Discussion Papers
Department of Economics
University of Copenhagen

No. 11-29

Asymptotic theory for iterated one-step Huber-skip estimators

Søren Johansen & Bent Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Asymptotic theory for iterated one-step Huber-skip estimators

Søren Johansen[1] & Bent Nielsen[2]

16 November 2011

Summary: Iterated one-step Huber-skip M-estimators are considered for regression problems. Each one-step estimator is a reweighted least squares estimators with zero/one weights determined by the initial estimator and the data. The asymptotic theory is given for iteration of such estimators using a tightness argument. The results apply to stationary as well as non-stationary regression problems.

Keywords: Huber-skip, iteration, one-step M-estimators, unit roots.
JEL Classification: C32.

1 Introduction

In regression analysis it is often an important concern to be able to detect outliers or other unsuspected structures. A very simple algorithm addressing this is first to obtain an initial estimator of the parameters, use this to discard observations with large residuals, and then run the regression. This is the one-step Huber-skip estimator. It is a special case of the one-step M-estimator in which the criterion function is not convex. The one-step Huber-skip estimator could be used as a new initial estimator when re-running the regression. We give an asymptotic fixed point result for such iterations of one-step Huber-skip estimators when the model has no outliers. The result is based on a tightness argument and allows regressors which are fixed, stationary, and non-stationary.

One-step M-estimators have been analysed previously in various situations: Bickel (1975), Jurečová and Sen (1996, Section 7.4) considered cases of smooth weight functions. Ruppert and Carroll (1980) considered one-step Huber-skip L-estimators. Welsh and Ronchetti (2002) analyse the one-step Huber-skip estimator when the initial estimator is the least squares estimator as well as one-step M-estimators with general initial estimator but smooth weight functions. Johansen and Nielsen (2009)

[1] Department of Economics, University of Copenhagen and CREATES, School of Economics and Management, Aarhus University, DK-8000 Aarhus C. E-mail: soren.johansen@econ.ku.dk.

[2] Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk. Financial support from the Institute for New Economic Thinking at the Oxford Martin School is gratefully acknowledged.

analyse one-step Huber-skip estimators for general initial estimators and stationary as well as non-stationary regressors.

Iterated one-step M-estimators are related to iteratively reweighted least squares estimators. Indeed the one-step Huber-skip estimator corresponds to a reweighted least squares estimator with weights of zero or unity. Dollinger and Staudte (1991) considered a situation with smooth weights, hence ruling out Huber-skips, and gave conditions for convergence. Their argument was cast in terms of influence functions. Our result for iteration of Huber-skip estimators is similar, but the employed tightness argument is different.

2 Definition of the one-step Huber-skip estimator

Consider the regression model

$$Y_t = \beta' X_t + \varepsilon_t \qquad t = 1, \ldots, T, \tag{2.1}$$

where X_t is a p-dimensional vector of regressors and the conditional distribution of the errors, ε_t, given $(X_1, \ldots, X_t, \varepsilon_1, \ldots, \varepsilon_{t-1})$ has density $\sigma^{-1}\mathsf{f}(\sigma^{-1}\varepsilon)$ so that $\sigma^{-1}\varepsilon_t$ are i.i.d. with known density f. The idea of the iterated one-step Huber-skip estimator is to start with some preliminary estimator $(\hat{\beta}, \hat{\sigma}^2)$ and seek to improve it through an iterative procedure by using it to identify outliers, discard these and then run a regression on the remaining observations.

The preliminary estimator $(\hat{\beta}, \hat{\sigma}^2)$ could be a least squares estimator on the full sample. Alternatively, the initial estimator could be chosen robustly. A candidate would be the least trimmed squares estimator of Rousseeuw (1984), Rousseeuw and Leroy (1987, p. 180). When the trimming proportion is at most a half this convergences in distribution at a usual $T^{1/2}$-rate as established by Víšek (2006).

The outliers are identified by first choosing a ψ giving the proportion of good, central observation and then introducing two critical values $\underline{c}$ and $\overline{c}$ so

$$\int_{\underline{c}}^{\overline{c}} \mathsf{f}(v)dv = \psi \qquad \text{and} \qquad \int_{\underline{c}}^{\overline{c}} v\mathsf{f}(v)dv = 0. \tag{2.2}$$

This can also be written as $\tau_0 = \psi$ and $\tau_1 = 0$, where τ_k are the truncated moments

$$\tau_k = \int_{\underline{c}}^{\overline{c}} v^k \mathsf{f}(v)dv \qquad \text{for } k \in \mathbb{N}_0.$$

Observations are retained if their residuals $Y_t - \hat{\beta}' X_t$ are in the interval from $\underline{c}\hat{\sigma}w_{Tt}$ to $\overline{c}\hat{\sigma}w_{Tt}$ where w_{Tt}^2 could be chosen for instance as 1 or as $1 - X_t'(\sum_{s=1}^T X_s X_s')^{-1} X_t$.

The one-step Huber-skip estimators, $\hat{\beta}_m$ and $\hat{\sigma}_m$, are the least squares estimator of Y_t on X_t among the retained observations. If $\hat{\beta}, \hat{\sigma}^2$ are denoted $\hat{\beta}_{m-1}, \hat{\sigma}^2_{m-1}$ then the one-step Huber-skip estimators, $\hat{\beta}_m$ and $\hat{\sigma}^2_m$, are defined recursively for $m \in \mathbb{N}$ as

$$\hat{\beta}_m = S_{xx}^{-1} S_{xy}, \qquad \hat{\sigma}^2_m = (\tau_2/\psi)^{-1} S_{11}^{-1}(S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}) \tag{2.3}$$

where, for $g_t, h_t \in (1, X_t, Y_t)$, then

$$S_{gh} = \sum_{t=1}^{T} g_t h_t' 1_{(\hat{\sigma}_{m-1} w_{Tt} \underline{c} \leq Y_t - X_t' \hat{\beta}_{m-1} \leq \hat{\sigma}_{m-1} w_{Tt} \bar{c})}. \tag{2.4}$$

The correction factor $(\tau_2/\psi)^{-1}$ is needed to obtain consistency. The m times iterated one-step Huber-skip estimator will be considered. Note that the iterateration has the property that the set of retained observations can change in each iteration step.

The main asymptotic results concern the convergence with increasing m when T is sufficiently large. Thus a normalisation matrix N_T in T is needed to normalize the regressors. If (Y_t, X_t) is stationary then $N_T = T^{-1/2} I_p$. If (Y_t, X_t) is trending a different normalisation is needed. For a linear trend component the normalisation would be $T^{3/2}$ and for a random walk component it would be T. Limiting matrices Σ, μ can then be introduced so

$$N_T \sum_{t=1}^{T} X_t X_t' N_T' \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0, \qquad T^{-1/2} N_T \sum_{t=1}^{T} X_t \overset{\mathsf{D}}{\to} \mu.$$

Note that Σ and μ may be stochastic as for instance when X_t is a random walk. The estimation errors are denoted

$$\hat{u}_{m,T} = \left\{ \begin{array}{c} (N_T^{-1})'(\hat{\beta}_m - \beta) \\ T^{1/2}(\hat{\sigma}_m - \sigma) \end{array} \right\}. \tag{2.5}$$

Introduce also coefficient matrices

$$\Psi_1 = \begin{pmatrix} \psi\Sigma & 0 \\ 0 & 2\tau_2 \end{pmatrix}, \qquad \Psi_2 = \begin{pmatrix} \xi_1 \Sigma & \xi_2 \mu \\ \zeta_2 \mu' & \zeta_3 \end{pmatrix},$$

where $\xi_n = (\bar{c})^n \mathsf{f}(\bar{c}) - (\underline{c})^n \mathsf{f}(\underline{c})$ and $\zeta_n = \xi_n - \xi_{n-2}\tau_2/\psi$, so that

$$\rho = \Psi_1^{-1} \Psi_2 = \begin{pmatrix} \psi^{-1} \xi_1 I_p & \psi^{-1} \xi_2 \Sigma^{-1} \mu \\ (2\tau_2)^{-1} \zeta_2 \mu' & (2\tau_2)^{-1} \zeta_3 \end{pmatrix},$$

along with a kernel

$$K_T = \Psi_1^{-1} \sum_{t=1}^{T} \left\{ \begin{array}{c} N_T X_t \varepsilon_t \\ T^{-1/2}(\varepsilon_t^2 - \sigma^2 \tau_2/\psi) \end{array} \right\} 1_{(\underline{c}\sigma \leq \varepsilon_t \leq \sigma \bar{c})}. \tag{2.6}$$

The asymptotic analysis of Johansen and Nielsen (2009) shows that the one-step estimators $\hat{\beta}_m, \hat{\sigma}^2_m$ satisfy the one-step equation

$$\hat{u}_{m,T} = \rho\hat{u}_{m-1,T} + K_T + R_T(\hat{u}_{m-1,T}), \tag{2.7}$$

for some remainder term $R_T(\hat{u}_{m-1,T})$. In this notation it is emphasised that the remainder term is a function of the previous estimator $\hat{u}_{m-1,T}$. Indeed, $R_T(\hat{u}_{m-1,T})$ is defined from the equation (2.7) where $\hat{u}_{m,T}$ is a function of the data and $\hat{u}_{m-1,T}$ through (2.3), (2.4) and K_T is a function of the innovations. A precise definition is given in Lemma A.1 in the Appendix.

Moreover, it will be shown that through infinite iteration then, for fixed T, and $m \to \infty$ it holds

$$\hat{u}_{m,T} \overset{\mathsf{P}}{\to} \hat{u}^*_T$$

where $\hat{u}^*_T = (I_{1+p} - \rho)^{-1} K_T$ satisfies the equation

$$\hat{u}^*_T = \rho\hat{u}^*_T + K_T. \tag{2.8}$$

3 The fixed point result

The fixed point result is primarily a tightness results. Thus, for the moment, only tightness of the kernel K_T is needed, and it is not necessary to establish the limiting distribution. The necessary assumptions are therefore fairly general. The Euclidean norm for vectors x is denoted $|x|$.

Assumption A *Suppose the initial estimator satisfies*

$$T^{1/2}(\hat{\sigma}^2_0 - \sigma^2), (N_T^{-1})'(\hat{\beta}_0 - \beta) = \mathsf{O}_\mathsf{P}(1).$$

Assumption B *Consider the model* (2.1). *Suppose there exists weights* $w_{t,T}$, *and non-stochastic normalisation matrices* $N_T \to 0$, *so that*
(i) *The weights satisfy* $\max_{t\leq T} T^{1/2}|w_{tT} - 1| = \mathsf{o}_\mathsf{P}(1)$.
(ii) *The regressors satisfy, jointly,*
(a) $N_T \sum_{t=1}^T X_t X_t' N_T' \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0$,
(b) $T^{-1/2} N_T \sum_{t=1}^T X_t \overset{\mathsf{D}}{\to} \mu$,
(c) $\max_{t\leq T} \mathsf{E}|T^{1/2} N_T X_t|^4 = \mathsf{O}(1)$.
(iii) *The density* f *has continuous derivative* f' *and satisfies*
(a) $\sup_{v\in\mathbb{R}}\{(1+v^4)\mathsf{f}(v) + (1+v^2)|\mathsf{f}'(v)|\} < \infty$,
(b) *it has mean zero, variance one, and finite fourth moment,*
(c) $\overline{c}, \underline{c}$ *are chosen so* $\tau_0 = \psi$ *and* $\tau_1 = 0$.

The first result is a tightness result for the kernel. The proof uses Chebychev's inequality. The details of the proof are given in the appendix.

Theorem 3.1 *Suppose Assumption B$(iic, iiib)$ holds. Then* $K_T = \mathrm{O}_{\mathsf{P}}(1)$.

Next, the remainder term $R_T(u)$ is shown to vanish uniformly in $|u| < U$. The proof involves a chaining argument which was given in Johansen and Nielsen (2009), but the result is written in a slightly different way as discussed in the appendix.

Theorem 3.2 *Suppose Assumption B holds. Then, for all $U > 0$ and $T \to \infty$ it holds*

$$\sup_{|u|\leq U} |R_T(u)| = \mathrm{o}_{\mathsf{P}}(1).$$

As a corollary to this result equation (2.7) reduces to

$$\hat{u}_{1,T} = \rho \hat{u}_{0,T} + K_T + \mathrm{o}_{\mathsf{P}}(1),$$

when Assumptions A, B are satisfied.

The fixed point result is now given. Initially a tight estimator $(\hat{\beta}_0, \hat{\sigma}_0^2)$ is available. This is iterated through the one-step equation (2.7). Theorem 3.3 shows that the estimator converges in probability to the solution of the fixed point equation (2.8).

Theorem 3.3 *Suppose Assumptions A, B hold and* $\max|\mathrm{eigen}(\rho)| < 1$. *Then*

$$\limsup_{m\to\infty} |\hat{u}_{m,T} - \hat{u}_T^*| = \mathrm{o}_{\mathsf{P}}(1).$$

The idea of iterating the one step estimator is also found in Cavaliere and Georgiev (2011, Theorem 4). They consider, however, a completely different setup of a first order autoregression with infinite variance innovations, a root close to one, and known scale. The idea of the proof of Theorem 3.3 is to argue that if the initial estimator $\hat{u}_{0,T}$ takes values in a large compact set with large probability then, due to the iteration, outcomes of $\hat{u}_{m,T}$ takes values in the same compact set while $|\hat{u}_{m,T} - (I_{p+1} - \rho)^{-1} K_T|$ is the sum of two terms vanishing exponentially and in probability, respectively. The details are given in the appendix. A necessary condition for the result is that the autoregressive coefficient matrix ρ is contracting. Therefore ρ is analyzed next.

Theorem 3.4 *The autoregressive coefficient matrix ρ has $p-1$ eigenvalues equal to $\xi_1\psi$ and two eigenvalue solving*

$$\lambda^2 - (\frac{\zeta_3}{2\tau_2} + \frac{\xi_1}{\psi})\lambda + \frac{1}{2\tau_2\psi}(\zeta_3\xi_1 - \zeta_2\xi_2\mu'\Sigma^{-1}\mu) = 0.$$

When f is symmetric then $\xi_2 = 0$ and ρ is the diagonal matrix $\mathrm{diag}\{I_p\xi_1/\psi, \zeta_3/(2\tau_2)\}$. Further results can then be given about the eigenvalues.

Theorem 3.5 *Suppose* f *is symmetric with third moments,* $\mathsf{f}'(c) \leq 0$ *for* $c > 0$ *and* $\lim_{c\to 0} \mathsf{f}''(c) < 0$. *Then*
(a) $0 < \xi_1/\psi < 1$ *for* $0 < \psi < 1$ *while* $\lim_{\psi\to 0} \xi_1/\psi = 1$ *and* $\lim_{\psi\to 1} \xi_1/\psi = 0$;
(b) $0 < \zeta_3^\psi/(2\tau_2^\psi)$ *for* $0 < \psi < 1$ *and* $\lim_{\psi\to 0} \zeta_3/(2\tau_2) = 1$ *and* $\lim_{\psi\to 1} \zeta_3/(2\tau_2) = 0$;
(c) if $[c\{\log \int_0^c \mathsf{f}(x)dx\}']' < 0$ *for* $c > 0$ *then* $\zeta_3/(2\tau_2) < 1$ *for* $0 < \psi < 1$;
(d) $\{\log \mathsf{f}(c)\}'' < 0 \Rightarrow [c\{\log \mathsf{f}(c)\}']' < 0 \Rightarrow [c\{\log \int_0^c \mathsf{f}(x)dx\}']' < 0$.

The condition $[c\{\log \int_0^c \mathsf{f}(x)dx\}']' < 0$ is satisfied for the Gaussian density which is log-concave and by t-densities which are not log-concave but satisfy $[c\{\log \mathsf{f}(c)\}']' < 0$. In the robust statistics literature Rousseuw (1982) uses the condition $[c\{\log \mathsf{f}(c)\}']' < 0$ when discussing change-of-variance curves for M-estimators and assumes log concave densities.

A consequence of Theorem 3.5 is that the roots of the coefficient matrix ρ are bounded away from unity for all compact subsets of the half open set $0 < \psi \leq 1$. The uniform distribution on $[-a, a]$ provides an example where ρ is not contracting since in this situation $\xi_1 = \psi$ over the entire support. However, the weak unimodality condition $\mathsf{f}'(c) \leq 0$ in Theorem 3.5 is not necessary as long as the mode at the origin is large in comparison to other modes.

Remark 3.6 *In the robustness literature there has been considerable discussion of the pure location case where the scale is known so* $\sigma = 1$. *The above results carry through. To write down the new result let*

$$\hat{b}_{m,T} = (N_T^{-1})'(\hat{\beta}_m - \beta), \qquad K_{b,T} = (\psi\Sigma)^{-1}\sum_{t=1}^T N_T X_t \varepsilon_t 1_{(\underline{c}\sigma<\varepsilon_t\leq\overline{c}\sigma)},$$

so that the 1-step equation (2.7) becomes

$$\hat{b}_{m,T} = \psi^{-1}\xi_1\hat{b}_{m-1,T} + K_{b,T} + R_{b,T}(\hat{b}_{m-1,T}), \tag{3.1}$$

where $\sup_{|b|<U} |R_{b,T}(b)| = \mathrm{o}_{\mathsf{P}}(1)$. *This equation is therefore the same as equation (2.7) with the estimation error for the scale set to zero. The fixed point equation (2.8) becomes*

$$\hat{b}_T^* = \psi^{-1}\xi_1\hat{b}_T^* + K_{b,T}. \tag{3.2}$$

This equation is the same as the location part of the general location-scale fixed point equation (2.8) when either the density is symmetric or the estimation uncertainty for the scale is set to zero. It has solution

$$\hat{b}_T^* = \frac{1}{\psi - \xi_1}\Sigma^{-1}K_{b,T}. \tag{3.3}$$

4 Distribution of the kernel

Due to the fixed point equation (2.8) the fully iterated one-step estimator is

$$u_T^* = (I_{p+1} - \rho)^{-1} K_T.$$

Thus for the distributional analysis it suffices to analyse the distribution of the kernel K_T. We do this in a few situations.

Stationary case. Suppose the regressors are fixed or arise from a stationary time series model. Then the limits Σ, μ in Assumption B(i) are deterministic. The Central Limit Theorem then shows that

$$K_T \overset{\mathsf{D}}{\to} \Psi_1^{-1} \mathsf{N}_{p+1}(0, \Phi), \tag{4.1}$$

where

$$\Phi = \left[\begin{array}{cc} \Sigma\sigma^2\tau_2 & \mu\sigma^3\tau_3 \\ \mu'\sigma^3\tau_3 & \sigma^4\{\tau_4 - (\tau_2)^2\psi^{-1}\} \end{array} \right]. \tag{4.2}$$

As a consequence the fully iterated estimator has limiting distribution

$$u_T^* = (\Psi_1 - \Psi_2)^{-1}\Psi_1 K_T \overset{\mathsf{D}}{\to} (\Psi_1 - \Psi_2)^{-1} \mathsf{N}_{p+1}(0, \Phi). \tag{4.3}$$

In the special case where the errors are symmetric then the fully iterated estimator has limiting distribution

$$(N_T^{-1})'(\hat{\beta}^* - \beta) = \frac{\Sigma^{-1}}{(\psi - \xi_1)} \sum_{t=1}^T N_T X_t \varepsilon_t 1_{(-c\sigma \le \varepsilon_t \le \sigma c)} \overset{\mathsf{D}}{\to} \mathsf{N}_p\{0, \frac{\sigma^2\tau_2\Sigma^{-1}}{(\psi - \xi_1)^2}\}, \tag{4.4}$$

noting that $\psi > \xi_1$ is satisfied for symmetric, unimodal distributions by Theorem 3.5(a). This limiting distribution also applies in the symmetric, pure location case, see Remark 3.6. It is also seen elsewhere in the robust statistic literature.

First, Víšek (2006, Theorem 1, p. 215) analysed the least trimmed squares estimator of Rousseeuw (1984). The estimator is given by

$$\hat{\beta}^{LTS} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{\mathrm{int}(T\psi)} r_{(t)}^2$$

where $r_{(1)}^2 < \cdots < r_T^{(2)}$ are the ordered squared residuals $r_t = Y_t - X_t'\beta$. The estimator has the property that it does not depend on the scale of the problem. Víšek showed that in the symmetric case the least trimmed squares estimator satisfies

$$(N_T^{-1})'(\hat{\beta}^{LTS} - \beta) = \frac{\Sigma^{-1}}{(\psi - \xi_1)} \sum_{t=1}^T N_T X_t \varepsilon_t 1_{(-c\sigma \le \varepsilon_t \le c\sigma)} + \mathsf{o}_{\mathsf{P}}(1).$$

With Remark 3.6 in mind it is seen that the leading term of $\hat{\beta}^{LTS}$ solves the fixed point equation (3.2). Thus, if in the case of known scale $\hat{\beta}^{LTS}$ is chosen as the initial estimator, then the distribution of the 1-step M-estimator equals that of the initial estimator apart from terms which are $\mathsf{o}_{\mathsf{P}}(1)$.

Secondly, Huber (1964, p. 79) considered a pure location problem without regressors so $X_t = 1$ and $\sigma = 1$. He suggested estimating the location β by the M-estimator, which in the symmetric case, minimizes the equation

$$\hat{\beta}^M = \arg\min_{\beta} \sum_{t=1}^T (Y_t - \beta)^2 1_{(-c<Y_t-\beta\leq c)}.$$

He conjectured that the variance of the limiting distribution would be $\tau_2/(\psi - \xi_1)^2$, matching the limit distribution of the iterated 1-step M-estimator as found in (4.4). A formal theory is given in Jurečová and Sen (1996, Theorem 5.3.3) showing that

$$T^{1/2}(\hat{\beta}^M - \beta) = \frac{T^{-1/2}}{(\psi - \xi_1)} \sum_{t=1}^T \varepsilon_t 1_{(-c\leq\varepsilon_t\leq c)} + \mathsf{O}_{\mathsf{P}}(T^{-1/4}).$$

Thus, as a complement to Theorem 3.3, it follows that

$$\limsup_{m\to\infty} \mathsf{P}(|\hat{\beta}_{m,T} - \hat{\beta}_T^M| > \eta) < \epsilon.$$

A consequence of this result is that the iterated 1-step M-estimator has the same limiting distribution as the M-estimator.

Deterministic trends. As a simple example consider the regression

$$Y_t = \beta_1 + \beta_2 t + \varepsilon_t,$$

where $\varepsilon_t \in \mathbb{R}$ satisfies Assumption B(iii). Define the normalisation

$$N_T = \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix}.$$

Then Assumption B(ii) is met with $X_t = (1, t)'$ and

$$\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}, \qquad \mu = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix},$$

and $\max_{t\leq T} \mathsf{E}|T^{1/2} N_T X_t|^4 \leq 4$. The kernel then has a limiting distribution given by (4.1) where the matrix Φ in (4.2) is computed in terms of the Σ and μ derived immediately above.

Trend stationary autoregressions. The derivation is in principle similar to the deterministic trend case but involve a notationally tedious detrending argument. The argument is similar to that of Johansen and Nielsen (2009, Section 1.5.1).

Unit roots. Consider the autoregression $Y_t = \beta Y_{t-1} + \varepsilon_t$ where $\beta = 1$. To derive the asymptotic distribution of the kernel note that the autoregression implies that $X_t = Y_{t-1} = Y_0 + \sum_{s=1}^{t-1} \varepsilon_s$. Thus let $N_T = T^{-1}$. By the functional Central Limit Theorem then

$$T^{-1/2} \sum_{t=1}^{\text{int}(Tu)} \left\{ \begin{array}{l} \varepsilon_t \\ \varepsilon_t 1_{(\underline{c}\sigma \leq \varepsilon_t \leq \sigma\overline{c})} \\ (\varepsilon_t^2 - \sigma^2\tau_2/\psi) 1_{(\underline{c}\sigma \leq \varepsilon_t \leq \sigma\overline{c})} \end{array} \right\} \overset{\mathsf{D}}{\to} \left(\begin{array}{c} W_{x,u} \\ W_{1,u} \\ W_{2,u} \end{array} \right),$$

where the limit is a Brownian motion with zero mean and variance

$$\Phi_W = \left[\begin{array}{ccc} \sigma^2 & \sigma^2\tau_2 & \sigma^3\tau_3 \\ \sigma^2\tau_2 & \sigma^2\tau_2 & \sigma^3\tau_3 \\ \sigma^3\tau_3 & \sigma^3\tau_3 & \sigma^4\{\tau_4 - (\tau_2)^2/\psi\} \end{array} \right].$$

Thus the limiting variables Σ and μ in Assumption B(ii) are

$$\Sigma = \int_0^1 W_{x,u}^2 du, \qquad \mu = \int_0^1 W_{x,u} du,$$

while the kernel has limiting distribution

$$K_T \overset{\mathsf{D}}{\to} \Psi_1^{-1} \left(\begin{array}{c} \int_0^1 W_{x,u} dW_{1,u} \\ W_{2,1} \end{array} \right).$$

Thus, when the density of ε_t is symmetric, the fully iterated estimator for β will have limiting distribution

$$T(\hat{\beta}^* - \beta) \overset{\mathsf{D}}{\to} \frac{\int_0^1 W_{x,u} dW_{1,u}}{(\psi - \xi_1) \int_0^1 W_{x,u}^2 du}.$$

When $\psi \to 1$ then $\xi_1 \to 0$ and $\tau_2 \to 1$ so $W_{1,u}$ and $W_{x,u}$ become identical and the limiting distribution becomes the usual Dickey-Fuller distribution. See also Johansen and Nielsen (2009, Section 1.5.4) for a related and more detailed derivation.

5 Discussion

The iteration result in Theorem 3.3 will have a variety of applications. An issue of interest in the literature is whether a slow initial convergence rate can be improved upon through iteration. This would open up for using robust estimators converging for instance at a $T^{1/3}$ rate as initial estimator. Such a result would complement the result of He and Portney (1992) who find that the convergence rate cannot be improved in a single step. The key would be to show that the remainder term of the one-step estimator in Theorem 3.2 remains small in an appropriate neighbourhood. The proof of Theorem 3.3 will then apply more or less in the same way leading to the same fixed point result.

A related algorithm is the *Forward Search* of Atkinson, Riani and Cerioli (2004, 2010). This involves finding an initial set of 'good' observations using for instance the least trimmed squares estimator of Rousseeuw (1984) and then increase the number of 'good' observations using a recursive test procedure. The algorithm involves iteration of one-step Huber-skip estimators, see Johansen and Nielsen (2010). Again the key to its analysis would be to improve Theorem 3.2, in this instance to hold uniformly in the cut-off fraction ψ. We are currently working on proving such generalisations of Theorem 3.2. Another algorithm of interest would be to analyse algorithms such as *Autometrics* of Hendry and Krolzig (2005) and Doornik (2009) which involves selection over observations as well as regressors.

A Proofs

Proof of Theorem 3.1. Chebychev's inequality gives $\mathsf{P}(|K_T| > C) \leq C^{-2}\mathsf{E}|K_T|^2$. Since K_T is a martingale then $\mathsf{E}|K_T|^2 = \sum_{t=1}^T \mathsf{E}(N_T X_t X_t' N_T')\mathsf{E}\{\varepsilon_t^2 1_{(\varepsilon_t < x\sigma)}\}$. Due to assumptions $(iic), (iiib)$ this is bounded. Thus, for all $\epsilon > 0$ then C can be chosen so large that $\mathsf{P}(|K_T| > C) < \epsilon$. ■

The key to the proving Theorem 3.2 is to understand the remainder terms of the moment matrices. This was done in Johansen and Nielsen (2009). As that paper was concerned only with the convergence of the 1-step estimator the main Theorem 1.1 simply stated that the remainder terms vanishes as $T \to \infty$. A more detailed result can, however, be extracted from the proof. To draw that out let a and b be the scale and location coordinates of u, respectively, and define product moment matrices

$$\tilde{S}_{gh}(u) = \sum_{t=1}^T g_t h_t' 1_{\{(\sigma + T^{-1/2}a)w_{Tt}\underline{c} < \varepsilon_t - X_t' N_T' b \leq (\sigma + T^{-1/2}a)w_{Tt}\overline{c}\}},$$

for $g_t, h_t \in (1, X_t, Y_t)$. The original product moment matrices in (2.4) then satisfy $S_{gh} = \tilde{S}_{gh}\{(N_T^{-1})'(\hat{\beta} - \beta), T^{1/2}(\hat{\sigma} - \sigma)\}$.

Lemma A.1 *Suppose Assumption B holds. Define the remainder terms* $R_{11}(u)$,

$R_{XX}(u)$, $R_{X1}(u)$, $R_{X\varepsilon}(u)$, and $R_{\varepsilon\varepsilon}(u)$ by the equations

$$\begin{aligned}
T^{-1}\tilde{S}_{11}(u) &= \psi + R_{11}(u), \\
N_T\tilde{S}_{XX}(u)N_T' &= \psi\Sigma + R_{XX}(u), \\
T^{-1/2}N_T\tilde{S}_{X1}(u) &= \psi\mu' + R_{X1}(u), \\
\left[\begin{array}{c} N_T\tilde{S}_{X\varepsilon}(u) \\ T^{-1/2}\{\psi\tilde{S}_{\varepsilon\varepsilon}(u) - \sigma^2\tau_2\tilde{S}_{11}(u)\} \end{array}\right] &= \sum_{t=1}^{T}\left\{\begin{array}{c} N_TX_t\varepsilon_t \\ T^{-1/2}(\psi\varepsilon_t^2 - \sigma^2\tau_2) \end{array}\right\}1_{(\underline{c}\sigma<\varepsilon_t\leq\bar{c}\sigma)} \\
&\quad + \begin{pmatrix} \xi_1\Sigma & \xi_2\mu \\ \zeta_2\mu' & \zeta_3\psi \end{pmatrix}\begin{pmatrix} b \\ a \end{pmatrix} + \left\{\begin{array}{c} R_{X\varepsilon}(u) \\ R_{\varepsilon\varepsilon}(u) \end{array}\right\},
\end{aligned}$$

where, for notational convenience, the dependence of T is suppressed. Then for all $U > 0$ and $T \to \infty$ it holds that

$$\sup_{|u|<U}\{|R_{11}(u)| + |R_{XX}(u)| + |R_{X1}(u)| + |R_{X\varepsilon}(u)| + |R_{\varepsilon\varepsilon}(u)|\} = \mathrm{o}_{\mathsf{P}}(1). \quad \text{(A.1)}$$

Proof of Lemma A.1. Theorem 1.1 in Johansen and Nielsen (2009) states that $|R_{11}(u)|$, $|R_{XX}(u)|$, $|R_{X1}(u)|$, $|R_{\varepsilon}(u)|$, $|R_{\varepsilon\varepsilon}(u)|$ vanish when u is evaluated at $\hat{u} = \{(N_T^{-1})'(\hat{\beta} - \beta), T^{1/2}(\hat{\sigma} - \sigma)\}$ under the assumption that $\hat{u} = \mathrm{O}_{\mathsf{P}}(1)$, as $T \to \infty$. The proof of that result then progresses by noting that assumption $\hat{u} = \mathrm{O}_{\mathsf{P}}(1)$ means that for all $\epsilon > 0$ then a U exists so $\mathsf{P}(|u| \geq U) < \epsilon$ and therefore for it suffices to prove that (A.1) holds. Therefore the proof of that theorem continues to prove precisely the statement (A.1), which is the desired result here. ■

Proof of Theorem 3.2. The updated estimator is defined in (2.3) in terms of the product moment statistics $S_{vw} = \tilde{S}_{vw}(\hat{u})$ where $\hat{u} = \{(N_T^{-1})'(\hat{\beta} - \beta), T^{1/2}(\hat{\sigma} - \sigma)\}$ and it is given by

$$\begin{aligned}
(N_T^{-1})'(\hat{\beta}_m - \beta) &= (N_TS_{XX}N_T')^{-1}N_TS_{X\varepsilon}, \\
T^{1/2}(\hat{\sigma}_m^2 - \sigma^2) &= (\tau_2S_{11})^{-1}T^{1/2} \\
&\qquad \times\{\psi S_{\varepsilon\varepsilon} - \sigma^2\tau_2S_{11} - \psi S_{\varepsilon X}N_T'(N_TS_{XX}N_T')^{-1}N_TS_{X\varepsilon}\}.
\end{aligned}$$

Insert the definitions from Lemma A.1 to get

$$\begin{aligned}
(N_T^{-1})'(\hat{\beta}_m - \beta) &= \{\psi\Sigma + R_{XX}(\hat{u})\}^{-1} \\
&\qquad \times \{\sum_{t=1}^{T}(N_TX_t\varepsilon_t)1_{(\underline{c}\sigma<\varepsilon_t\leq\bar{c}\sigma)} + \xi_1\Sigma\hat{b} + \xi_2\mu\hat{a} + R_{X\varepsilon}(\hat{u})\}.
\end{aligned}$$

Since $\sum_{t=1}^{T}(N_TX_t\varepsilon_t)1_{(\underline{c}\sigma<\varepsilon_t\leq\bar{c}\sigma)}$ is tight by Theorem 3.1, $\hat{u}$ is $\mathrm{O}_{\mathsf{P}}(1)$ and the remainders are vanishing by Lemma A.1 for $T \to \infty$, then

$$(N_T^{-1})'(\hat{\beta}_m - \beta) = (\psi\Sigma)^{-1}\sum_{t=1}^{T}(N_TX_t\varepsilon_t)1_{(\underline{c}\sigma<\varepsilon_t\leq\bar{c}\sigma)} + (\psi\Sigma)^{-1}(\xi_1\Sigma\hat{b} + \xi_2\mu\hat{a}) + R_{b,T}(\hat{u}),$$

where $\sup_{|u|<U} |R_{b,T}(u)| = \mathsf{o}_{\mathsf{P}}(1)$. A similar argument shows

$$T^{1/2}(\hat{\sigma}_m^2 - \sigma^2) = (\psi\tau_2)^{-1}T^{-1/2}\sum_{t=1}^{T}(\psi\varepsilon_t^2 - \sigma^2\tau_2)1_{(\sigma\underline{c}<\varepsilon_t\leq\sigma\overline{c})} \\ + \tau_2^{-1}(\zeta_2\mu'\hat{b} + \zeta_3\hat{a}) + R_{a,T}(\hat{u}),$$

where $\sup_{|u|<U} |R_{a,T}(u)| = \mathsf{o}_{\mathsf{P}}(1)$. Since $\hat{\sigma}_m^2 - \sigma^2$ vanishes, then Taylor expanding $(y + \sigma^2)^{1/2} - \sigma = y/2 + \mathrm{O}(y^2)$ shows that $\hat{\sigma}_j - \sigma$ and $(\hat{\sigma}_j^2 - \sigma^2)/2$ have the same limiting behaviour. ■

Proof of Theorem 3.3. We want to show that for all $\eta, \epsilon > 0$ there is a T_0 and m_0 so that for $T \geq T_0$ and $m \geq m_0$ we have for $u_T^* = (I_{p+1} - \rho)^{-1}K_T$, and prove

$$P(|\hat{u}_{m,T} - (I_{p+1} - \rho)^{-1}K_T| > \eta) < \epsilon, \tag{A.2}$$

and we start by showing

$$\sup_{0\leq m<\infty} |\hat{u}_{m,T}| = \mathrm{O}_{\mathsf{P}}(1). \tag{A.3}$$

Matrix norm: For matrices M choose the spectral norm $||M|| = \max\{\text{eigen}(M'M)\}^{1/2}$, so $||x|| = |x|$ for vectors x. We will use that the spectral norm and the Euclidean norm are compatible so $|Mx| \leq ||M||\,|x|$ as well as Gelfand's formula $\lim_{m\to\infty} ||M^m||^{1/m} = \max\{\text{eigen}(M)\}$, see Varga (2000, Theorems 1.5, 3.4).

Proof of (A.3): From the recursion (2.7) we find the representation

$$\hat{u}_{m+1,T} = \rho^{m+1}\hat{u}_{0,T} + \sum_{\ell=0}^{m}\rho^{\ell}\{K_T + R_T(\hat{u}_{m-\ell,T})\} \tag{A.4}$$

and the evaluation

$$|\hat{u}_{m+1,T}| \leq ||\rho^{m+1}||\,|\hat{u}_{0,T}| + (|K_T| + \max_{0\leq\ell\leq m}|R_T(\hat{u}_{\ell,T})|)\sum_{\ell=0}^{m}||\rho^{\ell}||.$$

By assumption a δ exists so $\max|\text{eigen}(\rho)| < \delta < 1$. Gelfand's formula then shows there is an $m_0 > 0$ so for all $m > m_0$ then $||\rho^m|| \leq \delta^m$. This in turn implies for some $c > 1$ then $\max_{0\leq m<\infty}||\rho^m|| < c$ and $\sum_{\ell=0}^{\infty}||\rho^{\ell}|| < c$, and hence

$$|\hat{u}_{m+1,T}| \leq c\{|\hat{u}_{0,T}| + |K_T| + \max_{0\leq\ell\leq m}|R_T(\hat{u}_{\ell,T})|\}. \tag{A.5}$$

Because it is assumed that $\hat{u}_{0,T}$ is tight, and K_T is tight by Theorem 3.1, and $\max_{|u|\leq U_1}|R_T(u)| = \mathsf{o}_{\mathsf{P}}(1)$ by Theorem 3.2, then constants $U_0, T_0 > 0$ exist so that for $T \geq T_0$, the set

$$\mathcal{A}_T = (c|\hat{u}_{0,T}| \leq U_0) \cap (c|K_T| \leq U_0) \cap (c\max_{|u|\leq 3U_0}|R_T(u)| \leq \eta/2)$$

has probability larger than $1-\epsilon$.

An induction over m is now used to show that $\sup_{0\leq m<\infty}|\hat{u}_{m,T}|\leq 3U_0$ on the set $\mathcal{A}_T$. As induction start, for $m=0$, then $|\hat{u}_{0,T}|\leq c^{-1}U_0<3U_0$ by the tightness assumption to $\hat{u}_{0,T}$ and $c>1$. The induction assumption is that $\max_{0\leq\ell\leq m}|\hat{u}_{\ell,T}|\leq 3U_0$. This implies that on the set $\mathcal{A}_T$ then $c\max_{0\leq\ell\leq m}|R_T(\hat{u}_{\ell,T})|\leq\eta/2$. Thus, the bound (A.5) becomes $|\hat{u}_{m+1,T}|\leq 2U_0+\eta/2\leq 3U_0$. It follows that $\max_{0\leq\ell\leq m+1}|\hat{u}_{\ell,T}|\leq 3U_0$. This proves (A.3).

Proof of (A.2)*:* In order to show (A.2) note that $\sum_{\ell=0}^{m}\rho^\ell=(I_{p+1}-\rho^{m+1})(I_{p+1}-\rho)^{-1}$ where $(I_{p+1}-\rho)^{-1}=\sum_{\ell=0}^{\infty}\rho^\ell$. Therefore equation (A.4) shows that the deviation $\hat{\Delta}_{m+1,T}=\hat{u}_{m+1,T}-(I_{p+1}-\rho)^{-1}K_T$ has the representation

$$\hat{\Delta}_{m+1,T}=\rho^{m+1}\{\hat{u}_{0,T}-(I_{p+1}-\rho)^{-1}K_T\}+\sum_{\ell=0}^{m}\rho^\ell R_T(\hat{u}_{m-\ell,T}).$$

To bound this, note first that $||(I_{p+1}-\rho)^{-1}||=||\sum_{\ell=0}^{\infty}\rho^\ell||\leq\sum_{\ell=0}^{\infty}||\rho^\ell||<c$. Thus on the set $\mathcal{A}_T$ it holds

$$|\hat{\Delta}_{m+1,T}|\leq||\rho^{m+1}||(c^{-1}U_0+U_0)+c\max_{0\leq\ell\leq m}|R_T(\hat{u}_{\ell,T})|\leq||\rho^{m+1}||2U_0+\eta/2.$$

Now, for $m\geq m_0$ then $||\rho^m||\leq\delta^m$. Since δ^m declines exponentially then m_0 can be chosen so large that it also holds that $||\rho^{m+1}||2U_0\leq\eta/2$. Thus $\mathsf{P}(|\hat{\Delta}_{m+1,T}|\geq\eta)<\epsilon$, for $m\geq m_0$ and $T\geq T_0$ which proves (A.2). ■

Proof of Theorem 3.4. The matrices ρ and $\rho-\lambda I_{p+1}$ are of the form

$$A=\begin{pmatrix} aI_p & b \\ c' & d \end{pmatrix}.$$

It suffices to show that $\det(A)=a^{p-1}(ad-c'd)$. If $b=0$ or $c=0$ then A is triangular and the result follows. Otherwise, define $\{p\times(p-1)\}$-matrices $b_\perp,c_\perp$ so $(b,b_\perp)$ and $(c,c_\perp)$ are regular and $b'b_\perp=c'c_\perp=0$. The skew projection identity $I_p=c_\perp(b_\perp'c_\perp)^{-1}b_\perp'+b(c'b)^{-1}c'$ implies

$$\det(A)=\det[\begin{pmatrix} b_\perp' & 0 \\ c' & 0 \\ 0 & 1 \end{pmatrix}A\left\{\begin{matrix} c_\perp(b_\perp'c_\perp)^{-1} & b(c'b)^{-1} & 0 \\ 0 & 0 & 1 \end{matrix}\right\}]=\det\begin{pmatrix} aI_{p-1} & 0 & 0 \\ 0 & a & c'b \\ 0 & 1 & d \end{pmatrix},$$

which is seen to have the correct determinant. ■

Proof of Theorem 3.5. (a) For $c>0$ then $\mathsf{f}(x)1_{(|x|\leq c)}\geq\mathsf{f}(c)1_{(|x|\leq c)}$ because f is symmetric and non-increasing. Integration gives

$$\psi=\int_{-c}^{c}\mathsf{f}(x)dx\geq 2c\mathsf{f}(c)=\xi_1,$$

where equality holds for $\mathsf{f}(x) = \mathsf{f}(c)$ for $|x| \leq c$, by continuity of f. This is, however, ruled out by assuming $\lim_{\psi\to 0} \mathsf{f}''(c) < 0$. It holds $\lim_{\psi\to 0} c^{-1} \int_0^c \mathsf{f}(x)dx = \mathsf{f}(0)$ and $\lim_{c\to 0} \xi_1/(2c) = \mathsf{f}(0)$ so $\lim_{\psi\to 0} \xi_1/\psi = 1$. Similarly, $\int_0^\infty \mathsf{f}(x)dx = 1/2$ and $\lim_{\psi\to 1} c\mathsf{f}(c) \to 0$ so $\lim_{\psi\to 1} \xi_1/\psi = 0$.

(b) Let $g(c) = \xi_3/(2\tau_2) - \xi_1/(2\tau_0)$. Since f is symmetric then $\tau_{2k} = 2\int_0^c x^{2k}\mathsf{f}(x)dx$ and $\xi_{2k+1} = c\tau'_{2k} = 2c^{2k+1}\mathsf{f}(c)$ so $2g(c) = c\tau'_2/\tau_2 - c\tau'_0/\tau_0$. It holds $(c\tau'_{2k})' = \tau'_{2k}\{2k + 1 + c(\log\mathsf{f})'\}$. Therefore l'Hôpital's rule gives

$$\lim_{\psi\to 0} \frac{c\tau'_{2k}}{\tau_{2k}} = \lim_{\psi\to 0} \frac{(c\tau'_{2k})'}{\tau'_{2k}} = 2k + 1.$$

As a conseqence $\lim_{\psi\to 0} g(c) = 1$. Assuming that f has third moments then $\lim_{\psi\to 1} \tau_{2k} < \infty$ while $\lim_{\psi\to 1} c\tau'_{2k} = 0$ for $k = 0, 1$. As a consequence $\lim_{\psi\to 1} g(c) = 0$.

(c) Rewrite $g(c)$ as N/D where $N = c\tau'_0(c^2\tau_0 - \tau_2)$ and $D = 2\tau_2\tau_0$. Then $g(c) < 1$ holds if and only if $N - D < 0$. It is convenient to write $N - D = \tau_0 M$ where

$$M = c\frac{\tau'_0}{\tau_0}(c^2\tau_0 - \tau_2) - 2\tau_2.$$

As $\tau_0 > 0$ for $c > 0$ it has to be shown that $M < 0$. Now $\lim_{\psi\to 0} M = 0$ since $\lim_{\psi\to 0} c\tau'_0/\tau_0 = 1$ and $\lim_{\psi\to 0} \tau_{2k} = 0$ so it suffices to show that $M' < 0$. But $M' = (c\tau'_0/\tau_0)'(c^2\tau_0 - \tau_2)$ for which it holds that $c^2\tau_0 - \tau_2 = \int_0^c (c^2 - x^2)\mathsf{f}(x)dx > 0$ and $(c\tau'_0/\tau_0)' = [c\{\log\int_0^c \mathsf{f}(x)dx\}']' < 0$ by assumption.

(d) First, assume $\{\log\mathsf{f}(c)\}'' < 0$ and $\mathsf{f}'(c) < 0$ for $c > 0$. Then

$$[c\{\log\mathsf{f}(c)\}']' = \{\log\mathsf{f}(c)\}' + c\{\log\mathsf{f}(c)\}'' = \frac{\mathsf{f}'(c)}{\mathsf{f}(c)} + c\{\log\mathsf{f}(c)\}'' < 0.$$

Secondly, assume $[c\{\log\mathsf{f}(c)\}']' < 0$. Denote $\mathsf{F}(c) = \int_0^c \mathsf{f}(x)dx$. Then

$$[c\{\log\mathsf{F}(c)\}']' = \frac{\{c\mathsf{f}(c)\}'\mathsf{F}(c) - c\{\mathsf{f}(c)\}^2}{\{\mathsf{F}(c)\}^2} = \frac{\mathsf{f}(c)}{\{\mathsf{F}(c)\}^2}L,$$

where $L = [1 + c\{\log\mathsf{f}(c)\}']\mathsf{F}(c) - c\mathsf{f}(c)$. Since $\mathsf{f}(c) \geq 0$ and $\mathsf{F}(c) > 0$ for $c > 0$ it has to be argued that $L < 0$. Now $\lim_{\psi\to 0} L = 0$ so it suffices to argue that $L' < 0$ for $c > 0$. But $L' = [c\{\log\mathsf{f}(c)\}']'\mathsf{F}(c)$ which is negative by assumption. ■

References

Atkinson, A.C., Riani, M. and Cerioli, A. (2004) *Exploring Multivariate Data with the Forward Search.* New York: Springer.

Atkinson, A.C., Riani, M. and Ceroli, A. (2010) The forward search: Theory and data analysis. Discussion paper *Journal of Korean Statistical Society* 39, 117–134.

Bickel, P. J. (1975) One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428–434.

Cavaliere, G. and Georgiev, I. (2011) Exploiting infinite variance through dummy variables in an autoregressive model. Mimeo.

Dollinger, M.B. and Staudte, R.G. (1991) Influence functions of iteratively reweighted least squares estimators. *Journal of the American Statistical Association* 86, 709–716.

Doornik, J.A. (2009) Autometrics. In Castle, J.L. and Shephard, N. (eds.) *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry*, pp. 88–121. Oxford: Oxford University Press.

He, X. and Portney, S. (1992) Reweighted LS estimators converge at the same rate as the initial estimator. *Annals of Statistics* 20, 2161–2167.

Hendry, D.F. and Krolzig, H.-M. (2005) The properties of automatic Gets modelling. *Economic Journal* 115, C32–C61.

Huber, P.J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.

Johansen, S. and Nielsen, B. (2009) An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J.L. and Shephard, N. (eds.) *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry*, pp. 1–36. Oxford: Oxford University Press.

Johansen, S. and Nielsen, B. (2010) Discussion: The forward search: Theory and data analysis. *Journal of the Korean Statistical Society* 39, 137–145.

Jurečová, J. and Sen, P.K. (1996) *Robust Statistical Procedures.* New York: Wiley.

Rousseeuw, P.J. (1982) Most robust M-estimators in the infinitesimal sense. *Zeitschrift für Warhscheinlichkeitstheorie und verwandte Gebiete* 61, 541–551.

Rousseeuw, P.J. (1984) Least median of squares regression. *Journal of the American Statistical Association* 79, 871–880.

Ruppert, D. and Carroll, R.J. (1980) Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828–838.

Varga, R.S. (2000) *Matrix Iterative Analysis*, 2nd edition. Berlin: Springer.

Víšek, J.Á. (2006) The least trimmed squares. Part III: Asymptotic normality. *Kybernetika* 42, 203–224.

Welsh, A.H. and Ronchetti, E. (2002) A journey in single steps: robust one step M-estimation in linear regression. *Journal of Statistical Planning and Inference* 103, 287–310.