Discussion Papers
Department of Economics
University of Copenhagen

No. 16-05

Tightness of M-estimators for multiple linear regression in time for multiple linear regression in time series

Søren Johansen

Bent Nielsen

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

ISSN: 1601-2461 (E)

Tightness of M-estimators for multiple linear regression in time series

Søren Johansen[1][2] & Bent Nielsen[3]

28 May 2016

Abstract: We show tightness 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. Tightness 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 an 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.
JEL Classification: 22.

1 Introduction and summary

We show tightness for a class of regression M-estimators, where the objective function can be non-monotonic 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 tightness 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. 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

$$y_i = \mu + x_{ni}'\alpha + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

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

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

[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, and to Steffen Lauritzen for useful comments.

[2]Department 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.

[3]Nuffield College & Department of Economics, University of Oxford & Programme for Economic Modelling. Address: Nuffield College, Oxford OX1 INF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk.

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 criterion function, as well as quantile regression estimators, in particular the least absolute deviation, and least squares estimator, which all have a convex criterion function.

The asymptotic theory of the regression M-estimator is well understood for nice criterion functions ρ. Maronna, Martin, and Yohai (2006, §10.9) provide an asymptotic theory for regression M-estimators and show existence and uniqueness for the case of a convex, differentiable criterion function. Chen and Wu (1988) give two results on tightness (and consistency) for more general criterion functions. In both cases the criterion function $\rho(u)$ is continuous, non-decreasing in $u > 0$ and non-increasing for $u < 0$. Their Theorem 1 shows tightness of $\hat{\mu}$ and $\hat{\alpha}$ in the regression $y_i = \mu + x_i'\alpha + \varepsilon_i$, when (y_i, x_i') are i.i.d. and $\mathsf{E}\rho(y_i - \mu - x_i'\alpha)$ has a unique minimum. Their Theorem 4 shows tightness when x_i is deterministic and satisfies a condition on the frequency of small regressors.

In this paper we generalize the result of Chen and Wu (1988). 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 criterion function $h(v)$, 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 tightness 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.

It is worth noting that 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 $(1-\epsilon)\mathsf{f}_0 + \epsilon\mathsf{f}_1$, 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 tightness 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, $i = 1, \ldots, n$, 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

$$y_i = \mu + x_{ni}'\alpha + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{2.1}$$

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 of

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

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 $\rho(u) = \frac{1}{2}u^2 1_{(|u|\leq c)} + c(|u - c| + \frac{1}{2}c)1_{(|u|>c)}$. The formulation also covers least squares regression, $\rho(u) = u^2/2$, and quantile regression, $\rho(u) = -(1-p)u1_{(u<0)} + pu1_{(u>0)}$, 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 $\rho(u) = -\ln \mathsf{f}(u)$ gives the maximum likelihood estimators.

2.2 The assumptions and the result on tightness

For the tightness 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 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 on the dimension of the regressors, see Assumption 1(iic). 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 (iib), but it need not rise monotonically from the origin. Lower semi-continuity in (iia) 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

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

Assumption 1 (i) *Let* $\mathcal{F}_i, i = 1, \ldots, n$ *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* $\liminf_{v\to u} \rho(v) \geq \rho(u)$ *for all* $u \in \mathbb{R}$*;*
(b) *Let* $0 < h(v) = \mathsf{E}\{\rho(\varepsilon_i - v)\} < \infty$, *and let* $\mu_*, u_* \in \mathbb{R}$ *exist so that*

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

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

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

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

Suppose

(*a*) $\lim_{(a,n)\to(0,\infty)}\mathsf{P}[\sup_{|\delta|=1}\{F_{n\delta}(a)-F_{n\delta}(0)\}\geq\epsilon]\to 0$ *for all* $\epsilon>0$;

(*b*) *a* $0<\xi<1$ *exists such that* $\lim_{n\to\infty}\mathsf{P}\{F_n(0)\geq\xi\}=0$.

We give some remarks on Assumption 1.

Remark 2.1 *Assumption 1(iiib) implies that* $\hat{\Sigma}_n=n^{-1}\sum_{i=1}^{n}z_{in}z_{in}'$ *is positive definite, because if* $\delta'\hat{\Sigma}_n\delta=0$, *then* $z_{ni}'\delta=0$ *for* $i=1,\ldots,n$, *and* $F_{n\delta}(0)=1$.

Remark 2.2 *Assumption 1(iiia, b) implies that* $\hat{\Sigma}_n$ *is bounded away from zero in large samples. We prove this by noting that*

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

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

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

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

Remark 2.3 *In some situations we get an inverse of the result in Remark 2.2. 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* $\liminf_{n\to\infty}\lambda_n>0$ *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 2.4 *If* $F_n(a)=\mathrm{o}_{\mathsf{P}}(1)$ *as* $(a,n)\to(0,\infty)$ *then Assumption 1(iii) is satisfied, because* $F_{n\delta}(a)-F_{n\delta}(0)\leq F_n(a)$. *To be precise, it suffices that for all* $\epsilon>0,\eta>0$ *there exist* $a_0,n_0>0$ *such that*

$$\mathsf{P}\{F_n(a)\geq\eta\}\leq\epsilon\qquad\text{for } a\leq a_0, n\geq n_0. \tag{2.4}$$

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

Remark 2.5 *If Assumption 1(iiia, b) are satisfied and the regressors are deterministic, it holds that* $\limsup_{(a,n)\to(0,\infty)}F_n(a)\leq\xi$, *see Davies (1990). We return to this issue in Section 3.1 and Example 3.5.*

We now give the main result on the evaluation of the objective function and the result on tightness.

Theorem 2.1 ***Tightness.*** *Under Assumption 1, we can for all $\epsilon > 0$ find $B, n_0 > 0$, and a set $\mathbb{C}_n$ 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*

$$|\hat{\beta}| = \mathrm{O}_{\mathsf{P}}(1).$$

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 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 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 $(\hat{\mu}, \hat{\alpha}) \to (\mu_0, \alpha_0)$ *a.s.* under the following conditions. The regressors are deterministic, the criterion function is bounded, $0 < \rho(\infty) = \rho(-\infty) < \infty$, and $F_n(a) \to 0$ as $(a, n) \to (0, \infty)$, 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 tightness for symmetric density f and deterministic regressors. He defines for $0 < \xi < 1$

$$\lambda_n(\xi) = \min_{|S|=\mathrm{int}(n\xi)} \min_{|\delta|=1} \max_{i \in S} |z_{ni}' \delta|, \tag{3.1}$$

where S are subsets of the indices $i = 1, \ldots, n$. It is a consequence of his Theorem 3, that if $\liminf_{n\to\infty} \lambda_n(\xi) > 0$ 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 the condition $\liminf_{n\to\infty} \lambda_n(\xi) > 0$ with Assumption 1$(iiib)$ that $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi$, and show that $\lambda_n(\xi)$ is almost the inverse of $F_n(a)$, see Remark 2.5.

Theorem 3.1 *Let the regressors be deterministic.*

(i) *For* $0 \leq \xi \leq 1$,

$$\{F_n(a) > \mathrm{int}(n\xi)/n\} \subset \{\lambda_n(\xi) \leq a\} \subset \{F_n(a) \geq \mathrm{int}(n\xi)/n\}.$$

(ii) *The condition* $\liminf_{n\to\infty} \lambda_n(\xi^*) > 0$ *for some* $0 < \xi^* < 1$ *is equivalent to the condition that there exists* $0 < \xi < 1$ *for which* $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi$.

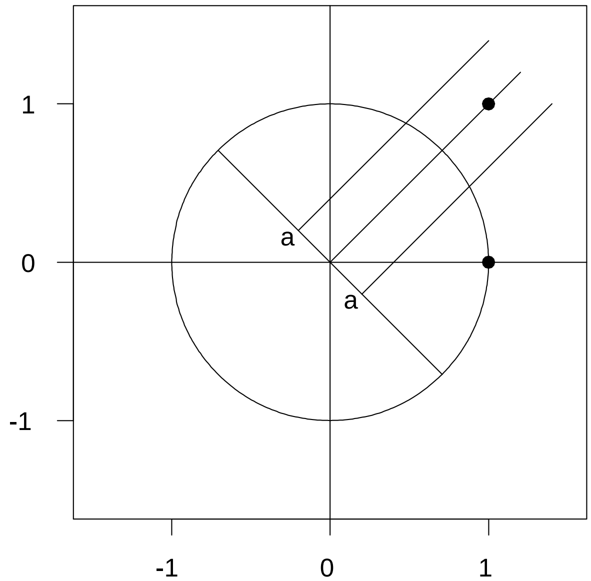

Figure 1: Illustration of $F_{n\delta}$ in the dummy variable case

3.2 Regression with deterministic regressors

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* $0 \leq c \leq 1/2$, $|\theta| < \pi/2$

$$|-\sin\theta + x\cos\theta| \leq c \qquad \Rightarrow \qquad \frac{1}{\cos\theta} \leq 2(1+|x|). \tag{3.2}$$

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* $F_n(a) \to 0$, $(a, n) \to (0, \infty)$, *used by Chen and Wu (1988), is not satisfied, see also Remark 2.4.*

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*

$$\frac{-a+\sin\theta}{\cos\theta} \leq (i/n)^q \leq \frac{a+\sin\theta}{\cos\theta}. \tag{3.3}$$

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 (3.2), 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 $F_n(a) \leq 4a/(1-a) \to 0$, for $(a,n) \to (0,\infty)$, 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 $F_n(a) \leq c\{2a/(1-a^2)^{1/2}\}^{1/q} \to 0$, for $(a,n) \to (0,\infty)$.

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 $\sin\theta = 1/\sqrt{2} - a + \mathrm{o}(a)$ and $\cos\theta = \{1-(1/\sqrt{2}-a)^2\}^{1/2}\{1+\mathrm{o}(a)\} = 1/\sqrt{2}+\mathrm{o}(a)$. This gives the bound $F_n(a) \leq c\{(1/\sqrt{2})^{1/q} - (1/\sqrt{2}-2a)^{1/q}\}/(1/\sqrt{2})^{1/q} \to 0$, for $(a,n) \to (0,\infty)$.

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 $\sin\theta = 1/\sqrt{2}+a+\mathrm{o}(a)$ and $\cos\theta = \{1-(1/\sqrt{2}+a)^2\}^{1/2}\{1+\mathrm{o}(a)\} = 1/\sqrt{2}+\mathrm{o}(a)$. In this case the interval for i becomes*

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

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

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

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

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

In Theorem 2.1, Assumption $1(iii,a)$ to F_n is a sufficient condition for tightness of $\hat{\beta}$. The necessity of Assumption $1(iii,a)$ for tightness 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 tight, such that also $\hat{\alpha} = \varepsilon_n - \hat{\mu}$ is tight. On the other hand we find for $0<a<1$, the function $F_n(a) = n^{-1}\sum_{i=1}^n 1_{(|z_i| \leq a)} = (n-1)/n \longrightarrow 1$, for $(a,n) \longrightarrow (0,\infty)$.

Example 3.5 *In the regression $z_{ni} = (1,1), i \leq n/4$ and $z_{ni} = (1, 2^{-i})$, $n/4 < i \leq n$ we find $F_n(0) = \mathrm{int}(n/4)/n \to 1/4$ and Assumption 1(iiib) is satisfied, but $\lim_{(a,n)\to(0,\infty)}\sup_{|\delta|=1}\{F_{n\delta}(a) - F_{n\delta}(0)\} = 3/4$ so Assumption 1(iiia) is not satisfied. However, $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi = 3/4$, so the condition of Davies is satisfied.*

3.3 Regression with multiple stochastic regressors

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

Theorem 3.2 *(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* $i = 1, \ldots, n$.

Then $F_n(a) = \mathsf{o}_\mathsf{P}(1)$, *for* $(a, n) \to (0, \infty)$, *such that Assumption 1(iii) holds.*

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* $\mathcal{G}_{i-1} = \sigma(x_1, \ldots, x_{i-1})$ *be bounded uniformly in* $(x_1, \ldots, x_{i-1}, x_i)$ *and* $|\gamma| = 1$, $\gamma \in \mathbb{R}^m$.

Then $F_n(a) = \mathsf{o}_\mathsf{P}(1)$, *for* $(a, n) \to (0, \infty)$, *such that Assumption 1(iii) holds.*

Theorems 3.2 and 3.3 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 3.2 is satisfied for a random walk with normal innovations. Indeed if* η_j *are independent normal* $\mathsf{N}_m(0, \Phi)$ *with positive definite covariance* Φ, *then* $\gamma' x_i/(i\gamma'\Phi\gamma)^{1/2}$ *is* $\mathsf{N}(0, 1)$ *for any* $|\gamma| = 1$.

Example 3.7 *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.* $\mathsf{N}_m(0, \Phi)$ *with positive definite variance* Φ, *then* $\gamma' x_i$ *given* $\mathcal{G}_{i-1}$ *is* $\mathsf{N}(\gamma'\alpha x_{i-1}, \gamma'\Phi\gamma)$. *The conditional density is bounded in the mean, while the variance* $\gamma'\Phi\gamma$ *is finite and bounded away from zero when* $|\gamma| = 1$.

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.

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* $\mathsf{E}(u_{n\ell i}^{2^r}) < \infty$, $r \in \mathbb{N}$, $\ell = 1, \ldots, L$, $i = 1, \ldots, n$, *and let* ς *and* λ *be positive real numbers defined by*

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

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

Then, if ν *is chosen such that*

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

it holds that

$$\max_{1\le \ell\le L} |\sum_{i=1}^{n} (u_{n\ell i} - \mathsf{E}_{i-1} u_{n\ell i})| = \mathsf{o}_\mathsf{P}(n^\nu). \tag{4.3}$$

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 is applied in the proof of Theorem 3.3 for stationary regressors.

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 $\mathsf{E}|u_{ni}(\kappa)| < \infty$ for $\kappa \in \mathcal{K}$, and normalized by $u_{ni}(\kappa_0) = 0$ for some $\kappa_0 \in \mathcal{K}$. We define the martingales*

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

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 $\phi \geq 0$ and $\kappa \in \mathcal{K}$,

$$\mathsf{E}_{i-1} \sup_{\tilde{\kappa} \in \mathbb{B}(\kappa, Bn^{-\phi}) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})|^p \leq n^{-\phi} A_{ni}(\kappa). \tag{4.5}$$

Let η, ν satisfy either of

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

Suppose

$$n^{-1} \sum_{i=1}^{n} \mathsf{E}\{ \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa)\} \leq C. \tag{4.6}$$

Then it holds

$$\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |M_n(\kappa)| = \mathrm{o}_{\mathsf{P}}(n^{\nu}). \tag{4.7}$$

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 $\mathsf{E} \sum_{i=1}^{n} |u_i|^{2^r} = \mathrm{O}(n)$. Then*

$$\sup_{|\delta|=1} |\sum_{i=1}^{n} u_i 1_{(\delta' z_{ni} = 0)}| = \mathrm{o}_{\mathsf{P}}(n^{\nu}).$$

Note that in Theorem 4.1, Assumption (4.5) implies that for $\phi > 0$, $\mathsf{E}_{i-1}|u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})|^p$ is smooth in $(\kappa, \tilde{\kappa})$, whereas in Theorem 4.2 we find

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

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 tightness 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 tightness (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 tightness

Proof of Theorem 2.1. (a) *Behavior of the criterion function for large* $|\beta|$. Using $y_i = \beta_0' z_{ni} + \varepsilon_i$ we find

$$\rho(y_i - z_{ni}'\beta) = \rho\{\varepsilon_i - (\mu-\mu_0) - x_{ni}'(\alpha-\alpha_0)\} = \rho\{\varepsilon_i - z_{ni}'(\beta-\beta_*) - \mu_*\},$$

where

$$\beta_* = (\mu_0+\mu_*, \alpha_0')'.$$

Assumption 1(iib) shows that $\mu_*, u_* \in \mathbb{R}$ exists so that $h_* = h(\mu_*) = \mathsf{E}\rho(\varepsilon_i - \mu_*) < \rho_* = \inf_{|u|\geq|u_*|}\rho(u)$. We then give a condition for $\rho\{\varepsilon_i - z_{ni}'(\beta-\beta_*) - \mu_*\}$ to be greater than ρ_* for large $|\beta-\beta_*|$. Define the direction $\delta = (\beta-\beta_*)/|\beta-\beta_*|$, and length $\lambda = |\beta-\beta_*|$, so that $\beta - \beta_* = \lambda\delta$, $|\delta| = 1$, and $z_{ni}'(\beta-\beta_*) = \lambda z_{ni}'\delta$.

Then, for $|\varepsilon_i| \leq A$ and $|z_{ni}'\delta| \geq a$, and $\lambda \geq (A+|u_*|+|\mu_*|)/a$,

$$|\varepsilon_i - z_{ni}'(\beta-\beta_*) - \mu_*| \geq |z_{ni}'(\beta-\beta_*)| - |\varepsilon_i| - |\mu_*| \geq \lambda a - A - |\mu_*| \geq |u_*|.$$

Hence $\rho(\varepsilon_i - z_{ni}'(\beta-\beta_*) - \mu_*) \geq \rho_*$ for $|z_{ni}'\delta| \geq a$ and large λ, but not uniformly in δ.

(b) *A set $\mathbb{C}_n$ with large probability.* We define $m_i = \rho(\varepsilon_i - \mu_*) - h_*$ and the martingale

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

Assumption 1(iic) implies that the first term is $\mathrm{o}_{\mathsf{P}}(1)$ by the Law of Large Numbers for martingales, and the second term is $\mathrm{o}_{\mathsf{P}}(1)$ uniformly in δ by Theorem 4.2 used with $\nu = 1$.

Next we find by the Law of Large Numbers that $n^{-1}\sum_{i=1}^n 1_{(|\varepsilon_i|\geq A)} \to \mathsf{P}(|\varepsilon_1| \geq A)$, which is small for large A. Moreover, $\mathsf{P}[\sup_{|\delta|=1}\{F_{n\delta}(a) - F_{n\delta}(0)\} \geq \epsilon] \to 0$ while, for some $\xi < 1$, $\mathsf{P}\{\sup_{|\delta|=1} F_{n\delta}(0) \geq \xi\} \to 0$ by Assumption 1(iii). Collecting these results, we define below sets $\mathbb{C}_n$ with large probability. That is, for all $\epsilon, \eta > 0$ there exists $A_0, a_0, n_0 > 0$ such that for all $A \geq A_0, n \geq n_0$, $a \leq a_0$ the sets $\mathbb{C}_n$ defined by the inequalities

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

$$n^{-1}\sum_{i=1}^n 1_{(|\varepsilon_i|\geq A)} \leq \eta, \tag{A.2}$$

$$F_n(0) = \sup_{|\delta|=1} F_{n\delta}(0) \leq \xi, \tag{A.3}$$

$$\sup_{|\delta|=1}\{F_{n\delta}(a) - F_{n\delta}(0)\} \leq \eta. \tag{A.4}$$

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

(c) *Reformulation of* $R_n(\beta) - R_n(\beta_*)$. We note that $R_n(\beta) - R_n(\beta_*)$ does not depend on the terms with $z_{ni}'\delta = 0$, and therefore define

$$\tilde{R}_n(\beta) = n^{-1}\sum_{i=1}^n \rho(\varepsilon_i - \lambda z_{ni}'\delta - \mu_*) 1_{(|z_{ni}'\delta|>0)},$$

such that $R_n(\beta) - R_n(\beta_*) = \tilde{R}_n(\beta) - \tilde{R}_n(\beta_*)$.

(d) *Lower bound for $\tilde{R}_n(\beta)$ on $\mathbb{C}_n$ for large $|\beta - \beta_*|$ uniformly in δ.* Delete terms for which $|\varepsilon_i| \geq A$ or $|z_{ni}'\delta| \leq a$ and take $|\beta - \beta_*| = \lambda \geq (A + u_* + |\mu_*|)/a$, so that $\rho(\varepsilon_i - \lambda z_{ni}'\delta - \mu_*) \geq \rho_*$ by item *(a)*. Then,

$$\tilde{R}_n(\beta) \geq n^{-1}\sum_{i=1}^n \rho(\varepsilon_i - \lambda z_{ni}'\delta - \mu_*) 1_{(|\varepsilon_i|<A)} 1_{(|\delta' z_{ni}|>a)} \geq \rho_* n^{-1}\sum_{i=1}^n 1_{(|\varepsilon_i|<A)} 1_{(|\delta' z_{ni}|>a)}.$$

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

$$\tilde{R}_n(\beta) \geq \rho_*\{1 - n^{-1}\sum_{i=1}^n 1_{(|\varepsilon_i|\geq A)} - F_{n\delta}(a)\}.$$

Now, on the set $\mathbb{C}_n$ use (A.2) and (A.4) so that

$$\begin{aligned}\tilde{R}_n(\beta) &\geq \rho_*\{1 - \eta - F_{n\delta}(a)\} \\ &\geq \rho_*[1 - \eta - \sup_{|\delta|=1}\{F_{n\delta}(a) - F_{n\delta}(0)\} - F_{n\delta}(0)] \geq \rho_*\{1 - F_{n\delta}(0) - 2\eta\}.\end{aligned}$$

(e) *Upper bound for* $\tilde{R}_n(\beta_*)$ *on* $\mathbb{C}_n$. Using $m_i = \rho(\varepsilon_i - \mu_*) - h_*$ we find

$$\tilde{R}_n(\beta_*) = n^{-1}\sum_{i=1}^{n}\{\rho(\varepsilon_i - \mu_*) - h_* + h_*\}1_{(|z_{ni}'\delta|>0)} = M_n(\delta) + h_* n^{-1}\sum_{i=1}^{n}1_{(|z_{ni}'\delta|>0)}.$$

Recall the definition of $F_{n\delta}(0)$ and use (A.1) to get

$$\tilde{R}_n(\beta_*) \leq \eta + h_* n^{-1}\sum_{i=1}^{n}1_{(|z_{ni}'\delta|>0)} = \eta + h_*\{1 - F_{n\delta}(0)\}.$$

(f) *Combine (d) and (e) to get a uniform positive lower bound for* $R_n^*(\beta) - R_n^*(\beta_*)$.

$$\begin{aligned} R_n(\beta) - R_n(\beta_*) &\geq \rho_*\{1 - F_{n\delta}(0) - 2\eta\} - h_*\{1 - F_{n\delta}(0)\} - \eta \\ &= -\eta(2\rho_* + 1) + \rho_*\{1 - F_{n\delta}(0)\}(1 - h_*/\rho_*). \end{aligned}$$

The bound (A.3) to $F_{n\delta}(0) \leq F_n(0) \leq \xi < 1$ then shows that, uniformly in $|\delta| = 1$ and $|\beta - \beta_*| > (A + u_* + |\mu_*|)/a = B_0$, say, then

$$R_n(\beta) - R_n(\beta_*) \geq -\eta(2\rho_* + 1) + \rho_*(1 - \xi)(1 - h_*/\rho_*).$$

By Assumption 1(*iib*), $1 - h_*/\rho_* > 0$, such that the lower bound is positive when $\eta < \rho_*(1 - \xi)(1 - h_*/\rho_*)/(2\rho_* + 1)$. Thus, $\inf_{|\beta-\beta_*|\geq B_0} R_n(\beta) > R_n(\beta_*)$ but $R_n(\beta_*) \geq \inf_{|\beta-\beta_*|\leq B_0} R_n(\beta)$ on $\mathbb{C}_n$.

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

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 tight. ■

A.2 Proof of martingale results

Proof of Theorem 4.1. We study the martingale $M_n(\kappa) = \sum_{i=1}^{n}\{u_{ni}(\kappa) - \mathsf{E}_{i-1}u_{ni}(\kappa)\}$, see (4.4), on sets of the form $\mathbb{B}(\kappa_0, Bn^{-\eta})\cap\mathcal{K}$. We prove (4.7) in the situations

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

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

$$1/2 < \zeta < (2^{r-1} - 1)/(m + 1). \tag{A.5}$$

For $0 \leq \eta < 1/2$, cover $\mathbb{B}(\kappa_0, Bn^{-\eta})\cap\mathcal{K}$ by $L = \mathrm{O}\{n^{(\zeta-\eta)(m+1)}\}$ 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 $\ell = 1, \ldots, L$. Note that $L = \mathrm{O}\{n^{(\zeta-\eta)(m+1)}\} \to \infty$. 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

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

We need a martingale bound for $\mathcal{R}_{n2}$. Define

$$k_{ni}(\kappa_\ell, n^{-\zeta}) = \sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\kappa_\ell)|.$$

Applying the triangle inequality and then $|X - \mathsf{E}X| \leq (|X| - \mathsf{E}|X|) + 2\mathsf{E}|X|$ and defining

$$\widetilde{\mathcal{M}}_{n2\ell} = \sum_{i=1}^{n} \{k_{ni}(\kappa_\ell, n^{-\zeta}) - \mathsf{E}_{i-1} k_{ni}(\kappa_\ell, n^{-\zeta})\}, \qquad \overline{\mathcal{M}}_{n2\ell} = \sum_{i=1}^{n} \mathsf{E}_{i-1} k_{ni}(\kappa_\ell, n^{-\zeta})$$

gives the bound $\mathcal{R}_{n2} \leq \max_{1 \leq \ell \leq L}(\widetilde{\mathcal{M}}_{n2\ell} + 2\overline{\mathcal{M}}_{n2\ell})$. We then prove (4.7) 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 $u_{n\ell i} = u_{ni}(\kappa_\ell)$. Let $\lambda = (\zeta - \eta)(m+1)$ in (4.1). We argue that $\varsigma = 1 - \eta$ in (4.2). Indeed, we find from (4.5) with $\phi = \eta$, and $1 \leq p \leq 2^r$, because $u_{ni}(\kappa_0) = 0$,

$$\mathsf{E}_{i-1} \sup_{\kappa_l \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa_\ell)|^p = \mathsf{E}_{i-1} \sup_{\kappa_l \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa_\ell) - u_{ni}(\kappa_0)|^p \leq n^{-\eta} A_{ni}(\kappa_0),$$

uniformly in ℓ, p. Since $\sum_{i=1}^{n} \mathsf{E} A_{ni}(\kappa_0) = \mathrm{O}(n)$ by (4.6) we get, see (4.2),

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

We then check the conditions of Lemma 4.1 :

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

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.5), because for $0 \leq \eta$ and $\nu \geq 1/2$

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

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

(c) *The term* $\max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell}$. Use (4.5) for $\phi = \zeta$ to get

$$\mathsf{E}_{i-1} k_{ni}^{p}(\kappa_\ell, n^{-\zeta}) \leq n^{-\zeta} A_{ni}(\kappa_\ell) \leq n^{-\zeta} \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa), \tag{A.6}$$

uniformly in κ_ℓ and $1 \leq p \leq 2^r$. We then find from (4.6) that

$$\mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} k_{ni}^{p}(\kappa_\ell, n^{-\zeta}) \leq n^{-\zeta} \sum_{i=1}^{n} \mathsf{E}\{\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa)\} = \mathrm{O}(n^{1-\zeta}) = \mathrm{o}(n^{\nu}), \tag{A.7}$$

since $\nu \geq 1/2 > 1 - \zeta$ by (A.5). In particular for $p = 1$ we find $\mathsf{E} \max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell} = \mathrm{o}(n^{\nu})$ so that, by the Chebychev inequality, $\max_{1 \leq \ell \leq L} \overline{\mathcal{M}}_{n2\ell} = \mathrm{o}_{\mathsf{P}}(n^{\nu})$.

(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.7) 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 $\max_{1 \leq \ell \leq L} \widetilde{\mathcal{M}}_{n2\ell} = \mathrm{o}_{\mathsf{P}}(n^{\nu})$. ∎

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 $1 \le \ell_1 \le n+1$

$$M_{n\ell_1} = \sum_{i=1}^{\ell_1 - 1} u_i 1_{(z_i=0)} + \sum_{i=\ell_1}^{n} u_i 1_{(z_{\ell_1} \neq 0, (z_{\ell_1})'_\perp z_i = 0)}, \tag{A.8}$$

and find

$$\sup_{|\delta|=1} |S_n(\delta)| \le \max_{1 \le \ell_1 \le n+1} |M_{n\ell_1}|. \tag{A.9}$$

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

$$M_{n,\ell_1,\ell_2,\dots,\ell_m} = \sum_{k=0}^{m} \sum_{i=\ell_k}^{\ell_{k+1}-1} u_i 1_{\{z_{\ell_1}^{(0)} \neq 0, \dots, z_{\ell_k}^{(k-1)} \neq 0,\ z_i^{(k)} = 0\}}, \tag{A.10}$$

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

$$\max_{|\delta|=1} |S_n(\delta)| \le \max_{1 \le \ell_1 \le \dots \le \ell_m \le n+1} |M_{n\ell_1,\dots,\ell_m}|. \tag{A.11}$$

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 $s_1^\delta = n+1$. For $i < s_1^\delta$ this means that 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 $\sum_{i=1}^{s_1^\delta - 1} u_i 1_{(\delta' z_i = 0)} = \sum_{i=1}^{s_1^\delta - 1} u_i 1_{(z_i=0)}$. For $i = s_1^\delta$ 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 $i \ge s_1^\delta$. 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

$$S_n(\delta) = \sum_{i=1}^{n} u_i 1_{(\delta' z_i = 0)} = \sum_{i=1}^{s_1^\delta - 1} u_i 1_{(z_i=0)} + \sum_{i=s_1^\delta}^{n} u_i 1_{\{z_{s_1^\delta} \neq 0, (z_{s_1^\delta})'_\perp z_i = 0\}}. \tag{A.12}$$

We now compare with the martingales $M_{n_{\ell_1}}$, see (A.8). For all sample paths and all δ we have the evaluation

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

and hence (A.9) 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.12), we then find

$$\sum_{i=1}^{n} u_i 1_{\{\delta' z_i^{(0)}=0\}} = \sum_{i=1}^{s_1^\delta - 1} u_i 1_{\{z_i^{(0)}=0\}} + \sum_{i=s_1^\delta}^{n} u_i 1_{\{z_{s_1^\delta}^{(0)} \neq 0, \delta_1' z_i^{(1)}=0\}}. \tag{A.13}$$

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 $k = 2, \ldots, m$

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

By repeated application of (A.13) we then find, using the notation $s_0^\delta = 1$ and $s_{m+1}^\delta = n+1$,

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

and the inequality, using the martingales defined in (A.10),

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

which proves (A.11). ∎

Proof of Theorem 4.2. We apply Lemma A.1 for $m+1 = \dim z$ to see that we must evaluate the maximum of $|M_{n,\ell_1,\ell_2,\ldots,\ell_m}|$, see (A.10), where the summands

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

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

$$\max_{1 \leq q \leq r} \mathsf{E}(\max_{1 \leq \ell_1 < \cdots < \ell_m \leq n} \sum_{i=1}^{n} \mathsf{E}_{i-1} u_{n\ell_1,\ldots,\ell_m,i}^{2^q}) \leq \max_{1 \leq q \leq r} \sum_{i=1}^{n} \mathsf{E} u_i^{2^q} \leq \sum_{i=1}^{n} \mathsf{E}(1 + u_i^{2^r}) = \mathrm{O}(n),$$

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 $\max_{1 \leq \ell_1 \leq \cdots \leq \ell_m \leq n+1} |M_{n,\ell_1,\ell_2,\ldots,\ell_m}| = \mathrm{o}_{\mathsf{P}}(n^\nu)$. ∎

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 3.2 and 3.3, that the condition for small regressors is satisfied for random walk and stationary regressors.

Proof of Theorem 3.1. (ia) We first prove that

$$\{\lambda_n(\xi) \leq a\} \subset \{F_n(a) \geq \text{int}(n\xi)/n\},\ 0 \leq \xi \leq 1. \tag{A.14}$$

If $\xi < 1/n$, $\text{int}(n\xi)/n = 0$, and $\{F_n(a) \geq \text{int}(n\xi)/n\}$ is the full set, so the relation trivially holds. Consider therefore $1/n \leq \xi \leq 1$. If $\lambda_n(\xi) \leq a$, there exists a non empty subset $\mathbb{S}$ of $(1, \ldots, n)$ with $|\mathbb{S}| = \text{int}(n\xi)$ elements such that $\min_{|\delta|=1} \max_{i\in\mathbb{S}} |z_{ni}'\delta| \leq a$, and therefore, by continuity, a δ with $\delta'\delta = 1$, such that for all $i \in \mathbb{S}$ we have $|z_{ni}'\delta| \leq a$. Hence the number of i for which $|z_{ni}'\delta| \leq a$ must be greater than or equal $|\mathbb{S}| = \text{int}(n\xi)$, that is, $F_{n\delta}(a) \geq \text{int}(n\xi)/n$ and hence $F_n(a) \geq \text{int}(n\xi)/n$. which proves (A.14).

(ib) We next prove

$$\{F_n(a) > \text{int}(n\xi)/n\} \subset \{\lambda_n(\xi) \leq a\}. \tag{A.15}$$

This is obvious for $\xi = 1$, because $\{F_n(a) > \text{int}(n\xi)/n\}$ is empty, so the relation trivially holds. Assume therefore $\xi < 1$. If $F_n(a) = \sup_{|\delta|=1} F_{n\delta}(a) > \text{int}(n\xi)/n$, there is a δ for which $F_{n\delta}(a) > \text{int}(n\xi)/n$ or $\sum_{i=1}^n 1_{(|z_{ni}'\delta|\leq a)} > \text{int}(n\xi)/n$, and therefore the number of i for which $|z_{ni}'\delta| \leq a$ is at least $\text{int}(n\xi)$ and hence we can find $\mathbb{S}$ with $|\mathbb{S}| = \text{int}(n\xi)$ for which $|z_{ni}'\delta| \leq a$ for $i \in \mathbb{S}$. Thus $\lambda_n(\xi) \leq a$, which shows (A.15).

(ii) First, if $\liminf_{n\to\infty} \lambda_n(\xi^*) > 0$ for some $\xi^* \leq 1$, then $\lambda_n(\xi^*) > a_0 > 0$ for some a_0, and all $n \geq n_0$. It then follows from (A.15) that $F_n(a_0) \leq \text{int}(n\xi^*)/n \leq \xi^*$, $n \geq n_0$, and hence $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi^*$ and we define $\xi = \xi^*$. Second, if $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi < 1$ then for $\eta > 0$, there is a_0,n_0 such that for $n \geq n_0$ and $a \leq a_0$, $F_n(a_0) < \xi+\eta < \text{int}\{n(\xi+2\eta)\}/n$. Thus, we choose η so small that $\xi + 2\eta < 1$ and find from (A.14) that $\lambda_n(\xi + 2\eta) > a_0$, so we choose $\xi^* = \xi + 2\eta < 1$, such that $\liminf_{n\to\infty} \lambda_n(\xi^*) \geq a_0$, which proves (ii). ■

Before we prove Theorem 3.2 we show an intermediate result.

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.* $(0, \Phi)$, $j = 1, \ldots, n$, *and assume that the density of* $\gamma' x_i/(i\gamma'\Phi\gamma)^{1/2}$ *is bounded uniformly in* $|\gamma| = 1$ *and* $i = 1, \ldots, n$. *Then the sets*

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

satisfy

$$\mathsf{E}(n^{-1}\sum_{i=1}^n 1_{\mathbb{B}_i})^{m+1} \leq n^{-1} + CM^{m+1}. \tag{A.16}$$

Proof of Lemma A.2. We find

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

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.16).

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

$$(m+1)! \sum_{1\leq i_1<\cdots<i_m\leq n} \mathsf{E}\left\{(\prod_{j=1}^{m} 1_{\mathbb{B}_{i_j}}) \sum_{i_{m+1}=i_m+1}^{n} \mathsf{E}(1_{\mathbb{B}_{i_{m+1}}}|\mathcal{G}_m)\right\}. \tag{A.17}$$

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

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

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

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

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

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

Inserting this into (A.17) we get

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

Repeating the argument we find the result in (A.16). ■

Proof of Theorem 3.2. 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 $F_n(a) = \mathrm{o}_{\mathsf{P}}(1)$ for $(a,n) \to (0,\infty)$, see (2.4).

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 $\delta_\ell, \ell = 1, \ldots, L$, 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

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

where

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

We discuss these in turn.

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

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

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

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 (3.2) 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

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

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

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

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

$$\mathcal{P}_{1n} \leq CL2a(A+1)\frac{1}{n}\sum_{i=1}^{n}(\frac{n}{i})^{1/2} \leq CLAa. \tag{A.18}$$

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

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

The same holds multiplying by $1_{(|x_i|\leq A)}$. 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 (3.2) for $b = \pm a$, $c = a + \eta(A+1) < 1/2$. We find that $(\cos\theta_\ell)^{-1} \leq 2(A+1)$, and therefore

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

say. By Chebychev's inequality

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

From Lemma A.2 with $M = 2\eta(1+A)^2$, we find from (A.16) that

$$\mathcal{P}_{2n} \leq 2L(\frac{4}{\eta})^{m+1}\{n^{-1} + C\eta^{m+1}(1+A)^{2(m+1)}\} \leq CLn^{-1} + CL\eta^{m+1}(1+A)^{2(m+1)}. \tag{A.19}$$

We therefore find from (A.18) and (A.19), using $\eta^{m+1} = L^{-1-1/m}$, that

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

By tightness of $\max_{1\leq i\leq n} |x_i| n^{-1/2}$ we can choose $A > 0$ so large that $\mathsf{P}\{\max_{1\leq i\leq n} |x_i| n^{-1/2} > A\} \leq \epsilon/4$ 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 $CLAa \leq \epsilon/4$, and finally n so large that $CLn^{-1} \leq \epsilon/4$. This proves (2.4) and hence Theorem 3.2. ∎

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 $F_n(a) = \mathrm{o}_{\mathsf{P}}(1)$ for $(a, n) \longrightarrow (0, \infty)$, see (2.4).

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

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

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

$$\mathsf{P}\{\sup_{|\delta|=1} R_{1n\delta}(a) > \eta\} \leq \mathsf{P}\{n^{-1}\sum_{i=1}^{n} 1_{(|x_i|>A)} \geq \eta\} \leq \frac{1}{\eta}\mathsf{P}(|x_1| \geq A), \quad (A.20)$$

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 (3.2) 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

$$\mathsf{E}\{1_{(|z_i'\delta|\leq a, |x_i|\leq A)}|\mathcal{G}_{i-1}\} \leq Ca(A+1), \quad (A.21)$$

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

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

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 (4.5) 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 $|\delta - \tilde{\delta}| \leq Qn^{-\phi}$,

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

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

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

We can then bound

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

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

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

so that (4.5) holds with $A_{ni}(\delta) = C(1+A)^2$, and hence (4.6) holds. Theorem 4.1, case 1, with $\eta = 0$ and $\nu = 1$ now shows

$$\sup_{|\delta|=1} |\frac{1}{n}\sum_{i=1}^n [u_{ni}(\delta) - \mathsf{E}\{u_{ni}(\delta)|\mathcal{G}_{i-1}\}]| = \mathrm{o}_{\mathsf{P}}(1). \tag{A.22}$$

Combining (A.20), (A.21), and (A.22) 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 $\mathsf{P}(\sup_{|\delta|=1} |M_{n\delta}| \geq \eta/3) \leq \epsilon/3$. This proves (2.4) and hence Theorem 3.3. ∎

References

Chen, X.R. and Wu, Y.H. (1988) Strong consistency of M-estimates in linear models. *Journal of Multivariate Analysis* 27, 116–130.

Davies, L. (1990) The asymptotics of S-estimators in the linear regression model. *Annals of Statistics* 18, 1651–1675.

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.

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