Journal of Multivariate Analysis 102 (2011) 828–846

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

$\sqrt{n}$-Consistent robust integration-based estimation

Sung Jae Jun *, Joris Pinkse, Yuanyuan Wan

Center for the Study of Auctions, Procurements and Competition Policy, Department of Economics, The Pennsylvania State University, United States

ARTICLE INFO

Article history:
Received 14 June 2010
Available online 8 January 2011

AMS subject classifications:
62F10
62F35
62J05

Keywords:
Robust regression
Linear model
Integration-based estimator
High breakdown point estimator

ABSTRACT

We propose a new robust estimator of the regression coefficients in a linear regression model. The proposed estimator is the only robust estimator based on integration rather than optimization. It allows for dependence between errors and regressors, is $\sqrt{n}$-consistent, and asymptotically normal. Moreover, it has the best achievable *breakdown point* of regression invariant estimators, has bounded *gross error sensitivity*, is both *affine invariant* and *regression invariant*, and the number of operations required for its computation is linear in n. An extension would result in bounded *local shift sensitivity*, also.

1. Introduction

We propose a new estimator for the regression coefficients in a linear regression model, which is robust to 'contamination'. Our estimator is inspired by the *least median of squares* (LMS) estimator of Rousseeuw [27] and the *Laplace estimator* of Chernozhukov and Hong [3]; see also Jun et al. [18]. Like Laplace estimators, our estimator is defined as the ratio of two integrals involving an exponential transform of (in our case) the LMS objective function, but this is where the similarity ends.

Suppose that the parameter vector of interest θ_0 is the unique minimizer of a population objective function $\varOmega$ over a compact parameter space $\varTheta$. Laplace estimators then employ the fact that θ_0 satisfies

$$\theta_0 = \lim_{n\to\infty} \frac{\int \theta \varpi(\theta) \exp\{-\alpha_n \varOmega(\theta)\} \mathrm{d}\theta}{\int \varpi(\theta) \exp\{-\alpha_n \varOmega(\theta)\} \mathrm{d}\theta}, \tag{1.1}$$

where ϖ is a pseudo-prior defined on $\varTheta$ and $\{\alpha_n\}$ is a scalar-valued deterministic sequence diverging to infinity with the sample size n. Note here that the density $\varpi(\theta)\exp\{-\alpha_n\varOmega(\theta)\}/\int \varpi(\theta)\exp\{-\alpha_n\varOmega(\theta)\}\mathrm{d}\theta$ becomes more concentrated around θ_0 as α_n increases. Replacing $\varOmega$ in (1.1) with its sample analog $\hat{\boldsymbol{\varOmega}}$[1] results in a Laplace estimator. If a quadratic expansion of $\hat{\boldsymbol{\varOmega}}$ is available then the Laplace estimator is generally $\sqrt{n}$-consistent [3] and the divergence rate of α_n is of lesser importance. In the absence of such a quadratic expansion, as in the case of the LMS estimator, the resulting estimator is not $\sqrt{n}$-consistent, and the divergence rate of α_n partly determines the convergence rate of the Laplace estimator [18].

We, instead, use the fact that in our case $\varOmega$ is symmetric around θ_0, which implies that

$$\theta_0 = \frac{\int \theta \exp\{-\varOmega(\theta)\}\mathrm{d}\theta}{\int \exp\{-\varOmega(\theta)\}\mathrm{d}\theta}, \tag{1.2}$$

* Corresponding address: 303 Kern Graduate Building, University Park, PA 16802, United States.
E-mail addresses: sjun@psu.edu, suj14@psu.edu (S.J. Jun), joris@psu.edu (J. Pinkse), yxw162@psu.edu (Y. Wan).

1 We use bold face for random variables.

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.01.003

Table 1
Comparison of robust estimators of the coefficients in a linear regression model.

Estimation method	Acronym	BDP = 0.5	GES finite	LSS finite	$\sqrt{n}$ rate	Normal	Comp. # oper.	Equivariance		
								Scale	Affine	Regr.
Huber [17]	HUB				✓	✓	?		✓	✓
Koenker and Bassett [21]	LAD				✓	✓	n [a]	✓	✓	✓
Krasker [22]	HK		✓		✓	✓	?		✓	✓
Siegel [32][b]	RM	✓	✓		✓	✓	n^d [c]	✓		✓
Mallows [23]	MAL		✓	✓	✓	✓	?		✓	✓
Rousseeuw [27]	LMS	✓	✓[d]				n^d [e]	✓	✓	✓
Rousseeuw [27]	LTS	✓	✓		✓	✓	$n \log n$	✓	✓	✓
Rousseeuw and Yohai [29]	SEST	✓			✓	✓	$n^2 \log n$	✓	✓	✓
Yohai [33]	MM	✓			✓	✓	?		✓	✓
Yohai and Zamar [34]	TAU	✓			✓	✓	?		✓	✓
Croux et al. [5]	GS	✓			✓	✓	$n^2 \log n$ [f]		✓	✓
Hossjer [16]	LTA	✓	✓		✓	✓	$n \log n$	✓	✓	✓
Chang et al. [2]	HBRR	✓	✓[g]	?	✓	✓	?	✓	✓	✓
Zinde-Walsh [35]	SLMS	✓				✓	?		✓	✓
Čížek [4]	GTE	✓	✓		✓	✓	?	✓	✓	✓
New		✓	✓	[h]	✓	✓	n		✓	✓

[a] With preprocessing; see Portnoy and Koenker [24].
[b] Asymptotics are due to Hossjer [16].
[c] d is the number of regressors.
[d] If the constant is not varied, infinite if varied; see Davies [7].
[e] See Croux et al. [5].
[f] See Croux et al. [5]; d is the number of regressors.
[g] If the constant is not varied.
[h] Can be modified to have a finite LSS.

where the integrals are taken over the entire Euclidean space. There are four fundamental differences between (1.1) and (1.2): in (1.2) there is no limit, there is no α_n, there is no compact parameter space requirement, and there is no ϖ. As there is no limit in (1.2), α_n is not needed anymore. Since the symmetry of Ω around θ_0 is used, the parameter space should not be artificially restricted and no prior can be used. Our estimator $\hat{\boldsymbol{\theta}}$ is obtained by replacing Ω in (1.2) with $\hat{\boldsymbol{\Omega}}$.

In this paper we focus our attention on the case in which $\hat{\boldsymbol{\Omega}}$ is the LMS objective function, or a close relative thereof. We show that, subject to assumptions outlined in subsequent sections, $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent and asymptotically normal with many robustness properties, which will be further explained below. Please note that although our estimator resembles a Bayes estimator, it is quite different in that with $\hat{\boldsymbol{\Omega}}$ being the LMS objective function, $\exp(-\hat{\boldsymbol{\Omega}})$ is not a likelihood.

Instead of basing an estimator on (1.2), as we do in this paper, one could alternatively consider $\hat{\boldsymbol{\theta}}_{\boldsymbol{L}}$, the Laplace estimator using $\hat{\boldsymbol{\Omega}}$. However, as the LMS objective function does not allow for a quadratic expansion [19], $\hat{\boldsymbol{\theta}}_{\boldsymbol{L}}$ will not be $\sqrt{n}$-consistent. Indeed, this scenario is similar to the one studied in [18] for the objective functions of other $\sqrt[3]{n}$-consistent estimators.

The pioneering work of Huber [17] has spawned an abundance of papers proposing estimators with ever more desirable robustness properties. The main differences between the estimators are their robustness properties, their asymptotic behavior absent contamination, their equivariance properties, and their degree of computational complexity. These properties are summarized in Table 1. Our estimator is attractive in all four respects, as the exposition below will make apparent.

One notion of robustness is the finite sample *breakdown point* [8],[2] which is the fraction of the sample that must be changed to push the value of an estimator arbitrarily far. The breakdown point of the least squares estimator equals $1/n$ and the breakdown point of the least absolute deviations estimator [21] depends on the regressor distribution and can be arbitrarily close to zero in large samples [15, p. 328]. Most estimators, however, have a finite sample breakdown point close to 0.5 if the regressors are in *general position* [27]. Notable exceptions are Huber [17], Krasker [22], Mallows [23]. Our estimator has the best achievable breakdown point of regression invariant estimators, determined in [27].

Since the requirement that regressors are in general position is strong, we provide results that are more general than that. Specifically, it can be preferable (from a breakdown point perspective) to use a quantile q other than the median. Details can be found in Section 3.

Other commonly used notions of robustness are the *gross error sensitivity* (GES) and the *local shift sensitivity* (LSS), both due to Hampel [12,14]. The GES of an estimator is finite if its *influence function* [12,14] is bounded. Many, but not all, robust estimators have a bounded influence function, including ours.

The LSS is finite if the partial derivative of the influence function with respect to regressor and regressand values is bounded.[3] We know of only one estimator, namely Mallows [23], which is known to have a finite LSS. The proposed estimator

[2] An asymptotic version can be found in [13] and a different breakdown point concept in [30,31].
[3] The definition of the LSS is more general in that it allows for left and right derivatives to be different.