



AARHUS UNIVERSITY



Cover sheet

This is the accepted manuscript (post-print version) of the article.

The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication

Please cite the final published version:

Johansen, S., & Nielsen, B. (2019). Boundedness of M-estimators for linear regression in time series. Econometric Theory, 35(3), 653-683. <https://doi.org/10.1017/S0266466618000257>

Publication metadata

Title:	Boundedness of M-estimators for linear regression in time series.
Author(s):	Johansen, Søren ; Nielsen, Bent
Journal:	Econometric Theory
DOI/Link:	10.1017/S0266466618000257
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Boundedness of M-estimators for linear regression in time series

Søren Johansen^{1,2} & Bent Nielsen³

7 September 2017

Summary: We show boundedness in probability uniformly in sample size of a general M-estimator for multiple linear regression in time series. The positive criterion function for the M-estimator is assumed lower semi-continuous and sufficiently large for large argument. Particular cases are the Huber-skip and quantile regression. Boundedness requires an assumption on the frequency of small regressors. We show that this is satisfied for a variety of deterministic and stochastic regressors, including stationary and random walks regressors. The results are obtained using a detailed analysis of the condition on the regressors combined with some recent martingale results.

Keywords: M-estimator, robust statistics, martingales, Huber-skip, quantile estimation, boundedness.

1 Introduction and summary

We show boundedness in probability uniformly in sample size, n , for a class of regression M-estimators. Thus we show tightness of their non-standardized distributions, see Billingsley (1968). The objective function can be non-convex and non-continuous. A prominent example of an estimator with a non-convex objective functions is the skip estimator suggested by Huber (1964), where each observation contributes to the objective function through a criterion function, which is quadratic in the central part and horizontal otherwise. The boundedness result addresses a difficulty which is often met in asymptotic analysis of problems, where the objective function is non-convex. A very common solution is to assume a compact parameter space, see for instance the analysis of M-estimators for general regression by Liese and Vajda (1994), LTS-estimators by Čížek (2005), Víšek (2006), MM-estimators by Fasano, Maronna, Sued and Yohai (2012) and M-estimators for unit root processes by Knight (1989) and Lucas (1995). Such an assumption circumvents the problem through a condition on the unknown parameter and it is therefore rarely satisfactory from an applied viewpoint. Instead, our result only requires an assumption that can be justified by inspecting the observed regressors and the objective function.

We consider the multiple linear regression and use the notation

$$y_i = \mu + x'_{ni}\alpha + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

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

²Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. Address: Department of Economics University of Copenhagen, Øster Farimagsgade 5 Building 26, 1353 Copenhagen K, Denmark. E-mail: soren.johansen@econ.ku.dk.

³Nuffield College & Department of Economics, University of Oxford & Programme for Economic Modelling. E-mail: bent.nielsen@nuffield.ox.ac.uk.

see (2.1) for assumptions on regressors and error term. The M-estimator for the $(m + 1)$ -dimensional parameter $(\mu, \alpha)'$ is the minimizer of the objective function

$$R_n(\mu, \alpha) = \frac{1}{n} \sum_{i=1}^n \rho(y_i - \mu - x'_{ni}\alpha), \quad (1.2)$$

for some criterion function ρ . M-estimators were originally introduced for location problems by Huber (1964) but later extended to regression models, see Maronna, Martin, and Yohai (2006), Huber and Ronchetti (2009), or Jurečková, Sen, and Picek (2012) for recent monographs on the topic. The class of M-estimators considered includes the Huber-skip estimator, which has a non-convex and non-differentiable criterion function, as well as quantile regression estimators and least squares estimators, which are based on convex criterion functions.

The asymptotic theory of the regression M-estimator is well understood for convex criterion functions ρ . An example could be median regression or least absolute deviation regression where $\rho(u) = |u|$ or more generally quantile regression. The consistency of such estimators could be argued as follows. In a finite sample, minimizers to (1.2) are finite as long as the design matrix given by $(1, x_{ni})$ has full rank, while the criterion $\rho(u)$ is a continuous, nondecreasing and unbounded function of $|u|$, see Maronna, Martin, and Yohai (2006, Theorem 10.14). For median regression, the minimizer may not be unique, and measurability is an issue, see Jennrich (1969). In large samples additional conditions are needed to ensure that the sequence of solutions does not diverge. Knight (1989, Lemma A) presents an elegant result that applies for convex ρ . To apply this result consider R_n as a convex process on \mathbb{R}^{m+1} . If the finite dimensional distributions of R_n converge and the limit has a unique minimizer, then any (measurable) minimizer of R_n converges in distribution. Knight (1989, Theorem 2) apply this idea to median regressions with a random walk regressor albeit without an intercept. Koenker and Xiao (2004) generalize this work to quantile regressions.

For non-convex criterion functions the asymptotic theory is more complicated. Boundedness and consistency is now harder to establish and is often assumed, which is not appealing from an applied viewpoint. Examples include Knight (1991) and Lucas (1995), who consider M-estimation with a random walk regressor, while Abadir and Lucas (2000) consider unit root testing. In all cases the criterion ρ is twice differentiable, but not necessarily convex. This class of functions include Tukey's biweight function, but not the Huber-skip function. Their results apply to minimizers $\hat{\mu}, \hat{\alpha}$ that are assumed to be consistent.

Boundedness and consistency has, however, been established for non-convex ρ in some situations. Chen and Wu (1988) give two further results on boundedness and consistency for more general criterion functions. In both cases the criterion function $\rho(u)$ is continuous and nondecreasing in $|u| > 0$. It need not to be differentiable, so the Huber-skip function is now in consideration. Their Theorem 1 shows boundedness of $\hat{\mu}, \hat{\alpha}$ when (y_i, x'_{ni}) are i.i.d. and $E\rho(y_i - \mu - x'_{ni}\alpha)$ has a unique minimum. Their Theorem 4 shows boundedness when x_{ni} is deterministic and satisfies a condition on the frequency of small regressors. The advantage of this result is that the condition on the frequency of small regressors can be justified in particular examples at least in principle. It appears that this condition has not been explored that much. Our contribution is therefore to generalize the Chen and Wu result to situations with stochastic regressors and to explore the condition on the regressors for particular regressors including various deterministic regressors, stationary processes and random walks.

In the present paper we focus on boundedness of the minimizer. This is for two reasons. Boundedness is typically harder to establish than consistency. The conditions for boundedness

and for consistency given boundedness are somewhat different, as they are based on the criterion function and its (pseudo-) derivatives, respectively. By focusing on boundedness we can work with a clean set of conditions and discuss how these can be verified in particular situations. For this, we assume ρ is semi-continuous and nonnegative with a minimum at zero and greater than $\rho_* > 0$ for large values of the argument. We also need an extra condition on the expected shifted criterion function $h(v) = \mathbb{E}\rho(\varepsilon_i - v)$, which is assumed to take a value below ρ_* somewhere in the central part of the distribution of the error term. The only condition to the regressors is a condition on the frequency of small regressors, which is weaker than the condition of Chen and Wu (1988), albeit stronger than the conditions for the boundedness of least square estimators. The latter illustrates the price we pay by leaving the least squares criterion. The condition is related to a condition for deterministic regressor used by Davies (1990) for S-estimators. Our condition is, however, formulated in a slightly different way, which seems to be easier to check for particular regressors. Indeed, we check the condition for a few situations. We give a number of examples with deterministic regressors to illustrate the condition. We also show that the condition is satisfied for stationary regressors and for random walk regressors under suitable conditions on the conditional density or the density respectively.

The proof of the boundedness builds on a result concerning the supremum of a family of martingales on a compact set. In a way this result contributes to replacing the convexity lemma of Knight (1989) mentioned above. The result is proved using the iterated martingale inequality by Johansen and Nielsen (2016), which we present as Lemma 4.1 for convenience. We present a second such martingale result, which is applied when checking the boundedness conditions for stationary processes.

It is worth noting that for the general results, the innovations are neither required to have a zero expectation nor a continuous density. Thus, the results apply both when the innovations follow a non-contaminated reference distribution with density f_0 , say, and when they are contaminated so that they follow a mixture distribution with density $(1 - \epsilon)f_0 + \epsilon f_1$, say. For simplicity we will, however, require that the innovations are identically distributed. This assumption could potentially be relaxed as the proofs use martingale techniques rather than results designed for an i.i.d. situation. All proofs are given in the appendix.

2 Model, assumptions and main result

We define the model and some notation and then give the assumptions and the boundedness result.

2.1 Formulation of the multiple regression model

To define the multiple regression model we consider a filtration \mathcal{F}_i , and errors ε_i , $i = 1, \dots, n$, and assume ε_i is \mathcal{F}_i measurable and independent of \mathcal{F}_{i-1} and i.i.d. The model is defined by the equations

$$y_i = \mu + x_{ni}'\alpha + \varepsilon_i, \quad i = 1, \dots, n. \quad (2.1)$$

The m -dimensional regressors x_{ni} may be deterministic, stationary or even stochastically or deterministically trending. If x_{ni} is stochastic, we assume that it is adapted to \mathcal{F}_{i-1} .

This notation is chosen to cover a number of cases. The leading case is $y_i = \mu + x_i'\alpha + \varepsilon_i$, where the regressors do not depend on n , but in $y_i = \mu + \alpha 1_{(i \leq \tau n)} + \varepsilon_i$, the regressor $1_{(i \leq \tau n)}$

depends on n . If the regressors are $(1, i)$, we normalize the regressor as $(1, i/n)$ and consider $y_i = \mu + \alpha(i/n) + \varepsilon_i$, and if x_i is a random walk we consider $y_i = \mu + \alpha(x_i/n^{1/2}) + \varepsilon_i$.

An M-estimator $(\hat{\mu}, \hat{\alpha})'$ is a minimizer for $\mu \in \mathbb{R}$ and $\alpha \in \mathbb{R}^m$ of

$$R_n(\mu, \alpha) = \frac{1}{n} \sum_{i=1}^n \rho(y_i - \mu - x'_{ni}\alpha). \quad (2.2)$$

Special criterion functions are the Huber-skip defined by $\rho(u) = \min(u^2, c^2)/2$, and the Huber estimator defined by the convex function $\rho(u) = \frac{1}{2}u^2\mathbf{1}_{(|u| \leq c)} + c(|u - c| + \frac{1}{2}c)\mathbf{1}_{(|u| > c)}$. The formulation also covers least squares regression, $\rho(u) = u^2/2$, and quantile regression, $\rho(u) = -(1-p)u\mathbf{1}_{(u < 0)} + pu\mathbf{1}_{(u > 0)}$, for some $0 < p < 1$. In particular for $p = 1/2$ we get the least absolute deviation. Note that the two Huber estimators require that the scale is known, whereas this is not a requirement for least squares and quantile regression. Finally if f is the density of the errors then $\rho(u) = -\ln f(u)$ gives the maximum likelihood estimators.

2.2 The assumptions and the result on boundedness

For the boundedness result, we need a condition on the frequency of small regressors, see Assumption 1(*iii*). This is related to the assumptions of Chen and Wu (1988) and Davies (1990), see Section 3, where we also discuss how to check the condition in some specific situations.

The proof of the uniform boundedness of the estimator relies on a bound on the supremum of a family of martingales indexed by a continuous parameter in a compact set, which is evaluated using a recent martingale result, see Lemma 4.1 or Johansen and Nielsen (2016, Lemma 5.2). The proof requires a moment condition that depends linearly on the dimension of the regressors, see Assumption 1(*ic*). We refrain from exploring the heterogeneity allowed by the martingale theory and require i.i.d. innovations for specificity in Assumption 1(*i*).

The required assumptions on the criterion function ρ are modest. It must exceed a threshold for large values of u , see (*ib*), but it need not rise monotonically from the origin. Lower semi-continuity in (*ia*) is used to ensure the existence of a minimizer on a compact set, and continuity is needed to find a measurable minimizer. The value of μ_* is chosen so the shifted criterion function $\rho(\varepsilon_i - \mu_*)$ has expectation less than $\rho(u_*)$.

For the formulation of the assumptions and results we use the notation

$$z_{ni} = \begin{pmatrix} 1 \\ x_{ni} \end{pmatrix} \in \mathbb{R}^{m+1}, \quad \beta = \begin{pmatrix} \mu \\ \alpha \end{pmatrix}, \quad \Sigma_n = n^{-1} \sum_{i=1}^n z_{ni} z'_{ni}.$$

Assumption 1 (*i*) Let $\mathcal{F}_i, i = 1, \dots, n$ be a filtration and assume ε_i is measurable with respect to \mathcal{F}_i and independent of \mathcal{F}_{i-1} and i.i.d. with variance σ^2 .

(*ii*) The **Criterion function** satisfies $\rho(u) \geq 0$, $\rho(0) = 0$ and the conditions

(a) ρ is lower semi-continuous so that $\liminf_{v \rightarrow u} \rho(v) \geq \rho(u)$ for all $u \in \mathbb{R}$;

(b) Let $0 < h(v) = \mathbf{E}\{\rho(\varepsilon_i - v)\} < \infty$, and let $\mu_*, u_* \in \mathbb{R}$ exist so that

$$0 < h(\mu_*) < \rho_* = \inf_{|u| \geq |u_*|} \rho(u);$$

(c) $\mathbf{E}\{\rho(\varepsilon_i - \mu_*)\}^{2^r} < \infty$ for some $r \in \mathbb{N}$ so that $2^r > m + 1 = \dim z_{ni}$.

(*iii*) **Frequency of small regressors.** Define

$$F_n(a) = \sup_{|\delta|=1} F_{n\delta}(a) = \sup_{|\delta|=1} n^{-1} \sum_{i=1}^n \mathbf{1}_{(|z'_{ni}\delta| \leq a)}. \quad (2.3)$$

Suppose

- (a) $\lim_{(a,n) \rightarrow (0,\infty)} \mathbf{P}[\sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\} \geq \epsilon] = 0$ for all $\epsilon > 0$;
- (b) a $0 < \xi < 1$ exists such that $\lim_{n \rightarrow \infty} \mathbf{P}\{F_n(0) \geq \xi\} = 0$.

We note that Assumption 1 does not involve the unknown parameter and in particular not any compactness of the parameter space. Indeed, Assumption 1 is concerned with (i) the properties of the innovations, which in principle can be checked through specification testing; (ii) the choice of criterion function and its expectation, (iii) a condition on the regressors, which can be checked by inspecting the regressors.

The density of the innovations is not directly constrained by Assumption 1 as this only constrains ε_i indirectly through the criterion function ρ in condition (ii). In particular, the density of ε_i is not required to be continuous. This is relevant for quantile regressions where Koenker and Bassett (1978) and Koenker and Xiao (2004) assume continuity. A certain number of moments is required for $\rho(\varepsilon_i - \mu_*)$ in condition (iic). This condition is, however, not binding for robust estimators such as the Huber-skip estimator where the criterion function is bounded.

The condition to the regressors given in condition (iii) involves the function $F_{n\delta}(a)$. For a fixed δ this is a discrete distribution function, in the sense that it takes values on $(0, 1, \dots, n)/n$ and it is continuous from the right with limits from the left. We show that $F_n(a)$ is also a discrete distribution function and give a detailed analysis of condition (iii) in §3.

For now, we give the main result on boundedness.

Theorem 2.1 Boundedness. *Under Assumption 1, we can for all $\epsilon > 0$ find $B, n_0 > 0$, and sets \mathbb{C}_n with $P(\mathbb{C}_n) \geq 1 - \epsilon$ for $n \geq n_0$, such that on \mathbb{C}_n a minimizer $\hat{\beta}$ of $R_n(\beta)$ exists on the set $(\beta : |\beta| \leq B)$ and any minimizer satisfies*

$$|\hat{\beta}| \leq B.$$

Theorem 2.2 Measurability. *Under Assumption 1, and if ρ is continuous, a measurable minimizer $\hat{\beta}$ of $R_n(\beta)$ exists and satisfies*

$$|\hat{\beta}| = O_{\mathbf{P}}(1).$$

A feature of Assumption 1 is that it separates the conditions on the criterion function and the innovations on the one hand and the regressors on the other hand. Abandoning that aim it is possible to formulate weaker assumptions to the regressors depending on the nature of the criterion function and the innovations, see Remarks 3.5, A.6.

3 The assumption to the frequency of small regressors

In this section, we illustrate Assumption 1(iii) concerning the frequency of small regressors through some examples. We start with some general remarks and then proceed to relate it to the quantity $\lambda_n(\xi)$ of Davies (1990), who considered S -estimators for fixed regressors, and to a condition in Chen and Wu (1988). We show that our condition is satisfied for a number of different regressors including random walks and stationary processes with a boundedness condition on a conditional density.

3.1 Some general remarks

The assumption to the frequency of small regressors is a little complicated. We start by relating it to the behaviour of the sum of squared regressors. We then proceed to comment briefly on the conditions by Chen and Wu (1988) and Davies (1990) before giving a more detailed analysis in the next subsection.

Remark 3.1 *Assumption 1(iiib) implies that $\mathbb{P}(\sum_{i=1}^n z_{in}z'_{in}$ is positive definite) $\rightarrow 1$, because if $\delta'\hat{\Sigma}_n\delta = 0$, then $z'_{ni}\delta = 0$ for $i = 1, \dots, n$, and $F_{n\delta}(0) = 1$.*

Remark 3.2 *A common condition for consistency of least squares estimators is that $\hat{\Sigma}_n = n^{-1} \sum_{i=1}^n z_{in}z'_{in}$ is bounded away from zero in large samples. We argue that this is implied by the conditions to F_n . We prove this by noting that*

$$\delta'\hat{\Sigma}_n\delta \geq n^{-1} \sum_{i=1}^n \delta'z_{ni}z'_{ni}\delta 1_{(|z'_{ni}\delta|>a)} \geq a^2 n^{-1} \sum_{i=1}^n 1_{(|z'_{ni}\delta|>a)} = a^2 \{1 - F_{n\delta}(a)\}.$$

Adding and subtracting $F_{n\delta}(0)$ and taking supremum over δ gives the further bound

$$\delta'\hat{\Sigma}_n\delta \geq a^2 [1 - \sup_{|\delta|=1} F_{n\delta}(0) - \sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\}] \geq a^2 (1 - \xi - \epsilon) > 0,$$

with large probability for large n , for $\epsilon < 1 - \xi$ chosen according to Assumption 1(iiia, b).

Remark 3.3 *When the regressors are deterministic and bounded, then the least squares condition that Σ_n is bounded away from zero implies the F_n condition (iiib). This represents the opposite implication of the result in Remark 3.2. If the regressors are deterministic and bounded and λ_n , the smallest eigenvalue of $\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$, satisfies $\liminf_{n \rightarrow \infty} \lambda_n > 0$ then $a > 0$, $\xi < 1$ exist so that $F_n(a) \leq \xi$ and the Assumption 1(iiib) is satisfied, see Chen and Wu (1988, Lemma 6). The argument depends critically on the boundedness of the regressors.*

Remark 3.4 *If $F_n(a) = o_{\mathbb{P}}(1)$ as $(a, n) \rightarrow (0, \infty)$ then Assumption 1(iii) is satisfied. This is because $F_{n\delta}$ is an distribution function so that $F_{n\delta}(a) - F_{n\delta}(0) \leq F_{n\delta}(a)$ while $F_{n\delta}(a) \leq \max_{|\delta|=1} F_{n\delta}(a) = F_n(a)$. To be precise, it suffices that for all $\epsilon > 0, \eta > 0$ there exist $a_0, n_0 > 0$ such that*

$$\mathbb{P}\{F_n(a) \geq \eta\} \leq \epsilon \quad \text{for } a \leq a_0, n \geq n_0. \quad (3.1)$$

Chen and Wu (1988, Theorem 4) assume the regressors are deterministic and that $F_n(a) = o(1)$ as $(a, n) \rightarrow (0, \infty)$.

Remark 3.5 *If Assumption 1(iiia, b) are satisfied and the regressors are deterministic, it holds that $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi$, see Davies (1990). We return to this issue in Section 3.2 and Example 3.5.*

3.2 Relation to conditions in the literature

Chen and Wu (1988, Theorem 4) show in the regression $y_i = \mu + \alpha x_i + \varepsilon_i$, that $(\hat{\mu}, \hat{\alpha}) \rightarrow (\mu_0, \alpha_0)$ a.s. under the following conditions. The regressors are deterministic, the criterion function is bounded, $0 < \rho(\infty) = \rho(-\infty) < \infty$, and $F_n(a) \rightarrow 0$ as $(a, n) \rightarrow (0, \infty)$, noting that

F_n is deterministic when the regressors are deterministic. These conditions are relaxed in this paper, albeit we only consider weak consistency. We allow quite general time series regressors, drop the condition $\rho(\infty) = \rho(-\infty) < \infty$, and give a weaker Assumption 1(iii).

Davies (1990) considers S-estimators rather than M-estimators and proves boundedness for symmetric density f and deterministic regressors. He defines for $0 < \xi < 1$

$$\lambda_n(\xi) = \min_{|S|=\text{int}(n\xi)} \min_{|\delta|=1} \max_{i \in S} |z'_{ni}\delta|, \quad (3.2)$$

where S are subsets of the indices $i = 1, \dots, n$. It is a consequence of his Theorem 3, that if $\liminf_{n \rightarrow \infty} \lambda_n(\xi) > 0$ for some $\xi > 0$, then the S-estimator for β is consistent. If $\xi = 0$ then $\lambda_n(0) = 0$, and for $\xi = 1$ then $\lambda_n(1) = \min_{|\delta|=1} \max_{1 \leq i \leq n} |z'_{ni}\delta|$.

We next give a result that compares Davies' function λ_n with the function F_n . We first show that F_n and λ_n are distribution functions, and that λ_n is a type of inverse for F_n .

Theorem 3.1 (i) *The function F_n is a discrete distribution function.*

(ii) *The function λ_n is a right continuous, piecewise constant function satisfying*

$$\{\lambda_n(k/n) \leq a\} = \{F_n(a) \geq k/n\}, \text{ for } k = 0, \dots, n \text{ and } a \geq 0. \quad (3.3)$$

(iii) *Consequently, $\lambda_n(\xi) = \inf\{a : F_n(a) \geq \text{int}(n\xi)/n\}$.*

Theorem 3.1(iii) shows that λ_n is a type of inverse for F_n although it is neither the lower quantile function nor the upper quantile functions as those are defined as $q_n^{\text{lower}}(\xi) = \inf\{a : F_n(a) \geq \xi\}$ and $q_n^{\text{upper}}(\xi) = \inf\{a : F_n(a) > \xi\}$ respectively. For $\xi = k/n$ we have that $\lambda_n(k/n) = q_n^{\text{lower}}(k/n)$. Next, we compare the condition $\liminf_{n \rightarrow \infty} \lambda_n(\xi) > 0$ with Assumption 1(iii) that $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi$, see also Remark 3.5.

Theorem 3.2 *Consider an array of deterministic regressors z_{ni} . Then there is equivalence between the condition that $\liminf_{n \rightarrow \infty} \lambda_n(\xi^*) > 0$ for some $0 < \xi^* < 1$ and that there exists $0 < \xi < 1$ for which $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi$.*

In the following we give some examples of simple regressors and show that Assumption 1(iii) is satisfied. We apply a simple evaluation given in the next result.

Lemma 3.1 *For any $0 \leq c \leq 1/2$, $|\theta| < \pi/2$*

$$|-\sin \theta + x \cos \theta| \leq c \quad \Rightarrow \quad \frac{1}{\cos \theta} \leq 2(1 + |x|). \quad (3.4)$$

We consider first the regression $y_i = \mu + \alpha 1_{(i \leq \tau n)} + \varepsilon_i$, and then the regressions $y_i = \mu + \alpha(i/n)^q + \varepsilon_i$, for different q .

Example 3.1 *The regression $y_i = \mu + \alpha 1_{(i \leq \tau n)} + \varepsilon_i$ for some $0 < \tau < 1$. We show that $z_{ni} = \{1, 1_{(i \leq \tau n)}\}'$ satisfies Assumption 1(iii). We use a geometric proof illustrated by Figure 1. The regressors take values in the points $(1, 0)$ and $(1, 1)$ with frequency $1 - \tau$ and τ , respectively. The direction $\delta = (-\sin \theta, \cos \theta)'$ for $\theta = \pi/4$ is illustrated with a diameter. The radial through $(1, 1)$ is the orthogonal complement with angle θ . The two parallel lines at a distance of a to the radial indicate which points are counted towards $F_{n\delta}(a)$. Thus, by varying θ , and thereby turning the diameter, we see that if a is sufficiently small, $0 \leq a < 1/2$ say, then $F_n(a) = \sup_{|\delta|=1} F_{n\delta}(a) = \max(\tau, 1 - \tau)$. In particular $F_{n\delta}(a) - F_{n\delta}(0) = 0$ and $F_n(0) = \max(\tau, 1 - \tau) < 1$ so that Assumption 1(iii) is satisfied. However, since $F_n(0) > 0$, the assumption $F_n(a) \rightarrow 0$, $(a, n) \rightarrow (0, \infty)$, used by Chen and Wu (1988), is not satisfied, see also Remark 3.4.*

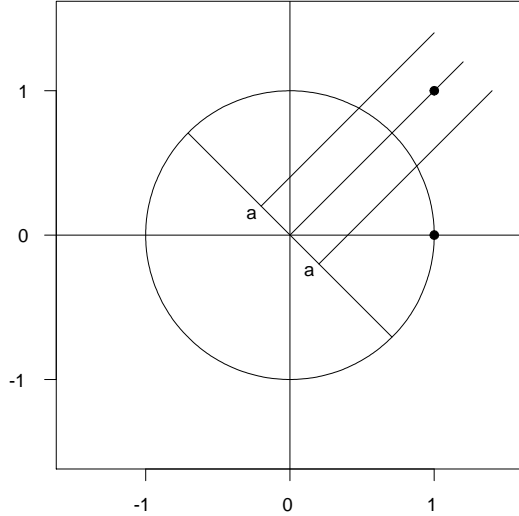


Figure 1: Illustration of $F_{n\delta}$ in the dummy variable case

Example 3.2 The regression $y_i = \mu + \alpha(i/n)^q + \varepsilon_i$, with $q > 0$. In terms of Figure 1, the points $z_{ni} = \{1, (i/n)^q\}'$ are spaced on the line between the points $(1, 0)$ and $(1, 1)$. For large n , their distribution can be described by the density $q^{-1}x^{1/q-1}$, $x \in [0, 1]$. For $|\delta'z_{ni}| = |-\sin\theta + (i/n)^q \cos\theta| \leq a$, $\cos\theta > 0$, the basic inequality is in all cases

$$\frac{-a + \sin\theta}{\cos\theta} \leq (i/n)^q \leq \frac{a + \sin\theta}{\cos\theta}. \quad (3.5)$$

This describes an interval for i of length $n\{(a + \sin\theta)^{1/q} - (-a + \sin\theta)^{1/q}\}/(\cos\theta)^{1/q}$ for $\sin\theta > a$.

For $q = 1$, the density $q^{-1}x^{1/q-1}$ is uniform on $[0, 1]$, and we can use the inequality, see (3.4), that $(\cos\theta)^{-1} \leq 2/(1-a)$. It follows that the length of the interval is bounded by $2na/\cos\theta \leq 4na/(1-a)$, such that $F_n(a) \leq 4a/(1-a) \rightarrow 0$, for $(a, n) \rightarrow (0, \infty)$, and Assumption 1(iii) is satisfied.

For $q > 1$, the density $q^{-1}x^{1/q-1}$ has most mass close to $x = 0$ and the largest number of points in the interval we find for θ small. The smallest value is found for $(\sin\theta - a)/\cos\theta = n^{-q}$, such that $\sin\theta - a = O(n^{-q})$, and $\cos\theta = (1 - a^2)^{1/2}\{1 + O(n^{-q})\}$. It follows that $F_n(a) \leq c\{2a/(1 - a^2)^{1/2}\}^{1/q} \rightarrow 0$, for $(a, n) \rightarrow (0, \infty)$.

Finally if $0 < q < 1$ the density gives most mass to points close to 1, so we choose an interval using $\sin\theta$ close to $\pi/4$, that is $(\sin\theta + a)/\cos\theta = 1$. This implies $\sin\theta = 1/\sqrt{2} - a + o(a)$ and $\cos\theta = \{1 - (1/\sqrt{2} - a)^2\}^{1/2}\{1 + o(a)\} = 1/\sqrt{2} + o(a)$. This gives the bound $F_n(a) \leq c\{(1/\sqrt{2})^{1/q} - (1/\sqrt{2} - 2a)^{1/q}\}/(1/\sqrt{2})^{1/q} \rightarrow 0$, for $(a, n) \rightarrow (0, \infty)$.

Example 3.3 The regression $y_i = \mu + \alpha(i/n)^q + \varepsilon_i$, with $-1/2 < q < 0$. The density of the points is now $|q|^{-1}x^{1/q-1}$ on the interval $[1, \infty[$. This has most mass close to $x = 1$ and again we should choose θ close to $\pi/4$ such that $(\sin\theta - a)/\cos\theta = 1$. This implies $\sin\theta = 1/\sqrt{2} + a + o(a)$ and $\cos\theta = \{1 - (1/\sqrt{2} + a)^2\}^{1/2}\{1 + o(a)\} = 1/\sqrt{2} + o(a)$. In this case the interval for i becomes

$$n\left(\frac{a + \sin\theta}{\cos\theta}\right)^{1/q} \leq i \leq \left(\frac{-a + \sin\theta}{\cos\theta}\right)^{1/q}n = n\{1 + o(a)\},$$

and we find an upper bound of the form $F_n(a) \leq c\{(1/\sqrt{2})^{1/q} - (2a + 1/\sqrt{2})^{1/q}\}/(1/\sqrt{2})^{1/q} \rightarrow 0$, for $(a, n) \rightarrow (0, \infty)$.

In Examples 3.1, 3.2, and 3.3, the normalization is such that $n^{-1} \sum_{i=1}^n x_{ni}^2 = O(1)$. Thus for $q > -1/2$ we have

$$n^{-1} \sum_{i=1}^n x_{ni}^2 = n^{-1} \sum_{i=1}^n (i/n)^{2q} \rightarrow \int_0^1 u^{2q} du = (1 + 2q)^{-1}.$$

In these cases the regression is $y_i = \mu + \alpha(i/n)^q + \varepsilon_i$ and Theorem 2.1 proves boundedness of $(\hat{\mu}, \hat{\alpha})$.

In Theorem 2.1, Assumption 1(iii, a) to F_n is a sufficient condition for boundedness of $\hat{\beta}$. The necessity of Assumption 1(iii, a) for boundedness of $\hat{\beta}$ depends on the choice of criterion function. For a least squares criterion it is not necessary. For a Huber-skip criterion it is also not necessary. We give an example.

Example 3.4 Let ρ be the Huber-skip function and let $z_{ni} = (1, 1_{(i=n)})'$, such that

$$\sum_{i=1}^n \rho\{\varepsilon_i - \mu - \alpha 1_{(i=n)}\} = \sum_{i=1}^{n-1} \rho(\varepsilon_i - \mu) + \rho(\varepsilon_n - \mu - \alpha),$$

which shows that $\hat{\alpha}(\mu) = \varepsilon_n - \mu$. Inserting this we find the objective function for the Huber-skip location problem (with $n - 1$ observations). It follows from Theorem 2.1, that $\hat{\mu}$ is bounded, such that also $\hat{\alpha} = \varepsilon_n - \hat{\mu}$ is bounded. On the other hand we find for $0 \leq a < 1$, the function $F_n(a) = n^{-1} \sum_{i=1}^n 1_{(|z_i| \leq a)} = (n - 1)/n \rightarrow 1$, for $(a, n) \rightarrow (0, \infty)$.

We find an example where Assumption 1(iiii) is not satisfied but the condition of Davies that $\liminf_{n \rightarrow \infty} \lambda_n(\xi^*) > 0$ for some $\xi^* > 0$ is satisfied, where λ_n is defined in (3.2). Having said that, Davies considers a particular criterion function with more stringent assumptions to the innovations which leaves some room for a weaker condition than Assumption 1(iiii), see also Remark A.6.

Example 3.5 Let the $z_{ni} = (1, x_i)$ where the second coordinate takes values $1, 1/2, 1/4, 1/8, 1, 1/16, 1/32, 1/64, 1, \dots$. Equivalently, we can reorder the regressors so $z_{ni} = (1, 2^{-i})$ for $1 < i \leq 3n/4$ and $z_{ni} = (1, 1)$ for $3n/4 < i \leq n$. In this case we find $F_n(0) = \text{int}(n/4)/n \rightarrow 1/4$ and Assumption 1(iii) is satisfied, but $\lim_{(a,n) \rightarrow (0,\infty)} \sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\} = 3/4$ so Assumption 1(iiii) is not satisfied. However, $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi = 3/4$. Thus, Theorem 3.2 shows that a $\xi^* > 0$ exists so that $\liminf_{n \rightarrow \infty} \lambda_n(\xi^*) > 0$.

3.3 Regression with multiple stochastic regressors

For this case we give two examples, where in first example x_{ni} is a stationary process and in the second case x_{ni} is a random walk normalized by $n^{-1/2}$. In these cases, we give conditions on the density for Assumption 1(iii) to be satisfied. The proofs involve two rather different chaining arguments.

Theorem 3.3 (Stationary regressor) Let $z_{ni} = z_i = (1, x_i)'$ where x_i is stationary of dimension m . Let the conditional density of $\gamma'x_i$ given $\mathcal{G}_{i-1} = \sigma(x_1, \dots, x_{i-1})$ be bounded uniformly in $(x_1, \dots, x_{i-1}, x_i)$ and $|\gamma| = 1$, $\gamma \in \mathbb{R}^m$.

Then $F_n(a) = o_{\mathbb{P}}(1)$, for $(a, n) \rightarrow (0, \infty)$, such that Assumption 1(iii) holds.

Theorem 3.4 (*Random walk regressor*) Let $z_{ni} = (1, n^{-1/2}x_i)'$ where x_i is a multivariate random walk $x_i = \sum_{j=1}^i \eta_j$ and η_j are i.i.d. $(0, \Phi)$ of dimension m . Assume the density of $\gamma'x_i/(i\gamma'\Phi\gamma)^{1/2}$ is bounded uniformly in $|\gamma| = 1$ and $i = 1, \dots, n$.

Then $F_n(a) = o_{\mathbb{P}}(1)$, for $(a, n) \rightarrow (0, \infty)$, such that Assumption 1(iii) holds.

Theorems 3.3 and 3.4 involve conditions to certain conditional densities. These are satisfied in a variety of situations. We give some simple examples.

Example 3.6 *The assumption on the conditional density in Theorem 3.3 is satisfied for a stationary autoregressive process x_i with Gaussian innovations. Indeed, if $x_i = \alpha x_{i-1} + \eta_i$ with η_i i.i.d. $\mathbf{N}_m(0, \Phi)$ with positive definite variance Φ , then $\gamma'x_i$ given \mathcal{G}_{i-1} is $\mathbf{N}(\gamma'\alpha x_{i-1}, \gamma'\Phi\gamma)$. The conditional density is bounded in the mean, while the variance $\gamma'\Phi\gamma$ is finite and bounded away from zero when $|\gamma| = 1$.*

Example 3.7 *The assumption on the conditional density in Theorem 3.4 is satisfied for a random walk with normal innovations. Indeed if η_j are independent normal $\mathbf{N}_m(0, \Phi)$ with positive definite covariance Φ , then $\gamma'x_i/(i\gamma'\Phi\gamma)^{1/2}$ is $\mathbf{N}(0, 1)$ for any $|\gamma| = 1$.*

4 On the supremum of families of martingales

We will need some results bounding the supremum of a family of martingales indexed by a parameter in a compact set of \mathbb{R}^m . These results build on the following result taken from Johansen and Nielsen (2016) concerning the maximum of finitely many martingales. This result in turn is based on an iteration of the exponential martingale inequality by Bercu and Touati (2008).

Lemma 4.1 (*Johansen and Nielsen 2016, Theorem 5.2*) Let \mathcal{F}_i be an increasing sequence of σ -fields and let $u_{n\ell i}$ be \mathcal{F}_i adapted with $\mathbb{E}(u_{n\ell i}^{2r}) < \infty$, $r \in \mathbb{N}$, $\ell = 1, \dots, L$, $i = 1, \dots, n$, and let ς and λ be positive real numbers defined by

$$L = O(n^\lambda), \tag{4.1}$$

$$\max_{1 \leq q \leq r} \mathbb{E}(\max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbb{E}_{i-1} u_{n\ell i}^{2q}) = O(n^\varsigma). \tag{4.2}$$

Then, if ν is chosen such that

$$(i) : \varsigma < 2\nu, \quad (ii) : \varsigma + \lambda < \nu 2^r,$$

it holds that

$$\max_{1 \leq \ell \leq L} \left| \sum_{i=1}^n (u_{n\ell i} - \mathbb{E}_{i-1} u_{n\ell i}) \right| = o_{\mathbb{P}}(n^\nu). \tag{4.3}$$

We prove a similar result for a family of martingales with a parameter $\kappa \in \mathbb{R}^{m+1}$ which lies in the intersection of a compact subset \mathcal{K} and a ball $\mathbb{B}(\kappa_0, Bn^{-\phi})$ centered in κ_0 and with radius $Bn^{-\phi}$. This result has two cases. The first is applied in the proof of Theorem 3.3 for stationary regressors. The second will be used for asymptotic expansions in a follow-up paper.

Theorem 4.1 *Let \mathcal{F}_i be an increasing sequence of σ -fields while \mathcal{K} is a compact subset of \mathbb{R}^{m+1} . Consider a family of \mathcal{F}_i measurable random variables $u_{ni}(\kappa)$ with $\mathbf{E}|u_{ni}(\kappa)| < \infty$ for $\kappa \in \mathcal{K}$, and normalized by $u_{ni}(\kappa_0) = 0$ for some $\kappa_0 \in \mathcal{K}$. We define the martingales*

$$M_n(\kappa) = \sum_{i=1}^n \{u_{ni}(\kappa) - \mathbf{E}_{i-1}u_{ni}(\kappa)\} \quad \text{for } \kappa \in \mathcal{K}. \quad (4.4)$$

Suppose there exist a set Ω^ with probability one and \mathcal{F}_{i-1} -measurable random variables $\bar{u}_{ni}^*(\kappa)$ such that for all $0 \leq \phi \leq 1$ and $\kappa \in \mathcal{K}$,*

$$\sup_{\tilde{\kappa} \in \mathbb{B}(\kappa, B\phi) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})| \leq \phi \bar{u}_{ni}^*(\kappa). \quad (4.5)$$

Further, we can choose $B > 0$ and r such that $2^r > 3 + m$, and \mathcal{F}_{i-1} -measurable random variables $A_{ni}(\kappa)$, such that, for all $1 \leq p \leq 2^r$ and $\kappa \in \mathcal{K}$,

$$\mathbf{E}_{i-1}\{\bar{u}_{ni}^*(\kappa)\}^p \leq A_{ni}(\kappa) \quad \text{a.s.} \quad (4.6)$$

Let η, ν satisfy either of

$$\text{Case 1: } \eta = 0, \nu > 1/2, \quad \text{Case 2: } 0 < \eta < 1/2, \nu = 1/2.$$

Furthermore, there exist \mathcal{F}_{i-1} -measurable random variables A_{ni}^ such that*

$$\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa) \leq A_{ni}^* \quad (4.7)$$

and for some $C > 0$

$$n^{-1} \sum_{i=1}^n \mathbf{E}(A_{ni}^*) \leq C. \quad (4.8)$$

Then, there exist random variables $M_n^ = o_{\mathbf{P}}(1)$ such that, on the set Ω^* ,*

$$n^{-\nu} \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |M_n(\kappa)| \leq M_n^*. \quad (4.9)$$

Lemma 4.1 is also used to prove the next result concerning a special class of martingales needed in the proof of Theorem 2.1.

Theorem 4.2 *Let u_i be an \mathcal{F}_i martingale difference sequence while $z_{ni} \in \mathbb{R}^{m+1}$ is \mathcal{F}_{i-1} adapted, where $m \in \mathbb{N}$. Choose $\nu > 1/2$ and r so that $m + 1 < \nu 2^r$. Let $\mathbf{E} \sum_{i=1}^n |u_i|^{2^r} = O(n)$. Then*

$$\sup_{|\delta|=1} \left| \sum_{i=1}^n u_i 1_{(\delta' z_{ni}=0)} \right| = o_{\mathbf{P}}(n^\nu).$$

Note that in Theorem 4.1, Assumption (4.6) implies that for $\phi > 0$, $\mathbf{E}_{i-1}|u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})|^p$ is smooth in $(\kappa, \tilde{\kappa})$, whereas in Theorem 4.2 we find

$$\mathbf{E}_{i-1}|u_i 1_{(\delta' z_{ni}=0)} - u_i 1_{(\tilde{\delta}' z_{ni}=0)}|^p = |1_{(\delta' z_{ni}=0)} - 1_{(\tilde{\delta}' z_{ni}=0)}|^p \mathbf{E}|u_i|^p,$$

which is not smooth in $(\delta, \tilde{\delta})$. The analysis in Theorem 4.2 of this situation is made possible by the very explicit dependence on δ , which is analyzed in Lemma A.1.

5 Conclusion and discussion

We have investigated boundedness for M-estimators for the multiple regression model with stochastic regressors and unrestricted parameters. The leading case of a robust M-estimator is the Huber-skip proposed by Huber in (1964). As an assumption for the main result on boundedness (Theorem 2.1) we introduced a condition on the frequency of small regressors to show that the objective function is uniformly bounded away from zero for large parameter values. This applies for random regressors. It is weaker than the condition given by Chen and Wu (1988) for deterministic regressors. It is related to the condition of Davies (1990) for S-estimators with deterministic regressors. This condition is not so easy to check in specific examples, but it is verified for some deterministic regressors and stochastic regressors that are either stationary or random walks.

A Appendix

We have here collected all the proofs of the results in the previous sections.

Proof of Theorem 3.1. With $|\theta| < \pi/2$ we have $\cos \theta > 0$. Writing $\tan \theta = y$ and $(-\sin \theta, \cos \theta) = (-y, 1)/\sqrt{1+y^2}$, noting $0 \leq c \leq 1/2$, the inequality is therefore equivalent to

$$|x - y| \leq c\sqrt{1+y^2} \leq 1/2\sqrt{1+y^2} \leq 1/2(1+|y|),$$

using $\sqrt{1+y^2} \leq 1+|y|$. Further, using first the triangle inequality and then the above inequalities shows

$$1+|y| \leq 1+|x|+|y-x| \leq 1+|x|+1/2(1+|y|),$$

so that $(1+|y|) \leq 2(1+|x|)$, and hence $1/\cos \theta = \sqrt{1+y^2}$ as bounded by first $1+|y|$ and then $2(1+|x|)$. ■

A.1 Proof of boundedness

Proof of Theorem 2.1. We start by rewriting the criterion function and the objective function. We then give an overview of what we have to prove. Then the main algebraic arguments follow. The arguments are then wrapped up in a probabilistic framework.

(a) *Criterion function.* We rewrite the criterion function as

$$\rho(y_i - z'_{ni}\beta) = \rho\{\varepsilon_i - (\mu - \mu_0) - x'_{ni}(\alpha - \alpha_0)\},$$

recalling the model equation $y_i = \beta'_0 z_{ni} + \varepsilon_i$. Assumption 1(iib) shows that $\mu_*, u_* \in \mathbb{R}$ exists so that $h_* = h(\mu_*) = \mathbb{E}\rho(\varepsilon_i - \mu_*) < \rho_* = \inf_{|u| \geq |u_*|} \rho(u)$. For that μ_* we define

$$\beta_* = (\mu_0 + \mu_*, \alpha'_0)',$$

so that the criterion function satisfies

$$\rho(y_i - z'_{ni}\beta) = \rho\{\varepsilon_i - z'_{ni}(\beta - \beta_*) - \mu_*\}.$$

(b) *Objective function.* Minimizing for $R_n(\beta)$ is equivalent to minimizing

$$S_n(\beta, \beta_*) = R_n(\beta) - R_n(\beta_*) = n^{-1} \sum_{i=1}^n [\rho\{\varepsilon_i - z'_{ni}(\beta - \beta_*) - \mu_*\} - \rho(\varepsilon_i - \mu_*)].$$

The summands, for which $z'_{ni}(\beta - \beta_*) = 0$, do not contribute to the sums. Thus, we can write $R_n(\beta) - R_n(\beta_*) = S_n(\beta, \beta_*)$ where

$$S_n(\beta, \beta_*) = n^{-1} \sum_{i=1}^n [\rho\{\varepsilon_i - z'_{ni}(\beta - \beta_*) - \mu_*\} - \rho(\varepsilon_i - \mu_*)] 1_{\{z'_{ni}(\beta - \beta_*) \neq 0\}}.$$

This function is zero when $\beta = \beta_*$ and we need to find a lower bound when $|\beta - \beta_*|$ is bounded away from zero. For such β we introduce polar coordinates with direction $\delta = (\beta - \beta_*)/|\beta - \beta_*|$ and length $\lambda = |\beta - \beta_*|$ so that $\beta - \beta_* = \lambda\delta$ and $|\delta| = 1$ while $z'_{ni}(\beta - \beta_*) = \lambda z'_{ni}\delta$. We can then write $S_n(\beta, \beta_*) = S_n(\delta, \lambda)$ and decompose $S_n = S_n^{(1)} - S_n^{(2)}$ where

$$S_n^{(1)}(\delta, \lambda) = n^{-1} \sum_{i=1}^n \rho(\varepsilon_i - \lambda z'_{ni}\delta - \mu_*) 1_{(z'_{ni}\delta \neq 0)}, \quad S_n^{(2)}(\delta) = n^{-1} \sum_{i=1}^n \rho(\varepsilon_i - \mu_*) 1_{(z'_{ni}\delta \neq 0)}.$$

The main idea of the proof is to find an upper bound for $S_n^{(1)}$ and a lower bound for $S_n^{(2)}$ so that $S_n^{(1)} - S_n^{(2)}$ is bounded away from zero.

(c) *Lower bound for ρ under constraints to $\varepsilon_i, z'_{ni}\delta$ and λ .* Assumption 1(iib) requires existence of μ_*, u_* so that $0 < \mathbf{E}\{\rho(\varepsilon_i - \mu_*)\} < \rho_* = \inf_{|u| \geq |u_*|} \rho(u)$. Choose A_0, a_0 and consider $|\varepsilon_i| \leq A_0$ and $|z'_{ni}\delta| \geq a_0$, and $\lambda \geq B_0$ where $B_0 = (A_0 + |u_*| + |\mu_*|)/a_0$. Then

$$|\varepsilon_i - \lambda z'_{ni}\delta - \mu_*| \geq \lambda |z'_{ni}\delta| - |\varepsilon_i| - |\mu_*| \geq \lambda a_0 - A_0 - |\mu_*| \geq |u_*|.$$

Hence $\rho(\varepsilon_i - \lambda z'_{ni}\delta - \mu_*) \geq \rho_*$ for $|z'_{ni}\delta| \geq a_0$ and large λ .

(d) *Lower bound for $S_n^{(1)}$ for large λ uniformly in δ .* Choose A_0, a_0 and delete summands of $S_n^{(1)}$ for which $|\varepsilon_i| \geq A_0$ or $|z'_{ni}\delta| \leq a_0$ so that

$$S_n^{(1)}(\delta, \lambda) \geq n^{-1} \sum_{i=1}^n \rho(\varepsilon_i - \lambda z'_{ni}\delta - \mu_*) 1_{(z'_{ni}\delta \neq 0)} 1_{(|\varepsilon_i| < A_0)} 1_{(|\delta' z_{ni}| > a_0)},$$

which is valid for any $\beta \neq \beta_*$. Now, for large $\lambda \geq B_0$ we can apply item (c) to get the further bound

$$S_n^{(1)}(\delta, \lambda) \geq \rho_* n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| < A_0)} 1_{(|\delta' z_{ni}| > a_0)}.$$

Use that for sets \mathbb{A} and \mathbb{B} , $1_{\mathbb{A} \cap \mathbb{B}} \geq 1 - 1_{\mathbb{A}^c} - 1_{\mathbb{B}^c}$ so that

$$S_n^{(1)}(\delta, \lambda) \geq \rho_* \left\{ 1 - n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| < A_0)} - n^{-1} \sum_{i=1}^n 1_{(|\delta' z_{ni}| > a_0)} \right\} = \rho_* \left\{ 1 - n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A_0)} - F_{n\delta}(a_0) \right\},$$

for any $a_0, A_0 > 0$ and $\lambda \geq B_0 = (A_0 + |u_*| + |\mu_*|)/a_0$.

(e) *Upper bound for $S_n^{(2)}$.* Let $h_* = \mathbf{E}\rho(\varepsilon_i - \mu_*)$ and introduce martingale differences $m_i = \rho(\varepsilon_i - \mu_*) - h_*$ and the family of martingales $M_n(\delta) = n^{-1} \sum_{i=1}^n \{\rho(\varepsilon_i - \mu_*) - h_*\} 1_{(|z'_{ni}\delta| > 0)}$. We then find the martingale decomposition

$$S_n^{(2)}(\delta) = M_n(\delta) + h_* n^{-1} \sum_{i=1}^n 1_{(|z'_{ni}\delta| > 0)} = M_n(\delta) + h_* \{1 - F_{n\delta}(0)\}.$$

Now evaluate $M_n(\delta)$ by $\sup_{|\delta|=1} |M_n(\delta)|$, to get

$$S_n^{(2)}(\delta) \leq h_* \{1 - F_{n\delta}(0)\} + \sup_{|\delta|=1} |M_n(\delta)|.$$

(f) *Lower bound for S_n .* Combine (d) and (e) to get for any $a_0, A_0 > 0$ and $\lambda \geq B_0$ that

$$\begin{aligned} S_n(\delta, \lambda) &\geq \rho_* \{1 - n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A_0)} - F_{n\delta}(a_0)\} - h_* \{1 - F_{n\delta}(0)\} - \sup_{|\delta|=1} |M_n(\delta)| \\ &= F_{n\delta}^*(a_0) - \rho_* n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A_0)} - \sup_{|\delta|=1} |M_n(\delta)|, \end{aligned}$$

with $F_{n\delta}^*(a_0) = \rho_* \{1 - F_{n\delta}(a_0)\} - h_* \{1 - F_{n\delta}(0)\}$. Taking infimum over δ gives

$$\inf_{|\delta|=1} S_n(\delta, \lambda) \geq \inf_{|\delta|=1} F_{n\delta}^*(a_0) - \rho_* n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A_0)} - \sup_{|\delta|=1} |M_n(\delta)|. \quad (\text{A.1})$$

The function $F_{n\delta}^*$ involves both the criterion function through ρ_* , h_* and the regressors through the $F_{n\delta}$ function. We would like to disentangle these. Thus rewrite

$$F_{n\delta}^*(a_0) = (\rho_* - h_*) \{1 - F_{n\delta}(0)\} - \rho_* \{F_{n\delta}(a_0) - F_{n\delta}(0)\},$$

so that when taking infimum and recalling $F_n(0) = \sup_{|\delta|=1} F_{n\delta}(0)$ we get

$$\inf_{|\delta|=1} F_{n\delta}^*(a) \geq (\rho_* - h_*) \{1 - F_n(0)\} - \rho_* \sup_{|\delta|=1} \{F_{n\delta}(a_0) - F_{n\delta}(0)\}.$$

Thus, we will proceed with the lower bound

$$\begin{aligned} \inf_{|\delta|=1} S_n(\delta, \lambda) &\geq (\rho_* - h_*) \{1 - F_n(0)\} - \rho_* \sup_{|\delta|=1} \{F_{n\delta}(a_0) - F_{n\delta}(0)\} \\ &\quad - \rho_* n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A_0)} - \sup_{|\delta|=1} |M_n(\delta)| \end{aligned}$$

(g) *Probability argument.* We construct large probability sets \mathbb{C}_n as follows. Write

$$M_n(\delta) = n^{-1} \sum_{i=1}^n m_i 1_{(|z'_{ni} \delta| > 0)} = n^{-1} \sum_{i=1}^n m_i - n^{-1} \sum_{i=1}^n m_i 1_{(|z'_{ni} \delta| = 0)}.$$

Assumption 1(i, iic) implies that the first term is $\text{op}(1)$ by the Law of Large Numbers for martingales, and the second term is $\text{op}(1)$ uniformly in δ by Theorem 4.2 used with $\nu = 1$ and Assumption 1(i, iic). Thus $\sup_{|\delta|=1} |M_n(\delta)|$ is small.

The Law of Large Numbers implies that $n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \geq A)} \rightarrow \mathbf{P}(|\varepsilon_1| \geq A)$, which is small for large A_0 .

Assumption 1(iii) states $\mathbf{P}[\sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\} \geq \epsilon] \rightarrow 0$ while, for some $\xi < 1$, $\mathbf{P}\{F_n(0) \geq \xi\} \rightarrow 0$.

Collecting these results we find that for all $\epsilon, \eta > 0$ there exists $A_0, a_0, n_0 > 0$ such that for all $n \geq n_0$ the sets \mathbb{C}_n defined by the inequalities

$$\sup_{|\delta|=1} |M_n(\delta)| \leq \eta, \quad (\text{A.2})$$

$$n^{-1} \sum_{i=1}^n \mathbf{1}_{(|\varepsilon_i| \geq A_0)} \leq \eta, \quad (\text{A.3})$$

$$F_n(0) = \sup_{|\delta|=1} F_{n\delta}(0) \leq \xi, \quad (\text{A.4})$$

$$\sup_{|\delta|=1} \{F_{n\delta}(a_0) - F_{n\delta}(0)\} \leq \eta, \quad (\text{A.5})$$

have probability $\mathbb{P}(\mathbb{C}_n) \geq 1 - \epsilon$.

(h) *Lower bound for S_n on \mathbb{C}_n .* Apply the constraints defining \mathbb{C}_n to the lower bound for S_n in item (f) to get the bound

$$S_n(\beta, \beta_*) \geq (\rho_* - h_*)(1 - \xi) - \rho_*\eta - \rho_*\eta - \eta,$$

for $\lambda > B_0$ and uniformly in δ . Since $\rho_* > h_* \geq 0$ and $\xi < 1$ this lower bound is positive for small η .

(i) *Existence of minimizer.* The objective function is lower semi-continuous on the compact $(|\beta - \beta_*| \leq B_0)$ by Assumption 1(iiia), and therefore attains its minimum, and any minimizer is in the set $(|\beta| \leq B)$ for $B = B_0 + |\beta_*|$. ■

Remark A.6 *The point of Theorem 2.1 is to find conditions for boundedness that apply separately to criterion function and innovations on the one hand and to the regressors on the other hand. The proof shows Assumption 1(iii) are sufficient conditions for arguing that*

$$F_{n\delta}^*(a) = \rho_*\{1 - F_{n\delta}(a)\} - h_*\{1 - F_{n\delta}(0)\}. \quad (\text{A.6})$$

is bounded away from zero uniformly. Foregoing that aim it would of course be possible to replace Assumption 1(iii) with the weaker condition

(iii') Suppose a $0 < \xi^$ exists such that $\lim_{(a,n) \rightarrow (0,\infty)} \mathbb{P}\{\inf_{|\delta|=1} F_{n\delta}^*(a_0) < \xi^*\} = 0$.*

This may be useful for particular combinations of criterion functions, innovations and regressors. Indeed, Example 3.5 gives some regressors violating Assumption 1(iii, a). However, condition (iii') can be demonstrated for instance for the case of Huber-skip with cut-off $c > 2$ and standard normal innovations.

Proof of Theorem 2.2. If further ρ is continuous we can apply the argument of Jennrich (1969) and construct a measurable minimizer, $\hat{\beta}$, with value in the compact set $(|\beta| \leq B)$, such that $\hat{\beta}$ is bounded. ■

A.2 Proof of martingale results

Proof of Theorem 4.1. We study the martingale $M_n(\kappa) = \sum_{i=1}^n \{u_{ni}(\kappa) - \mathbb{E}_{i-1} u_{ni}(\kappa)\}$, see (4.4), on sets of the form $\mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}$. We prove (4.9) in the situations

$$\text{Case 1 : } \eta = 0, \nu > 1/2, \quad \text{Case 2 : } 0 < \eta < 1/2, \nu = 1/2.$$

(a) *Chaining argument.* With the assumption $2^r > 2 + m + 1$ we can choose ζ such that

$$1/2 < \zeta < (2^{r-1} - 1)/(m + 1). \quad (\text{A.7})$$

For $0 \leq \eta < 1/2$, cover $\mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}$ by $L = O\{n^{(\zeta-\eta)(m+1)}\}$ balls $\mathbb{B}(\kappa_\ell, n^{-\zeta})$ with radius $n^{-\zeta}$ and centers $\kappa_\ell \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}$ for $\ell = 1, \dots, L$. Note that $L = O\{n^{(\zeta-\eta)(m+1)}\} \rightarrow \infty$. For all $\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}$, we choose κ_ℓ such that $\mathbb{B}(\kappa_\ell, Bn^{-\zeta})$ covers κ . Use chaining to get

$$\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |M_n(\kappa)| \leq \max_{1 \leq \ell \leq L} |M_n(\kappa_\ell)| + \max_{1 \leq \ell \leq L} \sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |M_n(\kappa) - M_n(\kappa_\ell)| = \mathcal{R}_{n1} + \mathcal{R}_{n2}.$$

The variable \mathcal{R}_{n2} may be measurable, but we will find measurable martingale bounds on the set Ω^* . On Ω^* we apply the bound (4.5) with $\phi = n^{-\zeta}$ and get

$$\sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\kappa_\ell)| \leq n^{-\zeta} \check{u}_{ni}^*(\kappa).$$

Applying the triangle inequality to get, on Ω^* ,

$$\sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |M_n(\kappa) - M_n(\kappa_\ell)| \leq \sum_{i=1}^n \sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\kappa_\ell)| \leq \sum_{i=1}^n n^{-\zeta} \check{u}_{ni}^*(\kappa),$$

which is measurable. In order to make a martingale decomposition we define

$$\widetilde{\mathcal{M}}_{n2\ell} = \sum_{i=1}^n \{\check{u}_{ni}^*(\kappa) - \mathbf{E}_{i-1} \check{u}_{ni}^*(\kappa)\}, \quad \overline{\mathcal{M}}_{n2\ell} = \sum_{i=1}^n \mathbf{E}_{i-1} \check{u}_{ni}^*(\kappa).$$

Then the inequality $|X - \mathbf{E}X| \leq (|X| - \mathbf{E}|X|) + 2\mathbf{E}|X|$ implies the bound $\mathcal{R}_{n2} \leq n^{-\zeta} \max_{1 \leq \ell \leq L} (\widetilde{\mathcal{M}}_{n2\ell} + 2\overline{\mathcal{M}}_{n2\ell})$. We then prove (4.9) by applying Lemma 4.1 to the martingales $M_n(\kappa_\ell)$ and $\widetilde{\mathcal{M}}_{n2\ell}$ while bounding $\overline{\mathcal{M}}_{n2\ell}$.

(b) *The term* $\max_{1 \leq \ell \leq L} |M_n(\kappa_\ell)| = \mathcal{R}_{n1}$. We apply Lemma 4.1 to the martingale $M_n(\kappa_\ell)$ defining $u_{n\ell i} = n^{-\eta} \check{u}_{ni}^*(\kappa_\ell)$. Let $\lambda = (\zeta - \eta)(m + 1)$ in (4.1). We argue that $\varsigma = 1 - \eta$ in (4.2). We start by finding a bound for $u_{ni}(\kappa_\ell)$ that is uniform in ℓ . On Ω^* we apply the bound (4.5) with $\phi = n^{-\zeta}$ noting $u_{ni}(\kappa_0) = 0$ to get

$$\sup_{\kappa_\ell \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa_\ell)| = \sup_{\kappa_\ell \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa_\ell) - u_{ni}(\kappa_0)| \leq n^{-\eta} \check{u}_{ni}^*(\kappa_0),$$

which is measurable. We note that this bound is uniform in ℓ . From (4.6) we find, for $1 \leq p \leq 2^r$,

$$\mathbf{E}_{i-1} \{\check{u}_{ni}^*(\kappa_0)\}^p \leq A_{ni}(\kappa_0).$$

By (4.7) then $A_{ni}(\kappa_0) \leq A_{ni}^*$ on Ω^* so that $\sum_{i=1}^n \mathbf{E} A_{ni}(\kappa_0) = O(n)$ by (4.8). Recalling $u_{n\ell i} = n^{-\eta} \check{u}_{ni}^*(\kappa_\ell)$ and $\varsigma = 1 - \eta$ we get, see (4.2),

$$\max_{1 \leq p \leq 2^r} \mathbf{E} \left(\max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} |u_{n\ell i}|^p \right) \leq n^{-\eta} \sum_{i=1}^n \mathbf{E} \{A_{ni}(\kappa_0)\} \leq C n^{1-\eta} = O(n^\varsigma).$$

We then check the conditions of Lemma 4.1:

$$(i) : 0 < \varsigma = 1 - \eta < 2\nu, \quad (ii) : \varsigma + \lambda = 1 - \eta + (\zeta - \eta)(m + 1) < \nu 2^r.$$

Condition (i) is satisfied in Case 1 since $1 - \eta = 1$ and $2\nu > 1$ and in Case 2 since $1 - \eta < 1$ and $2\nu = 1$. Condition (ii) is satisfied in Case 1 and 2 by the choice of ζ in (A.7), because for $0 \leq \eta$ and $\nu \geq 1/2$

$$\zeta + \lambda = 1 - \eta + (\zeta - \eta)(m + 1) \leq 1 + \zeta(m + 1) < 1 + \frac{2^{r-1} - 1}{m + 1}(m + 1) = 2^{r-1} \leq \nu 2^r.$$

Applying (4.3) of Lemma 4.1, we get $\max_{1 \leq \ell \leq L} |M_n(\kappa_\ell)| = \mathcal{R}_{n1} = o_{\mathbb{P}}(n^\nu)$ in both cases.

(c) The term $\max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell}$. Use (4.6) for $\phi = n^{-\zeta}$ to get

$$\begin{aligned} & \mathbf{E}_{i-1} \check{u}_{ni}^*(\kappa) \\ & \mathbf{E}_{i-1} k_{ni}^p(\kappa_\ell, n^{-\zeta}) \leq n^{-\zeta} A_{ni}(\kappa_\ell) \leq n^{-\zeta} \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa), \end{aligned} \quad (\text{A.8})$$

uniformly in κ_ℓ and $1 \leq p \leq 2^r$. We then find from (4.8) that

$$\mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} k_{ni}^p(\kappa_\ell, n^{-\zeta}) \leq n^{-\zeta} \sum_{i=1}^n \mathbf{E} \left\{ \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa) \right\} = O(n^{1-\zeta}) = o(n^\nu), \quad (\text{A.9})$$

since $\nu \geq 1/2 > 1 - \zeta$ by (A.7). In particular for $p = 1$ we find $\mathbf{E} \max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell} = o(n^\nu)$ so that, by the Chebychev inequality, $\max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell} = o_{\mathbb{P}}(n^\nu)$.

(d) The term $\max_{1 \leq \ell \leq L} \widetilde{\mathcal{M}}_{n2\ell}$. We apply Lemma 4.1 to the martingale $\widetilde{\mathcal{M}}_{n2\ell}$ using $u_{n\ell i} = k_{ni}(\kappa_\ell, Bn^{-\zeta})$. Due to (A.9) we can choose $\lambda = (\zeta - \eta)(m + 1)$ and $\varsigma = 1 - \zeta$. Noting that $\zeta > 1/2 > \eta$ then $\varsigma = 1 - \zeta \leq 1 - \eta$, which was the value of ς chosen in item (b). We can then proceed as in (b) to get $\max_{1 \leq \ell \leq L} \widetilde{\mathcal{M}}_{n2\ell} = o_{\mathbb{P}}(n^\nu)$. ■

For the proof of Theorem 4.2 we need the following Lemma. For any $0 \neq v \in \mathbb{R}^{m+1}$ we define v_\perp as an $(m + 1) \times m$ matrix of rank m for which $v'v_\perp = 0$.

Lemma A.1 For $i = 1, \dots, n$, let u_i be random variables while z_i are random vectors in \mathbb{R}^{m+1} . Let

$$S_n(\delta) = \sum_{i=1}^n u_i 1_{(\delta' z_i = 0)}.$$

For $m = 0$ we define $M_n = \sum_{i=1}^n u_i 1_{(z_i = 0)}$ and find $S_n(\delta) = M_n$.

For $m = 1$ we define for $1 \leq \ell_1 \leq n + 1$

$$M_{n\ell_1} = \sum_{i=1}^{\ell_1-1} u_i 1_{(z_i=0)} + \sum_{i=\ell_1}^n u_i 1_{(z_{\ell_1} \neq 0, (z_{\ell_1})'_\perp z_i = 0)}, \quad (\text{A.10})$$

and find

$$\sup_{|\delta|=1} |S_n(\delta)| \leq \max_{1 \leq \ell_1 \leq n+1} |M_{n\ell_1}|. \quad (\text{A.11})$$

For $m > 1$ we define for $1 \leq \ell_1 \leq \dots \leq \ell_m \leq n + 1$ the martingales

$$M_{n, \ell_1, \ell_2, \dots, \ell_m} = \sum_{k=0}^m \sum_{i=\ell_k}^{\ell_{k+1}-1} u_i 1_{\{z_{\ell_1}^{(0)} \neq 0, \dots, z_{\ell_k}^{(k-1)} \neq 0, z_i^{(k)} = 0\}}, \quad (\text{A.12})$$

where we write $\ell_0 = 1$ and $\ell_{m+1} = n + 1$. Here $z_i^{(0)} = z_i$, and for $k = 1, \dots, m$ we define recursively for $z_{\ell_k}^{(k-1)} \neq 0$, $z_i^{(k)} = (z_{\ell_k}^{(k-1)})'_\perp z_i^{(k-1)}$ for $i = \ell_k, \dots, \ell_{k+1} - 1$. Then

$$\max_{|\delta|=1} |S_n(\delta)| \leq \max_{1 \leq \ell_1 \leq \dots \leq \ell_m \leq n+1} |M_{n\ell_1, \dots, \ell_m}|. \quad (\text{A.13})$$

Proof of Lemma A.1. (a) *The case $m = 0$.* If $z_i \in \mathbb{R}$, then $\delta = \pm 1$, and $\delta' z_i = 0$ if and only if $z_i = 0$, such that $S_n(\delta) = M_n$.

(b) *The case $m = 1$.* In order to find an expression for $S_n(\delta)$, we define the stopping time s_1^δ with respect to the filtration generated by z_i , as the first index i for which $\delta' z_i = 0$ and $z_i \neq 0$. If no such i exists, we let $s_1^\delta = n + 1$. Consider first the case $i < s_1^\delta$, which includes the situation where $s_1^\delta = n + 1$. In this case either $\delta' z_i \neq 0$ or $z_i = 0$. If $\delta' z_i \neq 0$, then $1_{(\delta' z_i=0)} = 1_{(z_i=0)} = 0$ and if $z_i = 0$ then $1_{(\delta' z_i=0)} = 1_{(z_i=0)} = 1$, such that $\sum_{i=1}^{s_1^\delta-1} u_i 1_{(\delta' z_i=0)} = \sum_{i=1}^{s_1^\delta-1} u_i 1_{(z_i=0)}$. Consider now the case $i = s_1^\delta$ so that $s_1^\delta \leq n$. In this case we find $z_{s_1^\delta} \neq 0$ and $\delta' z_{s_1^\delta} = 0$, and because $m = 1$, $\delta \in \mathbb{R}^2$ is proportional to $(z_{s_1^\delta})'_\perp$, and therefore $\delta' z_i \propto (z_{s_1^\delta})'_\perp z_i$ for $i \geq s_1^\delta$. Thus $\sum_{i=s_1^\delta}^n u_i 1_{(\delta' z_i=0)} = \sum_{i=s_1^\delta}^n u_i 1_{\{z_{s_1^\delta} \neq 0, (z_{s_1^\delta})'_\perp z_i = 0\}}$, and

$$S_n(\delta) = \sum_{i=1}^n u_i 1_{(\delta' z_i=0)} = \sum_{i=1}^{s_1^\delta-1} u_i 1_{(z_i=0)} + \sum_{i=s_1^\delta}^n u_i 1_{\{z_{s_1^\delta} \neq 0, (z_{s_1^\delta})'_\perp z_i = 0\}}. \quad (\text{A.14})$$

We now compare with the martingales $M_{n\ell_1}$, see (A.10). For all sample paths and all δ we have the evaluation

$$|S_n(\delta)| = \left| \sum_{i=1}^{s_1^\delta-1} u_i 1_{(z_i=0)} + \sum_{i=s_1^\delta}^n u_i 1_{\{z_{s_1^\delta} \neq 0, (z_{s_1^\delta})'_\perp z_i = 0\}} \right| \leq \max_{1 \leq \ell_1 \leq n+1} |M_{n\ell_1}|,$$

and hence (A.11) holds for $m = 1$.

(c) *The case $m > 1$.* We use the stopping time s_1^δ , and define the regressors $z_i^{(0)} = z_i$, $i < s_1^\delta$ and $z_i^{(1)} = \{z_{s_1^\delta}^{(0)}\}'_\perp z_i^{(0)}$, $s_1^\delta \leq i$. Then $z_{s_1^\delta}^{(0)} \neq 0$ and $\delta' z_{s_1^\delta}^{(0)} = 0$, so that $\delta = \delta_0 = \{z_{s_1^\delta}^{(0)}\}'_\perp \delta_1$ for some $\delta_1 \in \mathbb{R}^{m-1}$ and hence $\delta' z_i^{(0)} = \delta_1' \{z_{s_1^\delta}^{(0)}\}'_\perp z_i^{(0)} = \delta_1' z_i^{(1)}$. By the argument leading to (A.14), we then find

$$\sum_{i=1}^n u_i 1_{\{\delta' z_i^{(0)}=0\}} = \sum_{i=1}^{s_1^\delta-1} u_i 1_{\{z_i^{(0)}=0\}} + \sum_{i=s_1^\delta}^n u_i 1_{\{z_{s_1^\delta}^{(0)} \neq 0, \delta_1' z_i^{(1)}=0\}}. \quad (\text{A.15})$$

Then we have expressed the sum from 1 to n involving an $m + 1$ dimensional parameter δ and regressors $z_i^{(0)}$ in terms of a sum from s_1^δ to n involving an m dimensional parameter δ_1 and regressors $z_i^{(1)}$. Therefore we define recursively stopping times s_k^δ , parameters $\delta_k \in \mathbb{R}^{m+1-k}$, and regressors $z_i^{(k)}$, for $k = 2, \dots, m$

$$\begin{aligned} s_k^\delta &= \min\{i : \delta_{k-1}' z_i^{(k-1)} = 0, z_i^{(k-1)} \neq 0\}, \\ \delta_{k-1} &= \{z_{s_k^\delta}^{(k-1)}\}'_\perp \delta_k, \\ z_i^{(k)} &= \{z_{s_k^\delta}^{(k-1)}\}'_\perp z_i^{(k-1)}, i = s_k^\delta, \dots, n. \end{aligned}$$

By repeated application of (A.15) we then find, using the notation $s_0^\delta = 1$ and $s_{m+1}^\delta = n + 1$,

$$\sum_{i=1}^n u_i 1_{\{\delta' z_i^{(0)} = 0\}} = \sum_{k=0}^m \sum_{i=s_k^\delta}^{s_{k+1}^\delta - 1} u_i 1_{\{z_{s_1^\delta}^{(0)} \neq 0, \dots, z_{s_k^\delta}^{(k-1)} \neq 0, z_i^{(k)} = 0\}},$$

and the inequality, using the martingales defined in (A.12),

$$\left| \sum_{i=1}^n u_i 1_{\{\delta' z_i^{(0)} = 0\}} \right| \leq \max_{1 \leq \ell_1 \leq \dots \leq \ell_m \leq n+1} |M_{n, \ell_1, \ell_2, \dots, \ell_m}|,$$

which proves (A.13). ■

Proof of Theorem 4.2. We apply Lemma A.1 for $m + 1 = \dim z$ to see that we must evaluate the maximum of $|M_{n, \ell_1, \ell_2, \dots, \ell_m}|$, see (A.12), where the summands

$$u_{n\ell_1, \dots, \ell_m, i} = u_i 1_{\{z_{n\ell_1}^{(0)} \neq 0, \dots, z_{n\ell_k}^{(k-1)} \neq 0, z_{ni}^{(k)} = 0\}}, \quad \ell_k \leq i < \ell_{k+1},$$

are now martingale difference sequences. We apply Lemma 4.1. The number of martingales is $L = O(n^m)$ and we choose $\lambda = m \in \mathbb{N}$, see (4.1). We find

$$\max_{1 \leq q \leq r} \mathbb{E} \left(\max_{1 \leq \ell_1 < \dots < \ell_m \leq n} \sum_{i=1}^n \mathbb{E}_{i-1} u_{n\ell_1, \dots, \ell_m, i}^{2^q} \right) \leq \max_{1 \leq q \leq r} \sum_{i=1}^n \mathbb{E} u_i^{2^q} \leq \sum_{i=1}^n \mathbb{E} (1 + u_i^{2^r}) = O(n),$$

such that $\varsigma = 1$, see (4.2). We then apply Lemma 4.1 and find for $\nu > 1/2$, that $1 = \varsigma < 2\nu$ and $m + 1 = \varsigma + \lambda < \nu 2^r$, so that $\max_{1 \leq \ell_1 \leq \dots \leq \ell_m \leq n+1} |M_{n, \ell_1, \ell_2, \dots, \ell_m}| = o_{\mathbb{P}}(n^\nu)$. ■

A.3 Proof of results regarding the frequency of small regressors

We prove Theorem 3.1 which relates the condition of small regressors in Assumption 1(iii) to the condition of Davies (1990) for deterministic regressors. Finally we show in Theorems 3.4 and 3.3, that the condition for small regressors is satisfied for random walk and stationary regressors.

Proof of Theorem 3.1. We use the notation $D_n = \{0, 1, \dots, n\}$ and $D_{nn} = D/n$.

(i) The function of (δ, a) given by $F_{n\delta}(a) = n^{-1} \sum_{i=1}^n 1_{(|\delta' z_{ni}| \leq a)}$ has values in the finite set D_{nn} and is nondecreasing in a . Therefore $F_n(a) = \sup_{|\delta|=1} F_{n\delta}(a)$ is a nondecreasing function of a with values in D_{nn} . As a consequence $\sup_{|\delta|=1} F_{n\delta}(a) = k/n$, if and only if there is a δ_a such that $F_{n\delta_a}(a) = k/n$ and $F_{\delta_n}(a) \leq k/n$ for all $|\delta| = 1$.

To prove that $F_n(a)$ is right continuous at a_0 , we take a sequence $a_m \downarrow a_0$ and want to show that $F_n(a_m) \rightarrow F_n(a_0)$. For each a_m we can find δ_m such that

$$F_n(a_m) = F_{n\delta_m}(a_m) = n^{-1} \sum_{i=1}^n 1_{(|z'_{ni} \delta_m| \leq a_m)} = k_m/n,$$

say. Thus, there is a k_m -subset $S_m \subseteq D_n$, for which

$$|z'_{ni} \delta_m| \leq a_m, \quad i \in S_m \quad \text{and} \quad |z'_{ni} \delta_m| > a_m, \quad i \notin S_m. \quad (\text{A.16})$$

The sequence k_m is nonincreasing and contained in the finite set D_n , such that we can find a subsequence m' for which $k_{m'} = k$ for some $k \in D_n$. There is only a finite number of k -subsets, so we can find a further subsequence, m'' , for which $S_{m''} = S$ for some k -subset S . By compactness of $|\delta| = 1$, we can finally find a further subsequence, m''' such that $\delta_{m'''} \rightarrow \delta_0$ for some δ_0 . Then it follows that

$$F_n(a_0) \leq F_n(a_{m'''}) = F_{n\delta_{m'''}}(a_{m'''}) = k_{m'''} / n = k/n.$$

To prove the opposite inequality, we find from (A.16) and $m''' \rightarrow \infty$, that

$$|z'_{ni}\delta_0| \leq a_0, i \in S \text{ and } |z'_{ni}\delta_0| \geq a_0, i \notin S.$$

Therefore

$$F_n(a_0) \geq F_{n\delta_0}(a_0) = n^{-1} \sum_{i=1}^n \mathbf{1}_{(|z'_{ni}\delta_0| \leq a_0)} = k/n.$$

Thus $F_n(a_0) = k/n$, and $F_n(a_{m'''}) \rightarrow F_n(a_0)$.

(ii) It follows from the definition that λ_n is nonincreasing and that $\lambda_n(\xi) = \lambda_n(k/n)$ for $k/n \leq \xi < (k+1)/n$, so that λ_n right continuous with left limits.

To prove (3.3) we first choose a such that $\lambda_n(k/n) \leq a$. Then there is a k -set S and a δ with $|\delta| = 1$, such that $\max_{i \in S} |z'_{ni}\delta| \leq a$. This implies that

$$nF_n(a) = \sum_{i=1}^n \mathbf{1}_{(|z'_{ni}\delta| \leq a)} \geq \sum_{i \in S} \mathbf{1}_{(|z'_{ni}\delta| \leq a)} = k/n.$$

Next choose a such that $F_n(a) \geq k/n$, and choose δ with $|\delta| = 1$, such that

$$nF_n(a) = nF_{n\delta}(a) = \sum_{i=1}^n \mathbf{1}_{(|z'_{ni}\delta| \leq a)} \geq k.$$

The last inequality shows that for this δ there exists a k -set S such that $\max_{i \in S} |z'_{ni}\delta| \leq a$. Therefore $\lambda_n(k/n) = \min_{|S|=k} \min_{|\delta|} \max_{i \in S} |z'_{ni}\delta| \leq a$, which proves (3.3).

(iii) Define the function $q_n(\xi) = \inf\{a : F_n(a) \geq \text{int}(n\xi)/n\}$. Let ξ satisfy $k \leq n\xi < k+1$ so that $q_n(\xi) = q_n(k/n)$ while $\lambda_n(\xi) = \lambda_n(k/n)$ due to the right continuity established in (ii).

We show $\lambda_n(\xi) \leq q_n(\xi)$. If we take a sequence $a_m \downarrow q_n(\xi)$ such that $F_n(a_m) \geq k/n$ then part (ii) shows that $\lambda_n(k/n) \leq a_m$ and in turn $\lambda_n(\xi) = \lambda_n(k/n) \leq a_m$. For $m \rightarrow \infty$ then $\lambda_n(\xi) \leq q_n(\xi)$.

We show $\lambda_n(\xi) \geq q_n(\xi)$. Let $a = \lambda_n(\xi)$ so that $\lambda_n(\xi) = \lambda_n(k/n) \leq a$. From part (ii) we find that $F_n(a) \geq k/n$ so that $F_n\{\lambda_n(k/n)\} \geq k/n$. The definition of $q_n(\xi)$ shows that $\lambda_n(\xi) \geq q_n(\xi)$. ■

Proof of Theorem 3.2. First, assume $\liminf_{n \rightarrow \infty} \lambda_n(\xi^*) > 0$ for some $\xi^* \leq 1$. Let $k^* = \text{int}(n\xi^*)$, then Theorem 3.1(ii) shows that $\lambda_n(k^*/n) = \lambda_n(\xi^*) > a_0 > 0$ for some a_0 and all $n \geq n_0$. It then follows from (3.3) that $F_n(a_0) < k^*/n \leq \xi^*$ for $n \geq n_0$, and hence $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi^*$ and we can choose $\xi = \xi^*$.

Second, if $\limsup_{(a,n) \rightarrow (0,\infty)} F_n(a) \leq \xi < 1$ then for each $\eta > 0$, there is a_0, n_0 such that for $n \geq n_0$ and $a \leq a_0$ then $F_n(a_0) < \xi + \eta$. Next note that $\text{int}\{n(\xi + 2\eta)\}/n \rightarrow \xi + 2\eta < 1$, if we choose η small. Thus we can choose n so large that $F_n(a_0) < \xi + \eta < \text{int}\{n(\xi + 2\eta)\}/n$, and

therefore by (3.3) we find $\lambda_n(\xi + 2\eta) > a_0$. It follows that $\liminf_{n \rightarrow \infty} \lambda_n(\xi + 2\eta) \geq a_0$, so we choose $\xi^* = \xi + 2\eta < 1$ as desired. ■

Theorem 3.3 demonstrates the F_n condition for stationary regressors. The proof involves a martingale decomposition. The martingale is analyzed using the inequality in Theorem 4.1 case 1, which in turn was proved using a chaining argument and the iterated martingale inequality in Lemma 4.1.

Proof of Theorem 3.3. We assume that $z_i = (1, x_i)'$, where x_i is a stationary process. We want to prove that $F_n(a) = o_{\mathbb{P}}(1)$ for $(a, n) \rightarrow (0, \infty)$, see (3.1).

We truncate each of the stationary regressors at A and decompose $F_{n\delta}(a)$ as follows

$$\begin{aligned} F_{n\delta}(a) &= n^{-1} \sum_{i=1}^n [1_{(|z'_i \delta| \leq a, |x_i| \leq A)} - \mathbb{E}\{1_{(|z'_i \delta| \leq a, |x_i| \leq A)} | \mathcal{G}_{i-1}\}] \\ &\quad + n^{-1} \sum_{i=1}^n 1_{(|z'_i \delta| \leq a, |x_i| > A)} + n^{-1} \sum_{i=1}^n \mathbb{E}\{1_{(|z'_i \delta| \leq a, |x_i| \leq A)} | \mathcal{G}_{i-1}\} \\ &= M_{n\delta}(a) + R_{1n\delta}(a) + R_{2n\delta}(a). \end{aligned}$$

We have to prove that the terms $M_{n\delta}(a)$, $R_{1n\delta}(a)$, $R_{2n\delta}(a)$ vanish in probability uniformly in $|\delta| = 1$ for suitable choices of a , A , and n .

The remainder term $R_{1n\delta}$. From $1_{(|z'_i \delta| \leq a, |x_i| > A)} \leq 1_{(|x_i| > A)}$, we find by Chebychev's inequality

$$\mathbb{P}\{\sup_{|\delta|=1} R_{1n\delta}(a) > \eta\} \leq \mathbb{P}\{n^{-1} \sum_{i=1}^n 1_{(|x_i| > A)} \geq \eta\} \leq \frac{1}{\eta} \mathbb{P}(|x_1| \geq A), \quad (\text{A.17})$$

which can be made arbitrary small by choosing A large.

The remainder term $R_{2n\delta}$. Since $|\delta| = 1$ we can write $z'_i \delta = -\sin \theta + \gamma' x_i \cos \theta$ for $\cos \theta > 0$ and $|\gamma| = 1$. Thus, using (3.4) with $c = a \leq 1/2$ we get $(\cos \theta)^{-1} \leq 2(A + 1)$. Then

$$(|z'_i \delta| \leq a, |x_i| \leq A) = \{|-\sin \theta + (\gamma' x_i) \cos \theta| \leq a, |x_i| \leq A\} \subset \{|-\tan \theta + \gamma' x_i| \leq 2a(A + 1)\}.$$

Further, the density of $\gamma' x_i$ (and hence of $|\tan \theta + \gamma' x_i|$) given \mathcal{G}_{i-1} is bounded by assumption, and we find

$$\mathbb{E}\{1_{(|z'_i \delta| \leq a, |x_i| \leq A)} | \mathcal{G}_{i-1}\} \leq Ca(A + 1), \quad (\text{A.18})$$

which can be made arbitrarily small for fixed A by choosing a small.

The martingale term $M_{n\delta}$. Define the compact set $\mathcal{K} = \{\delta' \delta = 1\} \subset \mathbb{R}^{m+1}$, choose $\delta_0 \in \mathcal{K}$, let $u_{ni}(\delta) = 1_{(|z'_i \delta| \leq a, |x_i| \leq A)} - 1_{(|z'_i \delta_0| \leq a, |x_i| \leq A)}$ so that $u_{ni}(\delta_0) = 0$ and write

$$M_{n\delta}(a) = \frac{1}{n} \sum_{i=1}^n [1_{(|z'_i \delta_0| \leq a, |x_i| \leq A)} - \mathbb{E}\{1_{(|z'_i \delta_0| \leq a, |x_i| \leq A)} | \mathcal{G}_{i-1}\}] + \frac{1}{n} \sum_{i=1}^n [u_{ni}(\delta) - \mathbb{E}\{u_{ni}(\delta) | \mathcal{G}_{i-1}\}].$$

The first term does not depend on δ and vanishes by the Law of Large Numbers for martingales. For the second term we apply Theorem 4.1 case 1 with $\eta = 0$ and $\nu = 1$. To check condition (4.6) we must bound

$$|u_{ni}(\delta) - u_{ni}(\tilde{\delta})| = |1_{(|z'_i \tilde{\delta}| \leq a, |x_i| \leq A)} - 1_{(|z'_i \delta| \leq a, |x_i| \leq A)}|.$$

Replacing $\tilde{\delta}$ by $\delta + (\tilde{\delta} - \delta)$ and using the triangle inequality we get, for $|\delta - \tilde{\delta}| \leq Qn^{-\phi}$,

$$|u_{ni}(\delta) - u_{ni}(\tilde{\delta})| \leq \mathbf{1}_{(|z'_i\delta - a| \leq Qn^{-\phi}|z_i|, |x_i| \leq A)} + \mathbf{1}_{(|z'_i\delta + a| \leq Qn^{-\phi}|z_i|, |x_i| \leq A)}.$$

As before, we can write $z'_i\delta = -\sin\theta + x'_i\gamma \cos\theta$ for $\cos\theta > 0$ and $|\gamma| = 1$. Since $|z_i| \leq 1 + |x_i| \leq 1 + A$ we get

$$\{|z'_i\delta \pm a| \leq Qn^{-\phi}|z_i|, |x_i| \leq A\} \subset \{|-\sin\theta + x'_i\gamma \cos\theta \pm a| \leq Qn^{-\phi}(1 + A), |x_i| \leq A\}.$$

Then, (3.4) with $c = a + Qn^{-\phi}(1 + A) < 1/2$ shows $(\cos\theta)^{-1} \leq 2(A + 1)$ so that

$$\{|z'_i\delta \pm a| \leq Qn^{-\phi}|z_i|, |x_i| \leq A\} \subset \left\{ \left| -\frac{\sin\theta \pm a}{\cos\theta} + x'_i\gamma \right| \leq 2Qn^{-\phi}(1 + A)^2 \right\} = S_i^\pm.$$

We can then bound

$$\sup_{\tilde{\delta}: |\delta - \tilde{\delta}| \leq Qn^{-\phi}} |u_{ni}(\delta) - u_{ni}(\tilde{\delta})|^p \leq (1_{S_i^-} + 1_{S_i^+})^p \leq C(1_{S_i^-} + 1_{S_i^+}).$$

Because $(\cos\theta)^{-1}$ is bounded and the conditional density of $\gamma'x_i$ given \mathcal{G}_{i-1} is bounded in $|\gamma| = 1$, it follows that

$$\mathbf{E} \left\{ \sup_{\tilde{\delta}: |\delta - \tilde{\delta}| \leq Qn^{-\phi}} |u_{ni}(\delta) - u_{ni}(\tilde{\delta})|^p \middle| \mathcal{G}_{i-1} \right\} \leq CQn^{-\phi}(1 + A)^2,$$

so that (4.6) holds with $A_{ni}(\delta) = C(1 + A)^2$, and hence (4.8) holds. Theorem 4.1, case 1, with $\eta = 0$ and $\nu = 1$ now shows

$$\sup_{|\delta|=1} \left| \frac{1}{n} \sum_{i=1}^n [u_{ni}(\delta) - \mathbf{E}\{u_{ni}(\delta) | \mathcal{G}_{i-1}\}] \right| = o_{\mathbf{P}}(1). \quad (\text{A.19})$$

Combining (A.17), (A.18), and (A.19) we find that for any $\epsilon > 0$, we first take A so large that $P(\sup_{|\delta|=1} R_{1n\delta} \geq \eta/3) \leq \epsilon/3$, and then a and Q so small that $a + Q(1 + A) < 1/2$, and $P(\sup_{|\delta|=1} R_{2n\delta} \geq \eta/3) \leq \epsilon/3$, and finally n so large that $P(\sup_{|\delta|=1} |M_{n\delta}| \geq \eta/3) \leq \epsilon/3$. This proves (3.1) and hence Theorem 3.3. ■

Theorem 3.4 demonstrates the F_n condition for random regressors. This requires a chaining argument, but without a martingale decomposition. We therefore need a new concentration inequality which is presented before the proof of the Theorem.

Lemma A.2 *Consider the random walk $x_i = \sum_{j=1}^i \eta_j \in \mathbb{R}^m$, where η_j i.i.d. $(0, \Phi)$, $j = 1, \dots, n$, and assume that the density of $\gamma'x_i / (i\gamma'\Phi\gamma)^{1/2}$ is bounded uniformly in $|\gamma| = 1$ and $i = 1, \dots, n$. Then the sets*

$$\mathbb{B}_i = \left\{ \left| -\frac{\sin\theta + a}{\cos\theta} + \frac{\gamma'x_i}{n^{1/2}} \right| \leq M \right\},$$

satisfy

$$\mathbf{E} \left(n^{-1} \sum_{i=1}^n \mathbf{1}_{\mathbb{B}_i} \right)^{m+1} \leq n^{-1} + CM^{m+1}. \quad (\text{A.20})$$

Proof of Lemma A.2. We find

$$\mathbb{E}\left(\sum_{i=1}^n 1_{\mathbb{B}_i}\right)^{m+1} = \sum_{1 \leq i_1, \dots, i_{m+1} \leq n} \mathbb{E}\left(\prod_{j=1}^{m+1} 1_{\mathbb{B}_{i_j}}\right) = \sum_{\text{at least two } i_j \text{ equal}} \mathbb{E}\left(\prod_{j=1}^{m+1} 1_{\mathbb{B}_{i_j}}\right) + \sum_{\text{all } i_j \text{ different}} \mathbb{E}\left(\prod_{j=1}^{m+1} 1_{\mathbb{B}_{i_j}}\right)$$

The first sum contains at most n^m terms which are all bounded by 1 and hence the contribution is at most n^m , which accounts for the term n^{-1} in (A.20).

Conditioning on the σ -field $\mathcal{G}_m = \sigma\{\eta_j, j \leq i_m\}$ we can express the second sum as

$$(m+1)! \sum_{1 \leq i_1 < \dots < i_m \leq n} \mathbb{E}\left\{ \left(\prod_{j=1}^m 1_{\mathbb{B}_{i_j}} \right) \sum_{i_{m+1}=i_{m+1}}^n \mathbb{E}(1_{\mathbb{B}_{i_{m+1}}} | \mathcal{G}_m) \right\}. \quad (\text{A.21})$$

Let $\sigma_{m+1}^2 = \text{Var}(\gamma' \sum_{j=i_{m+1}}^{i_{m+1}} \eta_j) = (i_{m+1} - i_m) \gamma' \Phi \gamma$ be the conditional variance of $\gamma' \sum_{j=1}^{i_{m+1}} \eta_j$, given \mathcal{G}_m . Then

$$\mathbb{E}(1_{\mathbb{B}_{i_{m+1}}} | \mathcal{G}_m) = \mathbb{P}\left(\left| -\frac{\sin \theta + a}{\cos \theta} \frac{n^{1/2}}{\sigma_{m+1}} + \frac{\gamma' x_{i_m}}{\sigma_{m+1}} + \frac{\gamma' \sum_{j=i_{m+1}}^{i_{m+1}} \eta_j}{\sigma_{m+1}} \right| \leq M \frac{n^{1/2}}{\sigma_{m+1}} \right) | \mathcal{G}_m,$$

is the probability that the random component, $\gamma' \sum_{j=i_{m+1}}^{i_{m+1}} \eta_j / \sigma_{m+1}$, is contained in an interval of length $2Mn^{1/2} \sigma_{m+1}^{-1}$. Hence the assumption of a bounded density of the normalized random walk implies that

$$\mathbb{E}(1_{\mathbb{B}_{i_{m+1}}} | \mathcal{G}_m) \leq CMn^{1/2} \frac{1}{(i_{m+1} - i_m)^{1/2}}.$$

Summing over i_{m+1} we find the bound

$$\sum_{i_{m+1}=i_{m+1}}^n \mathbb{E}(1_{\mathbb{B}_{i_{m+1}}} | \mathcal{G}_m) \leq CMn^{1/2} (n - i_m)^{1/2} \leq CMn.$$

Inserting this into (A.21) we get

$$\sum_{1 \leq i_1 < \dots < i_{m+1} \leq n} \mathbb{E}\left(\prod_{j=1}^{m+1} 1_{\mathbb{B}_{i_j}}\right) \leq CMn \sum_{1 \leq i_1 < \dots < i_m \leq n} \mathbb{E}\left(\prod_{j=1}^m 1_{\mathbb{B}_{i_j}}\right).$$

Repeating the argument we find the result in (A.20). ■

Proof of Theorem 3.4. We assume that the regressors are $z_{ni} = (1, n^{-1/2} x_i')'$, where x_i is a random walk. We want to prove that $F_n(a) = o_{\mathbb{P}}(1)$ for $(a, n) \rightarrow (0, \infty)$, see (3.1).

To study the $\sup_{|\delta|=1} F_{n\delta}(a)$, we apply a chaining argument and let $m = \dim x$. We therefore consider $\delta \in \mathbb{R}^{m+1}$ and cover the m dimensional surface $\mathcal{K} = \{|\delta| = 1\}$ with $L = \eta^{-m}$ balls, $\mathbb{B}(\delta_\ell, \eta)$, of equal radius η and centers $\delta_\ell, \ell = 1, \dots, L$, and evaluate $\sup_{|\delta|=1} F_{n\delta}(a)$ as follows

$$\sup_{|\delta|=1} F_{n\delta}(a) \leq \max_{1 \leq \ell \leq L} F_{n\delta_\ell}(a) + \max_{1 \leq \ell \leq L} \sup_{\mathbb{B}(\delta_\ell, \eta) \cap \mathcal{K}} |F_{n\delta}(a) - F_{n\delta_\ell}(a)|.$$

We truncate the stochastic regressors $|n^{-1/2} x_i|$ by A and find, using Boole's inequality, that

$$\mathbb{P}\left\{ \sup_{|\delta|=1} F_{n\delta}(a) > \eta \right\} \leq \mathcal{P}_{0n} + \mathcal{P}_{1n} + \mathcal{P}_{2n},$$

where

$$\begin{aligned}\mathcal{P}_{0n} &= \mathbf{P}(\max_{1 \leq i \leq n} n^{-1/2}|x_i| > A), \\ \mathcal{P}_{1n} &= \sum_{\ell=1}^L \mathbf{P}\{F_{n\delta_\ell}(a) > \eta/2, \max_{1 \leq i \leq n} n^{-1/2}|x_i| \leq A\}, \\ \mathcal{P}_{2n} &= \sum_{\ell=1}^L \mathbf{P}\left\{\sup_{\mathbb{B}(\delta_\ell, \eta) \cap \mathcal{K}} |F_{n\delta}(a) - F_{n\delta_\ell}(a)| > \eta/2, \max_{1 \leq i \leq n} n^{-1/2}|x_i| \leq A\right\}.\end{aligned}$$

We discuss these in turn.

\mathcal{P}_{0n} . By boundedness of $\max_{1 \leq i \leq n} n^{-1/2}|x_i|$, \mathcal{P}_{0n} tends to zero for $A \rightarrow \infty$ uniformly in n .

\mathcal{P}_{1n} . We bound \mathcal{P}_{1n} by the Markov inequality

$$\mathcal{P}_{1n} \leq \frac{2}{\eta} \sum_{\ell=1}^L \mathbf{E}\{F_{n\delta_\ell}(a) \mathbf{1}_{(\max_{1 \leq i \leq n} n^{-1/2}|x_i| \leq A)}\} = \frac{2}{\eta} \sum_{\ell=1}^L \frac{1}{n} \sum_{i=1}^n \mathbf{P}(|z'_{ni}\delta_\ell| \leq a, \max_{1 \leq i \leq n} n^{-1/2}|x_i| \leq A).$$

Write $\delta'_\ell z_{ni} = -\sin \theta_\ell + n^{-1/2}\gamma'x_i \cos \theta_\ell$ where $\cos \theta_\ell > 0$ and $|\gamma| = 1$. From (3.4) with $|x|$ replaced by $n^{-1/2}|\gamma'x_i| \leq A$ while $c = a < 1/2$, we find $(\cos \theta_\ell)^{-1} \leq 2(1+A)$ and therefore, when dividing by $\cos \theta_\ell$, and leaving out the intersection with $(\max_{1 \leq i \leq n} n^{-1/2}|x_i| \leq A)$ we get the further bound

$$\mathcal{P}_{1n} \leq \frac{2}{\eta} \sum_{\ell=1}^L \frac{1}{n} \sum_{i=1}^n \mathbf{P}\{|-\tan \theta_\ell + n^{-1/2}\gamma'x_i| \leq 2a(1+A)\}.$$

Dividing by $n^{-1/2}(i\gamma'\Phi\gamma)^{1/2}$ gives

$$\mathcal{P}_{1n} \leq 2\eta^{-1} \sum_{\ell=1}^L \frac{1}{n} \sum_{i=1}^n \mathbf{P}\left\{\left|-\frac{n^{1/2}}{(i\gamma'\Phi\gamma)^{1/2}} \tan \theta_\ell + \frac{\gamma'_\ell x_i}{(i\gamma'_\ell\Phi\gamma_\ell)^{1/2}}\right| \leq \frac{2a(A+1)n^{1/2}}{(i\gamma'\Phi\gamma)^{1/2}}\right\}.$$

The random variable $(\gamma'_\ell x_i)(i\gamma'_\ell\Phi\gamma_\ell)^{-1/2}$ is assumed to have a bounded density and the probability that it is contained in an interval of length $4(A+1)n^{1/2}(i\gamma'_\ell\Phi\gamma_\ell)^{-1/2}$, then gives the inequality

$$\mathcal{P}_{1n} \leq CL2a(A+1) \frac{1}{n} \sum_{i=1}^n \left(\frac{n}{i}\right)^{1/2} \leq CLAa. \quad (\text{A.22})$$

\mathcal{P}_{2n} . Let $z_{ni} = (1, x'_i n^{-1/2})'$ and note $|\delta_\ell - \delta| < \eta$ resulting in the inequality

$$|\mathbf{1}_{\{|z'_{ni}\delta_\ell| \leq a\}} - \mathbf{1}_{\{|z'_{ni}\delta| \leq a\}}| = |\mathbf{1}_{\{|z'_{ni}\delta_\ell| \leq a\}} - \mathbf{1}_{\{|z'_{ni}\delta_\ell + z'_{ni}(\delta - \delta_\ell)| \leq a\}}| \leq \mathbf{1}_{\{|z'_{ni}\delta_\ell - a| \leq \eta|z_{ni}|\}} + \mathbf{1}_{\{|z'_{ni}\delta_\ell + a| \leq \eta|z_{ni}|\}}.$$

The same holds multiplying by $\mathbf{1}_{\{|x_i| \leq A\}}$. Introducing $z'_{ni}\delta_\ell = -\sin \theta_\ell + \cos \theta_\ell(\gamma'_\ell x_i n^{-1/2})$ and the bound $|z_{ni}| \leq 1 + |x_i|n^{-1/2} \leq 1 + A$, we apply (3.4) for $b = \pm a$, $c = a + \eta(A+1) < 1/2$. We find that $(\cos \theta_\ell)^{-1} \leq 2(A+1)$, and therefore

$$\{|z'_{ni}\delta_\ell \pm a| \leq \eta|z_{ni}|, |x_i|n^{-1/2} \leq A\} \subset \left\{\left|-\frac{\sin \theta_\ell \pm a}{\cos \theta_\ell} + \frac{\gamma'_\ell x_i}{n^{1/2}}\right| \leq 2\eta(1+A)^2\right\} = \mathbb{B}_{i,\ell}^\pm,$$

say. By Chebychev's inequality

$$\begin{aligned} \mathcal{P}_{2n} &\leq \sum_{\ell=1}^L \{ \mathbb{P}(n^{-1} \sum_{i=1}^n 1_{\mathbb{B}_{i,\ell}^+} > \eta/4) + \mathbb{P}(n^{-1} \sum_{i=1}^n 1_{\mathbb{B}_{i,\ell}^-} > \eta/4) \} \\ &\leq \left(\frac{4}{\eta n}\right)^{m+1} \left\{ \sum_{\ell=1}^L \mathbb{E} \left(\sum_{i=1}^n 1_{\mathbb{B}_{i,\ell}^+} \right)^{m+1} + \mathbb{E} \left(\sum_{i=1}^n 1_{\mathbb{B}_{i,\ell}^-} \right)^{m+1} \right\}. \end{aligned}$$

From Lemma A.2 with $M = 2\eta(1+A)^2$, we find from (A.20) that

$$\mathcal{P}_{2n} \leq 2L \left(\frac{4}{\eta}\right)^{m+1} \{n^{-1} + C\eta^{m+1}(1+A)^{2(m+1)}\} \leq CLn^{-1} + CL\eta^{m+1}(1+A)^{2(m+1)}. \quad (\text{A.23})$$

We therefore find from (A.22) and (A.23), using $\eta^{m+1} = L^{-1-1/m}$, that

$$\mathbb{P}\{\sup_{|\delta|=1} F_{n\delta}(a) > \eta\} \leq \mathbb{P}\{\max_{1 \leq i \leq n} |x_i|n^{-1/2} > A\} + CLa + CLn^{-1} + CL^{-1/m}(1+A)^{2(m+1)}.$$

By boundedness of $\max_{1 \leq i \leq n} |x_i|n^{-1/2}$ we can choose $A > 0$ so large that $\mathbb{P}\{\max_{1 \leq i \leq n} |x_i|n^{-1/2} > A\} \leq \epsilon/4$ for all n . Next choose a small and L large so that $a + \eta(1+A) = a + L^{-1/m}(1+A) \leq 1/2$ and $CL^{-1/m}(1+A)^{2(m+1)} \leq \epsilon/4$, then a so small that $CLa \leq \epsilon/4$, and finally n so large that $CLn^{-1} \leq \epsilon/4$. This proves (3.1) and hence Theorem 3.4. ■

References

- Abadir, K.M. and Lucas, A. (2000) Quantiles for t-statistics based on M-estimators and unit roots. *Economic Letters* 67, 131–137.
- Bercu, B. and Touati, A. (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848–1869.
- Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
- Chen, X.R. and Wu, Y.H. (1988) Strong consistency of M-estimates in linear models. *Journal of Multivariate Analysis* 27, 116–130.
- Čížek, P. (2005) Least trimmed squares in nonlinear regression under dependence. *Journal of Statistical Planning and Inference* 136, 3967–3988.
- Davies, L. (1990) The asymptotics of S -estimators in the linear regression model. *Annals of Statistics* 18, 1651–1675.
- Fasano, M.V., Maronna, R.A., Sued, M. and Yohai, V.J. (2012) Continuity and differentiability of regression M functionals. *Bernoulli* 18, 1284–1309.
- 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*. Wiley, New York.
- Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.
- Johansen, S. and Nielsen, B. (2016) Analysis of the Forward Search using some new results for martingales and empirical processes. *Bernoulli* 22, 1131–1183.
- Jurečková, J., Sen, P.K. and Picek, J. (2012) *Methodological Tools in Robust and Nonparametric Statistics*. London: Chapman & Hall/CRC Press.

- Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261–278.
- Knight, K. (1991) Limit theory for M-estimates in an integrated infinite variance process. *Econometric Theory* 7, 200–212.
- Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R. and Xiao, Z. (2004) Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99, 775–787.
- Liese, F. and Vajda, I. (1994) Consistency of M-estimates in general regression models. *Journal of Multivariate Analysis* 50, 93–114.
- Lucas, A. (1995) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.
- Maronna, R.A., Martin, D.R. and Yohai, V.J. (2006) *Robust Statistics: Theory and Methods*. Wiley, New York.
- Višek, J.Á. (2006) The least trimmed squares. Part I: Consistency. *Kybernetika* 42, 1–36.