RESEARCH REPORT

ROBUST STANDARD ERRORS FOR ROBUST ESTIMATORS

CHRISTOPHE CROUX • GEERT DHAENE • DIRK HOORELBEKE

OR 0367

Robust Standard Errors for Robust Estimators

Christophe Croux* K.U.Leuven
Geert Dhaene K.U.Leuven
Dirk Hoorelbeke K.U.Leuven

Abstract

A regression estimator is said to be robust if it is still reliable in the presence of outliers. On the other hand, its standard error is said to be robust if it is still reliable when the regression errors are autocorrelated and/or heteroskedastic. This paper shows how robust standard errors can be computed for several robust estimators of regression, including MM-estimators. The improvement relative to non-robust standard errors is illustrated by means of large-sample bias calculations, simulations, and a real data example. It turns out that non-robust standard errors of robust estimators may be severely biased. However, if autocorrelation and heteroscedasticity are absent, non-robust standard errors are more efficient than the robust standard errors that we propose. We therefore also present a test of the hypothesis that the robust and non-robust standard errors have the same probability limit.

Keywords: robust regression, robust standard errors, autocorrelation, heteroskedasticity

*Corresponding author. Address: K.U.Leuven, Department of Applied Economics, Naamsestraat 69, 3000 Leuven, Belgium. Tel. +32 16 326958. Email: christophe.croux@econ.kuleuven.ac.be

1 Introduction

It is well known that ordinary least squares estimation in the linear regression model is not robust to outliers. A single atypical observation can in fact cause this estimator to break down. Moreover, the consistency of this estimator requires a moment condition on the error distribution. In recent decades, robust regression estimators have been introduced to overcome these problems, and they have become a standard tool in regression analysis. Consider the regression model

$$Y_t = X_t'\beta + u_t, \quad \text{for } t = 1, \ldots, T, \tag{1.1}$$

where Y_t is the dependent variable, X_t is the vector of covariates and T is the number of observations. The error terms are often supposed to be distributed according to

$$u_t \overset{\text{i.i.d.}}{\sim} F_\sigma, \tag{1.2}$$

where the error distribution $F_\sigma(u) = F(u/\sigma)$ depends on a dispersion parameter σ. The distribution F can either be specified, e.g. $F = N(0,1)$, or be left unspecified. Normality of the error terms is in fact not needed; nor does any moment need to exist when applying robust estimators.

The robust statistics literature takes the view that a vast majority of the data are generated by the above-described model, while a smaller part of the data may not follow the model. The fraction of outliers that an estimator can cope with is then, roughly speaking, the estimator's breakdown point. Most of the literature in robust regression (e.g. Rousseeuw and Leroy, 1987) deals with robustness with respect to outliers, and with the identification of these outliers. The standard approach to statistical inference based on robust regression methods is to derive the limiting distribution of the robust estimator from assumption (1.2), and to compute the standard errors of the estimated regression coefficients from the formula for the asymptotic variance matrix.

In the econometric literature less attention is given to robust estimators of regression, but the concept of *robust standard errors* is well established and can be found even in introductory textbooks (see e.g. Stock and Watson, 2003, p. 504; Greene, 2003, p. 267). Here the estimator being used is often the ordinary least squares (OLS) estimator, but its standard errors are estimated without relying on assumption (1.2). As such, these so-called robust standard errors remain valid when the error terms are not i.i.d., but suffer from heteroskedasticity or autocorrelation. A robust standard error consistently estimates the true standard error even for non i.i.d. error terms. The most popular robust standard errors in econometrics are the White or Eicker-White standard errors (after Eicker, 1967, and White, 1980), which protect against heteroskedasticity, and the Newey-West standard errors (Newey and West, 1987), which are heteroskedasticity and

autocorrelation consistent (HAC) estimates of the standard error. An important property of robust standard errors is that the form of the heteroskedasticity and/or autocorrelation does not need to be specified. On the other hand, these standard errors will not be robust against outliers, since they are based on the OLS estimator.

In the robustness literature the problem of reliable estimation of the standard errors is considered as well, but most often in the sense of their robustness against outliers. While the point estimator for β is robust against outlying observations, the expression for the standard errors also needs to be reliable, in the sense of not being overly biased by the presence of outliers. Robust estimation of the variance matrix of an estimator $\hat{\beta}$ was used in this sense by Simpson, Ruppert and Carroll (1992) and Croux, Van Aelst and Dehon (2003). Similarly, when using the bootstrap to compute the standard error of a robust estimator, the issue of protection against the repeated occurrence of outliers in bootstrap samples has led to the development of robust bootstrap procedures for robust estimators. We refer to Salibián-Barrera and Zamar (2002) for a review and a new proposal. Again, robust bootstrapping is mainly used in the sense of robustness against outliers.

This paper considers the computation of robust standard errors for robust estimators, where the standard error estimates are designed to be robust against heteroskedasticity, autocorrelation and the presence of outliers. Explicit formulas for robust standard errors are given, so recourse to bootstrap techniques is not necessary. Advantage will be taken of results known from the Generalized Method of Moments (GMM) literature (Hansen, 1982). The standard errors proposed by Simpson, Ruppert and Carroll (1992), being robust only against heteroskedasticity for symmetric error terms, appear as a special case.

We focus on three different classes of estimators. The first is the class of M-estimators (Huber, 1981). These estimators are easy to compute, but their breakdown point equals zero. A smooth high-breakdown estimator for the regression model is the S-estimator (Rousseeuw and Yohai, 1984). This estimator is fast to compute using the algorithm of Ruppert (1992), has a breakdown point which can be up to 50%, and is asymptotically normal. However, its statistical efficiency is rather low. Therefore, Yohai (1987) introduced MM-estimators, which are in fact M-estimators with an auxiliary scale estimate obtained from an S-estimator. MM-estimators combine high efficiency with high breakdown and require the same computation time as S-estimators. They are now well established and are implemented in S-Plus. We will present robust standard errors for these three classes of estimators. Let us mention that in a recent paper of Field and Zhou (2003), HAC standard errors for M-estimators with known scale were studied. The hypothesis of known scale, however, is unrealistic in practice and eliminates the

variability of the residual scale estimator when deriving the expressions for the standard errors.

The different estimators are defined in Section 2, and the link with GMM estimation is made there. A general expression for the variance matrix of $\hat{\beta}$ under general conditions is presented in Section 3. A detailed discussion will follow on how this expression simplifies if certain conditions like homoskedasticity or absence of autocorrelation are met. In Section 4 it is shown by means of theoretical calculations that using non-robust standard errors for robust estimators can lead to severe biases (even when no outliers are present). A simulation study presented in Section 5 confirms again the necessity of using HAC standard errors. The price to be paid for using the robust standard errors is an increased variability of the standard error estimate. Hence in the absence of heteroskedasticity or autocorrelation there is a gain in using less robust standard errors. Motivated by this trade-off between robustness and efficiency of the estimator of the standard error, Section 6 outlines how a test can be carried out for testing which expression for the variance matrix of $\hat{\beta}$ is most appropriate. An example in Section 7 illustrates the use of this test. Section 8 concludes.

2 Robust regression estimators as GMM estimators

Let Y_t be the scalar dependent variable and X_t the p-vector of covariates, observed for $t = 1, \ldots, T$. The observations $(X_1, Y_1), \ldots, (X_T, Y_T)$ do not need to be independent, but are supposed to be generated by a stationary and ergodic process H. We shall study the form of the variance matrix of the M-, S-, and MM-estimators of the regression of Y_t on X_t, under a variety of assumptions regarding H. We first make clear that these estimators are first-order equivalent with exactly-identified GMM estimators and then take advantage of the results established for GMM (Hansen, 1982).

We start recalling the definition of a GMM estimator. Suppose that we would like to estimate the functional $\theta(H)$ that is implicitly defined by the equation

$$E\left[m_t(\theta(H))\right] = 0 \tag{2.1}$$

for $t = 1, \ldots, T$. Here m_t is a known k-valued function depending on Y_t and X_t, and $E[\cdot]$ denotes the mathematical expectation with respect to H. If k equals the dimension of the estimand $\theta(H)$, as will be the case in our setting, then we have an exactly-identified GMM estimation problem. An estimator of $\theta(H)$ is then simply obtained by solving the sample analogue of (2.1), that is,

$$\frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\theta}) = 0. \tag{2.2}$$

We prefer to call estimators of the above type exactly-identified GMM estimators, instead of general M-estimators as discussed in Huber (1981), since we will make use of the standard notations in the GMM literature. Generalized method of moments estimators have been extensively studied in econometrics. Hayashi (2000) offers a synthesis of parametric estimation and testing from a GMM perspective.

Let us now return to the regression problem. The regression functional $\beta(H)$ and the scale functional $\sigma(H)$ corresponding with M-estimators of regression are given by the solutions of the equations

$$E\left[\psi\left(\frac{Y_t - X_t'\beta(H)}{\sigma(H)}\right)X_t\right] = 0, \tag{2.3}$$

$$E\left[\rho\left(\frac{Y_t - X_t'\beta(H)}{\sigma(H)}\right) - b\right] = 0. \tag{2.4}$$

The functions ψ and ρ (which are chosen by the statistician) are non-constant, scalar-valued and a.e. differentiable. Furthermore ψ is odd, ρ is even and non-decreasing on $[0, \infty[$, with $\rho(0) = 0$, and b is a selected constant. The M-estimator of regression, $\hat{\beta}_M$, and the M-estimator of scale, $\hat{\sigma}_M$, then solve the sample analogues of (2.3)–(2.4), that is,

$$\frac{1}{T}\sum_{t=1}^{T}\psi\left(\frac{Y_t - X_t'\hat{\beta}_M}{\hat{\sigma}_M}\right)X_t = 0, \tag{2.5}$$

$$\frac{1}{T}\sum_{t=1}^{T}\rho\left(\frac{Y_t - X_t'\hat{\beta}_M}{\hat{\sigma}_M}\right) - b = 0. \tag{2.6}$$

It is clear that the M-estimator is an exactly-identified GMM estimator for $\theta = (\beta', \sigma)'$, with

$$m_t(\theta) = \begin{pmatrix} \psi\left(\frac{Y_t - X_t'\beta}{\sigma}\right)X_t \\ \rho\left(\frac{Y_t - X_t'\beta}{\sigma}\right) - b \end{pmatrix}. \tag{2.7}$$

S-estimators of regression and scale depend only on the chosen function ρ and on the constant b. This regression estimator is defined as minimising an M-estimator of scale computed from the residuals. So

$$\hat{\beta}_S = \arg\min_{\beta} \hat{\sigma}(\beta) \tag{2.8}$$

where $\hat{\sigma}(\beta)$ solves

$$\frac{1}{T}\sum_{t=1}^{T}\rho\left(\frac{Y_t - X_t'\beta}{\hat{\sigma}(\beta)}\right) - b = 0. \tag{2.9}$$

The scale estimate is then simply given as $\hat{\sigma}_S = \hat{\sigma}(\hat{\beta}_S)$. It was shown by Rousseeuw and

Yohai (1984) that $\hat{\beta}_S$ and $\hat{\sigma}_S$ satisfy the first-order conditions

$$\frac{1}{T}\sum_{t=1}^{T}\rho'\left(\frac{Y_t - X_t'\hat{\beta}_S}{\hat{\sigma}_S}\right)X_t = 0, \tag{2.10}$$

$$\frac{1}{T}\sum_{t=1}^{T}\rho\left(\frac{Y_t - X_t'\hat{\beta}_S}{\hat{\sigma}_S}\right) - b = 0. \tag{2.11}$$

(Thoughout the paper, a prime on a scalar-valued function denotes its derivative, otherwise it denotes transposition.) Note that the above equations are of the same form as (2.5) and (2.6). Hence an S-estimator is first-order equivalent with an M-estimator with $\psi = \rho'$, and has the same asymptotic distribution (Rousseeuw and Yohai, 1984). As a result, S-estimators are first-order equivalent with GMM-estimators. Note that the function ρ defining the S-estimator needs to be bounded to get a positive breakdown point for the regression estimator. But if ρ is bounded, ρ' will be redescending and (2.10) may have multiple solutions. Therefore one uses (2.8) to compute the S-estimate, but to determine the asymptotic distribution one typically uses (2.10).

We shall focus on MM-estimators of regression, because these are at once highly efficient and highly robust. First one needs to compute S-estimators $(\hat{\beta}_S, \hat{\sigma}_S)$ for a given function ρ and a constant b. Then, for a given function ψ, the MM-estimator of regression solves

$$\frac{1}{T}\sum_{t=1}^{T}\psi\left(\frac{Y_t - X_t'\hat{\beta}_{MM}}{\hat{\sigma}_S}\right)X_t = 0. \tag{2.12}$$

In case (2.12) has multiple solutions, one takes the solution with the smallest value for $\hat{\sigma}(\beta)$. For more detail on the conditions on ψ and ρ we refer to Yohai (1987). Note that ψ needs to be different from ρ' – otherwise the MM-estimator would be equivalent with an S-estimator and share the low efficiency of the latter.

Let us introduce some notation that will be used throughout the paper. Define, for given β and σ, and for given β_0 and σ, the standardised error terms

$$\varepsilon_t = \frac{Y_t - X_t'\beta}{\sigma} \qquad \text{and} \qquad \varepsilon_{0t} = \frac{Y_t - X_t'\beta_0}{\sigma}, \tag{2.13}$$

respectively. Write $\psi_t = \psi(\varepsilon_t)$, $\rho_t = \rho(\varepsilon_t)$ and $\rho_{0t} = \rho(\varepsilon_{0t})$. For the MM-estimator, let $\theta = (\beta', \beta_0', \sigma)'$, where the first parameter will be estimated by $\hat{\beta}_{MM}$ and the latter two by $\hat{\beta}_S$ and $\hat{\sigma}_S$. By (2.12), (2.10) and (2.11), these estimators are first-order equivalent with exactly-identified GMM estimators with moment function

$$m_t(\theta) = \begin{pmatrix} \psi\left(\frac{Y_t - X_t'\beta}{\sigma}\right)X_t \\ \rho'\left(\frac{Y_t - X_t'\beta_0}{\sigma}\right)X_t \\ \rho\left(\frac{Y_t - X_t'\beta_0}{\sigma}\right) - b \end{pmatrix} = \begin{pmatrix} \psi_t X_t \\ \rho_{0t}' X_t \\ \rho_{0t} - b \end{pmatrix}. \tag{2.14}$$

Here and later, we omit the functional dependency of the parameters on H, but this should be kept in mind.

3 Variance matrices

Under regularity conditions detailed in Hansen (1982), the GMM estimator $\hat{\theta}$ defined in (2.2) has a limiting normal distribution given by

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{\mathrm{d}} N(0, V), \tag{3.1}$$

where

$$V = \left(G'\Omega^{-1}G\right)^{-1}, \tag{3.2}$$

with

$$G = E\left[\frac{\partial m_t(\theta)}{\partial \theta'}\right] \quad \text{and} \quad \Omega = \sum_{j=-\infty}^{\infty} E[m_t(\theta)m_{t-j}(\theta)]. \tag{3.3}$$

In the exactly-identified case, $V = G^{-1}\Omega G'^{-1}$.

Throughout the paper, we will suppose that the M-, S-, and MM-estimators are asymptotically normal with the same limit distribution as their GMM counterpart. Asymptotic normality of M-estimators under i.i.d. assumptions for the error terms was already studied by Yohai and Maronna (1979) and for MM-estimators by Yohai (1987). Under fairly general conditions, allowing also for heteroskedasticity, asymptotic normality for S and MM-estimators was shown by Salibián-Barrera and Zamar (2004) in the location case. To our knowledge, exact conditions for the asymptotic normality of MM-estimators in the regression case with non-independent, non-identically distributed error terms have not yet been stated in the literature. For M-estimators we can refer to Hansen (1982).

For the M-estimator, in obvious notation, we have $\sqrt{T}(\hat{\theta}_M - \theta_M) \xrightarrow{\mathrm{d}} N(0, V_M)$, with $V_M = G_M^{-1}\Omega_M G_M'^{-1}$, and similarly for the S- and MM-estimators. As a first step to compute V_M, V_S and V_{MM}, the matrices G and Ω need to be calculated.

From (2.7) it readily follows that for the M-estimator

$$G_M = -\frac{1}{\sigma}E\begin{pmatrix} \psi_t' X_t X_t' & \psi_t' X_t \varepsilon_t \\ \rho_t' X_t' & \rho_t' \varepsilon_t \end{pmatrix} \text{and } \Omega_M = \sum_{j=-\infty}^{\infty} E\begin{pmatrix} \psi_t \psi_{t-j} X_t X_{t-j}' & \psi_t \rho_{t-j} X_t \\ \rho_t \psi_{t-j} X_{t-j}' & \rho_t \rho_{t-j} - b^2 \end{pmatrix}. \tag{3.4}$$

For the S-estimator, $\psi = \rho'$ in the above formula, leading to a simplification since $E(\rho_t' X_t) = 0$ by (2.10). Hence

$$G_S = -\frac{1}{\sigma}E\begin{pmatrix} \rho_t'' X_t X_t' & \rho_t'' X_t \varepsilon_t \\ 0 & \rho_t' \varepsilon_t \end{pmatrix} \quad \text{and} \quad \Omega_S = \sum_{j=-\infty}^{\infty} E\begin{pmatrix} \rho_t' \rho_{t-j}' X_t X_{t-j}' & \rho_t' \rho_{t-j} X_t \\ \rho_t \rho_{t-j}' X_{t-j}' & \rho_t \rho_{t-j} - b^2 \end{pmatrix}. \tag{3.5}$$

For the MM-estimator, $\theta = (\beta', \beta_0', \sigma)'$ consists of 3 blocks of components, and (2.14) yields

$$G_{MM} = -\frac{1}{\sigma} E \begin{pmatrix} \psi_t' X_t X_t' & 0 & \psi_t' X_t \varepsilon_t \\ 0 & \rho_{0t}'' X_t X_t' & \rho_{0t}'' X_t \varepsilon_{0t} \\ 0 & 0 & \rho_{0t}' \varepsilon_{0t} \end{pmatrix} \tag{3.6}$$

and

$$\Omega_{MM} = \sum_{j=-\infty}^{\infty} E \begin{pmatrix} \psi_t \psi_{t-j} X_t X_{t-j}' & \psi_t \rho_{0,t-j}' X_t X_{t-j}' & \psi_t \rho_{0,t-j} X_t \\ \rho_{0t}' \psi_{t-j} X_t X_{t-j}' & \rho_{0t}' \rho_{0,t-j}' X_t X_{t-j}' & \rho_{0t}' \rho_{0,t-j} X_t \\ \rho_{0t} \psi_{t-j} X_{t-j}' & \rho_{0t} \rho_{0,t-j}' X_{t-j}' & \rho_{0t} \rho_{0,t-j} - b^2 \end{pmatrix}. \tag{3.7}$$

Deriving V_M, V_S and V_{MM} is now straightforward, and the upper left $p \times p$ submatrices of these matrices are exactly the asymptotic variances $\text{Avar}(\hat{\beta}_M)$, $\text{Avar}(\hat{\beta}_S)$ and $\text{Avar}(\hat{\beta}_{MM})$ of interest. We first calculate these variance matrices under general conditions and then show which assumptions on H are needed to make them simplify to the expressions that are commonly encountered and used in practice.

From now on our focus will be on the MM-estimator. For the S- and M-estimator we refer to the Appendix. Calculating the product $G_{MM}^{-1} \Omega_{MM} G_{MM}'^{-1}$ and applying the formula for the inverse of a partitioned matrix yields the asymptotic variance of $\hat{\beta}_{MM}$ as

$$\begin{aligned} \text{Avar}(\hat{\beta}_{MM}) = A \sum_{j=-\infty}^{\infty} E(\psi_t \psi_{t-j} X_t X_{t-j}') A - a \sum_{j=-\infty}^{\infty} E(\rho_{0t} \psi_{t-j} X_{t-j}') A \\ - A \sum_{j=-\infty}^{\infty} E(\psi_t \rho_{0,t-j} X_t) a' + \sum_{j=-\infty}^{\infty} E(\rho_{0t} \rho_{0,t-j} - b^2) a a' \end{aligned} \tag{3.8}$$

where

$$A = \sigma [E(\psi_t' X_t X_t')]^{-1} \quad \text{and} \quad a = A \frac{E(\psi_t' X_t \varepsilon_t)}{E(\rho_{0t}' \varepsilon_{0t})}. \tag{3.9}$$

This expression for $\text{Avar}(\hat{\beta}_{MM})$ is robust, in the sense that it has been derived without any assumptions of homoskedasticity or absence of autocorrelation. To consistently estimate $\text{Avar}(\hat{\beta}_{MM})$ we simply take its empirical counterpart, $\widehat{\text{Avar}}(\hat{\beta}_{MM})$, applying the following rules:

1. Replace (β, β_0, σ) by $(\hat{\beta}_{MM}, \hat{\beta}_S, \hat{\sigma}_S)$;

2. Replace $E(\cdot)$ by $T^{-1} \sum_{t=1}^{T} (\cdot)$ and put any term outside the observation window (i.e. when $t - j$ is smaller than 1 or larger than T) equal to zero;

3. Replace the infinite sum $\sum_{j=-\infty}^{\infty} (\cdot)$ by the truncated weighted sum $\sum_{j=-q}^{q} w_j (\cdot)$, using Bartlett weights $w_j = 1 - |j|/(q+1)$ and $q = q(T) \to \infty$ at a slow rate in T. Newey and West (1987) show that these weights ensure positive semi-definiteness of the variance matrix estimate.

For example, the first term of $\widehat{\text{Avar}}(\hat{\beta}_{MM})$ is given by

$$\hat{A} \sum_{j=-q}^{q} \frac{1}{T} \sum_{t=1}^{T} w_j \hat{\psi}_t \hat{\psi}_{t-j} X_t X_{t-j}' \hat{A},$$

with

$$\hat{A} = \hat{\sigma} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\psi}_t X_t X_t' \right)^{-1},$$

and where $\hat{\psi}_t$ and $\hat{\psi}_t'$ are ψ_t and ψ_t' with $(\beta, \sigma) = (\hat{\beta}_{MM}, \hat{\sigma}_S)$. Using standard asymptotic arguments, it can be shown that following these rules indeed yields a consistent estimate of $\text{Avar}(\hat{\beta}_{MM})$. From $\widehat{\text{Avar}}(\hat{\beta}_{MM})$, standard errors for the regression coefficients are obtained in the usual way:

$$\widehat{\text{se}}(\hat{\beta}_{MM,j}) = \sqrt{\frac{1}{T} \widehat{\text{Avar}}(\hat{\beta}_{MM})_{jj}}$$

for $j = 1, \ldots, p$.

Note that if there are observations with large residuals with respect to the robust MM-fit, then $\hat{\psi}_t$ has a small value when ψ is a redescending function. The usual choices for ψ in MM-estimation have the property that they are zero for large arguments. Hence, if we have bad leverage points in the sample, then their X_t-value will be large, but at the same time $\hat{\psi}_t$ will be zero. It is easy to verify that bad leverage points and vertical outliers make only a limited contribution to the estimate of the asymptotic variance, and so the resulting standard errors can be called robust with respect to these types of outliers.

We will call $\widehat{\text{Avar}}(\hat{\beta})$ a HAC estimator, since it is consistent under heteroskedasticity and autocorrelation. Now we introduce restrictions on H that simplify the expressions for the variance matrix.

Absence of autocorrelation. It is obvious that the infinite sums in the expression for the asymptotic variance matrix (3.8) disappear when H is an independent process.

Condition C_1 : the observations (X_t, Y_t), $t = 1, \ldots, T$, are independent.

If C_1 holds, $\text{Avar}(\hat{\beta}_{MM})$ simplifies to

$$\begin{aligned} \text{Avar}_1(\hat{\beta}_{MM}) = {} & AE(\psi_t^2 X_t X_t')A - aE(\rho_{0t} \psi_t X_t')A - AE(\rho_{0t} \psi_t X_t)a' \\ & + E(\rho_{0t}^2 - b^2)aa'. \end{aligned}$$

This simpler expression is then estimated by its empirical counterpart $\widehat{\text{Avar}}_1(\hat{\beta}_{MM})$, by applying the rules (1) and (2) given earlier. Salibián-Barrera (2000, p. 164–165) also considered the asymptotic variance of MM-estimators under C_1 and obtained a similar formula.

Absence of heteroskedasticity. A further simplification occurs when there is no heteroskedasticity.

Condition C_2: the processes X_t and $(\varepsilon_t, \varepsilon_{0t})$ are independent.

Imposing C_2 without C_1 admits only a marginal simplification of the variance, which we omit. If C_1 and C_2 hold, then $\mathrm{Avar}(\hat{\beta}_{MM})$ becomes

$$\begin{aligned}\mathrm{Avar}_2(\hat{\beta}_{MM}) = {} & E(\psi_t^2) A_2 E(X_t X_t') A_2 - E(\rho_{0t}\psi_t) a_2 E(X_t') A_2 \\ & - E(\rho_{0t}\psi_t) A_2 E(X_t) a_2' + E(\rho_{0t}^2 - b^2) a_2 a_2'\end{aligned}$$

where

$$A_2 = \sigma \frac{[E(X_t X_t')]^{-1}}{E(\psi_t')} \quad \text{and} \quad a_2 = A_2 \frac{E(\psi_t' \varepsilon_t) E(X_t)}{E(\rho_{0t}' \varepsilon_{0t})}.$$

Taking the empirical counterpart yields $\widehat{\mathrm{Avar}}_2(\hat{\beta}_{MM})$. We do, however, advise against the use of this variance matrix estimator in practice, even when C_1 and C_2 hold. The reason is that this estimator will not be robust with respect to outliers. If a bad leverage point is present in the sample, then it will have a huge impact on the standard error estimate. For example, $\hat{A}_2$ is proportional to the inverse of an empirical second moment matrix of the observations X_t. Bad leverage points are outlying in the covariate space, and will have a strong influence on $\hat{A}_2$. This can even lead $\widehat{\mathrm{Avar}}_2(\hat{\beta}_{MM})$ to break down, where breakdown of a variance matrix estimator means that the latter has a determinant close to zero or enormously large. We refer to Simpson, Ruppert and Carroll (1992) who consider standard error breakdown. When using the bootstrap, standard error breakdown was studied by Singh (1998).

Symmetric error terms. A condition that is often imposed in the literature is symmetry of the error distribution. If this condition is met, the different expressions simplify considerably. In fact, this condition implies that the regression parameter estimator and the estimator of residual scale are asymptotically independent. Under the symmetry assumption, the expressions for the M-, S- (with $\psi = \rho'$), and the MM-estimator all have the same form.

Condition C_s: the distribution of ε_t, given X_t, is symmetric.

The simplification comes from the fact that $a = 0$ under C_s, where a was defined in (3.9). If only C_s holds, $\hat{\beta}_{MM}$ has aymptotic variance

$$\begin{aligned}\mathrm{Avar}_s(\hat{\beta}_{MM}) &= A \sum_{j=-\infty}^{\infty} E(\psi_t \psi_{t-j} X_t X_{t-j}') A \qquad (3.10) \\ &= \sigma^2 [E(\psi_t' X_t X_t')]^{-1} \sum_{j=-\infty}^{\infty} E(\psi_t \psi_{t-j} X_t X_{t-j}') [E(\psi_t' X_t X_t')]^{-1}.\end{aligned}$$

Under C_1 and C_s, the asymptotic variance becomes

$$\begin{aligned}\text{Avar}_{1s}(\hat{\beta}_{MM}) &= AE(\psi_t^2 X_t X_t')A \\ &= \sigma^2[E(\psi_t' X_t X_t')]^{-1}E(\psi_t^2 X_t X_t')[E(\psi_t' X_t X_t')]^{-1}. \end{aligned} \tag{3.11}$$

The empirical counterpart of the latter expression, $\widehat{\text{Avar}}_{1s}$, was used in Simpson, Ruppert and Carroll (1992) (in fact, they used it for the one-step GM estimator). It is also used for estimating the standard errors of an MM-estimator (Marazzi, Joss and Randriamiharisoa, 1993) in the statistical software packages S-Plus and R. The estimate $\widehat{\text{Avar}}_{1s}(\hat{\beta}_{MM})$ is robust against outliers and heteroskedasticity, but not against autocorrelation. Moreover, it relies on symmetry, which we believe to be too strong a condition.

When all of C_1, C_2 and C_s hold, then $\hat{\beta}_{MM}$ has asymptotic variance

$$\text{Avar}_{2s}(\hat{\beta}_{MM}) = \sigma^2 \frac{E(\psi_t^2)}{[E(\psi_t')]^2}[E(X_t X_t')]^{-1}, \tag{3.12}$$

since $a_2 = 0$. This is the expression for the variance of the robust regression estimator that was derived in Yohai (1987) for the MM-estimator, in Rousseeuw and Yohai (1984) for the S-estimator and in Huber (1981) for the M-estimator. However, for the same reason as for $\widehat{\text{Avar}}_2$, we cannot recommend its use in practice, since on top of not being HAC it lacks robustness with respect to outliers.

Specified error distribution. A final simplification occurs when the distribution of the standardised error term, $F(\varepsilon)$, is known or assumed. The estimate can then be improved upon because $E(\psi_t^2)$ and other terms can be calculated analytically rather than estimated. Assuming C_1 and C_2 hold, we can rewrite $\text{Avar}_2(\hat{\beta}_{MM})$ as

$$\text{Avar}_2(\hat{\beta}_{MM}) = c_1\sigma^2[E(X_t X_t')]^{-1} + c_2\sigma^2[E(X_t X_t')]^{-1}E(X_t)E(X_t')[E(X_t X_t')]^{-1},$$

where

$$c_1 = \frac{E(\psi_t^2)}{[E(\psi_t')]^2}$$

and

$$c_2 = \frac{E(\psi_t'\varepsilon_t)}{[E(\psi_t')]^2 E(\rho_{0t}'\varepsilon_{0t})}\left[\frac{E(\psi_t'\varepsilon_t)E(\rho_t^2 - b^2)}{E(\rho_{0t}'\varepsilon_{0t})} - 2E(\rho_{0t}\psi_t)\right]$$

are constants that only depend on F. A common practice is to take the distribution of ε_t as the standard normal distribution, for which $c_2 = 0$, in view of the symmetry of the normal distribution. The resulting expression for the asymptotic variance of $\hat{\beta}_{MM}$ is then

$$\widehat{\text{Avar}}_3(\hat{\beta}_{MM}) = c_1\hat{\sigma}^2\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}.$$

This simple expression resembles the formula for the variance matrix of the OLS estimator (under Gauss-Markov conditions), but it lacks any form of robustness.

Let us summarise the above findings. First of all, most of the above formulas are robust with respect to outliers. If a bounded ψ is used, then $\widehat{\text{Avar}}$, $\widehat{\text{Avar}}_1$, $\widehat{\text{Avar}}_2$ (and their versions under symmetry) are robust with respect to vertical outliers. If a redescending ψ is used, then $\widehat{\text{Avar}}$ and $\widehat{\text{Avar}}_1$ are robust with respect to bad leverage points, but $\widehat{\text{Avar}}_2$ is not. In a time-series setting we advocate using $\widehat{\text{Avar}}$, since it gives full protection against autocorrelation, heteroskedasticity and outliers. In the absence of serial dependence between observations, $\widehat{\text{Avar}}_1$ is appropriate. The simplified expression $\widehat{\text{Avar}}_{1s}$ relies on symmetry of the errors, an assumption we feel is too strong. It can be found in Simpson, Ruppert and Carroll (1992) and is currently being used. Therefore, it will be included in our simulation and example section, for reasons of comparison. It will turn out from the simulation study that, even when symmetry is present, there is no gain in using $\widehat{\text{Avar}}_{1s}$ compared to $\widehat{\text{Avar}}_1$, so why not use $\widehat{\text{Avar}}_1$ in any case. We do not recommend using $\widehat{\text{Avar}}_2$ nor its symmetric counterpart, since these estimators are not robust with respect to bad leverage points. However, since $\widehat{\text{Avar}}_{2s}$ has been considered before in the literature, we will include it in the Monte Carlo study. The same remark applies to $\widehat{\text{Avar}}_3$. We do not recommend it in practice, but we include it as a point of reference.

4 Large-sample bias

In this section the following question is addressed: "Do we make a big mistake by using one of the simplified formulas for the standard errors instead of the heteroskedasticity and autocorrelation consistent ones?" In simple settings it is possible to carry out theoretical calculations to answer this question. Throughout the examples, we assume that the symmetry condition C_s holds.

Let $\hat{\beta}$ be a scalar estimate (or redefine $\hat{\beta}$ to be the element of interest from a vector of estimates), which can be an M, S, or MM-estimator. Consider standard error estimates of $\hat{\beta}$ based on $\widehat{\text{Avar}}_{1s}(\hat{\beta})$ and $\widehat{\text{Avar}}_{2s}(\hat{\beta})$ as alternative, more simple standard errors than those based on $\widehat{\text{Avar}}(\hat{\beta})$. We know that $\widehat{\text{Avar}}_{js}(\hat{\beta})$ is consistent for $\text{Avar}_{js}(\hat{\beta})$, for $j = 1, 2$. The (asymptotic) proportional bias of any standard error estimate, say $\widehat{\text{se}}_{2s}(\hat{\beta})$, is then defined as

$$\text{Abias}\left(\widehat{\text{se}}_{2s}(\hat{\beta})\right) = \frac{1}{2}\log\left(\frac{\text{Avar}_{2s}(\hat{\beta})}{\text{Avar}(\hat{\beta})}\right). \tag{4.1}$$

A consistent estimate of the true asymptotic variance thus delivers a standard error which has zero proportional bias. We show in two stylised cases where heteroskedasticity or autocorrelation is present that the proportional bias of $\widehat{\text{se}}_{2s}(\hat{\beta})$ and $\widehat{\text{se}}_{1s}(\hat{\beta})$ may become infinite. These examples are illustrative of the fact that the validity of the assumptions on which standard error

estimates are based does matter for the validity of the inference conducted. So it deserves careful consideration in any serious application.

4.1 Heteroskedasticity

Let H be an independent process, with $Y_t = X_t\beta + \sigma\varepsilon_t$, X_t univariate and symmetric around a zero mean, and ε_t, given X_t, symmetrically distributed around zero. The presence of heteroskedasticity is not excluded, and we would like to compute the proportional bias when using the estimate $\widehat{\text{Avar}}_{2s}$, based on assumption C_2. Since C_s and C_1 hold in this example, $\text{Avar}_{1s} = \text{Avar}$ and from (3.11) and (3.12) it readily follows that

$$\text{Abias}\left(\widehat{\text{se}}_{2s}(\hat{\beta})\right) = \frac{1}{2}\log\left(\frac{K^2}{L}\right), \tag{4.2}$$

where

$$K = \frac{E\left(\psi_t' X_t^2\right)}{E\left(\psi_t'\right) E\left(X_t^2\right)} \qquad \text{and} \qquad L = \frac{E(\psi_t^2 X_t^2)}{E(\psi_t^2)E(X_t^2)}. \tag{4.3}$$

Neither K nor L are bounded above by any constant, and the proportional bias may become infinite. To make this more clear, consider for example error terms given by $\varepsilon_t = \pm|X_t|$, each with probability $\frac{1}{2}$, and take a redescending ψ. Then, while K and L are finite, one still has

$$0 \leq \frac{K^2}{L} \leq M\left[E(X_t^2)\right]^{-1},$$

for some $M > 0$. Hence, for a covariate with an infinite variance, the proportional bias will be minus infinity. Note that if $\psi(u) = u$, then the M-estimator equals the OLS estimator, and (4.2) returns $\frac{1}{2}\log(\text{kurtosis}X_t)$.

To develop a sense for the likely sign of the bias, we will make use of the relations

$$K < 1 \Leftrightarrow \text{Cov}(\psi_t', X_t^2) > 0 \quad \text{and} \quad L > 1 \Leftrightarrow \text{Cov}(\psi_t^2, X_t^2) > 0.$$

Take ψ concave on $[0, \infty)$, such that ψ' is non-increasing and ψ^2 is non-decreasing on $[0, \infty)$. Let us model the heteroskedasticity as $\sigma\varepsilon_t = \gamma(|X_t|)\eta_t$ with η_t an i.i.d. $N(0,1)$ sequence independent of X_t. So $\gamma(x)$ is nothing else but the conditional variance function. If γ is increasing in $|X_t|$, then it is not difficult to see that ψ_t^2 will be positively correlated with X_t^2 and ψ' negatively correlated with X_t^2. From the relations stated earlier, it then follows that the proportional bias will be negative. So ignoring the existence of an increasing conditional variance function $\gamma(|X_t|)$ when computing standard errors will yield an underestimation of the standard errors. The consequence is that confidence intervals will be too short, as will be confirmed in the simulation study in Section 5. If, instead, $\gamma(|X_t|)$ is decreasing in $|X_t|$ (an assumption that is much less likely to hold), this conclusion is reversed.

4.2 Autocorrelation

Let $Y_t = X_t\beta + \varepsilon_t$, with X_t a univariate stationary process with zero mean. Now $\{X_t\}$ and $\{\varepsilon_t\}$ are supposed to be independent, but there is serial correlation in $\{\varepsilon_t\}$. Suppose, for the sake of simplicity again, that $\{\varepsilon_t\}$ follows an ARMA process with symmetrically distributed innovations. Then C_s holds and, in view of the independence between $\{X_t\}$ and $\{\varepsilon_t\}$, it follows from (3.10) and (3.11) that

$$\text{Abias}\left(\widehat{\text{se}}_{1s}(\hat{\beta})\right) = \frac{1}{2}\log\left(\frac{1}{L}\right), \tag{4.4}$$

where

$$L = \frac{\sum_{j=-\infty}^{\infty} E(\psi_t\psi_{t-j}X_tX_{t-j}')}{E(\psi_t^2)E(X_t^2)} = 1 + 2\sum_{j=1}^{\infty}\rho_j^{\psi}\rho_j^{X}, \tag{4.5}$$

with

$$\rho_j^{\psi} = \frac{E(\psi_t\psi_{t-j})}{E(\psi_t^2)} \qquad \text{and} \qquad \rho_j^{X} = \frac{E(X_tX_{t-j})}{E(X_t^2)}. \tag{4.6}$$

So it can be seen that the proportional bias depends only on the autocorrelation coefficients of the processes X_t and ε_t.

Clearly, L is only bounded by $0 < L < \infty$, and the bias of $\widehat{\text{se}}_{2s}(\hat{\beta})$ can go beyond all limits. To illustrate this, suppose that X_t follows an AR(1) process with parameter ρ^X and ε_t follows another AR(1) process with parameter ρ^{ψ}. Then the proportional bias equals $\log((1-\rho)/(1+\rho))$ with $\rho = \rho^{\psi}\rho^X$. If this parameter ρ approaches one, then the proportional bias becomes minus infinity, leading to severe underestimation of the standard errors.

An interesting observation is that when all autocorrelations of X_t are zero, L reduces to 1 and there is no proportional bias anymore. Under the assumptions being made, correcting the standard errors for autocorrelation is only necessary when both the errors and the covariates are correlated over time.

5 Simulations

Here we present Monte Carlo evidence of the fact that (i) non-robust standard errors of robust regression estimates can be highly misleading (thereby confirming the large-sample analysis of Section 4); (ii) robust standard errors perform well, notwithstanding their higher degree of complexity. Monte Carlo results are, of course, design-specific, but in *all* the cases that we studied, the non-robust standard errors *were* misleading, while the robust standard errors were acceptable.

We investigated the properties of the standard errors of the slope estimate $\hat{\beta}_1$ in the simple regression model $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ $(t = 1, ..., 1000)$. Since the regression and scale estimates

are equivariant, the results are invariant with respect to β_0, β_1, and σ. So we set $Y_t = \varepsilon_t$ without loss of generality. We generated $\{X_t\}$ and $\{\varepsilon_t\}$ under five different designs:

1. i.i.d. errors: $\{\varepsilon_t\}$ and $\{X_t\}$ and independent, i.i.d. $N(0,1)$ processes;
2. heteroskedastic errors: $\varepsilon_t = |X_t|u_t$, where $\{X_t\}$ and $\{u_t\}$ are independent, i.i.d. $N(0,1)$ processes;
3. autocorrelated errors (AR): $X_t = .7X_{t-1} + v_t$ and $\varepsilon_t = .7\varepsilon_{t-1} + u_t$, where $\{v_t\}$ and $\{u_t\}$ are independent, i.i.d. $N(0,1)$ processes;
4. autocorrelated errors (MA): $X_t = .9X_{t-1} + v_t$ and $\varepsilon_t = u_t + .9u_{t-1}$, where $\{v_t\}$ and $\{u_t\}$ are independent, i.i.d. $N(0,1)$ processes;
5. heteroskedastic and autocorrelated errors: $X_t = .7X_{t-1} + v_t$ and $\varepsilon_t = |X_t|u_t$, where $u_t = .7u_{t-1} + w_t$, and $\{v_t\}$, and $\{w_t\}$ are independent, i.i.d. $N(0,1)$ processes.

The covariate values were not held fixed, but were randomly redrawn for each new Monte Carlo run. For the designs with serial correlation in the error terms, the covariate was also correlated over time, first of all because we think this is a realistic situation in many applications, and second because it was shown in the previous section that no correction for the standard error is needed (at least asymptotically) when serial correlation is present only in the error term and not in the covariates.

We considered the following robust estimators of regression:

1. An M-estimator using Huber's ψ-function, $\psi_b^H(u) = \min(b, \max(u, -b))$ and $\rho_c^H(u) = (\psi_c^H(u))^2$. We choose $b = 1.35$ and $c = 2.38$ to attain 95% efficiency at i.i.d. normal errors, both for the regression and for the scale estimator.
2. The S-estimator defined by Tukey's biweight function,

$$\rho_c^B(u) = \begin{cases} \frac{u^2}{2} - \frac{u^4}{2c^2} + \frac{u^6}{6c^4}, & \text{if } |u| \le c; \\ \frac{c^2}{6}, & \text{if } |u| > c. \end{cases} \tag{5.1}$$

   Here we put $c = 2.94$, so that that the estimator has a 25% breakdown point.
3. An MM-estimator using an initial residual scale estimate based on a robust S-estimator defined by ρ_c^B with $c = 1.55$ (yielding a 50% breakdown point). Then take $\psi(u) = \rho_c^B(u)$, where $c = 4.69$ to have an efficiency of 95% at the Gaussian model.

The above choices of ρ and ψ are standard. For all the estimators considered we calculate the standard errors using the methods described in Section 3, namely: $\widehat{\text{se}}$, $\widehat{\text{se}}_1$, $\widehat{\text{se}}_{1s}$, $\widehat{\text{se}}_{2s}$ and $\widehat{\text{se}}_3$. Recall that $\widehat{\text{se}}$ is the HAC estimate of the standard error, and $\widehat{\text{se}}_1$ the one we advocate to use for non time-series problems. The other estimators for the standard errors are included for comparison and have already been considered in the literature: $\widehat{\text{se}}_{1s}$ requires symmetry of the errors and absence of autocorrelation and heteroskedasticity; for $\widehat{\text{se}}_{2s}$ one also needs independence of the error terms and the covariates, and $\widehat{\text{se}}_{3s}$ is valid for i.i.d. normal errors, independent of the covariates.

When implementing $\widehat{\text{se}}$, a truncation lag q for the infinite sums needs to selected. The choice of q in $\widehat{\text{Avar}}$ is, in itself, a research domain. We keep a low profile in this debate and choose q equal to the integer part of $4(T/100)^{2/9}$, as suggested by Newey and West (1994), and as used in the popular econometric software package EViews.

The simulation study was carried out with $R = 10000$ Monte Carlo runs. We focus on three distinct aspects of standard errors:

1. The quality of the standard error as an estimate of the standard deviation of the regression estimate, as measured by the proportional bias. We define the finite-sample proportional bias of a standard error estimate, say $\widehat{\text{se}}(\hat{\beta}) = [T^{-1}\widehat{\text{Avar}}(\hat{\beta})]^{1/2}$, as

$$\text{PB}\left(\widehat{\text{se}}(\hat{\beta})\right) = E\log\left(\frac{\widehat{\text{se}}(\hat{\beta})}{\text{sd}(\hat{\beta})}\right),$$

   where $\text{sd}(\hat{\beta}) = [E(\hat{\beta} - E(\hat{\beta}))^2]^{1/2}$ is the "true" standard error, which we approximate by the Monte Carlo simulations as

$$\text{sd}(\hat{\beta}) = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\beta})(\hat{\beta}_r - \bar{\beta})',$$

   where $\bar{\beta} = (1/R)\sum_{r=1}^{R}\hat{\beta}_r$;

2. The quality of the standard error as an estimate of the standard deviation of the regression estimate, as measured by the mean squared error of the estimate of the true standard error. Indeed, among consistent estimates of $\text{sd}(\hat{\beta})$ one preferably chooses the most precise one. The simulation study can give insight into efficiency issues. The root-mean-squared proportional error is defined as

$$\text{RMSE}\left(\widehat{\text{se}}(\hat{\beta})\right) = \left[E\log^2\left(\frac{\widehat{\text{se}}(\hat{\beta})}{\text{sd}(\hat{\beta})}\right)\right]^{1/2} = \left[E\left(\log(\widehat{\text{se}}(\hat{\beta})) - \log(\text{sd}(\hat{\beta}))\right)^2\right]^{1/2};$$

3. The rejection probability of a 5%-level t-test of $H_0 : \beta_1 = 0$ (recall that β_1 is the slope parameter). When, say, $\widehat{\text{se}}$ is used, the test rejects if $|\hat{\beta}_1|/\widehat{\text{se}}(\hat{\beta}_1)$ exceeds the 0.975-quantile of the t_{T-2} distribution. This rejection probability is obtained by computing the number of times that the null hypothesis was rejected. Since we will simulate samples under H_0, we expect this value to be close to 5%.

For each robust regression estimate and each method for calculating the standard error we report the bias, RMSE, and rejection probability, all estimated from 10000 Monte Carlo runs. The expected values in the formula for RMSE and PB are approximated by the Monte Carlo average.

Table 1 provides a comparison of the different standard error estimates for the MM-estimator under different sampling schemes. The results for the M- and S-estimators can be found in Tables 5 and 6 in the Appendix. First consider the ideal setting where we have i.i.d. Gaussian error terms. As measured by the proportional bias and rejection probability, all estimators work fine here: no significant bias is present and the associated tests have an almost exact 5% level. To distinguish between the different estimators, we need to look at their precision, measured by the RMSE. Here we can clearly rank the estimators: $\widehat{\text{se}}_3$ and $\widehat{\text{se}}_{2s}$ outperform the other standard error estimates in terms of RMSE. The price paid for robustness against heteroskedasticity or/and autocorrelation is an efficiency loss in the Gaussian model. The HAC estimate $\widehat{\text{se}}$ exhibits the greatest loss of precision. Note that we are not speaking about the efficiency of $\hat{\beta}$, but about the efficiency of $\widehat{\text{se}}$ as an estimator of the true standard error. When using the OLS estimator, the loss of efficiency of $\widehat{\text{se}}_1$ with respect to $\widehat{\text{se}}_2$ under i.i.d. error terms was discussed in Kauermann and Carroll (2001). Also notice that explicitly using the normality assumption, as in $\widehat{\text{se}}_3$, offers no significant advantage over the similar formula $\widehat{\text{se}}_{2s}$. Furthermore the results for $\widehat{\text{se}}_{1s}$ and $\widehat{\text{se}}_1$ are virtually identical, and this is true for all sampling schemes that are considered here. There seems to be no efficiency gain in using $\widehat{\text{se}}_{1s}$ over $\widehat{\text{se}}_1$, even when the error distribution is truly symmetric. The message from this is to always use the formulas that allow for asymmetry, even if they are more complicated.

For the M- and S- estimator similar conclusions can be drawn. From Table 6, in the i.i.d. normal case it can be seen that some finite-sample bias is present for the S-estimator, together with a size distortion for the test procedure. Here the Gaussian efficiency of the S-estimator of regression is equal to 75.9%, much less than the 95% efficiency for both the M- and MM-procedures. It is no surprise that a more precise inference can be derived from more efficient regression estimators.

Once heteroskedasticity is introduced, the situation changes. Both $\widehat{\text{se}}_3$ and $\widehat{\text{se}}_{2s}$ have a severe

Table 1: Performance of alternative estimates of standard errors of MM-estimates (1000 observations)

	$\widehat{\text{se}}_3$	$\widehat{\text{se}}_{2s}$	$\widehat{\text{se}}_{1s}$	$\widehat{\text{se}}_1$	$\widehat{\text{se}}$
	i.i.d. normal				
PB	0.0005	0.0023	0.0022	0.0029	-0.0031
RMSE	0.0374	0.0327	0.0495	0.0495	0.0657
RP	0.0521	0.0511	0.0503	0.0503	0.0521
	heteroscedasticity				
PB	1.1543	-0.9974	-0.0092	-0.0048	-0.0105
RMSE	1.1552	0.9985	0.1164	0.1168	0.1249
RP	0.5346	0.4678	0.0545	0.0531	0.0567
	autocorrelation: AR(1)				
PB	-0.5184	-0.5162	-0.5184	-0.5173	-0.1034
RMSE	0.5217	0.5192	0.5235	0.5224	0.1414
RP	0.2418	0.2405	0.2423	0.2423	0.0775
	autocorrelation: MA(1)				
PB	-0.3091	-0.3072	-0.3091	-0.3082	-0.0456
RMSE	0.3183	0.3160	0.3208	0.3199	0.1168
RP	0.1445	0.1445	0.1472	0.1468	0.0630
	heteroscedasticity and autocorrelation				
PB	-1.4642	-1.3075	-0.3236	-0.3186	-0.0662
RMSE	1.4654	1.3088	0.3474	0.3430	0.1619
RP	0.6545	0.5977	0.1590	0.1572	0.0747

Results are based on 10000 Monte Carlo runs. Data are generated by $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ $(t = 1, ..., 1000)$ under 5 different sampling schemes, described in the main text.

PB: Proportional Bias.

RMSE: Root-Mean-Squared Proportional Error.

RP: Rejection Probability of 5%-level t-test on slope parameter.

downward bias and strongly overreject the null hypothesis. We cannot rely on them anymore, even when there are no outliers but only heteroskedastic error terms. The estimators $\widehat{\text{se}}_{1s}$ and $\widehat{\text{se}}_1$ are doing their work; they indeed remain robust when condition C_2 is violated. The HAC estimator $\widehat{\text{se}}$, too, still has the right size and no bias, and the loss in RMSE relative to $\widehat{\text{se}}_1$ is rather small. This holds as well for the M- and S-estimators. Note that the RMSE of $\widehat{\text{se}}_1$ and $\widehat{\text{se}}$ for the M-estimator are smaller than for the MM procedure. The reason is that while the M- and MM-regression estimators were calibrated to have equal efficiencies in a Gaussian homoskedastic model, the M-estimator will have a greater efficiency in this heteroskedastic model. Intuitively this is clear since the M-estimator uses a non-decreasing ψ, while MM is based on a redescending ψ. The heteroskedasticity generated by this model will generate observations resembling outliers, and these are downweighted too much by the S- and MM-estimators.

By introducing autocorrelation into the error terms, and in the last, worst-case sampling scheme even combined with heteroskedasticity, all standard error estimators have huge negative biases; hence they severely underestimate the true sampling variance and the tests reject the null far too often. The only exception is the HAC estimate $\widehat{\text{se}}$, which continues to have a moderate bias and acceptable rejection probabilities. Although the results for the HAC estimator are not perfect, they are far better than those of methods that don't correct for the presence of autocorrelation.

When computing the HAC estimator for the variance matrix an infinite sum is replaced by a finite weighted sum, where autocorrelations of higher orders receive less weight. In this way, an approximation error is made, which is small if there is not much persistence in the residual process. We see that the PB, RMSE and RP are better for the MA(1) than for AR(1) errors, the latter being more persistent. Moreover, this approximation also improves when the sample size increases. Here the sample size is already quite large ($T = 1000$), and hence we decided to repeat the simulation experiment with a smaller sample size, $T = 200$. From Table 7 in the Appendix, one sees that the performance of $\widehat{\text{se}}$ indeed deteriorates due to the smaller sample size, but it still outperforms all other estimates of the standard error, which all break down in the presence of autocorrelation. Now the difference in RMSE between $\widehat{\text{se}}$ and $\widehat{\text{se}}_1$ when C_1 holds is slightly increased. This confirms the intuition that when only heteroskedasticity is present, for example in non time-series problems, robust standard errors of the type $\widehat{\text{se}}_1$ are to be preferred.

So far, only robustness of the standard errors under heteroskedasticity and autocorrelation has been investigated, but not yet with respect to outliers. Two different sampling schemes are considered, and the results are reported in Table 2. For the first sampling scheme the error terms were generated as i.i.d Cauchy, instead of i.i.d normal. Due to the heavy tails of the Cauchy

distribution, this will generate vertical outliers. We see that the PB and RP remain very good, confirming that $\widehat{se}$, $\widehat{se}_1$ and $\widehat{se}_{2s}$ are robust with respect to vertical outliers and that we do not need normality at all. There is a larger bias for $\widehat{se}_3$, since the latter uses normality. The RMSE increases compared to the case of normal errors, due to the fat tails of the Cauchy distribution.

In a second sampling scheme bad leverage points have been generated. The same data generating processes as for i.i.d. normal errors were taken, but now bad leverage points are added by putting 10% of the data (X_t, Y_t) equal to (10,18). For the MM-estimator the results remain very good for $\widehat{se}$ and $\widehat{se}_1$, but $\widehat{se}_{2s}$ and $\widehat{se}_3$ completely break down in the presence of the bad leverage points, as expected. The MM-estimator itself, of course, does not break down. For the M-estimator the results are very bad, because here the regression estimator itself breaks down, resulting in a 100% rejection probability. In addition, here the bad leverage points inflate the estimate of scale, and, in turn, the standard error estimates. Hence, using robust estimators of the standard errors combined with non-robust regression estimators yields a non-robust procedure.

Table 2: Performance of alternative estimates of standard errors of MM- and M-estimates in the presence of outliers (1000 observations)

	$\widehat{se}_3$	$\widehat{se}_{2s}$	$\widehat{se}_{1s}$	$\widehat{se}_1$	$\widehat{se}$
	i.i.d. Cauchy; MM-estimator				
PB	-0.0623	-0.0117	-0.0116	-0.0092	-0.0143
RMSE	0.0812	0.0553	0.0756	0.0755	0.0879
RP	0.0622	0.0505	0.0497	0.0493	0.0509
	bad leverage points; MM-estimator				
PB	-1.0497	-1.1472	-0.0004	0.0002	-0.0059
RMSE	1.0502	1.1475	0.0502	0.0502	0.0679
RP	0.4825	0.5222	0.0507	0.0507	0.0538
	bad leverage points; M-estimator				
PB	0.7647	0.7500	0.4741	0.5162	1.3018
RMSE	0.7650	0.7505	0.4762	0.5180	1.3027
RP	1.0000	1.0000	1.0000	1.0000	1.0000

Results are based on 10000 Monte Carlo runs. Data are generated by $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ $(t = 1, ..., 1000)$ with i.i.d. Cauchy errors or with i.i.d. normal errors with 10% bad leverage points.
PB: Proportional Bias.
RMSE: Root-Mean-Squared Proportional Error.
RP: Rejection Probability of 5%-level t-test on slope parameter.

6 Testing the difference between variance estimates

In this section we propose a procedure for testing whether a variance estimate of $\hat{\beta}$ is significantly different from the HAC estimator. To fix ideas, suppose we would like to know whether it is

appropriate to use $\widehat{\text{Avar}}_{2s}$ instead of the HAC estimator. If this is indeed appropriate, then it is preferable to use $\widehat{\text{Avar}}_{2s}$, since it gives some efficiency gains. A test statistic will then be based on the difference

$$\widehat{\text{Avar}}_{2s} - \widehat{\text{Avar}}.$$

If the difference is small, then one could use $\widehat{\text{Avar}}_{2s}$, but one could also test whether $\widehat{\text{Avar}}_3$ is perhaps appropriate as well. This leads to tests of the type

$$H_0^j : \text{Avar}_j = \text{Avar},$$

where in our case $j = 2s$. Denote the difference between the estimates as

$$\hat{D}_j = \text{vech}(\widehat{\text{Avar}}_j - \widehat{\text{Avar}}), \tag{6.1}$$

where $\text{vech}(A)$ is the vectorised lower triangular part of A. To compute the limit distribution of the quadratic form associated with $\hat{D}_j$, we shall in fact employ the stronger conditions stated earlier that led up to Avar_j.

One can expect (deriving the exact conditions would lead us too far here) that $\hat{D}_j$ is asymptotically normal with mean zero and asymptotic variance matrix V_j. The associated quadratic form statistic is then

$$T_j = T\hat{D}_j' V_j^{-1} \hat{D}_j,$$

(recall that T is the number of observations) and has an asymptotic χ_q^2 distribution with q degrees of freedom, where q is the rank of V_j, here at most $q = p(p+1)/2$. To put the above formula into practice, one needs an estimate of the variance matrix V_j. For this we resort to a bootstrapping procedure, since analytic computation of V_j seems too tough here. The bootstrap samples are to be generated under the corresponding null hypothesis. This leads us to consider different kinds of bootstrap sampling schemes, depending on the conditions that are invoked:

- C_1: *non-parametric bootstrap of pairs* (X_t, Y_t). For testing H_0^1 one resamples with replacement from the set $\{(X_t, Y_t) : t = 1, \ldots, T\}$, which is the standard non-parametric bootstrap in regression;

- C_1 and C_s: a version of the *wild bootstrap* of Liu (1988). We first compute the standardised residuals $\hat{\varepsilon}_t = (Y_t - X_t'\hat{\beta})/\hat{\sigma}$. Then we generate independent drawings $Z_{t,b}$, for $t = 1, \ldots, T$ and $b = 1, \ldots, B$, with B the number of bootstrap samples such that $Z_{t,b}$ equals 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$. The wild bootstrap samples are then obtained as $Y_{t,b} = X_t\hat{\beta} + \hat{\sigma} Z_{t,b}\hat{\varepsilon}_t$. The covariates are not being resampled here;

- C_1 and C_2: *residual bootstrap.* A bootstrap observation is formed as $Y_{t,b} = X_t'\hat{\beta} + \hat{\sigma}e_{t,b}$, where $e_{t,b}$ is sampled with replacement from the collection $\{\hat{\varepsilon}_t : t = 1, \ldots, T\}$. This is the standard residual bootstrap;

- C_1, C_2 and C_s: *randomly signed residual bootstrap.* Here the bootstrap observations are generated as $Y_{t,b} = X_t'\hat{\beta} + \hat{\sigma}e_{t,b}$, where $e_{t,b}$ is sampled with replacement from $\{\pm\varepsilon_t : t = 1, \ldots, T\}$;

- C_1, C_2 and normality: *parametric bootstrap.* If one assumes i.i.d. normal errors, independent of the covariates, then a bootstrap sample is generated as $Y_{t,b} = X_t'\hat{\beta} + \hat{\sigma}e_{t,b}$, where $e_{t,b}$ is drawn from the $N(0,1)$ distribution. This is a fully parametric bootstrap method.

A bootstrap estimate of the variance matrix V_j is obtained as follows:

1. For $b = 1, \ldots, B$ generate a bootstrap sample under the appropriate sampling scheme. Compute for each bootstrap sample $\hat{D}_{j,b}$ as in (6.1);

2. Estimate V_j by a robust variance matrix estimator computed from $\{\hat{D}_{j,b} | 1 \leq b \leq B\}$. Call this estimate $\hat{V}_{j,B}$.

It is necessary to use a robust variance matrix, since by bootstrapping it is possible that the number of outliers gets multiplied, which can cause breakdown of the regression MM-estimator and hence also of its variance matrix estimate. So some of the $D_{j,b}$ might be contaminated, and could strongly bias a sample variance matrix computed from them. As robust variance matrix estimator we use a Multivariate S-estimator (Rousseeuw and Leroy, 1987), but other choices are possible. The following step is to construct a quadratic form statistic

$$T_{j,B} = T\hat{D}_j' \left(\hat{V}_{j,B}\right)^{-1} \hat{D}_j.$$

The asymptotic distribution of $T_{j,B}$, as $T \to \infty$ and $B \to \infty$, is again χ^2_q.

7 Application

We apply some of the standard error estimators proposed in Section 3 to the salinity data set (Rousseeuw and Leroy, 1987, p. 82). This data set has been used by many authors to illustrate the robustness of the regression estimator, but the associated standard errors were almost never discussed. The dependent variable here is water salinity (salt concentration) in Pamlico Sound, and the explanatory variables include a constant, lagged salinity (x_1), a trend (x_2) and the volume of river discharge into the sound (x_3). The regression model is estimated by the MM-estimator with a 50% breakdown point and 95% efficiency, and a plot of the standardised

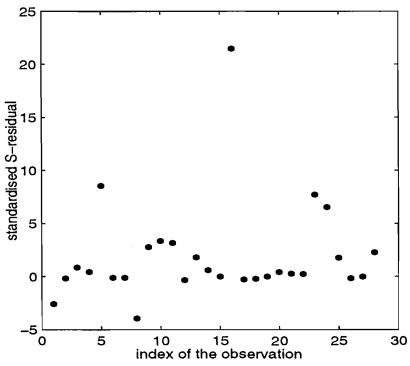

Figure 1: Robust residual plot of salinity data

residuals is given in Figure 1. The residuals are robust, since they are residuals from an MM-fit and scaled by the robust estimator of residual scale. In most textbooks one discusses the presence of the outliers. One of them, observation 16, is known to be a bad leverage point. However, a closer inspection of the residual plot against time also reveals that autocorrelation is likely to be present.

Table 3 contains the estimated coefficients and their associated standard errors, estimated using five different methods: $\widehat{\text{Avar}}$, $\widehat{\text{Avar}}_1$, $\widehat{\text{Avar}}_{1s}$, $\widehat{\text{Avar}}_{2s}$ and $\widehat{\text{Avar}}_3$. From Table 3 it is clear that $\widehat{\text{se}}_3$ and $\widehat{\text{se}}_{2s}$ differ substantially from the other 3 estimated standard errors. This difference is explained by the presence of outliers and the fact that $\widehat{\text{se}}_2$ and $\widehat{\text{se}}_{2s}$ are not outlier-robust. The robust standard errors obtained by the other methods are quite comparable. Assuming symmetry or not makes no difference for $\widehat{\text{se}}_1$, while the standard errors $\widehat{\text{se}}$ are somewhat bigger for most coefficients. Given the appearance of serial correlation, we have most confidence in the HAC standard errors.

Table 3: Salinity data: MM estimates and standard errors

	x_1	x_2	x_3	constant
$\hat{\beta}_{MM}$	0.7271	-0.2116	-0.6191	18.2110
$\widehat{\text{se}}_3$	0.0433	0.0809	0.0537	1.5704
$\widehat{\text{se}}_{2s}$	0.0604	0.1128	0.0748	2.1894
$\widehat{\text{se}}_{1s}$	0.0428	0.1613	0.2196	5.3837
$\widehat{\text{se}}_1$	0.0418	0.1612	0.2198	5.3762
$\widehat{\text{se}}$	0.0492	0.1547	0.2333	5.7289

The MM estimator has a 50% breakdown point and 95% efficiency in the Gaussian model. Variables: lagged salinity (x_1), a trend (x_2) and the volume of river discharge into the sound (x_3).

Using the procedure from the previous section, we would like to know whether the variance matrices obtained by the different methods are statistically different from the baseline HAC

estimator $\widehat{\text{Avar}}$. More precisely, we will perform a sequence of tests, based on $T_{j,B}$ with $j = 3, 2s, 1s, 1$, for the sequence of hypotheses H_0^j. The number of bootstrap samples was taken as $B = 1000$. Results are reported in Table 4, giving the values of the statistics for the MM-estimator. These values can be compared with the 1% critical value from the χ^2_{10} distribution, equal to 23.209. Since the sample size is rather small here, the critical value should not be taken too literally.

First we test the null hypothesis that there is no serial correlation nor heteroskedasticity and that the error term is normally distributed, meaning that we could use $\widehat{\text{Avar}}_3$ to estimate the variance of the estimated regression parameters. Testing H_0^3, using $T_{3,B}$, results in a strong rejection of this hypothesis. Similarly for H_0^{2s}. So one can conclude that the use of $\widehat{\text{Avar}}_3$ or $\widehat{\text{Avar}}_{2s}$ is not appropriate here. Testing H_0^1, which amounts to the absence of autocorrelation and hence implies the consistency of $\widehat{\text{Avar}}_1$, results in a clear rejection as well. This seems to be a bit surprising, since the standard errors computed under C_1 are close to the HAC ones in table 6. But let us not forget that the test statistic computes the difference between two estimates of asymptotic variance matrices, and not only of their diagonal elements. For instance, we see in this example that $\widehat{\text{Avar}}_{1,14} = -0.04275$ while $\widehat{\text{Avar}}_{14} = -0.10592$, a difference which is already greater. To conclude, the tests confirm that the HAC estimator is to be preferred here.

Table 4: Salinity data: testing alternative variance formulas for the MM-estimator

j	$T_{j,B}$
3	3423.6
$2s$	542.36
$1s$	92.186
1	38.072

χ^2_{10} test statistics for $H_0^j : \text{Avar}_j = \text{Avar}$.

8 Conclusion

In this paper the problem of the estimation of the standard errors of robust estimators is studied. A general expression for the asymptotic variance matrix of the regression estimator is presented, which is consistent under both heteroskedasticity and autocorrelation. Moreover, it can cope with regression outliers. Some theoretical calculations in Section 4 and the simulation study in Section 5 clearly showed the necessity for using such an estimator in order to avoid biased estimation of the limiting variance matrix. Incorrectly estimated standard errors bias the whole inference procedure. All expressions were obtained supposing that the regularity conditions for asymptotic normality were fulfilled. The formulas presented for estimating the variance

matrices of M-, S-, and MM-estimators can be directly used by practitioners and programmers of statistical software.

The three estimators of regression considered have an unbounded influence function with respect to good leverage points. To achieve a bounded influence a weighted version of the MM-estimator can be computed, where the weights measure the leverage of each observation. One of the first papers to consider robust leverage weights was Coackley and Hettmansperger (1993), for a (one-step) M-estimator. It can be shown that all formulas for the standard errors remain valid in this case, after inserting the weights for the leverages in all sums.

Simplifications of the asymptotic variance matrix arise when extra conditions are imposed on the error term. For example, when the data are collected in such a way that serial correlation can be excluded, a more simple estimator should be used, thereby yielding an efficiency gain. The simplest expression is obtained under the assumption that the errors are i.i.d. normal. When normality really holds, using the corresponding estimator decreases the variance of the estimate of the standard error (as shown by the simulations). To check whether the more simple expressions for the standard errors may be used, the testing procedure outlined in Section 6 can be used. We do not suggest such tests should be applied systematically, but in case of doubt between different expressions for the standard errors, the proposed test statistics provide guidance in selecting the most appropriate one. These tests are similar in spirit to the information matrix test, proposed by White (1982). White proposed a test for testing the equality of the variance matrices of a maximum likelihood and a quasi-maximum likelihood estimator.

The heteroskedasticity and autocorrelation consistent (HAC) estimator of the standard error does not require any functional specification of the heteroskedasticity nor any modelling of the correlation structure in the error terms. In Bianco, Boente and Di Rienzo (2000) robust estimation of regression models was discussed for a specified conditional variance function. Robust estimation of ARMA models has been considered, e.g. by Bustos and Yohai (1986).

The contributions of this paper are that (i) it gives an overview of different estimates of the standard errors of M-, S-, and MM-estimators and recommends the use of $\widehat{\text{Avar}}$ or $\widehat{\text{Avar}}_1$; (ii) it studies the bias and the efficiency loss of the different estimates through large sample calculations and simulations; (iii) it proposes a test procedure for testing the equality of two different expressions for the standard errors. Robust estimation of standard errors, in the HAC sense, has been the topic of numerous papers in econometrics. Also in generalised linear modelling one is aware of the problem, and so-called sandwich estimators of the standard errors are advocated . Strangely enough, robust standard errors for robust estimators have never been thoroughly studied. This paper tries to fill this gap.

Appendix

Asymptotic variance of the S-estimator

The most general expression is

$$\begin{aligned}\mathrm{Avar}(\hat{\beta}_S) = A_S \sum_{j=-\infty}^{\infty} E(\rho_t' \rho_{t-j}' X_t X_{t-j}') A_S - a_S \sum_{j=-\infty}^{\infty} E(\rho_t \rho_{t-j}' X_{t-j}') A_S \\ - A_S \sum_{j=-\infty}^{\infty} E(\rho_t' \rho_{t-j} X_t) a_S' + a_S \sum_{j=-\infty}^{\infty} E(\rho_t \rho_{t-j} - b^2) a_S',\end{aligned}$$

where

$$A_S = \sigma[E(\rho_t'' X_t X_t')]^{-1} \quad \text{and} \quad a_S = A_S \frac{E(\rho_t'' X_t \varepsilon_t)}{E(\rho_t' \varepsilon_t)}.$$

If condition C_1 holds, then

$$\begin{aligned}\mathrm{Avar}_1(\hat{\beta}_S) = A_S E(\rho_t'^2 X_t X_t') A_S - a_S E(\rho_t \rho_t' X_t') A_S - A_S E(\rho_t \rho_t' X_t) a_S' \\ + E(\rho_t^2 - b^2) a_S a_S'.\end{aligned}$$

If in addition C_2 holds, then

$$\begin{aligned}\mathrm{Avar}_2(\hat{\beta}_S) = \sigma^2 \frac{E(\rho_t'^2)}{[E(\rho_t'')]^2} [E(X_t X_t')]^{-1} + \sigma^2 \frac{E(\rho_t'' \varepsilon_t)}{[E(\rho_t'')]^2 E(\rho_t' \varepsilon_t)} \\ \left\{ \frac{E(\rho_t'' \varepsilon_t) E(\rho_t^2 - b^2)}{E(\rho_t' \varepsilon_t)} - 2E(\rho_t \rho_t') \right\} [E(X_t X_t')]^{-1} E(X_t) E(X_t') [E(X_t X_t')]^{-1}.\end{aligned}$$

Under C_s the expressions are the same as those for the MM-estimator, listed in Section 3, with $\psi = \rho'$.

Asymptotic variance of the M-estimator

Here the expressions are less explicit. The asymptotic variance $\mathrm{Avar}_j(\hat{\beta}_M)$ ($j = 1$ when C_1 holds; $j = 2$ when C_1 and C_2 hold) are given by the upper left $p \times p$ block of $G_j^{-1} \Omega_j (G_j^{-1})'$. If C_1 holds, then

$$\Omega_1 = E \begin{pmatrix} \psi_t^2 X_t X_t' & \rho_t \psi_t X_t \\ \rho_t \psi_t X_t' & \rho_t^2 - b^2 \end{pmatrix}$$

and G_1 is as in (3.4). If in addition C_2 holds, then

$$\Omega_2 = \begin{pmatrix} E(\psi_t^2) E(X_t X_t') & E(\rho_t \psi_t) E(X_t) \\ E(\rho_t \psi_t) E(X_t') & E(\rho_t^2) - b^2 \end{pmatrix}$$

and

$$G_2 = -\frac{1}{\sigma} \begin{pmatrix} E(\psi_t') E(X_t X_t') & E(\psi_t' \varepsilon_t) E(X_t) \\ E(\rho_t') E(X_t') & E(\rho_t' \varepsilon_t) \end{pmatrix}.$$

Under C_s the expressions are the same as those for the MM-estimator, listed in Section 3.

Table 5: Performance of alternative estimates of standard errors of M-estimates (1000 observations)

	$\widehat{\text{se}}_3$	$\widehat{\text{se}}_{2s}$	$\widehat{\text{se}}_{1s}$	$\widehat{\text{se}}_1$	$\widehat{\text{se}}$
	i.i.d. normal				
PB	0.0024	0.0024	0.0026	0.0029	-0.0031
RMSE	0.0321	0.0345	0.0517	0.0517	0.0671
RP	0.0499	0.0494	0.0503	0.0504	0.0518
	heteroskedasticity				
PB	-0.5669	-0.6778	0.0002	0.0011	-0.0047
RMSE	0.5679	0.6787	0.0627	0.0627	0.0761
RP	0.2660	0.3179	0.0509	0.0508	0.0536
	autocorrelation: AR(1)				
PB	-0.5169	-0.5160	-0.5185	-0.5179	-0.1041
RMSE	0.5198	0.5191	0.5238	0.5232	0.1426
RP	0.2412	0.2403	0.2438	0.2435	0.0791
	autocorrelation: MA(1)				
PB	-0.3074	-0.3071	-0.3090	-0.3086	-0.0460
RMSE	0.3160	0.3160	0.3211	0.3206	0.1177
RP	0.1452	0.1447	0.1477	0.1476	0.0621
	heteroskedasticity and autocorrelation				
PB	-1.0443	-1.1548	-0.4804	-0.4792	-0.0947
RMSE	1.0455	1.1559	0.4864	0.4852	0.1425
RP	0.4947	0.5413	0.2208	0.2205	0.0769

Results are based on 10000 Monte Carlo runs. Data are generated by $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ ($t = 1, ..., 1000$) under 5 different sampling schemes, described in the main text.

PB: Proportional Bias.

RMSE: Root-Mean-Squared Proportional Error.

RP: Rejection Probability of 5%-level t-test on slope parameter.

Table 6: Performance of alternative estimates of standard errors of S-estimates (1000 observations)

	$\widehat{\text{se}}_3$	$\widehat{\text{se}}_{2s}$	$\widehat{\text{se}}_{1s}$	$\widehat{\text{se}}_1$	$\widehat{\text{se}}$
	i.i.d. normal				
PB	-0.0491	-0.0499	-0.0482	-0.0473	-0.0529
RMSE	0.0589	0.0628	0.0838	0.0833	0.0966
RP	0.0616	0.0620	0.0624	0.0620	0.0630
	heteroskedasticity				
PB	-0.8231	-1.1183	-0.0197	-0.0168	-0.0224
RMSE	0.8238	1.1193	0.1415	0.1412	0.1482
RP	0.3842	0.5213	0.0614	0.0604	0.0627
	autocorrelation: AR(1)				
PB	-0.4907	-0.4913	-0.4913	-0.4901	-0.1133
RMSE	0.4937	0.4950	0.4999	0.4988	0.1623
RP	0.2301	0.2306	0.2310	0.2303	0.0799
	autocorrelation: MA(1)				
PB	-0.3034	-0.3040	-0.3042	-0.3032	-0.0739
RMSE	0.3122	0.3135	0.3204	0.3195	0.1430
RP	0.1454	0.1457	0.1462	0.1456	0.0691
	heteroskedasticity and autocorrelation				
PB	-1.0877	-1.3817	-0.2863	-0.2833	-0.0650
RMSE	1.0888	1.3830	0.3227	0.3201	0.1809
RP	0.5095	0.6234	0.1434	0.1422	0.0754

Results are based on 10000 Monte Carlo runs. Data are generated by $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ $(t = 1, ..., 1000)$ under 5 different sampling schemes, described in the main text.

PB: Proportional Bias.

RMSE: Root-Mean-Squared Proportional Error.

RP: Rejection Probability of 5%-level t-test on slope parameter.

Table 7: Performance of alternative estimates of standard errors of MM-estimates (200 observations)

	$\widehat{\text{se}}_3$	$\widehat{\text{se}}_{2s}$	$\widehat{\text{se}}_{1s}$	$\widehat{\text{se}}_1$	$\widehat{\text{se}}$
	i.i.d. normal				
PB	-0.0203	-0.0081	-0.0120	-0.0079	-0.0254
RMSE	0.0889	0.0754	0.1138	0.1139	0.1406
RP	0.0546	0.0497	0.0528	0.0510	0.0572
	heteroskedasticity				
PB	-1.1864	-1.0215	-0.0510	-0.0290	-0.0475
RMSE	1.1905	1.0268	0.2710	0.2735	0.2862
RP	0.5435	0.4782	0.0782	0.0731	0.0794
	autocorrelation: AR(1)				
PB	-0.5451	-0.5319	-0.5439	-0.5382	-0.1979
RMSE	0.5609	0.5459	0.5668	0.5616	0.2806
RP	0.2549	0.2499	0.2541	0.2511	0.1133
	autocorrelation: MA(1)				
PB	-0.3492	-0.3374	-0.3450	-0.3403	-0.1223
RMSE	0.3876	0.3745	0.3924	0.3885	0.2497
RP	0.1550	0.1496	0.1560	0.1541	0.0816
	heteroskedasticity and autocorrelation				
PB	-1.4821	-1.3171	-0.3648	-0.3414	-0.1375
RMSE	1.4876	1.3239	0.4646	0.4503	0.3556
RP	0.6524	0.5963	0.1813	0.1738	0.1048

Results are based on 10000 Monte Carlo runs. Data are generated by $Y_t = \beta_0 + \beta_1 X_t + \sigma\varepsilon_t$ $(t = 1, ..., 200)$ under 5 different sampling schemes, described in the main text.
PB: Proportional Bias.
RMSE: Root-Mean-Squared Proportional Error.
RP: Rejection Probability of 5%-level t-test on slope parameter.

References

[1] Bianco, A., Boente, G. & de Rienzo, J. (2000) Some results for robust GM-based estimators in heteroskedastic regression models. *Journal of Statistical Planning and Inference* 89, 215-242.

[2] Bustos, O.H. & Yohai, V.J. (1986) Robust estimates of ARMA models. *Journal of the American Statistical Association* 81, 155-169.

[3] Coakley, C.W. & T.P. Hettmansperger (1993) A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association* 88, 872-880.

[4] Croux, C., Van Aelst, S., & C. Dehon (2003) Bounded influence regression using high breakdown scatter matrices. *Annals of the Institute of Statistical Mathematics* 55, 265-285.

[5] Eicker, F. (1967) Limit theorems for regression with unequal and dependent regressors. In: L. leCam & J. Neyman, *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California, 59-82.

[6] Field, C., & J. Zhou (2003) Confidence intervals based on robust regression. *Journal of Statistical Planning and Inference* 115, 425-439.

[7] Greene, W.H. (2003) *Econometric analysis.* Englewood Cliffs (New Jersey): Prentice Hall.

[8] Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

[9] Hayashi, F. (2000) *Econometrics.* Princeton: Princeton University Press.

[10] Huber, P.J. (1981) *Robust Statistics.* New York: Wiley.

[11] Kauermann, G. & Carroll R.J. (2001) A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 96, 1387-1396.

[12] Liu, R.Y. (1988) Bootstrap procedures under some non-i.i.d. models. The Annals of Statistics 16, 1696-1708.

[13] Marazzi, A., Joss, J. & A. Randriamiharisoa (1993) *Algorithms, Routines, and S-Functions for Robust Statistics: the FORTRAN library ROBETH with an interface to S-PLUS.* New York: Chapman and Hall.

[14] Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

[15] Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-653.

[16] Rousseeuw, P.J. & A. Leroy (1987) *Robust regression and outlier detection.* New York: John Wiley.

[17] Rousseeuw, P.J. & V.J. Yohai (1984) Robust regression by means of S-estimators. In: J. Franke, W. Härdle & R.D. Martin (eds.), Robust and nonlinear time series analysis. *Lecture notes in Statistics* 26. New York: Springer.

[18] Ruppert, D. (1992) Computing S-estimators for regression and multivariate location and dispersion. *Journal of Computational Graphical Statistics* 1, 253-270.

[19] Salibián-Barrera, M. (2000) Contributions to the Theory of Robust Inference, Ph.D. thesis, The University of British Columbia.

[20] Salibián-Barrera, M. & R.H. Zamar (2004) Uniform asymptotics for robust location estimates when the scale is unknown. *Annals of Statistics*, to appear.

[21] Salibián-Barrera, M. & R.H. Zamar (2002) Bootstrapping robust estimates of regression. *The Annals of Statistics* 30, 556-582.

[22] Simpson, D.G., Ruppert, D. & R.J. Carroll (1992) On one-step GM estimates and stability of inferences in linear regression. *Journal of the American Statistical Association* 87, 439-450.

[23] Singh, K. (1998) Breakdown theory for bootstrap quantiles. *Annals of Statistics* 26, 1719-1732.

[24] Stock, J.H. & M.W. Watson (2003) *Introduction to Econometrics.* Boston: Addison-Wesley.

[25] White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817-838.

[26] White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-26.

[27] Yohai, V.J (1987) High breakdown point and high efficiency robust estimates for regression. *Annals of Statistics* 15, 642-656.

[28] Yohai, V.J. & Maronna, R. (1979) Asymptotic behavior of M-estimates for the linear model. *The Annals of Statistics* 7, 258-268.