Econometric Theory, **35**, 2019, 653–683.
doi:10.1017/S0266466618000257

BOUNDEDNESS OF M-ESTIMATORS FOR LINEAR REGRESSION IN TIME SERIES

SØREN JOHANSEN
University of Copenhagen, and CREATES
Aarhus University

BENT NIELSEN
University of Oxford

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 semicontinuous 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.

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 nonstandardized distributions, see Billingsley (1968). The objective function can be nonconvex and noncontinuous. A prominent example of an estimator with a nonconvex objective function 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 nonconvex. 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 Čižek (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

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. Address correspondence to Bent Nielsen, Nuffield College & Department of Economics, University of Oxford & Programme for Economic Modelling Nuffield College, Oxford, OX1 1NF, UK; e-mail: bent.nielsen@nuffield.ox.ac.uk.

© Cambridge University Press 2018

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $(m+1)$-dimensional parameter $(\mu, \alpha')'$ is the minimizer of the objective function

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

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 nonconvex and nondifferentiable 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 nonconvex 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 includes 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 nonconvex ρ in some situations. Chen and Wu (1988) give two further results on boundedness

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 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 $\mathsf{E}\rho(y_i - \mu - x_{ni}'\alpha)$ 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 article we focus on boundedness of the minimizer. This is for two reasons. Boundedness is typically harder to establish than consistency. 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 lower semicontinuous 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 $h(v) = \mathsf{E}\rho(\varepsilon_i - v)$, 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 deterministics 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 martingale arrays 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 4.2 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 noncontaminated reference distribution P_0, say, and when they are contaminated so that they follow a mixture distribution $(1-\epsilon)\mathsf{P}_0 + \epsilon\mathsf{P}_1$, say. For simplicity we will, however, require that the innovations are identically distributed. This assumption could potentially be relaxed as

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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

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

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 $y_i = \mu + \alpha 1_{(i \le \tau n)} + \varepsilon_i$, the regressor $1_{(i \le \tau n)}$ 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

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

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 $\rho(u) = \frac{1}{2}u^2 1_{(|u| \le c)} + c(|u - c| + \frac{1}{2}c)1_{(|u| > c)}$. 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.

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 martingale arrays indexed by a continuous parameter in a compact set, which is evaluated using a recent martingale result, see

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

Lemma 4.2 or Johansen and Nielsen (2016, Lemma 5.2). The proof requires a moment condition that depends approximately linearly 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 semicontinuity 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_* in (iib) 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.

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

(a) ρ is lower semicontinuous 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. We will consider sequences or rather nets, $s_{a,n}$, indexed by $a > 0$ and $n \in \mathbb{N}$. We write $\lim_{(a,n) \to (0,\infty)} s_{a,n} = s$ if $\forall \epsilon > 0$, $\exists a_0, n_0, \forall a \leq a_0, n \geq n_0$ then $|s_{a,n} - s| < \epsilon$. 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) $\lim_{(a,n) \to (0,\infty)} \mathsf{P}[\sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\} \geq \epsilon] = 0$ 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 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).

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $(0, 1, \ldots, n)/n$ and it is continuous from the right with limits from the left. We show that $F_n(a)$ is a random variable for fixed a and is a discrete distribution function as a function in a and give a detailed analysis of condition (iii) in Section 3.

For now, we give the main result on boundedness.

THEOREM 2.1 (Boundedness). *Under Assumption 1, we can for all $\epsilon > 0$ find $B > 0, n_0 > 0$, 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 on $\mathbb{C}_n$ 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 any measurable minimizer satisfies*

$$|\hat{\beta}| = \mathrm{O_P}(1).$$

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 Example 3.5 and Remark A.6.

The structure of the rest of the article is that we discuss in Section 3 the content of Assumption $1(iii)$ on the frequency of small regressors, and give in Section 4 some results on the supremum of martingale arrays indexed by a continuous parameter. All proofs are collected in the Appendix.

3. THE ASSUMPTION CONCERNING 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 marginal and conditional density, respectively.

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 Section 3.2.

Remark 3.1. Assumption $1(iiib)$ implies that $\mathrm{P}(\sum_{i=1}^n z_{in} z_{in}'$ is positive definite$) \to 1$, 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 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 $\liminf_{n\to\infty} \lambda_n > 0$ 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 3.4. If $F_n(a) = \mathrm{o_P}(1)$ as $(a, n) \to (0, \infty)$, then Assumption $1(iii)$ is satisfied. This is because $F_{n\delta}$ is a 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

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

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

Remark 3.5. 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 3.2 and Example 3.5.

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $(\hat{\mu}, \hat{\alpha}) \to (\mu_0, \alpha_0)$ $a.s.$ 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 article, 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 $0 < \xi < 1$

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

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 $\lambda_n(1) = \min_{|\delta|=1} \max_{1 \le i \le n} |z_{ni}' \delta|$.

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 then that λ_n is a type of inverse for F_n.

THEOREM 3.1. (i) *$F_n(a)$ is a random variable and F_n is a discrete distribution function.*
(ii) *The function λ_n is a right continuous, piecewise constant function satisfying*

$$\{\lambda_n(k/n) \le a\} = \{F_n(a) \ge k/n\}, \text{ for } k = 0, \ldots, n \text{ and } a \ge 0. \tag{3.3}$$

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

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 $q_n^{lower}(\xi) = \inf\{a : F_n(a) \ge \xi\}$ and $q_n^{upper}(\xi) = \inf\{a : F_n(a) > \xi\}$, respectively. For $\xi = k/n$ we have that $\lambda_n(k/n) = q_n^{lower}(k/n)$.

Next, we compare the condition $\liminf_{n\to\infty} \lambda_n(\xi) > 0$ with the condition in Assumption $1(iiib)$ that $\limsup_{(a,n)\to(0,\infty)} F_n(a) \le \xi$, see also Remark 3.5.

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

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.

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

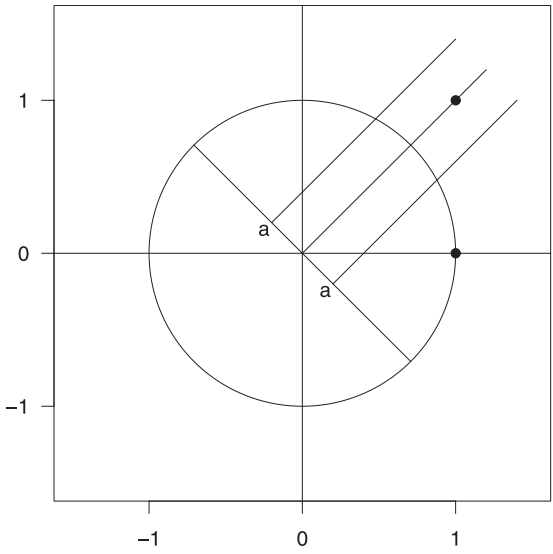


FIGURE 1. Illustration of $F_{n\delta}$ in the dummy variable case.

LEMMA 3.1. *For any* $0 \le c \le 1/2$, $|\theta| < \pi/2$

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

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

Example 3.1
The regression $y_i = \mu + \alpha 1_{(i\le\tau n)} + \varepsilon_i$ for some $0 < \tau < 1$. We show that $z_{ni} = \{1, 1_{(i\le\tau n)}\}'$ 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, $0 \le a < 1/2$ 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 3.4.

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

large n, their distribution can be described by the density $q^{-1}x^{1/q-1}, x \in [0,1]$. For $|z'_{ni}\delta| = |-\sin\theta + (i/n)^q \cos\theta| \le a,\ \cos\theta > 0$, the basic inequality is in all cases

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

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 $(\cos\theta)^{-1} \le 4$. It follows that the length of the interval is bounded by $2na/\cos\theta \le 8na$, such that $F_n(a) \le 8a \to 0$, 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 of θ so $F_{n\delta}(0) > 0$ 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 $F_n(a) \le c\{2a/(1-a^2)^{1/2}\}^{1/q} \to 0$, 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 $\sin\theta = 1/\sqrt{2} - a + o(a)$ and $\cos\theta = \{1-(1/\sqrt{2}-a)^2\}^{1/2}\{1+o(a)\} = 1/\sqrt{2} + o(a)$. This gives the bound $F_n(a) \le c\{(1/\sqrt{2})^{1/q} - (1/\sqrt{2} - 2a)^{1/q}\}/(1/\sqrt{2})^{1/q} \to 0$, 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 $\sin\theta = 1/\sqrt{2} + a + o(a)$ 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

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

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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $0 \leq a < 1$, and $\delta = (0, 1)'$ that

$$F_n(a) \geq F_{\delta n}(a) = n^{-1} \sum_{i=1}^{n} 1_{(|z_{ni}'\delta| \leq a)} = (n-1)/n \to 1,$$

for $(a, n) \to (0, \infty)$.

We find an example where Assumption $1(iii a)$ is not satisfied, but the condition of Davies that $\liminf_{n\to\infty} \lambda_n(\xi^*) > 0$ 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 $1(iii a)$, 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, ... 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 $F_n(0) = \text{int}(n/4)/n \to 1/4$ and Assumption $1(iii b)$ is satisfied, but $\lim_{(a,n)\to(0,\infty)} \sup_{|\delta|=1} \{F_{n\delta}(a) - F_{n\delta}(0)\} = 3/4$ so Assumption $1(iii a)$ is not satisfied. However, $\limsup_{(a,n)\to(0,\infty)} F_n(a) \leq \xi = 3/4$. Thus, Theorem 3.2 shows that a $\xi^* > 0$ exists so that $\liminf_{n\to\infty} \lambda_n(\xi^*) > 0$.

Example 3.6
Explosive regressors. Consider the regression $y_i = \mu + \alpha z_{ni} + \varepsilon_i$ where $z_{ni} = 2^{i-1-n}\sqrt{n}$, $i = 1, \ldots, n$. This is an example where the condition number, see Lai and Wei (1982, Theorem 1), for strong consistency of the least squares estimator, $\hat{\alpha}_{OLS}$ say, fails. It can be shown that $n^{1/2}(\hat{\alpha}_{OLS} - \alpha)$ converges in distribution, if the innovations ε_i are normal using techniques tailored to the explosiveness. We show that the condition to F_n fails in this situation. Most of the regressors z_{ni} are close to zero. In this case $|z_{ni}'\delta| = |-\sin\theta + z_{ni}\cos\theta| \leq a$ implies

$$\frac{-a + \sin\theta}{\cos\theta} \leq z_{ni} = 2^{i-1-n}\sqrt{n} \leq \frac{a + \sin\theta}{\cos\theta}. \tag{3.6}$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

For given $a > 0$, the smallest value of θ that gives an $F_{n\delta}(a) > 0$ satisfies $-a + \sin\theta = z_{ni}\cos\theta$ for $i = 1$, such that $\sin\theta = a + 2^{-n}\sqrt{n} + \mathrm{o}(a + 2^{-n}\sqrt{n})$ and $\cos\theta = 1 + \mathrm{o}(a + 2^{-n}\sqrt{n})$. Insert that θ in the upper bound in (3.6) to get

$$\frac{a + \sin\theta}{\cos\theta} = 2a + 2^{-n}\sqrt{n} + \mathrm{o}(a + 2^{-n}\sqrt{n}).$$

Isolating i in the upper inequality in (3.6) then gives

$$i \le n + 1 + \frac{\log\{2a/\sqrt{n} + 2^{-n} + \mathrm{o}(2a/\sqrt{n} + 2^{-n})\}}{\log 2}.$$

In turn we get

$$F_n(a) \ge F_{n\delta}(a) = 1 + \frac{1}{n} + \frac{\log\{2a/\sqrt{n} + 2^{-n} + \mathrm{o}(2a/\sqrt{n} + 2^{-n})\}}{n\log 2}.$$

Now for any a_0 we get for large n, that

$$F_{n\delta}(a_0) = 1 + \frac{1}{n} + \frac{\log\{2a_0/\sqrt{n} + \mathrm{o}(n^{-1/2})\}}{n\log 2} \to 1.$$

Therefore it is not possible to find a_0, n_0 so that for any $a < a_0$ and $n > n_0$ then $F_n(a) \le \xi$ for some $\xi < 1$. The condition in Assumption $1(iiib)$ then fails.

3.3. Regression with Multiple Stochastic Regressors

For this case we give two examples, where in the 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 $\mathcal{G}_{i-1} = \sigma(x_1, \ldots, x_{i-1})$ be bounded uniformly in $(x_1, \ldots, x_{i-1}, x_i)$ and $|\gamma| = 1$, $\gamma \in \mathbb{R}^m$.*

Then $F_n(a) = \mathrm{o_P}(1)$, for $(a, n) \to (0, \infty)$, 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 Φ is positive definite, and the density of $\gamma' x_i/(i\gamma'\Phi\gamma)^{1/2}$ is bounded uniformly in $|\gamma| = 1$ and $i = 1, \ldots, n$.*

Then $F_n(a) = \mathrm{o_P}(1)$, for $(a, n) \to (0, \infty)$, such that Assumption $1(iii)$ holds.

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $|\gamma| = 1$.

Example 3.8
The assumption on the marginal density in Theorem 3.4 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 $|\gamma| = 1$.

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 result build on the following iterated martingale inequality that can be proved by an iteration of the exponential martingale inequality by Bercu and Touati (2008).

LEMMA 4.1 (Johansen and Nielsen, 2016, Theorem 5.1). *For ℓ, $1 \le \ell \le L$, let $z_{\ell i}$ be $\mathcal{F}_i$-adapted and $\mathsf{E} z_{\ell i}^{2^r} < \infty$ for some $r \in N$. Let $D_p = \max_{1\le\ell\le L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell i}^{2^p}$ for $1 \le p \le r$. Then, for all $\omega_0, \omega_1, \ldots, \omega_r > 0$,*

$$\mathsf{P}\left\{ \max_{1\le\ell\le L} \left| \sum_{i=1}^n (z_{\ell i} - \mathsf{E}_{i-1} z_{\ell i}) \right| > \omega_0 \right\} \le L \frac{\mathsf{E} D_r}{\omega_r} + \sum_{p=1}^r \frac{\mathsf{E} D_p}{\omega_p} + 2L \sum_{p=0}^{r-1} \exp\left(-\frac{\omega_p^2}{14\omega_{p+1}} \right).$$

We use Lemma 4.1 to generalize Theorem 5.2 of Johansen and Nielsen (2016) concerning the maximum of finitely many martingales, to be valid for martingale arrays.

LEMMA 4.2. *Let $\mathcal{F}_i$ be an increasing sequence of σ-fields and let $u_{n\ell i}$ be $\mathcal{F}_i$-adapted with $\mathsf{E}(u_{n\ell i}^{2^r}) < \infty$, $r \in \mathbb{N}$, $\ell = 1, \ldots, L_n$, $i = 1, \ldots, n$, and let $\lambda > 0$ and ς be real numbers defined by*

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

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

Then, if $\nu \ge 0$ is chosen such that

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

it holds that

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

We prove a similar result for a family of martingale arrays indexed by $\kappa \in R^{m+1}$ 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 $Bn^{-\phi}$. For a filtration $\mathcal{F}_i$ and adapted functions $u_{ni}(\kappa)$ with finite expectation, we define the martingale array

$$M_{nk}(\kappa) = \sum_{i=1}^{k} \{u_{ni}(\kappa) - \mathsf{E}_{i-1} u_{ni}(\kappa)\}, \qquad 1 \le k \le n \text{ for } \kappa \in \mathcal{K},$$

and use the notation $M_n(\kappa) = M_{nn}(\kappa)$.

This result has two cases. The first is applied in the proof of Theorem 3.3 for stationary regressors. Both cases will be used for asymptotic analysis of consistency and expansions in a follow-up article.

Below we evaluate the supremum of martingale arrays indexed by a parameter κ in a compact set $\mathcal{K}$. Such a supremum need not be measurable, but we show that on a set with probability one, we can bound it using a measurable function, which is then used in the further analysis.

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 $\kappa_0 \in \mathcal{K}$, and let*

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

Suppose there exist a set $\Omega^\dagger$ with probability one and $\mathcal{F}_i$-measurable random variables $u^\dagger_{ni}(\kappa)$ such that for all $0 \le \phi \le 1$ and $\kappa \in \mathcal{K}$,

$$\sup_{\tilde{\kappa} \in \mathbb{B}(\kappa, B\phi) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})| \le \phi u^\dagger_{ni}(\kappa). \tag{4.5}$$

Furthermore, 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 $1 \le p \le 2^r$ and $\kappa \in \mathcal{K}$,

$$\mathsf{E}_{i-1}\{u^\dagger_{ni}(\kappa)\}^p \le A_{ni}(\kappa). \tag{4.6}$$

Let η, ν satisfy either of

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

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

$$\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} A_{ni}(\kappa) \le A^\dagger_{ni}, \tag{4.7}$$

and for some $C > 0$

$$n^{-1} \sum_{i=1}^{n} \mathsf{E}(A^\dagger_{ni}) \le C \text{ for all } n. \tag{4.8}$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

Then, there exist random variables $M_n^\dagger = o_P(1)$ *such that, on the set* $\Omega^\dagger$,

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

In Theorem 4.1 we take supremum in three places, (4.5), (4.7), and (4.9). The resulting object may not be measurable. We could then continue using outer measures as frequently done in the empirical process literature. However, in many application of this Theorem it is possible to find simple measurable bounds and proceed with standard measure theory. These bounds may not be the smallest possible bounds, but may be sufficient to find the order of magnitude of the terms involved. The notation for the bounds therefore involves a † to indicate that an outer measure argument could be applied.

Lemma 4.2 is also used to prove the next result concerning the supremum of a special class of martingale arrays 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* $\mathsf{E} \sum_{i=1}^n |u_i|^{2^r} = O(n)$. *Then*

$$n^{-\nu} \sup_{|\delta|=1} |\sum_{i=1}^n u_i 1_{(z_{ni}'\delta = 0)}| = o_P(1).$$

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

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

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.

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.

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

REFERENCES

Abadir, K.M. & A. Lucas (2000) Quantiles for t-statistics based on M-estimators and unit roots. *Economic Letters* 67, 131–137.

Bercu, B. & A. Touati (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848–1869.

Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley.

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

Chow, Y.S. (1965) Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics* 36, 552–558.

Čižek, 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., R.A. Maronna, M. Sued, & V.J. Yohai (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. & E.M. Ronchetti (2009) *Robust Statistics*. Wiley.

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.

Johansen, S. & B. Nielsen (2016) Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli* 22, 1131–1183.

Jurečková, J., P.K. Sen, & J. Picek (2012) *Methodological Tools in Robust and Nonparametric Statistics.* 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. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.

Koenker, R. & Z. Xiao (2004) Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99, 775–787.

Lai, T.L. & C.Z. Wei (1982) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10, 154–166.

Liese, F. & I. Vajda (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., D.R. Martin, & V.J. Yohai (2006) *Robust Statistics*: *Theory and Methods.* Wiley.

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

APPENDIX

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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 $h_* = h(\mu_*) = \mathsf{E}\rho(\varepsilon_i - \mu_*) < \rho_* = \inf_{|u| \geq |u_*|} \rho(u)$. 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 $R_n(\beta)$ 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

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

The main idea of the proof is to find a lower bound for $S_n^{(1)}$ and an upper 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 $0 < \mathsf{E}\{\rho(\varepsilon_i - \mu_*)\} < \rho_* = \inf_{|u| \geq |u_*|} \rho(u)$. Choose A_0, a_0 and consider $|\varepsilon_i| \leq A_0$ and $|z_{ni}'\delta| > a_0$, and $\lambda \geq B_0$ where $B_0 = (A_0 + |u_*| + |\mu_*|)/a_0$. Then

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

Hence $\rho(\varepsilon_i - \lambda z_{ni}'\delta - \mu_*) \geq \rho_*$ for $|\varepsilon_i| \leq A_0$, $|z_{ni}'\delta| > a_0$ 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 $|\varepsilon_i| > A_0$ or $|z_{ni}'\delta| \leq a_0$ so that

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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

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

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

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

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 $h_* = \mathsf{E}\rho(\varepsilon_i - \mu_*)$ and introduce martingale differences $m_i = \rho(\varepsilon_i - \mu_*) - h_*$ and the martingale arrays indexed by δ are $M_n(\delta) = n^{-1}\sum_{i=1}^n \{\rho(\varepsilon_i - \mu_*) - h_*\} 1_{(|z'_{ni}\delta|>0)}$. 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

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

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

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

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

$$F^*_{n\delta}(a_0) = (\rho_* - h_*)\{1 - F_{n\delta}(0)\} - \rho_*\{F_{n\delta}(a_0) - F_{n\delta}(0)\}, \tag{A.2}$$

so that when taking infimum and recalling $F_n(0) = \sup_{|\delta|=1} F_{n\delta}(0)$ and $\rho_* > h_*$ 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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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

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

where $m_i = \rho(\varepsilon_i - \mu_*) - h_*$ are i.i.d. Assumption $1(i,iic)$ implies that the first term is $o_P(1)$ by Kolmogorov Law of Large Numbers using Assumption $1(iib)$, and the second term is $o_P(1)$ 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 $n^{-1}\sum_{i=1}^{n} 1_{(|\varepsilon_i|>A_0)} \to \mathsf{P}(|\varepsilon_1| > A_0)$, which is small for large A_0.

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

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

$$\sup_{|\delta|=1}|M_n(\delta)| \leq \eta, \tag{A.3}$$

$$n^{-1}\sum_{i=1}^{n} 1_{(|\varepsilon_i|>A_0)} \leq \eta, \tag{A.4}$$

$$F_n(0) = \sup_{|\delta|=1} F_{n\delta}(0) \leq \xi, \tag{A.5}$$

$$\sup_{|\delta|=1}\{F_{n\delta}(a_0) - F_{n\delta}(0)\} \leq \eta, \tag{A.6}$$

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

(*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 semicontinuous 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 bounded. ■

Remark A.6. The point of Theorems 2.1 and 2.2 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

$$F^*_{n\delta}(a) = \rho_*\{1 - F_{n\delta}(a)\} - h_*\{1 - F_{n\delta}(0)\}, \tag{A.7}$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

see (A.2), 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 $\lim_{(a,n)\to(0,\infty)} \mathsf{P}\{\inf_{|\delta|=1} F^*_{n\delta}(a_0) < \xi^*\} = 0$.

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.

A.2. Proof of Results Regarding the Frequency of Small Regressors

We prove Theorems 3.1 and 3.2 which relate the condition of small regressors in Assumption $1(iii)$ to the condition of Davies (1990) for deterministic regressors. Next we prove Lemma 3.1, which is used in the examples. Finally we show in Theorems 3.3 and 3.4, that the condition for small regressors is satisfied for stationary and random walk regressors.

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

(i) We first prove that F_n is a discrete distribution function. The function of (δ, a) given by $F_{n\delta}(a) = n^{-1}\sum_{i=1}^n 1_{(|z'_{ni}\delta|\le a)}$ 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 $F_{\delta n}(a) \le k/n$ 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 $F_n(a_m) \to F_n(a_0)$. For each a_m we can find δ_m such that

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

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

$$|z'_{ni}\delta_m| \le a_m, i \in S_m \text{ and } |z'_{ni}\delta_m| > a_m, i \notin S_m. \tag{A.8}$$

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 $\delta_{m'''} \to \delta_0$ for some δ_0. Then it follows that

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

To prove the opposite inequality, we find from (A.8) and $m''' \to \infty$, that

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

Therefore

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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

We next prove measurability. The function $z \to 1_{\{|z'\delta|<a\}}$ is lower semicontinuous, hence $(z_1,\ldots,z_n) \to \sum_{i=1}^n 1_{\{|z_i'\delta|<a\}}$ is lower semicontinuous, such that also the function $(z_1,\ldots,z_n) \to \sup_{|\delta|=1} \sum_{i=1}^n 1_{\{|z_i'\delta|<a\}}$ is lower semicontinuous and hence Borel measurable. We define the $\mathcal{F}_n$ measurable function $G_n(a) = \sup_{|\delta|=1} n^{-1} \sum_{i=1}^n 1_{\{|z_{ni}'\delta|<a\}}$, and note that for $\epsilon_k \to 0$ we have the inequalities

$$F_n(a) \le G_n(a+\epsilon_k) \le F_n(a+\epsilon_k).$$

For $k \to \infty$ we use the right continuity of F_n and find

$$F_n(a) \le \lim_{k\to\infty} G_n(a+\epsilon_k) \le \lim_{k\to\infty} F_n(a+\epsilon_k) = F_n(a).$$

Hence $F_n(a)$ is the limit of a sequence of measurable functions and therefore measurable.

This shows that $F_n(a) = \sup_{|\delta|=1} F_{n\delta}(a)$ is measurable and a similar proof shows that $\sup_{|\delta|=1}\{F_{n\delta}(a) - F_{n\delta}(0)\}$ is measurable, see Assumption $1(iii,a,b)$.

(ii) It follows from the definition (3.2) that λ_n is nonincreasing and that $\lambda_n(\xi) = \lambda_n(k/n)$ for $k/n \le \xi < (k+1)/n$, so that λ_n right continuous with left limits.

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

$$nF_n(a) = \sum_{i=1}^n 1_{(|z_{ni}'\delta|\le a)} \ge \sum_{i\in S} 1_{(|z_{ni}'\delta|\le a)} = k.$$

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

$$nF_n(a) = nF_{n\delta}(a) = \sum_{i=1}^n 1_{(|z_{ni}'\delta|\le a)} \ge k.$$

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

(iii) Define the function $q_n(\xi) = \inf\{a : F_n(a) \ge \mathrm{int}(n\xi)/n\}$. Let ξ satisfy $k \le n\xi < k+1$ 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 $\lambda_n(\xi) \le q_n(\xi)$. If we take a sequence $a_m \downarrow q_n(\xi)$ such that $F_n(a_m) \ge k/n$ then part (ii) shows that $\lambda_n(k/n) \le a_m$ and in turn $\lambda_n(\xi) = \lambda_n(k/n) \le a_m$. For $m \to \infty$ then $\lambda_n(\xi) \le q_n(\xi)$.

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

Proof of Theorem 3.2. First, assume $\liminf_{n\to\infty} \lambda_n(\xi^*) > 0$ for some $0 < \xi^* < 1$. Let $k^* = \mathrm{int}(n\xi^*)$, then the right continuity and piecewise constancy of Theorem $3.1(ii)$ show that there are $a_0, n_0 > 0$ such that for $n \ge n_0$, $\lambda_n(k^*/n) = \lambda_n(\xi^*) > a_0 > 0$. It then follows from (3.3) that $F_n(a_0) < k^*/n \le \xi^*$ for $n \ge n_0$, and hence $\limsup_{(a,n)\to(0,\infty)} F_n(a) \le \xi^*$ and we can choose $\xi = \xi^*$.

Second, if $\limsup_{(a,n)\to(0,\infty)} F_n(a) \le \xi$ for some $0 < \xi < 1$, then for each $\eta > 0$, there are a_0, $n_0 > 0$ such that for $n \ge n_0$ and $a \le a_0$, $F_n(a) < \xi + \eta$. Next note that

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

$\text{int}\{n(\xi+2\eta)\}/n \to \xi+2\eta < 1$, 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 $\liminf_{n\to\infty}\lambda_n(\xi+2\eta) \geq a_0$, so we choose $\xi^* = \xi+2\eta < 1$ as desired. ■

Proof of Lemma 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 $\sqrt{1+y^2} \leq 1+|y|$. Furthermore, 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 $1/\cos\theta = \sqrt{1+y^2}$ is bounded by first $1+|y|$ and then $2(1+|x|)$. ■

Theorem 3.3 demonstrates that Assumption $1(iii)$ on the frequency of small regressors holds 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 will be proved in Appendix A.3 using a chaining argument and the iterated martingale inequality in Lemma 4.2.

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 $F_n(a) = o_P(1)$ for $(a,n) \to (0,\infty)$, see (3.1).

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

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

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

$$\mathsf{P}\{\sup_{|\delta|=1} R_{1n\delta}(a) > \eta\} \leq \mathsf{P}\{n^{-1}\sum_{i=1}^{n} 1_{(|x_i|>A)} \geq \eta\} \leq \frac{1}{\eta}\mathsf{P}(|x_1| > A), \tag{A.9}$$

which can be made arbitrarily 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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

Furthermore, 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

$$\mathsf{E}\{1_{(|z_i'\delta|\le a,|x_i|\le A)}|\mathcal{G}_{i-1}\}\le Ca(A+1), \tag{A.10}$$

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 $u_{ni}(\delta)=1_{(|z_i'\delta|\le a,|x_i|\le A)}-1_{(|z_i'\delta_0|\le a,|x_i|\le A)}$ so that $u_{ni}(\delta_0)=0$ and write

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

The first term does not depend on δ and vanishes by the Law of Large Numbers for martingales, see Chow (1965, Theorem 5). 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

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

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

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

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

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

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

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

We can then bound

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

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

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

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

$$\sup_{|\delta|=1}|\frac{1}{n}\sum_{i=1}^{n}[u_{ni}(\delta)-\mathsf{E}\{u_{ni}(\delta)|\mathcal{G}_{i-1}\}]|=o_{\mathsf{P}}(1). \tag{A.11}$$

Combining (A.9), (A.10), and (A.11) we find that for any $\epsilon>0$, we first take A so large that $P(\sup_{|\delta|=1}R_{1n\delta}\ge\eta/3)\le\epsilon/3$, and then a and Q so small that $a+Q(1+A)<1/2$,

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

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 (3.1) and hence Theorem 3.3. ■

Theorem 3.4 demonstrates the F_n condition for random walk 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.1. *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, for* $M > 0$, $a \in \mathbb{R}$, $|\theta| < \pi/2$, *the sets*

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

satisfy

$$\mathsf{E}(n^{-1} \sum_{i=1}^{n} 1_{\mathbb{B}_i})^{m+1} \leq n^{-1} + C M^{m+1}. \tag{A.12}$$

Proof of Lemma A.1. 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.12).

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

$$(m+1)! \sum_{1 \leq i_1 < \cdots < i_m \leq n} \mathsf{E}\left\{ (\prod_{j=1}^{m} 1_{\mathbb{B}_{i_j}}) \sum_{i_{m+1}=1+i_m}^{n} \mathsf{E}(1_{\mathbb{B}_{i_{m+1}}} | \mathcal{G}_m) \right\}. \tag{A.13}$$

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

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

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

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

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

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

Inserting this into (A.13) we get

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

Repeating the argument we find the result in (A.12). ■

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 $F_n(a) = o_{\mathsf{P}}(1)$ for $(a,n) \to (0,\infty)$, see (3.1).

We show $\lim_{(a,n)\to(0,\infty)} \mathsf{P}\{\sup_{|\delta|=1} F_{n\delta}(a) > \epsilon\} \le \epsilon$ for all $\epsilon > 0$. Let $a_0 \le 1/4$ and consider $a \le a_0$. We apply a chaining argument and let $m = \dim x$. We therefore consider $\delta \in \mathbb{R}^{m+1}$ and, for an $\eta > 0$ to be chosen later, cover the m-dimensional surface $\mathcal{K} = \{|\delta| = 1\}$ with $L = O(\eta^{-m})$ 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

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

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

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

where

$$\mathcal{P}_{0n} = \mathsf{P}(\max_{1\le i\le n} n^{-1/2}|x_i| > A),$$

$$\mathcal{P}_{1n} = \sum_{\ell=1}^{L} \mathsf{P}\{F_{n\delta_\ell}(a) > \epsilon/2,\ \max_{1\le i\le n} n^{-1/2}|x_i| \le A\},$$

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

We discuss these in turn.

$\mathcal{P}_{0n}$. Since $\max_{1\le i\le n} n^{-1/2}|x_i|$ converges in distribution, see Billingsley (1968, pp. 90–91), then $\mathcal{P}_{0n}$ tends to zero for $A \to \infty$ uniformly in n.

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

$$\mathcal{P}_{1n} \le \frac{2}{\epsilon} \sum_{\ell=1}^{L} \mathsf{E}\{F_{n\delta_\ell}(a) 1_{(\max_{1\le i\le n} n^{-1/2}|x_i|\le A)}\}$$

$$= \frac{2}{\epsilon} \sum_{\ell=1}^{L} \frac{1}{n} \sum_{i=1}^{n} \mathsf{P}(|z_{ni}'\delta_\ell| \le a,\ \max_{1\le i\le n} n^{-1/2}|x_i| \le A).$$

Write $z_{ni}'\delta_\ell = -\sin\theta_\ell + n^{-1/2}\gamma_\ell' x_i \cos\theta_\ell$ where $\cos\theta_\ell > 0$ and $|\gamma_\ell| = 1$. From (3.4) with $|x|$ replaced by $n^{-1/2}|\gamma_\ell' x_i| \le A$ while $c = a \le a_0 < 1/2$, we find $(\cos\theta_\ell)^{-1} \le$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

$2(1+A)$ and therefore, when dividing by $\cos\theta_\ell$, and leaving out the intersection with $(\max_{1\le i\le n} n^{-1/2}|x_i| \le A)$ we get the further bound

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

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

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

The random variable $(\gamma_\ell' x_i)(i\gamma_\ell'\Phi\gamma_\ell)^{-1/2}$ is assumed to have a bounded density. Since Φ is positive definite, $(\gamma'\Phi\gamma)^{-1}$ is bounded uniformly in $|\gamma|=1$. Thus, the probability that $(\gamma_\ell' x_i)(i\gamma_\ell'\Phi\gamma_\ell)^{-1/2}$ is contained in an interval of length $4a(A+1)n^{1/2}(i\gamma_\ell'\Phi\gamma_\ell)^{-1/2}$, gives the inequality

$$\mathcal{P}_{1n} \le \frac{1}{\epsilon}CLa(A+1)\frac{1}{n}\sum_{i=1}^{n}(\frac{n}{i})^{1/2} \le \frac{1}{\epsilon}CLAa. \tag{A.14}$$

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

$$\begin{aligned}|1_{\{|z_{ni}'\delta_\ell|\le a\}} - 1_{\{|z_{ni}'\delta|\le a\}}| &= |1_{(|z_{ni}'\delta_\ell|\le a)} - 1_{\{|z_{ni}'\delta_\ell + z_{ni}'(\delta-\delta_\ell)|\le a\}}| \\ &\le 1_{\{|z_{ni}'\delta_\ell - a|\le \eta|z_{ni}|\}} + 1_{\{|z_{ni}'\delta_\ell + a|\le \eta|z_{ni}|\}}.\end{aligned}$$

The same holds multiplying by $1_{(|x_i|\le A)}$. Introducing $z_{ni}'\delta_\ell = -\sin\theta_\ell + \cos\theta_\ell(\gamma_\ell' x_i n^{-1/2})$ and the bound $|z_{ni}| \le 1 + |x_i|n^{-1/2} \le 1 + A$, we apply (3.4) for $c = a + \eta(A+1) < 1/2$, so that for $a \le a_0 \le 1/4$ we require $\eta(A+1) < 1/4$. We find that $(\cos\theta_\ell)^{-1} \le 2(A+1)$, and therefore

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

say. By Chebychev's inequality

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

From Lemma A.1 with $M = 2\eta(1+A)^2$, we find from (A.12) that

$$\begin{aligned}\mathcal{P}_{2n} &\le 2L(\frac{4}{\epsilon n})^{m+1}\{n^{-1} + C\eta^{m+1}(1+A)^{2(m+1)}\} \\ &\le \frac{CL}{\epsilon^{m+1}n^{m+1}}\{n^{-1} + \eta^{m+1}(1+A)^{2(m+1)}\}.\end{aligned} \tag{A.15}$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

We therefore find from (A.14) and (A.15), using $L = O(\eta^{-m})$, that

$$\mathsf{P}\{\sup_{|\delta|=1} F_{n\delta}(a) > \epsilon\} \le \mathsf{P}\{\max_{1\le i\le n} |x_i| n^{-1/2} > A\} + C\eta^{-m} Aa + \frac{C}{\epsilon^{m+1} n^{m+1} \eta^m}\{n^{-1} + \eta^{m+1}(1+A)^{2(m+1)}\}.$$

Because $\max_{1\le i\le n} |x_i| n^{-1/2}$ converges in distribution, we can choose $A > 0$ so large that $\mathsf{P}\{\max_{1\le i\le n} |x_i| n^{-1/2} > A\} \le \epsilon/4$ for all n. Next choose η so small that $\eta(1+A) < 1/4$ and a so small that $C\eta^{-m} Aa \le \epsilon/4$. Finally choose n so large that $C\{n^{-1} + \eta^{m+1}(1+A)^{2(m+1)}\}/(\eta^m \epsilon^{m+1} n^{m+1}) \le \epsilon/4$. This proves (3.1) and hence Theorem 3.4. ■

A.3. Proof of Martingale Results

Proof of Lemma 4.2. For each n apply Lemma 4.1 with $z_{\ell i} = u_{n\ell i}$ and $\omega_p = (\omega n^\nu)^{2^p} (28\lambda \log n)^{1-2^p}$ for any $\omega, \lambda > 0$ and $0 \le p \le r$ while $L = L_n$. Note that $\omega_p^2/\omega_{p+1} = 28\lambda \log n$, and that $\mathsf{E} z_{\ell i}^{2^r} = \mathsf{E} u_{n\ell i}^{2^r} < \infty$. Let $D_p = \max_{1\le \ell \le L_n} \sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell i}^{2^p}$. By assumption $\max_{1\le p\le r} \mathsf{E} D_p \le C n^\varsigma$ and $L_n \le C n^\lambda$ for some $C > 0$ and some ς. Lemma 4.1 then gives

$$\mathcal{P}_n = \mathsf{P}\{\max_{1\le \ell \le L_n} |\sum_{i=1}^n (u_{n\ell i} - \mathsf{E}_{i-1} u_{n\ell i})| > \omega n^\nu\}$$

$$\le C\left\{ n^\lambda \frac{n^\varsigma (28\lambda \log n)^{2^r - 1}}{(\omega n^\nu)^{2^r}} + \sum_{p=1}^r \frac{n^\varsigma (28\lambda \log n)^{2^p - 1}}{(\omega n^\nu)^{2^p}} + 2n^\lambda \sum_{p=0}^{r-1} n^{-2\lambda} \right\}.$$

Exploit the conditions (i) $\varsigma < 2\nu$ and (ii) $\varsigma + \lambda < \nu 2^r$, to see that an $\epsilon > 0$ exists so that $\varsigma < 2\nu - \epsilon$ and $\varsigma + \lambda < \nu 2^r - \epsilon$ and in turn

$$\mathcal{P} \le C\left\{ \frac{n^{\nu 2^r - \epsilon} (\log n)^{2^r - 1}}{n^{\nu 2^r}} + \sum_{p=1}^r \frac{n^{2\nu - \epsilon} (\log n)^{2^p - 1}}{n^{\nu 2^p}} + 2rn^{-\lambda} \right\},$$

which vanishes for large n. ■

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

$$1/2 < \zeta < (2^{r-1} - 1)/(m+1). \tag{A.16}$$

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

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

$O\{n^{(\zeta-\eta)(m+1)}\} \to \infty$. 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

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

The variable $\mathcal{R}_{n2}$ may not be measurable, but we will find measurable martingale bounds on the set $\Omega^\dagger$. By the triangular inequality

$$|M_n(\kappa) - M_n(\kappa_\ell)| \le \sum_{i=1}^{n} \{|u_{ni}(\kappa) - u_{ni}(\kappa_\ell)| + \mathsf{E}_{i-1}|u_{ni}(\kappa) - u_{ni}(\kappa_\ell)|\}. \tag{A.17}$$

On $\Omega^\dagger$, we can apply the bound (A.17). Since the summands of the bound are $\mathcal{F}_i$-measurable we can take conditional expectations to get

$$|M_n(\kappa) - M_n(\kappa_\ell)| \le \sum_{i=1}^{n} n^{-\zeta} \{u_{ni}^\dagger(\kappa_\ell) + \mathsf{E}_{i-1} u_{ni}^\dagger(\kappa_\ell)\}.$$

This bound is uniform in $\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}$ and the summands are $\mathcal{F}_i$-measurable, so that

$$\sup_{\kappa \in \mathbb{B}(\kappa_\ell, Bn^{-\zeta}) \cap \mathcal{K}} |M_n(\kappa) - M_n(\kappa_\ell)| \le n^{-\zeta} \sum_{i=1}^{n} \{u_{ni}^\dagger(\kappa_\ell) + \mathsf{E}_{i-1} u_{ni}^\dagger(\kappa_\ell)\}.$$

The variables $u_{ni}^\dagger(\kappa_\ell) - \mathsf{E}_{i-1} u_{ni}^\dagger(\kappa_\ell)$ form an indexed family of martingale difference arrays and we define

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

For $X > 0$ the inequality $|X - \mathsf{E}X| \le (X - \mathsf{E}X) + 2\mathsf{E}X$ implies the bound $\mathcal{R}_{n2} \le n^{-\zeta} \max_{1 \le \ell \le L_n} (\widetilde{\mathcal{M}}_{n2\ell} + 2\overline{\mathcal{M}}_{n2\ell})$. We then prove (4.9) by applying Lemma 4.2 to $M_n(\kappa_\ell)$ in $\mathcal{R}_{n1}$ and $\widetilde{\mathcal{M}}_{n2\ell}$ in $\mathcal{R}_{n2}$, while bounding $\overline{\mathcal{M}}_{n2\ell}$.

(*b*) *The term* $\max_{1 \le \ell \le L_n} |M_n(\kappa_\ell)| = \mathcal{R}_{n1}$.

We apply Lemma 4.2 to $M_n(\kappa_\ell)$ with $u_{n\ell i} = u_{ni}(\kappa_\ell)$. We define $\lambda = (\zeta - \eta)(m+1)$ in (4.1) and argue that $\varsigma = 1 - \eta$ in (4.2).

We start by finding a bound for $u_{ni}(\kappa_\ell)$ that is uniform in ℓ. Note first that

$$\max_{1 \le \ell \le L_n} |u_{ni}(\kappa_\ell)| \le \sup_{\kappa \in \mathbb{B}(\kappa_0 Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa)|.$$

On $\Omega^\dagger$ we apply the bound (4.5) with $\kappa = \kappa_0, \tilde{\kappa} = \kappa$ and $\phi = n^{-\eta}$ noting $u_{ni}(\kappa_0) = 0$ to get

$$\sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa)| = \sup_{\kappa \in \mathbb{B}(\kappa_0, Bn^{-\eta}) \cap \mathcal{K}} |u_{ni}(\kappa) - u_{ni}(\kappa_0)| \le n^{-\eta} u_{ni}^\dagger(\kappa_0),$$

which is measurable. We note that this bound is uniform in ℓ and $\mathcal{F}_i$-measurable. From (4.6) and (4.7) we find, for $1 \le p \le 2^r$, the $\mathcal{F}_{i-1}$-measurable bound:

$$\mathsf{E}_{i-1}\{u_{ni}^\dagger(\kappa_0)\}^p \le A_{ni}(\kappa_0) \le A_{ni}^\dagger,$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

where $\sum_{i=1}^n \mathsf{E}A_{ni}^\dagger \le Cn$ by (4.8). Recalling $u_{n\ell i} = u_{ni}(\kappa_\ell)$ and $\varsigma = 1-\eta$ we get, see (4.2),

$$\max_{1\le p\le 2^r} \mathsf{E}\Big(\max_{1\le \ell\le L_n} \sum_{i=1}^n \mathsf{E}_{i-1}|u_{n\ell i}|^p\Big) \le n^{-\eta}\sum_{i=1}^n \mathsf{E}A_{ni}^\dagger \le Cn^{1-\eta} = \mathrm{O}(n^\varsigma).$$

We then check the conditions of Lemma 4.2:

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

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.16), because for $0 \le \eta$ and $\nu \ge 1/2$

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

Applying (4.3) of Lemma 4.2, we get $\max_{1\le\ell\le L_n}|M_n(\kappa_\ell)| = \mathcal{R}_{n1} = \mathrm{o}_{\mathsf{P}}(n^\nu)$ in both cases.

(*c*) *The term* $\max_{1\le\ell\le L_n}\overline{\mathcal{M}}_{n2\ell} = \mathrm{o}_P(n^{\nu+\zeta})$. Use (4.6), (4.7) to get

$$\mathsf{E}_{i-1}\{u_{ni}^\dagger(\kappa_\ell)\}^p \le A_{ni}(\kappa_\ell) \le \sup_{\kappa\in\mathbb{B}(\kappa_0, Bn^{-\eta})\cap\mathcal{K}} A_{ni}(\kappa) \le A_{ni}^\dagger \tag{A.18}$$

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

$$\mathsf{E}\max_{1\le\ell\le L_n}\sum_{i=1}^n \mathsf{E}_{i-1}\{u_{ni}^\dagger(\kappa_\ell)\}^p \le \sum_{i=1}^n \mathsf{E}(A_{ni}^\dagger) = \mathrm{O}(n) = \mathrm{o}(n^{\nu+\zeta}) \tag{A.19}$$

since $\nu \ge 1/2 > 1-\zeta$ by (A.16) so that $\nu+\zeta > 1$. In particular for $p=1$ we find $n^{-\zeta}\mathsf{E}\max_{1\le\ell\le L_n}\overline{\mathcal{M}}_{n2\ell} = \mathrm{o}(n^\nu)$ so that, by the Markov inequality, $n^{-\zeta}\max_{1\le\ell\le L_n}\overline{\mathcal{M}}_{n2\ell} = \mathrm{o}_{\mathsf{P}}(n^\nu)$.

(*d*) *The term* $\max_{1\le\ell\le L_n} n^{-\zeta}\widetilde{\mathcal{M}}_{n2\ell}$. We apply Lemma 4.2 to $\widetilde{\mathcal{M}}_{n2\ell}$ using $u_{n\ell i} = n^{-\zeta}u_{ni}^\dagger(\kappa_\ell)$. Due to (A.19) we can choose $\lambda = (\zeta-\eta)(m+1)$ and $\varsigma = 1-\zeta$. Noting that $\zeta > 1/2 > \eta$ then $\varsigma = 1-\zeta \le 1-\eta$, which was the value of ς chosen in item (b). Since λ is the same as in (b), the conditions of Lemma 4.2 are satisfied as in (b) so that $n^{-\zeta}\max_{1\le\ell\le L_n}\widetilde{\mathcal{M}}_{n2\ell} = \mathrm{o}_{\mathsf{P}}(n^\nu)$. ■

Proof of Theorem 4.2. For notational convenience we replace z_{ni} by z_i in the proof. Let $S_n(\delta) = \sum_{i=1}^n u_i 1_{(z_i'\delta=0)}$. We show that $\max_{|\delta|=1}|S_n(\delta)| = \mathrm{o}_{\mathsf{P}}(n^{-\nu})$ for $\nu > 1/2$.

Proof of measurability of $\max_{|\delta|=1}|S_n(\delta)|$*:* Let $1 \le d \le m$. For $i_1 < \cdots < i_d$ let $(z_{i_1},\ldots,z_{i_d}) \subset \{z_1,\ldots,z_n\} = K$, say, denote d linearly independent vectors. We will prove the identity

$$\max_{|\delta|=1}|S_n(\delta)| = \max_{0\le d\le m}\max_{(z_{i_1},\ldots,z_{i_d})\subset K}\Big|\sum_{i=1}^n u_i 1_{\{z_i\in \mathrm{sp}(z_{i_1},\ldots,z_{i_d})\}}\Big|, \tag{A.20}$$

where $d=0$ is interpreted as $\sum_{i=1}^n u_i 1_{\{z_i=0\}}$. The function on the right hand side of (A.20) is measurable, since the maximum is taken over finitely many sums, each of which is measurable.

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

We prove $\leq$ in (A.20). Define for a given $\delta \in \mathbb{R}^{m+1}$, $|\delta| = 1$, the subspace

$$V_\delta = \mathrm{sp}(z_i \in K \text{ for which } z_i'\delta = 0),$$

which has dimension $d = \dim(V_\delta)$ for some $d = 0, \ldots, m$. Choose a basis for V_δ of linearly independent vectors $(z_{i_1}, \ldots, z_{i_d})$, such that $\mathrm{sp}(z_{i_1}, \ldots, z_{i_d}) = V_\delta$, and express $S_n(\delta)$ as

$$S_n(\delta) = \sum_{i=1}^{n} u_i 1_{(z_i'\delta=0)} = \sum_{i=1}^{n} u_i 1_{(z_i \in V_\delta)} = \sum_{i=1}^{n} u_i 1_{\{z_i \in \mathrm{sp}(z_{i_1},\ldots,z_{i_d})\}}.$$

This construction applies for each δ so that $\leq$ in (A.20) follows.

We prove $\geq$ in (A.20). Take any set of linearly independent vectors with $i_1 < \cdots < i_d$ such that $(z_{i_1},\ldots,z_{i_d}) \subset K$. Then there is a $\delta_0 \in \mathbb{R}^{m+1}$ with $|\delta_0| = 1$ for which $z'\delta_0 = 0$ for $z \in \mathrm{sp}(z_{i_1},\ldots,z_{i_d})$ and $z_k'\delta_0 \neq 0$ for all $z_k \in K$, which do not belong to $\mathrm{sp}(z_{i_1},\ldots,z_{i_d})$. This construction applies to any selection of d vectors, so that $\geq$ in (A.20) follows.

Proof of martingale bound for $\max_{|\delta|=1} |S_n(\delta)|$*:* Let $0 \leq d \leq m$, and define for any increasing sequence of deterministic indices ℓ_i, so that $1 \leq \ell_1 < \cdots < \ell_d = n$, the martingale array, using $\ell_0 = 1$ and $\ell_{d+1} = n+1$,

$$M_{n,d,\ell_1,\ell_2,\ldots,\ell_d} = \sum_{k=0}^{d} \sum_{\ell_k \leq i < \ell_{k+1}} u_i 1_{\{z_i \in \mathrm{sp}(z_{\ell_1},\ldots,z_{\ell_k})\}}, n = 1, 2, \ldots, \tag{A.21}$$

where $1_{\{z_i \in \mathrm{sp}(\varnothing)\}}$ is interpreted as $1_{(z_i=0)}$. An equivalent expression is

$$M_{n,d,\ell_1,\ell_2,\ldots,\ell_d} = \sum_{1 \leq i < \ell_1} u_i 1_{(z_i=0)} + \sum_{\ell_1 \leq i < \ell_2} u_i 1_{\{z_i \in \mathrm{sp}(z_{\ell_1})\}} + \cdots + \sum_{\ell_d \leq i \leq n} u_i 1_{\{z_i \in \mathrm{sp}(z_{\ell_1},\ldots,z_{\ell_d})\}}.$$

The process $M_{n,d,\ell_1,\ldots,\ell_d}$, $n = 1, 2, \ldots$, is a martingale array. We prove that

$$\max_{|\delta|=1} |S_n(\delta)| \leq \max_{0 \leq d \leq m} \max_{1 \leq \ell_1 < \cdots < \ell_d \leq n} |M_{n,d,\ell_1,\ldots,\ell_d}|. \tag{A.22}$$

To see this, define the subspaces

$$V_{\delta,i} = \mathrm{sp}(z_j \in K \text{ for which } z_j'\delta = 0 \text{ and } j \leq i), i = 1, \ldots, n$$

with dimension $d_i = \dim(V_{\delta,i})$ and $0 \leq d_1 \leq \cdots \leq d_n = d = \dim(V_{\delta,n})$. Define the stopping time, $s_{\delta,k}$, $k = 1, \ldots, d$, as the subscript, where the dimension of $\mathrm{sp}(z_1, \ldots, z_i)$ changes from $k-1$ to k, that is,

$$s_{\delta,k} = \min(i : d_i = k).$$

Note that $s_{\delta,k}$ is a stopping time because the event $(s_{\delta,k} = i)$ is $\mathcal{F}_{i-1}$-adapted since z_i is $\mathcal{F}_{i-1}$-adapted. For completeness define $s_{\delta,0} = 0$ and $s_{\delta,d+1} = n+1$. This gives the representation

$$S_n(\delta) = \sum_{i=1}^{n} u_i 1_{(z_i'\delta=0)} = \sum_{k=0}^{d} \sum_{s_{\delta,k} \leq i < s_{\delta,k+1}} u_i 1_{\{z_i \in \mathrm{sp}(z_j:\, j=s_{\delta,1},\ldots,s_{\delta,k})\}}. \tag{A.23}$$

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press

Comparing (A.23) and (A.21), it is seen that for any given outcome, one can define the increasing sequence $\ell_k = s_{\delta,k}$ such that

$$S_n(\delta) = \sum_{k=0}^{d} \sum_{s_{\delta,k} \le i < s_{\delta,k+1}} u_i 1_{\{z_i \in \mathrm{sp}(z_j:\, j=s_{\delta,1},\ldots,s_{\delta,k})\}} = \sum_{k=0}^{d} \sum_{\ell_k \le i < \ell_{k+1}} u_i 1_{\{z_i \in \mathrm{sp}(z_j:\, j=\ell_1,\ldots,\ell_k)\}} = M_{n,d,\ell_1,\ell_2,\ldots,\ell_d},$$

such that

$$|S_n(\delta)| \le \max_{0\le d\le m} \max_{1\le \ell_1 < \cdots < \ell_d \le n} |M_{n,d,\ell_1,\ldots,\ell_d}|$$

and the same holds for supremum over $|\delta| = 1$, which proves (A.22).

Proof of $\max_{|\delta|=1} |S_n(\delta)| = o_P(n^\nu)$: From (A.22) it is seen that we need to evaluate the maximum of the absolute value of the martingale arrays $|M_{n,d,\ell_1,\ell_2,\ldots,\ell_d}|$, see (A.21), using Lemma 4.2. The number of martingales is $L_n = O(n^m)$ and we choose $\lambda = m \in \mathbb{N}$, see (4.1) and for $1 \le q \le r$, we choose

$$u_{n\ell i}^{2q} = u_i^{2q} 1_{\{z_i \in \mathrm{sp}(z_j:\, j=\ell_1,\ldots,\ell_k)\}} \le u_i^{2q} \le 1 + u_i^{2r}.$$

Then,

$$\max_{1\le q\le r} \mathsf{E}[\max_{0\le d\le m} \max_{1\le \ell_1 < \cdots < \ell_d \le n} \sum_{k=0}^{d} \sum_{\ell_k \le i < \ell_{k+1}} \mathsf{E}_{i-1} u_i^{2q} 1_{\{z_i \in \mathrm{sp}(z_j:\, j=\ell_1,\ldots,\ell_k)\}}]$$

$$\le \sum_{i=1}^{n} \mathsf{E}(1 + u_i^{2r}) = O(n),$$

so that $\varsigma = 1$, see (4.2). We then apply Lemma 4.2 and find for $\nu > 1/2$, that $1 = \varsigma < 2\nu$ and $m + 1 = \varsigma + \lambda < \nu 2^r$ by assumption, so that $\max_{0\le d\le m} \max_{1\le \ell_1 < \cdots < \ell_d \le n} |M_{n,d,\ell_1,\ldots,\ell_d}| = o_P(n^\nu)$. ■

https://doi.org/10.1017/S0266466618000257 Published online by Cambridge University Press