Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator

Søren Johansen[1,2] & Bent Nielsen[3]

26 January 2013

Summary: In regression we can delete outliers based upon a preliminary estimator and reestimate the parameters by least squares based upon the retained observations. We study the properties of an iteratively defined sequence of estimators based on this idea. We relate the sequence to the Huber-skip estimator.

We provide a stochastic recursion equation for the estimation error in terms of a kernel, the previous estimation error and a uniformly small error term.

The main contribution is the analysis of the solution of the stochastic recursion equation as a fixed point, and the results that the normalized estimation errors are tight and are close to a linear function of the kernel, thus providing a stochastic expansion of the estimators, which is the same as for the Huber-skip. This implies that the iterated estimator is a close approximation of the Huber-skip.

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

1 Introduction and main results

Outlier detection in regression is an important topic in econometrics. The idea is to find an estimation method that is robust to the presence of outliers, and the statistical literature abounds in robust methods, since the introduction of M-estimators by Huber (1964), see also the monographs Maronna, Martin, and Yohai (2006), Huber and Ronchetti (2009), and Jurečková, Sen, and Picek (2012). Recent contributions are the impulse indicator saturation method, see Hendry, Johansen and Santos (2008) and Johansen & Nielsen (2009), and the Forward Search, see Atkinson, Riani, and Cerioli (2004).

The present paper is a contribution to the theory of the robust estimators, where we focus on the Huber (1964) skip-estimator that minimizes

$$\sum_{i=1}^{n} \rho(y_i - \beta' X_i)$$

where the objective function, ρ, is given by

$$\rho(z) = \frac{1}{2}\min(z^2, c^2) = \frac{1}{2}(z^2 1_{(|z|\leq c)} + c^2 1_{(|z|>c)}).$$

[1]The first author is grateful to CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

[2]Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University. E-mail: soren.johansen@econ.ku.dk.

[3]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.

This estimator removes the observations with large residuals, something that, at least in the analysis of economic time series, appears a reasonable method.

It is seen that ρ is absolutely continuous with derivative $\rho'(z) = z1_{(|z|\leq c)}$, but $\rho'(z)$ is neither monotone nor absolutely continuous, which makes the calculation of the minimizer somewhat tricky, and the asymptotic analysis rather difficult.

Thus the estimator is often replaced by the Winsorized estimator, which has convex objective function

$$\rho_1(z) = \frac{1}{2}z^2 1_{(|z|\leq c)} + c(|z| - \frac{1}{2}c)1_{(|z|>c)}$$

with derivative

$$\rho_1'(z) = z1_{(|z|\leq c)} + c\text{sign}(z)1_{(|z|>c)},$$

which is both monotone and absolutely continuous and hence a lot easier to analyse, see Huber (1964). Note, however, that the function ρ_1 replaces the large residuals by $\pm c$, instead of removing the observation. This is a less common method in time series econometrics.

An alternative simplification is formulated by Bickel (1975), who suggested applying a preliminary estimator $\hat{\beta}_{n0}$ and define the one-step estimator, $\hat{\beta}_{n1}$, by linearizing the first order condition. He also suggested iterating this by using $\hat{\beta}_{n1}$ as initial estimator for $\hat{\beta}_{n2}$ etc. but no results are given.

In the analysis of the Huber-skip, derived from ρ, we shall replace β by a preliminary estimator in the indicator function, which leads to eliminating the outlying observations and run a regression on retained observations. We shall do so iteratively and study the sequence of recursively defined estimators $\hat{\beta}_{nm}$. We prove under fairly general assumptions on regressors and distribution, that for $(m, n) \to \infty$, the estimator $\hat{\beta}_{nm}$ has the same asymptotic expansion as the Huber-skip, and in this sense $\hat{\beta}_{nm}$, which is easy to calculate, is a very good approximation to the Huber-skip.

One-step M-estimators have been analysed previously in various situations: Apart from Bickel (1975), who considered a situation with fixed regressors and weight functions satisfying certain smoothness and integrability conditions, Ruppert and Carroll (1980) considered one-step Huber-skip L-estimators. Welsh and Ronchetti (2002) analysed 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 with a function ρ with absolutely continuous derivative. Recently Cavaliere and Georgiev (2011) analysed a sequence of Huber-skip estimators for the parameter of an $AR(1)$ model with infinite variance errors in case the autoregressive coefficient is 1. Johansen and Nielsen (2009) analyse one-step Huber-skip estimators for general $n^{1/2}$ consistent initial estimators and stationary as well as some 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 because of the non-smooth weight function.

Notation: The Euclidean norm for vectors x is denoted $|x|$. We write $(m, n) \to \infty$ if both m and n tend to infinity. We use the notation $o_P(1)$ and $O_P(1)$ implicitly assuming that $n \to \infty$. For matrices M we choose the spectral norm $||M|| = \max\{\text{eigen}(M'M)\}^{1/2}$, so that $||x|| = |x|$ for vectors x. We use that the spectral norm and the Euclidean norm are compatible,

$|Mx| \leq ||M||\,|x|$, 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).

2 The model and the definition of the one-step Huber-skip

We consider the regression model

$$y_i = \beta' X_i + \varepsilon_i,\ i = 1, \ldots, n \tag{2.1}$$

where X_i is p-dimensional, and ε_i is assumed independent of $(X_1, \ldots, X_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$ are i.i.d. with known density f, which need not be symmetric. These assumptions allow for both deterministic and stochastic regressors. In particular X_i can be the lagged dependent variables as for an autoregrsssive process, and the process can be stationary or nonstationary.

We consider estimation of both β and σ^2. Thus we start with some preliminary estimator $(\hat{\beta}_{n0}, \hat{\sigma}^2_{n0})$ 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 technical assumptions are listed in Assumption A see §2.2 below, and allows the regressors to be deterministic or stochastic and stationary or trending.

The preliminary estimator $(\hat{\beta}_{n0}, \hat{\sigma}^2_{n0})$ could be a least squares estimator on the full sample, although that is not a good idea from a robustness viewpoint, see Welsh and Ronchetti (2002). Alternatively, the initial estimator could be chosen as a robust estimator, as for instance 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 $n^{1/2}$-rate, see Víšek (2006*a,b,c*).

The outliers are identified by first choosing a ψ giving the proportion of good, central observations and then, because f is not assumed symmetric, introducing two critical values $\underline{c}$ and $\bar{c}$ so

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

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

$$\tau_k = \int_{\underline{c}}^{\bar{c}} v^k \mathsf{f}(v)dv \qquad \text{for } k \in \mathbb{N}_0. \tag{2.3}$$

If f is symmetric we find $c = -\underline{c} = \bar{c}$ and $\tau_{2k+1} = 0, k \in \mathbb{N}_0$. Observations are retained based on $(\hat{\beta}_{n0}, \hat{\sigma}^2_{n0})$ if their residuals $y_i - \hat{\beta}'_{n0} X_i$ are in the interval $[\underline{c}\hat{\sigma}_{n0}, \bar{c}\hat{\sigma}_{n0}]$ and otherwise deleted from the sample.

The Huber-skip, $\hat{\beta}_{nH}$, is defined by minimizing

$$\frac{1}{2}\sum_{i=1}^{n}[(y_i - X_i'\beta)^2 1_{(\underline{c}\sigma \leq y_i - X_i'\beta \leq \bar{c}\sigma)} + \underline{c}^2 1_{(y_i - X_i'\beta \leq \underline{c}\sigma)} + \bar{c}^2 1_{(\bar{c}\sigma \leq y_i - X_i'\beta)}],$$

for a given σ. If the minimum is attained at a point of differentiability of the objective function, then the solution solves the equation

$$\hat{\beta}_{nH} = (\sum_{i=1}^{n} X_i X_i' 1_{(\underline{c}\sigma \leq y_i - X_i'\hat{\beta}_{nH} \leq \bar{c}\sigma)})^{-1} \sum_{i=1}^{n} X_i y_i 1_{(\underline{c}\sigma \leq y_i - X_i'\hat{\beta}_{nH} \leq \bar{c}\sigma)}.$$

We apply this to propose a sequence of recursively defined estimators $(\hat{\beta}_{nm}, \hat{\sigma}^2_{nm})$ by starting with $(\hat{\beta}_{n0}, \hat{\sigma}^2_{n0})$ and defining for $m, n = 1, 2, \ldots$

$$\begin{aligned} \mathcal{S}_{n,m-1} &= \{i : \underline{c}\hat{\sigma}_{n,m-1} \leq y_i - X_i'\hat{\beta}_{n,m-1} \leq \overline{c}\hat{\sigma}_{n,m-1}\}, && (2.4) \\ \hat{\beta}_{nm} &= (\sum_{i \in \mathcal{S}_{n,m-1}} X_i X_i')^{-1} \sum_{i \in \mathcal{S}_{n,m-1}} X_i y_i, && (2.5) \\ \hat{\sigma}^2_{nm} &= \psi\tau_2^{-1}(\sum_{i \in \mathcal{S}_{n,m-1}} 1)^{-1} \sum_{i \in \mathcal{S}_{n,m-1}} (y_i - X_i'\hat{\beta}_{n,m-1})^2. && (2.6) \end{aligned}$$

Thus, the iterated one-step Huber-skip estimators $\hat{\beta}_{nm}$ and $\hat{\sigma}_{nm}$, are the least squares estimator of y_i on X_i among the retained observations in $\mathcal{S}_{n,m-1}$ based upon $\hat{\beta}_{n,m-1}$ and $\hat{\sigma}^2_{n,m-1}$. The bias correction factor $\psi\tau_2^{-1}$ in $\hat{\sigma}^2_{nm}$ is needed to obtain consistency.

2.1 Asymptotic results

To obtain asymptotic results we need a normalisation matrix N for the regressors. If X_i is stationary then $N = n^{-1/2}I_p$. If X_i is trending a different normalisation is needed. For a linear trend component the normalisation is $n^{3/2}$ and for a random walk component it is n. We assume that N has been chosen such that matrices Σ and μ exist for which

$$N' \sum_{i=1}^{n} X_i X_i' N \xrightarrow{\mathsf{D}} \Sigma \overset{a.s.}{>} 0, \qquad n^{-1/2} N' \sum_{i=1}^{n} X_i \xrightarrow{\mathsf{D}} \mu.$$

Note that Σ and μ may be stochastic as for instance when X_i is a random walk and $N = n^{-1}$.

The estimation errors are denoted

$$\hat{u}_{nm} = \left\{ \begin{array}{c} N^{-1}(\hat{\beta}_{nm} - \beta) \\ n^{1/2}(\hat{\sigma}_{nm} - \sigma) \end{array} \right\}, \tag{2.7}$$

and the recursion defined in (2.4), (2.5), and (2.6) can expressed as

$$\hat{u}_{nm} = G_n(\hat{u}_{n,m-1}). \tag{2.8}$$

If the Huber-skip is a point of differentiability of the objective function, it is a fixed point of G_n.

We introduce coefficient matrices

$$\Psi_1 = \left(\begin{array}{cc} \psi\Sigma & 0 \\ 0 & 2\tau_2 \end{array} \right), \qquad \Psi_2 = \left(\begin{array}{cc} \xi_1\Sigma & \xi_2\mu \\ \zeta_2\mu' & \zeta_3 \end{array} \right), \tag{2.9}$$

where

$$\xi_n = (\overline{c})^n \mathsf{f}(\overline{c}) - (\underline{c})^n \mathsf{f}(\underline{c}), n = 0, \ldots, 3 \text{ and } \zeta_n = \xi_n - \xi_{n-2}\tau_2/\psi, \ n = 2, 3, \tag{2.10}$$

and τ_2 is defined in (2.3), and define

$$\Gamma = \Psi_1^{-1}\Psi_2 = \left(\begin{array}{cc} \psi^{-1}\xi_1 I_p & \psi^{-1}\xi_2\Sigma^{-1}\mu \\ (2\tau_2)^{-1}\zeta_2\mu' & (2\tau_2)^{-1}\zeta_3 \end{array} \right). \tag{2.11}$$

When f is symmetric we let $c = -\underline{c} = \bar{c}$ and find $\xi_2 = 0$, so that Γ is diagonal. Moreover from $\xi_{2k+1} = 2c^{2k+1}\mathsf{f}(c)$, we find $\zeta_1 = \xi_1/\psi = 2c\mathsf{f}(c)/\psi$ and $\zeta_3/(2\tau_2) = c^3\mathsf{f}(c)/\tau_2 - c\mathsf{f}(c)/\psi$ and therefore $\Gamma = \text{diag}\{2c\mathsf{f}(c)/\psi I_p, c\mathsf{f}(c)(c^2/\tau_2 - 1/\psi)\}$.

Finally we define a kernel

$$K_n = \Psi_1^{-1} \sum_{i=1}^{n} \left\{ \begin{array}{c} N' X_i \varepsilon_i \\ n^{-1/2}(\varepsilon_i^2 - \sigma^2\tau_2/\psi) \end{array} \right\} 1_{(\underline{c}\sigma \le \varepsilon_i \le \sigma\bar{c})}. \tag{2.12}$$

The analysis of the one-step estimator in Johansen and Nielsen (2009) shows that, by linearizing G_n, the one-step estimation errors $\hat{u}_{n,m}$ satisfy the recursion equation

$$\hat{u}_{n,m} = G_n(\hat{u}_{n,m-1}) = \Gamma\hat{u}_{n,m-1} + K_n + R_n(\hat{u}_{n,m-1}), \tag{2.13}$$

for some remainder term $R_n(\hat{u}_{n,m-1})$. In this notation it is emphasized that the remainder term is a function of the previous estimator $\hat{u}_{n,m-1}$, see Lemma A.1 in the Appendix for a precise formulation.

It will be shown in Section 3 that if $\max|\text{eigen}(\Gamma)| < 1$, so that Γ is a contraction, then

$$\hat{u}_{n,m} - (I_{1+p} - \Gamma)^{-1}K_n \xrightarrow{\mathsf{P}} 0 \text{ for } (m,n) \to \infty,$$

that is, for any η and $\epsilon > 0$ there exists m_0 and n_0 such that for $m \ge m_0$ and $n \ge n_0$ it holds that

$$\mathsf{P}(|\hat{u}_{n,m} - (I_{1+p} - \Gamma)^{-1}K_n| \ge \eta) \le \epsilon.$$

We therefore define $\hat{u}_{n*} = (I_{1+p} - \Gamma)^{-1}K_n$ and note that it satisfies the equation

$$\hat{u}_{n*} = \Gamma\hat{u}_{n*} + K_n, \tag{2.14}$$

and in this sense the estimation error of (β, σ) has the same limit distribution as the fixed point of the linear function $\Gamma u + K_n$.

Moreover it follows from Johansen and Nielsen (2013*b*) that the Huber skip has the stochastic expansion

$$\hat{\beta}_{nH} = (I_{1+p} - \Gamma)^{-1}K_n + o_{\mathsf{P}}(1)$$

and hence the same asymptotic distribution as $\hat{u}_{n*}$ and moreover it holds that

$$n^{1/2}(\hat{\beta}_{nH} - \hat{\beta}_{nm}) \xrightarrow{\mathsf{P}} 0 \text{ for } (n,m) \to \infty.$$

Finally the asymptotic distribution of K_n, and therefore $\hat{u}_{n*}$, is discussed in Section 4.

2.2 Assumptions for the asymptotic analysis

The assumptions are fairly general, in particular we do not assume that f is symmetric.

Assumption A *Consider model (2.1). Assume*
(*i*) *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*) $\bar{c}, \underline{c}$ *are chosen so* $\tau_0 = \psi$ *and* $\tau_1 = 0$
(*ii*) *For a suitable normalization matirx* $N \to 0$, *the regressors satisfy, jointly,*

(*a*) $N' \sum_{i=1}^{n} X_i X_i' N \xrightarrow{\mathsf{D}} \Sigma \overset{a.s.}{>} 0$,
(*b*) $n^{-1/2} N' \sum_{i=1}^{n} X_i \xrightarrow{\mathsf{D}} \mu$,
(*c*) $\max_{i \leq n} \mathsf{E}|n^{1/2} N' X_i|^4 = \mathrm{O}(1)$.
(*iii*) *The initial estimator error satisfies*

$$(N^{-1}(\hat{\beta}_{n0} - \beta), n^{1/2}(\hat{\sigma}_{n0} - \sigma)) = \mathrm{O}_{\mathsf{P}}(1).$$

3 The fixed point result

The fixed point result is primarily a tightness result. Thus, for the moment, only tightness of the kernel K_n is needed, and it is not necessary to establish the limit distribution, which is discussed in Section 4. The first result is a tightness result for the kernel, see (2.12).

Theorem 3.1 *Suppose Assumption A*(ib, iic) *holds. Then* K_n*, see (2.9) and (2.12), is tight, that is,*

$$K_n = \Psi_1^{-1} \sum_{i=1}^{n} \left\{ \begin{array}{c} N' X_i \varepsilon_i \\ n^{-1/2}(\varepsilon_i^2 - \sigma^2 \tau_2/\psi) \end{array} \right\} 1_{(\underline{c}\sigma \leq \varepsilon_i \leq \sigma \bar{c})} = \mathrm{O}_{\mathsf{P}}(1).$$

The proof follows from Chebychev's inequality and the details are given in the appendix.

The next result discusses one step of the iteration (2.13), and it is shown that the remainder term $R_n(u)$ in (2.13) vanishes in probability uniformly in $|u| < U$.

Theorem 3.2 *Let* m *be fixed. Suppose Assumption A holds for the initial estimator* $\hat{u}_{n,m-1}$*, see (2.7). Then, for all* $U > 0$*, it holds that*

$$\hat{u}_{n,m} = \Gamma \hat{u}_{n,m-1} + K_n + R_n(\hat{u}_{n,m-1}),$$

where the remainder term satisfies

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

The proof involves a chaining argument which was given in Johansen and Nielsen (2009), although there, the result was written up in a slightly different, way as discussed in the appendix.

The iterated estimators start with an initial estimator $(\hat{\beta}_{n0}, \hat{\sigma}_{n0})$ with tight estimation error, see Assumption A(*iii*). This is iterated through the one-step equation (2.13) and defines the sequence of estimation errors $\hat{u}_{n,m}$. We next show that this sequence is tight uniformly in m.

Theorem 3.3 *Suppose Assumption A holds. Then the sequence of estimation errors* $\hat{u}_{n,m}$ *is tight uniformly in* m

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

That is, for all $\epsilon > 0$ *there exists* $U > 0$ *and* $n_0 > 0$*, so that for all* $n \geq n_0$ *it holds that*

$$\mathsf{P}(\sup_{0 \leq m < \infty} |\hat{u}_{n,m}| > U) < \epsilon.$$

The proof is given in the appendix, but the idea of the proof is to write the solution of the recursive relation (2.13) as

$$\hat{u}_{n,m} = \Gamma^m \hat{u}_{n0} + \sum_{\ell=1}^{m} \Gamma^{\ell-1}\{K_n + R_n(\hat{u}_{n,m})\}. \tag{3.1}$$

Then, if the initial estimator $\hat{u}_{n0}$ takes values in a large compact set with large probability, it follows from (3.1), by finite induction, that also $\hat{u}_{n,m}$ takes values in the same compact set for all m, and therefore $\hat{u}_{n,m}$ is tight uniformly in m.

Finally we give the fixed point result. Theorem 3.4 shows that the estimator has the same limit distribution as the solution of equation (2.14), $\hat{u}_{n*} = (I_{p+1} - \Gamma)^{-1}K_n$, which is a fixed point of the linear function $\Gamma u + K_n$.

Theorem 3.4 *Suppose Assumption A holds and that* $\max|\text{eigen}(\Gamma)| < 1$, *so that* Γ *is a contraction. Then*

$$\hat{u}_{n,m} - \hat{u}_{n*} = \hat{u}_{n,m} - (I_{p+1} - \Gamma)^{-1}K_n \xrightarrow{\mathsf{P}} 0 \text{ for } (m,n) \to \infty.$$

That is, for all ϵ *and* $\eta > 0$, *an* $n_0 > 0$ *and* $m_0 > 0$ *exist so that for all* $n \geq n_0$ *and* $m \geq m_0$ *it holds*

$$\mathsf{P}(|\hat{u}_{n,m} - (I_{p+1} - \Gamma)^{-1}K_n| > \eta) < \epsilon.$$

Using $\sum_{\ell=1}^{m} \Gamma^{\ell-1} = (I_{p+1} - \Gamma)^{-1}(I_{p+1} - \Gamma^m)$ we find from (3.1) that

$$\hat{u}_{n,m} - (I_{p+1} - \Gamma)^{-1}K_n = \Gamma^m(\hat{u}_{n0} - (I_{p+1} - \Gamma)^{-1}K_n) + \sum_{\ell=1}^{m} \Gamma^{\ell-1}R_n(\hat{u}_{n,m-\ell}). \tag{3.2}$$

From (3.2) it can be seen that $|\hat{u}_{nm} - (I_{p+1} - \Gamma)^{-1}K_n|$ is the sum of two terms vanishing exponentially and in probability, respectively. The details are given in the Appendix.

In the special case where σ is known then $\hat{u}_{nm}$ reduces to $\hat{b}_{nm} = N^{-1}(\hat{\beta}_{nm} - \beta)$ and $\Gamma = \psi^{-1}\xi_1 I_p$. The estimator $\hat{b}_{n*} = (\psi - \xi_1)^{-1}\Sigma^{-1}\sum_{i=1}^{n} N' X_i \varepsilon_i 1_{(\underline{c}\sigma < \varepsilon_i \leq \bar{c}\sigma)}$ appears as the leading term for other robust estimators, such as the Least Trimmed Squares estimator discussed later on.

A necessary condition for the result is that the autoregressive coefficient matrix Γ is contracting. Therefore Γ is analyzed next.

Theorem 3.5 *The autoregressive coefficient matrix* Γ, *(2.11), has* $p-1$ *eigenvalues equal to* ξ_1/ψ *and two eigenvalues solving*

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

where the coefficients ζ_n *and* ξ_n *are given in (2.10).*

Further results can be given about the eigenvalues of Γ for symmetric densities, where $\xi_2 = 0$, and $\Gamma = \text{diag}(\xi_1\psi^{-1}I_p, \zeta_3/(2\tau_2))$. Note that the quantities $(c, \tau, \xi_n, \zeta_n)$ all depend on ψ, see (2.2), (2.3), and (2.10). If f is symmetric, we show below, (a), that $\xi_1 < \psi$ and a condition, (c), is given for $\zeta_3 < 2\tau_2$, in which case the eigenvalues of Γ are less than one, and Γ is a contraction. Finally (d) shows that Γ is a contraction if f is log concave.

Theorem 3.6 *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*) $0 < \zeta_3/(2\tau_2)$ *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, Rousseeuw (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 3.6 is that if f is symmetric, the roots of the coefficient matrix Γ are bounded away from unity for $\psi_0 \leq \psi \leq 1$ for all $\psi_0 > 0$. The uniform distribution on $[-a, a]$ provides an example where Γ 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 3.6 is not necessary as long as the mode at the origin is large in comparison to other modes.

4 Distribution of the kernel

It follows from Theorem 3.4 that $\hat{u}_{n*} = (I_{p+1} - \Gamma)^{-1}K_n$ has the same limit as $\hat{u}_{nm}$, and we therefore find the limit distribution of the kernel K_n in a few situations.

4.1 Stationary case

Suppose the regressors are a stationary time series. Then the limits Σ and μ in Assumption A(ia, ib) are deterministic. The central limit theorem then shows that

$$K_n \xrightarrow{\mathsf{D}} \mathsf{N}_{p+1}(0, \Phi), \tag{4.1}$$

where

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

As a consequence, the fully iterated estimator has limit distribution

$$\hat{u}_* = (I_{p+1} - \Gamma)^{-1}K_n \xrightarrow{\mathsf{D}} (I_{p+1} - \Gamma)^{-1}\mathsf{N}_{p+1}(0, \Phi). \tag{4.3}$$

In the special case where the errors are symmetric, we find

$$\begin{aligned} N^{-1}(\hat{\beta}_{n*} - \beta) &= \frac{1}{(\psi - \xi_1)}\Sigma^{-1}\sum_{i=1}^{n} N'X_i\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)} + o_{\mathsf{P}}(1) \xrightarrow{\mathsf{D}} \mathsf{N}_p\{0, \frac{\sigma^2\tau_2}{(\psi - \xi_1)^2}\Sigma^{-1}\}, \\ n^{1/2}(\hat{\sigma}^2_{n*} - \sigma^2\tau_\psi/\psi) &= \{1 - \zeta_3(2\tau_2)^{-1}\}^{-1}\sum_{i=1}^{n} n^{-1/2}(\varepsilon_i^2 - \sigma^2\tau_2\psi^{-1})1_{(|\varepsilon_i|\leq\sigma c)} + o_{\mathsf{P}}(1) \\ &\xrightarrow{\mathsf{D}} \mathsf{N}_p\{0, \frac{\sigma^4\tau_2^2(\tau_4 - \psi^{-1}\tau_2^2)}{(2\tau_2 - \zeta_3)^2}\} \end{aligned}$$

noting that $\psi > \xi_1$ and $\zeta_3 > 2\tau_2$ are satisfied for symmetric, unimodal distributions by Theorem 3.6(a, b).

The limiting distribution is also seen elsewhere in the robust statistics literature.

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

$$\hat{\beta}^{LTS} = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{\mathrm{int}(n\psi)} r^2_{(i)}(\beta)$$

where $r^2_{(1)}(\beta) < \cdots < r^2_{(n)}(\beta)$ are the ordered squared residuals $r_i = y_i - X_i'\beta$. The estimator has the property that it does not depend on the scale of the problem. Víšek argued that in the symmetric case, the least trimmed squares estimator satisfies

$$N^{-1}(\hat{\beta}^{LTS}_n - \beta) = \frac{1}{(\psi - \xi_1)}\Sigma^{-1}\sum_{i=1}^{n} N'X_i\varepsilon_i 1_{(|\varepsilon_i|\le c\sigma)} + \mathrm{o}_{\mathsf{P}}(1), \tag{4.4}$$

that is, the main term is the same as for $\hat{\beta}_{n*}$, and it follows from Theorem 3.4 that because $\hat{\beta}^{LTS}_n$ and $\hat{\beta}_{n*}$ have the same expansions we have

$$|N^{-1}(\hat{\beta}_{nm} - \hat{\beta}^{LTS}_n)| \xrightarrow{\mathsf{P}} 0$$

for $(m,n) \to \infty$. Thus $\hat{\beta}_{nm}$ can be seen as an approximation to the LTS estimator when there are no outliers.

Second, Jurečková, Sen, and Picek (2012, Theorem 5.5, page 176) considered a pure location problem with regressor $X_i = 1$ and known $\sigma = 1$, and found an asymptotic expansion like (4.4) for the Huber-skip, and Johansen and Nielsen (2013*b*) show the similar result for the general regression model. A consequence of this is that the iterated 1-step Huber-skip has the same limit distribution as the Huber-skip, and because $\hat{\beta}_{nm}$ and $\hat{\beta}_{nH}$ have the same expansion it follows from Theorem 3.4, that

$$n^{1/2}|\hat{\beta}_{nm} - \hat{\beta}_{nH}| \xrightarrow{\mathsf{P}} 0 \text{ for } (m,n) \to \infty, \tag{4.5}$$

so the iterated estimator is in this sense an approximation to the Huber-skip.

4.2 Deterministic trends

As a simple example with i.i.d. errors, consider the regression

$$y_i = \beta_1 + \beta_2 i + \varepsilon_i,$$

where $\varepsilon_i \in \mathbb{R}$ satisfies Assumption A(i). Define the normalisation

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

Then Assumption A(ii) is met with $X_i = (1, i)'$ and

$$\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}, \qquad \mu = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}, \tag{4.6}$$

and $\max_{i\le n} \mathsf{E}|n^{1/2}N'X_i|^4 \le 4$. The kernel has a limit distribution given by (4.1) where the matrix Φ in (4.2) is computed in terms of the Σ and μ derived in (4.6).

If the errors are autoregressive, the derivation is in principle similar, but involves a notationally tedious detrending argument. The argument is similar to that of Johansen and Nielsen (2009, Section 1.5.1), and (4.5) holds.

4.3 Unit roots

Consider as an example the autoregression $y_i = \beta y_{i-1} + \varepsilon_i, i = 1, \ldots, n$. If $\beta = 1$ then $X_i = y_{i-1} = y_0 + \sum_{s=1}^{i-1} \varepsilon_s$ and we have to choose $N = n^{-1}$. By the functional Central Limit Theorem

$$n^{-1/2} \sum_{i=1}^{\text{int}(nu)} \left\{ \begin{array}{l} \varepsilon_i \\ \varepsilon_i 1_{(\underline{c}\sigma \leq \varepsilon_i \leq \sigma \bar{c})} \\ (\varepsilon_i^2 - \sigma^2 \tau_2/\psi) 1_{(\underline{c}\sigma \leq \varepsilon_i \leq \sigma \bar{c})} \end{array} \right\} \xrightarrow{\mathsf{D}} \left(\begin{array}{c} W_{x,u} \\ W_{1,u} \\ W_{2,u} \end{array} \right),$$

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 limit variables Σ and μ in Assumption A(i) are

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

while the kernel has limit distribution

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

and (4.5) holds. Thus, when the density of ε_i is symmetric, $\hat{\beta}_*$ has limit distribution

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

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 limit 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 of possible extensions

The iteration result in Theorem 3.4 has a variety of extensions. 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 $n^{1/3}$ 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 by this procedure which applies least squares to the retained observation.

The key is to show that the remainder term of the one-step estimator in Theorem 3.2 remains small in an appropriately larger neighbourhood. The proof of Theorem 3.4 then applies the same way leading to the same fixed point result. The necessary techniques are developed in Johansen and Nielsen (2013*a*)

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 is to improve Theorem 3.2, in this instance to hold uniformly in the cut-off fraction ψ, see Johansen and Nielsen (2013a) for details.

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. Because K_n is a martingale, see (2.12), we find

$$\mathsf{E}K_nK_n' = \Psi_1^{-1}\begin{pmatrix} \sigma^2\tau_2\sum_{i=1}^n \mathsf{E}(N'X_iX_i'N) & \sigma^3\tau_3\sum_{i=1}^n \mathsf{E}(N'X_i) \\ \sigma^3\tau_3\sum_{i=1}^n \mathsf{E}(N'X_i)' & \sigma^4(\tau_4-\tau_2^2\psi^{-1}) \end{pmatrix}\Psi_1'^{-1}.$$

Due to assumptions $(iic), (iiib)$ this is bounded in n. Chebychev's inequality gives $\mathsf{P}(|K_n| > C) \leq C^{-2}\mathsf{E}|K_n|^2$. Thus, for all $\epsilon > 0$, C can be chosen so large that $\mathsf{P}(|K_n| > C) < \epsilon$. ■

The key to 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 $n \to \infty$. 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 $u = (b, a)$, respectively, and define, for $g_i, h_i \in (1, X_i, \varepsilon_i)$, the product moment matrices

$$\tilde{S}_{gh}(u) = \sum_{i=1}^n g_i h_i' 1_{\{(\sigma+n^{-1/2}a)\underline{c}<\varepsilon_i - X_i'Nb \leq (\sigma+n^{-1/2}a)\bar{c}\}}.$$

Lemma A.1 *Suppose Assumption A 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*

$$\begin{aligned} n^{-1}\tilde{S}_{11}(u) &= \psi + R_{11}(u), \\ N'\tilde{S}_{xx}(u)N &= \psi\Sigma + R_{xx}(u), \\ n^{-1/2}N'\tilde{S}_{x1}(u) &= \psi\mu + R_{x1}(u), \\ \begin{bmatrix} N'\tilde{S}_{x\varepsilon}(u) \\ n^{-1/2}\{\tilde{S}_{\varepsilon\varepsilon}(u) - \sigma^2\tau_2\psi^{-1}\tilde{S}_{11}(u)\} \end{bmatrix} &= \sum_{i=1}^n \begin{Bmatrix} N'X_i\varepsilon_i \\ n^{-1/2}(\varepsilon_i^2 - \sigma^2\tau_2\psi^{-1}) \end{Bmatrix} 1_{(\underline{c}\sigma<\varepsilon_i\leq\bar{c}\sigma)} \\ &\quad + \begin{pmatrix} \xi_1\Sigma & \xi_2\mu \\ \sigma\zeta_2\mu' & \sigma\zeta_3 \end{pmatrix}\begin{pmatrix} b \\ a \end{pmatrix} + \begin{Bmatrix} R_{x\varepsilon}(u) \\ R_{\varepsilon\varepsilon}(u) \end{Bmatrix}, \end{aligned}$$

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

$$\sup_{|u|<U}\{|R_{11}(u)| + |R_{xx}(u)| + |R_{x1}(u)| + |R_{x\varepsilon}(u)| + |R_{\varepsilon\varepsilon}(u)|\} = \mathrm{o}_{\mathsf{P}}(1). \tag{A.1}$$

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 $\hat{u} = \{N^{-1}(\hat{\beta}-\beta), n^{1/2}(\hat{\sigma}-\sigma)\}$ under the assumption that $\hat{u} = \mathrm{O}_{\mathsf{P}}(1)$, as $n \to \infty$. 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 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 $(\hat{\beta}_{nm}, \hat{\sigma}^2_{nm})$ is defined in (2.5) and (2.6), in terms of the initial estimator $(\hat{\beta}_{n,m-1}, \hat{\sigma}^2_{n,m-1})$ and we express them in terms of $S_{gh} = \tilde{S}_{gh}(\hat{u}_{n,m-1})$ where $\hat{u}_{n,m-1} = \{N^{-1}(\hat{\beta}_{n,m-1} - \beta), n^{1/2}(\hat{\sigma}_{n,m-1} - \sigma)\}$, as follows

$$\begin{aligned} N^{-1}(\hat{\beta}_{nm} - \beta) &= (N'S_{xx}N)^{-1}N'S_{x\varepsilon}, \\ n^{1/2}(\hat{\sigma}^2_{nm} - \sigma^2) &= \psi\tau_2^{-1}(S_{11})^{-1}n^{1/2}\{S_{\varepsilon\varepsilon} - S_{\varepsilon x}N(N'S_{xx}N)^{-1}N'S_{x\varepsilon} - \sigma^2\tau_2\psi^{-1}S_{11}\}. \end{aligned}$$

For $\hat{u}_{n,m-1} = (\hat{b}_{n,m-1}, \hat{a}_{n,m-1})$ we get, by inserting the definitions from Lemma A.1,

$$\hat{b}_{nm} = \{\psi\Sigma + R_{xx}(\hat{u}_{n,m-1})\}^{-1}\{\sum_{i=1}^{n}(N'X_i\varepsilon_i)1_{(\underline{c}\sigma<\varepsilon_i\leq\bar{c}\sigma)} + \xi_1\Sigma\hat{b}_{n,m-1} + \xi_2\mu\hat{a}_{n,m-1} + R_{x\varepsilon}(\hat{u}_{n,m-1})\}.$$

Since $\sum_{i=1}^{n}(N'X_i\varepsilon_i)1_{(\underline{c}\sigma<\varepsilon_i\leq\bar{c}\sigma)}$ is tight by Theorem 3.1, $\hat{u}_{n,m-1}$ is $\mathrm{O}_{\mathsf{P}}(1)$, and the remainders are vanishing by Lemma A.1 for $n \to \infty$, then

$$\hat{b}_{nm} = (\psi\Sigma)^{-1}\sum_{i=1}^{n}(N'X_i\varepsilon_i)1_{(\underline{c}\sigma<\varepsilon_i\leq\bar{c}\sigma)} + (\psi\Sigma)^{-1}(\xi_1\Sigma\hat{b}_{n,m-1} + \xi_2\mu\hat{a}_{n,m-1}) + R_{b,n}(\hat{u}_{n,m-1}),$$

where $\sup_{|u|<U}|R_{b,n}(u)| = \mathrm{o}_{\mathsf{P}}(1)$. From $n^{1/2}(\hat{\sigma}^2_{nm} - \sigma^2) = (\hat{\sigma}_{nm} + \sigma)n^{1/2}(\hat{\sigma}_{nm} - \sigma) = 2\sigma\hat{a}_{nm}(1 + \mathrm{o}_{\mathsf{P}}(1))$ we find that a similar argument shows

$$\hat{a}_{nm} = (2\sigma\tau_2)^{-1}n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i^2 - \psi^{-1}\sigma^2\tau_2)1_{(\sigma\underline{c}<\varepsilon_i\leq\sigma\bar{c})} + (2\tau_2)^{-1}(\zeta_2\mu'\hat{b}_{n,m-1} + \zeta_3\hat{a}_{n,m-1}) + R_{a,n}(\hat{u}_{n,m-1}),$$

where $\sup_{|u|<U}|R_{a,n}(u)| = \mathrm{o}_{\mathsf{P}}(1)$. ■

Proof of Theorem 3.3. We want to show that for all $\epsilon > 0$ there exists $U > 0$, and n_0 so that for $n \geq n_0$ it holds

$$\mathsf{P}(\sup_{0\leq m<\infty}|\hat{u}_{nm}| \leq U) \geq 1 - \epsilon. \tag{A.2}$$

From the recursion (2.13) we find the representation

$$\hat{u}_{nm} = \Gamma^m\hat{u}_{n0} + \sum_{\ell=1}^{m}\Gamma^{\ell-1}\{K_n + R_n(\hat{u}_{n,m-\ell})\} \tag{A.3}$$

and the evaluation

$$|\hat{u}_{nm}| \leq ||\Gamma^m||\,|\hat{u}_{n0}| + (|K_n| + \max_{0\leq\ell\leq m-1}|R_n(\hat{u}_{n\ell})|)\sum_{\ell=1}^{m}||\Gamma^{\ell-1}||.$$

By assumption a δ exists so that $\max|\mathrm{eigen}(\Gamma)| < \delta < 1$. Gelfand's formula, Varga (2000, Theorems 1.5, 3.4), then shows there is an $m_0 > 0$ so for all $m > m_0$ then $||\Gamma^m|| \leq \delta^m$. This in turn implies for some $c > 1$, that $\max_{0\leq m<\infty}||\Gamma^m|| < \sum_{\ell=0}^{\infty}||\Gamma^\ell|| < c$, and hence

$$|\hat{u}_{nm}| \leq c\{|\hat{u}_{n0}| + |K_n| + \max_{0\leq\ell\leq m-1}|R_n(\hat{u}_{n\ell})|\}. \tag{A.4}$$

Because it is assumed that $\hat{u}_{n0}$ is tight, and the sequence $\{K_n\}$ is tight by Theorem 3.1, and $\max_{|u|\leq U_1}|R_n(u)| = o_{\mathsf{P}}(1)$ by Theorem 3.2, then constants $U_0 > \eta/2, n_0 > 0$ exist so that for $n \geq n_0$, the set

$$\mathcal{A}_n = (c|\hat{u}_0| \leq U_0) \cap (c|K_n| \leq U_0) \cap (c \max_{|u|\leq 3U_0} |R_n(u)| \leq \eta/2) \tag{A.5}$$

has probability larger than $1-\epsilon$.

An induction over m is now used to show that $\sup_{0\leq m<\infty}|\hat{u}_{nm}| \leq 3U_0$ on the set $\mathcal{A}_n$. As induction start, for $m=0$, then $|\hat{u}_{n0}| \leq c^{-1}U_0 < 3U_0$ by the tightness assumption to $\hat{u}_0$ and $c > 1$. The induction assumption is that $\max_{0\leq\ell\leq m-1}|\hat{u}_{n\ell}| \leq 3U_0$. This implies that on the set $\mathcal{A}_n$ then $c\max_{0\leq\ell\leq m-1}|R_n(\hat{u}_{n\ell})| \leq c\max_{|u|\leq 3U_0}|R_n(u)| \leq \eta/2$. Thus, the bound (A.4) becomes $|\hat{u}_{nm}| \leq 2U_0 + \eta/2 \leq 3U_0$. It follows that $\max_{0\leq\ell\leq m}|\hat{u}_{n\ell}| \leq 3U_0$. This proves (A.2) for $U = 3U_0$. ■

Proof of Theorem 3.4. We want to show that for all $\eta, \epsilon > 0$ there is a n_0 and m_0 so that for $n \geq n_0$ and $m \geq m_0$ it holds that

$$\mathsf{P}(|\hat{u}_{n*} - (I_{p+1} - \Gamma)^{-1}K_n| > \eta) < \epsilon. \tag{A.6}$$

In order to show (A.6), note that $\sum_{\ell=1}^{m}\Gamma^{\ell-1} = (I_{p+1}-\Gamma^m)(I_{p+1}-\Gamma)^{-1}$ where $(I_{p+1}-\Gamma)^{-1} = \sum_{\ell=0}^{\infty}\Gamma^{\ell}$. Therefore equation (A.3) shows that

$$\hat{u}_{nm} - (I_{p+1}-\Gamma)^{-1}K_n = \Gamma^m\{\hat{u}_{n0} - (I_{p+1}-\Gamma)^{-1}K_n\} + \sum_{\ell=1}^{m}\Gamma^{\ell-1}R_n(\hat{u}_{n,m-\ell}).$$

To bound this, note first that $||(I_{p+1}-\Gamma)^{-1}|| = ||\sum_{\ell=0}^{\infty}\Gamma^{\ell}|| \leq \sum_{\ell=0}^{\infty}||\Gamma^{\ell}|| < c$. Thus on the set $\mathcal{A}_n$, see (A.5), it holds that

$$|\hat{u}_{nm} - (I_{p+1}-\Gamma)^{-1}K_n| \leq ||\Gamma^m||(c^{-1}U_0 + U_0) + c\max_{0\leq\ell\leq m-1}|R_n(\hat{u}_{n\ell})| \leq ||\Gamma^m||2U_0 + \eta/2.$$

Now, for $m \geq m_0$ then $||\Gamma^m|| \leq \delta^m$. Since δ^m declines exponentially, m_0 can be chosen so large that it also holds that $||\Gamma^m||2U_0 \leq \eta/2$. Thus $\mathsf{P}(|\hat{u}_{nm} - (I_{p+1}-\Gamma)^{-1}K_n| \geq \eta) < \epsilon$, for $m \geq m_0$ and $n \geq n_0$ which proves (A.6). ■

Proof of Theorem 3.5. The matrices Γ and $\Gamma - \lambda I_{p+1}$ are of the form

$$\begin{pmatrix} aI_p & b \\ c' & d \end{pmatrix},$$

and the result follows from the identity

$$a\det\begin{pmatrix} aI_p & b \\ c' & d \end{pmatrix} = \det\begin{pmatrix} I_p & 0 \\ -c' & a \end{pmatrix}\det\begin{pmatrix} aI_p & b \\ c' & d \end{pmatrix} = \det\begin{pmatrix} aI_p & b \\ 0 & ad-c'b \end{pmatrix} = a^p(ad-c'b).$$

■

Proof of Theorem 3.6. (*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 $\lim_{c\to 0} \mathsf{f}''(c) < 0$. It holds $\lim_{c\to 0} c^{-1} \int_0^c \mathsf{f}(x)dx = \mathsf{f}(0)$ and $\lim_{c\to 0} \xi_1/(2c) = \mathsf{f}(0)$ so $\lim_{c\to 0} \xi_1/\psi = 1$. Similarly, $\int_0^\infty \mathsf{f}(x)dx = 1$ and $\lim_{\psi\to 1} c\mathsf{f}(c) \to 0$ so $\lim_{\psi\to 1} \xi_1/\psi = 0$.

(*b*) We find

$$g(c) = \zeta_3/(2\tau_2) = \xi_3/(2\tau_2) - \xi_1/(2\tau_0) = \frac{2cf(c)\{\int_0^c (c^2 - x^2)f(x)dx\}}{\tau_2\tau_0} > 0. \tag{A.7}$$

For $c \to 0$, or $\psi \to 0$, we find the approximations for $k = 0, 1 : \tau_{2k} = 2\int_0^c x^{2k}\mathsf{f}(x)dx \approx 2c^{2k+1}\mathsf{f}(0)/(2k+1)$, which show that $g(c) \to 1$.

For $c \to \infty$, or $\psi \to 1$, we find $\tau_0 \to 1, \tau_2 \to 1$ and $g(c) \approx 2c\mathsf{f}(c)(c^2 - 1) \to 0$ because f is assumed to have finite third moment.

(*c*) Using $c\tau_0' = 2c\mathsf{f}(c)$ we find from (A.7) that $g(c) < 1$ if

$$h(c) = \frac{c\tau_0'}{\tau_0}(c^2\tau_0 - \tau_2) - \tau_2 = \frac{2c\mathsf{f}(c)}{\tau_0}\{\int_0^c (c^2 - x^2)f(x)dx\} - \tau_2 < 0,$$

and because the limit for $c \to 0$ is zero it is enough to show that $h'(c) < 0$.

We find

$$h'(c) = (\frac{c\tau_0'}{\tau_0})'(c^2\tau_0 - \tau_2) + \frac{c\tau_0'}{\tau_0}(2c\tau_0 + c^2\tau_0' - \tau_2') - \tau_2' = (\frac{c\tau_0'}{\tau_0})'(c^2\tau_0 - \tau_2),$$

because the extra term vanishes:

$$\frac{c\tau_0'}{\tau_0}(2c\tau_0 + c^2\tau_0' - \tau_2') - \tau_2' = 2c^2\mathsf{f}(c) + c^3\frac{(2\mathsf{f}(c))^2}{\tau_0} - \frac{2c^3\mathsf{f}(c)2\mathsf{f}(c)}{\tau_0} - 2c^2\mathsf{f}(c) = 0.$$

Because $c^2\tau_0 - \tau_2 > 0$ and $(\frac{c\tau_0'}{\tau_0})' = [c\{\log\int_0^c \mathsf{f}(x)dx\}']' < 0$ by assumption we find $g(c) < 1$.

(*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 $\lim_{c\to 0} L = 0$ 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.

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.

Cavaliere, G. and Georgiev, I. (2011) Exploiting infinite variance through dummy variables in an AR model. Discussion paper, Universidade Nova de Lisboa.

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.

Hendry, D.F., Johansen, S. and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23, 317–335 and Erratum 337–339.

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

Huber, P.J. and Ronchetti E.M. (2009) *Robust Statistics.* Second edition, New York: Wiley.

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.

Johansen, S. and Nielsen, B. (2013a) Asymptotic analysis of the Forward Search. Discussion paper

Johansen, S. and Nielsen, B. (2013b). A stochastic expansion of the Huber-skip estimator for regression analysis. Discussion paper.

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

Jurečková, J., Sen, P.K. and Picek, J. (2012) *Methodological Tools in Robust and Nonparametric Statistics. London,* Chapman & Hall/CRC Press.

Koenker, R. (2005) *Quantile Regression.* Cambridge: Cambridge Books, Cambridge University Press.

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

Rousseeuw, P.J. and Leroy, A. M. (1987) *Robust regression and outlier detection.* New Jersey : Wiley.

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. (2006a) The least trimmed squares. Part I: Consistency. *Kybernetika* 42, 1–36.

Víšek, J.Á. (2006b) The least trimmed squares. Part II: $\sqrt{n}$-consistency *Kybernetika,* 42, 181–202.

Víšek, J.Á. (2006c) 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.