizontal dotted line gives the 0.95 confidence interval).

It is worth noting, then, that by multiplying $\tilde{\Psi}_\tau$ by the factor $\omega(\tau, \lambda)$, the induced estimator for τ does not change, as well as its asymptotic behavior and its robustness properties.

The quasi-profile loglikelihood function is given by

$$\ell_{QP}(\tau) = \int^\tau \omega(t, \tilde{\lambda}_t) \Psi_\tau(t, \tilde{\lambda}_t) dt . \quad (7)$$

Function (7) has properties similar to (2). In particular, for setting confidence intervals or for testing hypothesis, it is possible to derive a quasi-likelihood ratio statistic

$$W_{QP}(\tau) = 2\{\ell_{QP}(\tilde{\tau}) - \ell_{QP}(\tau)\} . \quad (8)$$

Under the null hypothesis and usual regularity conditions, (8) is approximately χ_1^2 distributed, as in the classical framework. For instance, confidence intervals for τ with nominal level $1 - \alpha$ can be constructed as $\{\tau : W_{QP}(\tau) \leq \chi_{1;1-\alpha}^2\}$, where $\chi_{1;1-\alpha}^2$ is the $(1 - \alpha)$ -quantile of the χ_1^2 distribution.

In practice, the scale adjustment is necessary to obtain quasi-profile likelihood ratio tests based on (8) with the χ_1^2 asymptotic distribution, as in the classical framework. In fact, if this term is neglected, (8) does not present the standard asymptotic behavior unlike its classical counterpart (Heritier and Ronchetti, 1994; Hanfelt and Liang, 1995).

To illustrate the use of (8) for robust inference on the shape parameter in composite transformation models, some applications are provided with regard to some well known models. The behavior of (8) is compared with that of Wald-type test statistics $W_e(\tau)$ and score-type test statistics $W_u(\tau)$, as defined in Heritier and Ronchetti (1994) and with that of the OBRE.

Example 1: Gamma distribution. Let Y_i be n independent Gamma random variables with density $f(y_i; \tau, \lambda) = y_i^{\tau-1} e^{-y_i/\lambda} / (\lambda^\tau \Gamma(\tau))$, $i = 1 \dots n$. This is a composite scale model. Suppose the shape parameter τ is the parameter of interest and

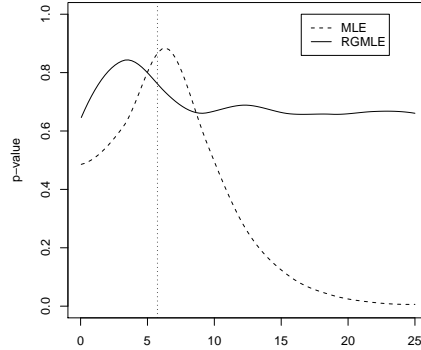


Figure 2: Simulated sample of size $n = 20$ from a Gamma random variable with shape $\tau = 2$ and scale $\sigma = 1$: sensitivity curves of the p-values for W_P and W_{QP} for testing the null hypothesis $H_0 : \tau = 2$.

		$n = 200$			$n = 100$		
		$1 - \alpha$			$1 - \alpha$		
		0.990	0.950	0.900	0.990	0.950	0.900
g1	OBRE (4.5)	0.981	0.934	0.882	0.984	0.936	0.884
	$W_{QP}(\tau)$	0.991	0.956	0.909	0.991	0.952	0.907
	$W_e(\tau)$	0.979	0.954	0.912	0.996	0.969	0.927
	$W_u(\tau)$	0.972	0.923	0.874	0.972	0.927	0.871
g2	OBRE (4.5)	0.930	0.887	0.761	0.929	0.835	0.766
	$W_{QP}(\tau)$	0.989	0.947	0.893	0.989	0.948	0.907
	$W_e(\tau)$	0.972	0.913	0.857	0.982	0.935	0.884
	$W_u(\tau)$	0.984	0.959	0.926	0.992	0.967	0.930

Table 1: Gamma distribution: empirical coverage probabilities of the confidence intervals for the shape parameter based on W_{QP} , W_e , W_u and the OBRE.

that inference is expected to be done by taking into account the possibility of mild contaminations. To eliminate the scale nuisance parameter λ , we use the MAD-type estimate $\tilde{\lambda}_\tau = \text{median}(y_i)/F^{-1}(0.5; \tau)$, which is Fisher consistent at the Gamma model for τ considered as fixed. The quasi-profile likelihood ratio test is obtained from the estimating function $\tilde{\Psi}_\tau = \sum_{i=1}^n h_i(\tau; c)\tilde{\ell}_{p^*}(\tau; y_i)$ defining the RGMLE $\tilde{\tau}_{RG}$ for τ , with $c = 1.5$.

A sample of size $n = 200$ has been generated from a Gamma random variable with shape parameter $\tau = 2$ and scale $\lambda = 1$. Then a contamination has been introduced on the right tail by replacing the five largest observations by even larger values.

Figure 1 shows the behavior of the ordinary profile likelihood ratio test $W_P(\tau) = 2\{\ell_P(\hat{\tau}) - \ell_P(\tau)\}$ and of the quasi-profile likelihood ratio test $W_{QP}(\tau)$ under both scenarios. The quasi-profile likelihood ratio test has been computed by 400 bootstrap samples. The likelihood ratio test statistic $W_P(\tau)$ shifts remarkably, whereas it does

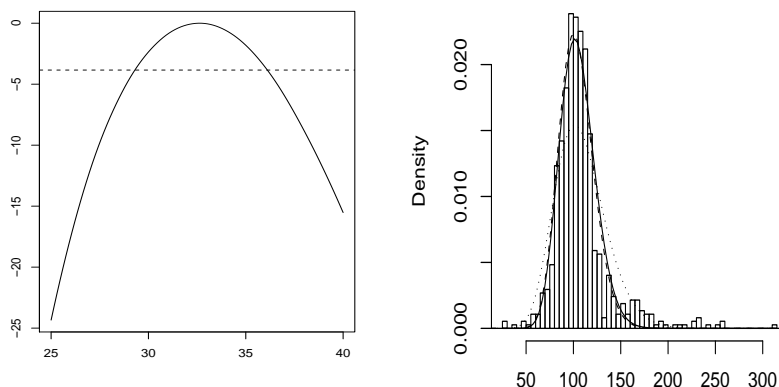


Figure 3: Left: Quasi-likelihood ratio test for the shape parameter (the horizontal dashed line gives the 0.95 confidence interval) for household income data. Right: histogram with estimated Gamma distribution by RGMLE (solid line), the OBRE (dashed line) and the MLE (dotted line).

not occur for $W_{QP}(\tau)$. Moreover, $W_{QP}(\tau)$ is very close to $W_P(\tau)$ under the true model. Note that the 0.95-level confidence interval for τ based on $W_P(\tau)$ under the contaminated sample does not include the true value of the parameter.

The robustness of $W_{QP}(\tau)$ can also be assessed by means of an empirical sensitivity analysis. We use a simulated sample of size $n = 20$. The value in the sample corresponding to the vertical dotted line in Figure 2 is perturbed and allowed to vary in the range $[0.05, 25]$. At each time $W_P(\tau_0)$ and $W_{QP}(\tau_0)$ for testing $H_0 : \tau = \tau_0$, where τ_0 is the true parameter value, are recomputed.

Figure 2 displays the behavior of the p-value associated to both test statistics. It is evident that $W_P(\tau_0)$ appears sensitive to outlying observations, whereas the p-value associated to $W_{QP}(\tau_0)$ is more stable, at the cost of less evidence for the null hypothesis in a small region around the original value.

A simulation experiment, based on 4000 Monte Carlo trials, has also been performed in order to evaluate the empirical coverages of the nominal $1 - \alpha$ confidence intervals for the shape parameter obtained by $W_{QP}(\tau)$. Data have been generated from a Gamma model with $\tau = 2$ and $\lambda = 1$ (g1) and allowing a mild contamination on the right tail, i.e. data have been generated with probability 0.02 from a Gamma distribution with scale equals to 5 (g2). The simulation results are given in Table 1 for nominal levels 0.990, 0.950 and 0.900. The quasi-profile likelihood ratio test $W_{QP}(\tau)$ performs well both under the true model and under the contaminated model. Indeed, it improves over Wald and score-type tests in terms of empirical coverage probabilities and highlights a stronger robustness, in the sense that empirical coverages are less affected by the contamination. Table 1 also reports the empirical coverage probabilities of the Wald-type confidence intervals obtained from the computation of the OBRE, with the scale parameter not treated as nuisance but estimated simultaneously with the shape. Setting $c = 4.5$, the obtained empirical coverage probabilities are clearly improved by the quasi-likelihood approach.

Example 2: Household incomes data. Consider the empirical distribution of

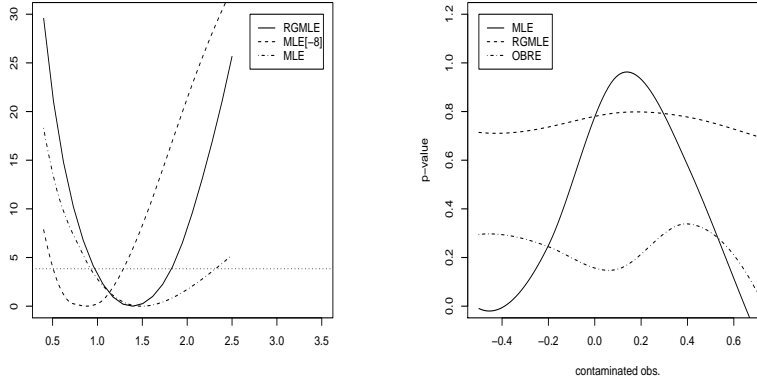


Figure 4: Left: Quasi-likelihood ratio test for the shape parameter κ for Martin Marietta data (the horizontal dotted line gives the 0.95 confidence interval). Right: sensitivity analysis for the p-value of the classical likelihood test, Wald test from OBRE and quasi-likelihood test from RGMLE.

household incomes in 1979 in UK. The aim is to fit a Gamma distribution to the majority of the data, so that inference is not influenced by extreme observations in the tails (Victoria-Feser and Ronchetti, 1994). Assume the interest concerns inference on the shape parameter. To this end we use the estimating function defining the RGMLE. The weighting function used to bound the generalised profile score function for the parameter of interest is chosen so that more importance is given to the most frequent observations, located in the centre of the distribution.

In Figure 3, $W_{QP}(\tau)$ (left) and the histogram (right) of the empirical distribution and the estimated Gamma distribution by the solution to the set of equations defining the RGMLE (solid line), the OBRE (dashed line) and MLE (dotted line) are plotted. The estimated curve according to the MLE tends to be influenced by extreme observations in the tails whereas the distributions estimated by RGMLE and OBRE catch the inequality structure of the majority of the data.

Example 3: Martin-Marietta data. Let us consider a data set concerning monthly excess return for a given security (x) and excess return on the market portfolio (y) on the New York Stock Exchange for the company Martin Marietta (Butler *et al.*, 1990). Here the focus is on a regression-scale and shape model. Actually, it is assumed that a simple linear model holds, with error distributed as an EP with shape parameter τ . Suppose the interest is on quasi-likelihood confidence intervals for the shape parameter τ , based on the estimating function $\tilde{\Psi}_\tau = \sum_{i=1}^n h_i(\tau; c) \tilde{\ell}_{p_*}(\tau; y_i)$, with $c = 2.5$. Figure 5 (left) displays the behavior of $W_{QP}(\tau)$, $W_P(\tau)$ and $W_P(\tau)$ when observation 8, which is the most influencing observation, is deleted. The quasi-profile likelihood ratio test has been evaluated by 500 bootstrap samples. At a significance level of 0.05, $W_{QP}(\tau)$ rejects the hypothesis that the shape parameter assumes values $\tau \leq 1$ and the value $\tau = 1$ even falls outside the 0.95-level likeli-

hood based confidence interval obtained after deleting observation 8. The 0.95-level confidence interval for τ obtained after having computed the OBRE with $c = 2$ is [0.95–1.57]. A sensitivity analysis has also been performed to assess the stability of the p-values associated to $W_{QP}(\tau)$ when the null hypothesis is $H_0 : \tau = 1.5$. Observation 8 is perturbed and allowed to vary in the range [-0.5 – 0.7]. It is evident from Figure 4 (right) that the p-value from the χ_1^2 asymptotic approximation for $W_{QP}(\tau)$ is stable and the evidence for the null hypothesis is always stronger than that obtained from the Wald-type test computed from the OBRE.

References

- Adimari, G., Ventura, L. (2002). Quasi-profile loglikelihood for unbiased estimating functions. *Ann. Inst. Statist. Math.*, **54**, 235–244.
- Butler, R.J., McDonald, J.B., Nelson, R.D. and White, S.B., Robust and partially adaptive estimation of regression models, *The Review of Economics and Statistics*, **72**, (1990) 321–327.
- Carroll, R.J., Ruppert, D., *Transformation and Weighting in Regression* (Chapman and Hall, London, 1988).
- Greco, L., Ventura, L. (2006). Bounded Estimation in the Presence of Nuisance Parameters. *Statistical Methods & Applications*, to appear.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., *Robust Statistics: The Approach Based on Influence Functions* (Wiley, New York, 1986).
- Hanfelt, J. J., Liang, K. Y. (1995). Approximate likelihood ratios for general estimating functions. *Biometrika*, **82**, 461–477.
- Heritier, S., Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.*, **89**, 897–904.
- Markatou, M., Ronchetti, E., Robust inference: The approach based on influence functions, In: *Handbook of Statistics*, Maddala, G.S., Rao, C.R. (Eds.), **Vol. 15**, North-Holland, Amsterdam, (1997) 49–75.
- McCullagh, P., Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. R. Statist. Soc. B* **52**, 325–344.
- Severini, T.A. , Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika* **85**, (1998) 507–522.
- Severini, T. A. (2000). *Likelihoods Methods in Statistics*. Oxford University Press, Oxford.
- Stefanski, L.A. and Boos, D.D., The calculus of M-estimation, *The American Statistician*, **56**, (2002) 29–38.
- Victoria-Feser, M. and Ronchetti, E., Robust methods for personal income models, *Canadian Journal of Statistics*, **22**, (1994) 247–258.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>

