

Asymptotic analysis of the Forward Search

Søren Johansen¹ & Bent Nielsen²

10 February 2013

Summary: The *Forward Search* is an iterative algorithm concerned with detection of outliers and other unsuspected structures in data. This approach has been suggested, analysed and applied for regression models in the monograph Atkinson and Riani (2000). An asymptotic analysis of the Forward Search is made. The argument involves theory for a new class of weighted and marked empirical processes, quantile process theory, and a fixed point argument to describe the iterative element of the procedure.

Keywords: Fixed point result, Forward Search, quantile process, weighted and marked empirical process.

1 Introduction

The *Forward Search* is concerned with detecting outliers and other unsuspected structures in data. This approach has been suggested, analysed and applied for regression models in the monograph Atkinson and Riani (2000), see also Atkinson, Riani and Cerioli (2010a) for a recent overview, while R and matlab code is freely available from www.riani.it. Riani, Atkinson and Cerioli (2009) discuss the application of the Forward Search to multivariate location-scale models. So far formal asymptotic analysis has not been undertaken and inferential procedures are relying on a calibrated distribution approximation, see Riani and Atkinson (2007). In the following we will provide an asymptotic analysis of the Forward Search. The analysis is conducted under the hypothesis that all observations follow a regression model, allowing for stationary as well as stochastically and deterministically trending regressors. The results can therefore be used for computing the confidence bands needed in the Forward Search.

The Forward Search involves an iteration combined with a comprehensive graphical representation of the iteration results. The iteration starts from a robust regression estimator in order to select a small set of m_0 observations without outliers. Examples of robust regression estimators are the least median squares estimator or the least trimmed squares estimator of Rousseeuw (1984). These estimators are known to have good breakdown properties, see Rousseeuw and Leroy (1987, §3.4), and an asymptotic theory for the least trimmed squares regression estimator is provided by Víšek (2006a,b,c). In the iteration step the regression estimator is based on a set of m observations. Using this estimator the residuals are computed for all observations and the set with the $m + 1$ smallest absolute residuals is then found. The largest of these $m + 1$ residuals is the forward residual. The fully iterated estimator is the full sample least squares regression estimator. We derive the asymptotic distribution of the resulting sequence of forward residuals.

A single step of the algorithm has been discussed for the location-scale case by Johansen and Nielsen (2010). The aim of this paper is to discuss the full iterative procedure. Broadly speaking, this requires three asymptotic tools. First, a theory for weighted and marked empirical

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

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

processes to describe the least squares statistics. Secondly, quantile processes to describe the forward residuals. Thirdly, a fixed point result to describe the iteration involved.

In the empirical process theory the weights represent functions of the regressors and the marks are functions of the regression error. The results generalise those of Johansen and Nielsen (2009) who did not allow variation in the quantiles and those of Koul and Ossiander (1994) who did not allow marks. The proof combines a chaining argument with an exponential inequality for martingales by Bercu and Touati (2008).

The quantile process theory draws on the exposition of Csörgő (1983). It is found that in the case of a known variance the forward residuals satisfy a Bahadur representation, so that, asymptotically, the forward residuals have the same distribution as the order statistics of the regression innovations. When the variance is estimated an additional term appears in the asymptotic distribution.

The last ingredient is a fixed point result to describe the iterative result. Each step of the iteration is an L-estimator defined from a Huber-skip objective function which assigns 0-1 weights to the observations. That is, sample quantiles are used to classify observations and least squares regression is performed for the observations not classified as outliers. Starting with Bickel (1975) there are a number of asymptotic 1-step results for L- and M-estimators. These are predominantly concerned with objective functions that have continuous derivatives, thereby excluding the Huber-skip function. One-step Huber-skip estimators have been studied by Rupert and Carroll (1980) and Johansen and Nielsen (2009, 2010). There appears to be less work on iteration of 1-step estimators. The case of smooth weights was considered by Dollinger and Staudte (1991), but the case of 0-1 weights does not appear to have been studied until recently. Cavaliere and Georgiev (2011) analysed a sequence of Huber-skip estimators for a first order autoregression with infinite variance errors, while Johansen and Nielsen (2011) analysed sequences of one-step Huber-skip M-estimators with a fixed critical value. Here, we generalise the fixed point result of the latter paper.

Outline of the paper: The model and the Forward Search algorithm are described in §2. The main asymptotic results are given in §3. The marked and weighted empirical process results are given in §4 with proofs following in §5, 6, 7. The proofs of the main results follow in §8.

2 Model and Forward Search algorithm

The regression model and the Forward Search algorithm are outlined at first. In order to motivate the subsequent asymptotic results, the Bahadur representation for quantile process is reviewed. The currently used calibrated distribution approximation to the Forward Search is then discussed.

2.1 Model

Suppose data (y_i, x_i) , $i = 1, \dots, n$ are available, satisfying the regression equation

$$y_i = x_i' \beta + \varepsilon_i. \quad (2.1)$$

The errors are assumed independent and identically distributed with scale σ so that ε_i/σ has known density f and distribution function $F(c) = \mathbf{P}(\varepsilon_i \leq \sigma c)$.

In practice, the distribution F will often be standard normal or at least symmetric. When discussing some general empirical result in §4-7 a more general distribution is allowed, otherwise f is symmetric. In the symmetric case the distribution function of the absolute errors $|\varepsilon_i|/\sigma$ is $G(c) = 2F(c) - 1$ with density $g(c) = 2f(c)$ and quantiles

$$c_\psi = G^{-1}(\psi) = F^{-1}\{(1 + \psi)/2\}. \quad (2.2)$$

2.2 Forward Search algorithm

The Forward Search is designed to detect outliers. It evolves around an algorithm initiated by a robust estimator. The algorithm generates a sequence of regression estimators, which will be the object of interest here. The $(m + 1)$ th step of the algorithm is given as follows.

Algorithm 2.1 *Generating a sequence of estimators.*

1. Given an estimator $\hat{\beta}^{(m)}$ compute residuals $\hat{\xi}_i^{(m)} = |y_i - x_i' \hat{\beta}^{(m)}|$.
2. Find the $(m + 1)$ -th smallest order statistics $\hat{z}^{(m)} = \hat{\xi}_{(m+1)}^{(m)}$.
3. Find set of $(m + 1)$ observations with smallest residuals $S^{(m+1)} = (i : \hat{\xi}_i^{(m)} \leq \hat{z}^{(m)})$.
4. Compute a new least squares statistic on $S^{(m+1)}$

$$\hat{\beta}^{(m+1)} = \left(\sum_{i \in S^{(m+1)}} x_i x_i' \right)^{-1} \left(\sum_{i \in S^{(m+1)}} x_i y_i \right). \quad (2.3)$$

The algorithm is initiated with an estimator $\hat{\beta}^{(m_0)}$. We will allow initial estimators $\hat{\beta}^{(m_0)}$ converging at a rate slower rate than the usual $n^{1/2}$ -rate - for the stationary case. An example is the Least Median Squares estimator which is $n^{1/3}$ -consistent in location-scale models, see Rousseeuw (1984). Applying the algorithm repeatedly results in sequences of order statistics $\hat{z}^{(m)}$, least squares estimators $\hat{\beta}^{(m)}$ as well as residual variances

$$(\hat{\sigma}^{(m)})^2 = \frac{1}{m} \sum_{i \in S^{(m)}} (y_i - x_i' \hat{\beta}^{(m)})^2 = \frac{1}{m} \left\{ \sum_{i \in S^{(m)}} y_i^2 - \sum_{i \in S^{(m)}} y_i x_i' \left(\sum_{i \in S^{(m)}} x_i x_i' \right)^{-1} \left(\sum_{i \in S^{(m)}} x_i y_i \right) \right\},$$

for $m = m_0 + 1, \dots, n$ along with the scaled forward residuals

$$\frac{\hat{z}^{(m)}}{\hat{\sigma}^{(m)}} = \frac{\hat{\xi}_{(m+1)}^{(m)}}{\hat{\sigma}^{(m)}}.$$

The plots of $\hat{\beta}^{(m)}$ and $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$ against m are *forward plots* of the estimator and of the scaled forward residuals, respectively, see Atkinson and Riani (2000, p.12). The primary objective of this paper is to derive the asymptotic distribution of the forward plot of scaled residuals $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$.

In the application the forward plot of for instance $\hat{z}^{(m)}$ is a process on $m = m_0 + 1, \dots, n$. It is useful to embed it in the space $D[0, 1]$ of right continuous process on $[0, 1]$ with limits from the left, endowed with the uniform norm since all limiting processes will be continuous. Thus, define

$$\hat{z}_\psi = \begin{cases} \hat{z}^{(m)} & \text{for } m = \text{int}(n\psi) \text{ and } m_0/n \leq \psi \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

Embed in a similar way $\hat{\beta}^{(m)}$, $\hat{\sigma}^{(m)}$ as $\hat{\beta}_\psi$, $\hat{\sigma}_\psi$.

We will show consistency of the bias corrected variance estimator

$$\hat{\sigma}_{\psi, \text{cor}}^2 = \hat{\sigma}_\psi^2 / \varsigma_\psi^2, \quad (2.5)$$

with a bias correction factor introduced in Johansen and Nielsen (2010), that is

$$\varsigma_\psi^2 = \frac{\tau_\psi}{\psi}, \quad \text{where } \tau_\psi = 2 \int_0^{c_\psi} \epsilon^2 f(\epsilon) d\epsilon. \quad (2.6)$$

When $f = \varphi$ is normal then $\tau_\psi = \psi - 2c_\psi \varphi(c_\psi)$.

2.3 Some known results from the theory of quantile processes

Johansen and Nielsen (2010, Theorems 5.1–5.3) analysed a single step of the Forward Search applied in a location-scale setting. The results show that the one-step version of the scaled residuals $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$ has an asymptotic representation involving an empirical process and a term arising from estimation error for the variance. The subsequent analysis shows how this result generalises to a fully iterated Forward Search. An interesting feature of these results is that in the case where the variance σ is known, these results reduce to a Bahadur representation linking a quantile process and empirical process. It is useful to recall this theory.

Introduce the empirical distribution function of the absolute errors, $|\varepsilon_i|/\sigma$, that is

$$\widehat{\mathbb{G}}_n(c) = \frac{1}{n} \sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma c)}. \quad (2.7)$$

The first result gives the asymptotic distribution of the empirical process

$$\mathbb{G}_n(c_\psi) = n^{1/2} \{ \widehat{\mathbb{G}}_n(c_\psi) - \psi \}$$

Theorem 2.1 *Billingsley (1968, Theorem 16.15). Let \mathbb{B} be a Brownian bridge so that $\mathbb{B}(\psi)$ is $N\{0, \psi(1-\psi)\}$ -distributed. Then, it holds $\mathbb{G}_n \xrightarrow{D} \mathbb{B}$ on $D[0, 1]$.*

The empirical quantiles of the absolute errors, $|\varepsilon_i|/\sigma$, are defined as

$$\hat{c}_\psi = \widehat{\mathbb{G}}_n^{-1}(\psi) = \inf\{c : \widehat{\mathbb{G}}_n(c) \geq \psi\}. \quad (2.8)$$

Empirical quantiles and empirical distribution functions are linked as follows.

Theorem 2.2 *Csörgő (1983, Corollaries 6.2.1, 6.2.2). Suppose that f is symmetric, differentiable, positive for $F^{-1}(0) < c < F^{-1}(1)$, satisfying $\gamma = \sup_{c>0} F(c)\{1-F(c)\}|f'(c)|/\{f(c)\}^2 < \infty$, and decreasing for large c . Then, for all $\zeta > 0$, it holds*

- (a) $\sup_{0 \leq \psi \leq 1} |2f(c_\psi)n^{1/2}(\hat{c}_\psi - c_\psi) + n^{1/2}\{\widehat{\mathbb{G}}_n(c_\psi) - \psi\}| = o_{\mathbb{P}}(n^{\zeta-1/4});$
- (b) $\sup_{0 \leq \psi \leq 1} |2f(c_\psi)n^{1/2}(\hat{c}_\psi - c_\psi) - n^{1/2}\{\mathbb{G}(\hat{c}_\psi) - \psi\}| = o_{\mathbb{P}}(n^{\zeta-1/2});$
- (c) $\sup_{0 \leq \psi \leq 1} |n^{1/2}\{\mathbb{G}(\hat{c}_\psi) - \psi\} + n^{1/2}\{\widehat{\mathbb{G}}_n(c_\psi) - \psi\}| = o_{\mathbb{P}}(n^{\zeta-1/4}).$

The result in Theorem 2.2(a) shows that the empirical quantile \hat{c}_ψ satisfies, for $0 < \psi < 1$,

$$n^{1/2}(\hat{c}_\psi - c_\psi) = \frac{1}{2f(c_\psi)} n^{1/2} \{ \psi - \widehat{\mathbb{G}}_n(c_\psi) \} + o_{\mathbb{P}}(1).$$

This is known as the Bahadur (1966) representation. The results in parts (b, c) combine to that of (a) and were studied by Kiefer (1967). More detail can be found in Csörgő (1983) who also gives almost sure, logarithmic rates.

Some weighted versions of the above results are also needed.

Theorem 2.3 *(Shorack 1979, Csörgő, 1983, Theorem 5.1.1). Let the function q_ψ be symmetric about $\psi = 1/2$ (it suffices if q_ψ is bounded below by such a function), such that on $0 \leq \psi \leq 1/2$ then q_ψ is increasing and continuous, and satisfies $q_\psi = \{\psi \log \log(1/\psi)\}^{1/2} g_\psi$ for a function g_ψ so $\lim_{\psi \rightarrow 0} g_\psi = \infty$. Then, a probability space exists on which one can define a Brownian bridge \mathbb{B}_n for each n , so that*

- (a) $\sup_{0 \leq \psi \leq 1} |\{\mathbb{G}_n(c_\psi) - \mathbb{B}_n(\psi)\}/q_\psi| = o_{\mathbb{P}}(1);$
- (b) $\sup_{1/(n+1) \leq \psi \leq n/(n+1)} |\{f(c_\psi)n^{1/2}(\hat{c}_\psi - c_\psi) - \mathbb{B}_n(\psi)\}/q_\psi| = o_{\mathbb{P}}(1)$ provided the assumptions of Theorem 2.2 hold.

In Theorem 2.3 a possible choice of q_w is $\{\psi(1 - \psi)\}^\alpha$ for $\alpha < 1/2$, which will be used in the proof of the main Theorem. Finally, a continuity property of the Brownian bridge is needed.

Theorem 2.4 (*Revuz and Yor, 1998, Theorem I.2.2*) *A Brownian motion \mathbb{W} is locally Hölder continuous of order α for all $\alpha < 1/2$. That is*

$$\sup_{0 \leq \psi < \psi^\dagger \leq 1} \frac{|\mathbb{W}(\psi^\dagger) - \mathbb{W}(\psi)|}{(\psi^\dagger - \psi)^\alpha} \stackrel{a.s.}{<} \infty.$$

Thus, for a Brownian bridge \mathbb{B} then $\lim_{\psi \rightarrow 0} \mathbb{B}(\psi)/\psi^\alpha = 0$ a.s.

2.4 A calibrated distribution approximation

Riani and Atkinson (2007) presented a distribution approximation to the deletion residuals $\hat{r}^{(m)}$ based on ordered \mathbf{t} -variates. Through simulation they show that this gives a rather good approximation in finite samples. However, we will show that this approximation is not capturing the asymptotic distribution.

The distribution approximation is derived by a heuristic argument, which is potentially correct up to an approximation. It approximates the distribution of $\hat{r}^{(m)}$ by the distribution of $\hat{v}^{(m)}$, which is the $(m + 1)$ th quantile of a sample of n scaled, absolute $\mathbf{t}_{m-\dim x}$ variables. To get a handle on the asymptotic distribution of $\hat{v}^{(m)}$ consider first the $(m + 1)$ -smallest order statistic, $\hat{w}^{(m)}$ say, from n draws of absolute standard normal variables. This satisfies

$$2\varphi(c_{m/n})n^{1/2}(\hat{w}^{(m)} - c_{m/n}) \xrightarrow{D} \mathbf{N}\{0, \psi(1 - \psi)\},$$

for $m \sim \psi n$ and $c_\psi = \mathbf{G}^{-1}(\psi)$ due to Theorems 2.1, 2.2(a). The absolute standard normal variables have distribution function $2\Phi(y) - 1$. For the \mathbf{t} -order statistic $\hat{v}^{(m)}$ it is useful to Edgeworth expand $\mathbf{P}(\mathbf{t}_{m-\dim x} \leq y) = 2\{\Phi(y) + O(n^{-1})\} - 1$, which indicates that the same asymptotic distribution arises as in the normal case. A more formal argument will keep track of the remainder terms. The starting point could be the expression for $\mathbf{P}(\hat{v}^{(m)} \leq y)$ in terms of the distribution of an F variate as given in Guenther (1977, equation 3). This can be expanded using the approximation to the log F distribution by Aroian (1941, Section 15). These considerations lead to the following result.

Theorem 2.5 *Let v_1, \dots, v_n be independent absolute $\mathbf{t}_{m-\dim x}$ distributed. Consider the $m + 1$ smallest order statistic $\hat{v}^{(m)}$. Suppose $\dim x$ is fixed while $m \sim \psi n$ for some $0 < \psi < 1$. Let φ be the standard normal density. Then as $n \rightarrow \infty$ it holds*

$$2\varphi(c_{m/n})n^{1/2}(\hat{v}^{(m)} - c_{m/n}) \xrightarrow{D} \mathbf{N}\{0, \psi(1 - \psi)\}.$$

We see that $\hat{v}^{(m)}/c_{m/n}$ is consistent for $c_{m/n}/c_{m/n}$. We will later see that the forward residual scaled by a known variance $\hat{z}^{(m)}/\sigma$ has the same asymptotic distribution, whereas the forward residual scaled by an estimated variance $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$ is consistent for $c_{m/n}/c_{m/n}$ but with a different asymptotic variance. The \mathbf{t} -approximation is therefore not useful in large samples.

3 The main results

The main results for the Forward Search are given. These are expressed in terms of a class of weighted and marked empirical distribution functions at first, then the assumptions are listed, the results are given, and some simulation results reported.

3.1 Absolute empirical process representation

Normalisations are needed for estimators and regressors. Depending on the stochastic properties of the regressor x_i choose a normalisation matrix N and define

$$\hat{b} = N^{-1}(\hat{\beta} - \beta), \quad x_{in} = N'x_i,$$

so that $\sum_{i=1}^n x_{in}x'_{in}$ converges, $n^{-1/2} \sum_{i=1}^n |x_{in}|$ is bounded, and $x'_i(\hat{\beta} - \beta) = x'_{in}b$. If, for example, (y_i, x_i) is stationary then $N = n^{-1/2}I_{\dim x}$ so that $b = n^{1/2}(\hat{\beta} - \beta)$ and $x_{in} = n^{-1/2}x_i$. If x_i is a random walk then $N = n^{-1}$.

Introduce matrix-valued weights g_{in} of the form 1, $n^{1/2}Nx_i$ or $nNx_ix'_iN$, so that $n^{-1} \sum_{i=1}^n |g_{in}|$ is bounded. In the stationary case g_{in} will be 1, x_i or $x_ix'_i$. When x_i is a random walk g_{in} is 1, $n^{-1/2}x_i$ or $n^{-1}x_ix'_i$.

Define the *weighted and marked absolute empirical distribution functions*

$$\widehat{\mathbf{G}}_n^{g,p}(b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p \mathbf{1}_{(|\varepsilon_i - x'_{in}b| \leq \sigma c)}, \quad (3.1)$$

for $b \in \mathbb{R}^{\dim x}$ and $c \geq 0$ and with weights g_{in} and marks ε_i^p . Four combinations of weights and marks are of interest in the analysis of the Forward Search. The deletion residuals involve $g_{in} = 1$, $p = 0$. The least squares estimator involves $g_{in} = n^{1/2}N'x_i$, $p = 1$ and $g_{in} = nN'x_ix'_iN$, $p = 0$. The variance estimator involves the mentioned terms as well as $g_{in} = 1$, $p = 2$. When $p = 0$ the marks are $\varepsilon_i^0 = 1$ so that $\widehat{\mathbf{G}}_n^{g,0}$ is a weighted absolute empirical distribution function similar to those studied by Koul and Ossiander (1994). When also $b = 0$ then $\widehat{\mathbf{G}}_n^{1,0}$ equals the empirical distribution function $\widehat{\mathbf{G}}_n$ of (2.7).

The Forward Search Algorithm 2.1 can now be cast as follows. Step $(m + 1)$ results in an order statistic

$$\hat{z}^{(m)} = \sigma \inf \left\{ c : \widehat{\mathbf{G}}_n^{1,0}(\hat{b}^{(m)}, c) \geq \frac{m+1}{n} \right\}, \quad (3.2)$$

where $g_{in} = 1$, $p = 0$, so that

$$\frac{m+1}{n} = \widehat{\mathbf{G}}_n^{1,0}(\hat{b}^{(m)}, \frac{\hat{z}^{(m)}}{\sigma}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(|\varepsilon_i - x'_{in}\hat{b}^{(m)}| \leq \hat{z}^{(m)})} = \frac{1}{n} \sum_{i \in S^{(m+1)}} 1. \quad (3.3)$$

The least squares estimator has estimation error

$$\hat{b}^{(m+1)} = N^{-1}(\hat{\beta}^{(m)} - \beta) = \left\{ \widehat{\mathbf{G}}_n^{xx,0}(\hat{b}^{(m)}, \frac{\hat{z}^{(m)}}{\sigma}) \right\}^{-1} \widehat{\mathbf{G}}_n^{x,1}(\hat{b}^{(m)}, \frac{\hat{z}^{(m)}}{\sigma}), \quad (3.4)$$

while the bias corrected least squares variance estimator satisfies

$$n^{1/2} \left\{ (\hat{\sigma}_{cor}^{(m+1)})^2 - \sigma^2 \right\} = \frac{n^{1/2}}{\tau_{m/n}} \left[\widehat{\mathbf{G}}_n^{1,2}(\hat{b}^{(m)}, \frac{\hat{z}^{(m)}}{\sigma}) - \left\{ \hat{b}^{(m+1)} \right\}' \widehat{\mathbf{G}}_n^{xx,0}(\hat{b}^{(m)}, \frac{\hat{z}^{(m)}}{\sigma}) \left\{ \hat{b}^{(m+1)} \right\} \right]. \quad (3.5)$$

3.2 Assumptions

In the following a series of sufficient assumptions are listed for the asymptotic theory of the Forward Search. When using the Forward Search the density f is assumed known. In practice the normal distribution is used, although a distribution with heavier tails could also be used. To accommodate that intended use, the listed Assumptions are somewhat stronger than they need to be for the sake of parsimony. As a part of the proof, a class of weighted and marked empirical

processes are analysed in §4 and at that point somewhat weaker assumptions are introduced, see Assumption 4.1.

Assumption 3.1(i) is satisfied for the normal distribution. For other distributions the regularity conditions involve a trade-off between four features: η , which indicates the rate of the initial estimator, κ , which indicates the order of magnitude of maximum of the normalised regressors, and $\dim x$, the dimension of the regressor. From these quantities a number r is defined, which controls the number of moments and the smoothness required for the density f . The number r is increasing in κ and $\dim x$ and decreasing in η . The number of required moments is larger than 8 in order to control the estimation error for the variance.

Assumption 3.1 *Let \mathcal{F}_i be an increasing sequence of σ fields so ε_{i-1} and x_i are \mathcal{F}_{i-1} -measurable and ε_i is independent of \mathcal{F}_{i-1} with symmetric, continuous, differentiable density f which is positive for $F^{-1}(0) < c < F^{-1}(1)$. For some $0 \leq \kappa < \eta \leq 1/4$ choose an $r \geq 2$ so $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim x)$. Let $q_0 = 1 + 2^{r+1}$. Suppose*

(i) *density satisfies*

- (a) *tail monotonicity: $c^q f(c)$, $|c^{q-1} f'(c)|$ are decreasing for large c and some $q \in \mathbb{R}$ so $q > q_0$;*
- (b) *quantile process condition: $\gamma = \sup_{c>0} F(c)\{1 - F(c)\}|f'(c)|/\{f(c)\}^2 < \infty$;*
- (c) *unimodality: $f'(c) \leq 0$ for $c > 0$ and $\lim_{c \rightarrow 0} f''(c) < 0$;*
- (d) *logarithmic derivative: $\Delta(c) = [c\{\log f(c)\}]' < 0$ for $c > 0$;*
- (e) *strong quantile process condition: $\{1 - F(c)\}/\{cf(c)\} = O(1)$ for $c \rightarrow \infty$;*

(ii) *regressors x_i are \mathcal{F}_{i-1} -measurable and satisfy*

- (a) $\Sigma_n = N' \sum_{i=1}^n x_i x_i' N \xrightarrow{D} \Sigma \stackrel{a.s.}{>} 0$;
- (b) $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = O_P(1)$ for some $\kappa < \eta$;
- (c) $n^{-1} \mathbf{E} \sum_{i=1}^n |n^{1/2} N' x_i|^{q_0} = O(1)$;

(iii) *initial estimator: $N^{-1}(\hat{\beta}^{(m_0)} - \beta) = O_P(n^{1/4-\eta})$ for some $\eta > 0$.*

Assumption 3.1 is satisfied in a range of situations. First some general comments. Condition (ia) is more severe than normally seen in empirical process theory due to the marks ε_i^p . Condition (ib) was used in Theorem 2.2. Conditions (ic, id) are needed for controlling the iterative aspect of the Forward Search. Condition (id) to $\Delta(c)$ is also used in Rousseeuw (1982) when discussing change-of-variance curves for M-estimators and assumes log concave densities. It is also the cross derivative of the log likelihood for location-scale families. Condition (ie) to Mill's ratio is milder than the condition employed for kernel density estimation by Csörgő (1983, p. 139). Condition (iia) is standard in regression analysis. Condition (iib) is discussed below.

Example 3.1 *Assumption 3.1(i) to the reference distribution f .*

(a) **Standard normal distribution**, so $f = \varphi$. Condition (i) is satisfied: (ia) holds since $c^q \varphi(c) = -c^{q-1} \varphi'(c)$ is decreasing for large c for any q . (ib) holds with $\gamma = 1$, noting $\varphi'(c) = -c\varphi(c)$ and the Mill's ratio result $\{(4 + c^2)^{1/2} - c\}/2 < \{1 - \Phi(c)\}/\varphi(c) < 1/c$, see Sampford (1953). (id) holds with $\Delta(c) = -2c$. (ie) holds since $\{1 - \Phi(c)\}/\{c\varphi(c)\} < 1/c^2 \rightarrow 0$ as $c \rightarrow \infty$. (b) **t-distribution** with $d > 2^{r+1}$ degrees of freedom and density $f(c) = C_d(1 + c^2/d)^{-(d+1)/2}$ and $C_d = \Gamma\{(d+1)/2\}/\{(d\pi)^{1/2}\Gamma(d/2)\}$. It holds $f'(c) = -\gamma f(c)/h(c)$ with $\gamma = 1 + d^{-1}$, defining $h(c) = c/(1 + c^2/d)$ so that $h(c) \sim d/c$ for large c . Condition (i) is satisfied: (ia) holds since $c^q f(c) \sim C c^{q-d-1}$ is declining for q chosen so $d+1 > q > q_0$, and $c^{q-1}|f'(c)|$ and $c^q f(c)$ are declining. (ib) holds with the stated γ since $1 - c^{-2}d/(d+2) < h(c)\{1 - F(c)\}/f(c) < 1$, see Soms (1976, equation 3.2). (ic) is well-known to hold. (id) holds with $\Delta(c) = -2\gamma\{h(c)\}^2/c < 0$. (ie) holds since $\{1 - F(c)\}/\{cf(c)\} < 1/\{ch(c)\} \rightarrow 1/d$ as $c \rightarrow \infty$.

Example 3.2 Assumption 3.1(ii) to the regressors x_i .

(a) **Stationary regressors.** Let $N = n^{-1/2}I_{\dim x}$ and $g_{in} = x_i$. To ensure (iic) it is necessary that $E|x_i|^{q_0} < \infty$. By Boole's inequality and the triangle inequality then $n^{1/2-\kappa} \max_{1 \leq i \leq n} |N'x_i| = O_P(n^{1-\kappa q_0})$ so (iib) holds for all $\eta > \kappa = q_0^{-1}$.

(b) **Deterministic regressors** such as $x_i = (1, i)'$. Let $N = \text{diag}(n^{-1/2}, n^{-3/2})$. Then $n^{1/2}N'x_i = (1, i/n)'$. Thus condition (ii) follows with $\kappa = 0$.

(c) **Random walk regressors** such as $x_i = \sum_{s=1}^{i-1} \varepsilon_s$. Let $N = n^{-1}$. Then $n^{-1}x_{\text{int}(n\psi)}$ converges to a Brownian motion by Donsker's invariance principle, see Billingsley (1968). Condition (iia, iib) follows from the continuous mapping theorem with $\kappa = 0$. As x_i is defined in terms of ε_i which has moments of order q_0 , so has x_i and (iic) follows.

3.3 The results

The main results are described in terms of two processes. The first process is

$$\mathbb{G}_n(c_\psi) = n^{1/2}\{\widehat{\mathbb{G}}_n^{1,0}(0, c_\psi) - \psi\} = n^{1/2}\{\widehat{\mathbb{G}}_n(c_\psi) - \psi\}, \quad (3.6)$$

which behaves asymptotically as a Brownian bridge as discussed in Theorems 2.1, 2.2. The second process is asymptotically Gaussian and given by

$$\mathbb{H}_n(c_\psi) = n^{1/2}\{\sigma^{-2}\widehat{\mathbb{G}}_n^{1,2}(0, c_\psi) - \tau_\psi\} = n^{1/2}\{\sigma^{-2}\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \mathbf{1}_{(|\varepsilon_i| \leq \sigma c_\psi)} - \tau_\psi\}. \quad (3.7)$$

The two first results are a Bahadur representation for the forward scaled residuals $\sigma^{-1}\hat{z}_\psi$ with known variance along with a representation for the bias corrected variance estimator.

Theorem 3.1 *Suppose Assumption 3.1 holds. Let $\psi_0 > 0$. Then*

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} |2\mathbf{f}(c_\psi)n^{1/2}(\sigma^{-1}\hat{z}_\psi - c_\psi) + \mathbb{G}_n(c_\psi)| \xrightarrow{P} 0.$$

Theorem 3.2 *Suppose Assumption 3.1 holds. Let $\psi_0 > 0$. Then*

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} |n^{1/2}\left(\frac{\hat{\sigma}_{\psi, \text{cor}}^2}{\sigma^2} - 1\right) - \frac{1}{\tau_\psi}\{\mathbb{H}_n(c_\psi) - c_\psi^2\mathbb{G}_n(c_\psi)\}| = o_P(1).$$

Remark 3.1 *In Theorems 3.1, 3.2 the supremum is taken over a smaller interval for ψ than the unit interval. A left end point larger than 0 is needed to ensure consistency. The results potentially hold with a right end point equal to 1. Proving this would, however, add significantly to the length of the proof without practical benefit since the last forward residual is based on the set $S^{(n-1)}$ with $n-1$ selected points.*

Remark 3.2 *The least squares estimator for the variance is $\hat{\sigma}_1^2 = \hat{\sigma}_{1, \text{cor}}^2$, noting that $\tau_1 = 1$. Least squares theory shows that $n^{1/2}(\hat{\sigma}_{1, \text{cor}}^2/\sigma^2 - 1) = n^{-1}\sum_{i=1}^n(\varepsilon_i^2/\sigma^2 - 1) + o_P(1)$. To see that Theorem 3.2 matches this result note that the leading term of the least squares approximation is $\lim_{\psi \rightarrow 1} \mathbb{H}_n(c_\psi)$. It is therefore necessary that $\lim_{\psi \rightarrow 1} c_\psi^2\mathbb{G}_n(c_\psi) = o_P(1)$. Since ε_i has more than 8 moments then $c_\psi^2 = o\{(1-\psi)^{-1/4}\}$, see also item 5 of the proof of Lemma 8.11. Combine this with Theorems 2.3(a), 2.4 to see that $\lim_{\psi \rightarrow 1} c_\psi^2\mathbb{G}_n(c_\psi) = o_P(1)$.*

Combining Theorem 3.1 and 3.2 gives an asymptotic representation of the forward residuals scaled by the estimated variance.

Theorem 3.3 *Suppose Assumption 3.1 holds. Let $c_\psi = \mathbf{G}^{-1}(\psi)$ and $\psi_0 > 0$. Then the bias corrected scaled forward residuals has the expansion*

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} \left| 2\mathbf{f}(c_\psi)n^{1/2}\left(\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi,cor}} - c_\psi\right) + \left(1 - \frac{c_\psi^3 \mathbf{f}(c_\psi)}{\tau_\psi}\right)\mathbf{G}_n(c_\psi) + \frac{c_\psi \mathbf{f}(c_\psi)}{\tau_\psi}\mathbb{H}_n(c_\psi) \right| = o_{\mathbf{P}}(1).$$

The scaled forward residuals $\hat{z}_\psi/\hat{\sigma}_\psi$ have a similar expansion. To see this, divide by ς_ψ and note that $\hat{\sigma}_{\psi,cor}\varsigma_\psi = \hat{\sigma}_\psi$ for $\psi_0 \leq \psi \leq n/(n+1)$.

The above results generalise those of Johansen and Nielsen (2010, Theorems 5.1, 5.3) which hold for a single forward step for location-scale models. It is interesting to note that the results do not depend on the type of regressors for the model. In particular, the results do not depend on whether the regressors include an intercept or not, which sets the results aside from empirical processes of residuals, see for instance Engler and Nielsen (2009).

To apply Theorem 3.3 in a point-wise fashion the asymptotic covariance matrices of $\mathbf{G}_n(c_\psi)$ and $\mathbb{H}_n(c_\psi)$ have to be found. In parallel to the integral τ_ψ from (2.6) introduce

$$\tau_\psi = 2 \int_0^{c_\psi} x^2 \mathbf{f}(x) dx, \quad \varkappa_\psi = 2 \int_0^{c_\psi} x^4 \mathbf{f}(x) dx.$$

The asymptotic covariance of $\mathbf{G}_n(c_\psi)$ and $\mathbb{H}_n(c_\psi)$ is then

$$\begin{aligned} \begin{pmatrix} \omega_{\psi,GG} & \omega_{\psi,GH} \\ \omega_{\psi,GH} & \omega_{\psi,HH} \end{pmatrix} &= \text{Var} \left\{ \begin{array}{c} \mathbf{G}_n(c_\psi) \\ \mathbb{H}_n(c_\psi) \end{array} \right\} \\ &= \text{Var} \left\{ \begin{array}{c} 1_{(|\varepsilon_i| \leq \sigma c_\psi)} - \psi \\ (\varepsilon_i^2/\sigma^2)1_{(|\varepsilon_i| \leq \sigma c_\psi)} - \tau_\psi \end{array} \right\} = \begin{pmatrix} \psi(1-\psi) & \tau_\psi(1-\psi) \\ \tau_\psi(1-\psi) & \varkappa_\psi - \tau_\psi^2 \end{pmatrix}. \end{aligned}$$

The following pointwise results arise for $\psi_0 \leq \psi \leq \psi_1$, for some $\psi_0 > 0$ and $\psi_1 < 1$,

$$n^{1/2}\left(\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi,cor}} - c_\psi\right) \xrightarrow{\mathbf{D}} \mathbf{N}(0, \omega_\psi), \quad n^{1/2}\left(\frac{\hat{z}_\psi}{\hat{\sigma}_\psi} - \frac{c_\psi}{\varsigma_\psi}\right) \xrightarrow{\mathbf{D}} \mathbf{N}\left(0, \frac{\omega_\psi}{\varsigma_\psi^2}\right), \quad (3.8)$$

where $\varsigma_\psi = \tau_\psi/\psi$ and

$$4\tau_\psi^2 \omega_\psi = \left\{ \frac{\tau_\psi}{\varphi(c_\psi)} - c_\psi^3 \right\}^2 \omega_{\psi,GG} + 2\left\{ \frac{\tau_\psi}{\varphi(c_\psi)} - c_\psi^3 \right\} c_\psi \omega_{\psi,GH} + c_\psi^2 \omega_{\psi,HH}.$$

Using l'Hôpital's rule it is seen that $c_0/\varsigma_0 = \sqrt{3}$.

Example 3.3 *Some particular reference distributions.*

(a) **Standard normal distribution.** *If $\mathbf{f} = \varphi$ then $c_\psi = \Phi^{-1}\{(1+\psi)/2\}$ and*

$$\begin{aligned} \tau_\psi &= 2 \int_0^{c_\psi} x^2 \varphi(x) dx = 2\{\Phi(x) - x\varphi(x)\}|_0^{c_\psi} = \psi - 2c_\psi \varphi(c_\psi), \\ \varkappa_\psi &= 2 \int_0^{c_\psi} x^4 \varphi(x) dx = 2\{3\Phi(x) - (x^3 + 3x)\varphi(x)\}|_0^{c_\psi} = 3\psi - 2(c_\psi^3 + 3c_\psi)\varphi(c_\psi). \end{aligned}$$

(b) **t-distribution** *with d degrees of freedom. Then $c_\psi = F_d^{-1}\{(1+\psi)/2\}$. Let*

$$\begin{aligned} H_{d-2}(c_\psi) &= \frac{d-1}{d-2} [2F_{d-2}\{c_\psi(\frac{d-2}{d})^{1/2}\} - 1], \\ H_{d-4}(c_\psi) &= \frac{(d-1)(d-3)}{(d-2)(d-4)} [2F_{d-4}\{c_\psi(\frac{d-4}{d})^{1/2}\} - 1]. \end{aligned}$$

Then it holds

$$\tau_\psi = d\{H_{d-2}(c_\psi) - \psi\}, \quad \varkappa_\psi = d^2\{H_{d-4}(c_\psi) - 2H_{d-2}(c_\psi) + \psi\}.$$

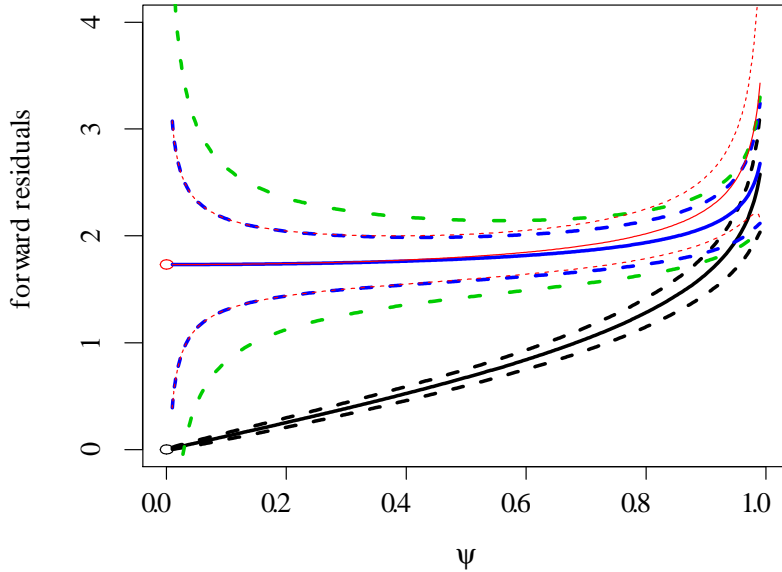


Figure 1: Illustration of the asymptotic results (3.8) for $n = 100$.

³Figure 1 illustrates the asymptotic results (3.8) when f is standard normal or t_5 and $n = 100$. For the standard normal case of Example 3.3(a) the asymptotic results of the forward residuals $\hat{z}_\psi/\hat{\sigma}_\psi$ based on the biased estimator $\hat{\sigma}_\psi$ are shown with blue. The bold, solid line is the asymptotic mean c_ψ/ς_ψ , noting that $c_0/\varsigma_0 = \sqrt{3}$. The bold, dashed lines are the 5% and 95% quantiles $\{c_\psi \pm 2(\omega_\psi/n)^{1/2}\}/\varsigma_\psi$, which contain 90% of the scaled forward residuals with $n = 128$, chosen for comparability with the data example in Riani and Atkinson (2007, Figure 1). In a similar way the bold, black lines show the results for bias-corrected forward residuals $\hat{z}_\psi/\hat{\sigma}_{\psi,cor}$. The biased, blue and unbiased, black quantiles are equivalent. The blue 5% and 95% quantiles fan out for small ψ due to the biased estimate of the variance $\hat{\sigma}^{(m)}$. For large ψ the quantiles diverge for both the biased and the unbiased variance estimates. The dash-dot green lines show the 5% and 95% quantiles of the forward residuals with known scale \hat{z}_ψ/σ . These bounds are the same as those arising from the asymptotic version of the t -approximation as discussed in Theorem 2.5. Note that the variance of \hat{z}_ψ/σ is wider than that of $\hat{z}_\psi/\hat{\sigma}_{\psi,cor}$. This phenomenon is also seen for empirical processes of estimated residuals, see Engler and Nielsen (2009, equation 2.10). For the t case of Example 3.3(b) the degrees of freedom are chosen to be 5 and the result are shown with thin, red lines. With 5 degrees of freedom Assumption 3.1 is not met. For higher degrees of freedom the results will be in between the t_5 and the normal results.

3.4 Simulation evidence

The quality of the asymptotic distribution approximation to $\hat{z}_\psi/\hat{\sigma}_\psi$ has been investigated through simulation. The asymptotic result in Theorem 3.3 does not depend on the choice of regressors x_i and parameters β, σ^2 . The simulations of Riani and Atkinson (2007, Figures 3, 12) suggest that this is largely the case in finite samples. We therefore report a simulation study based on a very simple data generating process without exploring finite sample variation with x_i, β .

Quantiles of the asymptotic approximation and the t -distribution approximation were computed analytically. The finite sample probability of not exceeding these values were computed

³Simulations were done using Ox 6.3, see Doornik (2007). Graphics were done using R 2.13, see R Development Core Team (2011).

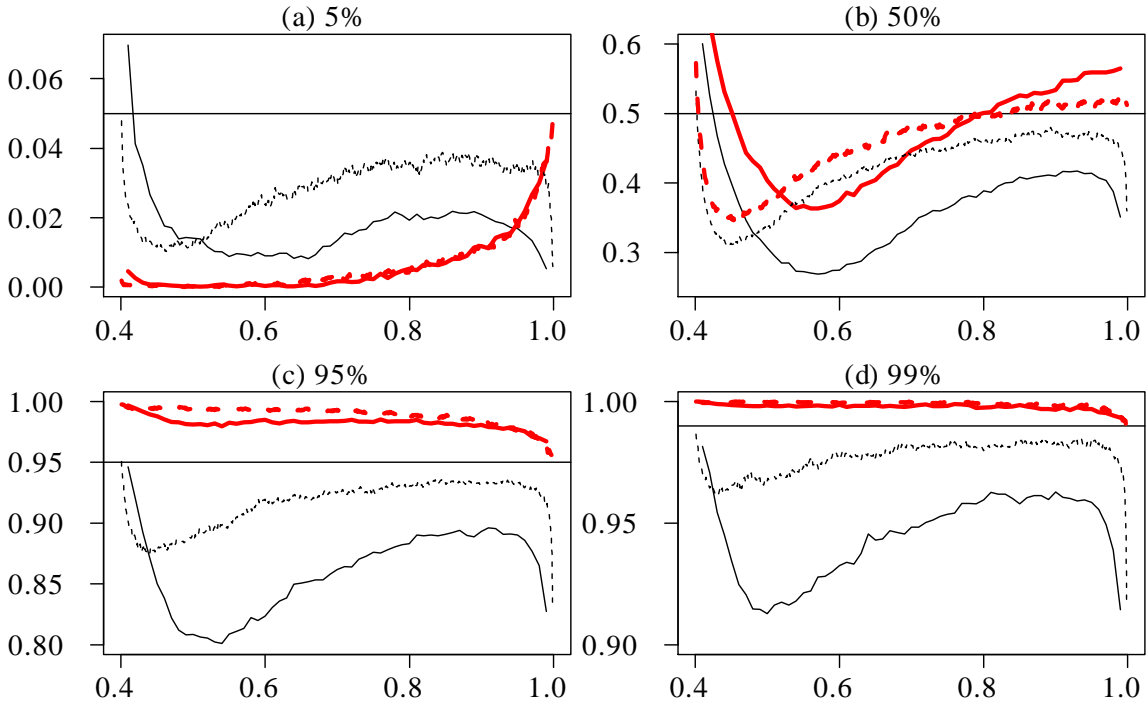


Figure 2: The finite sample performance of the asymptotic approximation and the t-approximation to the distribution of $\hat{z}_\psi/\hat{\sigma}_\psi$ is evaluated. The four plots consider different target p -values: 5%, 50%, 95%, 99%. In each plot the horizontal line gives the target p -value. The asymptotic approximation is evaluated by the thin, black, solid and dotted lines for $n = 100$ and $n = 1000$, respectively. In a similar way, the t-approximation is evaluated by the thick, red, solid and dotted lines. In all cases the Forward Search is initialised at $m_0 = 0.4n$ using the full sample average.

by simulation. In the simulation the model was $y_i = \beta + \varepsilon_i$, so the regressor is $x_i = 1$ with sample sizes of either $n = 100$ or $n = 1000$. The data generating process had $\beta = 0$ and ε_i independent standard normal. Two choices of initial estimators were considered: the average of all observations and the average of the first $n^{2/3}$ observations. Neither is robust, but since no outlier is present, both are consistent at rates of $n^{1/2}$ and $n^{1/3}$, respectively. The size of the initial set of observations was either 4% or 40% of n . The number of repetitions was 10^4 . Figures 2, 3 consider various combinations of these setups.

Figure 2 has four panels considering the performance of the 5%, 50%, 95%, and 99% quantile approximations, respectively. In each panel the asymptotic approximation and the t-distribution approximation are evaluated for $n = 100$ and $n = 1000$ with $m_0 = 40$ and using the full sample average as initial estimator. The consistency of the asymptotic approximation and the inconsistency of the t-distribution approximation with increasing n is seen. In current practice the Forward Search is terminated at the first exit from the pointwise 95% or 99% bands. The evaluation in panels (c), (d) shows that for $n = 100$ the actual p -values vary in the intervals 80–95% and 91–99%. For $n = 1000$ this improves to 88–95% and 96–99%. While this is not perfect it is comparable with the performance of other statistical procedures and it suggests that a second order expansion may give rather accurate results.

Figure 3 considers the performance of the 95% quantile approximation when the initial estimator and the size of the initial set of observations vary. Panel (a) has m_0 given by 4% and 40% of $n = 100$. It is seen that initially the performance of the distribution approximation deteriorates

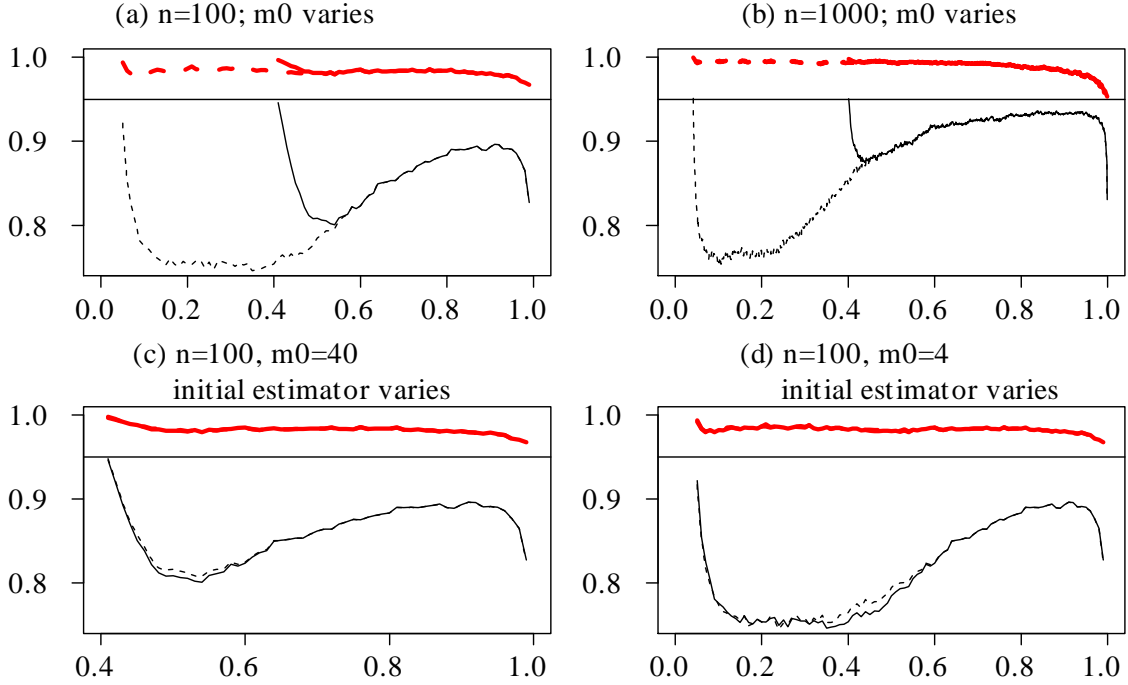


Figure 3: The effect on the distribution of $\hat{z}_\psi/\hat{\sigma}_\psi$ from varying the size of initial set of observations, m_0 , and the choice of initial estimator, $\hat{\beta}^{(m_0)}$, is evaluated for a target p -value of 95%. Plots (a), (b) have varying m_0 while using the full sample average as initial estimator for $n = 100$ and $n = 1000$, respectively. The asymptotic approximation is evaluated by the thin, black, solid and dotted lines for $m_0 = 0.4n$ and $m_0 = 0.04n$, respectively. In a similar way, the t-approximation is evaluated by the thick, red, solid and dotted lines. Plots (c), (d) have $n = 100$ and varying initial estimator for $m_0 = 0.4n$ and $m_0 = 0.04n$, respectively. The asymptotic approximation is evaluated by the thin, black, solid and dotted lines for $\hat{\beta}^{(m_0)}$ chosen as the average of all n observations and the first $n^{2/3}$ observations, respectively, noting that $100^{2/3} = 22$. In a similar way, the t-approximation is evaluated by the thick, red, solid and dotted lines.

with decreasing m_0 , but for large m the quality of the approximation is not varying much with m_0 . Panel (b) has the same evaluation for $n = 1000$. Panel (c) considers the two choices of initial estimators which are $n^{1/2}$ and $n^{1/3}$ -consistent, respectively, for m_0 given by 40% of $n = 100$. There is not much variation with the estimator. Panel (d) has the same evaluation for m_0 given by 4% of $n = 100$.

4 A class of auxiliary weighted and marked empirical processes

It is useful to consider an auxiliary class of weighted and marked empirical distribution functions for errors ε_i as opposed to absolute errors $|\varepsilon_i|$. The analysis of this class generalises that of Koull and Ossiander (1994) in two respects. First, the standardised estimation error b is permitted to diverge at a rate of $n^{1/4-\eta}$ rather than being bounded. Secondly, non-bounded marks of the type ε_i^p are allowed. These results are therefore of independent interest. This class of *weighted and marked empirical distribution functions* is defined for $b \in \mathbb{R}^{\dim x}$ and $c \in \mathbb{R}$ by

$$\widehat{\mathbf{F}}_n^{g,p}(b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p \mathbf{1}_{(\varepsilon_i \leq \sigma c + x'_{in} b)}, \quad (4.1)$$

with $(\varepsilon_{i-1}, \dots, \varepsilon_1, x_i, \dots, x_1)$ -measurable weights g_{in} and marks ε_i^p .

4.1 Assumptions

We will keep track of the assumptions in a more explicit way than done above. In the analysis of the one-sided empirical processes the density f is not necessarily symmetric.

Assumption 4.1 *Let \mathcal{F}_i be an increasing sequence of σ fields so $\varepsilon_{i-1}, x_i, g_{in}$ are \mathcal{F}_{i-1} -measurable and ε_i is independent of \mathcal{F}_{i-1} with continuous, differentiable density f which is positive for $F^{-1}(0) < c < F^{-1}(1)$. Let p, r, η, κ, ν be given so $p, r \in \mathbb{N}_0$, $0 \leq \kappa < \eta \leq 1/4$ and $\nu \leq 1$.*

Suppose

(i) *density satisfies:*

- (a) *moments:* $\int_{-\infty}^{\infty} |\varepsilon|^{2rp/\nu} f(\varepsilon) d\varepsilon < \infty$;
- (b) *boundedness:* $\sup_{c \in \mathbb{R}} \{(1 + |c|^{2r})f(c) + (1 + |c|^{2r})|f'(c)|\} < \infty$;
- (c) *smoothness:* a $C_H \in \mathbb{N}$ exist so that for all $c > 0$

$$\frac{\sup_{\varepsilon \geq c} (1 + \varepsilon^{2rp})f(\varepsilon)}{\inf_{0 \leq \varepsilon \leq c} (1 + \varepsilon^{2rp})f(\varepsilon)} \leq C_H, \quad \frac{\sup_{\varepsilon \leq -c} (1 + |\varepsilon|^{2rp})f(\varepsilon)}{\inf_{-c \leq \varepsilon \leq 0} (1 + |\varepsilon|^{2rp})f(\varepsilon)} \leq C_H.$$

(ii) *regressors x_i satisfy $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' x_i| = O_P(1)$;*

(iii) *weights g_{in} are matrix valued and satisfy*

- (a) $n^{-1} \mathbf{E} \sum_{i=1}^n |g_{in}|^{2r} (1 + |n^{1/2} N' x_i|) = O(1)$;
- (b) $n^{-1} \sum_{i=1}^n |g_{in}| (1 + |n^{1/2} N' x_i|^2) = O_P(1)$.

Remark 4.1 *Some discussion of Assumption 4.1 is given*

(a) **The case of no marks** $p = 0$. This is the situation discussed in Koul and Ossiander (1994). The primary role of r is to control the tail behaviour of the density. When $p = 0$ then $2^r p = 0$ for all $r \in \mathbb{N}_0$, so r can be chosen as $r = 0$ and the assumption simplifies considerably.

(b) **The tail condition in Assumption 4.1(ia)** is used for some $\nu < 1$ for the tightness result in Theorem 4.4. Otherwise $\nu = 1$ suffices.

(c) **The smoothness of density in Assumption 4.1(ic)** is satisfied if $h_r(c) = (1 + \varepsilon^{2rp})f(\varepsilon)$ is monotone for $|c| > d_1$ for some $d_1 \geq 0$. Indeed, choose $d_2 \geq d_1$ so that $\sup_{c \geq d_2} h_r(c) = \inf_{0 \leq c \leq d_2} h_r(c) = h_r(d_2)$. Then choose C_H larger than $\sup_{0 \leq c \leq d_2} h_r(c) / \inf_{0 \leq c \leq d_2} h_r(c)$. A similar argument applies for $c < 0$. Note, that the smoothness condition implies that the density has connected support.

(d) **Sufficient condition for Assumption 4.1(i)**. If f is symmetric and differentiable with $c^q f(c)$, $c^{q-1} |f'(c)|$ both decreasing for large c for some $q > 1 + 2^r p$, then Assumption 4.1(i) holds. Indeed, (ia) holds, since when $c^q f(c)$ is decreasing, then $c^{2rp/\nu} f(c)$ is integrable for some $\nu < 1$. Further, (ib) holds, since, first, the continuity and decreasingness of $c^q f(c)$ and hence of $f(c)$ implies $(1 + |c|^{1+2^r p})f(c)$ is bounded, and, secondly, since $f'(c) < 0$ so that $|c^{q-1} f'(c)|$ decreases then $(1 + |c|^{2^r p})|f'(c)|$ is bounded. Finally, (ic) holds due to the remark (c) above.

4.2 The empirical process results

The weighted and marked empirical distribution function $\widehat{F}_n^{g,p}(b, c)$ defined in (4.1) is analysed through martingale arguments. Thus, introduce the sum of conditional expectations

$$\overline{F}_n^{g,p}(b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \{ \varepsilon_i^p 1_{(\varepsilon_i \leq \sigma c + x_{in}' b)} \}, \quad (4.2)$$

and the *weighted and marked empirical process*

$$\mathbb{F}_n^{g,p}(b, c) = n^{1/2} \{ \widehat{\mathbb{F}}_n^{g,p}(b, c) - \overline{\mathbb{F}}_n^{g,p}(b, c) \}. \quad (4.3)$$

Three results follows. These are proved in the subsequent sections 5, 6, 7. The first result shows that the dependence of $\mathbb{F}_n^{g,p}$ on the estimation error b is negligible.

Theorem 4.1 *Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 4.1(i, ii, iii) holds with $\nu = 1$, some $\eta > 0$ and an r so $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim x)$. Then, for any $B > 0$ and $n \rightarrow \infty$, it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4 - \eta} B} |\mathbb{F}_n^{g,p}(b, c_\psi) - \mathbb{F}_n^{g,p}(0, c_\psi)| = o_{\mathbb{P}}(1).$$

For the standard empirical process with weights $g_{in} = 1$ and marks $\varepsilon_i^p = 1$ the order of the remainder term can be improved as follows. In terms of the Assumption 4.1 note that when $p = 0$ then the r will be irrelevant except for the condition on the regressors in part (iii).

Theorem 4.2 *Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 4.1(i, ii, iii) holds with $\nu = 1$, $p = 0$, $r = 2$ and some $\eta > 0$. Then, for any $B > 0$, any $\omega < \eta - \kappa \leq 1/4$ and $n \rightarrow \infty$, it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4 - \eta} B} |\mathbb{F}_n^{1,0}(b, c_\psi + n^{\kappa - 1/2} d) - \mathbb{F}_n^{1,0}(0, c_\psi)| = o_{\mathbb{P}}(n^{-\omega}).$$

The next results presents a linearization of $\overline{\mathbb{F}}_n^{g,p}(b, c)$.

Theorem 4.3 *Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 4.1(ib, iiib) holds with $r = 0$ and some $\eta > 0$. Then, for all $B > 0$ and $n \rightarrow \infty$, it holds that*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4 - \eta} B} |n^{1/2} \{ \overline{\mathbb{F}}_n^{g,p}(b, c_\psi) - \overline{\mathbb{F}}_n^{g,p}(0, c_\psi) \} - \sigma^{p-1} c_\psi^p \mathbf{f}(c_\psi) n^{-1} \sum_{i=1}^n g_{in} n^{1/2} x'_{in} b| = O_{\mathbb{P}}(n^{-2\eta}).$$

Finally, the weighted and marked empirical process $\mathbb{F}_n^{g,p}(0, c_\psi)$ in (4.3) is tight. It holds by construction that $\mathbb{F}_n^{g,p}(0, 0) = 0$. Following Billingsley (1968, Theorem 15.5) tightness in the space $D[0, 1]$ endowed with the uniform metric, then follows from the next result.

Theorem 4.4 *Let $c_\psi = F^{-1}(\psi)$. Suppose Assumption 4.1(ia, iiiia) holds with $r = 2$ and some $\nu < 1$. Then, for all $\epsilon > 0$, it holds*

$$\lim_{\phi \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{0 \leq \psi \leq \psi^\dagger \leq 1: \psi^\dagger - \psi \leq \phi} |\mathbb{F}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0, c_\psi)| > \epsilon \right\} \rightarrow 0.$$

The proofs of these results are given in Section 7, but first we establish some martingale results and discuss a metric on \mathbb{R} which is applied in the chaining argument needed in the proofs. Finally, the proofs of the main results, Theorems 3.1, 3.2, 3.3, are given in Section 8.

5 Martingale results

Chaining arguments will be used to handle tightness properties of the empirical processes. This reduces the problem to a problem of finding the tail probability for the maximum of a certain family of martingales. Initially a general bound to the tail probability is constructed using the martingale inequality by Bercu and Touati (2008). Subsequently, two special cases are analysed where the number of elements in the martingale family is increasing and where it is fixed.

Lemma 5.1 *For ℓ so $1 \leq \ell \leq L$ let $z_{\ell,i}$ be \mathcal{F}_i -adapted so $\mathbb{E}z_{\ell,i}^{2\bar{r}} < \infty$ for some $\bar{r} \in \mathbb{N}$. Let $D_r = \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbb{E}_{i-1} z_{\ell,i}^{2r}$ for $1 \leq r \leq \bar{r}$. Then, for all $\kappa_0, \kappa_1, \dots, \kappa_{\bar{r}} > 0$, it holds*

$$\mathbb{P}\left\{\max_{1 \leq \ell \leq L} \left| \sum_{i=1}^n (z_{\ell,i} - \mathbb{E}_{i-1} z_{\ell,i}) \right| > \kappa_0\right\} \leq L \frac{\mathbb{E}D_{\bar{r}}}{\kappa_{\bar{r}}} + \sum_{r=1}^{\bar{r}} \frac{\mathbb{E}D_r}{\kappa_r} + 2L \sum_{r=0}^{\bar{r}-1} \exp\left(-\frac{\kappa_r^2}{14\kappa_{r+1}}\right).$$

Proof of Lemma 5.1. 1. *Notation.* For $0 \leq r \leq \bar{r}$ define $A_{\ell,r} = \sum_{i=1}^n (z_{\ell,i}^{2r} - \mathbb{E}_{i-1} z_{\ell,i}^{2r})$ and

$$\mathcal{P}_r(\kappa_r) = \mathbb{P}\left(\max_{1 \leq \ell \leq L} A_{\ell,r} > \kappa_r\right), \quad \mathcal{Q}_r(\kappa_r) = \mathbb{P}\left(\max_{1 \leq \ell \leq L} |A_{\ell,r}| > \kappa_r\right),$$

where $\mathcal{Q}_0(\kappa_0)$ is the probability of interest, while $\mathcal{P}_r(\kappa_r) \leq \mathcal{Q}_r(\kappa_r)$.

2. *The terms $\mathcal{Q}_r(\kappa_r)$ for $0 \leq r < \bar{r}$.* Argue that for any $\kappa_r, \kappa_{r+1} > 0$ then

$$\mathcal{Q}_r(\kappa_r) \leq 2L \exp\left(-\frac{\kappa_r^2}{14\kappa_{r+1}}\right) + \mathcal{P}_{r+1}(\kappa_{r+1}) + \frac{\mathbb{E}D_{r+1}}{\kappa_{r+1}}. \quad (5.1)$$

The idea is now to apply the following inequality, for sets \mathcal{A}, \mathcal{B} ,

$$\mathbb{P}(\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}^c) \leq \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{B}^c).$$

In the first term, \mathcal{A} relates to the tails of a martingale and \mathcal{B} to the central part of the distribution of the quadratic variation. Thus the first term can be controlled by a martingale inequality. In the second term, \mathcal{B}^c relates to the tail of the quadratic variation. The sum of the predictable and the total quadratic variation of $A_{\ell,r}$ is $B_{\ell,r} = \sum_{i=1}^n B_{\ell,r,i}$ where $B_{\ell,r,i} = (z_{\ell,i}^{2r} - \mathbb{E}_{i-1} z_{\ell,i}^{2r})^2 + \mathbb{E}_{i-1} (z_{\ell,i}^{2r} - \mathbb{E}_{i-1} z_{\ell,i}^{2r})^2$. It holds

$$\mathcal{Q}_r(\kappa_r) \leq \mathbb{P}\left\{\left(\max_{1 \leq \ell \leq L} |A_{\ell,r}| > \kappa_r\right) \cap \left(\max_{1 \leq \ell \leq L} B_{\ell,r} < 7\kappa_{r+1}\right)\right\} + \mathbb{P}\left(\max_{1 \leq \ell \leq L} B_{\ell,r} \geq 7\kappa_{r+1}\right). \quad (5.2)$$

Consider the first term in (5.2), $\mathcal{S}_{1,r}$ say. By Boole's inequality this satisfies

$$\mathcal{S}_{1,r} \leq \sum_{\ell=1}^L \mathbb{P}\left\{\left(|A_{\ell,r}| > \kappa_r\right) \cap \left(\max_{1 \leq \ell \leq L} B_{\ell,r} < 7\kappa_{r+1}\right)\right\}.$$

Noting that $(\max_{1 \leq \ell \leq L} B_{\ell,r} \leq 7\kappa_{r+1}) \subset (B_{\ell,r} \leq 7\kappa_{r+1})$ gives the further bound

$$\mathcal{S}_{1,r} \leq \sum_{\ell=1}^L \mathbb{P}\left\{\left(|A_{\ell,r}| > \kappa_r\right) \cap (B_{\ell,r} < 7\kappa_{r+1})\right\}.$$

Since $A_{\ell,r}$ is a martingale the exponential inequality of Bercu and Touati (2008) shows

$$\mathbb{P}\left\{\left(|A_{\ell,r}| > \kappa_r\right) \cap (B_{\ell,r} < 7\kappa_{r+1})\right\} \leq 2 \exp\left\{-\frac{\kappa_r^2}{14\kappa_{r+1}}\right\}.$$

Taken L times, this gives the first term in (5.1).

Consider the second term in (5.2), $\mathcal{S}_{2,r}$ say. Ignore the indices on $B_{\ell,r,i}, z_{\ell,i}^{2^r}$ and apply the inequality $(z - \mathbf{E}z)^2 \leq 2(z^2 + \mathbf{E}^2z)$ along with $\mathbf{E}^2z \leq \mathbf{E}z^2$ and $\mathbf{E}(z - \mathbf{E}z)^2 \leq \mathbf{E}z^2$ to get that $B = (z - \mathbf{E}z)^2 + \mathbf{E}(z - \mathbf{E}z)^2 \leq 2z^2 + 3\mathbf{E}z^2 = 2(z^2 - \mathbf{E}z^2) + 5\mathbf{E}z^2$. Thus,

$$\mathcal{S}_{2,r} \leq \mathbf{P}\left\{\max_{1 \leq \ell \leq L} \sum_{i=1}^n (z_{\ell,i}^{2^{r+1}} - \mathbf{E}_{i-1} z_{\ell,i}^{2^{r+1}}) \geq \kappa_{r+1}\right\} + \mathbf{P}\left(\max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^{2^{r+1}} \geq \kappa_{r+1}\right).$$

Use the notation from above and then the Markov inequality to get

$$\mathcal{S}_{2,r} \leq \mathcal{P}_{r+1}(\kappa_{r+1}) + \mathbf{P}(D_{r+1} \geq \kappa_{r+1}) \leq \mathcal{P}_{r+1}(\kappa_{r+1}) + \frac{1}{\kappa_{r+1}} \mathbf{E}D_{r+1},$$

which are the last terms of (5.1).

3. *The term $\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}})$.* Apply the inequality $|z| - \mathbf{E}_{i-1}|z| \leq |z|$ and then Boole's and Markov's inequalities to get

$$\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}}) \leq \mathbf{P}\left(\max_{1 \leq \ell \leq L} \sum_{i=1}^n z_{\ell,i}^{2^{\bar{r}}} > \kappa_{\bar{r}}\right) \leq L \max_{1 \leq \ell \leq L} \mathbf{P}\left(\sum_{i=1}^n z_{\ell,i}^{2^{\bar{r}}} > \kappa_{\bar{r}}\right) \leq \frac{L}{\kappa_{\bar{r}}} \max_{1 \leq \ell \leq L} \mathbf{E} \sum_{i=1}^n z_{\ell,i}^{2^{\bar{r}}}.$$

Apply iterated expectations and interchange maximum and expectation to get

$$\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}}) \leq \frac{L}{\kappa_{\bar{r}}} \max_{1 \leq \ell \leq L} \mathbf{E} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^{2^{\bar{r}}} \leq \frac{L}{\kappa_{\bar{r}}} \mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^{2^{\bar{r}}} = \frac{L}{\kappa_{\bar{r}}} \mathbf{E}D_{\bar{r}}.$$

4. *Combine expressions.* Since $\mathcal{P}_{r+1}(\kappa_{r+1}) \leq \mathcal{Q}_{r+1}(\kappa_{r+1})$ then write (5.1) as

$$\begin{aligned} \mathcal{Q}_r(\kappa_r) &\leq 2L \exp\left(-\frac{\kappa_r^2}{14\kappa_{r+1}}\right) + \mathcal{Q}_{r+1}(\kappa_{r+1}) + \frac{\mathbf{E}D_{r+1}}{\kappa_{r+1}} && \text{for } r = 0, \dots, \bar{r} - 2, \\ \mathcal{Q}_r(\kappa_r) &\leq 2L \exp\left(-\frac{\kappa_r^2}{14\kappa_{r+1}}\right) + \mathcal{P}_{r+1}(\kappa_{r+1}) + \frac{\mathbf{E}D_{r+1}}{\kappa_{r+1}} && \text{for } r = \bar{r} - 1. \end{aligned}$$

Then sum from $r = 0$ to $\bar{r} - 1$ and insert the bound $\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}}) \leq \kappa_{\bar{r}}^{-1} \mathbf{E}D_{\bar{r}}$. ■

Lemma 5.2 *For ℓ so $1 \leq \ell \leq L$ let $z_{\ell,i}$ be \mathcal{F}_i -adapted so $\mathbf{E}z_{\ell,i}^{2^{\bar{r}}} < \infty$ for some $\bar{r} \in \mathbb{N}$. Let $D_r = \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^{2^r}$ for $1 \leq r \leq \bar{r}$. Suppose, for some $\lambda, \lambda \geq 0$ and $v > 0$ so $\varsigma < 2v$ and $\varsigma + \lambda < v2^{\bar{r}}$, that (i) $L = O(n^\lambda)$; (ii) $\mathbf{E}D_r = O(n^\varsigma)$ for $r \leq \bar{r}$. Then, for all $\kappa > 0$, it holds as $n \rightarrow \infty$ that*

$$\lim_{n \rightarrow \infty} \mathbf{P}\left\{\max_{1 \leq \ell \leq L} \left| \sum_{i=1}^n (z_{\ell,i} - \mathbf{E}_{i-1} z_{\ell,i}) \right| > \kappa n^v\right\} = 0.$$

Proof of Lemma 5.2. Apply Lemma 5.1 with $\kappa_q = (\kappa n^v)^{2^q} (28\lambda \log n)^{1-2^q}$ for any $\kappa > 0$ so that $\kappa_0 = \kappa n^v$ and $\kappa_q^2 / \kappa_{q+1} = 28\lambda \log n$ and exploit conditions (i, ii) to see that the probability of interest satisfies

$$\mathcal{P}_{n=O} \left\{ n^\lambda \frac{n^\varsigma (\log n)^{2^{\bar{r}}-1}}{n^{v2^{\bar{r}}}} + \sum_{r=1}^{\bar{r}} \frac{n^\varsigma (\log n)^{2^r-1}}{n^{v2^r}} + 2n^{\lambda \bar{r}} n^{-2\lambda} \right\} = o(1),$$

as desired since $\varsigma + \lambda < v2^{\bar{r}}$ and $\varsigma < 2v \leq v2^r$ for $r \geq 1$. ■

Lemma 5.3 For ℓ so $1 \leq \ell \leq L$ let $z_{\ell,i}$ be \mathcal{F}_i -adapted so $\mathbf{E}z_{\ell,i}^4 < \infty$. Suppose $\mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^{2q} \leq Cn$ for $q = 1, 2$ and some $C > 0$. Then it holds, for all $\theta > 0$,

$$\mathbf{P}\left\{\max_{1 \leq \ell \leq L} \left| \sum_{i=1}^n (z_{\ell,i} - \mathbf{E}_{t-1} z_{\ell,i}) \right| > \kappa n^{1/2}\right\} \leq \frac{(L+1)\theta^3 C}{\kappa n} + \frac{\theta C}{\kappa} + 4L \exp\left(-\frac{\kappa\theta}{14}\right).$$

Proof of Lemma 5.3. Apply Lemma 5.1 with $\kappa_q = \kappa n^{2q-1} \theta^{1-2q}$ for any $\kappa, \theta > 0$ so that $\kappa_0 = \kappa n^{1/2}$ and $\kappa_q^2 / \kappa_{q+1} = \kappa\theta$ to get the bound

$$\mathcal{P} \leq \frac{(L+1)\theta^3}{\kappa n^2} \mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^4 + \frac{\theta}{\kappa n} \mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1} z_{\ell,i}^2 + 4L \exp\left(-\frac{\kappa\theta}{14}\right).$$

Exploit the moment conditions to get the desired result. ■

6 A metric on \mathbb{R} and some inequalities

A metric is set up that will be used for the chaining argument. Then a number of inequalities are shown, mostly related to this metric.

Introduce the function

$$J_{i,p}(x, y) = (\varepsilon_i / \sigma)^p \{1_{(\varepsilon_i \leq \sigma y)} - 1_{(\varepsilon_i \leq \sigma x)}\}, \quad (6.1)$$

where $p \in \mathbb{N}_0$ and ε_i / σ has density f . We will be interested in powers of $J_{i,p}(x, y)$ of order 2^r where $r \in \mathbb{N}$ was chosen in Assumption 4.1(i). Note that $2^r p$ is even for $p \in \mathbb{N}_0$ and $r \in \mathbb{N}$ so that $\varepsilon_i^{2^r p}$ is non-negative. Thus, define the increasing function

$$\mathbf{H}_r(x) = \int_{-\infty}^x (1 + \varepsilon^{2^r p}) f(\varepsilon) d\varepsilon,$$

with derivative $\mathbf{H}'_r(x) = (1 + x^{2^r p})f(x)$, along with the constant

$$H_r = \mathbf{H}_r(\infty) = \int_{-\infty}^{\infty} (1 + \varepsilon^{2^r p}) f(\varepsilon) d\varepsilon.$$

It follows that, for $x \leq y$ and $0 \leq s \leq r$ then

$$0 \leq |\mathbf{E}\{J_{i,p}(x, y)\}^{2^s}| \leq \mathbf{E}\{|J_{i,p}(x, y)|^{2^s}\} < \mathbf{H}_r(y) - \mathbf{H}_r(x). \quad (6.2)$$

noting that, for $q \geq p \geq 0$ and $\varepsilon \in \mathbb{R}$, then $|\varepsilon^q| < 1 + |\varepsilon|^q$.

For the chaining, partition the range of $\mathbf{H}_r(c)$ into K intervals of equal size. That is, partition the support into K intervals defined by the endpoints

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \quad (6.3)$$

and for $1 \leq k \leq K$,

$$\mathbf{E}[\{J_{i,p}(c_{k-1}, c_k)\}^{2^r}] = \mathbf{H}_r(c_k) - \mathbf{H}_r(c_{k-1}) = \frac{H_r}{K}.$$

Let $c_{-k} = c_0$ for $k \in \mathbb{N}$.

The number of intervals K will be chosen so large that c_-, c_+ exists which are (weakly) separated from zero by grid points in the sense that $c_{k-1} \leq c_- \leq c_k \leq 0$ and $0 \leq c_{k+1} \leq c_+ \leq c_k$ and so that

$$\mathbf{H}'_r(c_-) = \mathbf{H}'_r(c_+) = H_r / (C_H K^{1/2}). \quad (6.4)$$

This can be done for sufficiently large K since f is continuous and since the function $\mathbf{H}'_r(c)$ is integrable by Assumption 4.1(ia).

The first inequality concerns the \mathbf{H}_r -distance of certain perturbations of the $]c_{k-1}, c_k]$ intervals.

Lemma 6.1 *Suppose Assumption 4.1(i) holds with $\nu = 1$ only. Then a constant $C > 0$ exists so that for all K satisfying (6.4) then*

$$\sup_{1 \leq k \leq K} \sup_{|d| \leq K^{-1/2}} \{H_r(c_k + d) - H_r(c_{k-1} + d)\} \leq CH_r/K.$$

Proof of Lemma 6.1. 1. *Definitions.* Consider positive c_k only with a similar argument for negative c_k . Let $\mathcal{H} = H_r(c_k + d) - H_r(c_{k-1} + d)$. Let $H'_r(c) = (1 + c^{2^r p})f(c)$ and

$$\underline{H}'_r(c) = \inf_{0 \leq d \leq c} H'_r(d), \quad \overline{H}'_r(c) = \sup_{d \geq c} H'_r(d),$$

which are decreasing in c . Assumption 4.1(ic) then implies

$$C_H^{-1} \overline{H}'_r(c) \leq \underline{H}'_r(c) \leq H'_r(c) \leq \overline{H}'_r(c) \leq C_H H'_r(c). \quad (6.5)$$

Since $H''_r(c) = 2^r p c^{2^r p - 1} f(c) + (1 + c^{2^r p}) f'(c)$ then Assumption 4.1(ib) gives

$$\sup_{c \in \mathbb{R}} |H''_r(c)| < \infty. \quad (6.6)$$

2. *Apply the mean-value theorem* to get, for some c_ℓ^* so $c_{\ell-1} \leq c_\ell^* \leq c_\ell$, that

$$H_r/K = H_r(c_\ell) - H_r(c_{\ell-1}) = (c_\ell - c_{\ell-1}) H'_r(c_\ell^*). \quad (6.7)$$

Two inequalities for $H'_r(c)$ arise from (6.5) and condition (6.4). These are

$$H'_r(c) \leq \overline{H}'_r(c) \leq \overline{H}'_r(c_+) \leq C_H H'_r(c_+) = H_r/K^{1/2} \quad \text{for } c \geq c_+, \quad (6.8)$$

$$H'_r(c) \geq \underline{H}'_r(c) \geq \underline{H}'_r(c_+) \geq \overline{H}'_r(c_+)/C_H \geq H'_r(c_+)/C_H = H_r/(C_H^2 K^{1/2}) \quad \text{for } 0 \leq c \leq c_+. \quad (6.9)$$

In parallel to (6.9) which is derived for positive c it holds for negative c that

$$H'_r(c) \geq H_r/(C_H^2 K^{1/2}) \quad \text{for } 0 \geq c \geq c_-. \quad (6.10)$$

3. *Small arguments* $c_- \leq c_k^* \leq c_+$. Combine (6.7), (6.9) and (6.10) to get

$$c_k - c_{k-1} = H_r/\{K H'_r(c_k^*)\} \leq C_H^2/K^{1/2}. \quad (6.11)$$

Two second order Taylor expansions give

$$\begin{aligned} H_r(c_k + d) - H_r(c_k) &= d H'_r(c_k) + (d^2/2) H''_r(c_k^{**}), \\ H_r(c_{k-1} + d) - H_r(c_{k-1}) &= d H'_r(c_{k-1}) + (d^2/2) H''_r(c_{k-1}^{**}), \end{aligned}$$

where c_k^{**}, c_{k-1}^{**} satisfy $\max(|c_k^* - c_k|, |c_{k-1}^* - c_{k-1}|) \leq |d| \leq K^{-1/2}$. The difference is, when recalling the definition of \mathcal{H} in item 1,

$$\mathcal{H} - \{H_r(c_k) - H_r(c_{k-1})\} = d\{H'_r(c_k) - H'_r(c_{k-1})\} + (d^2/2)\{H''_r(c_k^{**}) - H''_r(c_{k-1}^{**})\}.$$

It holds $H_r(c_k) - H_r(c_{k-1}) = H_r/K$. The mean-value theorem gives that for a \tilde{c}_k so $c_{k-1} \leq \tilde{c}_k \leq c_k$ then $H'_r(c_k) - H'_r(c_{k-1}) = (c_k - c_{k-1}) H''_r(\tilde{c}_k)$. Insert this and rearrange to get

$$0 \leq \mathcal{H} = \frac{H_r}{K} + d(c_k - c_{k-1}) H''_r(\tilde{c}_k) + \frac{d^2}{2} \{H''_r(c_k^{**}) - H''_r(c_{k-1}^{**})\}.$$

Using the bound to $c_k - c_{k-1} \leq C_{\mathbb{H}}^2/K^{1/2}$ from (6.11) and the bound $|d| \leq K^{-1/2}$ it follows that $0 \leq \mathcal{H} \leq C/K$ where $C = H_r + (C_{\mathbb{H}}^2 + 1) \sup_{c \in \mathbb{R}} |H_r''(c)|$ does not depend on K .

4. *Large arguments* $c_k^* \geq c_+$ so $k \geq k_+ + 2$. Expansion (6.7) and inequality (6.8) imply

$$c_k - c_{k-1} = H_r/\{KH_r'(c_k^*)\} \geq K^{-1/2} \geq |d|.$$

The same holds for $c_{k+1} - c_k$ and $c_{k-1} - c_{k-2}$. Therefore

$$\begin{aligned} c_k + d &\leq c_k + |d| \leq c_k + c_{k+1} - c_k = c_{k+1}, \\ c_{k-1} + d &\geq c_{k-1} - |d| \geq c_{k-1} - (c_{k-1} - c_{k-2}) = c_{k-2}. \end{aligned}$$

It then holds that $0 \leq \mathcal{H} \leq H_r(c_{k+1}) - H_r(c_{k-2}) = C/K$, where $C = 3H_r$ does not depend on K .

5. *Intermediate arguments* $c_k^* \geq c_+$ and $k_+ \leq k \leq k_+ + 1$. In this case $c_{k_+-1} \leq c_+ \leq c_k^* \leq c_k \leq c_{k_++1}$. Consider the length of the interval $]c_{k_+-1}, c_+]$. The mean-value theorem shows

$$H_r/K = H_r(c_{k_+}) - H_r(c_{k_+-1}) \geq H_r(c_+) - H_r(c_{k_+-1}) = (c_+ - c_{k_+-1})H_r'(c_{k_+}^*),$$

for $0 \leq c_{k_+-1} \leq c_{k_+}^* \leq c_+$. Insert the inequality (6.9) and rearrange to get

$$c_+ - c_{k_+-1} \leq C_{\mathbb{H}}^2 K^{-1/2}. \quad (6.12)$$

Now, rewrite $0 \leq \mathcal{H} = \mathcal{H}_1 - \mathcal{H}_2$ where

$$\mathcal{H}_1 = H_r(c_k + d) - H_r(c_+), \quad \mathcal{H}_2 = H_r(c_{k-1} + d) - H_r(c_+).$$

For the term \mathcal{H}_1 note that following the argument in item 5 then $c_k - c_{k-1}$ and $c_{k+1} - c_k$ are greater than $|d|$, so that $c_{k_+-1} \leq c_{k-1} \leq c_k + d \leq c_{k+1} \leq c_{k_++2}$. Since $c_{k_+-1} \leq c_+ \leq c_{k_+}$ it holds $|\mathcal{H}_1| \leq H_r(c_{k_++2}) - H_r(c_{k_+-1}) \leq 3H_r/K$.

For the term \mathcal{H}_2 use the mean value theorem to get $\mathcal{H}_2 = \delta_{k,d}H_r'(c_+) + (\delta_{k,d}^2/2)H_r''(c^{**})$, where $\delta_{k,d} = c_{k-1} + d - c_+$ while c^{**} satisfies $|c^{**} - c_+| \leq |\delta_{k,d}|$. For the linear term note that (6.12) and the bound $|d| \leq K^{-1/2}$ imply $|\delta_{k,d}| \leq (C_{\mathbb{H}}^2 + 1)K^{-1/2}$, whereas (6.4) shows $H_r'(c_+) = H_r/(C_{\mathbb{H}}K^{1/2})$. For the quadratic term note that $\delta_{k,d}^2 \leq (C_{\mathbb{H}}^2 + 1)^2K^{-1}$, while $H_r''(c^{**})$ is bounded by (6.6). Therefore $|\mathcal{H}_2| \leq K^{-1}\{(C_{\mathbb{H}}^2 + 1)H_r/C_{\mathbb{H}} + (C_{\mathbb{H}}^2 + 1)^2 \sup_{c \in \mathbb{R}} |H_r''(c)|/2\}$.

Combine to get $\mathcal{H} \leq |\mathcal{H}_1| + |\mathcal{H}_2| \leq C/K$ for some constant C not depending on K . ■

The next lemma shows how small fluctuations in the arguments of the function $J_{i,p}$ can be controlled in terms of $J_{i,p}$ functions defined on the grid points. The proof uses Lemma 6.1.

Lemma 6.2 *Suppose Assumption 4.1(i) holds with $\nu = 1$ only. For any c let c_k be a right grid point for $k < K$ that is $c_{k-1} < c \leq c_k$ and let c_k is a left grid point for $c > c_{K-1}$ so $k = K - 1$. Then an integer $k_J > 0$ exists so that for all K satisfying (6.4) and all $c, d, d_m \in \mathbb{R}$ so $|d| \leq K^{-1/2}$ and $|d - d_m| \leq K^{-1}$ then integers k^\dagger, k^\ddagger exists so*

$$|J_{i,p}(c, c + d) - J_{i,p}(c_k, c_k + d_m)| \leq |J_{i,p}(c_{k-k_J}, c_k)| + |J_{i,p}(c_{k^\dagger-k_J}, c_{k^\dagger})| + |J_{i,p}(c_{k^\ddagger-k_J}, c_{k^\ddagger})|.$$

Proof of Lemma 6.2. 1. *Decomposition.* Only the case $k < K$ is proved. The proof for $k = K$ is similar. Let $\sigma = 1$ for notational simplicity. Write

$$\mathcal{J} = J_{i,p}(c, c + d) - J_{i,p}(c_k, c_k + d_m) = \varepsilon_i^p(\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3),$$

in terms of indicator functions $\mathcal{I}_1 = 1_{(c < \varepsilon_i \leq c_k)}$, $\mathcal{I}_2 = 1_{(\varepsilon_i \leq c_k + d)} - 1_{(\varepsilon_i \leq c_k + d_m)}$ and $\mathcal{I}_3 = 1_{(c + d < \varepsilon_i \leq c_k + d)}$. It follows that $|\mathcal{J}| \leq |\varepsilon_i^p|(|\mathcal{I}_1| + |\mathcal{I}_2| + |\mathcal{I}_3|)$.

2. *Bound for \mathcal{I}_1 .* Since $c_{k-1} < c \leq c_k$ then $0 \leq \mathcal{I}_1 = 1_{(c < \varepsilon_i \leq c_k)} \leq 1_{(c_{k-1} < \varepsilon_i \leq c_k)}$.

3. *Bound for \mathcal{I}_2 .* Write $d = d_m + (d - d_m)$ where $|d - d_m| \leq K^{-1}$. Let $c^\dagger = c_k + d_m$. Then it holds $|\mathcal{I}_2| \leq 1_{(c^\dagger - K^{-1} < \varepsilon_i \leq c^\dagger + K^{-1})}$. Using first this inequality and then the mean value theorem it holds

$$\mathcal{E}_2 = \mathbb{E}(|\varepsilon_i^p \mathcal{I}_2|) \leq H_r(c^\dagger + K^{-1}) - H_r(c^\dagger - K^{-1}) \leq 2H_r^{-1} \sup_{c \in \mathbb{R}} H'_r(c) H_r / K.$$

Therefore, a k^\dagger exists so $|\mathcal{I}_2| \leq 1_{(c_{k^\dagger - k_J} < \varepsilon_i \leq c_{k^\dagger})}$ where $k_J \leq 2H_r^{-1} \sup_{c \in \mathbb{R}} H'_r(c) + 2$.

4. *Bound for \mathcal{I}_3 .* Since $c_{k-1} < c \leq c_k$ then $\mathcal{I}_3 \leq 1_{(c_{k-1} + d < \varepsilon_i \leq c_k + d)}$. Using first this inequality and then Lemma 6.1 noting that $|d| \leq K^{-1/2}$ it holds

$$\mathcal{E}_3 = \mathbb{E}(|\varepsilon_i^p \mathcal{I}_3|) \leq H_r(c_k + d) - H_r(c_{k-1} + d) \leq CH_r / K.$$

Therefore, a k^\ddagger exists so $|\mathcal{I}_3| \leq 1_{(c_{k^\ddagger - k_J} < \varepsilon_i \leq c_{k^\ddagger})}$ where $k_J \leq C + 1$. ■

The next inequality gives a tightness type result for the function H_r .

Lemma 6.3 *Let $c_\psi = F^{-1}(\psi)$. For all densities satisfying Assumption 4.1(ia) for some $\nu < 1$, then a $C_\nu > 0$ exists so that for all $0 \leq \phi \leq 1$ it holds*

$$\max_{0 \leq \psi \leq 1 - \phi} \{H_r(c_{\psi + \phi}) - H_r(c_\psi)\} \leq C_\nu \phi^{1 - \nu}.$$

Proof of Lemma 6.3. Let $\psi_0 = F(0)$. Note that $2^r p$ is even for $r \in \mathbb{N}$, $p \in \mathbb{N}_0$.

1. *Let $\psi \geq \psi_0$.* Then $H_r(c_{\psi + \phi}) - H_r(c_\psi)$ is increasing in ψ since

$$\frac{d}{d\psi} \{H_r(c_{\psi + \phi}) - H_r(c_\psi)\} = \frac{H'_r(c_{\psi + \phi})}{f(c_{\psi + \phi})} - \frac{H'_r(c_\psi)}{f(c_\psi)} = c_{\psi + \phi}^{p2^r} - c_\psi^{p2^r} > 0.$$

Thus, $\max_{\psi_0 \leq \psi \leq 1 - \phi} \{H_r(c_{\psi + \phi}) - H_r(c_\psi)\} \leq H_r(\infty) - H_r(c_{1 - \phi})$. This bound satisfies

$$H_r(\infty) - H_r(c_{1 - \phi}) = \int_{c_{1 - \phi}}^{\infty} (1 + \varepsilon^{p2^r}) f(\varepsilon) d\varepsilon = \phi + \int_{c_{1 - \phi}}^{\infty} \varepsilon^{p2^r} f(\varepsilon) d\varepsilon.$$

Assumption 4.1(ia) shows $\mathbb{E} \varepsilon^{p2^r / \nu} \leq C$ for some $C > 0$ so $1 - F(\varepsilon) \leq C \varepsilon^{-p2^r / \nu}$ by the Chebychev inequality. Hence, $\varepsilon^{p2^r} \leq C^\nu \{1 - F(\varepsilon)\}^{-\nu}$, so that

$$H_r(\infty) - H_r(c_{1 - \phi}) \leq \phi + C^\nu \int_{c_{1 - \phi}}^{\infty} \{1 - F(\varepsilon)\}^{-\nu} f(\varepsilon) d\varepsilon$$

Substituting $\psi = F(\varepsilon)$ so $d\psi = f(\varepsilon) d\varepsilon$ gives

$$H_r(\infty) - H_r(c_{1 - \phi}) \leq \phi + C^\nu \int_{1 - \phi}^1 (1 - x)^{-\nu} dx = \phi + \frac{C^\nu}{1 - \nu} \phi^{1 - \nu}.$$

2. *Let $\psi \leq \psi_0 - \phi$.* Apply a similar argument as in item 1, to show that $H_r(c_{\psi + \phi}) - H_r(c_\psi)$ is decreasing because $c_\psi < c_{\psi + \phi} \leq 0$. Thus, $H_r(c_\phi) - H_r(-\infty)$ satisfies the same bound.

3. *Let $\psi_0 - \phi \leq \psi \leq \psi_0$.* Then

$$\mathcal{H} = \max_{\psi_0 - \phi \leq \psi \leq \psi_0} \{H_r(c_{\psi + \phi}) - H_r(c_\psi)\} \leq H_r(c_{\psi_0 + \phi}) - H_r(c_{\psi_0 - \phi}).$$

Using the mean value theorem then, for some ψ^* so $\psi_0 - \phi \leq \psi^* \leq \psi_0 + \phi$,

$$\mathcal{H} \leq H_r\{F^{-1}(\psi_0 + \phi)\} - H_r\{F^{-1}(\psi_0 - \phi)\} = H_r\{F^{-1}(\psi^*)\} 2\phi \leq 2H_r \phi.$$

4. *Combine results.* Note that $\phi \leq \phi^{1 - \nu}$. Let $C_\nu = \max\{2H_r, 1 + C^\nu / (1 - \nu)\}$. ■

7 Proofs of auxillary Theorems 4.1–4.4

Proof of Theorem 4.1. Without loss of generality let $\sigma = 1$. Let $\tilde{R}(b, c_\psi) = \mathbb{F}_n^{g,p}(b, c_\psi) - \mathbb{F}_n^{g,p}(0, c_\psi)$ and $\mathcal{R}_n = \sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta}B} |\mathbb{F}_n^{g,p}(b, c_\psi) - \mathbb{F}_n^{g,p}(0, c_\psi)|$.

1. *Partition the support.* For $\delta, n > 0$ partition axis as laid out in (6.3) with $K = \text{int}(H_r n^{1/2}/\delta)$ using Assumption 4.1(*ia*) with $\nu = 1$ only.

2. *Assign c_ψ to the partitioned support.* Consider $1/2 \leq \psi \leq 1$ only, noting that a similar argument can be made for $0 \leq \psi \leq 1/2$. Thus, for each c_ψ there exists c_{k-1}, c_k so $c_{k-1} < c_\psi \leq c_k$.

3. *Construct b -balls.* For a $\zeta > \kappa$ cover the set $|b| \leq n^{1/4-\eta}B$ with $M = O\{n^{(1/4-\eta+\zeta)\dim x}\}$ balls of radius $n^{-\zeta}$ with centers b_m . Thus, for any b there exists a b_m so $|b - b_m| < n^{-\zeta}$.

4. *Apply chaining.* For $k < K$ so $c_\psi \leq c_{K-1}$ then relate c_ψ to the nearest right grid point so $\tilde{R}(b, c_\psi) = \tilde{R}(b_m, c_k) + \{\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_k)\}$, whereas for $k = K$ so $c_\psi > c_{K-1}$ related c to nearest left grid point so $\tilde{R}(b, c_\psi) = \tilde{R}(b_m, c_{K-1}) + \{\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_{K-1})\}$. Therefore $\mathcal{R}_n \leq \sum_{j=1}^3 \mathcal{R}_{n,j}$, where

$$\begin{aligned} \mathcal{R}_{n,1} &= \max_{1 \leq k < K} \max_{1 \leq m \leq M} |\tilde{R}(b_m, c_k)|, \\ \mathcal{R}_{n,2} &= \max_{1 \leq k < K} \max_{1 \leq m \leq M} \sup_{c_{k-1} < c_\psi \leq c_k} \sup_{|b-b_m| < n^{-\zeta}} |\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_k)| \\ &\quad + \max_{1 \leq m \leq M} \sup_{c_{K-1} < c_\psi} \sup_{|b-b_m| < n^{-\zeta}} |\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_{K-1})|. \end{aligned}$$

Thus, it suffices to show that $P(\mathcal{R}_{n,j} > \gamma)$ vanishes for $j = 1, 2$.

5. *The term $\mathcal{R}_{n,1}$.* Use Lemma 5.2 to see that $\mathcal{R}_{n,1} = o_{\mathbb{P}}(1)$. To see this let $v = 1/2$ and let g_{in} have coordinates g_{in}^* . Then write $\tilde{R}(b_m, c_k)$ as $n^{-1/2} \sum_{i=1}^n (z_{li} - \mathbf{E}_{i-1} z_{li})$ with $z_{li} = g_{in}^* J_{i,p}(c_k, c_k + \sigma^{-1} x'_{in} b_m)$, see definition in (6.1), and where ℓ represents the indices k, m . The conditions of Lemma 5.2 need to be demonstrated.

The parameter λ . The set of indices ℓ has size $L = O(n^\lambda)$ where $\lambda = 1/2 + (1/4 - \eta + \zeta) \dim x$ since $K = O(n^{1/2})$ and $M = O\{n^{(1/4-\eta+\zeta)\dim b}\}$.

The parameter ς . Since $|1_{(\varepsilon_i \leq c_k + x'_{in} b_m)} - 1_{(\varepsilon_i \leq c_k)}| \leq 1_{(c_k - |x_{in}| |b_m| < \varepsilon_i \leq c_k + |x_{in}| |b_m|)}$ then, for $1 \leq q \leq r$,

$$\mathbf{E}_{i-1}(J_{i,p})^{2q} \leq H_r(c_k + |x_{in}| |b_m|) - H_r(c_k - |x_{in}| |b_m|) \leq 2|x_{in}| |b_m| \sup_{v \in \mathbb{R}} H'_r(v),$$

when using the mean-value theorem. Since $|b_m| \leq n^{1/4-\eta}B$ while $\sup_{v \in \mathbb{R}} H'_r(v) < \infty$ by Assumption 4.1(*ib*) then

$$D_q = \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1}(z_{li})^{2q} \leq C_1(n^{-1} \sum_{i=1}^n |g_{in}^*|^{2q} |n^{1/2} x_{in}|) n^{3/4-\eta}. \quad (7.1)$$

Thus, $\mathbf{E} D_q = O(n^\varsigma)$ where $\varsigma = 3/4 - \eta$ by Assumption 4.1(*iii*).

Condition (i) is that $\varsigma < 2v$. This holds since $0 < \eta$ so that $\varsigma = 3/4 - \eta < 1 = 2v$.

Condition (ii) is that $\varsigma + \lambda < 2v$. If $\zeta > \kappa$ is chosen sufficiently small then

$$\varsigma + \lambda = 1 + (1/4 + \kappa - \eta)(1 + \dim x) + (\zeta - \kappa) \dim x - \kappa < v 2^r = 2^{r-1},$$

provided r is chosen so $2^{r-1} \geq 1 + (1/4 + \kappa - \eta)(1 + \dim x)$.

6. *Decompose $\mathcal{R}_{n,2}$.* It will be argued that $\mathcal{R}_{n,2} \leq 3(\tilde{\mathcal{R}}_{n,2} + 2\overline{\mathcal{R}}_{n,2}) + o_{\mathbb{P}}(1)$, where

$$\tilde{\mathcal{R}}_{n,2} = \max_{1 \leq k \leq K} n^{-1/2} \sum_{i=1}^n |g_{in}| \{ |J_{i,p}(c_{k-k_J}, c_k)| - \mathbf{E}_{i-1} |J_{i,p}(c_{k-k_J}, c_k)| \}, \quad (7.2)$$

$$\overline{\mathcal{R}}_{n,2} = \max_{1 \leq k \leq K} n^{-1/2} \sum_{i=1}^n |g_{in}| \mathbf{E}_{i-1} |J_{i,p}(c_{k-k_J}, c_k)|. \quad (7.3)$$

To see this, let c_k denote nearest right grid point for $c_\psi \leq c_{K-1}$ while $c_k = c_{K-1}$ for $c_\psi > c_{K-1}$. Note first that $\tilde{R}_F^p(b, c_\psi) - \tilde{R}_F^p(b_m, c_k)$ involves the functions

$$\mathcal{J}_i = J_{i,p}(c_\psi, c_\psi + x'_{in}b) - J_{i,p}(c_k, c_k + x'_{in}b_m).$$

Assumption 4.1(ii) gives that $\max_{1 \leq i \leq n} |x_{in}| = O_P(n^{\kappa-1/2})$. Thus, for all $\epsilon > 0$ an $C_x > 0$ exists so that the set $(\max_{1 \leq i \leq n} |x_{in}| \leq n^{\kappa-1/2}C_x)$ has probability of at least $1 - \epsilon$. On that set and with $d = x'_{in}b$ and $d = x'_{in}b$ then $|d| = O(n^{-1/4+\kappa-\eta}) = o(K^{-1/2})$ for $\eta - \kappa > 0$ and $|d - d_m| = O(n^{-1/2+\kappa-\zeta}) = o(K^{-1})$ for $\zeta - \kappa > 0$. Thus, for sufficiently large n then $|d| < K^{-1/2}$ and $|d - d_m| < K^{-1}$. Lemma 6.2 using Assumption 4.1(i) then shows that a k_J exists so that for all c, d, d_m there exist k^\dagger, k^\ddagger so

$$|\mathcal{J}_i| \leq |J_{i,p}(c_{k-k_J}, c_k)| + |J_{i,p}(c_{k^\dagger-k_J}, c_{k^\dagger})| + |J_{i,p}(c_{k^\ddagger-k_J}, c_{k^\ddagger})|, \quad (7.4)$$

As a consequence it holds, as desired, $\mathcal{R}_{n,2} \leq 3(\tilde{\mathcal{R}}_{n,2} + 2\bar{\mathcal{R}}_{n,2}) + o_P(1)$.

7. *The term $\tilde{\mathcal{R}}_{n,2}$* is $o_P(1)$ by Lemma 5.2. To see this note that $\tilde{\mathcal{R}}_{n,2}$ is the maximum of a family of martingale of the required form with $\ell = k$ so $L = K$ and $z_{\ell i} = |g_{in}| |J_{i,p}(c_{k-k_J}, c_k)|$ and it suffices to set $\bar{r} = 2$.

Condition (i) holds with $\lambda = 1/2$ since $K = \text{int}(H_r n^{1/2}/\delta)$.

Condition (ii) holds with $\zeta = 1/2$ since $\mathbf{E}_{i-1}(J_{i,p})^{2^{\bar{r}}} \leq H_r(c_k) - H_r(c_{k-k_J}) = k_J H_r / K$ so that $\sum_{i=1}^n \mathbf{E}_{i-1}(J_{i,p})^{2^{\bar{r}}} = O(n^{1-1/2})$, uniformly in ℓ, i .

It holds that $\lambda + \zeta = 1$ which is less than $2^{\bar{r}} = 4$.

8. *Bounding $\bar{\mathcal{R}}_{n,2}$* . Note $\mathbf{E}_{i-1}|J_{i,p}(c_{k-k_J}, c_k)| \leq 2k_J \delta n^{-1/2}$ uniformly in i, k by the same argument as in item 7. It follows that $\bar{\mathcal{R}}_{n,2} \leq 2k_J \delta n^{-1} \sum_{i=1}^n |g_{in}|$. Here $n^{-1} \sum_{i=1}^n |g_{in}| = O_P(1)$ by Markov's inequality and Assumption 4.1(iiii), so that $\bar{\mathcal{R}}_{n,2} = O_P(\delta)$. Thus, choosing δ sufficiently small then $\bar{\mathcal{R}}_{n,2}$ is small in probability. ■

Proof of Theorem 4.2. It suffices to show, for all $\omega < \eta - \kappa$ where $\eta - \kappa \leq 1/4$, that

$$\begin{aligned} \mathcal{S}_1 &= \sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta} B} \sup_{d \in \mathbb{R}} |\mathbb{F}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{F}_n^{1,0}(0, c_\psi + n^{\kappa-1/2}d)| = o_P(n^{-\omega}), \\ \mathcal{S}_2 &= \sup_{0 \leq \psi \leq 1} \sup_{|d| \leq n^{1/4-\eta} B} |\mathbb{F}_n^{1,0}(0, c_\psi + n^{\kappa-1/2}d) - \mathbb{F}_n^{1,0}(0, c_\psi)| = o_P(n^{-\omega}). \end{aligned}$$

For each term the proof of Theorem 4.1 is used with minor modifications. Since $p = 0$ then $2^r p = 0$ for all r , which simplifies the assumptions.

A. *The term \mathcal{S}_1* . The steps of the proof of Theorem 4.1 are modified as follows.

1. Choose $K = \text{int}(H_r n^{1/2+\omega}/\delta)$ where $\omega < \eta - \kappa \leq 1/4$.
2. For each $c_\psi + n^{\kappa-1/2}d$ there exists c_{k-1}, c_k depending on n so $c_{k-1} < c_\psi + n^{\kappa-1/2}d \leq c_k$.
3. Choose $\zeta \geq \eta$ which implies $\zeta > \kappa$ since $\kappa < \eta$. The b -set is now $|b| \leq n^{1/4+\kappa-\eta} B$ so that the number of b -balls is $M = O\{n^{(1/4+\kappa-\eta+\zeta)\dim x}\}$.
4. Note that in the chaining argument c_ψ is replaced by $c_\psi + n^{\kappa-1/2}d$. This only affects $\mathcal{R}_{n,2}$.
5. The term $\mathcal{R}_{n,1}$. Use Lemma 5.2 to see that $\mathcal{R}_{n,1} = o_P(n^{-\omega})$, now using $v = 1/2 - \omega > 1/2 + \kappa - \eta$. Define $z_{\ell i}$ as before. Since $p = 0$, $g_{in} = 0$ then $|J_{i,p}(x, y)|^{2^r} = |J_{i,p}(x, y)|$ and $|z_{\ell i}^{2^r}| = |z_{\ell i}|$ for any $r \in \mathbb{N}_0$. The inequality (7.1) for D_q holds as before, uniformly in $q \in \mathbb{N}$ so $\zeta = 3/4 - \eta$, but $\lambda = 1/2 + \omega + (1/4 + \kappa - \eta + \zeta)\dim x$. Condition (i) holds since $\eta \leq 1/4$ and $\kappa \geq 0$ so $\zeta = 3/4 - \eta \leq 1 + \kappa - 2\eta < 2v$. Condition (ii) holds since $\zeta + \lambda < \infty$ while $v > 0$. Thus, for any ζ and sufficiently large \bar{r} then $\zeta + \lambda < v2^{\bar{r}}$.

6. Lemma 6.2 is an analytic result holding in finite samples. So the argument is not affect the dependence of c_k on n through $c_\psi + n^{\kappa-1/2}d$. In particular, (7.4) holds as stated and therefore the decomposition of $\mathcal{R}_{n,2}$ holds, noting that K is now chosen differently.

7. Apply Lemma 5.2 with $\bar{r} = 2$, but with λ, ς chosen differently. *Condition (i)* holds with $\lambda = 1/2 + \omega$ holds since $K = \text{int}(H_r n^{1/2+\omega}/\delta)$. *Condition (ii)* holds with $\varsigma = 1/2 - \omega$ since $\mathbf{E}_{i-1}(J_{i,p})^4 = \mathbf{E}_{i-1}(J_{i,p}) \leq H_r(c_k) - H_r(c_{k-k_J}) = k_J H_r/K$ so that $\sum_{i=1}^n \mathbf{E}_{i-1}(J_{i,p})^4 = O(n^{1-1/2-\omega})$, uniformly in ℓ, i . It holds that $\lambda + \varsigma = 1$ which is less than $2^2(1/2 - \omega)$ for all $\omega < 1/4$. Lemma 5.2 then shows $\tilde{\mathcal{R}}_{n,2} = o_{\mathbf{P}}(n^{-\omega})$ for all $\omega < 1/4$.

8. Note $\mathbf{E}_{i-1}|J_{i,p}(c_{k-k_J}, c_k)| \leq 2k_J \delta n^{-\omega-1/2}$ uniformly in i, k by the same argument as in item 7. It follows that $\tilde{\mathcal{R}}_{n,2} \leq (n^{-1} \sum_{i=1}^n |g_{in}|^{2r}) n^{-\omega} = o_{\mathbf{P}}(n^{-\omega})$.

B. The term \mathcal{S}_2 . Rewrite

$$\mathcal{S}_2 = \sup_{0 \leq \psi \leq 1} \sup_{|d| \leq n^{1/4-\eta} B} |\mathbb{F}_n^{1,0}(0, c_\psi + n^{\kappa-1/2}d) - \mathbb{F}_n^{1,0}(0, c_\psi)|.$$

Choosing the regressor as $x_{in}^* = n^{\kappa-1/2}$, then $\mathbb{F}_n^{1,0}(0, c_\psi + n^{\kappa-1/2}d) = \mathbb{F}_n^{1,0}(d, c_\psi)$. Apply the argument of part A. ■

Proof of Theorem 4.3. The expression of interest is

$$R(b, c_\psi) = n^{1/2} \{ \bar{\mathbb{F}}_n^{g,p}(b, c_\psi) - \bar{\mathbb{F}}_n^{g,p}(0, c_\psi) \} - \sigma^{p-1} c_\psi^p \mathbf{f}(c_\psi) n^{-1} \sum_{i=1}^n g_{in} n^{1/2} x'_{in} b.$$

Recalling the definition of $\bar{\mathbb{F}}_n^{g,p}$ from (4.2) this satisfies $R(b, c_\psi) = n^{-1/2} \sum_{i=1}^n g_{in} \mathcal{S}_i(b, c_\psi)$ where

$$\mathcal{S}_i(b, c_\psi) = \mathbf{E}_{i-1}[\varepsilon_i^p \{ 1_{(\varepsilon_i \leq \sigma c_\psi + b' x_{in})} - 1_{(\varepsilon_i \leq \sigma c_\psi)} \}] - \sigma^{p-1} x'_{in} b c_\psi^p \mathbf{f}(c_\psi).$$

A bound is needed for $\mathcal{S}_i(b, c_\psi)$. Let $h_{in} = \sigma^{-1} x'_{in} b$ and $g(c) = c^p \mathbf{f}(c)$. Write $\mathcal{S}_i(b, c_\psi)$ as an integral and Taylor expand to second order to get

$$\mathcal{S}_i(b, c_\psi) = \int_{c_\psi}^{c_\psi + h_{in}} g(c) dc - h_{in} g(c_\psi) = \frac{1}{2} h_{in}^2 g'(c^*),$$

for an intermediate point so $|c^* - c_\psi| \leq |h_{in}|$. Exploit the bound $|b| \leq n^{1/4-\eta} B$ to get

$$|\mathcal{S}_i(b, c_\psi)| \leq \frac{1}{2} \sigma^{-2} |b|^2 |x_{in}|^2 \sup_{c \in \mathbb{R}} |g'(c^*)| = |x_{in}|^2 \sup_{c \in \mathbb{R}} |g'(c)| O(n^{1/2-2\eta}).$$

Thus, by the triangular inequality then

$$|R(b, c_\psi)| \leq n^{-1/2} \sum_{i=1}^n |g_{in}| |\mathcal{S}_i(b, c_\psi)| \leq O(n^{-2\eta}) n^{-1} \sum_{i=1}^n |g_{in}| |n^{1/2} x_{in}|^2 \sup_{c \in \mathbb{R}} |g'(c)|.$$

Due to Assumption 4.1(*ib, iib*), this expression is of order $O_{\mathbf{P}}(n^{-2\eta})$ uniformly in ψ, b . ■

Proof of Theorem 4.4. 1. *Coefficients $\sigma, \epsilon, \phi, r$.* Without loss of generality let $\sigma = 1$ and $0 < \phi < 1$ and $\epsilon < 1$. Since $\psi^\dagger - \psi \leq \phi$ then Lemma 6.3 with Assumption 4.1(*ia*) shows that $0 < \nu < 1$ and $C_1 > 0$ exist so $H_r(c_{\psi^\dagger}) - H_r(c_\psi) \leq C_1 \phi^{1-\nu}$. Now, take $0 < \epsilon$ and n as well as $0 < \phi^{(1-\nu)/4} \leq \epsilon^2$ as given. Throughout, constants $C_j > 0$ for $j = 1, 2, \dots$ will be found not depending on ϕ, n, ϵ . Let $r = 2$.

2. *Fine grid.* Given ϵ, ϕ, n let \bar{m} satisfy $2^{-\bar{m}} \leq n^{-1/2} \epsilon \phi^{(1-\nu)/4} \leq 2^{1-\bar{m}}$.

3. *Coarse grid.* Let \underline{m} satisfy $2^{-\underline{m}-1}H_r < C_1\phi^{1-\nu} \leq 2^{-\underline{m}}H_r$. For large n then $\bar{m} > \underline{m}$.

4. *Partition support.* For each of $m = \underline{m}, \dots, \bar{m}$ partition axis as laid out in (6.3) with $K_m = 2^m$ points. For each m , points $c_{k_m, m}$ and $c_{k_m^\dagger, m}$ exist so $\underline{c}_m = c_{k_m-1, m} < c_\psi \leq c_{k_m, m} = \bar{c}_m$ and $\underline{c}_m^\dagger = c_{k_m^\dagger-1, m} < c_{\psi^\dagger} \leq c_{k_m^\dagger, m} = \bar{c}_m^\dagger$. It holds that $\bar{c}_{m-1} = c_{k_{m-1}, m-1}$ equals either $\bar{c}_m = c_{k_m, m}$ or $c_{k_m+1, m}$ so that $\bar{c}_{m-1} \geq \bar{c}_m$ and $H(\bar{c}_{m-1}) - H(\bar{c}_m)$ is either zero or $2^{-m}H_r$. There is at most one \underline{m} -grid point in the interval c_ψ, c_{ψ^\dagger} .

5. *Decompose $J_{i,p}(c_\psi, c_{\psi^\dagger})$,* see definition in (6.1). Split the c_ψ, c_{ψ^\dagger} interval into three intervals where the partitioning points are \bar{c}_m and \underline{c}_m^\dagger which are the fine grid points to the right of c_ψ and to the left of c_{ψ^\dagger} , respectively. Note, that if c_ψ, c_{ψ^\dagger} are in the same \bar{m} -interval then $\bar{c}_m > \underline{c}_m^\dagger$ and if they are in neighbouring \bar{m} -interval then $\bar{c}_m = \underline{c}_m^\dagger$. Thus,

$$J_{i,p}(c_\psi, c_{\psi^\dagger}) = J_{i,p}(c_\psi, \bar{c}_m) + J_{i,p}(\underline{c}_m^\dagger, c_{\psi^\dagger}) - 1_{(\bar{c}_m > \underline{c}_m^\dagger)} J_{i,p}(\underline{c}_m, \bar{c}_m) + 1_{(\bar{c}_m < \underline{c}_m^\dagger)} J_{i,p}(\bar{c}_m, \underline{c}_m^\dagger).$$

Consider the fourth term. An iterative argument can be made. Since $\bar{c}_m < \underline{c}_m^\dagger$ then the coarser $(\bar{m} - 1)$ -grid satisfies $\bar{c}_m \leq \bar{c}_{m-1} \leq \underline{c}_{m-1}^\dagger \leq \underline{c}_m^\dagger$, so that

$$J_{i,p}(\bar{c}_m, \underline{c}_m^\dagger) = J_{i,p}(\bar{c}_m, \bar{c}_{m-1}) + J_{i,p}(\bar{c}_{m-1}, \underline{c}_{m-1}^\dagger) + J_{i,p}(\underline{c}_{m-1}^\dagger, \underline{c}_m^\dagger).$$

If $\bar{c}_{m-1} = \underline{c}_{m-1}^\dagger$ then $J_{i,p}(\bar{c}_{m-1}, \underline{c}_{m-1}^\dagger) = 0$ and the iteration stops noting that for $m < \bar{m} - 1$ then m -grid points cross over so $\bar{c}_m \geq \bar{c}_{m-1} = \underline{c}_{m-1}^\dagger \geq \underline{c}_m^\dagger$. If $\bar{c}_{m-1} < \underline{c}_{m-1}^\dagger$ then the argument can be made again for $J_{i,p}(\bar{c}_{m-1}, \underline{c}_{m-1}^\dagger)$. In the m -th step the iteration continues if $\bar{c}_m < \underline{c}_m^\dagger$, so that if there are no other m -grid points between $\bar{c}_m, \underline{c}_m^\dagger$ the contribution from the $(m - 1)$ -step is zero. Since there is at most one \underline{m} -point in the interval c_ψ, c_{ψ^\dagger} , then the \underline{m} -step will either give a zero contribution or the grid points will have crossed over at an earlier stage. Therefore the fourth term satisfies

$$1_{(\bar{c}_m < \underline{c}_m^\dagger)} J_{i,p}(\bar{c}_m, \underline{c}_m^\dagger) = \sum_{m=\underline{m}+1}^{\bar{m}} 1_{(\bar{c}_m < \underline{c}_m^\dagger)} \{J_{i,p}(\bar{c}_m, \bar{c}_{m-1}) + J_{i,p}(\underline{c}_{m-1}^\dagger, \underline{c}_m^\dagger)\}.$$

6. *Decompose $\mathcal{S} = n^{1/2}\{\tilde{\mathbf{F}}(0, 0, c_{\psi^\dagger}) - \tilde{\mathbf{F}}(0, 0, c_\psi)\}$.* Due to the decomposition of $J_{i,p}(c_\psi, c_{\psi^\dagger})$ in item 5 then $|\mathcal{S}| \leq |Z_1| + |Z_2| + |Z_3| + |Z_4| + |Z_5|$, where

$$\begin{aligned} Z_1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(c_\psi, \bar{c}_m) - \mathbf{E}_{i-1}\{J_{i,p}(c_\psi, \bar{c}_m)\}], \\ Z_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(\underline{c}_m^\dagger, c_{\psi^\dagger}) - \mathbf{E}_{i-1}\{J_{i,p}(\underline{c}_m^\dagger, c_{\psi^\dagger})\}], \\ Z_3 &= 1_{(\bar{c}_m > \underline{c}_m^\dagger)} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(\underline{c}_m, \bar{c}_m) - \mathbf{E}_{i-1}\{J_{i,p}(\underline{c}_m, \bar{c}_m)\}], \\ Z_4 &= \sum_{m=\underline{m}+1}^{\bar{m}} 1_{(\bar{c}_m < \underline{c}_m^\dagger)} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(\bar{c}_m, \bar{c}_{m-1}) - \mathbf{E}_{i-1}\{J_{i,p}(\bar{c}_m, \bar{c}_{m-1})\}], \\ Z_5 &= \sum_{m=\underline{m}+1}^{\bar{m}} 1_{(\bar{c}_m < \underline{c}_m^\dagger)} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(\underline{c}_{m-1}^\dagger, \underline{c}_m^\dagger) - \mathbf{E}_{i-1}\{J_{i,p}(\underline{c}_{m-1}^\dagger, \underline{c}_m^\dagger)\}]. \end{aligned}$$

7. *The term Z_1 .* Since $|J_{i,p}(c_\psi, \bar{c}_m)| \leq |J_{i,p}(\underline{c}_m, \bar{c}_m)|$ it holds

$$|Z_1| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n |g_{in}| [|J_{i,p}(\underline{c}_m, \bar{c}_m)| + \mathbf{E}_{i-1}\{|J_{i,p}(\underline{c}_m, \bar{c}_m)|\}].$$

Since $\mathbf{E}_{i-1}\{|J_{i,p}(\underline{c}_{\bar{m}}, \bar{c}_{\bar{m}})|\} \leq H_r(\bar{c}_{\bar{m}}) - H_r(\underline{c}_{\bar{m}}) = 2^{-\bar{m}}H_r$ then Assumption 4.1(*ia, iii*) shows, for some $C_2 > 0$, that

$$\mathbf{E} \sum_{i=1}^n |g_{in}| \mathbf{E}_{i-1}\{|J_{i,p}(\underline{c}_{\bar{m}}, \bar{c}_{\bar{m}})|\} \leq nC_2 2^{-\bar{m}}H_r. \quad (7.5)$$

Noting that $2^{-\bar{m}} \leq n^{-1/2}\epsilon\phi^{(1-\nu)/4}$ and using the Markov inequality then

$$\mathbf{P}(|Z_1| > \epsilon) \leq \frac{2}{\epsilon\sqrt{n}} \mathbf{E} \sum_{i=1}^n |g_{in}| \mathbf{E}_{i-1}\{|J_{i,p}(\underline{c}_{\bar{m}}, \bar{c}_{\bar{m}})|\} \leq 2C_2H_r\phi^{(1-\nu)/4}.$$

8. *The terms Z_2 and Z_3 .* Apply the same argument as in item 7.

9. *The term Z_4 : finding martingale.* Introduce martingales

$$M_{\ell,m,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_{in} [J_{i,p}(c_{\ell,m}, c_{\ell+1,m}) - \mathbf{E}_{i-1}\{J_{i,p}(c_{\ell,m}, c_{\ell+1,m})\}].$$

Recall that for instance $\bar{c}_m = c_{k_m,m}$ while \bar{c}_{m-1} either equals $c_{k_m,m}$ or $c_{k_{m+1},m}$ so that \bar{c}_m, \bar{c}_{m-1} are at most 1 step apart in the m -grid. It then holds that

$$|Z_4| \leq \sum_{m=\underline{m}+1}^{\bar{m}} |M_{k_m,m,n}|.$$

The point $c_{k_m,m}$ satisfies $\underline{c}_m < c_\psi < c_{k_m,m} \leq \bar{c}_m$. Decompose the interval $\underline{c}_m, \bar{c}_m$ of length $2^{-m}H_r$ into $2^{m-\underline{m}}$ intervals of length $2^{-m}H_r$ with left endpoint $c_{k_m+\ell}$ for $0 \leq \ell < 2^{m-\underline{m}}$, say. The interval $c_{k_m,m}, c_{k_{m+1},m}$ is one of those. This gives rise to a further bound

$$|Z_4| \leq \sum_{m=\underline{m}+1}^{\bar{m}} \max_{k_m \leq \ell < k_m + 2^{m-\underline{m}}} |M_{\ell,m,n}|.$$

Note that $\sum_{m=\underline{m}+1}^{\bar{m}} 2^{(m-\underline{m})/4} \leq \sum_{j=1}^{\infty} 2^{-j/4} = (2^{1/4} - 1)^{-1} < 6$. It therefore holds

$$\mathbf{P}(|Z_4| > \epsilon) \leq \mathbf{P} \bigcup_{m=\underline{m}+1}^{\bar{m}} \left\{ \max_{k_m \leq \ell < k_m + 2^{m-\underline{m}}} |M_{\ell,m,n}| > \frac{2^{(m-\underline{m})/4}\epsilon}{6} \right\}$$

Using Boole's inequality then

$$\mathbf{P}(|Z_4| > \epsilon) \leq \sum_{m=\underline{m}+1}^{\bar{m}} \mathbf{P} \left\{ \max_{k_m \leq \ell < k_m + 2^{m-\underline{m}}} |M_{\ell,m,n}| > \frac{2^{(m-\underline{m})/4}\epsilon}{6} \right\}.$$

10. *The term Z_4 : apply Lemma 5.3* with $z_{\ell,i} = g_{in}^* J_{i,p}(c_{\ell-1,m}, c_{\ell,m})$ where g_{in}^* is a coordinated of g_{in} and with $L = 2^{m-\underline{m}}$ while $\kappa = 2^{(m-\underline{m})/4}\epsilon/6$. Noting that for $q = 1, 2$ then $\mathbf{E}_{i-1}|J_{i,p}(c_{\ell-1,m}, c_{\ell,m})|^{2q} \leq 2^{-m}H_r$. Therefore the moment condition holds with $C = 2^{-m}H_rC_2$ since

$$\mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathbf{E}_{i-1}|z_{\ell,i}|^q \leq \mathbf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n |g_{in}|^q 2^{-m}H_r = n2^{-m}H_r(n^{-1}\mathbf{E} \sum_{i=1}^n |g_{in}|^q) \leq n2^{-m}H_rC_2,$$

as in the argument leading to (7.5) under Assumption 4.1(*ia, iii*). The Lemma then shows that for all $\theta_m > 0$ then

$$\mathbf{P}(|Z_4| > \epsilon) \leq \sum_{m=\underline{m}+1}^{\bar{m}} (\mathcal{A}_m \mathcal{B}_m + \mathcal{C}_m),$$

where

$$\mathcal{A}_m = \frac{C\theta_m}{\kappa}2^{-m/2}, \quad \mathcal{B}_m = 2^{-m/2}\{1 + (L+1)\frac{\theta_m^2}{n}\}, \quad \mathcal{C}_m = 4L \exp\left(-\frac{\kappa\theta_m}{14}\right).$$

Choose $\theta_m = 14\kappa^{-1}\{\log(4^{m-\underline{m}}) + \log\phi^{-1}\}$. If $\mathcal{A}_m, \mathcal{B}_m$ are exponentially decreasing in $m - \underline{m}$ and proportional to ϕ^α for some $\alpha > 0$ while \mathcal{C}_m is bounded then $\mathbf{P}(|Z_4| > \epsilon) < C\phi^\alpha$ for some constant $C > 0$.

11. *The term \mathcal{A}_m .* Use that $\kappa = C2^{(m-\underline{m})/4}\epsilon$, the definition of θ_m to get

$$\mathcal{A}_m = C\{\log(4^{m-\underline{m}}) + \log\phi^{-1}\}2^{-(m-\underline{m})/2}\epsilon^{-2}2^{-m/2}.$$

Use the bounds $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $2^{-m/2} < C\phi^{(1-\nu)/2}$ to get

$$\mathcal{A}_m < C\{(m - \underline{m}) \log 4 + \log\phi^{-1}\}2^{-(m-\underline{m})/2}\phi^{(1-\nu)/4}.$$

Since $\phi^{(1-\nu)/8} \log\phi^{-1}$ is bounded and $\phi^{(1-\nu)/8} < 1$ then

$$\mathcal{A}_m < C\{(m - \underline{m}) + 1\}2^{-(m-\underline{m})/2}\phi^{(1-\nu)/8},$$

which is exponentially decreasing in $m - \underline{m}$ and proportional to $\phi^{(1-\nu)/8}$.

12. *The term \mathcal{B}_m .* Use that $2^{-m/2} \leq 1$ and $(L+1) \leq 2L$. Note that $\theta_m^2 = 2^m\kappa^2\mathcal{A}_m^2$ where \mathcal{A}_m is bounded due to item 11. Therefore

$$\mathcal{B}_m \leq 1 + C2^{-m/2}2^{m-\underline{m}}n^{-1}2^m\kappa^2 = 1 + C2^{m/2}2^{m-\underline{m}}n^{-1}\kappa^2.$$

Use that $n^{-1} \leq C2^{-\bar{m}}\epsilon^{-2}\phi^{-(1-\nu)/2}$ and $\kappa^2 = C2^{(m-\underline{m})/2}\epsilon^2$ to get

$$\mathcal{B}_m - 1 \leq C2^{m/2}2^{m-\underline{m}}2^{-\bar{m}}\epsilon^{-2}\phi^{-(1-\nu)/2}2^{-(m-\underline{m})/2}\epsilon^2 = C2^{-(\bar{m}-m)}2^{-\underline{m}/2}\phi^{-(1-\nu)/2} \leq C,$$

since $2^{-(\bar{m}-m)} \leq 2$ and $\phi^{-(1-\nu)/2} < C2^{\underline{m}/2}$.

13. *The term \mathcal{C}_m .* Insert expression for θ_m to get

$$\mathcal{C}_m = C2^{m-\underline{m}} \exp\{-\log(4^{m-\underline{m}}) - \log\phi^{-1}\} = C2^{-(m-\underline{m})}\phi,$$

which is exponentially decreasing in $m - \underline{m}$ and proportional to $\phi \leq \phi^{(1-\nu)/8}$.

14. *The terms Z_5 .* Apply the same argument as for Z_4 .

15. *Combine* the bounds from items 7,8,10,14 to get

$$\mathbf{P}(|\mathcal{S}| > \epsilon) \leq \sum_{j=1}^5 \mathbf{P}(|Z_j| > \epsilon) \leq 3(2C_1H_r\phi^{(1-\nu)/4}) + 2C\phi^{(1-\nu)/8}.$$

uniformly in $\underline{m}, \bar{m}, n$. For a given $\epsilon > 0$ the only constraint to ϕ is that $0 < \phi^{(1-\nu)/4} \leq \epsilon^2$. Thus, the probability vanishes as $\phi \downarrow 0$. ■

8 Proofs of main Theorems 3.1, 3.2 and 3.3

The main results for the forward search are proved. It involves four types of arguments. First, the weighted and marked absolute empirical distribution function $\widehat{\mathbf{G}}_n$ is analysed using the results from Section 4. Secondly, the corresponding quantile processes are analysed. Thirdly, a single step of the Forward Search is analysed using these results. Fourthly, the iteration of the Forward Search is analysed. The analysis is performed using Assumption 3.1 requiring for instance symmetry of the innovation density f . Only the four combinations of g_{in}, p are now considered as outlined in Section 3.1. When checking Assumption 4.1 it suffices to check the conditions for the hybrid case where $g_{in} = nN'x_ix'_iN$ and $p = 2$.

8.1 The absolute empirical distribution

The weighted and marked absolute empirical distribution function was introduced in (3.1) as

$$\widehat{\mathbb{G}}_n^{g,p}(b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \varepsilon_i^p \mathbf{1}_{(|\varepsilon_i - x'_{in} b| \leq \sigma c)}.$$

Throughout this section g, p can take the four combination needed for (3.2), (3.4), (3.5), these are 1, 0 and $n^{1/2}N'x_i$, 1 and $nN'x_i x'_i N$, 0 and 1, 2. In the above expression b represents the regression estimation error $N^{-1}(\widehat{\beta} - \beta_0)$. The process $\widehat{\mathbb{G}}_n$ can be expressed in terms of $\widehat{\mathbb{F}}_n$ quite easily by

$$\widehat{\mathbb{G}}_n^{g,p}(b, c) = \widehat{\mathbb{F}}_n^{g,p}(b, c) - \lim_{c^+ \downarrow c} \widehat{\mathbb{F}}_n^{g,p}(b, -c^+). \quad (8.1)$$

The asymptotic arguments are made on the probability scale $\psi = \mathbb{G}(c_\psi)$. When \mathbf{f} is symmetric then the probability scales of \mathbb{G} and \mathbb{F} are related in a simple linear fashion, see (2.2), so that (8.1) translates into

$$\widehat{\mathbb{G}}_n^{g,p}\{b, \mathbb{G}^{-1}(\psi)\} = \widehat{\mathbb{F}}_n^{g,p}\{b, \mathbb{F}^{-1}\left(\frac{1+\psi}{2}\right)\} - \lim_{\psi^+ \downarrow \psi} \widehat{\mathbb{F}}_n^{g,p}\{b, \mathbb{F}^{-1}\left(\frac{1-\psi^+}{2}\right)\}. \quad (8.2)$$

Therefore, results for $\widehat{\mathbb{F}}_n$ transfer to $\widehat{\mathbb{G}}_n$. The corresponding conditional mean process is

$$\overline{\mathbb{G}}_n^{g,p}(b, c) = \frac{1}{n} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \{\varepsilon_i^p \mathbf{1}_{(|\varepsilon_i - x'_{in} b| \leq \sigma c)}\}, \quad p = 0, 1, 2. \quad (8.3)$$

Form also the empirical process

$$\mathbb{G}_n^{g,p}(b, c) = n^{1/2} \{\widehat{\mathbb{G}}_n^{g,p}(b, c) - \overline{\mathbb{G}}_n^{g,p}(b, c)\}. \quad (8.4)$$

For later use note that $\mathbf{E}_{i-1} \{\varepsilon_i^p \mathbf{1}_{(|\varepsilon_i| \leq \sigma c)}\} = 0$ for odd p since \mathbf{f} is symmetric and $b = 0$. Errors in estimating the quantile are denoted $d = n^{1/2}(c_\psi^b - c_\psi)$. Estimation errors represented by b, d vanish uniformly as shown in the next result. Due to the two-sidedness of the absolute residuals and symmetry of \mathbf{f} , only one of the error terms $x'_{in} b$ and $n^{-1/2}d$ enters the asymptotic expansion depending on the choice of p .

Lemma 8.1 *For each ψ let $c_\psi = \mathbb{G}^{-1}(\psi)$. Suppose Assumption 3.1(ia, iib, iic) holds for some $0 \leq \kappa < \eta \leq 1/4$. Then, for all $B, \epsilon > 0$ and all $\omega < \eta - \kappa \leq 1/4$, it holds*

- (a) $\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4 - \eta} B} |n^{1/2} \{\overline{\mathbb{G}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) - \overline{\mathbb{G}}_n^{g,p}(0, c_\psi)\} - 2\sigma^{p-1} c_\psi^p \mathbf{f}(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} \{1_{(p \text{ odd})} x'_{in} b + 1_{(p \text{ even})} n^{\kappa-1/2} d\}| = \mathcal{O}_{\mathbb{P}}\{n^{2(\kappa-\eta)}\};$
- (b) $\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4 - \eta} B} |\mathbb{G}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{G}_n^{g,p}(0, c_\psi)| = \mathcal{O}_{\mathbb{P}}(1);$
- (b') $\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4 - \eta} B} |\mathbb{G}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{G}_n^{1,0}(0, c_\psi)| = \mathcal{O}_{\mathbb{P}}(n^{-\omega});$
- (c) $\lim_{\phi \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\{\sup_{0 \leq \psi \leq \psi^\dagger \leq 1: \psi^\dagger - \psi \leq \phi} |\mathbb{G}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{G}_n^{g,p}(0, c_\psi)| > \epsilon\} \rightarrow 0.$

Proof of Lemma 8.1. (a) Assumption 3.1(ia, iic) implies Assumption 4.1(ib, iiib) with $r = 0$, $p \leq 2$ and $g_{in} = 1, n^{1/2}x_{in}$ or $n x_{in} x'_{in}$, and hence the assumptions of Theorem 4.3. First, we want to apply this result to $\overline{\mathbb{F}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d)$. Thus, rewrite

$$\overline{\mathbb{F}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) = n^{-1} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \varepsilon_i^p \mathbf{1}_{\{\varepsilon_i - x'_{in} b \leq \sigma(c_\psi + n^{\kappa-1/2}d)\}} = n^{-1} \sum_{i=1}^n g_{in} \mathbf{E}_{i-1} \varepsilon_i^p \mathbf{1}_{\{\varepsilon_i - \bar{x}'_{in} \bar{b} \leq \sigma c_\psi\}},$$

for $\bar{b} = (b', n^\kappa d)'$ and $\bar{x}_{in} = (x'_{in}, n^{-1/2}\sigma)'$, where $|\bar{b}| \leq 2n^{1/4+\kappa-\eta}B$ while \bar{x}_{in} satisfies Assumption 4.1(iii) because $|\bar{x}_{in}|^2 = |x_{in}|^2 + n^{-1}\sigma^2$. Therefore we find, using that $\bar{\mathbb{G}}_n^{g,p}$ can be expressed in terms of $\bar{\mathbb{F}}_n^{g,p}$ as in (8.1), that $n^{1/2}\{\bar{\mathbb{G}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) - \bar{\mathbb{G}}_n^{g,p}(0, c_\psi)\}$ has correction term

$$\begin{aligned} & \sigma^{p-1}c_\psi^p f(c_\psi) n^{-1} \sum_{i=1}^n g_{in} n^{1/2} (x'_{in} b + n^{\kappa-1/2} \sigma d) - \sigma^{p-1} (-c_\psi)^p f(-c_\psi) n^{-1} \sum_{i=1}^n g_{in} n^{1/2} (x'_{in} b - n^{\kappa-1/2} \sigma d) \\ & = \sigma^{p-1} c_\psi^p f(c_\psi) n^{-1/2} \sum_{i=1}^n g_{in} [\{1 - (-1)^p\} x'_{in} b + \{1 + (-1)^p\} n^{\kappa-1/2} \sigma d], \end{aligned}$$

due to the symmetry of f . This reduces as desired.

(b) Let $c_\psi^\dagger = c_\psi + n^{\kappa-1/2}d$. Rewrite $\mathcal{G} = \mathbb{G}_n^{g,p}(b, c_\psi^\dagger) - \mathbb{G}_n^{g,p}(0, c_\psi)$ as $\mathcal{G} = \mathcal{G}_1 + \mathcal{G}_2$ where

$$\mathcal{G}_1 = \mathbb{G}_n^{g,p}(b, c_\psi^\dagger) - \mathbb{G}_n^{g,p}(0, c_\psi^\dagger), \quad \mathcal{G}_2 = \mathbb{G}_n^{g,p}(0, c_\psi^\dagger) - \mathbb{G}_n^{g,p}(0, c_\psi).$$

The term \mathcal{G}_1 is $\text{op}(1)$ uniformly in $|b| \leq n^{1/4-\eta}B$, $0 \leq \psi \leq 1$. To see this, expand $\mathbb{G}_n^{g,p}$ in a similar fashion to (8.1). Apply Theorem 4.1, noting that Assumption 3.1(ia, ib, ic) implies Assumption 4.1(i, ii, iii) with $p \leq 2$, $g_{in} = 1, n^{1/2}x_{in}$ or $n x_{in} x'_{in}$ and the chosen r .

The term \mathcal{G}_2 . Apply Theorem 4.4 noting that Assumption 3.1(ia, ic) implies Assumption 4.1(ia, iii) with $r = 2$ and some $\nu < 1$.

(b') Similar to (b), but using Theorem 4.2.

(c) Assumption 3.1(ia, ic) implies Assumption 4.1(ia, iii) using the Cauchy-Schwarz inequality. Expand $\mathbb{G}_n^{g,p}$ and apply Theorem 4.4. ■

8.2 A first analysis of the order statistics

The Forward Search evolves around order statistics $\hat{z}^{(m)}$ defined in (3.2). A process version gives quantiles

$$\hat{c}_\psi^b = \inf\{c : \widehat{\mathbb{G}}_n^{1,0}(b, c) \geq \psi\}. \quad (8.5)$$

Setting $b = 0$ gives $\hat{c}_\psi^0 = \widehat{\mathbb{G}}_n^{-1}(\psi)$ as defined in (2.8) and studied in Theorem 2.2. The first result gives an algebraic bound to the distance between \hat{c}_ψ^b and \hat{c}_ψ^0 . Probabilistic bounds follow.

Lemma 8.2 *It holds, for all b, ψ , that $\sigma|\hat{c}_\psi^b - \hat{c}_\psi^0| < 2|b| \max_{1 \leq i \leq n} |x_{in}|$.*

Proof of Lemma 8.2. 1. *A property of $\widehat{\mathbb{G}}_n$.* The quantile $\sigma\hat{c}_\psi^0$ is the left-continuous inverse of the right-continuous function $\widehat{\mathbb{G}}_n^{1,0}(0, c) = \widehat{\mathbb{G}}_n(c)$ in (2.8). Thus,

$$\widehat{\mathbb{G}}_n(y) \leq \widehat{\mathbb{G}}_n(\hat{c}_\psi^0) \leq \widehat{\mathbb{G}}_n(z) \quad \Rightarrow \quad y \leq \hat{c}_\psi^0 \leq z. \quad (8.6)$$

2. *A lower bound.* Let $x_{\max} = \max_{1 \leq i \leq n} |x_{in}|$. It holds that

$$\mathcal{S}_i = [-\sigma\hat{c}_\psi^b + x'_{in}b, \sigma\hat{c}_\psi^b + x'_{in}b] \subset [-\sigma\hat{c}_\psi^b - x_{\max}|b|, \sigma\hat{c}_\psi^b + x_{\max}|b|] = \mathcal{S},$$

so that for all $0 \leq \psi \leq 1$ then, with $z = \hat{c}_\psi^b + \sigma^{-1}x_{\max}|b|$ it holds

$$\widehat{\mathbb{G}}_n^{1,0}(b, \hat{c}_\psi^b) \leq \frac{1}{n} \sum_{i=1}^n 1_{(|\varepsilon_i| \leq \sigma z)} = \widehat{\mathbb{G}}_n^{1,0}(0, z) = \widehat{\mathbb{G}}_n(z).$$

Since, $\widehat{\mathbb{G}}_n^{1,0}(b, \hat{c}_\psi^b) = n^{-1} \text{int}(\psi n)$ for all b, ψ then

$$0 = \widehat{\mathbb{G}}_n^{1,0}(b, \hat{c}_\psi^b) - \widehat{\mathbb{G}}_n^{1,0}(0, \hat{c}_\psi^0) \leq \widehat{\mathbb{G}}_n(z) - \widehat{\mathbb{G}}_n(\hat{c}_\psi^0),$$

which implies that $\sigma z = \sigma \hat{c}_\psi^b + x_{\max}|b| \geq \sigma \hat{c}_\psi^0$ by inequality (8.6).

3. An upper bound for $\psi < 1$. It holds, for $y = \hat{c}_\psi^b - \sigma^{-1}2x_{\max}|b|$ that

$$\mathcal{S}_i = [-\sigma \hat{c}_\psi^b + x'_{in}b, \sigma \hat{c}_\psi^b + x'_{in}b] \supset [-\sigma y, \sigma y] = \mathcal{S},$$

noting that the smaller set is empty if $y < 0$. It will therefore hold that

$$\widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^b) \geq \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(|\varepsilon_i| \leq \sigma y)} = \widehat{\mathbf{G}}_n(y).$$

Actually, this inequality must be strict. Indeed, at least one i^\dagger exists so that $\sigma \hat{c}_\psi^b = |\varepsilon_{i^\dagger} - x'_{i^\dagger n}b|$. For this (these) i^\dagger it holds that $\varepsilon_{i^\dagger} \in \mathcal{S}_i$ but $\varepsilon_{i^\dagger} \notin \mathcal{S}$. Thus, it holds $\widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^b) > \widehat{\mathbf{G}}_n(y)$. Proceed as before to see that

$$0 = \widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^b) - \widehat{\mathbf{G}}_n^{1,0}(0, \hat{c}_\psi^0) > \widehat{\mathbf{G}}_n(y) - \widehat{\mathbf{G}}_n(\hat{c}_\psi^0), \quad (8.7)$$

which implies that $y = \hat{c}_\psi^b - \sigma^{-1}2x_{\max}|b| < \hat{c}_\psi^0$ by inequality (8.6). ■

The next result introduces a convergence rate for $\hat{c}_\psi^b - \hat{c}_\psi^0$.

Lemma 8.3 *Suppose Assumptions 3.1(ia, iib, iic) holds. Then, for all $\omega < \eta - \kappa$,*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta}B} n^{1/2} |\mathbf{f}(\hat{c}_\psi^0)(\hat{c}_\psi^b - \hat{c}_\psi^0)| = o_{\mathbf{P}}(n^{-\omega}).$$

Proof of Lemma 8.3. By definition $\widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^b) = \widehat{\mathbf{G}}_n^{1,0}(0, \hat{c}_\psi^0) = n^{-1} \text{int}(n\psi)$. Combine the inequality of Lemma 8.2 with Assumption 3.1 (iib) showing $\max_{1 \leq i \leq n} |x_{in}| = O_{\mathbf{P}}(n^{\kappa-1/2})$ to get that $\hat{c}_\psi^b - \hat{c}_\psi^0 = O_{\mathbf{P}}(n^{-1/4+\kappa-\eta})$ for $|b| \leq n^{1/4-\eta}B$. Thus, for any $\epsilon > 0$ a $C > 0$ exists so that the set $\mathcal{C}_n = \{|n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)| \leq n^{1/4-\eta}C\}$ has probability $\mathbf{P}(\mathcal{C}_n) > 1 - \epsilon$. On this set it holds, with $d = n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)$, that

$$0 = \widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^0 + n^{\kappa-1/2}d) - \widehat{\mathbf{G}}_n^{1,0}(0, \hat{c}_\psi^0).$$

Lemma 8.1(a) using Assumption 3.1(ia, iic) shows that

$$n^{1/2} \{ \overline{\mathbf{G}}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \overline{\mathbf{G}}_n^{1,0}(0, c_\psi) \} - 2\sigma^{-1} \mathbf{f}(c_\psi) n^\kappa d = O_{\mathbf{P}}(n^{2\kappa-2\eta}) = o_{\mathbf{P}}(n^{-\omega}),$$

uniformly in $0 \leq \psi \leq 1$ and $|b|, |d| \leq n^{1/4+\kappa-\eta}B$, for all $\omega < \eta - \kappa < 2(\eta - \kappa)$. Lemma 8.1(b') using Assumption 3.1(ia, iib, iic) shows that, uniformly in $0 \leq \psi \leq 1$ and $|b|, |d| \leq n^{1/4-\eta}B$,

$$\mathbb{G}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{G}_n^{1,0}(0, c_\psi) = o_{\mathbf{P}}(n^{-\omega}),$$

for all $\omega < \eta - \kappa$. Using the definition $\mathbb{G}_n^{1,0} = n^{1/2}(\widehat{\mathbf{G}}_n^{1,0} - \overline{\mathbf{G}}_n^{1,0})$ then

$$0 = n^{1/2} \{ \widehat{\mathbf{G}}_n^{1,0}(b, \hat{c}_\psi^0 + n^{\kappa-1/2}d) - \widehat{\mathbf{G}}_n^{1,0}(0, \hat{c}_\psi^0) \} = 2\sigma^{-1} \mathbf{f}(\hat{c}_\psi^0) n^\kappa d + o_{\mathbf{P}}(n^{-\omega}).$$

Inserting $d = n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)$ we get the desired result. ■

The next result provides a modification of Csörgő (1983, equation 2.8).

Lemma 8.4 *Let $c_\psi = \mathbf{G}^{-1}(\psi)$. Suppose \mathbf{f} is symmetric and decreasing for large c and that Assumption 3.1(ib) holds. Then, for all ψ^* so $|\psi^* - \psi| \leq |\mathbf{G}(\hat{c}_\psi^0) - \psi|$, it holds*

- (a) $\sup_{0 \leq \psi \leq c_n} |1 - \mathbf{f}(c_\psi)/\mathbf{f}(c_{\psi^*})| = o_{\mathbf{P}}(1)$, for any sequence $c_n \rightarrow 0$ so $nc_n \rightarrow \infty$;
- (b) $\sup_{0 \leq \psi \leq n/(n+1)} |1 - \mathbf{f}(c_\psi)/\mathbf{f}(c_{\psi^*})| = O_{\mathbf{P}}(1)$.

Proof of Lemma 8.4. (a) By (2.2) then $G^{-1}(\psi) = F^{-1}(y)$ for $y = (1 + \psi)/2$ varying in $1/2 \leq y \leq 1 - (2n + 2)^{-1}$. Let $\gamma = \sup_{c \in \mathbb{R}} F(c)\{1 - F(c)\}|f'(c)|/\{f(c)\}^2$ which is finite by Assumption 3.1(ib). It is first argued that for all $\epsilon > 0$ and $0 < c < 1$ and all n then

$$\mathbf{P}\left\{\sup_{1/2+c \leq y \leq 1-c} \left| \frac{f\{F^{-1}(y)\}}{f\{\widehat{F}_n^{-1}(y^*)\}} - 1 \right| > \epsilon\right\} \leq 4\{1 + \text{int}(\gamma)\}\{\exp(-nch_1) + \exp(-nch_2)\}, \quad (8.8)$$

where, with $h(\lambda) = \lambda + \log(1/\lambda) - 1$ then $h_1 = h[(1 + \epsilon)^{\{1 + \text{int}(\gamma)\}/2}]$ and $h_2 = h[1/(1 + \epsilon)^{\{1 + \text{int}(\gamma)\}/2}]$. This is nearly the statement of Theorem 1.5.1 of Csörgö (1983), which, however, has the denominator $f(\theta_{y,n})$ instead of $f\{\widehat{F}_n^{-1}(y^*)\}$ where $\theta_{y,n}$ is a particular intermediate point between $\widehat{F}_n^{-1}(y)$ and $F^{-1}(y)$ rather than any intermediate point. Csörgö states that the proof of this Theorem is similar to that of his Theorem 1.4.3. Equation (1.4.18.2) of that proof uses a bound only depending on $\widehat{F}_n^{-1}(y)$ and $F^{-1}(y)$ and not on the particular intermediate point $\theta_{y,n}$. This proves (8.8).

The inequality (8.8) implies that for any sequence $c_n \rightarrow 0$ so $nc_n \rightarrow \infty$ then

$$\mathbf{P}\left\{\sup_{1/2+c_n \leq y \leq 1-c_n} \left| \frac{f\{F^{-1}(y)\}}{f\{\widehat{F}_n^{-1}(y^*)\}} - 1 \right| > \epsilon\right\} \rightarrow 0.$$

The reason is that $h(\lambda) > 0$ for all $\lambda > 0$ so $\lambda \neq 1$. Consider the tails.

Left hand tail. Use that c_n vanishes, that $G(\hat{c}_\psi^0) - \psi = O_P(n^{-1/2})$ by Theorem 2.1, and that f is uniformly continuous in a neighbourhood of zero because f is bounded, positive and continuous.

(b) *Right hand tail.* It suffices to argue that

$$\lim_{\epsilon \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P}\left\{\sup_{1-c_n \leq y \leq 1-(2n+2)^{-1}} \left| \frac{f\{F^{-1}(y)\}}{f\{\widehat{F}_n^{-1}(y^*)\}} - 1 \right| > \epsilon\right\} = 0. \quad (8.9)$$

Apply the inequality (8.8) with $c = (2n + 2)^{-1}$ so that $nc \sim 1/2$. Then use that $h_1, h_2 \rightarrow \infty$ for $\epsilon \rightarrow \infty$ since $h(\lambda) \rightarrow \infty$ for $\lambda \rightarrow \infty$. ■

The next result relates \hat{c}_ψ^0 to c_ψ .

Lemma 8.5 *Suppose Assumptions 3.1(ia, ib) holds with $q = 1$ only. Then*

$$\sup_{0 \leq \psi \leq 1} |(\hat{c}_\psi^0)^k f(\hat{c}_\psi^0) - (c_\psi)^k f(c_\psi)| = o_P(1) \quad \text{for } k = 0, 1.$$

Proof of Lemma 8.5. 1. Consider ψ so $0 \leq \psi \leq 1 - 1/z_n$ for any sequence $0 < z_n < o(n^{1/2})$. Rewrite the process of interest as

$$(\hat{c}_\psi^0)^k f(\hat{c}_\psi^0) - (c_\psi)^k f(c_\psi) = \{(\hat{c}_\psi^0)^k - (c_\psi)^k\} f(c_\psi) + (\hat{c}_\psi^0)^k f(\hat{c}_\psi^0) \left\{1 - \frac{f(c_\psi)}{f(\hat{c}_\psi^0)}\right\}. \quad (8.10)$$

The first term is zero for $k = 0$ and $O_P(n^{-1/2})$ for $k = 1$ due to Lemmas 2.1, 2.2(a) using Assumption 3.1(ib). For the second term, note that $(\hat{c}_\psi^0)^k f(\hat{c}_\psi^0)$ is bounded uniformly in $0 \leq \psi \leq 1$ due to Assumption 3.1(ia) with $q = 1$, while $1 - f(c_\psi)/f(\hat{c}_\psi^0)$ vanishes by Lemma 8.4(a) using Assumption 3.1(ib).

2. Consider ψ so $\psi_n \leq \psi \leq 1$ for any sequence $\psi_n \rightarrow 1$. Assumption 3.1(ia) and the continuity of f implies that $(c_\psi)^k f(c_\psi)$ is continuous and convergent for $\psi \rightarrow 1$. Rewrite

$$\hat{c}_{\psi_n}^0 = G^{-1}\{G(\hat{c}_{\psi_n}^0)\} = G^{-1}[\psi_n + \{G(\hat{c}_{\psi_n}^0) - \psi_n\}] \geq G^{-1}(\psi_n - g_n),$$

where $g_n = \sup_{0 \leq \psi \leq 1} \{G(\hat{c}_\psi^0) - \psi\} = O_P(n^{-1/2})$ due to Lemmas 2.1, 2.2(c) using Assumption 3.1(ib). By the continuity of G^{-1} then $\hat{c}_{\psi_n}^0 \rightarrow G^{-1}(1)$ in probability and therefore $(\hat{c}_\psi^0)^k f(\hat{c}_\psi^0) - (c_\psi)^k f(c_\psi)$ vanishes in probability. ■

8.3 A one-step result for the least squares estimator

A one-step result for the least squares estimator now follows. Equation (3.4) represents the one-step least squares estimator $\hat{\beta}^{(m+1)}$ in terms of $\widehat{\mathbf{G}}_n^{g,p}$. That expression has the random quantities $\hat{b}^{(m)}$ and $\sigma^{-1}\hat{z}^{(m)}$ as arguments. Replacing these by a deterministic quantity b and the residual \hat{c}_ψ^b defined in (8.5) gives the following asymptotic uniform linearization result for the one-step least squares estimator if we insert the initial estimator $b = \hat{b}^{(m)}$.

Lemma 8.6 *Let $c_\psi = \mathbf{G}^{-1}(\psi)$ and*

$$\rho_\psi = 2c_\psi \mathbf{f}(c_\psi) / \psi. \quad (8.11)$$

Suppose Assumption 3.1(ia – ib, ii) hold for some $0 \leq \kappa < \eta \leq 1/4$. Then, for all $\psi_0 > 0$ it holds

$$(a) \sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta}B} |n^{1/2}\widehat{\mathbf{G}}_n^{x,1}(b, \hat{c}_\psi^b) - \mathbb{G}_n^{x,1}(0, c_\psi) - 2c_\psi \mathbf{f}(c_\psi) \Sigma_n b| = o_{\mathbf{P}}(1);$$

$$(b) \sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta}B} |\widehat{\mathbf{G}}_n^{xx,0}(b, \hat{c}_\psi^b) - \Sigma_n \psi| = O_{\mathbf{P}}(n^{\kappa-\eta-1/4});$$

$$(c) \sup_{\psi_0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta}B} |\{\widehat{\mathbf{G}}_n^{xx,0}(b, \hat{c}_\psi^b)\}^{-1} n^{1/2}\widehat{\mathbf{G}}_n^{x,1}(b, \hat{c}_\psi^b) - (\psi \Sigma_n)^{-1} \mathbb{G}_n^{x,1}(0, c_\psi) - \rho_\psi b| = o_{\mathbf{P}}(1).$$

Proof of Lemma 8.6. (a) The inequality of Lemma 8.2 implies that $\hat{c}_\psi^b - \hat{c}_\psi^0 = O_{\mathbf{P}}(n^{\kappa-\eta-1/4})$ uniformly in $0 \leq \psi \leq 1$ and $|b| \leq n^{1/4-\eta}B$ since $\max_{1 \leq i \leq n} |x_{in}| = O_{\mathbf{P}}(n^{\kappa-1/2})$ by Assumption 3.1 (iib). Start by expanding $\widehat{\mathbf{G}}_n^{x,1}$. By definition

$$n^{1/2}\widehat{\mathbf{G}}_n^{x,1}(b, c_\psi + n^{\kappa-1/2}d) = \mathbb{G}_n^{x,1}(b, c_\psi + n^{\kappa-1/2}d) + n^{1/2}\overline{\mathbf{G}}_n^{x,1}(b, c_\psi + n^{\kappa-1/2}d).$$

Lemma 8.1(a, b), using Assumption 3.1(ia, iib, iic) along with the definitions $g_{in} = n^{1/2}x_{in}$ and $\Sigma_n = \sum_{i=1}^n x_{in}x'_{in}$ gives, uniformly in $|b|, |d| \leq n^{1/4-\eta}B$ and $0 \leq \psi \leq 1$,

$$n^{1/2}\widehat{\mathbf{G}}_n^{x,1}(b, c_\psi + n^{\kappa-1/2}d) = \mathbb{G}_n^{x,1}(0, c_\psi) + n^{1/2}\overline{\mathbf{G}}_n^{x,1}(0, c_\psi) + 2c_\psi \mathbf{f}(c_\psi) \Sigma_n b + o_{\mathbf{P}}(1).$$

Note that $\overline{\mathbf{G}}_n^{x,1}(0, c_\psi) = 0$ due to the symmetry of \mathbf{f} . Replace c_ψ by \hat{c}_ψ^0 and d by $n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)$, which is $O_{\mathbf{P}}(n^{1/4-\eta})$. Thus it holds

$$n^{1/2}\widehat{\mathbf{G}}_n^{x,1}(b, \hat{c}_\psi^b) = \mathbb{G}_n^{x,1}(0, \hat{c}_\psi^0) + 2\hat{c}_\psi^0 \mathbf{f}(\hat{c}_\psi^0) \Sigma_n b + o_{\mathbf{P}}(1), \quad (8.12)$$

uniformly in $|b| \leq n^{1/4-\eta}B$ and $0 \leq \psi \leq 1$. The two terms are analysed in turn.

First term. Theorem 2.1 shows $a_\psi = n^{1/2}\{\mathbf{G}(\hat{c}_\psi^0) - \psi\}$ is tight. Expand

$$\hat{c}_\psi^0 = \mathbf{G}^{-1}\{\mathbf{G}(\hat{c}_\psi^0)\} = c_{\mathbf{G}(\hat{c}_\psi^0)} = c_{\psi+n^{-1/2}a_\psi}. \quad (8.13)$$

Lemma 8.1(c) using Assumption 3.1(ia, iib, iic) shows $\mathbb{G}_n^{x,1}(0, \hat{c}_\psi^0) = \mathbb{G}_n^{x,1}(0, c_\psi) + o_{\mathbf{P}}(1)$.

Second term. Use that $\hat{c}_\psi^0 \mathbf{f}(\hat{c}_\psi^0) = c_\psi \mathbf{f}(c_\psi) + o_{\mathbf{P}}(1)$ uniformly in ψ by Lemma 8.5 using Assumptions 3.1(ia, ib).

(b) An expansion as in (8.12) gives

$$\widehat{\mathbf{G}}_n^{xx,0}(b, \hat{c}_\psi^b) = n^{-1/2}\mathbb{G}_n^{xx,0}(0, \hat{c}_\psi^0) + \overline{\mathbf{G}}_n^{xx,0}(0, \hat{c}_\psi^0) + 2\sigma^{-1}\mathbf{f}(\hat{c}_\psi^0)\Sigma_n(\hat{c}_\psi^b - \hat{c}_\psi^0) + o_{\mathbf{P}}(n^{-1/2}),$$

uniformly in b, ψ . The three terms are analysed in turn.

First term. This is $n^{-1/2}\mathbb{G}_n^{xx,0}(0, \hat{c}_\psi^0) = n^{-1/2}\mathbb{G}_n^{xx,0}(0, c_\psi) + o_{\mathbf{P}}(n^{-1/2})$ by an argument as for the first term of (8.12).

Second term. Use the definition of Σ_n and Theorem 2.1, 2.2(c) using Assumption 3.1(ib) showing $\mathbf{G}(\hat{c}_\psi^0) = \psi + \mathcal{O}_{\mathbf{P}}(n^{-1/2})$ uniformly in ψ along with the tightness of Σ_n by Assumption 3.1(ia) to see that

$$\overline{\mathbf{G}}_n^{xx,0}(0, \hat{c}_\psi^0) = \frac{1}{n} \sum_{i=1}^n n x_{in} x'_{in} \mathbf{E}_{i-1} 1_{(|\varepsilon_i| \leq \sigma \hat{c}_\psi^0)} = \sum_{i=1}^n x_{in} x'_{in} \mathbf{G}(\hat{c}_\psi^0) = \Sigma_n \psi + \mathcal{O}_{\mathbf{P}}(n^{-1/2}).$$

Third term. This is $\mathcal{O}_{\mathbf{P}}(n^{\kappa-\eta-1/4})$ since $\mathbf{f}(\hat{c}_\psi^0) = \mathbf{f}(c_\psi) + \mathcal{O}_{\mathbf{P}}(1)$ uniformly in $0 \leq \psi \leq 1$ by Lemma 8.5 using Assumptions 3.1(ia, ib), while $\hat{c}_\psi^b - \hat{c}_\psi^0 = \mathcal{O}(n^{\kappa-\eta-1/4})$ and Σ_n is tight by Assumption 3.1(ia).

(c) Combine (a), (b). The denominator from (b) satisfies

$$\widehat{\mathbf{G}}_n^{xx,0}(b, \hat{c}_\psi^b) = \psi \Sigma_n \{1 + \mathcal{O}_{\mathbf{P}}(1)\},$$

for $\psi \geq \psi_0 > 0$ and since $\Sigma_n \rightarrow \Sigma$ in distribution where $\Sigma > 0$ a.s. by Assumption 3.1(ia). Combine with the expression for the numerator in (a). ■

8.4 The forward plot of least squares estimators

The Forward Plot of least squares estimators is now considered. The one-step result in Lemma 8.6 implies that the Forward Search iteration can be viewed as a fixed point problem. Indeed, the one-step result in Lemma 8.6 implies an autoregressive relation between the one-step updated estimation error $\hat{b}^{(m+1)}$ and the previous estimation error $\hat{b}^{(m)}$. It holds that

$$\hat{b}^{(m+1)} = \rho_\psi \hat{b}^{(m)} + (\psi \Sigma_n)^{-1} \mathbf{G}_n^{x,1}(0, c_\psi) + e_\psi(\hat{b}^{(m)}), \quad (8.14)$$

for $\psi = m/n + o(1)$, an “autoregressive coefficient” ρ_ψ defined in (8.11) and a vanishing remainder term e_ψ . This autoregressive representation generalises Theorem 5.2 of Johansen and Nielsen (2010) which was concerned with a location-scale model, a fixed $\psi \sim m/n$, and convergent initial estimators, $\hat{b}^{(m)} = \mathcal{O}(1)$.

It is first established that ρ_ψ has nice properties for unimodal densities \mathbf{f} .

Lemma 8.7 *Suppose Assumption 3.1(ic) holds. Then $0 < \rho_\psi < 1$ for $0 < \psi < 1$ while $\lim_{\psi \rightarrow 0} \rho_\psi = 1$ and $\lim_{\psi \rightarrow 1} \rho_\psi = 0$.*

Proof of Lemma 8.7. For $c > 0$ then $\mathbf{f}(x)1_{(|x| \leq c)} \geq \mathbf{f}(c)1_{(|x| \leq c)}$ because \mathbf{f} is symmetric and non-increasing by Assumption 3.1(ic). Integration gives

$$\psi = 2 \int_0^{c_\psi} \mathbf{f}(x) dx \geq 2c_\psi \mathbf{f}(c_\psi) = \rho_\psi \psi,$$

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

Lemma 8.8 *Suppose Assumption 3.1(id) holds. Then ρ_ψ is strictly decreasing.*

Proof of Lemma 8.8. Let $\tau_k = 2 \int_0^c x^k f(x) dx$ for $k \in \mathbb{N}_0$. It holds $\lim_{c \rightarrow 0} \tau_k = 0$ and $\tau_k > 0$ for $c > 0$. The derivatives with respect to c are

$$\tau'_k = 2c^k f, \quad \tau''_k = \tau'_{k-1} \left(k + c \frac{\partial}{\partial c} \log f \right).$$

Consider the ratio $R_k = \tau'_{k+1}/\tau_k$, noting that $R_0 = \xi_1^\psi/\psi$. l'Hôpital's rule gives

$$\lim_{c \rightarrow 0} R_k = \lim_{c \rightarrow 0} \frac{\tau''_{k+1}}{\tau'_k} = k + 1.$$

Moreover, R_k has derivative

$$R'_k = \frac{\tau''_{k+1}\tau_k - \tau'_{k+1}\tau'_k}{\tau_k^2} = \frac{\tau'_k}{\tau_k^2} M_k.$$

where $M_k = \{k + 1 + c(\log f)'\}\tau_k - \tau'_{k+1}$. It has to be argued that $R'_k < 0$ for $c > 0$. Since $\tau'_k, \tau_k > 0$ then $R'_k < 0$ if and only if $M_k < 0$. Now, $\lim_{c \rightarrow 0} M_k = 0$ so a sufficient condition is that $M'_k < 0$. But

$$M'_k = \tau_k \{(\log f)' + c(\log f)''\} = \tau_k \{c(\log f)'\}'$$

which is negative if and only if $\Delta(c) = \{c(\log f)'\}' < 0$. ■

The next result investigates the forward estimator $\hat{\beta}^{(m+1)}$. There are two results: first, the forward search preserves the order of the initial estimator, and, secondly, by infinite iteration a slowly converging initial estimator can be improved to consistency at a standard rate. The proof of this result is related to that of Johansen and Nielsen (2011, Theorem 3.3).

Lemma 8.9 *Suppose Assumption 3.1(ia – id, ii, iii) holds. Then, for all $\psi_1 > \psi_0 > 0$ so $m_0/n = \psi_0 + o(1)$, it holds*

- (a) $\sup_{\psi_0 \leq \psi \leq 1} |N^{-1}(\hat{\beta}_\psi - \beta)| = O_{\mathbb{P}}(n^{1/4-\eta})$;
- (b) $\sup_{\psi_1 \leq \psi \leq 1} |N^{-1}(\hat{\beta}_\psi - \beta)| = O_{\mathbb{P}}(1)$.

Proof of Lemma 8.9. Due to the embedding (2.4) it suffices to evaluate $N^{-1}(\hat{\beta}_\psi - \beta)$ at the grid points $\psi = m/n$. Introduce notation $K_\psi^n = \Sigma_n^{-1} \mathbb{G}_n^{x,1}(0, c_\psi)$.

(a) Solve the autoregressive equation (8.14) recursively to get

$$\hat{b}^{(m+1)} = \sum_{k=m_0}^m \left(\prod_{\ell=k}^{m-1} \rho_{\ell/n} \right) \left\{ \frac{n}{k} K_{k/n}^n + e_{k/n}(\hat{b}^{(k)}) \right\} + \left(\prod_{k=m_0}^m \rho_{k/n} \right) \hat{b}^{(m_0)}.$$

with the convention that an empty product equals unity. Lemmas 8.7, 8.8 using Assumption 3.1(ic, id) show that the coefficient ρ_ψ is strictly decreasing and less than unity. For $m \geq \psi_0 n$ then $\rho_{m/n} \leq \rho_{m_0/n} = \rho_0$ for some $\rho_0 < 1$ giving the bound

$$|\hat{b}^{(m+1)}| \leq \left(\sum_{k=m_0}^m \rho_0^{m-k} \right) \left\{ \sup_{\psi_0 \leq \psi \leq 1} |\psi^{-1} K_\psi^n| + \max_{m_0 \leq k < m} |e_{k/n}(\hat{b}^{(k)})| \right\} + \rho_0^{m-m_0+1} |\hat{b}^{(m_0)}|. \quad (8.15)$$

For $\psi \geq \psi_0 > 0$ then $\psi^{-1} K_\psi^n$ is tight by Lemma 8.1(c) using Assumption 3.1(ia, iib, iic). The bound $\sum_{k=m_0}^m \rho_0^{m-k} \leq \sum_{k=0}^\infty \rho_0^k = C$ is finite, while $\hat{b}^{(m_0)} = O(n^{1/4-\eta})$ by Assumption 3.1(iii).

Moreover, $\sup_{\psi_0 \leq \psi \leq 1} \sup_{|b| \leq 3n^{1/4-\eta}B} |e_\psi(b)| = o_{\mathbf{P}}(1)$ for any $B > 0$ by Lemma 8.6 using Assumption 3.1(*ia, ib, ii*). Thus, for all $\epsilon, \zeta > 0$ constants $B, n_0 > 0$ exist so that for $n \geq n_0$, the set

$$\mathcal{A}_n = (C|\hat{b}^{(m_0)}| \leq n^{1/4-\eta}B) \cap (C \sup_{0 \leq \psi \leq 1} |K_\psi^n| \leq B) \cap (C \sup_{\psi_0 \leq \psi \leq 1} \sup_{|b| \leq 3n^{1/4-\eta}B} |e_\psi(b)| \leq \zeta/2)$$

has probability larger than $1 - \epsilon$. An induction over m is now used to prove that

$$\max_{m_0 \leq k \leq m} |\hat{b}^{(k)}| \leq 3n^{1/4-\eta}B \quad \text{for } m = m_0, \dots, n,$$

on the set \mathcal{A}_n , which implies the desired result. As induction start, for $m + 1 = m_0$, then $|\hat{b}^{(m_0)}| \leq n^{1/4-\eta}B$ on the set \mathcal{A}_n . Suppose the result holds for some m . This implies that

$$C \sup_{\psi_0 \leq \psi \leq 1} \max_{m_0 \leq k < m} |e_\psi(\hat{b}^{(k)})| \leq \zeta/2 \quad (8.16)$$

on the set \mathcal{A}_n . Thus, the bound (8.15) becomes $|\hat{b}^{(m)}| \leq 2n^{1/4-\eta}B + \zeta/2 \leq 3n^{1/4-\eta}B$. Thus, the result holds for $m + 1$.

(b) Consider (8.15). Here $\sum_{k=0}^n \rho_0^k$ is finite, $\sup_{\psi_0 \leq \psi \leq 1} |\psi^{-1}K_\psi^n| = O_{\mathbf{P}}(1)$ due to tightness and $\sup_{0 \leq \psi \leq 1} \max_{m_0 \leq k < n} |e_\psi(\hat{b}^{(k)})| = o_{\mathbf{P}}(1)$ due to (8.16). Let $m \geq \psi_1 n$ and $m_0 = \psi_0 n + O(1)$ for some $\psi_1 > \psi_0 > 0$. Since $\rho_0^{m-m_0}$ declines exponentially then $\rho_0^{m-m_0} < n^{-1/4}$ for large n so that $\max_{m \geq m_1} \rho_0^{m-m_0} |\hat{b}^{(m_0)}| = o_{\mathbf{P}}(1)$. ■

8.5 Proof of Theorem 3.1

Lemmas 2.2, 8.2 are now combined to show that the forward residuals scaled with a known variance, $\sigma^{-1}\hat{z}_\psi$, have the same Bahadur representation as the quantile process for the innovations $\sigma^{-1}\varepsilon_i$. This is the main Theorem 3.1 stated with slightly weaker conditions.

Lemma 8.10 *Suppose Assumption 3.1(*ia - id, ii, iii*) holds. Let $\psi_0 > 0$. Then*

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} |2\mathbf{f}(c_\psi)n^{1/2}(\sigma^{-1}\hat{z}_\psi - c_\psi) + \mathbb{G}_n^{1,0}(c_\psi)| = o_{\mathbf{P}}(1).$$

Proof of Lemma 8.10. Due to the embedding (2.4) it suffices to evaluate the forward residuals at the grid points $\psi = m/n$. It is first argued that the forward plot of the estimators is bounded in the sense that for all $\epsilon > 0$ a $B > 0$ exists so that the set $\mathcal{C}_n = (\sup_{\psi_0 \leq \psi \leq 1} |N^{-1}(\hat{\beta}_\psi - \beta)| \leq n^{1/4-\eta}B)$ has $\mathbf{P}(\mathcal{C}_n) \geq 1 - \epsilon$. This follows from Lemma 8.9 using Assumption 3.1(*ia - id, ii, iii*). Now, on \mathcal{C}_n it holds that $\sigma^{-1}\hat{z}_\psi = \hat{c}_\psi^b$, see (3.2), for some $|b| \leq n^{1/4-\eta}B$. Thus it suffices to show that

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} \sup_{|b| \leq n^{1/4-\eta}B} |\mathbb{C}_\psi^b| = o_{\mathbf{P}}(1) \quad \text{for} \quad \mathbb{C}_\psi^b = 2\mathbf{f}(c_\psi)n^{1/2}(\hat{c}_\psi^b - c_\psi) + \mathbb{G}_n^{1,0}(c_\psi).$$

Now, write $(\hat{c}_\psi^b - c_\psi) = (\hat{c}_\psi^0 - c_\psi) + (\hat{c}_\psi^b - \hat{c}_\psi^0)$, so that

$$\mathbb{C}_\psi^b = \{2\mathbf{f}(c_\psi)n^{1/2}(\hat{c}_\psi^0 - c_\psi) + \mathbb{G}_n^{1,0}(c_\psi)\} + 2\frac{\mathbf{f}(c_\psi)}{\mathbf{f}(\hat{c}_\psi^0)}n^{1/2}\mathbf{f}(\hat{c}_\psi^0)(\hat{c}_\psi^b - \hat{c}_\psi^0).$$

The first term is $o_{\mathbf{P}}(n^{\zeta-1/4})$ for all $\zeta > 0$ uniformly in $0 \leq \psi \leq 1$ by Theorem 2.2(a) using Assumption 3.1(*ib*). In the second term the ratio $\mathbf{f}(c_\psi)/\mathbf{f}(\hat{c}_\psi^0)$ is $O_{\mathbf{P}}(1)$ uniformly in $0 \leq \psi \leq n/(n+1)$ by Lemma 8.4 using Assumption 3.1(*ia, ib*), while $n^{1/2}\mathbf{f}(\hat{c}_\psi^0)(\hat{c}_\psi^b - \hat{c}_\psi^0) = o_{\mathbf{P}}(1)$ uniformly in $0 \leq \psi \leq 1$ by Lemma 8.3 using Assumption 3.1(*ia, iib, iic*) ■

8.6 Proofs of Theorems 3.2 and 3.3

The above theory for $\sigma^{-1}\hat{z}_\psi$ involves the population variance σ^2 . The next results gives an asymptotic expansion for $\hat{\sigma}_{\psi,cor}^2$, recalling, from (2.5) that

$$\hat{\sigma}_{\psi,cor}^2 - \sigma^2 = \frac{1}{\tau_\psi} [\widehat{\mathbb{G}}_n^{1,2}(\hat{b}, \hat{c}_\psi^b) - \{\widehat{\mathbb{G}}_n^{x,1}(\hat{b}, \hat{c}_\psi^b)\}' \{\widehat{\mathbb{G}}_n^{xx,0}(\hat{b}, \hat{c}_\psi^b)\}^{-1} \{\widehat{\mathbb{G}}_n^{x,1}(\hat{b}, \hat{c}_\psi^b)\}].$$

Compare also the definitions in (3.6), (3.7) with (8.4) to see

$$\mathbb{G}_n(c_\psi) = \mathbb{G}_n^{1,0}(0, c_\psi), \quad \mathbb{H}_n(c_\psi) = \sigma^{-2} \mathbb{G}_n^{1,2}(0, c_\psi).$$

The main Theorem 3.2 then follows immediately from the next result.

Lemma 8.11 *Suppose Assumption 3.1(ia, ib, ie, ii) holds. Then*

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} \sup_{|b| \leq n^{1/4-\eta}B} |n^{1/2}(\hat{\sigma}_{\psi,cor}^2 - \sigma^2) - \sigma^2 \tau_\psi^{-1} \{\mathbb{H}_n(c_\psi) - c_\psi^2 \mathbb{G}_n(c_\psi)\}| = o_{\mathbb{P}}(1).$$

Proof of Lemma 8.11. 1. *Regression correction term.* Lemma 8.6(a, b) using Assumption 3.1(ia, ib, ii) shows that

$$\begin{aligned} n^{1/2} \widehat{\mathbb{G}}_n^{x,1}(b, \hat{c}_\psi^b) &= \mathbb{G}_n^{1,0}(0, c_\psi) + 2c_\psi \mathbf{f}(c_\psi) \Sigma_n b + o_{\mathbb{P}}(1), \\ \widehat{\mathbb{G}}_n^{xx,0}(b, \hat{c}_\psi^b) &= \Sigma_n \psi + o_{\mathbb{P}}(1), \end{aligned}$$

uniformly in $|b| \leq n^{1/4-\eta}B$, $\psi_0 \leq \psi \leq 1$. Evaluate the terms of the expansion for $n^{1/2} \widehat{\mathbb{G}}_n^{x,1}(b, \hat{c}_\psi^b)$. The first term is $\mathbb{G}_n^{1,0}(0, c_\psi) = o_{\mathbb{P}}(1)$ since $\mathbb{G}_n^{1,0}$ is tight by Lemma 8.1(c) using Assumption 3.1(ia, iib, iic). The second term is $o_{\mathbb{P}}(n^{1/4-\eta})$ since b is of that order. Therefore $\widehat{\mathbb{G}}_n^{x,1}(b, \hat{c}_\psi^b) = o_{\mathbb{P}}(n^{-1/4-\eta})$. Note that $\Sigma_n \rightarrow \Sigma$ in distribution where $\Sigma > 0$ a.s. by Assumption 3.1(iaa). It follows that

$$\{\widehat{\mathbb{G}}_n^{x,1}(\hat{b}, \hat{c}_\psi^b)\}' \{\widehat{\mathbb{G}}_n^{xx,0}(\hat{b}, \hat{c}_\psi^b)\}^{-1} \{\widehat{\mathbb{G}}_n^{x,1}(\hat{b}, \hat{c}_\psi^b)\} = o_{\mathbb{P}}(n^{-1/2-2\eta}).$$

This vanishes, even when scaled by $n^{1/2}$.

2. *The leading term.* It remains to argue that

$$n^{1/2} \{\widehat{\mathbb{G}}_n^{1,2}(b, \hat{c}_\psi^b) - \tau_\psi \sigma^2\} - \sigma^2 \{\sigma^{-2} \mathbb{G}_n^{1,2}(0, c_\psi) - c_\psi^2 \mathbb{G}_n^{1,0}(c_\psi)\} = o_{\mathbb{P}}(1),$$

uniformly in $|b| \leq n^{1/4-\eta}B$, $\psi_0 \leq \psi \leq n/(n+1)$. Lemma 8.2 shows that $\sigma |\hat{c}_\psi^b - \hat{c}_\psi^0| < 2|b| \max_{1 \leq i \leq n} |x_{in}|$ for all b, ψ . Since $|b| \leq n^{1/4-\eta}B$ while $\max_{1 \leq i \leq n} |x_{in}| = o_{\mathbb{P}}(n^{\kappa-1/2})$ by Assumption 3.1(iib) then $\hat{c}_\psi^b - \hat{c}_\psi^0 = o_{\mathbb{P}}(n^{\kappa-\eta-1/4})$ uniformly in $0 \leq \psi \leq 1$ and $|b| \leq n^{1/4-\eta}B$. Thus, we start by expanding $n^{1/2} \widehat{\mathbb{G}}_n^{1,2}(b, \hat{c}_\psi^0 + n^{\kappa-1/2}d)$. By definition

$$n^{1/2} \widehat{\mathbb{G}}_n^{1,2}(b, c_\psi + n^{\kappa-1/2}d) = \mathbb{G}_n^{1,2}(b, c_\psi + n^{\kappa-1/2}d) + n^{1/2} \overline{\mathbb{G}}_n^{1,2}(b, c_\psi + n^{\kappa-1/2}d).$$

Apply Lemma 8.1(a, b) using Assumption 3.1(ia, iib, iic) to get

$$n^{1/2} \widehat{\mathbb{G}}_n^{1,2}(b, c_\psi + n^{\kappa-1/2}d) = \mathbb{G}_n^{1,2}(0, c_\psi) + n^{1/2} \overline{\mathbb{G}}_n^{1,2}(0, c_\psi) + 2\sigma(c_\psi)^2 \mathbf{f}(c_\psi) n^\kappa d + o_{\mathbb{P}}(1),$$

uniformly in $|b|, |d| \leq n^{1/4-\eta}B$, $0 \leq \psi \leq 1$. By definition it holds

$$n^{1/2} \widehat{\mathbb{G}}_n^{1,2}(b, c_\psi + n^{\kappa-1/2}d) = n^{1/2} \widehat{\mathbb{G}}_n^{1,2}(0, c_\psi) + 2\sigma(c_\psi)^2 \mathbf{f}(c_\psi) n^\kappa d + o_{\mathbb{P}}(1).$$

We can replace c_ψ by \hat{c}_ψ^0 . Moreover, since $\hat{c}_\psi^b - \hat{c}_\psi^0 = \mathcal{O}_P(n^{\kappa-\eta-1/4})$ we can replace $n^\kappa d$ by $n^{1/2}(\hat{c}_\psi^b - \hat{c}_\psi^0)$ on a set with large probability. When also subtracting $n^{1/2}\tau_\psi\sigma^2$ on both sides and adding and subtracting $n^{1/2}\tau_{G(\hat{c}_\psi^0)}\sigma^2$ on the right hand side we get

$$\begin{aligned} n^{1/2}\{\widehat{\mathbb{G}}_n^{1,2}(b, \hat{c}_\psi^b) - \tau_\psi\sigma^2\} &= n^{1/2}\{\widehat{\mathbb{G}}_n^{1,2}(0, \hat{c}_\psi^0) - \tau_{G(\hat{c}_\psi^0)}\sigma^2\} \\ &\quad + 2\sigma(\hat{c}_\psi^0)^2\mathbf{f}(\hat{c}_\psi^0)n^{1/2}(\hat{c}_\psi^b - \hat{c}_\psi^0) + \sigma^2n^{1/2}\{\tau_{G(\hat{c}_\psi^0)} - \tau_\psi\} + \mathcal{O}_P(1), \end{aligned} \quad (8.17)$$

uniformly in $|b| \leq n^{1/4-\eta}B$, $0 \leq \psi \leq 1$. The three terms are analysed in turn.

3. *First term of (8.17).* Since $\widehat{\mathbb{G}}_n^{1,2}(0, c) = \sigma^2\tau_{G(c)}$ the first term equals $\mathbb{G}_n^{1,2}(0, \hat{c}_\psi^0)$. Theorem 2.1 shows that $\hat{c}_\psi^0 = c_{\psi+n^{-1/2}\phi}$ where $\phi = n^{1/2}\{\mathbb{G}(\hat{c}_\psi^0) - \psi\}$ is tight. The tightness of $\mathbb{G}_n^{1,2}$ established in Lemma 8.1(c) using Assumption 3.1(*ia, iib, iic*) then implies that the first term equals $\mathbb{G}_n^{1,2}(0, c_\psi) + \mathcal{O}_P(1)$ uniformly in $0 \leq \psi \leq 1$.

4. *The order of \hat{c}_ψ^0 is $\mathcal{O}_P(n^{1/8})$.* The reason is that $\hat{c}_\psi^0 \leq \max_{i \leq n} |\varepsilon_i|$, that $\mathbf{E}|\varepsilon_i|^q < \infty$ for some $q > 8$ by Assumption 3.1(*ia*), and that Boole's and Markov's inequalities imply that $\mathbf{P}(\max_i |\varepsilon_i| > Cn^\nu) \leq \sum_{i=1}^n \mathbf{P}(|\varepsilon_i| > Cn^{1/8}) \leq n(Cn^{1/8})^{-q}\mathbf{E}|\varepsilon_i|^q$ vanishes.

5. *The order of c_ψ^2 is $\mathcal{O}\{(1-\psi)^{-1/4}\}$.* The reason is that $\mathbf{E}|\varepsilon_i|^q < \infty$ for some $q > 8$ by Assumption 3.1(*ia*) and that $1 - \mathbf{F}(c_\psi) = \mathbf{P}(|\varepsilon_i| > \sigma\psi)$ is bounded by $c_\psi^{-q}\mathbf{E}(|\varepsilon_i/\sigma|^q)$ by the Markov inequality. Thus, $c_\psi^2 = \mathcal{O}\{(1-\psi)^{-2/q}\}$. In particular, for $\psi \leq 1 - n^{-1}$ then $c_\psi^2 = \mathcal{O}(n^{2/q}) = \mathcal{O}(n^{1/4})$.

6. *Second term of (8.17).* It holds that $\mathbf{f}(\hat{c}_\psi^0)n^{1/2}(\hat{c}_\psi^b - \hat{c}_\psi^0) = \mathcal{O}_P(n^{-\omega})$ for all $\omega < \eta - \kappa$ uniformly in $0 \leq \psi \leq 1$, $b \leq n^{1/4-\eta}B$ by Lemma 8.3 using Assumption 3.1(*ia, iib, iic*). By item 4 then $(\hat{c}_\psi^0)^2 = \mathcal{O}_P(n^{2\nu})$ for all $\nu > \nu(\eta - \kappa)/2$ for some $\nu < 1$. Thus, the second term vanishes.

7. *Third term of (8.17).* Argue that this equals $-\sigma^2c_\psi^2\mathbb{G}_n^{1,0}(c_\psi) + \mathcal{O}_P(1)$. Recall the definition $\tau_\psi = 2 \int_0^{c_\psi} \varepsilon^2 \mathbf{f}(\varepsilon) d\varepsilon$ and expand

$$\mathcal{S}_3 = n^{1/2}(\tau_{\psi+n^{-1/2}\phi} - \tau_\psi) - c_\psi^2\phi = n^{1/2} \int_{c_\psi}^{c_{\psi+n^{-1/2}\phi}} (\varepsilon^2 - c_\psi^2) 2\mathbf{f}(\varepsilon) d\varepsilon.$$

If \mathcal{S}_3 can be proved to vanish when $\phi = n^{1/2}\{\mathbb{G}(\hat{c}_\psi^0) - \psi\}$ then the desired expression follows, noting that $\phi = -\mathbb{G}_n^{1,0}(c_\psi) + \mathcal{O}_P(1)$ by Theorem 2.2(c). Thus, consider \mathcal{S}_3 . Changing variable $y = \mathbb{G}(\varepsilon)$, $dy = 2\mathbf{f}(\varepsilon)d\varepsilon$, and Taylor expanding gives

$$\mathcal{S}_3 = n^{1/2} \int_\psi^{\psi+n^{-1/2}\phi} (c_y^2 - c_\psi^2) dy = \phi(c_{\psi^*}^2 - c_\psi^2),$$

for some ψ^* so $|\psi^* - \psi| \leq \phi$. Rewrite this, for some $\nu > 0$,

$$\mathcal{S}_3 = \{\psi(1-\psi)\}^{-2\nu} \left\{ \frac{\psi(1-\psi)}{\mathbf{f}(c_\psi)} \right\} (c_{\psi^*} + c_\psi) \left[\frac{\phi}{\{\psi(1-\psi)\}^{1/2-\nu}} \right] \left[\frac{\mathbf{f}(c_\psi)n^{1/2}(c_{\psi^*} - c_\psi)}{\{\psi(1-\psi)\}^{1/2-\nu}} \right] n^{-1/2}.$$

Insert $\phi = n^{1/2}\{\mathbb{G}(\hat{c}_\psi^0) - \psi\}$. Consider the five components of \mathcal{S}_3 individually. The first component is $\mathcal{O}(n^{2\nu})$ for $\psi_0 \leq \psi \leq n/(n+1)$. The second component is $\mathcal{O}(n^{1/8})$ for $\psi_0 \leq \psi \leq 1 - n^{-1}$. The reason is that $\psi(1-\psi)/\mathbf{f}(c_\psi) = \mathcal{O}(c_\psi) = \mathcal{O}(n^{1/8})$ by Assumption 3.1(*ie*) and item 5. The third component is $\mathcal{O}_P(n^{1/4})$ due to items 4,5. The fourth component equals $n^{1/2}\{\mathbb{G}(\hat{c}_\psi^0) - \psi\}/\{\psi(1-\psi)\}^{1/2-\nu}$, which is $\mathcal{O}_P(1)$ uniformly in $1/(n+1) \leq \psi \leq n/(n+1)$, see Theorem 2.3(a). The fifth component is seen to be $\mathcal{O}_P(1)$ by first bounding $|c_{\psi^*} - c_\psi| \leq |c_{\psi+n^{-1/2}\phi} - c_\psi| = |c_{\hat{\psi}} - c_\psi|$ where $\hat{\psi} = \mathbb{G}(\hat{c}_\psi^0)$ and then combining the result for the fourth component with the result that

$\{f(c_\psi)n^{1/2}(c_{\hat{\psi}} - c_\psi) - \phi\}/\{\psi(1 - \psi)\}^{1/2-v}$ is $\text{op}(1)$ uniformly in $1/(n+1) \leq \psi \leq n/(n+1)$, see Theorem 2.3(b) using Assumption 3.1(ib). The sixth component is $n^{-1/2}$. Overall it holds that \mathcal{S}_3 is of order $\text{op}(n^{2v+1/8+1/4+0+0-1/2}) = \text{op}(n^{2v-1/8}) = \text{op}(1)$ uniformly in $\psi_0 < \psi < 1 - n^{-1}$ since v can be chosen sufficiently small. ■

Proof of Theorem 3.3. Note, first the identity

$$\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi,cor}} - c_\psi = \frac{\hat{z}_\psi/\sigma - c_\psi}{\hat{\sigma}_{\psi,cor}/\sigma} - c_\psi \frac{\hat{\sigma}_{\psi,cor}^2 - \sigma^2}{\hat{\sigma}_{\psi,cor}(\hat{\sigma}_{\psi,cor} + \sigma)}.$$

Multiply this by $2f(c_\psi)n^{1/2}$. Use that $n^{1/2}(\hat{\sigma}_{\psi,cor}^2/\sigma^2 - 1)$ and $n^{1/2}(\hat{z}_\psi/\sigma - c_\psi)$ have the leading terms $\tau_\psi^{-1}\{\mathbb{H}_n(c_\psi) - c_\psi^2\mathbb{G}_n(c_\psi)\}$ and $\mathbb{G}_n(c_\psi)$, respectively, due to Theorems 3.1, 3.2. In particular $\hat{\sigma}_{\psi,cor}$ is consistent for σ . ■

9 Discussion

Theorem 3.3 presented an asymptotic theory for the scaled forward residuals $\hat{z}^{(m)}/\hat{\sigma}^{(m)} = \hat{\xi}_{(m+1)}^{(m)}/\hat{\sigma}^{(m)}$. In the earliest work on the Forward Search, such as Atkinson and Riani (2000), the focus was not so much on scaled forward residuals as on the minimum deletion residuals

$$\hat{r}^{(m)} = \min_{i \notin S^{(m)}} \frac{\hat{\xi}_i^{(m)}}{\hat{\sigma}^{(m)}\sqrt{1 + h_i}},$$

with a leverage factor $h_i = x_i'(\sum_{j \in S^{(m)}} x_j x_j')^{-1}x_i$, but with sets $S^{(m)}$ constructed exactly as here. In later work on the Forward Search the forward residuals $\hat{\xi}_{(m+1)}^{(m)}$ as opposed to $\min_{i \notin S^{(m)}} \hat{\xi}_i^{(m)}$ have gained in prominence. Indeed, the t-approximation discussed in Section 2.4 is of this form. Figure 1 of Atkinson, Riani and Cerioli (2010b) indicates that in small samples ($n = 200$) the 50% and 99% pointwise bands of $\hat{r}^{(m)}$ and $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$ are nearly indistinguishable. Here we have given a distribution theory for $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$ in terms of empirical distribution functions. In contrast, $\hat{r}^{(m)}$ will involve conditional empirical distribution functions which are harder to analyse; for a further discussion see Johansen and Nielsen (2010, Comment 2.3).

The presented results discuss the forward residuals when no outliers are present. Leading on from that, it would be of interest to analyse situations with outliers to describe how well the Forward Search captures those. In future work we will consider this issue.

References

- Aroian, L.A. (1941) A study of R.A. Fisher's z distribution and the related F distribution. *Annals of Mathematical Statistics* 12, 429–448.
- Atkinson, A. and Riani, M. (2000) *Robust Diagnostic Regression Analysis*. New York: Springer.
- Atkinson, A.C., Riani, M. and Cerioli, A. (2010a) The forward search: Theory and data analysis (with discussion). *Journal of the Korean Statistical Society* 39, 117–134.
- Atkinson, A.C., Riani, M. and Cerioli, A. (2010b) Rejoinder: The forward search: Theory and data analysis. *Journal of the Korean Statistical Society* 39, 161–163.

- Bahadur, R.R. (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics* 37, 577–580.
- Bercu, B. and Touati, A. (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848–1869.
- Bickel, P.J. (1975) One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428–434.
- Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
- Csörgő, M. (1983) *Quantile Processes with Statistical Applications*. CBMS-NFS Regional Conference Series in Applied Mathematics 42, Society for Industrial and Applied Mathematics.
- Cavaliere, G. and Georgiev, I. (2008) Exploiting infinite variance through dummy variables in an AR model. Discussion paper, Universidade Nova de Lisboa.
- Dollinger, M.B. and Staudte, R.G. (1991) Influence functions of iteratively reweighted least squares estimators. *Journal of the American Statistical Association* 86, 709–716.
- Doornik, J.A. (2007) *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.
- Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometrics Journal* 12, 367–381.
- Guenther, W.C. (1977) An easy method for obtaining percentage points of order statistics. *Technometrics* 19, 319–321.
- Johansen, S. and Nielsen, B. (2009) An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J.L. and Shephard, N. (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 1–36. Oxford: Oxford University Press.
- Johansen, S. and Nielsen, B. (2010) Discussion: The forward search: Theory and data analysis. *Journal of the Korean Statistical Society* 39, 137–145.
- Johansen, S. and Nielsen, B. (2011) Asymptotic theory for iterated one-step Huber-skip estimators. CREATES discussion paper 11-40.
- Kiefer, J. (1967) On Bahadur’s representation of sample quantiles. *Annals of Mathematical Statistics* 38, 1323–1342.
- Koul, H.L. and Ossiander, M. (1994) Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540–582.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revuz, D. and Yor, M. (1998) *Continuous Martingales and Brownian motion*. 3rd edition. Berlin: Springer.
- Riani, M. and Atkinson, A.C. (2007) Fast calibrations of the forward search for testing multiple outliers in regression. *Advances in Data Analysis and Classification* 1, 123–141.

- Riani, M., Atkinson, A.C. and Cerioli, A. (2009) Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society B*, 447–466.
- Rousseeuw, P.J. (1982) Most robust M-estimators in the infinitesimal sense. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 61, 541–551.
- Rousseeuw, P.J. (1984) Least median of squares regression. *Journal of the American Statistical Association* 79, 871–880.
- Rousseeuw, P.J. and Leroy, A.M. (1987) *Robust Regression and Outlier Detection*. New York: Wiley.
- Ruppert, D. and Carroll, R.J. (1980) Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828–838.
- Sampford, M.R. (1953) Some inequalities on Mill’s ratio and related functions. *Annals of Mathematical Statistics* 24, 130–132.
- Shorack, G.R. (1979) Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. *Stochastic Processes and their Applications* 9, 95–98.
- Soms, A.P. (1976) An asymptotic expansion for the tail area of the t-distribution. *Journal of the American Statistical Association* 71, 728–730.
- Víšek, J.Á. (2006a) The least trimmed squares. Part I: Consistency. *Kybernetika* 42, 1–36.
- Víšek, J.Á. (2006b) The least trimmed squares. Part II: \sqrt{n} -consistency. *Kybernetika* 42, 181–202.
- Víšek, J.Á. (2006c) The least trimmed squares. Part III: Asymptotic normality. *Kybernetika* 42, 203–224.