ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 664 January 2009
(Revised February 2011)

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‡

Revised on February 22, 2011

Abstract

The use of robust regression has gained popularity among applied econometricians. Unfortunately, most practitioners who have used these estimators seem to be unaware of the fact that their properties can be dramatically affected by both heteroskedasticity and skewness of the errors. In this paper we reconsider the interpretation of a specific robust regression estimator that has become popular in applied econometrics, and conclude that its use in this context cannot be generally recommended. Alternatively, quantile and mode regression could be used when the researcher wants to estimate conditional location functions that are robust to the presence of outliers.

JEL classification code: C13, C21.

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

*We are grateful to Marcus Chambers, Geert Dhaene, Gordon Kemp, Myoung-jae Lee, José Machado, Paulo Parente, Jon Temple, and Silvana Tenreyro for helpful comments and advice. We are especially grateful to David Strömberg for generously making available the data used in Section 4. Andrew Muller and Stuart Mestelman also kindly provided data used at earlier stages of the preparation of this paper. The usual disclaimer applies. Finally, we are grateful to RBB Economics for providing the stimulating environment that originated this research, and Santos Silva gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

†Stanford University. Email: baldauf@stanford.edu.

‡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 errors with heavy-tailed distributions.[1] 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.[2]

Nevertheless, over the past decade, a form of robust regression based on Huber's (1964) M-estimator was made available in popular software packages[3] and has been frequently used both in leading research publications and in industry.[4] The particular version of this estimator that has become popular in applied econometrics is based on

[1] For example, in an often-cited book, Hamilton (2008, p. 253) states: "Robust regression methods aim to achieve almost the efficiency of OLS with ideal data and substantially better-than-OLS efficiency in non-ideal (for example, nonnormal errors) situations".

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

[3] This robust regression estimator is available, for example, in Stata via the command *rreg* (StataCorp., 2009), 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).

[4] For examples of top academic publications using this M-estimator see, among many others, Alpizar, Carlsson and Johansson-Stenman (2008), Andersen and Aslaksen (2008), Baker and Hall (2004), Chan, Godby, Mestelman and Muller (2002), Crinò (2010), Croxson, Propper and Perkins (2001), Currie and Fahr (2004), Deschênes and Greenstone (2007), Freund and Bolaky (2008), Lang and Kahn (1998), Rogers (2008), 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.

the algorithm proposed by Li (1985), which is an iteratively reweighted least squares algorithm using biweights (Beaton and Tukey, 1974).

However, perhaps because of the lack of appropriate references on its use in econometrics, most practitioners who have used this estimator seem to be unaware of the fact that its properties depend on strong assumptions about the symmetry and homoskedasticity of the errors, and justify its use with misleading claims about its advantages.

In this paper we discuss the interpretation of the specific robust M-estimator that has become popular in applied econometrics, henceforth termed BWM-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 BWM-estimator, our results extend to other robust regression estimators as it is illustrated both in the simulations and in the empirical application we present.

The paper is organized as follows. In the next section we describe in detail the version of Huber's (1964) M-estimator that has been used in applied econometrics, and discuss its interpretation. Section 3 presents the results of simulation studies illustrating the pitfalls of using robust regression estimators when the errors are heteroskedastic and/or skewed. Section 4 revisits the study of Strömberg (2004) on the relation between mass media and public spending, and illustrates the importance of defining the location measure of interest and using a suitable estimator for it. Finally, Section 5 presents brief concluding remarks.

[5] We use this terminology in reference to the use of biweights and to distinguish this particular estimator from other M-estimators.

2. THE 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 a 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 conditional distribution of y is being 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, typically, 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.

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 $\mathrm{E}(\varepsilon_i|x_i) = 0$.

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

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

where δ is a scale parameter, and $\rho(\cdot)$ is an even function that is non-decreasing in the positive half-line. Generally speaking, $\hat{\beta}$ will be an estimate of the parameters of some location function of the conditional distribution of y given x and its properties will naturally depend of the particular form of $\rho(\cdot)$ that is adopted. 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.

The choice of $\rho(\cdot)$ is often based on robustness and computational considerations (see, e.g., Li, 1985). However, 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 (see, e.g., Li, 1985), but, more importantly, the interpretation of the estimates. Consequently, in order to be able to interpret the robust regression results that have appeared in the literature, it is now important to study in detail the particular algorithm used to compute them.

2.2. Li's algorithm

The algorithm described in Li (1985, pp. 335-6) has been used in virtually all econometric applications of the M-estimator. This algorithm starts with an OLS regression and proceeds with a set of iterations using weighted least squares regressions. These iterations use Huber (1964) weights of the form

$$w_i^H = \begin{cases} 1 & \text{if } |y_i - x_i'b| \leq s \times c_H \\ \frac{s \times c_H}{|y_i - x_i'b|} & \text{otherwise} \end{cases}, \tag{2}$$

where c_H is a tuning constant and s is proportional to the median absolute deviation defined as $\text{MAD} = \operatorname*{med}_i \left\{ \left| (y_i - x_i'b) - \operatorname*{med}_j \left(y_j - x_j'b \right) \right| \right\}$, where b is evaluated at the current estimate of β.[6] The purpose 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 is performed, this time using biweights (Beaton and Tukey, 1974) of the form

$$w_i^B = \begin{cases} \left[1 - \left(\frac{y_i - x_i'b}{s \times c_B}\right)^2\right]^2 & \text{if } |y_i - x_i'b| \leq s \times c_B \\ 0 & \text{otherwise} \end{cases},$$

where c_B is a new tuning constant and s is defined as before.[7] The use of this re-weighted least squares algorithm with biweights is equivalent to finding the vector b that minimizes the objective function

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

where $\boldsymbol{I}[e]$ is the indicator function for event e (see Li, 1985, p. 293).[8]

To gain further insight into this estimator, it is interesting to notice that minimizing (3) is equivalent to maximizing

$$\frac{1}{n\delta} \sum_{i=1}^{n} K_T\left(\frac{y_i - x_i'b}{\delta} \right), \tag{4}$$

where $\delta = s \times c_B$ and $K_T(u) = \frac{35}{32}\boldsymbol{I}[|u| \leq 1](1 - u^2)^3$ is the triweight kernel (see, e.g., Wand and Jones, 1995). Expression (4) is immediately recognizable as a non-parametric estimator of the density of y_i at $x_i'b$. Therefore, under appropriate reg-

[6] In all computations in sections 3 and 4, c_H is set to 1.349, and s is set to MAD/0.6745. These choices, which affect the properties the estimator, are the default in the Stata (StataCorp., 2009) command *rreg*, which was used in all the empirical applications we refer to.

[7] In all computations in sections 3 and 4, c_B is set to 4.685, which is also the default in *rreg* (StataCorp., 2009).

[8] The objective function defined by (3) can have multiple minima and that is why it is important to have good starting values and the first set of iterations is needed.

ularity conditions, the value of b that maximizes (4) 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 (3) 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 $\rho(\cdot) = a_1 - a_2 K(\cdot)$, where a_1 and $a_2 > 0$ are constants and $K(\cdot)$ is a kernel function such that $\int K(z)\,dz = 1$ and $\lim_{z\to\pm\infty} K(z) = 0$. The link between mode regression and the M-estimator defined by (3) is convenient because the conditions 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 M-estimator based on biweights

As in Lee (1989, 1993), the BWM-estimator suggested by Li (1985) treats δ as a fixed parameter. That is, δ 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 δ, the sufficient conditions for the estimator based on (3) 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 conditional density of ε_i is strictly unimodal with a finite mode at zero;

A2: Either of the following conditions holds:

(a) the conditional density of ε_i is symmetric around zero;[9]

(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 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. Indeed, 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.[10]

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

[10]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

Even if the errors are symmetrically distributed, heteroskedasticity is also likely to affect the efficiency of the BWM-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 (see Croux, Dhaene and Hoorelbeke, 2003). Therefore, the presence of heteroskedasticity greatly reduces the attractiveness of the BWM-estimator and, when coupled with skewed errors, it is likely to have devastating consequences.

Of course, if δ is allowed to go to zero as the sample size passes to infinity, the properties of the BWM-estimator based on (3) are very different. In this case, under suitable regularity conditions, 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, 2010). However, it is important to note that, although of interest in itself, the conditional mode does not generally coincide with the conditional mean and has very different properties.[11]

3. SIMULATION EVIDENCE

In this section, we perform two simulation studies to illustrate the performance of robust regression estimators when the errors of the regression model are heteroskedastic and/or skewed. The first set of experiments considers in detail the relative performance of the OLS and the BWM-estimator. The second set of experiments illustrates that the sensitivity of the BWM-estimator to skewness and heteroskedasticity extends to other robust regression estimators.

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

[11] 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.

3.1. Simulation design

The design of the experiments is inspired by the classic study of Arabmazar and Schmidt (1981). In particular, data are generated by the model

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

where x_i is a Bernoulli random variable with $\Pr(x_i = 0) = p$, 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.[12] 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 $h = -g/(g+1)$.[13] 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 estimators.

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 robust estimators, we conduct some 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 we consider $\alpha \in \{0.00, 0.01, 0.05, 0.10\}$. The second case we study considers errors with different degrees of asymmetry. Specifically, ε_i is generated from a $\chi^2_{(\nu)}$

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

[13] 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.

distribution, with $\nu \in \{3, 6, 12, 24, 48\}$.[14] As mentioned above, in all experiments ε_i is 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., 2009), which has been used by most applied econometricians to implement the BWM-estimator.[15]

3.2. Main simulation results

We start by considering the relative performance of the OLS and the BWM-estimator. 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 of β_1 obtained with the OLS and the BWM-estimator, as well as the ratio of the variance of the OLS to that of the BWM-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 BWM-estimator have means very close to 1, even for the heavily skewed $\chi^2_{(3)}$ errors. Moreover, the BWM-estimator has a smaller variance than the OLS for distributions with reasonable excess-kurtosis, i.e., for $\alpha \in \{0.01, 0.05, 0.10\}$ when the errors are generated as normal mixtures, and $\nu \in \{3, 6, 12\}$ when the errors have a $\chi^2_{(\nu)}$ distribution.

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

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

[15] The algorithm used in Stata (StataCorp., 2009), 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
	BWM	1.00004	1.00007	1.00021	1.00028	0.99966
	Variance ratio	0.86551	0.91170	0.95027	0.42778	0.17851
$\alpha = 0.01$	OLS	1.00007	1.00009	1.00017	1.00026	1.00028
	BWM	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
	BWM	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
	BWM	1.00002	1.00005	1.00014	1.00017	0.99991
	Variance ratio	1.31486	1.32430	1.42105	0.75468	0.32296

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 BWM-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 BWM-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 BWM-estimator are now smaller than in the homoskedastic case. For positive h, however, the variance of the BWM-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
	BWM	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
	BWM	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
	BWM	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
	BWM	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
	BWM	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 BWM-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 BWM-estimator can be more than 5 times larger than that of OLS.

What is more serious, however, is that now the means of the BWM-estimates of β_1 are often quite different from 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

BWM-estimator is particularly severe, e.g. in excess of -40% for $h = 4$ and $\nu = 3$. Even for the $\chi^2_{(48)}$ errors, which are almost symmetrical, the BWM-estimator can be severely biased in the presence of moderate heteroskedasticity.

3.3. Results with other robust estimators

The sensitivity of the BWM-estimator to skewness and heteroskedasticity extends to other robust estimators. This is obvious for estimators that explicitly depend on symmetry, such as the Gastwirth and trimean estimators introduced by Koenker and Bassett (1978), and the trimmed least squares estimator of Ruppert and Carroll (1980). However, more modern robust estimators are also sensitive to departures from the assumption of symmetric homoskedastic errors.

To illustrate this, we used the design described before to perform a small scale simulation experiment based only on $5,000$ replications, where we also studied the estimators implemented in Stata (StataCorp., 2009) by Verardi and Croux (2009) and by Jann (2010). The additional estimators considered are an M-estimator using Huber weights as defined in (2), the S-estimator of Rousseeuw and Yohai (1987), the MM-estimator of Yohai (1987), and the least median of squares and least trimmed squares of Rousseeuw (1984).[16]

In the interest of space, we do not report in detail the results of these experiments, which are available on request, but provide a brief overview of our findings. For homoskedastic errors, as expected, the means of the estimates of β_1 are always very close to 1 and, again, the robust estimators can be more efficient than OLS for high-kurtosis distributions. However, some robust estimators, notably the least median of squares, the least trimmed squares, and the S-estimators, are substantially less efficient that OLS for all the designs considered. With heteroskedastic symmetrical errors, the means of the estimates are again generally very close to 1, the exceptions

[16] To our knowledge, of these estimators, only the least trimmed squares has been used in economics (see, e.g., Temple, 1998, and Zaman, Rousseeuw and Orhan, 2001).

being the S- and MM-estimators as implemented by Verardi and Croux (2009), for which the mean of the estimates of β_1 can be as low as 0.941 when h is positive. Like in the main set of experiments, we find that the efficiency of the robust estimators is greatly affected by heteroskedasticity, especially when h is positive. For example, for $h = 4$ and $\alpha = 0.10$, only the M-estimator using Huber weights is more efficient than the OLS. Finally, for heteroskedastic skewed errors, the mean of the estimates of β_1 is very different from 1 for all robust estimators. Specifically, the means of β_1 can be substantially above 1 for $h < 0$, or substantially below 1 for $h > 0$.

This set of results confirms that, when the distribution of the errors is skewed and heteroskedastic, the so-called robust estimators do not identify the parameters of the conditional mean. Moreover, these estimators are also inconsistent for the parameters of the conditional median and mode. For example, for $\chi^2_{(3)}$ errors with $h = 4$, the slope parameters for the conditional median and mode are 0.828 and 0.458, respectively, whereas the mean of the BWM-estimates is 0.558. Therefore, unless very strong assumptions are made about the shape of the conditional density of the variate of interest, robust estimators like the BWM-estimator do not identify the parameters of any well understood measure of central tendency, making them, at best, difficult to interpret.

4. RADIO'S IMPACT ON PUBLIC SPENDING REVISITED

To illustrate the critical importance of defining the location function that is of interest and of choosing an appropriate estimator for it, we revisit the recent study by Strömberg (2004) on how mass media influences policy-making and public spending.

Strömberg (2004) develops an economic model that yields three testable implications: 1) government spending is higher on groups with better access to mass media, 2) government spending is higher on groups where a high percentage of people vote, and 3) turnout is higher in groups where many have access to media.

To test the first two implications of the model, Strömberg (2004) uses data on a cross-section of U.S. counties to study how radio penetration affected the distribution of funds in an important New Deal program providing unemployment relief between 1933 and 1935. The econometric model estimated by Strömberg (2004) has the form

$$\ln(z_c) = c_1 \ln(r_c) + c_2 \ln(t_c) + x_c'\beta + \mu_s + \varepsilon_c,$$

where z_c is the per capita cumulative spending in the program from April 1933 to December 1935 in county c, r_c is the share of households in county c with a radio, t_c denotes voter turnout in gubernatorial elections in county c, x_c is a set of control variables, μ_s are state fixed-effects, and ε_c is an error term uncorrelated with the regressors. According to the model, $c_1 > 0$ and $0 < c_2 < 1$.

The benchmark results of Strömberg (2004) are obtained by estimating the model by OLS using a sample with data from 2492 counties.[17] Table 3 provides a brief description of the variables used; for further details on the data, including a description of the sources, see Strömberg (2004). The first column of Table 4 replicates the benchmark estimates presented in Table II in Strömberg (2004, p. 206). In accordance with the predictions of the model, the estimate of the coefficient associated with the log of the share of radios per household his positive and the coefficient associated with the log of voter turnout is between zero and one.

As part of the checks to assess the robustness of these results, Strömberg (2004) also estimated the model using the BWM-estimator, whose results are presented in the second column of Table 4. Although the results obtained with the two estimators are generally close, there are a few instances where the differences are substantial, including the coefficient of one of the main variables of interest, $\ln(r_c)$, whose effect is halved when the BWM-estimator is used.[18]

[17] The author also considers instrumental variables estimation, but focuses on the least squares results when discussing the effects of radio on spending.

[18] In the interest of space, we do not report in detail the results obtained with other robust estimators, but for completeness we note that the estimated coefficients of $\ln(r_c)$ (and corresponding

Table 3: Description of the variables

z_c : Spending	Per capita cumulative disbursement within the program from April 1933 to December 1935
r_c : Radios	Share of families reporting radio sets in 1930
t_c : Turnout	Voter turnout in gubernatorial elections
Share illiterate	Share of persons ten years of age and over who are illiterate in 1930
School enrolment	Share of persons 7–18 years of age attending school in 1930
Marginal voter	Standard deviation of the county democratic vote shares in gubernatorial elections, 1922–1932
Unemp. 1930	Unemployment rate in 1930
Unemp. 1937	Unemployment rate in 1937
Bank deposits	Bank deposits per capita in 1934
%Δbank deposits	Percentage change in bank deposits per capita between 1930 and 1934.
Dwelling value	Median value of owner-occupied dwelling units in 1930
Farm value	Per capita value of farm buildings in 1930
Retail wage	Average wage in retail establishments in 1930
Crop value	Per capita value of all crops harvested in 1929
Rent	Median monthly rent of tenant-occupied dwelling units in 1930
Share 21+	Share of persons 21 years of age or older in 1930
Share 65+	Share of persons 65 years of age or older in 1930
Females	Percentage of females in 1930
Blacks	Percentage of African-Americans in 1930
Immigrants	Percentage of foreign-born white persons in 1930
Partisans	Share of voters who voted for the winning gubernatorial candidate
Urban	Share of urban population in 1930
Rural	1 for counties where share urban equals zero, 0 otherwise
Gas sales	Per capita sales of filling stations in 1934
Pop. density	Population per square mile in 1930
Population	0.6×population 1930+0.4×population 1940

t-statistics) obtained with the the M-estimator using Huber weights as defined in (2), the S-estimator of Rousseeuw and Yohai (1987) and the MS-estimator of Maronna and Yohai (2000) are, respectively, 0.082 (2.396), 0.015 (0.284), and 0.003 (0.067).

In view of the results presented before, it is interesting to investigate whether the difference between the estimates of the coefficient of $\ln(r_c)$ obtained with the two methods is the result of the OLS sensitivity to influential observations, or rather the consequence of applying the BWM-estimator in a situation where it does not identify the parameters of the conditional mean.

We start by checking whether the OLS results are critically affected by influential observations. Figure 1 displays the plot of the usual leverage indicator against the Studentized least squares residuals (see, e.g., Cook and Weisberg, 1982). This plot shows that there are both several high leverage points and some large residuals, especially in the left tail of the distribution. However, the observations with high leverage have reasonably small residuals, and the observations with large residuals tend to have little leverage, suggesting that none of the observations is particularly influent. To confirm this, we computed the percentage change of the fitted value of $\ln(z_c)$ that would result from deleting each single observation, and reestimated the model excluding the set of five observations whose deletion would lead to changes of over 5 percent of the fitted value. The estimates for the two main parameters of interest changed by less than 7 percent (less than 0.2 of a standard error), confirming that no single observation is particularly influential.

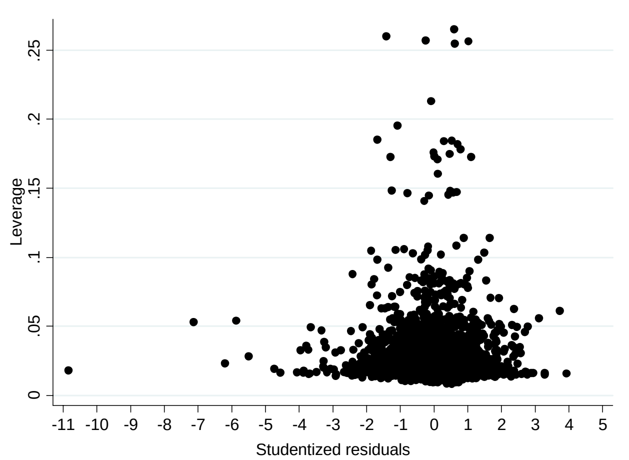

Figure 1: Leverage versus Studentized residual plot for the least squares regression.

Although these results suggest that the least squares results are not driven by a handful of influential observations, this kind of single-case diagnostics may not reveal the role of a group of influential observations masking each other's effects. Because there is no well-established method to deal with the possible masking effects of groups of influential observations, we developed a very simple bootstrap procedure to investigate this problem. In particular, we looked at the distribution of the estimates of the coefficient of $\ln(r_c)$ obtained using pairs-bootstrap. If all observations of the sample come from a common population, the pairs-bootstrap estimates of the coefficient of interest should have a distribution close to a normal with mean and standard error close to the OLS estimates obtained with the full sample. On the contrary, if there is a group of influential observations, the distribution of the pairs-bootstrap estimates of the coefficient of $\ln(r_c)$ should be bimodal, reflecting the fact that in bootstrap samples that exclude the influential observations the estimate of this parameter is substantially different from the result obtained with the full sample. Moreover, in this case, the mean of the bootstrap estimates should be between the OLS and BWM-estimates obtained with the full sample.

In this particular application we found that a standard test for the normality of the estimates of the coefficient of $\ln(r_c)$ in $5,000$ pair-bootstrap replicas has a p-value of 0.767. Moreover, the mean of these estimates is equal to 0.138, and its standard error is 0.037. These results are remarkably close to the OLS estimates obtained with the full sample and therefore suggest that indeed influential observations are not an issue.

Turning now to the conditions for the BWM-estimator to be consistent for the parameters of the conditional mean, we investigate the skewness and heteroskedasticity of the conditional distribution of $\ln(z_c)$. For this particular model and sample, the statistic for the non-normality-robust symmetry test of Godfrey and Orme (1991) has a value of 5.21, to which corresponds a p-value of 0.02, thus confirming the significance of the skewness revealed by Figure 1. As for heteroskedasticity, we use the non-normality-robust version of the Breusch and Pagan (1979) test proposed by Koenker (1981) to check whether the variance of ε_c depends on $\ln(r_c)$, the regressor of interest whose coefficient is more sensitive to the choice of estimator. The test statistic in this case has a value of 22.26, to which corresponds a p-value virtually equal to zero.[19] Therefore, in this particular application, there are signs of skewness and very strong heteroskedasticity, and therefore the OLS and the BWM-estimator are likely to identify measures of central tendency with different slope parameters.

The differences between the results of the OLS and BWM-estimator suggest that, even focusing only on measures of central tendency, one regression is not enough to understand the effect of radio penetration on the program expenditures. To further explore this issue, it is interesting to see how $\ln(r_c)$ and $\ln(t_c)$ affect the conditional median and the conditional mode, two measures of central tendency known for their robustness properties.

The last two columns of Table 4 display the estimation results for the conditional median and mode, respectively.[20] As noted by Strömberg (2004), the conditional median estimates (labelled LAD in Table 4) are remarkably close to those of the BWM-estimator, although the coefficient of $\ln(r_c)$ is not statistically significant in the conditional median. Therefore, although that is not a general rule, it looks like in this particular case the measure of location identified by the BWM-estimator is close to the conditional median.

The results obtained for the mode are, however, substantially different from those obtained for the other conditional locations measures. In particular, not only the estimated coefficients of several important control variables (e.g., School enrolment,

[19] Under the null, both test statistics are asymptotically distributed as $\chi^2_{(1)}$ variates.

[20] The conditional mode was estimated using the method described in Kemp and Santos Silva (2010), with smoothing parameter equal to $1.05\text{MAD}n^{-0.143}$, where as before MAD denotes the median absolute deviation of the residuals. Comparable results can be obtained with the *rreg* command in Stata (StataCorp., 2009) by using a tuning constant 7 times smaller that the default value.

Table 4: Estimation results

	OLS	BWM	LAD	Mode
ln (Radios)	0.138 (3.796)	0.068 (2.219)	0.066 (1.652)	-0.072 (2.397)
ln (Turnout)	0.165 (4.779)	0.133 (3.883)	0.129 (2.947)	0.115 (2.082)
Share illiterate	-1.111 (2.198)	-1.170 (2.453)	-1.309 (2.314)	-1.675 (3.617)
School enrolment	0.856 (2.817)	0.766 (2.853)	0.505 (1.537)	-1.034 (2.733)
Marginal voter	0.034 (0.129)	0.288 (1.288)	0.171 (0.560)	-0.490 (1.102)
Unemp. 1930	7.837 (4.506)	7.848 (5.109)	9.018 (4.405)	10.723 (6.258)
Unemp. 1937	9.750 (12.87)	9.706 (14.59)	9.872 (11.94)	11.380 (13.90)
Bank deposits	-0.093 (5.271)	-0.064 (4.286)	-0.081 (3.874)	0.005 (0.298)
%Δbank deposits	-0.013 (1.177)	-0.008 (0.507)	-0.008 (0.901)	0.075 (4.053)
Dwelling value	0.000 (0.009)	0.034 (0.881)	0.060 (1.234)	-0.073 (1.210)
Farm value	-0.144 (4.028)	-0.141 (4.982)	-0.106 (2.471)	-0.125 (3.652)
Retail wage	0.016 (0.181)	0.033 (0.408)	-0.021 (0.199)	0.588 (3.861)
Crop value	0.017 (0.710)	0.014 (0.761)	-0.002 (0.069)	0.018 (0.950)
Rent	-0.063 (1.052)	-0.086 (1.544)	-0.108 (1.548)	0.018 (0.352)
Share 21+	-1.908 (3.742)	-0.994 (2.428)	-0.937 (1.645)	1.908 (5.192)
Share 65+	-2.181 (1.323)	-2.854 (2.075)	-3.079 (1.683)	-8.367 (6.672)
Females	1.923 (1.713)	2.389 (2.889)	2.333 (2.165)	5.470 (8.393)
Blacks	0.105 (0.950)	-0.007 (0.072)	0.005 (0.046)	-0.291 (2.990)
Immigrants	0.319 (0.853)	0.772 (2.762)	0.639 (1.505)	1.527 (5.292)
Partisans	0.052 (0.438)	0.041 (0.439)	0.017 (0.156)	-0.066 (0.550)
Urban	0.994 (8.972)	0.861 (9.855)	0.930 (7.123)	0.217 (2.099)
Rural	0.253 (7.837)	0.203 (7.455)	0.203 (6.038)	0.181 (5.672)
Gas sales	0.015 (0.863)	0.021 (1.393)	0.012 (0.570)	0.030 (1.484)
Pop. density	-0.064 (2.592)	-0.052 (2.753)	-0.052 (2.003)	0.027 (1.178)
Population	-0.092 (3.462)	-0.127 (6.004)	-0.134 (4.465)	-0.222 (10.51)
Intercept	4.807 (4.314)	3.969 (4.981)	4.691 (3.655)	-2.000 (1.941)

Dependent variable is log of cumulative spending per capita from 1933 to 1935; t statistics in parentheses: robust for OLS, LAD, and Mode, standard for BWM; results based on 2492 observations; all regressions include state dummies.

Bank deposits, %Δbank deposits, Share 21+, Share 65+, Females, Blacks, and Immigrants) are substantially different in the mode regression, but more importantly the estimate of the coefficient of $\ln(r_c)$, one of the main regressors of interest, is now negative and statistically significant. Although at first sight it may be surprising to find that a regressor has coefficients with opposite signs in two conditional measures of central tendency, this is indeed entirely possible as a result of heterocliticity.[21]

Therefore, in this application, by changing the measure of central tendency that is estimated it is possible to make the effect of $\ln(r_c)$ on spending to go all the way from positive and significant to negative and significant. This clearly illustrates that, as noted by Portnoy and Welsh (1992), it is important to define which location function of the distribution is of interest because that determines the estimator to use, and the results obtained may depend critically on this choice. In the case of the model considered by Strömberg (2004), the maintained assumption is that ε_c is uncorrelated with the regressors, and therefore the functional of interest is either the conditional mean or a linear approximation to it. Consequently, OLS is the appropriate estimator for the parameters of the model proposed by Strömberg (2004).

Nevertheless, even if interest is mainly focused on the mean regression results, the estimation of other location measures of the conditional distribution provided additional information that enriched our understanding of the effects of radio penetration on the distribution of funds in the program being considered. Taking into account the results of all the estimators, it appears that an increase in radio penetration does not shift the conditional distribution of spending upwards, but rather that it changes its shape in complex ways so that the mean shifts upwards, as found by Strömberg (2004), but the mode moves in the opposite direction. Therefore, the positive effect of media access on public spending is not uniform, and the bulk of the counties may

[21] Kemp and Santos Silva (2010) note that it is possible that a regressor has a positive effect in all quantiles, but a negative effect on the mode. Indeed, that seems to be the case in one of the empirical examples they consider.

not benefit at all from an increased access to media. This heterogeneous effect of information on public spending raises interesting questions and deserves additional scrutiny, both theoretical and empirical. That endeavour is, however, beyond the scope of this paper.

5. CONCLUDING REMARKS

The BWM-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, the results presented in Sections 3 and 4 show that, in typical econometric problems, the BWM-estimates are difficult to interpret and can be very misleading. Therefore, the use of the BWM-estimator in econometrics cannot be generally recommended, and it certainly should not be used as an alternative to OLS.

This is perhaps why most modern textbooks in econometrics completely ignore the BWM-estimator. However, by ignoring it, these textbooks also fail to alert potential users to the pitfalls of this estimator in econometric applications. This lack of information on the potential drawbacks of the estimator, coupled with the attractive "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 BWM-estimator among applied econometricians.

It is, however, important to recall that the BWM-estimator was introduced at a time when there were no robust alternatives to least squares estimators. Given that in practice the presence of outliers is often a source of concern, these estimators were a very welcomed step in the long path towards the development of estimators for location measures that are less sensitive to the presence of atypical observations. The main limitation of this approach, however, is that it tries to obtain a robust estimator of the mean which, by definition, is not itself a "robust" location function. Consequently, the so-called robust regression methods are only valid under very stringent conditions. The natural next step in this path was the development of estimators for conditional location functions that are intrinsically robust, like the quantiles (Koenker and Bassett, 1978) and the mode (Lee, 1989, 1993, and Kemp and Santos Silva, 2010). These estimators combine the desired robustness to the presence of outliers with both a clear interpretation and validity under mild distributional assumptions. Moreover, as illustrated in Section 4, they also provide important informational gains in many contexts.

Therefore, practitioners have at their disposal appropriate tools to perform regression analysis when they want to shield their results from the effects of possible outliers. Indeed, both median and mode regression are consistent for the parameters of the conditional mean when the BWM-estimator is valid, and they continue to be consistent for interesting and clearly interpretable sets of parameters when the BWM-estimator is invalid. In view of this, it is recommended that practitioners should consider using both quantile and mode regression when the information provided by the standard OLS is somehow deemed inappropriate or insufficient.

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.

Breusch, T.S. and Pagan, A.R. (1979). "A simple test for heteroscedasticity and random coefficient variation," *Econometrica*, 47, 1287-1294.

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.

Cook, R.D., Weisberg, S. (1982). *Residuals and influence in regression.* New York (NY): Chapman and Hall.

Crinò, R. (2010). "Service Offshoring and White-Collar Employment," Review of Economic Studies, 77, 595-632.

Croux, C., Dhaene, G. and Hoorelbeke, D. (2003). Robust standard errors for robust estimators, K.U.Leuven, Center for Economic Studies, Discussions Paper Series 03.16.

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.

Godfrey, L.G. and Orme, C.D. (1991). "Testing for skewness of regression disturbances," *Economics Letters*, 37, 31-34.

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. (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.

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.

Jann, B. (2010). *robreg: Stata module providing robust regression estimators.* Available from http://ideas.repec.org/c/boc/bocode/s457114.html.

Kemp, G.C.R. and Santos Silva, J.M.C. (2010). Regression towards the mode, Department of Economics, University of Essex, Discussion Paper No 686.

Koenker, R. (1981). "A note on studentizing a test for heteroscedasticity," *Journal of Econometrics*, 17, 107-112.

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

Lang, K. and Kahn, S. (1998). ""The effect of minimum-wage laws on the distribution of employment: theory and evidence," *Journal of Public Economics*, 69, 67-82.

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.

Maronna, R.A., and Yohai, V.J. (2000). "Robust regression with both continuous and categorical predictors," *Journal of Statistical Planning and Inference*, 89, 197-214.

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

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

Portnoy, S. and Welsh, A.H. (1992). "Exactly what is being modelled by the systematic component in a heteroscedastic linear regression," Statistics & Probability Letters, 13, 253-258.

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. (1984). "Least median of squares regression," *Journal of the American Statistical Association*, 79, 871-880.

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

Rousseeuw, P.J. and Yohai, V. (1987). "Robust Regression by Means of S-estimators," in *Robust and Nonlinear Time Series Analysis*: 256-272, edited by Franke, J., Härdle, W., and Martin, D., Berlin: Springer Verlag.

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. (2009). *Stata Release 11. Statistical Software.* College Station (TX): StataCorp LP.

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.

Temple, J. (1998). “Robustness Tests of the Augmented Solow Model,” *Journal of Applied Econometrics*, 13, 361-375.

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.

Verardi, V. and Croux, C. (2009). “Robust Regression in Stata,” The Stata Journal, 9, 439-453.

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.

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

Yohai V.J. (1987). “High Breakdown Point and High Efficiency Robust Estimates for Regression,” *Annals of Statistics*, 15, 642-656.

Zaman, A. Rousseeuw, P.J. and Orhan, M. (2001). “Econometric applications of high-breakdown robust regression techniques,” *Economics Letters* 71, 1-8.