Stable Asymptotics for M-estimators

La Vecchia, Davide

Stable asymptotics for M-estimators

Davide La Vecchia*

Summary

We review some first- and higher-order asymptotic techniques for M-estimators and we study their stability in the presence of data contaminations. We show that the estimating function (ψ) and its derivative with respect to the parameter ($\nabla_{\theta^\top}\psi$) play a central role. We discuss in detail the first-order Gaussian density approximation, saddlepoint density approximation, saddlepoint test, tail area approximation via Lugannani-Rice formula, and empirical saddlepoint density approximation (a technique related to the empirical likelihood method). For all these asymptotics, we show that a bounded (in the Euclidean norm) ψ and a bounded (e.g., in the Frobenius norm) $\nabla_{\theta^\top}\psi$ yield stable inference in the presence of data contamination. We motivate and illustrate our findings by theoretical and numerical examples about the benchmark case of one-dimensional location model.

MSC: 62E17, 62E20, 62F05, 62F12, 62F35, G2G35

Key words: Edgeworth expansion, Empirical likelihood, Higher-order, Infinitesimal robustness, p-value, Redescending M- estimator, Relative error, Saddlepoint techniques, von Mises expansion.

1 Introduction

In the classical statistical approach, data are assumed to be a realization of a reference model. Then, inference is usually based on asymptotic techniques, whose purpose is twofold. First, from a practical standpoint, asymptotics define an approximate behavior of statistical quantities (e.g., estimators and/or tests) and allow the implementation of inferential procedures. Second,

*Monash University, Econometrics and Business Statistics Department, Melbourne, Australia. School of Economics and Political Science, Institute of Mathematics and Statistics, University of St. Gallen, Sankt Gallen, Switzerland. Email: davide.la-vecchia@monash.edu

asymptotic approximations can be applied theoretically to study the quality of the analyzed procedures (e.g., the efficiency of estimates and/or the power of tests). See, e.g., van der Vaart (1998) and reference therein for a book-length discussion.

For instance, assume we have to derive a confidence interval for the parameter in a one-dimensional location model. To achieve the goal, we need to know the distribution of the estimator. This quantity is seldom known exactly (e.g., this is the case of estimators which are linear function, like the sample mean, of the underlying Gamma observations), and more often only approximations are available. Typically, one defines a first-order approximation, based on a linearization of the estimator. Then, the behavior of the linearized statistic is studied, as the sample size (say n) goes to infinity. This leads, through the central limit theorem, to many asymptotic normality proofs and the resulting first-order asymptotic distribution can be used as an approximation to the exact distribution of the estimator. However, very often the accuracy of the first-order asymptotic distribution deteriorates quickly in small samples. Moreover, the Gaussian approximation tends to be inaccurate in the tails of the distribution.

To cope with the low accuracy of the firs-order asymptotic (Gaussian) approximation, several techniques have been developed to achieve higher-order accuracy: many distributional approximations that improve on the first-order asymptotic theory are nowadays available. These techniques are called *higher-order asymptotics*, since they are obtained considering higher-order (beyond the first linear term) expansions of the statistics, and they are accurate, even for $n = 1$.

As emphasized by Young (2009), the higher-order techniques are defined by two main approaches: (i) analytic methods based on small sample asymptotics (including, e.g., Edgeworth and saddlepoint techniques) and (ii) resampling methods (including, e.g., jackknife and bootstrap). The focus of this paper is on small sample techniques, but we remark that the small sample asymptotics and resampling methods have to be considered as complementary tools (rather than alternative); see, e.g., Davison and Hinkley (1988), Hall (1992), and Davison et al. (1995).

First- and higher-order asymptotic techniques represent helpful and powerful tools when the actual data distribution follows exactly the reference model; see Example 1. Unfortunately, see Example 2, many statistical procedures, which work well under the ideal model assumptions, can lose their stability in the presence of moderate deviations from the theoretical model — we call stable any statistical procedure such that a minor error in the reference model causes only limited error in the conclusions. This aspect is the key point of the paper and we state the following

> Research question (RQ): *"What is the behavior of the asymptotics when the actual data distribution slightly deviates (in some distributional sense) from the reference model?"*

The RQ is related to the robustness of the considered statistical procedures; see, e.g., Huber (1981), Hampel et al. (1986), and Rieder (1994). We here focus on statistical procedures for M-estimators, which are defined by an estimating function depending on the data and on the unknown parameter of interest.

To answer our RQ, we build on La Vecchia et al. (2012), who analyze the stability of some asymptotics of M-estimators, extending and adapting the analysis of La Vecchia et al. to other (higher-order) asymptotics.

The main contribution of this paper is to show that a bounded estimating function having bounded derivative with respect to (henceforth, wrt) the parameter ensures the stability of many asymptotics. Specifically, we study the stability of the: (i) saddlepoint density approximation and tail area of a real-valued function of the parameter; (ii) saddlepoint test of Robinson et al. (2003), and, for the benchmark case of location, we define a new robust test which guarantees the stability of the p-value in the presence of contamination; (iii) empirical saddlepoint approximation; (iv) empirical likelihood method (which is connected to the empirical saddlepoint). Finally, we illustrate numerically the role of the estimating function in the stability of the first-order asymptotics and empirical saddlepoint density approximation, in the one-dimensional location problem.

The message of our paper is in line with the robustness approach of Hampel et al. (1986): the characteristics (like, e.g., upper and lower bound, shape, and derivative wrt the parameter) of the estimating function determine the stability of the asymptotics.

The paper has the following structure. In §2, we provide some motivating examples. In §3 we define the setting and introduce the notation. In §4, we discuss the stability of the first- and higher-order asymptotics. In §5, we provide some numerical studies. In §6, we conclude and highlight some possible research topics. Technical details and proofs are available in Appendix.

2 Motivating examples

2.1 Edgeworth vs. Saddlepoint vs. Gaussian asymptotics

In the family of higher-order techniques, the Edgeworth and the saddlepoint approximations are the most popular methods.

One might try to improve on the first-order approximation by using the first few terms of an Edgeworth expansion (see, e.g., van der Vaart (1998), Ch. 23), which is a series in powers of $n^{-1/2}$, where the first term is the Gaussian density. In general, the Edgeworth expansion provides a good approximation in the center of the density, but can be inaccurate in the tails, where it can even become negative. Saddlepoint techniques overcome the problems of the Edgeworth expansions.

Since the seminal paper of Daniels (1954), saddlepoint expansions have been effectively applied to define confidence intervals (Tingley and Field (1990)), test statistics (Robinson et al. (2003)), and to approximate p-values in several inferential problems. Roughly speaking, in the Edgeworth techniques a higher-order approximation around the center of the distribution is applied. In the saddlepoint expansion, the higher-order approximation is replaced by a sequence of local low-order approximations; see, e.g., Goutis and Casella (1999) for an overview. Thanks to this approach, excellent small sample performances are obtained. See, among the others, Field and Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

Among the different higher-order techniques, this paper deals with saddlepoint techniques. Our choice is motivated by both theoretical and practical reasons. On a theoretical standpoint, the saddlepoint techniques exhibit relative error $O(n^{-1})$, which has to be compared to the absolute error $O(n^{-1/2})$ obtained using the Edgeworth type expansions. This implies, for instance, that the saddlepoint techniques provide very accurate numerical approximations for tail areas down to small sample sizes and/or out in the tail; see Example 1. On a more practical note, we remark that the excellent small sample behaviour is also due to the fact that saddlepoint approximations are density-like objects and do not show the polynomial-like waves exhibited by Edgeworth approximations. This simplify the implementation and implies that the saddlepoint density approximations cannot be negative.

Example 1. We illustrate the use of different asymptotics for the approximation of the density of the mean ($\bar{Z} = \sum_{i=1}^{n} Z_i/n$) of n independent and identically distributed random variables $Z_i \sim \chi^2(df)$, where df represents the degrees-of-freedom. We make a comparison using different sample sizes and different values of df. The true distribution of the mean is known in closed form. For each approximation and in each considered setting, we compute the percentage relative error, expressed as 100*(1-approximation/true). We consider three approximations: (i) first-order Gaussian asymptotic; (ii) Edgeworth; (iii) saddlepoint. In Figure 1 we plot the results. As highlighted by an anonymous Referee, in the considered case, the saddlepoint approximation is exact (see Kolassa (2006), page 68), namely it features zero error. The Edgeworth performs well near the center of the density, but it has large errors in the tails (especially in the left tail). Finally, the error entailed by the first-order asymptotic Gaussian is much bigger than the errors of the higher-order approximations.

2.2 Stability of the saddlepoint test

In the next example we illustrate the problem stated in our RQ, investigating the stability of the saddlepoint test introduced by Robinson et al. (2003); details are available in §4.

Example 2. Let us consider the mean of n independent and identically distributed random variables $Z_i \sim \chi^2(df)$. Along the lines of Example 1 of Robinson et al. (2003), to test the hypothesis $\mathcal{H}_0 : df_0 = 1$, we make use of the saddlepoint test statistic $2nh(\bar{Z})$, where $h(\bar{Z}) = (\bar{Z} - 1) - \log(\bar{Z})$. Under $\mathcal{H}_0$ the test has a $\chi^2(1)$ distribution and it has relative error $O(n^{-1})$. This yields accurate level, even for small sample sizes. However, in the presence of departures

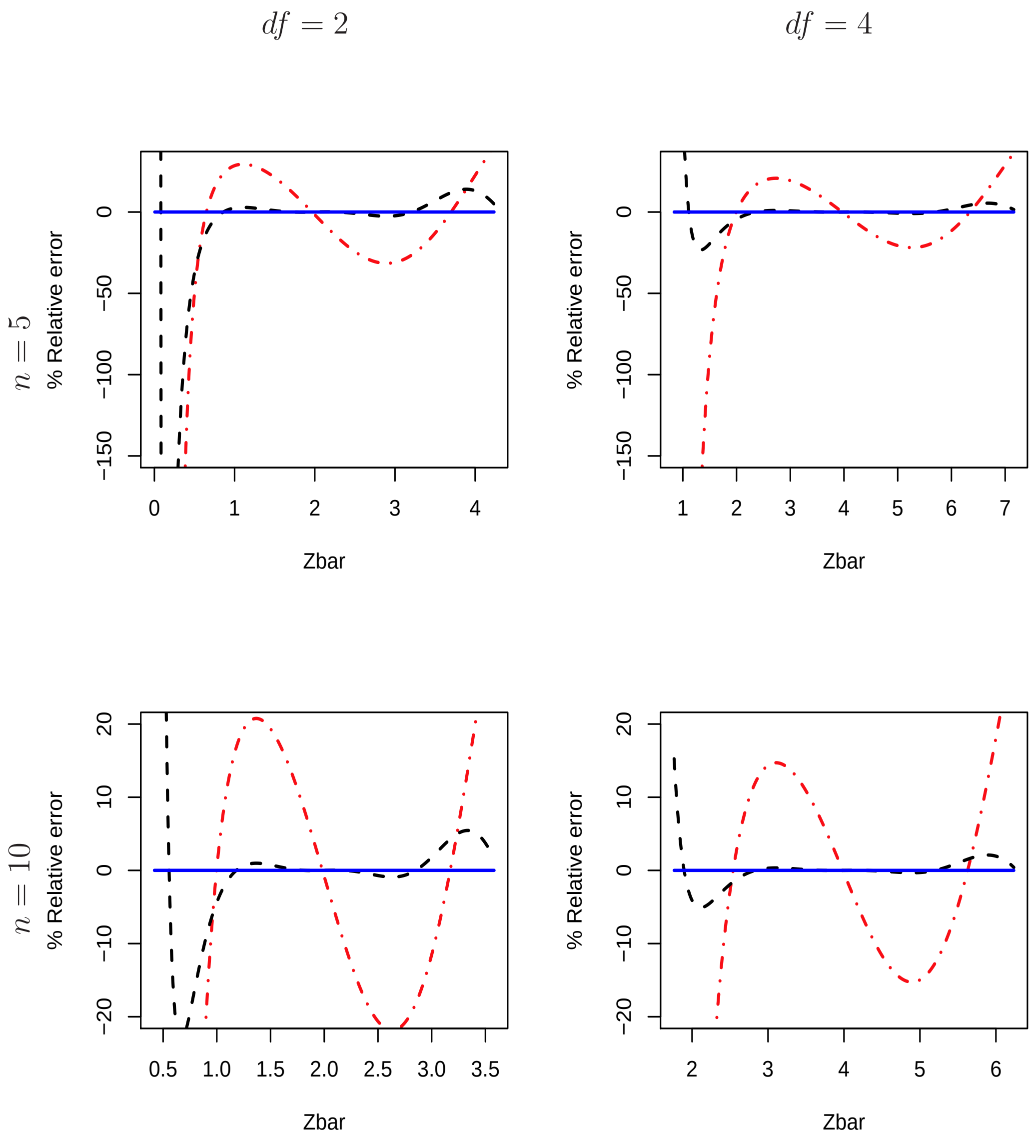

Figure 1: Percentage relative error of first and higher-order asymptotics, for the approximation of density of the mean ($\bar{Z}$) of n independent and identically distributed $Z_i \sim \chi^2(df)$ variables, for different sample sizes ($n = 5$ and $n = 10$, top and bottom panels, respectively) and different degrees-of-freedom ($df = 2$ and $df = 4$, left and right panels, respectively). In each plot: dash-dotted line represents the first-order Gaussian asymptotic, dashed line represents the Edgeworth, while continuous line is for the saddlepoint.

from the reference theoretical model, the level of the test can be distorted. As a numerical illustration, consider the case where we have $n = 30$ and two kinds of sample: (i) a clean sample, containing draws from a $\chi^2(1)$; (ii) a contaminated sample, in which we replace three observations of the clean sample by outliers having value 4 – this kind of sample represents a situation where the actual data distribution slightly deviates from the reference model. Then, we analyze the distribution of the test statistic, under clean and contaminated sample, in a Monte Carlo experiment having size 2,500. Figure 2 shows that the test has the desired theoretical accuracy for the clean sample. In contrast, in the presence of contamination, the sample mean is inflated by the outliers and the distribution of the test statistic is far from the theoretical one.

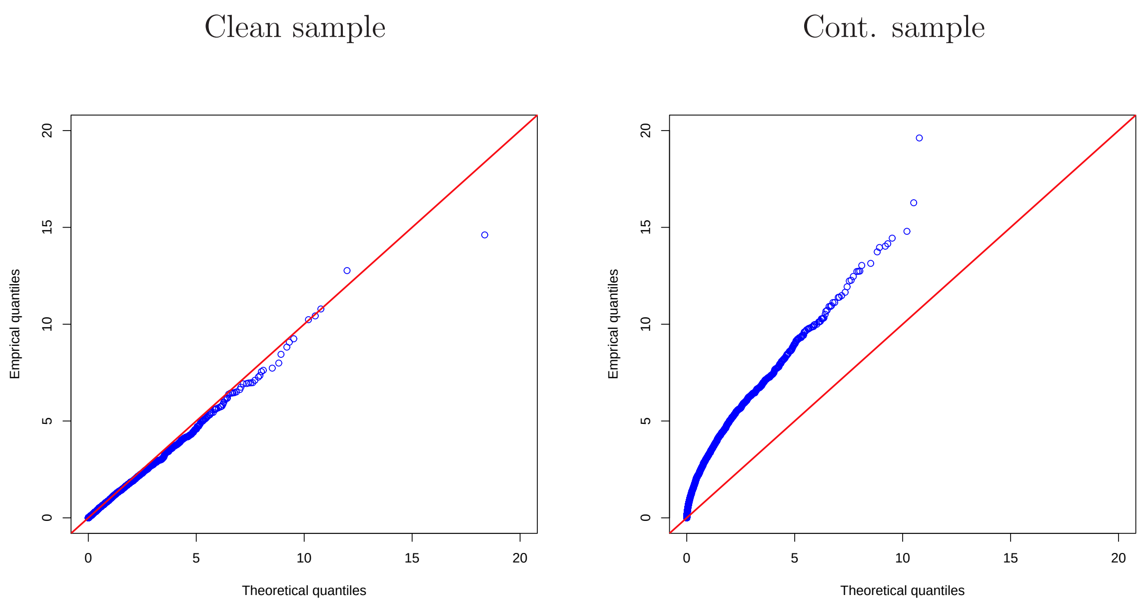

Figure 2: Saddlepoint test (distribution under the null) for the mean of n independent and identically distributed $Z_i \sim \chi^2(df)$ for the hypothesis $\mathcal{H}_0 : df_0 = 1$, based on the statistic $h(\bar{Z}) = (\bar{Z} - 1) - \log(\bar{Z})$.

2.3 Asymptotic bias in the linear model

To give a partial answer to our RQ, we analyze the impact that small deviations from the reference model can have on the asymptotic bias of different M-estimators of the parameters in a linear regression model — the asymptotic bias can be interpreted as a measure of the change in the performance of the estimator due to the presence of data contamination; a precise definition is available in §4. The aim of the example is to illustrate (only) one aspect related to the estimating function and its derivative wrt the parameter.

Example 3. Let us consider the simple regression model $Y = \theta_1 + \theta_2 X + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ and $\theta = (\theta_1, \theta_2)$. We generate a contaminated sample with size $n = 200$, as obtained introducing a small number outliers in the (x, y)-space (namely, we have outliers in both the response and the explanatory variable) — details about the simulation design are available in the caption of Figure 3. Then, we consider three estimators: the maximum likelihood (defined by an unbounded estimating function with unbounded derivative wrt θ), the Hampel-Krasker (defined by a bounded estimating function with unbounded derivative wrt θ, see Hampel et al. (1986)), and the second-order robust estimator (defined by a bounded estimating function and a bounded derivative wrt θ, see La Vecchia et al. (2012)). The functional form of these estimators is available in Appendix A, to which we refer for technical details. We here flag an important aspect: the maximum likelihood estimator is designed to perform well under the reference model, but it does not control the impact of outliers, while both the robust estimators, by design, limit the impact of anomalous observations. More precisely, the Hampel-Krasker estimator controls the impact of contaminated data on the first-order term of the asymptotic bias expansion, while the second order robust estimator controls both the first- and second-order terms. In Figure 3 we plot the results. The stability of the second order robust estimator (rightmost panel) is evident: the fitted regression line in the presence of contamination (continuous line) is almost indistinguishable from the theoretical reference line (dotted line), while the other estimators are more biased (with the Hampel-Krasker estimator less affected by the contamination).

The asymptotic bias analyzed in Example 3 represents only one characteristic (it is related to the location) of the distribution of the estimator in the presence of departures form the reference model. In the next sections, our aim is to gain further insights into other characteristics (e.g., the scale and more generally the whole distribution), in large and small samples.

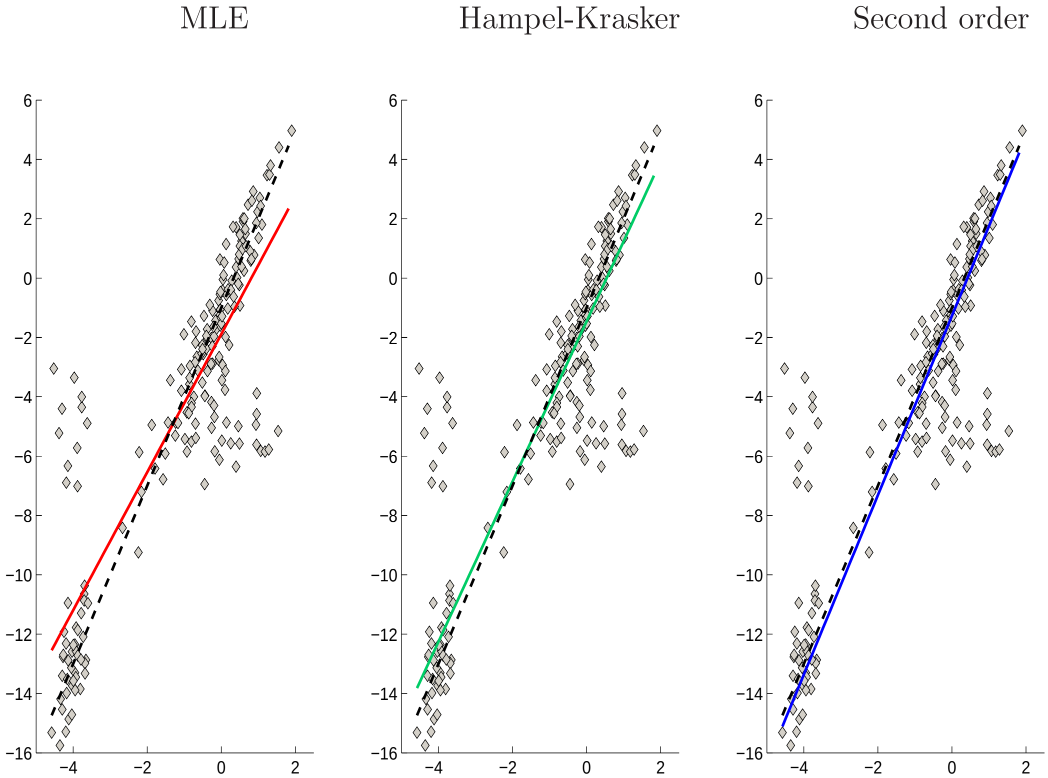

Figure 3: Linear model: Clean sample, generated accordingly to the regression model, with $\theta = (-1, 3)'$ and $\epsilon \sim \mathcal{N}(0, 1)$. Contaminated sample, generated by: replacing, with probability 5%, the original observations by $Z_y \sim \mathcal{N}(-5, 1)$ and using contaminated explanatory variable, as obtained replacing, with probability 25%, the original explanatory variable ($X \sim \mathcal{N}(0, 1)$) by $Z_x \sim \mathcal{N}(-4, 0.1)$. In each plot: diamonds represent the observations, the dashed line represents the theoretical regression line, the continuous line represents the fitted line. The tuning parameters for the robust estimators are: $b = 4.5$ for Hampel-Krasker, $b = 3.5$ and $c = 0.5$ for the second order robust estimator. With these tuning constants, the two robust estimators have a very similar performance in the clean sample.

3 Setting and notation

Let $\mathcal{M}$ be the family of all probability measures on $\mathcal{Z} \subset \mathbb{R}^m$, for $m \geq 1$, and let $T : \mathrm{dom(T)} \to \mathbb{R}^p$ be a statistical functional, defined on $\mathrm{dom(T)} \subset \mathcal{M}$ and taking values in $\mathbb{R}^p$, with $p \geq 1$. For $P \in \mathcal{M}$, the functional value $T(P)$ can represent any characteristic of P, e.g., the location, the scale, a tail area or, more generally, a quantity depending on P.

In the standard parametric approach, Z is assumed to have distribution function described by the parametric family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, where Θ is an open convex subset of $\mathbb{R}^p$, $p \geq 1$. A Fisher consistent M-functional $T(\cdot)$ is implicitly defined as the unique functional root of the system of p moment conditions:

$$E_{P_{\theta_0}}[\psi(Z; T(P_{\theta_0}))] = 0,$$

where $\psi : \mathbb{R}^m \times \Theta \to \mathbb{R}^p$ is the estimating function, $T(P_{\theta_0}) = \theta_0$ and $\theta_0 \in \Theta$.

Among the statistical functionals, the M-functionals play a central role for inferential purpose: given n independent and identically distributed (i.i.d.) observations $Z_1, Z_2, ..., Z_n$, an M-estimator is the sample counterpart of the M-functional. Precisely, an M-estimator is the implicit solution of the finite sample equations

$$\sum_{i=1}^{n} \psi(Z_i; \hat{\theta}_n) = 0,$$

where $\hat{\theta}_n = T(P_n) = T_n$, and P_n represents the empirical measure.

Infinitesimal robustness (see, e.g., Hampel et al. (1986) for a book-length introduction) weakens the parametric assumption $Z_i \sim P_{\theta_0}$, and allows the distribution of the actual data generating process to belong to the neighborhood:

$$\mathcal{U}_\eta(P_{\theta_0}) = \{P_{\epsilon,G} = (1-\epsilon)P_{\theta_0} + \epsilon G, \text{ for } G \in \mathcal{M}\} \;\; ; \;\; \theta_0 \in \Theta, \tag{1}$$

where $\epsilon \ll 0.5$, and $\mathcal{M}$ is the class of all measures on $\mathbb{R}^m$. Then, the stability over $\mathcal{U}_\eta(P_{\theta_0})$ of statistical functionals is analyzed. In the following, we call *stable* any statistical functional $T(\cdot)$ such that small deviations from P_{θ_0} imply bounded changes in the value of $T(\cdot)$. More precisely, for any $P_{\epsilon,G} \in \mathcal{U}_\eta(P_{\theta_0})$, such that $d_K(P_{\theta_0}; P_{\epsilon,G}) \leq \epsilon$ (d_K represents the Kolmogorv distance), we have $\|T(P_{\epsilon,G}) - T(P_{\theta_0})\| < \epsilon \cdot \text{const}$, namely the changes of the functional remain bounded by a constant.

4 Robustness and asymptotics

4.1 Main result via von Mises expansion

Our analysis of the behavior of an M-estimator over $\mathcal{U}_\eta(P_{\theta_0})$ makes use of the von Mises expansion (see von Mises (1947)), which can be interpreted as a kind of sophisticated Taylor expansion for statistical functionals.

The first (linear) order term of this expansion approximates the asymptotic bias:

$$Bias(\epsilon, G, P) = T(P_{\epsilon,G}) - \theta_0 = \epsilon T'(\epsilon, G) + Rem_1, \tag{2}$$

where $T'(\epsilon, G) = \lim_{\epsilon \to 0} \left[T(P_{\epsilon,G}) - \theta_0\right]/\epsilon$ is expressed by the Influence Function (IF), defined as:

$$T'(\epsilon, G) = \int_{\mathcal{Z}} IF(z; P_{\theta_0}) G(dz) \quad \text{where} \quad IF(z; P_{\theta_0}) = M^{-1}(\theta_0; P_{\theta_0})\psi(z; \theta_0). \tag{3}$$

and $M(\theta_0; P_{\theta_0}) = -E_{P_{\theta_0}}[\nabla_{\theta'}\psi(Z; \theta_0)]$.

Remark. Roughly speaking, we can interpret $T(P_{\epsilon,G})$ as the value of the estimator in the presence of contamination, while $T(P_{\theta_0})$ is the value of the estimator at the reference model. With this interpretation in mind and looking at (2), the IF represents a first-order measure of the effects that small perturbations have on the asymptotic bias. Moreover, (3) implies that IF $\propto \psi$: if $\sup_z \|\psi(z; \theta_0)\|$ is bounded (here $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^m$), then from (2) it follows that the M-estimator is B-robust, namely it has a bounded first-order asymptotic bias.

La Vecchia et al. (2012) investigate the robustness features of the second order (quadratic) term in the von Mises expansion, which reads as:

$$Bias^{(2)}(\epsilon, G, P) = T(P_{\epsilon,G}) - \theta_0 = \epsilon T'(\epsilon, G) + \frac{\epsilon^2}{2} T''(\epsilon, G) + Rem_2, \tag{4}$$

where

$$T''(\epsilon, G) \;=\; \int_{\mathcal{Z}\times\mathcal{Z}} \varphi_2(z_1, z_2; \theta_0) d(G - P)(z_1) d(G - P)(z_2).$$

The second-order kernel $\varphi_2(z_1, z_2; \theta_0)$ is defined in Fernholz (2001):

$$\begin{aligned} \varphi_2(z_1, z_2; \theta_0) &= IF(z_1; P_{\theta_0}) + IF(z_2; P_{\theta_0}) + M^{-1}(\psi; \theta_0)\Gamma(z_1, z_2; \theta_0) \\ &\quad + M^{-1}(\psi; \theta_0)\left\{\nabla_{\theta^\top}\psi(z_2; \theta_0)\varphi_1(z_1; \theta_0) + \nabla_{\theta^\top}\psi(z_1; \theta_0) IF(z_2; P_{\theta_0})\right\}, \end{aligned} \tag{5}$$

where

$$\Gamma(z_1, z_2; \theta_0)^\top = \left\{ \begin{array}{c} IF^\top(z_2; P_{\theta_0}) E_{P_{\theta_0}}\left(\frac{\partial^2}{\partial\theta\partial\theta^\top}\psi^{(1)}(Z; \theta_0)\right) IF(z_1; P_{\theta_0}) \\ \vdots \\ IF^\top(z_2; P_{\theta_0}) E_{P_{\theta_0}}\left(\frac{\partial^2}{\partial\theta\partial\theta^\top}\psi^{(p)}(Z; \theta_0)\right) IF(z_1; P_{\theta_0}) \end{array} \right\}, \tag{6}$$

where $\psi^{(j)}$ is the j-th component of the vector ψ.

The expression in (5) suggests that an estimating function ψ such that

$$\sup_z \|\psi(z; \theta_0)\| < const \quad \text{and} \quad \sup_z \|\nabla_{\theta^\top}\psi(z; \theta_0)\| < const$$

has bounded IF and bounded φ_2. Thus, from (4) it follows that $Bias^{(2)}(\epsilon, G, P)$ is uniformly bounded over $\mathcal{U}_\eta(P_{\theta_0})$. La Vecchia et al. (2012) label as second-order B-robust the M-estimators defined by such a class of estimating functions.

The next proposition characterizes the stability of two important quantities which are related to ψ and $\nabla_{\theta^\top}\psi$ and which are needed to define several first- and higher-order asymptotics of M-functionals — for instance, the expectation of $\nabla_{\theta^\top}\psi$ is a fundamental quantity in the first-order asymptotic variance of the M-estimator.

Proposition 1 *Let $Z_i \sim P_{\epsilon,G}$ and $\psi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^p$ be an estimating function satisfying:*

[A1] ψ is bounded, namely $\sup_z \|\psi_j(z; \theta_0)\| < const$, for $j = 1, ..., p$;

[A2] ψ has a bounded $\nabla_{\theta^\top}\psi$, namely $\sup_z \|\partial_{\theta_j}\psi_i(z; \theta_0)\| < const$, for $0 \leq i, j \leq p$,

where $\|\cdot\|$ *is the Euclidean norm. Then for* $\theta \in \Theta$

$$M(\theta; P_{\epsilon,G}) = E_{P_{\epsilon,G}}[-\nabla_{\theta^\top}\psi(Z; T(P_{\epsilon,G}))] \quad \textit{and} \quad \Sigma(\theta; P_{\epsilon,G}) = E_{P_{\epsilon,G}}[\psi(Z; T(P_{\epsilon,G}))\psi^\top(Z; T(P_{\epsilon,G}))] \tag{7}$$

are stable over $\mathcal{U}_\eta(P_{\theta_0})$*, namely:*

$$\sup_{G\in\mathcal{M}} \|\partial_\epsilon M_{ij}(\theta; P_{\epsilon,G})|_{\epsilon=0}\| < const \qquad \textit{and} \qquad \sup_{G\in\mathcal{M}} \|\partial_\epsilon \Sigma_{ij}(\theta; P_{\epsilon,G})|_{\epsilon=0}\| < const. \tag{8}$$

Interpretation. As in the case of the IF, the directional derivatives (∂_ϵ at $\epsilon = 0$) in (8) are tools which measure the changes that the expected value of $\nabla_{\theta^\top}\psi$ and the variance of ψ have in the presence of small perturbations of the reference model. Boundedness of these derivatives implies that the considered quantities undergo bounded changes when the actual data distribution slightly deviates from the reference model. The conditions in Assumptions [A1] and [A2] guarantee that such a boundedness is uniform over the whole $\mathcal{U}_\eta(P_{\theta_0})$. Thus, the message of (8) is clear: there is no distribution in $\mathcal{U}_\eta(P_{\theta_0})$ which can inflate or deflate M and Σ too much. Similar interpretation holds for the derivatives of the other quantities that we are going to consider in the next sections (see, e.g., Corollary 2 and Proposition 3).

4.2 First-order asymptotics

4.2.1 Stability in the presence of contamination

Suppose we are interested in conducting inference about a parameter $\eta_0 = j(\theta_0) \in \mathbb{R}^q$, where $q \leq p$ and $j(\cdot)$ is a known function. Computing $\hat{\eta} = j(\hat{\theta}_n)$ is the natural way to draw such an inference. The main issue with this approach is that the exact distribution of $\hat{\theta}_n$ is typically available only in few special cases, and in general situations one has to rely on an asymptotic approximation (e.g., via Δ-method), whose first-order (linear) term has an asymptotic Gaussian distribution (see, e.g., van der Vaart (1998), Chapter 3). Detailed conditions to ensure the asymptotic normality of general M-estimators are given in Appendix.

When $P_{\theta_0} \equiv P_{\epsilon,G}$, namely the reference model coincides with the actual distribution, the IF is useful to express the first-order asymptotic distribution of $\hat{\theta}_n$. For n sufficiently large, the empirical measure P_n is near to P_{θ_0} and (2) yields

$$T(P_n) - T(P_{\theta_0}) = (\hat{\theta}_n - \theta_0) = \frac{1}{n}\sum_{i=1}^{n} IF(z_i; P_{\theta_0}) + Rem_1. \tag{9}$$

This is the standard asymptotic result, where $P_n \Rightarrow P_{\theta_0}$ and (9) implies that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, V(\theta_0; P_{\theta_0})),$$

where

$$V(\theta_0; P_{\theta_0}) = M^{-1}(\theta_0; P_{\theta_0})\Sigma(\theta_0; P_{\theta_0})M^{-\top}(\theta_0; P_{\theta_0}), \tag{10}$$

and $\Sigma(\theta_0; P_{\theta_0}) = E_{P_{\theta_0}}[\psi(Z; P_{\theta_0})\psi^\top(Z; P_{\theta_0})]$.

When $P_{\epsilon,G} \neq P_{\theta_0}$, namely the distribution of the actual data generating process does not coincide with the reference model, the features of ψ affect the location/scatter of the asymptotic distribution. In this regard, Proposition 1 has implications on the stability of the first-order asymptotics of M-functionals and we state the following:

Corollary 2 *Let $Z_i \sim P_{\epsilon,G}$ and $\psi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^p$ be an estimating function defining an M-functional, with $M(\theta; P_{\epsilon,G})$ and $\Sigma(\theta; P_{\epsilon,G})$ as in (7). Under [A1] and [A2] of Proposition 1, the asymptotic variance functional at $P_{\epsilon,G}$ defined as:*

$$V(\theta; P_{\epsilon,G}) = M(\theta; P_{\epsilon,G})^{-1}\Sigma(\theta; P_{\epsilon,G})M(\theta; P_{\epsilon,G})^{-\top},$$

remain stable over $\mathcal{U}_\eta(P_{\theta_0})$.

Remark. Corollary 2 is related to V-robustness, see Hampel et al. (1986), which characterizes the stability of the asymptotic variance. Specifically, V-robust M-estimators have an asymptotic variance which neither shrinks to zero nor explodes over the neighborhood $\mathcal{U}_\eta(P_{\theta_0})$; see, e.g, Ferrari and La Vecchia (2012) and reference therein for a related discussion.

4.2.2 Discussion

Many M-estimators are defined by an unbounded ψ, having also unbounded $\nabla_{\theta^\top}\psi$. In the presence of data contamination, these characteristics entail: (i) an asymptotic bias which can be arbitrarily large (wrong location in the asymptotic distribution); (ii) an unstable asymptotic variance functional (wrong scatter in the asymptotic distribution).

First-order B-robust M-estimators limit the impact of the contamination, since they are defined by a bounded ψ, which yields a better control of the asymptotic bias. Nevertheless, Proposition 1 implies that there is no guarantee that also $\nabla_{\theta^\top}\psi$ remains bounded and Corollary 2 implies that the asymptotic variance can be inflated by outliers.

Second-order B-robust M-estimators have bounded ψ and bounded $\nabla_{\theta^\top}\psi$ and they feature V-robustness. Thus, Proposition 1 and Corollary 2 imply stable location and scatter in the asymptotic distribution. Moreover, an application of the chain rule ensures that the directional derivatives (∂_ϵ) of other asymptotic functionals related to the first-order asymptotic distribution (e.g., confidence intervals, score-type tests, and p-values) remain stable as well; see, e.g., Heritier and Ronchetti (1994).

4.3 Higher-order asymptotics

This section discusses the implications that Proposition 1 has on saddlepoint-type asymptotics, for some model-based and model-free techniques.

4.3.1 Model-based techniques

Saddlepoint density approximation. Theorem 4.5 in Field and Ronchetti (1990) shows that, for every distribution in $P_{\epsilon,G} \in \mathcal{U}_\eta(P_{\theta_0})$ satisfying the assumptions given in Appendix, the exact finite sample density $f(t; n, \varepsilon, G)$ of the M-functional can be approximated by a saddlepoint expansion of the form $f(t; n, \varepsilon, G) = g(t; n, \varepsilon, G)\{1 + O(n^{-1})\}$, for $t \in \mathbb{R}^p$. The saddlepoint density approximation is

$$g(t; n, \varepsilon, G) = (n/2\pi)^{p/2} c^{-n}(\alpha(t; P_{\varepsilon,G}); P_{\varepsilon,G}) \left|\det \tilde{M}(t; P_{\varepsilon,G})\right| \left|\det \tilde{\Sigma}(t; P_{\varepsilon,G})\right|^{-1/2}, \qquad (11)$$

where the vector $\alpha(t; P_{\varepsilon,G})$ is the saddlepoint, i.e., the solution of the saddlepoint equation $E_{h_{t,P_{\epsilon,G}}}[\psi(Z;t)] = 0$, while

$$\tilde{M}(t; P_{\epsilon,G}) = \left\{ E_{h_{t,P_{\epsilon,G}}} \left[- \left. \frac{\partial \psi_j(Z;t)}{\partial t_r} \right|_t \right] \right\}_{1 \le r,j \le p}, \tag{12}$$

and

$$\tilde{\Sigma}(t; P_{\epsilon,G}) = \left\{ E_{h_{t,P_{\epsilon,G}}} [\psi_j(Z;t)\psi_r(Z;t)] \right\}_{1 \le r,j \le p}. \tag{13}$$

In (12) and (13), we have

$$h_{t,P_{\epsilon,G}}(z) = c\{\alpha(t; P_{\epsilon,G}); P_{\epsilon,G}\} \exp\{\alpha^\top(t; P_{\epsilon,G})\psi(z;t)\},$$

where $h_{t,P_{\epsilon,G}}(z)$ is the so-called conjugate density and

$$c^{-1}\{\alpha(t; P_{\epsilon,G}); P_{\epsilon,G}\} = E_{P_{\epsilon,G}} \left[\exp\{\alpha^\top(t; P_{\epsilon,G})\psi(z;t)\} \right].$$

The detailed construction of the saddlepoint density is available in Field (1982) and needs a delicate use of the complex analysis. We here provide just the intuition of the whole procedure, highlighting three main steps. Assume we have to approximate the density of the estimator at a fix point θ. In the first step (also called Esscher's tilting), we re-center (by means of the conjugate density and of the saddlepoint) the unknown density of the estimator at θ. In the second step, we use (locally at θ) a low-order Edgeworth expansion (very accurate because it is at the center of the density) to approximate the tilted density. In the third step, we tilt back the resulting approximate density, relating it to the desired density. The whole procedure is repeated for each point $\theta \in \Theta$. The unusual characteristic of the resulting saddlepoint expansion is that the first few terms (or even just the first Gaussian term in the local Edgeworth expansion) often give very accurate approximations in the far tails of the distribution, even for small sample sizes.

To illustrate the role of ψ and $\nabla_{\theta^\top}\psi$ in the saddlepoint density approximation, let us assume that the actual data distribution is $P_{\epsilon,G} \in \mathcal{U}_\eta(P_{\theta_0})$. Under $P_{\epsilon,G}$, the exact density of the M-estimator is $f(t; n, \epsilon, G)$. In general, $P_{\epsilon,G}$ is unknown so that a saddlepoint of the density under the reference model P_{θ_0} (say, $g(t;n,P_{\theta_0})$) is applied. For a given $G \in \mathcal{M}$, the difference between $f(t;n,\epsilon,G)$ and $g(t;n,P_{\theta_0})$ gives the total error due to the saddlepoint approximation: $\text{err}_{tot}(t;n,\epsilon,G,P_{\theta_0}) = f(t;n,\epsilon,G) - g(t;n,P_{\theta_0})$. Following Ronchetti and Ventura (2001), we decompose this error as

$$\text{err}_{tot}(t;n,\epsilon,G,P_{\theta_0}) = \text{err}_{sad}(n) + \text{err}_{dev}(t;\epsilon,G,P_{\theta_0}),$$

where

$$\text{err}_{sad}(n) = f(t;n,\epsilon,G) - g(t;n,\varepsilon,G) = O(n^{-1}),$$

and

$$\text{err}_{dev}(t;\epsilon,G,P_{\theta_0}) = g(t;n,\varepsilon,G) - g(t;n,P_{\theta_0}). \tag{14}$$

The $\text{err}_{sad}(n)$ is related only to the sample size. Differently, $\text{err}_{dev}(t;\epsilon,G,P_{\theta_0})$ expresses the impact that the deviations from the theoretical model have on the saddlepoint approximation. This error can be unbounded over $\mathcal{U}_\eta(P_{\theta_0})$ unless the estimating function ψ is bounded and has bounded $\nabla_{\theta^\top}\psi$. The claim can be justified by considering that the saddlepoint approximation $g(t;n,\epsilon,G)$ depends upon the estimating function and its derivatives wrt to the parameter via $\tilde{M}$ and $\tilde{\Sigma}$ in (12) and (13), respectively. Thus, an unbounded ψ and/or an unbounded $\nabla_{\theta^\top}\psi$ implies an unbounded err_{dev}. More formally we state

Proposition 3 (La Vecchia et al. (2012), Proposition 3) *Let $\psi(z;\theta_0)$ be an estimating function satisfying the conditions in Proposition 1, then*

$$\sup_G |\partial_\varepsilon g(t;n,\varepsilon,G)|_{\varepsilon=0}| < \infty$$

Proposition 3 implies an additional property of second-order B-robust M-estimators: their saddlepoint density approximation (and its relative error of order $O(n^{-1})$, see La Vecchia et al. (2012)) undergo bounded changes in the presence of contamination.

Saddlepoint density and tail area approximations for a real valued function of the parameter. Gatto and Ronchetti (1996) define a saddlepoint approximation for the finite sample distribution of a real valued function $j(\cdot)$ of $T(P_n)$ and of $T(P_{\theta_0})$; for instance, $j(\cdot)$ can be just a component of $T(\cdot)$.

Gatto and Ronchetti (1996) show that the saddlepoint approximation to the density of $j\{T(P_n)\} - j\{T(P_{\theta_0})\}$ is:

$$p_n(t;\theta_0) = \left\{\frac{n}{2\pi \tilde{R}_n''(\alpha)}\right\}^{1/2} \exp\left\{n(\tilde{R}_n(\alpha) - \alpha t)\right\}. \tag{15}$$

In (15), $\tilde{R}_n(\alpha)$ is an approximate cumulant generating function of ψ and it reads as:

$$\tilde{R}_n(\alpha) = \mu_n \alpha + \frac{1}{2} n \sigma_n^2 \alpha^2 + \frac{1}{6} n^2 \kappa_{3n} \sigma_n^3 \alpha^3 + \frac{1}{24} n^3 \kappa_{4n} \sigma_n^4 \alpha^4$$

where α is the saddlepoint defined by $\tilde{R}_n'(\alpha) = v$, while $\mu_n, \sigma_n, \kappa_{3n}, \kappa_{4n}$ depend on the IF and φ_2. Moreover, Lugannani-Rice type formula yields the tail probability approximation:

$$P\left(j\{T(P_n)\} - j\{T(P_{\theta_0})\} > t\right) = 1 - \Phi(r) + \phi(r)\left(\frac{1}{s} - \frac{1}{r}\right), \tag{16}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and probability density function of the standard Normal, respectively, $s = \alpha\left\{\tilde{R}_n''(\alpha)\right\}^{1/2}$, $r = \text{sgn}(\alpha)\left\{2n\left(\alpha t - \tilde{R}_n(\alpha)\right)\right\}$.

Considerations analogous to the ones made for the saddlepoint density approximation in §4.3.1 lead to the conclusion that Gatto and Ronchetti's procedure is not stable over $\mathcal{U}_\eta(P_{\theta_0})$ when ψ does not satisfy [A1] and [A2] of Proposition 1. Indeed, (3) and (5) imply that a bounded ψ, having also a bounded $\nabla_{\theta^\top}\psi$, yields uniformly bounded IF and φ_2, which in turn yield stable $\tilde{R}_n(\alpha)$. Thus, a straightforward application of the chain rule implies that second-order B-robust M-estimators have stable saddlepoint density approximation and stable Lugannani-Rice tail area approximation.

Saddlepoint test. Let us consider the case of simple hypothesis $\mathcal{H}_0 : \theta = \theta_0 \in \mathbb{R}^p$ – the case of composite hypothesis is analyzed in the same way, but entails a cumbersome notation.

Robinson et al. (2003) propose the saddlepoint test statistic $2nh(T_n)$, where

$$h(t) = \sup_{\alpha}\{-K_\psi(\alpha, t)\} = -K_\psi(\alpha(t), t) \tag{17}$$

is the Legendre transform of $K_\psi(\alpha, t) = \log E_{P_{\theta_0}}[\exp \alpha^T \psi(Z, t)]$.

Under the null, the test has a $\chi^2(p)$ distribution and, if ψ is the likelihood score, it is asymptotically equivalent to the three classical tests: likelihood ratio, Wald, and score test. However, the test has relative error $O(n^{-1})$ which has to be compared to the absolute error of order $O(n^{-1/2})$ of the classical tests.

Robinson et al. (2003) define an approximation to p-value $= P_{\theta_0}(h(T_n) > h(t_n))$, where t_n is the observed value of T_n. In particular, Robinson et al. start from the saddlepoint density approximation and they integrate it (using a Laplace technique, see, e.g., Barndorff-Nielsen and Cox (1989)), evaluating

$$p\text{-value} = \int_{\tilde{A}} (n/2\pi)^{p/2} c^{-n}(\alpha(t; P_{\theta_0}); P_{\theta_0}) \left|\det \tilde{M}(t; P_{\theta_0})\right| \left|\det \tilde{\Sigma}(t; P_{\theta_0})\right|^{-1/2} \left(1 + O(n^{-1})\right) dt \tag{18}$$

over the area $\tilde{A} = \{t : h(t) \geq h(t_n)\}$.

A robust saddlepoint test has been derived in Lô and Ronchetti (2009) for the case of generalized linear model (GLM), in the presence of contamination in the response variable (y-space). However, Lô and Ronchetti's test is based on a first-order B-robust M-estimator. Similarly to the Hampel-Krasker estimator in the linear model of Example 3 in §2, the first-order B-robust GLM M-estimator ensures a bounded $\tilde{\Sigma}$, but it does not guarantee a uniformly bounded $\tilde{M}$. As a result, the stability of the p-value over $\mathcal{U}_\eta(P_{\theta_0})$ is still an open question.

We gain further insights on this question, considering the effect of data contamination on the integrand in (18). To this end, we need to evaluate the p-value assuming that the actual data distribution is $P_{\epsilon,G}$ (see Example 2). Within this setting, $\tilde{M}(t; P_{\varepsilon,G})$ and $\tilde{\Sigma}(t; P_{\varepsilon,G})$ are the key players. Once again, a joint application of the chain rule, in tandem with Proposition 1 and Proposition 3, show that model deviations can make the integrand arbitrarily large over $\mathcal{U}_\eta(P_{\theta_0})$, if ψ and/or $\nabla_{\theta^\top}\psi$ are unbounded. As a consequence, the p-value can be unstable. Clearly, tests based on second-order robust M-estimators guarantee the stability of the p-value.

In the next example, we illustrate the construction of a saddlepoint test, which makes use of a second order B-robust M-estimator. Differently from the results available in the literature (which rely on first-order B-robust M-estimators), our test is defined by an estimating function which imposes bounds on both $\tilde{M}$ and $\tilde{\Sigma}$. For illustrative purposes, we define the test for a well-known benchmark: the one-dimensional location problem.

Example 4 (stable saddlepoint test for univariate location model). Let $Z \sim \phi(z-\theta)$, where ϕ is the standard normal density and $\theta \in \Theta \subset \mathbb{R}$. For $b \in \mathbb{R}^+$, we set

$$\tilde{\psi}_{b,c}(z-\theta) = \begin{cases} c(z-\theta)\min\left(1; \frac{b}{|Ac(z-\theta)|}\right) & c \in (0,1) \\ (z-\theta)\min\left(1; \frac{b}{|A(z-\theta)|}\right) & c \geq 1. \end{cases} \tag{19}$$

Assume we have to test $\mathcal{H}_0 : \theta = 0$. For $c \geq 1$, the test based on $\tilde{\psi}_{b,c}$ is first-order equivalent to the robust tests already discussed in Hampel et al. (1986), Ch 3, and it has the same spirit as the test in Lô and Ronchetti (2009). For $c \in (0,1)$, we here flag the possibility to apply $\tilde{\psi}_{b,c}$ in the definition of a new saddlepoint test, which controls the bound on $\tilde{M}$.

The definition of the test via (17) requires the computation of the cumulant generating function of $\tilde{\psi}_{b,c}$ at the saddlepoint. To illustrate the procedure, we set $A = 1$, just for the sake of simplicity. Thus, the estimating function reads as:

$$\tilde{\psi}_{b,c}(z-\theta) = \begin{cases} -b & \mathcal{Z}_1 \\ c(z-\theta) & \mathcal{Z}_2 \\ b & \mathcal{Z}_3 \end{cases} \tag{20}$$

where $\mathcal{Z}_1 = \{z \in \mathbb{R} : -b > c(z-\theta)\}$, $\mathcal{Z}_3 = \{z \in \mathbb{R} : b < c(z-\theta)\}$, and $\mathcal{Z}_2 = \mathbb{R} \setminus (\mathcal{Z}_1 \sqcup \mathcal{Z}_3)$.

For $t \in \mathbb{R}$, the cumulant generating function $K_{\psi_{b,c}}(\alpha, t)$ under $\mathcal{H}_0$ is obtained as:

$$\begin{aligned} \log\left[\int_{\mathbb{R}} \exp\left\{\alpha\tilde{\psi}_{b,c}(z-t)\right\}\phi(dz)\right] &= \log\left[\int_{\mathcal{Z}_1} \exp\{-\alpha b\}\phi(dz) + \int_{\mathcal{Z}_2} \exp\{\alpha c(z-t)\}\phi(dz)\right. \\ &\left. + \int_{\mathcal{Z}_3} \exp\{\alpha b\}\phi(dz)\right] = \log\left[I_1 + I_2 + I_3\right]. \end{aligned}$$

After some algebra, we get:

$$I_1 = \Pr(\mathcal{Z}_1) \exp\{-\alpha b\}, \quad I_2 = \Pr(\mathcal{Z}_2) \exp\left\{-\alpha c \left(t + \tfrac{\alpha c}{2}\right)\right\}, \quad I_3 = \Pr(\mathcal{Z}_3) \exp\{\alpha b\},$$

where each $\Pr(\cdot)$ is obtained using the cumulative distribution function of the standard Normal – for instance, $\Pr(\mathcal{Z}_1) = \Phi(t - b/c)$.

Once the cumulant generating function is available, the saddlepoint is the solution to

$$\frac{\partial K_{\psi_{b,c}}(\alpha, t)}{\partial \alpha} = \frac{\partial_\alpha I_1 + \partial_\alpha I_2 + \partial_\alpha I_3}{I_1 + I_2 + I_3} = 0,$$

which has to be solved numerically; see, e.g., Kolassa (2006), section 4.3, for some hints based on Newton-Raphson or secant method, and/or on rescaling the cumulant generating function to make it closer to a quadratic function.

4.3.2 Model-free techniques

Empirical saddlepoint density approximation. The saddlepoint approximation in (11) has a parametric nature (it relies on P_{θ_0}) and we already know that $g(t; n, \varepsilon, G)$ can experience some robustness issues when the actual data distribution does not coincide with P_{θ_0}. One could conjecture that some gains in robustness are obtained using the nonparametric (empirical) saddlepoint density approximation, as derived in Ronchetti and Welsh (1994), where an empirical saddlepoint is defined via the empirical measure – thus avoiding any parametric modelling. We are going to show that this conjecture is wrong.

Under conditions (i)-(v) in Ronchetti and Welsh (1994), the empirical saddlepoint density approximation for $\hat{\theta}_n$ reads as:

$$\hat{g}(t; n, P_n) = (n/2\pi)^{p/2} \hat{c}^{-n}\{\hat{\alpha}(t; P_n); P_n\} \left|\det \hat{M}(t; P_n)\right| \left|\det \hat{\Sigma}(t; P_n)\right|^{-1/2}, \tag{21}$$

where

$$\hat{M}(t; P_n) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \left[- \left.\frac{\partial \psi_j(z_i; t)}{\partial t_r}\right|_t \right] \exp\{\hat{\alpha}^\top(t; P_n) \psi(z_i; t)\} \right\}_{1 \le r, j \le p}, \tag{22}$$

$$\hat{\Sigma}(t;P_n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\psi_j(z_i;t)\psi_r(z_i;t)\exp\{\hat{\alpha}^\top(t;P_n)\psi(z_i;t)\}\right\}_{1\le r,j\le p}, \tag{23}$$

and

$$\hat{c}^{-1}\{\hat{\alpha}(t;P_n);P_n\} = \frac{1}{n}\sum_{i=1}^{n}\exp\{\hat{\alpha}^\top(t;P_n)\psi(z_i;t)\}, \tag{24}$$

with $\hat{\alpha}(t;P_n)$ satisfying:

$$\frac{1}{n}\sum_{i=1}^{n}\psi(z_i;t)\exp\{\hat{\alpha}^\top(t;P_n)\psi(z_i;t)\} = 0. \tag{25}$$

Ronchetti and Welsh (1994) paper illustrates the accuracy of the empirical saddlepoint density approximation in many estimation problems, including the Huber-type M-estimator for the one-dimensional location problem and for linear regression model.

We remark that for some inferential problems, the good performance of $\hat{g}(t;n,P_n)$ can be perturbed by a small percentage of contaminated data, even for first-order B-robust M-estimators. Indeed, we notice that the quantities in (22)-(25) depend on ψ and/or on $\nabla_{\theta^\top}\psi$: estimators defined by an unbounded ψ and/or unbounded $\nabla_{\theta^\top}\psi$ can have large $\hat{M}$ and/or $\hat{\Sigma}$. This is the case, for instance, of the Hampel-Krasker M-estimator (see Appendix A), whose $\hat{g}(t;n,P_n)$ can be unstable over $\mathcal{U}_\eta(P_{\theta_0})$ because $\|\nabla_{\theta^\top}\psi\|$ can be made arbitrarily large in the presence of outliers in the (x,y)-space. In contrast, second-order B-robust M-estimators guarantee that the approximation in (21) remains stable.

Another remark is in order: Section 2 in Ronchetti and Welsh (1994) illustrates how to (i) implement the empirical saddlepoint and (ii) approximate the right tail area probabilities via numerical integration. We notice that, for both (i) and (ii), the estimate $\hat{\theta}_n$ is needed: only bounded ψ and/or bounded $\nabla_{\theta^\top}\psi$ imply that $\|\hat{\theta}_n - \theta_0\|$ remains stable in the presence of outliers.

The same arguments apply to the empirical version of the saddlepoint approximation defined in (15) – where, e.g., the expected values in μ_n or in κ_3 are computed using P_n; see Gatto and Ronchetti (1996), page 669.

Empirical likelihood. The empirical likelihood (EL) method of Owen (1990) leads to the construction of nonparametric confidence regions for the model parameter. In many estimation problems where the sample size is small, the EL confidence regions perform better than the ones obtained by the first-order asymptotics; see Owen (2001) for a book-length discussion.

By comparing the expansions of the empirical log-likelihood ratio and the empirical cumulant generating function calculated at the saddlepoint, Monti and Ronchetti (1993) illustrate the relationship between the EL and the empirical saddlepoint density approximation of Ronchetti and Welsh (1994). On one hand, nonparametric approximation to the density of multivariate M-estimators is obtained using the EL method; on the other hand, nonparametric EL confidence regions are defined using the empirical cumulant generating function evaluated at the saddlepoint

$$\hat{K}(t; P_n) = \ln\left[n^{-1}\sum_{i=1}^{n}\exp\{\hat{\alpha}^\top(t; P_n)\psi(z_i; t)\}\right]$$

where $\hat{\alpha}(t; P_n)$ is defined in (25).

Let $\hat{W}(t)$ be the empirical log-likelihood ratio of Owen (see Owen (1990) for its formal definition and page 332 of Monti and Ronchetti (1993) for its Taylor expansion). Monti and Ronchetti (1993) show that $\hat{W}(t)$ can be expressed using the empirical cumulant at the saddlepoint as

$$n\hat{K}(t; P_n) = -\frac{1}{2}\hat{W}(t) + \frac{1}{\sqrt{n}}\Gamma(u)/6 + O(n^{-1}),$$

where $u = n^{1/2}(t - T_n)$. The scalar quantity $\Gamma(u)$ is given by:

$$\Gamma(u) = -\frac{1}{n}\sum_{i=1}^{n}\left(u^\top\hat{V}^{-1}(T_n; P_n)IF(z_i; P_n)\right)^3,$$

where the matrix $\hat{V}(T_n; P_n) = \hat{M}^{-1}(T_n; P_n)\hat{\Sigma}(T_n; P_n)\hat{M}^{-\top}(T_n; P_n)$ is an estimate of the asymptotic variance, $\hat{\Sigma}(T_n; P_n) = n^{-1}\sum_{i=1}^{n}\psi(z_i; T_n)\psi^\top(z_i; T_n)$,

$$\hat{M}(T_n; P_n) = \frac{1}{n}\sum_{i=1}^{n}\partial\psi(z_i; t)/\partial_t|_{T_n} \quad \text{and} \quad IF(z_i; P_n) = -\hat{M}(T_n; P_n)^{-1}\psi(z_i; T_n).$$

Since $\Gamma(u)$ depends upon both ψ and $\nabla_{\theta^\top}\psi$, comments analogous to the ones made on the stability of the empirical saddlepoint density approximation imply that bounded ψ and bounded $\nabla_{\theta^\top}\psi$ yield stable empirical likelihood confidence intervals (and tests) in the presence of data contamination. Moreover, since the estimate T_n is needed to compute the EL confidence regions, we remark that for (first- and second-order) B-robust M-estimators $\|\hat{\theta}_n - \theta_0\|$ remains stable (at the first- and second-order, respectively) even in the presence of contamination.

Similar considerations apply for the stability analysis of the empirical exponential likelihood tests proposed by Robinson et al. (2003).

5 Numerical studies

Assume we are given two M-estimators of location, defined by two different, bounded estimating functions, having both bounded derivative wrt the parameter. An important practical question arises naturally: *Do different bounded estimating functions with bounded derivative wrt the parameter yield different stability in their asymptotics?* The following example answers this question for the one-dimensional location problem.

Setting and estimation. Let $z_1, ..., z_n$ be n i.i.d. real-valued observations of a random variable Z, having a Laplace ($\mathcal{L}$) distribution with location $\theta_0 = 0$ and scale one (thus, $\mathcal{L}(0,1) \equiv P_{\theta_0}$). We introduce deviations from the reference model, assuming that the actual data distribution is $P_{\epsilon,G} = (1-\epsilon)P_{\theta_0} + \epsilon\delta_z$, where δ_z is a Dirac with mass in 5.5, and ϵ controls for the degree of deviation from P_{θ_0}. Increasing levels of contamination are considered: $\epsilon = 0\%, 5\%$, and 10%.

Three M-estimators are analyzed: the estimator obtained by method of moments (sample mean), the widely-applied first-order (Huber), and the second-order B-robust M-estimators. They are defined by the estimating functions:

$$\psi_{\mathrm{MM}}(z) = z - \theta_0, \quad \psi_b = (z-\theta_0)w_b(z;\theta_0), \quad \text{and} \quad \psi_{b,c} = (z-\theta_0)w_b(z;\theta_0)w_c(z), \tag{26}$$

respectively, where $w_b(z;\theta_0) = \min(1; b/|z-\theta_0|)$, $w_c(z) = \min(1; c/z^4)$; see Proposition 5 in La Vecchia et al. (2012).

The estimating function ψ_{MM} is unbounded, whereas the two robust estimating functions (ψ_b and $\psi_{b,c}$) are both bounded. Moreover, all the estimating functions feature a bounded (in absolute value) $\partial_\theta \psi$, whose upper bound over $\mathcal{Z}$ is equal to one. Proposition 1 implies that both ψ_b and $\psi_{b,c}$ have stable asymptotics. However, thanks to the use of higher-order weights $w_c(z)$, the IF of $\psi_{b,c}$ is of redescending type, namely $\lim_{|z|\to\infty} \psi_{b,c} = 0$; see Figure 4. Thus, the second-order B-robust M-estimator has a finite rejection point (it gives zero weight to large outliers), which yields a better control of outliers; see Hampel et al. (1986).

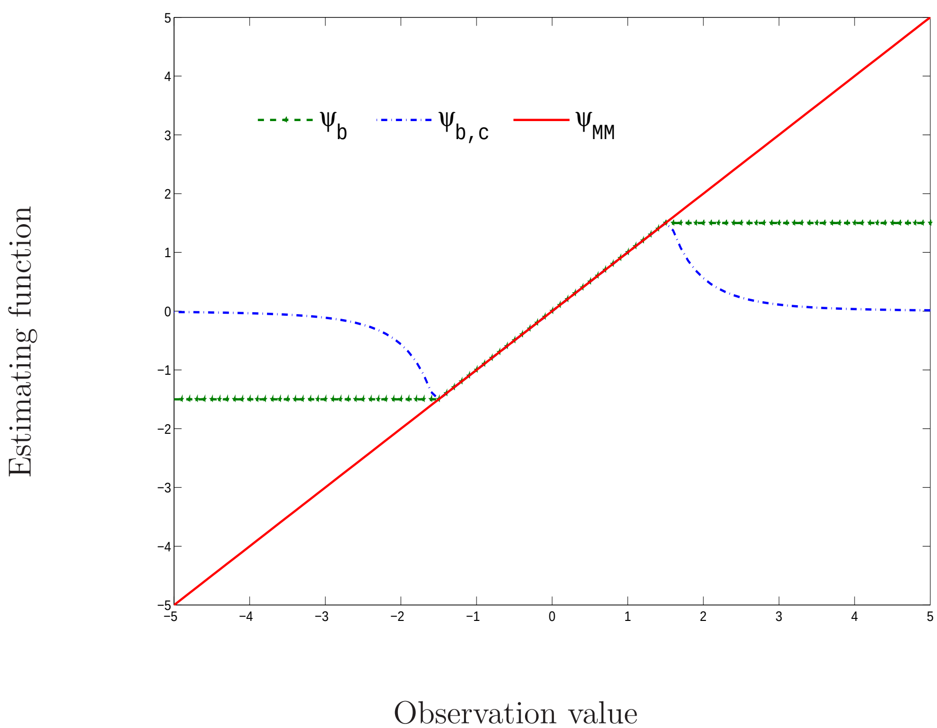

Figure 4: Estimating function for the considered M-estimators: the method of moments (ψ_{MM}, continuous line), the first-order robust (Huber estimator, ψ_b, dashed line) and the second-order robust ($\psi_{b,c}$, dot-dashed line). The tuning constants are $b = 1.5$ and $c = 4$.

First-order asymptotics. Table 1 illustrates numerically the different performance of the three estimators, in terms of their asymptotic bias and asymptotic variance, for $n = 30$ and $n = 200$. The Monte Carlo size is 2,500. When $\epsilon = 0$, all the M-estimators have essentially zero asymptotic bias. However, for both $n = 30$ and $n = 200$, the second-order B-robust M-estimator yields asymptotic bias and asymptotic variance which are smaller than the corresponding quantities related to the Huber-type estimator and to the estimator of the method of moments; similar considerations hold for larger sample sizes (e.g., $n = 2,000$, unreported). This evidence highlights an important point: different bounded estimating functions imply different degrees of stability in their first-order asymptotics. The higher-order weights in $\psi_{b,c}$, by construction, yield an additional flexibility to fine tune the level of robustness.

Higher-order asymptotics. We investigate the higher-order asymptotics of the three M-estimators using the empirical saddlepoint density approximation. To this end, we simulate one clean sample containing $n = 30$ observations from P_{θ_0} and one contaminated sample, having the same size as the clean one, but from $P_{\epsilon,G}$. We estimate the location using the three M-estimators as defined by the estimating functions in Eq. (26). Then, we implement the empirical saddlepoint density approximation for each M-estimator, using the clean and the contaminated sample.

In Figure 5 we display the effect of data contamination on the empirical saddlepoint of the estimator implied by $\psi_{b,c}$ under $P_{\epsilon,G}$ when $\epsilon = 0\%, 5\%$ and 10%: even in the presence of contamination, the density remains fairly stable.

For the sake of comparison, Figure 6 illustrates the stability of the saddlepoint density approximation for the three M- estimators, plotting the left- and right-tail error of the approximation due to deviation from the reference model, defined as

$$\mathrm{err}_{dev}(t;\epsilon, G, P_{\theta_0}) = g(t; n, P_n^{(\varepsilon,G)}) - g_n(t; n, P_n^{(\theta_0)}), \tag{27}$$

for $|t| \geq 0.3$, where $P_n^{(\varepsilon,G)}$ is the empirical measure under the contaminated model $P_{\epsilon,G}$, and $P_n^{(\theta_0)}$ represents the empirical measure under the reference model. The stability of the second-order B-robust M-estimator appears evident. Looking at the right panel, we conclude that, when the right-tail area probabilities are estimated via numerical integration of the empirical saddlepoint (as suggested in Ronchetti and Welsh (1994), page 317), the second-order robust M-estimator yields the most stable estimates in the presence of contamination. Similar considerations hold for the left-tail area probabilities, and, hence for the confidence regions derived from the empirical saddlepoint density approximation.

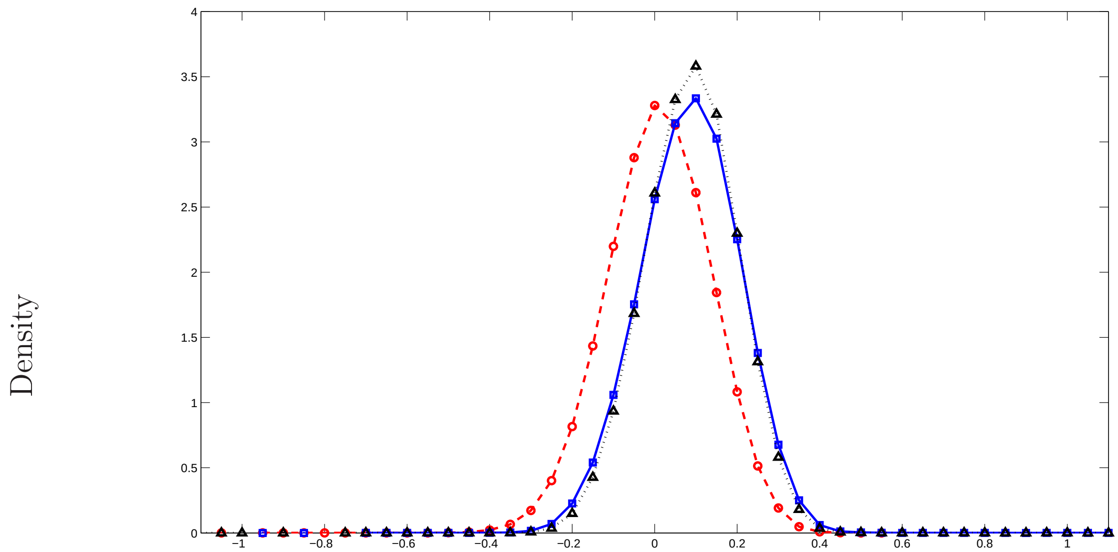

Figure 5: Approximate density of the second order robust M-estimator of location for a Laplace distribution, with mean $\theta_0 = 0$ and scale equal to 1. The dashed line with circles is the empirical saddlepoint density approximation at P_{θ_0}; the continuous line with squares is the empirical saddlepoint density approximation under $P_{\epsilon,G}$ with $\epsilon = 5\%$; the dotted line with triangles is for $\epsilon = 10\%$. The sample size is $n = 30$, G is a Dirac in 5.5. The constants tuning the degree of robustness are $b = 1.5$ and $c = 2$, as in Table 1.

6 Conclusion

The paper illustrates why the estimating function and its characteristics (e.g., its upper and lower bound and/or its derivative wrt the parameter) determine the behavior of the asymptotics of M-estimators. We illustrate that stable asymptotics can be obtained using a bounded ψ having also a bounded $\nabla_{\theta^\top}\psi$. Table 2 summarizes the findings for both model-based and model-free methods considered in this paper.

Our considerations can be adapted to other higher-order asymptotic techniques; like, e.g., the saddlepoint method derived in Easton and Ronchetti (1986) and the small sample confidence intervals of Tingley and Field (1990)). The interested Researcher can study the stability of other asymptotic methods building on the sketch of the proof available in the Appendix.

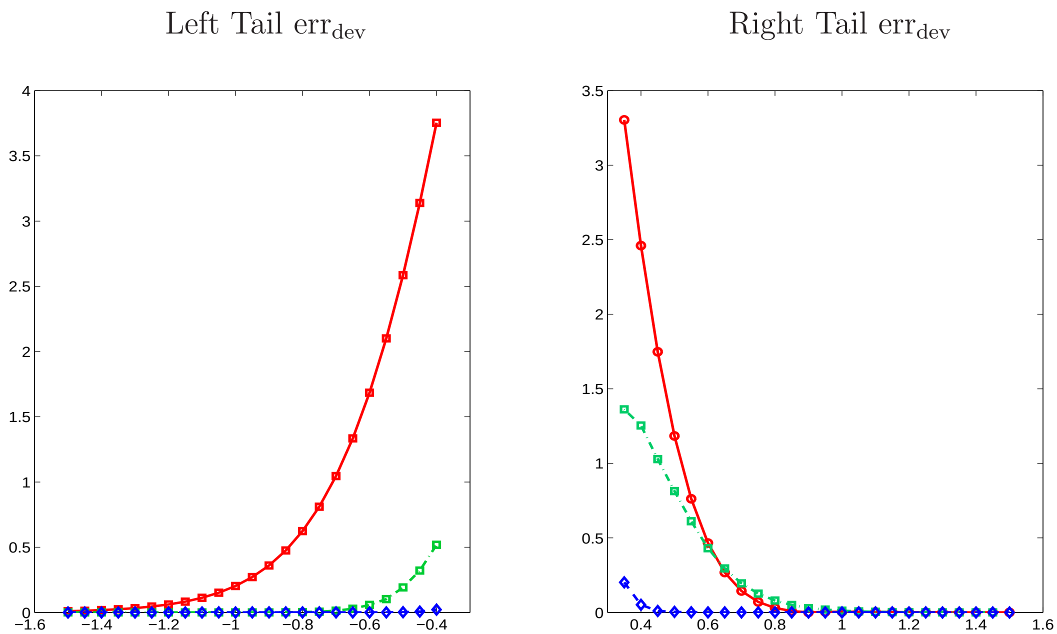

Figure 6: $\mathrm{err}_{\mathrm{dev}}$ in the left (left panel) and right (right panel) tail as in Eq. (27) of the empirical saddlepoint density approximation of the right tail of three M-estimators: method of moments (continuous line), first-order B-robust (Huber M-estimator, dash-dotted line), and second-order B-robust (dotted line). The contaminated distribution $P_{\epsilon,G}$ is obtained as in Eq. (1), where P_{θ_0} is $\mathcal{L}(0,1)$, G is a Dirac in 5.5 and $\epsilon = 10\%$.

7 Acknowledgement

The author thanks two reviewers, the editor and the co-editor-in-chief for their helpful comments and suggestions that have greatly improved this paper. The author acknowledges the Australian Research Council grant DP150100210 for financial support.

		$n=30$		
		ψ_{MM}	ψ_b	$\psi_{b,c}$
$P_{\epsilon,G}$, with $\epsilon=0$	Asympt. Bias	≈ 0	≈ 0	≈ 0
	Asympt. Variance	0.067	0.084	0.013
$P_{\epsilon,G}$, with $\epsilon=5\%$	Asympt. Bias	0.244	0.001	-0.001
	Asympt. Variance	0.096	0.082	0.015
$P_{\epsilon,G}$, with $\epsilon=10\%$	Asympt. Bias	0.503	0.015	-0.001
	Asympt. Variance	0.135	0.085	0.017
		$n=200$		
		ψ_{MM}	ψ_b	$\psi_{b,c}$
$P_{\epsilon,G}$, with $\epsilon=0$	Asympt. Bias	≈ 0	≈ 0	≈ 0
	Asympt. Variance	0.010	0.017	0.001
$P_{\epsilon,G}$, with $\epsilon=5\%$	Asympt. Bias	0.269	0.004	≈ 0
	Asympt. Variance	0.016	0.018	0.001
$P_{\epsilon,G}$, with $\epsilon=10\%$	Asympt. Bias	0.544	0.007	≈ 0
	Asympt. Variance	0.022	0.020	0.001

Table 1: Pure location model: first-order asymptotics. The reference model P_{θ_0} is a Laplace distribution with location $\theta_0 = 0$, scale parameter known and equal to 1 ($\mathcal{L}(0,1)$). The contaminated distribution is $P_{\epsilon,G} = (1-\epsilon)P_{\theta_0} + \epsilon\delta_z$, where δ_z is a Dirac with mass in 5.5. The considered contamination levels are: $\epsilon = 0$, $\epsilon = 0.05$, and $\epsilon = 0.10$. Three M-estimators are analyzed: the method of moments (ψ_{MM}), the first-order robust (Huber estimator, ψ_b) and the second-order robust ($\psi_{b,c}$). The constants are $b = 1.5$ and $c = 2$. The sample sizes are $n = 30$ and $n = 200$. Monte Carlo size is 2,500.

	Model-based techniques			
	Unb. ψ Unb. $\nabla_{\theta^\top}\psi$	Unb. ψ Bou. $\nabla_{\theta^\top}\psi$	Bou. ψ Unb. $\nabla_{\theta^\top}\psi$	Bou. ψ Bou. $\nabla_{\theta^\top}\psi$
Stable asy. bias	-	-	-/+	++
Stable asy. var.	-	-/+	-/+	++
Stable sadd. for $T(\cdot)$	-	-/+	-/+	++
Stable sadd. for a real-valued function $j(\cdot)$ of the parameter	-	-/+	-/+	++
Stable sadd. test	-	-/+	-/+	++
	Model-free techniques			
	Unb. ψ Unb. $\nabla_{\theta^\top}\psi$	Unb. ψ Bou. $\nabla_{\theta^\top}\psi$	Bou. ψ Unb. $\nabla_{\theta^\top}\psi$	Bou. ψ Bou. $\nabla_{\theta^\top}\psi$
Stable empirical sadd. density	-	-/+	-/+	++
Stable empirical sadd. for a real-valued function $j(\cdot)$ of the parameter	-	-/+	-/+	++
Stable empirical likelihood (EL) conf. int. & test	-	-/+	-/+	++
Stable empirical sadd. test	-	-/+	-/+	++

Table 2: Summary of the stability properties implied by ψ and $\nabla_{\theta^\top}\psi$. "Unb." means "unbounded", while "Bou." means "bounded". The symbol "-" represents a lack of the considered feature; the symbol "-/+" represents a partial fulfillment; the symbol "++" represents a complete fulfillment.

References

Barndorff-Nielsen, O. and Cox, D. (1989). *Asymptotic techniques for use in statistics*. Chapman and Hall London.

Brazzale, A., Davison, A. C., and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge University Press.

Butler, R. W. (2007). *Saddlepoint approximations with applications*. Cambridge University Press.

Daniels, H. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25:631–650.

Davison, A., Hinkley, D. V., and Worton, B. (1995). Accurate and efficient construction of bootstrap likelihoods. *Statistics and Computing*, 5:257–64.

Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*, 75(3):417–431.

Duncan, G. (1987). A simplified approach to M-estimation with application to two-stage estimators. *Journal of econometrics*, 34(3):373–389.

Easton, G. S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L-statistics. *Journal of the American Statistical Association*, 81(394):420–430.

Fernholz, L. T. (2001). On multivariate higher-order von Mises expansions. *Metrika*, 53:123–140.

Ferrari, D. and La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika*, 99(1):238–244.

Field, C. (1982). Small sample asymptotic expansions for multivariate M-estimates. *The Annals of Statistics*, pages 672–689.

Field, C. and Ronchetti, E. (1990). *Small sample asymptotics*, volume 13. Inst of Mathematical Statistic.

Gatto, R. and Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *Journal of the American Statistical Association*, Vol. 91, No. 434:666–673.

Goutis, C. and Casella, G. (1999). Explaining the saddlepoint approximation. *The American Statistician*, 53(3):216–224.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer.

Hampel, F., Ronchetti, E., Russeeuw, P., and Sthael, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

He, X. and Shao, Q. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, 24(6):2608–2630.

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

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

Jensen, J. L. (1995). *Saddlepoint Approximations*. Oxford University Press, U.K.

Kolassa, J. (2006). *Series approximation methods in statistics*, volume 88. Springer Verlag.

La Vecchia, D., Ronchetti, E., and Trojani, F. (2012). Higher-order infinitesimal robustness. *Journal of the American Statistical Association*, 107(500):1546–1557.

Lô, S. N. and Ronchetti, E. (2009). Robust and accurate inference for generalized linear models. *Journal of Multivariate Analysis*, 100(9):2126–2136.

Monti, A. and Ronchetti, E. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika*, 80(2):329–338.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (2001). *Empirical likelihood*. CHAPMAN & HALL, CRC press.

Rieder, H. (1994). Robust asymptotic statistics.

Robinson, J., Ronchetti, E., and Young, G. (2003). Saddlepoint approximations and tests based on multivariate M-estimates. *Annals of statistics*, pages 1154–1169.

Ronchetti, E. and Ventura, L. (2001). Between stability and higher-order asymptotics. *Statistics and Computing*, 11:67–73.

Ronchetti, E. and Welsh, A. (1994). Empirical saddlepoint approximations for multivariate M-estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 313–326.

Tingley, M. and Field, C. (1990). Small-sample confidence intervals. *Journal of the American Statistical Association*, Vol. 85, No. 410.:427–434.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, 18:309–348.

Young, G. (2009). Routes to higher-order accuracy in parametric inference. *Australian & New Zealand Journal of Statistics*, 51(2):115–126.

A Linear regression model

Let us consider the linear regression model $Y = \theta_1 + \theta_2 X + \epsilon$, where $\epsilon \sim \mathcal{N}(0,1)$. Set $\theta = (\theta_1, \theta_2)$, then, we have that:

- The maximum likelihood estimator is defined by

$$\psi(y, x; \theta) = ux$$

  with $u = y - \theta_1 - \theta_2 x$;

- The Hampel-Kasker (first-order B-robust) estimator is defined by:

$$\psi_b(y, x; \theta) = Axu \min\left(1; \frac{b}{|u|\|Ax\|}\right), \tag{28}$$

  where A is determined by

$$E_{P_{\theta_0}}[-\nabla_{\theta^\top} \psi_b(Y, X; \theta_0)] = I; \tag{29}$$

- The second-order robust M-estimator has an estimating function given by:

$$\psi_{b,c}(y, x; \theta) = Axu \min\left(1; \frac{b}{|u|\|Ax\|}\right) \min\left(1; \frac{c}{\|x\|^2}\right), \tag{30}$$

  where the matrix A is defined as in (29), with ψ_b replaced by $\psi_{b,c}$.

Estimating function (30) corresponds to the estimating function of an optimal B-robust Hampel-Krasker M-estimator, with additional *Mallows-type weights* on the x variable. Therefore, large observations in the x-space are typically down-weighted more than in the first-order B-robust M-estimator and this yields the control of the derivative of the estimating function wrt to the parameter.

B Proofs

Proof of Proposition 1. *Part (i).* Let ψ be an estimating function $\psi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^p$. For $P_{\epsilon,G} \in \mathcal{U}_\eta(P_\theta)$, we consider the quantities

$$\begin{aligned} M(\theta; P_{\epsilon,G}) &= (M_{ij}(\theta; P_{\epsilon,G}))_{1\le i,j\le p} = E_{P_{\epsilon,G}}\left[-\nabla_{\theta^\top}\psi(Z; T(P_{\epsilon,G}))\right] , \\ \Sigma(\theta; P_{\epsilon,G}) &= (\Sigma_{ij}(\theta; P_{\epsilon,G}))_{1\le i,j\le p} = E_{P_{\epsilon,G}}\left[\psi(Z, T(P_{\epsilon,G}))\psi^\top(Z, T(P_{\epsilon,G}))\right] . \end{aligned}$$

The stability of M over $\mathcal{U}_\eta(P_{\theta_0})$ is related to the directional derivative

$$\begin{aligned} \left.\frac{\partial M_{ij}(\theta; P_{\epsilon,G})}{\partial \epsilon}\right|_{\epsilon=0} &= \partial_\epsilon E_{P_{\epsilon,G}}\left[-\partial_{\theta_j}\psi_i(Z; T(P_{\epsilon,G}))\right]|_{\epsilon=0} \\ &= -\left\{E_G[\partial_{\theta_j}\psi_i(Z;\theta_0)] - E_{P_{\theta_0}}[\partial_{\theta_j}\psi_i(Z;\theta_0)]\right\} \\ &\quad -E_{P_{\theta_0}}[\nabla_{\theta^\top}\partial_{\theta_j}\psi_i(Z;\theta_0)]E_G[IF(Z;P_{\theta_0})]. \end{aligned} \tag{31}$$

Under A1, it follows that

$$\sup_{G\in\mathcal{M}} \|E_G[IF(Z; P_{\theta_0})]\| < \infty. \tag{32}$$

Under A2, we have that $\partial_{\theta_j}\psi_i$ is bounded for each $1 \le i, j \le p$. Thus,

$$\sup_{G\in\mathcal{M}} \|E_G[\partial_{\theta_j}\psi_i(Z;\theta_0)]\| < \infty. \tag{33}$$

Making use of Eq. (32) and (33) in Eq. (31), it follows

$$\sup_{G\in\mathcal{M}} \|\partial_\epsilon M_{ij}(\theta; P_{\epsilon,G})|_{\epsilon=0}\| < \infty. \tag{34}$$

As far as the derivative $\partial_\epsilon \Sigma(\theta; P_{\epsilon,G})$ is concerned, we have:

$$\begin{aligned}\left.\frac{\partial \Sigma_{ij}(\theta; P_{\epsilon,G})}{\partial \epsilon}\right|_{\epsilon=0} &= \partial_\epsilon E_{P_{\epsilon,G}}\left[\psi_i(Z; T(P_{\epsilon,G}))\psi_j(Z; T(P_{\epsilon,G}))\right]|_{\epsilon=0} \\ &= E_G\Big[\psi_i(Z;\theta_0)\psi_j(Z;\theta_0)\Big] - E_{P_{\theta_0}}\Big[\psi_i(Z;\theta_0)\psi_j(Z;\theta_0)\Big] \\ &\quad + E_{P_{\theta_0}}\Big[\nabla_{\theta^\top}(\psi_i(Z;\theta_0)+\psi_i(Z;\theta_0))\Big]E_G\Big[IF(Z;P_{\theta_0})\Big].\end{aligned}$$

Under A1 and A2, this derivative is bounded, thus

$$\sup_{G\in\mathcal{M}} \|\partial_\epsilon \Sigma_{ij}(\theta; P_{\epsilon,G})|_{\epsilon=0}\| < \infty. \tag{35}$$

That concludes the proof.

Proof of Corollary 2. The proof follows from the V-robustness of second-order robust M-estimators; see La Vecchia et al. (2012). We here provide a sketch of the proof. The behavior of $\partial_\epsilon V(\theta; P_{\epsilon,G})|_{\epsilon=0}$ is related to $\partial_\epsilon M(\theta; P_{\epsilon,G})|_{\epsilon=0}$ and $\partial_\epsilon \Sigma(\theta; P_{\epsilon,G})|_{\epsilon=0}$. Therefore, we need to compute these (functional) derivatives under the given assumptions A1 and A2. The derivative $\partial_\epsilon M(\theta; P_{\epsilon,G})|_{\epsilon=0}$ is given in Eq. (31) and from Eq. (32)-(33), we already know that a function ψ satisfying A1 and A2 is such that Eq. (34) holds. Moreover, Eq. (35) implies that derivative $\partial_\epsilon \Sigma(\theta; P_{\epsilon,G})|_{\epsilon=0}$ stay bounded over $\mathcal{U}_\eta(P_{\theta_0})$. That concludes the proof.

C Assumptions

C.1 Root-n consistency and asymptotic normality

The following general assumptions ensure root-n consistency and asymptotic normality of M-estimators. They are fairly sharp and they can be found in Huber (1981), p. 131. Stronger and more easily verifiable conditions are given in Duncan (1987). Weaker conditions to prove only asymptotic normality can be found in He and Shao (1996), Corollary 2.2.

*B*1. For each fixed θ, let be $\psi(z;\theta)$ a $p\times 1$-vector function on $\mathcal{Z}\times\Theta$, where $\mathcal{Z}\in\mathbb{R}^m$ and Θ is a compact subset of $\mathbb{R}^p$. The function $\psi(z;\theta)$ is separable in sense of Doob.

$B2$. The expected value $E_{P_{\theta_0}}[\psi(Z;\theta)]$ exists for all $\theta \in \Theta$ and has a uniques zero at $\theta = \theta_0$.

$B3$. The function ψ is continuous in θ:

$$\lim_{\|\tilde{\theta}-\theta\|\to 0} \|\psi(z;\tilde{\theta}) - \psi(z;\theta)\| = 0$$

Here and in the following assumptions, $\|\cdot\|$ represents the Euclidean norm.

$B4$. There exists a real-valued continuous function $b(\theta)$ that is bounded away from zero, $b(\theta) \geq b_0 > 0$, such that:

(i) $\sup_\theta \frac{\|\psi(z;\theta)\|}{b(\theta)}$ is P_{θ_0}-integrable;

(ii) $\liminf_{\theta\to\infty} \frac{\|E_{P_{\theta_0}}[\psi(Z;\theta)]\|}{b(\theta)} \geq 1$;

(iii) $E_{P_{\theta_0}}\left[\limsup_{\theta\to\infty} \frac{\|\psi(z;\theta)-E_{P_0}[\psi(Z;\theta)]\|}{b(\theta)}\right] < 1$.

In addition to $B1$.-$B4$., the following assumption is needed to show the asymptotic normality of the M-estimator implied by ψ.

$N4$. Define

$$u(z;\theta,d) = \sup_{\|\tilde{\theta}-\theta\|\leq d} \|\psi(z;\tilde{\theta}) - \psi(z,\theta)\|.$$

There are strictly positive numbers a_0, b_0, c_0, d_0 such that:

(i) $\|E_{P_{\theta_0}}[\psi(Z;\theta)]\| \geq a_0\|\theta-\theta_0\|$, for $\|\theta-\theta_0\| \leq d_0$;

(ii) $E_{P_{\theta_0}}[u(Z;\theta,d)] \leq bd$, for $\|\theta-\theta_0\| + d \leq d_0$;

(iii) $E_{P_{\theta_0}}[u(Z;\theta,d)^2] \leq cd$, for $\|\theta-\theta_0\| + d \leq d_0$.

Moreover, the expectations $E_{P_{\theta_0}}[\|\psi(Z;\theta)\|^2]$ and $M(\psi;\theta_0) = E_{P_{\theta_0}}[\nabla_{\theta^\top}\psi(Z;\theta_0)]$ are nonzero and finite. Finally, we also assume the following conditions about differentiability. Let ∂_{θ_j} denote differentiation w.r.t. θ_j, then the function $\psi(z;\theta)$ has (P_{θ_0}-a.s.) derivatives $\partial_{\theta_j}\psi_i(z;\theta)$, for

$1 \le i, j \le p$.

Condition $B3$. requires the continuity of ψ. This requirement can be weakened, considering P_0-a.s. continuity, namely

B3' The function ψ is P_{θ_0}-a.s. continuous in θ:

$$\lim_{\|\tilde{\theta}-\theta\|\to 0} \|\psi(z;\tilde{\theta}) - \psi(z;\theta)\| = 0, \quad P_{\theta_0} - a.s.$$

C.2 Saddlepoint density approximation

The next set of assumptions ensure the existence of the saddlepoint density approximation; see Field and Ronchetti (1990), p. 62.

*A*4.2*M* There exists an open subset U of $\mathbb{R}^m$ such that:

(i) for each $\theta \in \Theta$, $P_\theta(U) = 1$;

(ii) $\psi(z;\theta)$ has P_{θ_0}-a.s. derivatives $\partial_{\theta_k}\partial_{\theta_j}\psi_i(z;\theta)$ and $\partial_{\theta_l}\partial_{\theta_k}\partial_{\theta_j}\psi_i(z;\theta)$, for $1 \le i, j, k, l \le p$.

*A*4.3*M* For each compact $K \subset \Theta$:

(i) for $0 \le j, k \le p$, $1 \le i \le p$,

$$\sup_{\theta_0 \in K} E_{P_{\theta_0}} |\partial_{\theta_k}\partial_{\theta_j}\psi_i(Z;\theta_0)|^4 < \infty;$$

(ii) there exists an $\epsilon > 0$ such that for $1 \le i, j, k, l \le p$,

$$\sup_{\theta_0 \in K} E_{P_{\theta_0}} [\max_{\|\theta-\theta_0\|<\epsilon} |\partial_{\theta_l}\partial_{\theta_k}\partial_{\theta_j}\psi_i(Z;\theta)|]^3 < \infty.$$

*A*4.5*M* The functions $M(\psi;\theta)$ and $E_{P_\theta}[(\partial_{\theta_{k_1}}\partial_{\theta_{j_1}}\psi_{r1})(\partial_{\theta_{k_2}}\partial_{\theta_{j_2}}\psi_{r2})]$, $0 \le j_1, j_2, k_1, k_2 \le p$, $k_1 + j_1 \ge 1$, $k_2 + j_2 \ge 1$, $1 \le r_1, r_2 \le p$ are continuous on Θ.