ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 664 January 2009

On the use of robust regression in econometrics

Markus Baldauf and J.M.C. Santos Silva

On the use of robust regression in econometrics*

Markus Baldauf† J.M.C. Santos Silva‡

April 27, 2009

Abstract

The use of robust regression estimators has gained popularity among applied econometricians. The main argument invoked to justify the use of the robust estimators is that they provide efficiency gains in the presence of outliers or non-normal errors. Unfortunately, most practitioners seem to be unaware of the fact that heteroskedastic and skewed errors can dramatically affect the properties of these estimators. In this paper we reconsider the interpretation of the specific robust estimator that has become popular in applied econometrics, and conclude that its use in this context cannot be generally recommended.

JEL classification code: C13, C21.

Key words: Heteroskedasticity, Iteratively reweighted least squares, M-estimator, Mode regression, Skewness.

*We are grateful to Marcus Chambers, Geert Dhaene, Gordon Kemp, Myoung-jae Lee, José Machado, Paulo Parente and Silvana Tenreyro for helpful comments and advice. The usual disclaimer applies. Santos Silva also gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

†RBB Economics. E-mail: markus.baldauf@rbbecon.com.

‡Department of Economics, University of Essex and CEMAPRE. E-mail: jmcss@essex.ac.uk.

1. INTRODUCTION

The expression "robust regression" denotes a set of estimation techniques that are less sensitive than ordinary least squares (OLS) to the effect of possible influential observations. The main argument invoked to justify the use of robust regression is that it provides efficiency gains in the presence of outliers or heavy-tailed errors. In its various forms, robust regression has a well established tradition in statistics (see, e.g., Huber, 1981; Hampel, Ronchetti, Rousseeuw and Stahel, 1986; Rousseeuw and Leroy, 1987, and Maronna, Martin and Yohai, 2006). However, apart from median regression and quantile regression in general (Koenker and Bassett, 1978), robust regression was slow to gain popularity in economics and econometrics, and it is not covered in leading modern econometric textbooks.[1]

Nevertheless, over the past decade, a form of robust regression based on Huber's (1964) M-estimator has become popular among applied econometricians and has been frequently used both in leading research publications and in industry.[2] Following the work of Holland and Welsch (1977), this estimator is implemented using a iteratively reweighted least squares (IRLS) algorithm and is available in popular software packages.[3]

[1] A rare exception is Peracchi (2001).

[2] For examples of academic publications using the M-estimator see, among many others, Alpizar, Carlsson and Johansson-Stenman (2008), Andersen and Aslaksen (2008), Freund and Bolaky (2008), Rogers (2008), Deschênes and Greenstone (2007), Baker and Hall (2004), Currie and Fahr (2004) and Strömberg (2004). The recent merger appraisals of Ryanair/Aer Lingus and StatoilHydro/ConocoPhillips (European Commission, 2007 and 2008) are examples of the use of this estimator in industry. Baldauf advised Ryanair and StatoilHydro and Santos Silva provided economic advice to StatoilHydro.

[3] For example, robust regression is available in Stata via the command *rreg* (StataCorp., 2007), in SAS via *PROC ROBUSTREG* (SAS Institute Inc., 2008), in R and S-PLUS via *rlm* (Venables and Ripley, 2002) and *rreg* (Heiberger and Becker, 1992), and in Matlab via *robustfit* (Mathworks, 2008).

However, perhaps because of the lack of appropriate references on its use in econometrics, most practitioners seem to be unaware of the fact that the consistency of this estimator depends on assumptions about the symmetry and homoskedasticity of the errors and justify its use with misleading claims about its properties. For example, in an influential paper, Hamilton (1991, p. 23) suggests that this sort of estimator is able to handle skewed distributions and some authors go even further, justifying the use of this estimator as a way to deal with the likely presence of heteroskedasticity in the data (see, e.g., Croxson, Propper and Perkins, 2001, and Chan, Godby, Mestelman and Muller, 2002). In industry, an additional and surprising justification for the use of the robust M-estimator is found: "to improve the regression fit" (European Commission, 2007, p. 465, and 2008, p. 28).[4]

In this paper we discuss the interpretation of the specific robust M-estimator that has become popular in applied econometrics, henceforth termed IRLS M-estimator,[5] and give the conditions required for it to be consistent for the parameters of the conditional mean. In particular, we emphasize that in the presence of skewed heteroskedastic errors this M-estimator will be inconsistent for these parameters and note that its efficiency can be severely affected by heteroskedasticity. Although we focus on the IRLS M-estimator, our results extend naturally to other robust regression estimators, such as the Gastwirth and trimean estimators introduced by Koenker and Bassett (1978) and the trimmed least squares estimator of Ruppert and Carroll (1980).

The paper is organized as follows. In the next section we present the IRLS version of Huber's (1964) M-estimator that has been used in applied econometrics, and discuss in detail its interpretation. Section 3 presents the results of a small simulation study illustrating the pitfalls of using this estimator when the errors are heteroskedastic

[4]It is well-known that OLS maximizes the R^2, the standard goodness-of-fit measure in linear regression.

[5]We use this terminology in reference to the algorithm typically used in the implementation of the estimator and to distinguish it from other M-estimators such as OLS and quantile regression.

and/or skewed. Section 4 discusses the use of the IRLS M-estimator in econometric applications and, finally, Section 5 presents brief concluding remarks.

2. THE IRLS M-ESTIMATOR

2.1. Set-up and notation

We consider the problem of estimating a regression model of the form

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, ..., n,$$

where y_i is a scalar, x_i and β are k dimensional vectors with $k < n$, and ε_i is an uncorrelated random disturbance.

Part of the difficulty in interpreting the results obtained with robust regression estimators is that the authors are often vague about the properties of the error term and, consequently, about what location function of the distribution of y is estimated. For example, in his seminal contributions, Huber (1973, 1981) just states that the errors are independent with approximately identical distributions. However, Huber (1973, p. 800) adds that the desired estimate of β "will in some sense generalize a robust alternative to the sample mean," suggesting that $x_i'\beta = \mathrm{E}(y_i)$, for fixed regressors, or $x_i'\beta = \mathrm{E}(y_i|x_i)$, for random regressors. When further assumptions are made about the errors, it is typically added that they are identically distributed with $\mathrm{E}(\varepsilon_i) = 0$ (see, among others, Li, 1985, and Wu, 1985), confirming the idea that the objective is to make the usual mean regression more robust.

More rarely, it is additionally assumed that ε_i has a symmetric distribution (e.g., Hogg, 1979, Hampel et al., 1986). In this case, of course, there is no difficulty in interpreting the robust regression estimator because the location functions estimated by these methods coincide with the mean. However, the symmetry assumption is not explicitly mentioned in any of the empirical applications of this estimator that we came across, and it seems to be generally ignored by practitioners.

Because in economic applications it is generally more appropriate to treat the regressors as random, we will assume that $x_i'\beta$ is the conditional expectation of y_i given x_i, and consequently $\mathrm{E}(\varepsilon_i|x_i) = 0$.[6]

Following Huber (1973), a M-estimator of β is defined as

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i'b}{s}\right), \tag{1}$$

where s is a scale parameter and $\rho(\cdot)$ is an even function that is non-decreasing in the positive half-line. The properties of $\hat{\beta}$ will naturally depend of the particular form of $\rho(\cdot)$ that is adopted and the choice of function to use is often based on robustness and computational considerations (see, e.g., Li, 1985).

Generally speaking, $\hat{\beta}$ will be an estimate of the parameters of some location function of the conditional distribution of y given x. For example, it is well known that OLS and least absolute deviations are special cases of (1) that estimate the conditional mean and median, respectively. Therefore, it is clear that different choices of $\rho(\cdot)$ will affect not only the efficiency of the estimator and the convergence properties of the algorithm used in the minimization of the objective function, but, more importantly, the interpretation of the estimates.

2.2. Li's algorithm

It is now interesting to study the particular estimator described in Li (1985, pp. 335-6), which has been used in virtually all econometric applications of the IRLS M-estimator. This algorithm starts with an OLS regression and proceeds with a set of iterations using Huber (1964) weights. The objective of this first set of iterations is just to find suitable starting values for the minimization of the objective function of interest. After convergence with the first set of weights is achieved, a new set of iterations begins using biweights (Beaton and Tukey, 1974). That is, the objective

[6]See Goldberger (1991) and Wooldridge (2002) on the role of the conditional expectation in econometrics.

function to be minimized has the form

$$\sum_{i=1}^{n} \frac{s^2}{6} \left\{ 1 - \boldsymbol{I} \left[\left| \frac{y_i - x_i' b}{s} \right| \leq 1 \right] \left[1 - \left(\frac{y_i - x_i' b}{s} \right)^2 \right]^3 \right\}, \tag{2}$$

where $\boldsymbol{I}[e]$ is the indicator function for event e and s is proportional to the median absolute deviation defined as $\operatorname*{med}_i \left\{ \left| (y_i - x_i' b) - \operatorname*{med}_j \left(y_j - x_j' b \right) \right| \right\}$, where b is evaluated at the current estimate of β.[7]

To gain further insight into this estimator, it is interesting to notice that minimizing (2) is equivalent to maximizing

$$\frac{1}{ns} \sum_{i=1}^{n} K_T \left(\frac{y_i - x_i' b}{s} \right), \tag{3}$$

where $K_T(u) = \frac{35}{32} \boldsymbol{I}[|u| \leq 1] (1 - u^2)^3$ is the triweight kernel (see, e.g., Wand and Jones, 1995). Expression (3) is immediately recognizable as a non-parametric estimator of the density of y_i at $x_i' b$. Therefore, the value of b that maximizes (3) corresponds to the conditional mode of y_i, assumed to be a linear function of x_i.

Mode regression has been pioneered by Lee (1989 and 1993) and the estimator defined by (2) can be seen as a member of the family of mode regression estimators based on smooth kernels described in Lee and Kim (1998, pp. 214-5). Indeed, Lee and Kim (1998) explicitly consider the mode regression estimator based on the objective function of Andrews' (Andrews, Bickel, Hampel, Huber, Rogers, and Tukey, 1972) cosine M-estimator, and mention that the same approach can be used with related objective functions, such as the quartic (or biweight) kernel (see, e.g., Wand and Jones, 1995).

More generally, although that does not seem to have been recognized in the literature on robust regression, (1) can define a mode-regression estimator when the distribution of the errors ε_i has some degree of symmetry and $\rho(\cdot) = a - bK(\cdot)$, where a and $b > 0$ are constants and $K(\cdot)$ is a kernel function such that $\int K(z)\, dz = 1$ and

[7] The objective function defined by (2) can have multiple minima and that is why it is important to have good starting values and the first set of iterations is needed.

$\lim_{z\to\pm\infty} K(z) = 0$. The link between mode regression and the M-estimator defined by (2) is convenient because the conditions needed for it to be consistent for the parameters of the conditional mean can be explicitly found in the results given by Lee (1989, 1993).

2.3. Properties of the IRLS M-estimator based on biweights

As in Lee (1989, 1993), the IRLS M-regression algorithm suggested by Li (1985) treats s as a fixed parameter. That is, s is not allowed to depend on the sample size and its choice depends on the researcher's preferences with respect to the trade-off between efficiency and robustness.

For fixed s, the main requirements for the estimator based on (2) to be consistent for the parameters of the conditional expectation of y_i given x_i are as follows (see Lee, 1989, 1993, for further details):

A1: The density of ε_i is strictly unimodal with a finite mode at zero;

A2: Either of the following conditions holds:

(a) the density of ε_i is symmetric around zero;[8]

(b) ε_i is statistically independent of x_i.

Given A1, assumption A2 (a) is enough to ensure the consistent estimation of all parameters of the conditional expectation. Assumption A2 (b) only ensures the consistent estimation of the slope parameters (Lee, 1989), but the inconsistency of the intercept estimator is generally only a minor nuisance.

What A2 makes clear, however, is that under asymmetry consistent estimation of the slope parameters requires the statistical independence of ε_i and x_i, which rules

[8] Notice that for consistent estimation of the conditional mode the density of ε_i only needs to be symmetric around zero up to $\pm s$. However, this milder condition does not ensure that the conditional mode coincides with the conditional mean and therefore is it not enough to ensure consistent estimation of the conditional expectation.

out, for example, heteroskedasticity. Therefore, while A1 is possibly acceptable for most practitioners, A2 is clearly too strong to be generally accepted in econometric applications. This point will be pursued further in Section 4.

Whether or not the errors are symmetrically distributed, heteroskedasticity is also likely to affect the efficiency of the IRLS M-estimator relative to OLS. Although we present no formal results on this, the simulation evidence in section 3 clearly illustrates this point. A related consequence of the possible presence of heteroskedasticity is that it invalidates the estimator of the covariance matrix proposed by Street, Carroll and Ruppert (1988), which is generally used in practice. Therefore, the presence of heteroskedasticity greatly reduces the attractiveness of the IRLS M-estimator and, when coupled with skewed errors, it is likely to have devastating consequences.

Of course, if s is allowed to go to zero as the sample size passes to infinity, the properties of the IRLS M-estimator based on (2) are very different. In this case, it can be shown that the estimator is consistent for the conditional mode of y_i given x_i, even if the errors are skewed and heteroskedastic (see Kemp and Santos Silva, 2009). However, there are two points that are important to note. First, the estimator is consistent but its convergence is slower that the usual $\sqrt{n}$. Second, the conditional mode, although of interest in itself, does not generally coincide with the conditional mean and has very different properties.[9]

3. SIMULATION EVIDENCE

3.1. Simulation design

In this section, we perform a small simulation study to illustrate the performance of the IRLS M-estimator when the errors of the regression model are heteroskedastic and/or skewed. The design of the experiment is inspired by the classic study of

[9] For instance, the mean of a population can be obtained as the weighted average of the means of sub-populations, but the same is not true for the mode.

Arabmazar and Schmidt (1981). In particular, data is generated by the model

$$y_i = \beta_0 + \beta_1 x_i + k\left(1 + h x_i\right)\varepsilon_i, \qquad i = 1, ..., 500,$$

where x_i is a Bernoulli random variable with $\Pr\left(x_i = 0\right) = p$, ε_i is a random variable with zero mean and variance one, h is a parameter controlling the degree of heteroskedasticity, and k is set so that the population R^2 is one half.[10] Throughout, we set $\beta_0 = \beta_1 = 1$ and $p = 0.8$.

To explore the effects of heteroskedasticity, we perform simulations with $h \in \{-4/5, -2/3, 0, 2, 4\}$. Notice that, for g positive, the degree of heteroskedasticity is the same for $h = g$ and $h = -g/(g+1)$.[11] However, the two situations are quite different in that $h = g$ implies that the observations have the larger variance with probability $1 - p$, whereas when $h = -g/(g+1)$ the probability of the larger variance is p. Therefore, the designs with $h = g$ and $h = -g/(g+1)$ will have very different implications for the performance of the IRLS M-estimator. Indeed, when h is positive, this estimator tends to identify observations with $x_i = 1$ as influential and, consequently, downweights them. Therefore, the IRLS M-estimator will use very little information on the observations with $x_i = 1$, which by design are relatively rare. This makes the IRLS M-estimates of β_1 much more noisy when $h = g$ than when $h = -g/(g+1)$, because in the latter case the observations that are downweighted are those with $x_i = 0$.

To complete the design, it is necessary to define how ε_i is generated. We consider two cases. As it is standard in the analysis of the performance of M-estimators, we conduct a set of experiments in which ε_i is obtained from a contaminated normal. In particular, following Tukey (1960), we generate data such that, with probability $(1-\alpha)$, ε_i is drawn from a standard normal distribution and, with probability α, it is drawn from a normal distribution with zero mean and variance 9. In our experiments

[10] Specifically, $k = \sqrt{p(1-p)/\left[p + (1-p)(1+h)^2\right]}$.

[11] Arabmazar and Schmidt (1981) consider cases where the ratio between the larger and smaller variances goes up to 100. In our experiments, the maximum value for this ratio is 25.

we consider $\alpha \in \{0.00, 0.01, 0.05, 0.10\}$. The second set of experiments considers errors with different degrees of asymmetry. In particular, ε_i was generated from a $\chi^2_{(\nu)}$ distribution, with $\nu \in \{3, 6, 12, 24, 48\}$.[12] As mentioned above, in all experiments ε_i was centred and scaled so that it has zero mean and unit variance.

For each of the designs, y_i, x_i and ε_i were newly generated for each replication. All computations were performed using Stata (StataCorp., 2007), which has been used by most applied econometricians to implement the IRLS M-estimator.[13]

3.2. Simulation results

Tables 1 and 2 summarize the results obtained with 100.000 replications for each design point. To conserve space, we only report results for the more interesting parameter β_1. Specifically, for each design point, we report the mean of the estimates for β_1 obtained by OLS and the IRLS M-estimator, as well as the ratio of the variance of the OLS to that of the IRLS M-estimates, labelled variance ratio.

3.2.1. Homoskedastic errors

As expected, the results obtained with $h = 0$ confirm that under homoskedasticity the estimates for the slope parameter obtained with the IRLS M-estimator have means very close to 1, even for the heavily skewed $\chi^2_{(3)}$ errors. Moreover, the IRLS M-estimator has smaller variance that the OLS for distributions with reasonable excess-kurtosis, i.e., for $\alpha \in \{0.01, 0.05, 0.10\}$ in the first set of experiments and $\nu \in \{3, 6, 12\}$ in the second.

[12] The coefficient of skeness for the $\chi^2_{(\nu)}$ distribution is $\sqrt{8/\nu}$.

[13] The algorithm used in Stata (StataCorp., 2007), via the command *rreg,* is slightly different from the one described in subsection 2.2. above in that observations with Cook's (1977) distance larger than 1 are discarded after the initial OLS estimation (see Hamilton, 2008). However, with the particular design used in these experiments, that difference is immaterial.

Table 1: Results for β_1 with contaminated normal errors

		$h=-\frac{4}{5}$	$h=-\frac{2}{3}$	$h=0$	$h=2$	$h=4$
$\alpha=0.00$	OLS	1.00007	1.00009	1.00018	1.00028	1.00031
	IRLS	1.00004	1.00007	1.00021	1.00028	0.99966
	Variance ratio	0.86551	0.9117	0.95027	0.42778	0.17851
$\alpha=0.01$	OLS	1.00007	1.00009	1.00017	1.00026	1.00028
	IRLS	1.00004	1.00006	1.00019	1.00027	0.99978
	Variance ratio	0.92036	0.96476	1.00851	0.46143	0.19306
$\alpha=0.05$	OLS	1.00007	1.00008	1.00014	1.00022	1.00023
	IRLS	1.00003	1.00004	1.00016	1.00028	0.99999
	Variance ratio	1.12106	1.15442	1.21997	0.59467	0.25106
$\alpha=0.10$	OLS	1.00008	1.00010	1.00019	1.00031	1.00033
	IRLS	1.00002	1.00005	1.00014	1.00017	0.99991
	Variance ratio	1.31486	1.3243	1.42105	0.75468	0.32296

Therefore, under homoskedasticity, the IRLS M-estimator has clear advantages over OLS and these are the sort of results that have been used to advocate its use. However, the results obtained for $h \neq 0$ paint a very different picture.

3.2.2. Heteroskedastic symmetrical errors

For the experiments with the contaminated normal errors, we again find that the estimates of β_1 obtained with the IRLS M-estimator have means very close to 1, even when $h \neq 0$. However, the presence of heteroskedasticity has a detrimental effect on the performance of the IRLS M-estimator. For $h \in \{-4/5, -2/3\}$, the variance of this estimator is smaller than that of OLS only for $\alpha > 0.01$, but even in these cases the gains from the IRLS M-estimator are now smaller compared to the homoskedastic case. For positive h, however, the variance of the IRLS M-estimator is up to 5 times larger than that of the OLS estimator. Moreover, this advantage of the OLS is substantial even when there is noticeable excess-kurtosis.

Table 2: Results for β_1 with $\chi^2_{(\nu)}$ errors

		$h=-\frac{4}{5}$	$h=-\frac{2}{3}$	$h=0$	$h=2$	$h=4$
$\nu=3$	OLS	0.99997	0.99995	0.99986	0.99973	0.99970
	IRLS	1.10175	1.09147	1.00048	0.70863	0.55824
	Variance ratio	0.87080	0.99762	1.41690	1.32088	0.54299
$\nu=6$	OLS	1.00002	1.00003	1.00004	1.00005	1.00005
	IRLS	1.07640	1.06792	1.00051	0.76061	0.66009
	Variance ratio	0.83509	0.94588	1.14814	0.71333	0.24686
$\nu=12$	OLS	0.99988	0.99986	0.99975	0.99963	0.99961
	IRLS	1.05611	1.04900	1.00006	0.82090	0.75717
	Variance ratio	0.83725	0.93130	1.04387	0.52054	0.20013
$\nu=24$	OLS	1.00011	1.00012	1.00015	1.00017	1.00016
	IRLS	1.04084	1.03529	1.00040	0.87184	0.82916
	Variance ratio	0.84279	0.91717	0.99341	0.46127	0.18672
$\nu=48$	OLS	1.00000	1.00001	1.00001	1.00001	1.00001
	IRLS	1.02908	1.02499	1.00016	0.90917	0.87967
	Variance ratio	0.85442	0.91592	0.97345	0.44467	0.18342

3.2.3. Heteroskedastic skewed errors

With skewed errors the consequences of the heteroskedasticity are even more dramatic. First of all, with the $\chi^2_{(\nu)}$ errors, the variance of the IRLS M-estimator is larger than that of the OLS for all cases with $h \neq 0$, except when $h = 2$ and $\nu = 3$. Again, we find that for positive h, the variance of the IRLS M-estimator can be more than 5 times larger than that of OLS.

What is more serious, however, is that now the mean of the IRLS M-estimates are often quite different form 1. In particular, we observe that for $h < 0$ the estimator is biased upwards, with the reverse happening for $h > 0$. In this case, the bias of the IRLS M-estimator is particularly severe, being above 40% for $h = 4$ and $\nu = 3$. Even for the $\chi^2_{(48)}$ errors, which are almost symmetrical, the IRLS M-estimator can be severely biased in the presence of moderate heteroskedasticity.

It is important to notice that in the presence of heteroskedastic and skewed errors the IRLS M-estimator is inconsistent, not only for the parameters of the conditional expectation, but also for the parameters of the conditional median and mode. For example, for $h = 4$ and $\nu = 3$, the slope parameters for the conditional median and mode are 0.828 and 0.458, respectively, whereas the mean of the IRLS M-estimates is 0.558. Therefore, in the presence of heteroskedastic and skewed errors, the IRLS M-estimates can be extremely noisy and are, at best, difficult to interpret.

4. THE USE OF IRLS M-ESTIMATOR IN APPLIED ECONOMETRICS

The simulation results presented in the previous section confirm that the use of the IRLS M-estimator cannot be recommended when errors are even moderately skewed and heteroskedastic. However, skewness and heteroskedasticity are likely to be the norm, rather than the exception, in econometric applications.

The ubiquitous use of the Eicker-White standard errors (Eicker, 1963, 1967, White, 1980) suggests that in many applications the researcher is not willing to assume homoskedasticity. Moreover, the fact that in most econometric problems the variate of interest is non-negative suggests that skewness is also pervasive in this kind of applications. The widespread practice of logging the dependent variable can be seen as evidence that researchers often try to partially eliminate the skewness of the data. Of course, taking logs of the dependent variable not only makes it difficult to interpret the estimation results, but it also does not ensure that the resulting model has errors with a symmetrical distribution.[14]

[14]The work of Box and Cox (1964) is the leading reference in a vast literature on transformations of the dependent variable to achieve an approximately symmetrical distribution of the errors. In

Unfortunately, the literature on robust regression is essentially mute about the conditions for the consistency of the IRLS M-estimators of the conditional mean and most practitioners seem to be unaware of the devastating effects of heteroskedasticity and/or skewness on the properties of this estimator. Therefore, in econometric applications, the IRLS M-estimator is likely to be frequently used in situations where it is inconsistent for the parameters of interest and consequently it is not surprising to find that many authors report that it leads to estimates that are substantially different from those obtained with OLS. For example, in one of the earliest uses of the IRLS M-estimator in applied econometrics, Hall and Liebman (1998) report that the IRLS M-estimates are about 20 to 30 percent smaller (in absolute value) than the corresponding OLS estimates. Similarly, Strömberg (2004) finds that, depending on the sub-sample used, the IRLS M-estimates can be smaller or larger than the corresponding OLS estimates by about 50 percent. More recently, Deschênes and Greenstone (2007) report an OLS estimate of the parameter of interest of 1.29, and the corresponding IRLS M-estimate with a value of -0.24.

In view of the results presented before, these differences between the estimates obtained with the two methods, rather than suggesting some fragility of the OLS, strongly point to the inappropriate use of the IRLS M-estimator. Of course, to draw more substantiated conclusions it would be necessary to check for skewness and heteroskedasticity of the error terms. This is possible for the work by Deschênes and Greenstone (2007), who make available the data used in their study. For this particular data set, the two-degrees-of-freedom special case of White's (1980) test for homoskedasticity (see Wooldridge, 2002, p. 127) leads to a test statistic of 1454 and to a p-value virtually equal to zero. Furthermore, the coefficient of skewness of the OLS residuals is 2.46, which is extremely high and obviously statistically significant.

spite of this, skewness is rarely mentioned in econometric applications because it has little effect on the properties of the OLS estimator.

Therefore, in this particular application, it is clear that the IRLS M-estimator will be inconsistent for the parameters of the conditional mean.

Of course, when carefully used, the IRLS M-estimator may be a useful tool in applied econometrics. For example, it may serve to help identifying influential observations whose validity needs further investigation (see, e.g., Maronna, Martin and Yohai, 2006). However, its general use cannot be recommended and it certainly should not be used as an alternative to OLS.

5. CONCLUDING REMARKS

What distinguishes econometrics from other fields of applied statistics is the nature of the data that it uses. In particular, the fact that experimental data is rarely available means that most of the work econometricians do involves the study of identifiable characteristics of conditional distributions. Moreover, given the nature of the phenomena studied by econometricians, these conditional distributions are typically heteroskedastic and skewed. This is in stark contrast with other areas of applied statistics that rely heavily on the use of experimental data for which it may be reasonable to assume that the error terms are symmetrically distributed and independent of the covariates.

In any application, the statistical methods to use have to suit the characteristics of the data that are available. Thus, the distinctive nature of economic data implies that econometricians have to be cautious when adopting methods developed in other areas of statistics. The IRLS M-estimator has a long and well justified tradition of successful application in different areas of statistics. However, this is no guarantee that this particular estimator can also be generally useful in econometrics. On the contrary, in typical econometric problems the IRLS M-estimates are, at best, difficult to interpret and can be very misleading.

This is perhaps why most modern textbooks in econometrics completely ignore the IRLS M-estimator. However, by ignoring it, these textbooks also fail to alert practitioners to the pitfalls of the use of this estimator in econometric applications. This lack of information on the potential drawbacks of the estimator, coupled with the "robustness" label that is often attached to it and with its ready availability in popular software packages, helps to understand the recent rise in popularity of the IRLS M-estimator among applied econometricians. Unfortunately, this estimator is not generally appropriate in this context and all the econometric results based on it are highly questionable.

REFERENCES

Alpizar, F., Carlsson, F. and Johansson-Stenman, O. (2008). "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica," *Journal of Public Economics*, 92, 1047-1060.

Andrews D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). *Robust Estimates of Location: A Survey and Advances.* Princeton (NJ): Princeton University Press.

Andersen, J.J. and Aslaksen, S. (2008). "Constitutions and the Resource Curse," *Journal of Development Economics*, 87, 227-246.

Arabmazar, A. and Schmidt, P. (1981). "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity," *Journal of Econometrics*, 17, 253-258.

Baker, G. and Hall, B. (2004). "CEO Incentives and Firm Size," *Journal of Labor Economics*, 22, 767-798.

Beaton, A.E. and Tukey, J.W. (1974). "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," *Technometrics*, 16, 146-185.

Box, G.E.P. and Cox, D.R. (1964). "An Analysis of Transformations," *Journal of the Royal Statistical Society B*, 26, 211-243.

Chan, K.S., Godby, R., Mestelman, S. and Muller, R.A. (2002). "Crowding-out voluntary contributions to public goods," *Journal of Economic Behavior & Organization*, 48, 305-317.

Cook, R.D. (1977). "Detection of Influential Observations in Linear Regression," *Technometrics*, 19, 15-18.

Croxson, B., Propper, C. and Perkins, A. (2001). "Do Doctors Respond to Financial Incentives? UK Family Doctors and the GP Fundholder Scheme," *Journal of Public Economics*, 79, 375-398.

Currie, J. and Fahr, J. (2004). "Hospitals, Managed Care, and the Charity Caseload in California," *Journal of Health Economics*, 23, 421-442.

Deschênes, O. and Greenstone, M. (2007). "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather," *American Economic Review*, 97, 354-385.

Eicker, F. (1963). "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *The Annals of Mathematical Statistics*, 34, 447-456.

Eicker, F. (1967). "Limit Theorems for Regression with Unequal and Dependent Regressors," in: LeCam, L.M. and Neyman, J., *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley (CA): University of California, 59-82.

European Commission (2007). *Case No COMP/M.4439 – Ryanair/Aer Lingus*, available at:
http://ec.europa.eu/competition/mergers/cases/index/m88.html#m_4439.

European Commission (2008). *Case No COMP/M.4919 – StatoilHydro/ConocoPhillips*, available at:
http://ec.europa.eu/competition/mergers/cases/index/m98.html#m_4919.

Freund, C. and Bolaky, B. (2008). "Trade, Regulations, and Income," *Journal of Development Economics*, 87, 309-321.

Goldberger, A. (1991). *A Course in Econometrics.* Cambridge (MA): Harvard University Press.

Hall, B.J. and Liebman, J.B. (1998). "Are CEOS Really Paid Like Bureaucrats?," The Quarterly Journal of Economics, 113, 653-691.

Hamilton, L.C. (1991). "How robust is robust regression?," *Stata Technical Bulletin*, 2, 21-26.

Hamilton, L.C. (2008). *Statistics with Stata: Updated for Version 10.* Belmont (CA): Thomson Brooks/Cole.

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

Heiberger, R.M. and Becker, R.A. (1992). "Design of an S Function for Robust Regression Using Iteratively Reweighted Least Squares," *Journal of Computational and Graphical Statistics*, 1, 181-196.

Hogg, R.V. (1979). "Statistical Robustness: One View of its Use in Applications Today," *The American Statistician*, 33, 108-115.

Holland, P.W. and Welsch, R.E. (1977). "Robust Regression Using Iteratively Reweighted Least-Squares." *Communications in Statistics: Theory and Methods*, A6, 813-827.

Huber, P.J. (1964). "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P.J. (1973). "Robust Regression: Asymptotics, Conjectures and Monte Carlo." *Annals of Statistics*, 1, 799-821.

Huber, P.J. (1981). *Robust Statistics.* New York (NY): John Wiley & Sons.

Kemp, G.C.R. and Santos Silva, J.M.C. (2009). Smooth Mode Regression, mimeo, University of Essex.

Koenker, R. and Bassett Jr., G.S. (1978). "Regression Quantiles," *Econometrica*, 46, 33-50.

Lee, M.J. (1989). "Mode regression," *Journal of Econometrics*, 42, 337-349.

Lee, M.J. (1993). "Quadratic mode regression," *Journal of Econometrics*, 57, 1-19.

Lee, M.J. and Kim, H.J. (1998). "Semiparametric Econometric Estimators for a Truncated Regression Model: A review with an extension," *Statistica Neerlandica*, 52, 200-225.

Li, G. (1985). "Robust Regression," in Hoaglin, D.C, Mosteller, F. and Tukey, J.W. (eds.), *Exploring Data Tables, Trends and Shapes.* New York (NY): John Wiley & Sons, 281-340.

Maronna, R.A., Martin R.D. and Yohai, V.J., (2006). *Robust Statistics: Theory and Methods*, Chichester: John Wiley & Sons.

Mathworks. (2008). *Statistics Toolbox User's Guide, Version 7.* Natick (MA): The Mathworks Inc.

Peracchi, F. (2001). *Econometrics.* Chichester: John Wiley & Sons.

Rogers, M.L. (2008). "Directly Unproductive Schooling: How Country Characteristics Affect the Impact of Schooling on Growth," *European Economic Review*, 52, 356–385.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* New York (NY): John Wiley & Sons.

Ruppert, D. and Carroll, R.J. (1980). "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828-838.

SAS Institute Inc. (2008), *SAS/STAT 9.2 User's Guide*, Cary (NC): SAS Institute Inc.

StataCorp. (2007). *Statistical Software: Release 10.0.* College Station (TX): Stata Corporation.

Street, J.O., Carroll, R.J. and Ruppert, D. (1988). "A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares," *American Statistician*, 42, 152-154.

Strömberg, D. (2004). "Radio's Impact on Public Spending," *The Quarterly Journal of Economics*, 119, 189-221.

Tukey, J.W. (1960). "A Survey of Sampling from Contaminated Distributions," in Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G. and Mann, H.B., (eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford (CA): Stanford Univiversity Press, 448-485.

Venables, W.N. and Ripley, B. D. (2002). *Modern Applied Statistics with S, 4th ed.*, New York (NY): Springer.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, London: Chapman & Hall.

White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Wu, L.L. (1985). "Robust M-Estimation of Location and Regression," *Sociological Methodology*, 15, 316-388.