Submitted to Bernoulli
arXiv: arXiv:0000.0000

Analysis of the Forward Search using some new results for martingales and empirical processes

SØREN JOHANSEN[1,*] and BENT NIELSEN[2,**]

[1] *Department of Economics, University of Copenhagen Øster Farimagsgade 5, 1353 Copenhagen K, Denmark, and CREATES Aarhus University. E-mail:* [*] soren.johansen@econ.ku.dk

[2] *Nuffield College & Department of Economics, University of Oxford & Institute for New Economic Thinking at the Oxford Martin School. E-mail:* [**] bent.nielsen@nuffield.ox.ac.uk

1. Introduction

1.1. The Forward Search algorithm

The Forward Search algorithm was suggested for the multivariate location model by Hadi (1992) and for multiple regression by Hadi and Simonoff (1993) and developed further by Atkinson and Riani (2000), see also Atkinson, Riani and Cerioli (2010a) and Atkinson, Riani, and Cerioli (2010b). It is an algorithm for avoiding outliers in a regression analysis by recursively constructing subsets of 'good' observations. The algorithm starts with a robust estimate of the regression parameters based on all observations, and constructs the set of observations with the smallest m_0 absolute residuals. It continues by estimating the parameters by least squares based on the m_0 observations selected. From this estimate, the absolute residuals of all observations are computed and ordered. The $(m_0 + 1)'$st largest absolute residual is the forward residual and it is used to monitor the algorithm. The set of $m_0 + 1$ observations with the smallest absolute residuals is the starting point for the next iteration. The results of the analysis are plots of the recursively estimated forward residuals and robust parameter estimates. This paper provides an asymptotic theory for these forward plots when applied to multiple regression under the assumption of no outliers.

The Forward Search is used as a diagnostic tool in regression analysis. The idea is that most observations are 'good' in the sense that they conform with a regression model with symmetric, if not normal, errors. Some observations may not conform with the model - they are the outliers. When building a statistical model, the user can apply the Forward Search in combination with considerations about the substantive context to decide which observations are 'good' and how to treat the 'outliers' in the analysis. In order to use the algorithm we need to understand its properties when all observations are 'good' with symmetric or even normal errors. Currently this understanding comes from simulations

imsart-bj ver. 2014/10/16 file: Manus*attempt.tex date: November 24, 2014

reported in for instance the above mentioned papers. In the present paper, we analyse the algorithm using asymptotic tools. In the future, we hope to analyse the algorithm in the presence of outliers that may or may not be of a symmetric nature.

1.2. Purpose of paper and results.

In this paper, the forward plots are analysed for a multiple regression model. The model for the 'good' observations has symmetric zero mean errors with unknown scale, while the regressors can be stationary as well as stochastically and deterministically trending. The plots of forward residuals and estimators are embedded as stochastic processes in $D[0,1]$, and their asymptotic properties are derived using new results on empirical processes and martingales. The results can be applied to construct pointwise and simultaneous confidence bands for the forward plots.

The first result is that the process of forward residuals behaves asymptotically as if the parameters were known. That is, as the process of ordered absolute errors from an i.i.d. sample from the error distribution. Such empirical quantile processes are studied by analysing the empirical distribution function as an empirical process. In order to show that the estimation uncertainty is negligible, we introduce a class of weighted and marked empirical processes, where the weights represent functions of the regressors and the marks are functions of the regression error. A technical difficulty is, that because the empirical processes are constructed from estimated residuals, the argument of the empirical process is stochastically varying. We develop the theory of such processes, applying and generalizing the results of Koul and Ossiander (1994).

In the second result, the process of forward residuals is scaled by recursive estimates of the unknown standard error. The limiting process is Gaussian and the variance function is found.

In the study of weighted and marked empirical processes, the well known method of replacing the discontinuous processes by their smooth compensators is applied. The difference is a martingale. To justify this replacement, some new iterated exponential martingale inequalities for the variation of the maximum of finitely many martingales are developed by an iterative application of an exponential inequality of Bercu and Touati (2008).

1.3. History and background

The forward search starts with a robust estimator. Examples of robust regression estimators are the least median squares estimator and 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). We will allow initial estimators $\hat{\beta}^{(m_0)}$ converging at a rate slower than the usual $n^{1/2}$-rate, for the stationary case, as for example the least median squares estimator, which is $n^{1/3}$-consistent in location-scale models.

Broadly speaking, we require three asymptotic tools. First, a theory for weighted and marked empirical processes to describe the least squares statistics. Secondly, an analysis of the corresponding quantile processes to describe the forward residuals. Thirdly, a fixed point result to describe the iteration involved.

In 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 stochastic variation in the quantiles and those of Koul and Ossiander (1994) who did not allow marks. The proof combines a chaining argument with iterations of 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 absolute regression errors. 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. A single step of the algorithm has been discussed for the location-scale case by Johansen and Nielsen (2010). Starting with Bickel (1975), see also Simpson, Ruppert, and Carroll (1992), there are a number of asymptotic results for one-step L- and M-estimators. These are predominantly concerned with objective functions that have continuous derivatives, thereby excluding hard rejection as for the one-step Huber-skip function. The forward search gives a sequence of one-step estimators. Because the estimators are based on least squares in a sample selected by truncating the residuals, each estimator is a one-step Huber-skip estimator. Such estimators have been studied by Ruppert and Carroll (1980), Johansen and Nielsen (2009), Johansen and Nielsen (2013, Theorem 3.3), Ronchetti and Welsh (2002), and Hawkins and Olive (2002).

There appears to be less work on iteration of one-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 (2013) analysed a sequence of Huber-skip estimators for a first order autoregression with infinite variance errors, while Johansen and Nielsen (2013, Theorem 3.3) in their Theorem 3.3 analysed sequences of one-step Huber-skip estimators with a fixed critical value. Here we need a critical value which changes with m, the chosen number of observations, so we need a generalisation of the fixed point result of the latter paper.

Outline of the paper: The model and the Forward Search algorithm are defined in §2. The main asymptotic results are given in §3. The weighted and marked empirical process results are given in §4, while the iterated exponential martingale inequalities are presented in §5 with proofs following in Appendices A-C. The proofs of the main results follow in Appendix D. Finally, Appendix E gives a result on order statistics of t-distributed variables.

2. Model and Forward Search algorithm

The multiple regression model is presented, and the Forward Search algorithm is defined including the forward residual and forward deletion residual.

2.1. Model

We assume that (y_i, x_i), $i = 1, \ldots, n$ satisfy the multiple regression equation with regressors of dimension $\dim x$

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

The errors, ε_i, are assumed independent and identically distributed with mean zero and variance σ^2, and ε_i/σ has known density f and distribution function $\mathsf{F}(c) = \mathsf{P}(\varepsilon_i \leq \sigma c)$. In practice, the distribution F will often be standard normal.

The forward search is an algorithm based on ordering absolute residuals and calculation of least squares estimators from the selected observations. Both these choices implicitly assume a symmetric density. Because, unless symmetry is assumed, truncating the errors symmetrically gives in general an error distribution with mean different from zero and hence biased least squares estimators, at least for the location parameter.

The distribution function of the absolute errors $|\varepsilon_i|/\sigma$ of a symmetric density is $\mathsf{G}(c) = \mathsf{P}(|\varepsilon_1| \leq \sigma c) = 2\mathsf{F}(c) - 1$ with density $\mathsf{g}(c) = 2\mathsf{f}(c)$. We define the quantiles of the absolute errors as

$$c_\psi = \mathsf{G}^{-1}(\psi) = \mathsf{F}^{-1}\{(1+\psi)/2\}, \qquad \psi \in [0, 1[, \tag{2.2}$$

and the truncated moments

$$\tau_\psi = \int_{-c_\psi}^{c_\psi} u^2 \mathsf{f}(u) du \text{ and } \varkappa_\psi = \int_{-c_\psi}^{c_\psi} u^4 \mathsf{f}(u) du. \tag{2.3}$$

Then the conditional variance of ε_1/σ given $\{|\varepsilon_1| \leq \sigma c\}$ is

$$\varsigma_\psi^2 = \tau_\psi/\psi. \tag{2.4}$$

This will serve as a bias correction for the variance estimator based on the truncated sample. Using l'Hôpital's rule it is seen that

$$\varsigma_0^2 = 0, \qquad \frac{c_0^2}{\varsigma_0^2} = 3. \tag{2.5}$$

If $\mathsf{f} = \varphi$ is Gaussian, then $\varsigma_\psi^2 = 1 - 2c_\psi\varphi(c_\psi)/\psi$.

2.2. Forward Search algorithm

The Forward Search algorithm is designed to avoid outliers in a linear multiple regression. The first step is given by the choice of a robust estimator, $\hat{\beta}^{(m_0)}$, of the regression parameter, and the choice of the size m_0 of the initial set of 'good' observations. The algorithm generates a sequence of sets of 'good' observations and least squares regression estimators based on these. The $(m+1)'$st step of the algorithm is given as follows.

Algorithm 2.1. *(Forward Search)*

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

$$\hat{\beta}^{(m+1)} = (\textstyle\sum_{i \in S^{(m+1)}} x_i x_i')^{-1} (\textstyle\sum_{i \in S^{(m+1)}} x_i y_i), \tag{2.6}$$

$$(\hat{\sigma}^{(m+1)})^2 = \frac{1}{m+1} \textstyle\sum_{i \in S^{(m+1)}} (y_i - x_i'\hat{\beta}^{(m+1)})^2. \tag{2.7}$$

Note, that $\hat{\beta}^{(n)}$ and $(\hat{\sigma}^{(n)})^2$ are the full sample least squares estimators, and that for $n \to \infty, m/n \to \psi$, see Theorem 3.2,

$$(\hat{\sigma}^{(n)})^2 \overset{\mathsf{P}}{\to} \sigma^2 \tau_\psi / \psi.$$

We therefore introduce also the (asymptotically) bias corrected variance estimator using $\varsigma_{m/n}^2 = \tau_{m/n}/(m/n)$, see (2.4), so that

$$(\hat{\sigma}_{corr}^{(m)})^2 = \frac{(\hat{\sigma}^{(m)})^2}{\varsigma_{m/n}^2} \overset{\mathsf{P}}{\to} \sigma^2. \tag{2.8}$$

Applying the algorithm for $m = m_0, \ldots, n-1$, results in sequences of order statistics $\hat{z}^{(m)} = \hat{\xi}_{(m+1)}^{(m)}$, least squares estimators $(\hat{\beta}^{(m)}, (\hat{\sigma}^{(m)})^2)$, along with the scaled forward residuals

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

Atkinson and Riani (2000) propose to use the minimum deletion residual

$$\hat{d}^{(m)} = \min_{i \notin S^{(m)}} \hat{\xi}_i^{(m)},$$

instead of the forward residuals. Thus, the deletion residual is based on the smallest residual with respect to $\hat{\beta}^{(m)}$ among those observations that were not included in $S^{(m)}$ which in turn is based on $\hat{\beta}^{(m-1)}$, and the forward residual is the largest absolute residual in $S^{(m+1)}$ which is based on $\hat{\beta}^{(m)}$.

The plots of $\hat{\beta}^{(m)}$, $\hat{z}^{(m)}/\hat{\sigma}^{(m)}$, and $\hat{d}^{(m)}/\hat{\sigma}^{(m)}$ against m are called forward plots, see Atkinson and Riani (2000, p. 12–13). The primary objective of this paper is to derive the asymptotic distribution of these plots.

When the method was proposed by Hadi and Simonoff (1993), they also suggested scaling the residual by a leverage factor and replace the scaled residuals $\hat{\xi}_i^{(m)}/\hat{\sigma}^{(m)}$ above by

$$\frac{\hat{\xi}_i^{(m)}}{\hat{\sigma}^{(m)}\sqrt{1-h_i^{(m)}}} \text{ for } i \in S^{(m)}, \qquad \frac{\hat{\xi}_i^{(m)}}{\hat{\sigma}^{(m)}\sqrt{1+h_i^{(m)}}} \text{ for } i \notin S^{(m)},$$

where $h_i^{(m)} = x_i'(\sum_{j\in S^{(m)}} x_j x_j')^{-1} x_i$ is the leverage factor. Johansen and Nielsen (2009) prove that such a leverage factor does not change the asymptotic distribution for the one-step Huber-skip estimator, and the methods presented there can be used to prove a similar result for the forward search. Another small sample correction would be to replace $m+1$ with $m+1-dimx$ in 2.7, but we are mainly concerned with asymtotic properties in this paper.

3. The main results

Johansen and Nielsen (2010, Theorems 5.1–5.3) analysed a single step of the Forward Search applied in a location-scale setting. Those 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 the estimation error for the variance. The subsequent analysis shows how this result generalises to a fully iterated Forward Search. This section first gives the assumptions, then the results, and finally presents some simulations. The derivatives of f are denoted $\dot{\mathsf{f}}$ and $\ddot{\mathsf{f}}$ and for more complicated expressions by d/dx.

3.1. 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. The leading case is the normal density, φ, but the results are also discussed for the t-density.

Assumption 3.1. *Let $\mathcal{F}_i$ be an increasing sequence of σ fields such that ε_{i-1} and x_i are $\mathcal{F}_{i-1}$-measurable and ε_i is independent of $\mathcal{F}_{i-1}$ with symmetric, continuously differentiable density f which is positive on the support $\mathsf{F}^{-1}(0) < c < \mathsf{F}^{-1}(1)$ which contains 0. For some $0 \le \kappa < \eta \le 1/4$ choose an $r \ge 2$ so that $2^{r-1} \ge 1 + (1/4 + \kappa - \eta)(1 + \dim x)$. Let $q_0 = 1 + \max\{2^{r+1}, 2/(\eta - \kappa)\}$. Suppose*
(i) density satisfies
- *(a) tail monotonicity: $c^q\mathsf{f}(c)$, $|c^{q-1}\dot{\mathsf{f}}(c)|$ are decreasing for large c and some $q > q_0$;*
- *(b) quantile process condition: $\gamma = \sup_{c>0} \mathsf{F}(c)\{1-\mathsf{F}(c)\}|\dot{\mathsf{f}}(c)|/\{\mathsf{f}(c)\}^2 < \infty$;*
- *(c) unimodality: $\dot{\mathsf{f}}(c) \le 0$ for $c > 0$ and $\lim_{c\to 0}\ddot{\mathsf{f}}(c) < 0$;*
- *(d) tail condition: $\{1-\mathsf{F}(c)\}/\{c\mathsf{f}(c)\} = \mathrm{O}(1)$ for $c \to \infty$;*

(ii) regressors x_i are $\mathcal{F}_{i-1}$-measurable and a non-stochastic normalisation matrix N exists so that

(a) $\Sigma_n = N' \sum_{i=1}^n x_i x_i' N \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0$;

(b) $\max_{1\le i\le n} |n^{1/2-\kappa} N' x_i| = \mathrm{O}_{\mathsf{P}}(1)$;

(c) $n^{-1}\mathsf{E}\sum_{i=1}^n |n^{1/2} N' x_i|^{q_0} = \mathrm{O}(1)$;

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

Remark 3.1. ***The constant*** q_0 *involves the term $\eta - \kappa$ in two ways. Here κ is needed to control $N'x_i$. If the regressors are bounded, we can choose $\kappa = 0$. This is also the case if the regressors are deterministic or of random walk type, see Example 3.2 below. If $\kappa = 0$ and the initial estimator is convergent at the standard rate, $\eta = 1/4$, then q_0 reduces to $q_0 = 9$ and moments of order 8+ are sufficient. Depending on the trade-off between κ, η and $\dim x$, moments of order 8+ may suffice.*

Remark 3.2. ***Assumption 3.1****(i) is satisfied for the normal distribution, see Example 3.1 below. 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 the 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, $1 + 2^{r+1}$, is larger than 8 in order to control the estimation error for the variance. Condition (ia) is more severe than normally seen in empirical process theory due to the marks ε_i^p. Condition (ib) is used in Theorem D.2, which builds on Csörgő (1983). Condition (ic) is needed to ensure that the iterative element of the Forward Search is a contraction. The unimodality could be relaxed by assuming the conclusion of Lemma D.12. Condition (id) for Mill's ratio is milder than the condition employed for kernel density estimation by Csörgő (1983, p.139).*

Remark 3.3. ***Assumption 3.1****(ii). Condition (iia) is standard in regression analysis and allows for stationary, random walk, and deterministically trending regressors. Some specific examples are given in Example 3.2 below.*

As 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.

Example 3.1. ***Assumption 3.1(i) for the reference distribution*** f.
(a) ***Standard normal distribution***, $\mathsf{f} = \varphi$. *Condition (i) is satisfied: (ia) holds since $c^q\varphi(c) = -c^{q-1}\dot{\varphi}(c)$ is decreasing for large c for any q. (ib) holds with $\gamma = 1$, noting $\dot{\varphi}(c) = -c\varphi(c)$ 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 since $\{1-\Phi(c)\}/\{c\varphi(c)\} < 1/c^2 \to 0$ as $c \to \infty$.*
(b) ***Scaled distribution.*** *Consider a density $\mathsf{f}_\delta(c)$ that has variance δ^2 but otherwise satisfies condition (i). Then $\mathsf{f}(c) = \delta\mathsf{f}_\delta(c\delta)$ has unit variance, distribution function $\mathsf{F}(c) =$*

$\mathsf{F}_\delta(c\delta)$ and satisfies condition (i) with the same γ in part (b).
(c) ***Scaled t-distribution.*** *The t-distribution with $d > 2^{r+1}$ degrees of freedom has density $\mathsf{f}_d(c) = C_d(1 + c^2/d)^{-(d+1)/2}$ with $C_d = \Gamma\{(d + 1)/2\}/\{(d\pi)^{1/2}\Gamma(d/2)\}$ and variance $\delta_d^2 = d/(d-2)$. The reference density can be chosen as $\mathsf{f}(c) = \mathsf{f}_d(c\delta_d)\delta_d$. Due to part (b), it suffices to check condition (i) for f_d. It holds that $\dot{\mathsf{f}}_d(c) = -\gamma h(c)\mathsf{f}_d(c)$ where $\gamma = (d+1)/d$ and $h(c) = c/(1+c^2/d)$ so that $\frac{d}{dc}\log \mathsf{f}_d(c) = -\gamma h(c)$. Condition (ia): for some constants C, it holds that $c^q\mathsf{f}_d(c) \sim Cc^{q-d-1}$ and $c^{q-1}|\dot{\mathsf{f}}_d(c)| \sim Cc^{q-d-3}$ since $h(c) \sim c^{-1}$. Thus $c^q\mathsf{f}_d(c)$ and $c^{q-1}|\dot{\mathsf{f}}_d(c)|$ are both declining for large c, for q chosen so that $d + 1 > q > q_0$. (ib) holds with the stated γ since $1 - c^{-2}d/(d + 2) < h(c)\{1 - \mathsf{F}_d(c)\}/\mathsf{f}_d(c) < 1$, see Soms (1976, equation 3.2). (ic) is well-known to hold. (id) holds since $\{1 - \mathsf{F}_d(c)\}/\{c\mathsf{f}_d(c)\} < 1/\{ch(c)\} \to 1/d$ as $c \to \infty$.*

Example 3.2. ***Assumption 3.1(ii) for the regressors x_i.***
(a) ***Stationary and autoregressive regressors.*** *In this case x_i and ε_i have moments of the same order and $N = n^{-1/2}I_{\dim x}$. (iic) holds if $\mathsf{E}|x_i|^{q_0} < \infty$. (iib) holds due to the Boole and Markov inequalities if $\eta > \kappa > 1/q_0$.*
(b) ***Deterministic regressors*** *such as $x_i = (1, i)'$. Let $N = \mathrm{diag}(n^{-1/2}, n^{-3/2})$. 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 $n^{-1/2}x_{\mathrm{int}(n\psi)}$ converges to a Brownian motion by Donsker's invariance principle, see Billingsley (1999). 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.*

Example 3.3. ***Assumption 3.1(iii) for the initial estimator.***
The focus of this paper is the situation with no outliers. Thus, a wide range of $n^{1/2}$-consistent standard estimators or even $n^{1/3}$-consistent median based estimators can be used. Therefore, Assumption 3.1(iii) only becomes binding when analysing cases with outliers.

3.2. The results

The forward plot of for instance $\hat{z}^{(m)}$ is a process on $m = m_0, \dots, n-1$. 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

$$\hat{z}_\psi = \begin{cases} \hat{z}^{(m)} & \text{for } m = \mathrm{int}(n\psi) \text{ and } m_0/n \le \psi \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

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

The main results are described in terms of three processes

$$\mathbb{G}_n(c_\psi) = n^{-1/2}\textstyle\sum_{i=1}^n \{1_{(|\varepsilon_i/\sigma|\leq c_\psi)} - \psi\}, \tag{3.2}$$

$$\mathbb{L}_n(c_\psi) = \tau_\psi^{-1} n^{-1/2}\textstyle\sum_{i=1}^n [\{(\varepsilon_i/\sigma)^2 - c_\psi^2\}1_{(|\varepsilon_i/\sigma|\leq c_\psi)} - (\tau_\psi - c_\psi^2\psi)], \tag{3.3}$$

$$\mathbb{K}_n(c_\psi) = \textstyle\sum_{i=1}^n N' x_i \varepsilon_i 1_{(|\varepsilon_i/\sigma|\leq c_\psi)}, \tag{3.4}$$

The first two are asymptotically Gaussian processes and the same holds for the third if the regressors are stationary, see Theorem 3.6.

The main results give asymptotic representations of the forward residuals $\hat{z}_\psi/\sigma$ scaled with known scale, of the bias corrected variance, $\hat{\sigma}^2_{\psi,corr}$, and of the forward residuals $\hat{z}_\psi/\hat{\sigma}_{\psi,corr}$ scaled with the bias corrected variance estimator. Next, it is shown that the forward residuals, $\hat{z}_\psi$, and the deletion residuals, $\hat{d}_\psi$, have the same asymptotic representation after an initial burn-in period. Finally, an asymptotic representation is given for the forward plot of regression estimators, $\hat{\beta}_\psi$. Proofs of these results are given in Appendix D.

Theorem 3.1. *Suppose Assumption 3.1 holds. Let $\psi_0 > 0$ and $\omega < \eta - \kappa \leq 1/4$. Then*

$$\sup_{\psi_0\leq\psi\leq n/(n+1)} |2\mathsf{f}(c_\psi)n^{1/2}(\sigma^{-1}\hat{z}_\psi - c_\psi) + \mathbb{G}_n(c_\psi)| = \mathrm{o}_\mathsf{P}(n^{-\omega}). \tag{3.5}$$

Moreover, if $\hat{c}_{m/n}$ are the order statistics of $\xi_i/\sigma = |\varepsilon_i|/\sigma$, then

$$\sup_{\psi_0\leq\psi\leq n/(n+1)} |\mathsf{f}(c_\psi)n^{1/2}(\sigma^{-1}\hat{z}_\psi - \hat{c}_\psi)| = \mathrm{o}_\mathsf{P}(n^{-\omega}). \tag{3.6}$$

If β and σ were known, the residuals are the innovations, ε_i, and the ordering of the absolute residuals $\xi_i = |y_i - \beta' x_i| = |\varepsilon_i|$ can be done once, so that $\sigma^{-1}\hat{z}_m = \sigma^{-1}\xi_{(m+1)} = \hat{c}_{(m+1)/n}$, and the left hand side of (3.6) is trivially zero. In this situation, (3.5) reduces to the Bahadur (1966) representation for the order statistics of the errors ξ_i, see also Theorem D.2 in the Appendix. Theorem 3.1 therefore has the interpretation that the forward residuals $\hat{z}_m = \hat{\xi}^{(m)}_{(m+1)}$ behave asymptotically as the order statistics of the absolute innovations $\xi_i = |\varepsilon_i|$.

Theorem 3.2. *Let $\psi_0 > 0$. Under Assumption 3.1, the asymptotically biased corrected variance estimator has the representation*

$$\sup_{\psi_0\leq\psi\leq n/(n+1)} |n^{1/2}(\sigma^{-2}\hat{\sigma}^2_{\psi,corr} - 1) - \mathbb{L}_n(c_\psi)| = \mathrm{o}_\mathsf{P}(1).$$

Remark 3.4. *In Theorems 3.1 and 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 observations.*

Remark 3.5. *The least squares estimator for the variance is* $\hat{\sigma}^2_{1,corr} = \hat{\sigma}^2_1$*, noting that* $\tau_1 = 1$ *and* $\varsigma_1 = 1$*. Least squares theory shows that* $n^{1/2}(\hat{\sigma}^2_1/\sigma^2 - 1) = n^{-1/2}\sum_{i=1}^n(\varepsilon_i^2/\sigma^2 - 1) + \mathrm{o_P}(1)$*. To see that Theorem 3.2 matches this result, note that the leading term of the least squares approximation is* $\lim_{\psi\to 1}\tau_\psi^{-1}n^{-1/2}\sum_{i=1}^n\{(\varepsilon_i/\sigma)^2 1_{(|\varepsilon_i/\sigma|\le c_\psi)} - \tau_\psi\}$*. It is therefore necessary that the other term in* $\mathbb{L}_n(\psi)$ *satisfies*

$$\lim_{\psi\to 1}\tau_\psi^{-1}c_\psi^2 n^{-1/2}\sum_{i=1}^n\{1_{(|\varepsilon_i/\sigma|\le c_\psi)} - \psi\} = \lim_{\psi\to 1}c_\psi^2\mathbb{G}_n(\psi) = \mathrm{o_P}(1).$$

Because ε_i *has more than 8 moments,* $c_\psi^2 = \mathrm{o}\{(1-\psi)^{-1/4}\}$*, see also item 4 of the proof of Lemma D.11. Combine this with Theorems D.3(a), D.4 to see that* $\lim_{\psi\to 1}c_\psi^2\mathbb{G}_n(c_\psi) = \mathrm{o_P}(1)$*.*

Combining Theorem 3.1 and 3.2 gives an asymptotic representation of the forward residuals with a bias corrected scale.

Theorem 3.3. *Let* $c_\psi = \mathsf{G}^{-1}(\psi)$ *and* $\psi_0 > 0$*. Under Assumption 3.1, the bias corrected scaled forward residual has the expansion*

$$\sup_{\psi_0\le\psi\le n/(n+1)}|2\mathsf{f}(c_\psi)n^{1/2}(\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi,corr}} - c_\psi) + \mathbb{G}_n(c_\psi) + c_\psi\mathsf{f}(c_\psi)\mathbb{L}_n(c_\psi)| = \mathrm{o_P}(1).$$

The above results generalise those of Johansen and Nielsen (2010, Theorems 5.1, 5.3) 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, compare for instance Engler and Nielsen (2009, Theorem 2.1) and Lee and Wei (1999, Theorem 3.2).

In finite samples the forward residuals and the deletion residuals can be different, see for instance Johansen and Nielsen (2010, §2.2). The next result implies that $\widehat{d}^{(m)}$ and $\widehat{z}^{(m)}$ have the same asymptotic distribution.

Theorem 3.4. *It follows from the definitions that* $\widehat{d}^{(m)} \le \widehat{z}^{(m)}$*. Let* $m_0 = \mathrm{int}(n\psi_0)$ *where* $\psi_0 > 0$*, and let Assumption 3.1 hold. Then for all* ψ_1 *such that* $\psi_0 < \psi_1 < 1$

$$\sup_{\psi_1\le\psi\le n/(n+1)}|\mathsf{f}(c_\psi)n^{1/2}(\widehat{z}^{(m)} - \widehat{d}^{(m)})| = \mathrm{o_P}(1).$$

The last result is for the forward plot of the estimator error $N^{-1}(\widehat{\beta}^{(m)} - \beta)$, which can be analysed in two stages. First, it is established that $N^{-1}(\widehat{\beta}^{(m)} - \beta)$ satisfies a recursion of the form

$$N^{-1}(\widehat{\beta}^{(m+1)} - \beta) = \rho_{m/n}N^{-1}(\widehat{\beta}^{(m)} - \beta) + (\psi\Sigma_n)^{-1}\mathbb{K}_n(c_\psi) + e_{m/n}\{N^{-1}(\widehat{\beta}^{(m)} - \beta)\}, \quad (3.7)$$

where $\rho_\psi = 2c_\psi\mathsf{f}(c_\psi)/\psi$ is an 'autoregressive coefficient' and e_ψ is a vanishing remainder term. This result generalises the result for the location model in Johansen and Nielsen

(2010, Theorem 5.2). The unimodality required in Assumption 3.1 (*ic*) implies that ρ_ψ is bounded away from unity for $\psi \geq \psi_0$. The recursion (3.7) can then be iterated by generalising the argument in Johansen and Nielsen (2013, Theorem 3.3) for the iterated one-step Huber-skip estimator for a fixed ψ. The following result arises.

Theorem 3.5. *Suppose Assumption 3.1 holds. Let $m_0 = \text{int}(n\psi_0)$ where $\psi_0 > 0$. Then, for all ψ_1, $\psi_0 < \psi_1 < 1$, the forward plot of the estimator has the expansion*

$$\sup_{\psi_1 \leq \psi \leq 1} |N^{-1}(\widehat{\beta}_\psi - \beta) - \frac{1}{\psi - 2c_\psi \mathsf{f}(c_\psi)} \Sigma_n^{-1} \mathbb{K}_n(c_\psi)| = \mathrm{o}_{\mathsf{P}}(1).$$

3.3. Applications of the result for the forward residuals

The statements of Theorems 3.1, 3.3, 3.4 for the forward residuals and Theorem 3.2 do not depend on the type of regressor. Thus, to apply these theorems it suffices to analyse the asymptotically Gaussian processes $\mathbb{G}_n$ and $\mathbb{L}_n$ for the chosen reference distribution.

Theorem 3.6. *Suppose Assumption 4.1 holds. Then $\mathbb{G}_n$ and $\mathbb{L}_n$ converge on $D[0,1]$ to zero mean Gaussian processes, $\mathbb{G}$, $\mathbb{L}$. Their variances are given by*

$$\mathsf{Var}\{\mathbb{G}(c_\psi)\} = \psi(1-\psi), \tag{3.8}$$

$$\mathsf{Var}\{\mathbb{L}(c_\psi)\} = \frac{1}{\tau_\psi^2}\{\varkappa_\psi - \tau_\psi^2 + c_\psi^2(1-\psi)(c_\psi^2\psi - 2\tau_\psi)\}, \tag{3.9}$$

$$\mathsf{Cov}\{\mathbb{G}(c_\psi), \mathbb{L}(c_\psi)\} = \frac{1}{\tau_\psi}(\tau_\psi - c_\psi^2\psi)(1-\psi) < 0, \tag{3.10}$$

where the truncated moments τ_ψ and $\varkappa_\psi$ are given in (2.3).

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

$$n^{1/2}(\frac{\hat{z}_\psi}{\hat{\sigma}_\psi} - \frac{c_\psi}{\varsigma_\psi}) = n^{1/2}\frac{1}{\varsigma_\psi}(\frac{\hat{z}_\psi - c_\psi \hat{\sigma}_{\psi,corr}}{\hat{\sigma}_{\psi,corr}}) = n^{1/2}\frac{1}{\varsigma_\psi}(\frac{\hat{z}_\psi}{\hat{\sigma}_{\psi,corr}} - c_\psi) \overset{\mathsf{D}}{\to} \mathsf{N}(0, \omega_\psi), \tag{3.11}$$

where ω_ψ has contributions from $\hat{z}_\psi$, from $\hat{\sigma}_{\psi,corr}$, and from their covariance so that

$$\omega_\psi = \frac{1}{\{2\mathsf{f}(c_\psi)\}^2}\left[\mathsf{Var}\{\mathbb{G}(c_\psi)\} + 2c_\psi \mathsf{f}(c_\psi)\mathsf{Cov}\{\mathbb{G}(c_\psi), \mathbb{L}(c_\psi)\} + c_\psi^2 \mathsf{f}^2(c_\psi)\mathsf{Var}\{\mathbb{L}(c_\psi)\}\right].$$

The above results shed light on some previously suggested distributional approximations for the deletion residuals. The approximation of Atkinson and Riani (2006, Theorem 2) has an asymptotic variance that matches that of the process $\mathbb{G}$, while omitting the estimation error for the scale. Riani and Atkinson (2007) presented an approximation to the distribution of the deletion residuals that comes from order statistics of certain t-distributed variables. Due to Theorem E.1 in Appendix E, that approximation also has an asymptotic variance matching that of the process $\mathbb{G}_n$.

Figure 1. compares the asymptotic distribution of $\hat{z}_\psi/\hat{\sigma}_\psi$ for a normal reference distribution (thick line) with (*a*) $\hat{z}_\psi/\hat{\sigma}_{\psi,corr}$ using the corrected scale estimator, (*b*) $\hat{z}_\psi/(\sigma\varsigma_\psi)$ using the known scale, and (*c*) $\hat{z}_\psi/\hat{\sigma}_\psi$ for a $\mathsf{t}(5)$ reference distribution. The solid lines indicate the mean, the dashed lines indicate the 5% and 95% asymptotic quantiles for $n = 128$.

Example 3.4. ***Some particular reference distributions.***
(*a*) ***Standard normal distribution***. *If* $\mathsf{f} = \varphi$ *then* $c_\psi = \Phi^{-1}\{(1+\psi)/2\}$ *and*

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

(*b*) ***Scaled t-distribution*** *with d degrees of freedom of Example 3.1(c) has density* $\mathsf{f}(c) = \delta_d\mathsf{f}_d(c\delta_d)$ *where* f_d *is the t-density with d degrees of freedom and variance* $\delta_d^2 = d/(d-2)$. *Then* $c_\psi = \delta_d^{-1}\mathsf{F}_d^{-1}\{(1+\psi)/2\}$ *and* $\psi = 2\mathsf{F}_d(c_\psi\delta_d) - 1$, *and*

$$\tau_\psi = (d-1)\left\{2\mathsf{F}_{d-2}(c_\psi) - 1\right\} - (d-2)\left\{2\mathsf{F}_d(c_\psi\delta_d) - 1\right\}, \quad (3.12)$$
$$\varkappa_\psi = (d-2)^2\left[\frac{(d-1)(d-3)}{(d-2)(d-4)}\left\{2\mathsf{F}_{d-4}\left(\frac{c_\psi}{\delta_{d-2}}\right) - 1\right\}\right.$$
$$\left. -2\frac{d-1}{d-2}\left\{2\mathsf{F}_{d-2}(c_\psi) - 1\right\} + \left\{2\mathsf{F}_d(c_\psi\delta_d) - 1\right\}\right]. \quad (3.13)$$

Note that for $c_\psi \to \infty$, *the distribution functions approach unity so that*

$$\tau_\psi \to 1, \qquad \varkappa_\psi \to 3\frac{d-2}{d-4}, \quad (3.14)$$

which are the variance and the kurtosis of the scaled t-distribution.

Figure 1 compares the asymptotic distribution of $\hat{z}_\psi/\hat{\sigma}_\psi$ for a normal reference distribution with (*a*) $\hat{z}_\psi/\hat{\sigma}_{\psi,corr}$ using the corrected scale estimator, (*b*) $\hat{z}_\psi/(\sigma\varsigma_\psi)$ using the known scale, and (*c*) $\hat{z}_\psi/\hat{\sigma}_\psi$ for a $\mathsf{t}(5)$ reference distribution. The solid lines are the point-wise means, while the dashed lines are asymptotic 5% and 95% quantiles computed for $n = 128$. This value of n is chosen for comparability with the data example in Riani and Atkinson (2007, Figure 1). It is seen that the asymptotic mean c_ψ/ς_ψ for $\hat{z}_\psi/\hat{\sigma}_\psi$ approaches $\sqrt{3}$ for $\psi \to 0$, see (2.5). Further, the 5% and 95% quantiles for $\hat{z}_\psi/\hat{\sigma}_\psi$ and $\hat{z}_\psi/(\sigma\varsigma_\psi)$ diverge for $\psi \to 0$, which is a consequence of the division by ς_ψ since $\varsigma_0 = 0$, see (2.5). The quantiles also diverge for $\psi \to 1$ which is an extreme value effect.

In panel a, the forward residuals $\hat{z}_\psi/\hat{\sigma}_\psi$ are compared to the bias-corrected forward residuals $\hat{z}_\psi/\hat{\sigma}_{\psi,cor}$. These representations are equivalent, but the former may be preferable from a visual viewpoint.

Panel b compares situations with estimated and known variance. It is seen that estimating the variance contributes to reducing the uncertainty. This phenomenon is also seen for empirical processes of estimated residuals, see Engler and Nielsen (2009, equation 2.10).

Finally, panel c compares the result for $\mathsf{f} = \phi$ with the results for $\mathsf{f} = \mathsf{t}(5)$. 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. A striking feature of this panel is the excellent agreement between the curves when ψ is not too large. For larger ψ the long tails of the t-distribution have an increasing effect.[1]

3.4. Application of the result for the forward estimators

In an application of Theorem 3.5 for the forward estimators, the distribution of the kernel $\Sigma_n^{-1}\mathbb{K}_n(c_\psi)$ depends on the type of regressors. Building on the analysis in Johansen and Nielsen (2009, section 1.4,1.5), we present a result for the stationary case. For situations with deterministic trends or unit roots, see those papers. In the case of stationary and autoregressive regressors, we take $N = n^{-1/2}$ and the normalised matrix of squared regressors, $\Sigma_n = n^{-1}\sum_{i=1}^n x_i x_i'$, described in Assumption 3.1(iia), has a deterministic limit.

Theorem 3.7. *Suppose Assumption 4.1 holds and that x_i is stationary and autoregressive with finite variance. Then $\Sigma_n \overset{\mathsf{P}}{\to} \Sigma > 0$ and $\mathbb{K}_n$ converges on $D[0,1]$ to a zero mean Gaussian process $\mathbb{K}$ with variance given as*

$$\mathsf{Var}\{\mathbb{K}(c_\psi)\} = \tau_\psi \sigma^2 \Sigma. \tag{3.15}$$

Theorem 3.7 implies that

$$n^{1/2}(\widehat{\beta}_\psi - \beta) \overset{\mathsf{D}}{\to} \mathsf{N}[0, \frac{\tau_\psi \sigma^2}{\{\psi - 2c_\psi \mathsf{f}(c_\psi)\}^2}\Sigma^{-1}],$$

which generalises Johansen and Nielsen (2010, Corollary 5.3 2.2). The limiting distribution matches that of the least trimmed squares estimator with trimming ψ, see Víšek (2006c, Theorem 1).

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 Koul 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

[1]Graphics were done using R 2.13, see R Development Core Team (2011)

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

$$\widehat{\mathsf{F}}_n^{g,p}(b,c) = \frac{1}{n}\sum_{i=1}^n g_{in}\varepsilon_i^p 1_{(\varepsilon_i \le \sigma c + x_{in}' b)}, \tag{4.1}$$

with normalised regressors $x_{in} = n^{1/2}N'x_i$, weights g_{in} which are measurable with respect to $(\varepsilon_{i-1},\ldots,\varepsilon_1,x_i,\ldots,x_1)$, and marks ε_i^p. By proving results that hold uniformly in b, we can handle the Forward Search. This allows an analysis of the order statistics of the residuals at a given step m of the Forward Search, since the order statistics depend on the previous estimation error $\hat{b} = N^{-1}(\widehat{\beta}^{(m)} - \beta)$, but are scale invariant. In turn, we can apply the results for the estimation errors $N^{-1}(\hat{\beta}^{(m+1)} - \beta)$ and $n^{1/2}(\hat{\sigma}_{corr}^{(m+1)} - \sigma)$.

4.1. Assumptions

We will keep track of the assumptions in a more explicit way than 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 that $\varepsilon_{i-1}, x_i, g_{in}$ are $\mathcal{F}_{i-1}$-measurable and ε_i is independent of $\mathcal{F}_{i-1}$ with continuously differentiable density f which is positive on the support $\mathsf{F}^{-1}(0) < c < \mathsf{F}^{-1}(1)$ which contains 0. Let p, r, η, κ, ν be given such that $p, r \in \mathbb{N}_0$, $0 \le \kappa < \eta \le 1/4$ and $\nu \le 1$. Suppose*
(i) density satisfies:
- *(a) moments: $\int_{-\infty}^{\infty} |\varepsilon|^{2^r p/\nu}\mathsf{f}(u)du < \infty$;*
- *(b) boundedness: $\{(1+|c|^{\max(0,2^r p-1)})\mathsf{f}(c) + (1+|c|^{2^r p})|\dot{\mathsf{f}}(c)|\} < \infty$;*
- *(c) smoothness: a $C_{\mathsf{H}} \in \mathbb{N}$ exist such that for all $a > 0$*

$$\frac{\sup_{c\ge a}(1+c^{2^r p})\mathsf{f}(c)}{\inf_{0\le c\le a}(1+c^{2^r p})\mathsf{f}(c)} \le C_{\mathsf{H}}, \qquad \frac{\sup_{c\le -a}(1+|c|^{2^r p})\mathsf{f}(c)}{\inf_{-a\le u\le 0}(1+|c|^{2^r p})\mathsf{f}(c)} \le C_{\mathsf{H}}.$$

(ii) regressors x_i satisfy $\max_{1\le i\le n}|n^{1/2-\kappa}N'x_i| = \mathsf{O}_{\mathsf{P}}(1)$ for some non-stochastic normalisation matrix N;
(iii) weights g_{in} are matrix valued and satisfy
- *(a) $n^{-1}\mathsf{E}\sum_{i=1}^n |g_{in}|^{2^r}(1+|n^{1/2}N'x_i|) = \mathsf{O}(1)$;*
- *(b) $n^{-1}\sum_{i=1}^n |g_{in}|(1+|n^{1/2}N'x_i|^2) = \mathsf{O}_{\mathsf{P}}(1)$.*

Remark 4.1. *Discussion of Assumption 4.1.*
*(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 assumptions simplify considerably.*
*(b) **The moment 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* $\mathsf{h}_r(c) = (1 + \epsilon^{2^r p})\mathsf{f}(\epsilon)$ *is monotone for* $|c| > d_1$ *for some* $d_1 \geq 0$. *Indeed, choose* $d_2 \geq d_1$ *so that* $\sup_{c \geq d_2} \mathsf{h}_r(c) = \inf_{0 \leq c \leq d_2} \mathsf{h}_r(c) = \mathsf{h}_r(d_2)$. *Then choose*

$$C_{\mathsf{H}} > \sup_{0 \leq c \leq d_2} \mathsf{h}_r(c) / \inf_{0 \leq c \leq d_2} \mathsf{h}_r(c).$$

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

4.2. The empirical process results

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

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

and the weighted and marked empirical process

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

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

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

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

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. Note that when $p = 0$, then r will be irrelevant in Assumption 4.1(*i*), see also Remark 4.1(*a*).

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

$$\sup_{0 \leq \psi \leq 1} \sup_{|b|, |d| \leq n^{1/4+\kappa-\eta}B} |\mathbb{F}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{F}_n^{1,0}(0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(n^{-1/8-\omega/2}).$$

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

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

$$\sup_{0\le\psi\le1}\sup_{|b|\le n^{1/4-\eta}B}|n^{1/2}\{\overline{\mathsf{F}}_n^{g,p}(b,c_\psi)-\overline{\mathsf{F}}_n^{g,p}(0,c_\psi)\}-\sigma^{p-1}c_\psi^p\mathsf{f}(c_\psi)n^{-1}\textstyle\sum_{i=1}^n g_{in}n^{1/2}x_{in}'b|$$

is $\mathsf{O}_\mathsf{P}(n^{-2\eta})$.

Finally, we argue that the weighted and marked empirical process $\mathbb{F}_n^{g,p}(0,c_\psi)$ in (4.3) is tight when viewed as a sequence in n of processes on $D[0,1]$. Following Billingsley (1999, Theorem 15.5) we need to check two conditions. First, it holds by construction that $\mathbb{F}_n^{g,p}(0,c_0)=0$. Second, the next results shows that the modulus of continuity is small.

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

$$\lim_{\phi\downarrow0}\limsup_{n\to\infty}\mathsf{P}\left\{\sup_{0\le\psi\le\psi^\dagger\le1:\psi^\dagger-\psi\le\phi}|\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger})-\mathbb{F}_n^{g,p}(0,c_\psi)|>\epsilon\right\}\to0.$$

The proofs of these results are given in Appendix C.

5. Iterated exponential martingale inequalities

Chaining arguments will be used to handle tightness properties of the empirical processes. This reduces the tightness proof to a problem of finding the tail probability for the maximum of a certain family of martingales. We first give a general result on a bound of a finite number of martingales, which we prove by iterating a 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.

Theorem 5.1. *For* ℓ, $1\le\ell\le L$, *let* $z_{\ell,i}$ *be* $\mathcal{F}_i$*-adapted and* $\mathsf{E}z_{\ell,i}^{2^{\overline{r}}}<\infty$ *for some* $\overline{r}\in\mathbb{N}$*. Let* $D_r=\max_{1\le\ell\le L}\sum_{i=1}^n\mathsf{E}_{i-1}z_{\ell,i}^{2^r}$ *for* $1\le r\le\overline{r}$*. Then, for all* $\kappa_0,\kappa_1,\ldots,\kappa_{\overline{r}}>0$,

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

The proof is given in Appendix A.

Theorem 5.2. *For* ℓ, $1\le\ell\le L$, *let* $z_{\ell,i}$ *be* $\mathcal{F}_i$*-adapted and* $\mathsf{E}z_{\ell,i}^{2^{\overline{r}}}<\infty$ *for some* $\overline{r}\in\mathbb{N}$*. Let* $D_r=\max_{1\le\ell\le L}\sum_{i=1}^n\mathsf{E}_{i-1}z_{\ell,i}^{2^r}$ *for* $1\le r\le\overline{r}$*. Suppose, for some* $\varsigma\ge0$, $\lambda>0$, *that*

$L = \mathrm{O}(n^{\lambda})$ and $\mathsf{E}D_r = \mathrm{O}(n^{\varsigma})$ for $r \leq \overline{r}$. Then, if $\upsilon > 0$ is chosen such that
(i) $\varsigma < 2\upsilon$,
(ii) $\varsigma + \lambda < \upsilon 2^{\overline{r}}$,
it holds that for all $\kappa > 0$ and $n \to \infty$,

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

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

$$\mathcal{P}_n = \mathrm{O}\left\{n^{\lambda}\frac{n^{\varsigma}(\log n)^{2^{\overline{r}}-1}}{n^{\upsilon 2^{\overline{r}}}} + \textstyle\sum_{r=1}^{\overline{r}}\frac{n^{\varsigma}(\log n)^{2^{r}-1}}{n^{\upsilon 2^{r}}} + 2n^{\lambda}\overline{r}n^{-2\lambda}\right\} = \mathrm{o}(1),$$

as desired since $\varsigma + \lambda < \upsilon 2^{\overline{r}}$ and $\varsigma < 2\upsilon \leq \upsilon 2^{r}$ for $r \geq 1$. □

Theorem 5.3. *For ℓ, $1 \leq \ell \leq L$, let $z_{\ell,i}$ be $\mathcal{F}_i$-adapted and $\mathsf{E}z_{\ell,i}^4 < \infty$. Suppose $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{\ell,i}^{2^q} \leq Dn$ for $q = 1, 2$ and some $D > 0$. Then, for all $\theta, \kappa > 0$,*

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

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

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

Exploit the moment conditions to get the desired result. □

6. Conclusion

The intention of the Forward Search is to determine the number of outliers by looking at the forward plot of the forward residuals. The main results for the Forward Search, given in §3, describe the asymptotic distribution of that process in a situation where there are no outliers. We can therefore add pointwise confidence bands to the forward plot, using Theorem 3.3. These give an impression of the pointwise variation we would expect for the forward plot, if there were in fact no outliers. In practice we would want to make a simultaneous decision based on the entire graph. A theory is developed in Johansen and Nielsen (2014b) and implemented in the R-package ForwardSearch, see Nielsen (2014).

We suspect that the iterated martingale inequalities will be useful in a variety of situations. For instance, in ongoing research, we are finding that the inequalities are helpful in establishing consistency and asymptotic distribution results for general M-estimators, see Johansen and Nielsen (2014a).

The results and techniques in this paper could potentially also be used to shed light on other iterative 1-step methods in robust statistics such as those discussed in Bickel (1975), Simpson, Ruppert, and Carroll (1992), and Hawkins and Olive (2002). Another example would be to establish an asymptotic theory for the Forward Search applied to multivariate location and scatter, see Cerioli, Farcomini, and Riani (2014) for a discussion of consistency. Finally we mention Bellini (2012) for an application of the Forward Search to the cointegrated vectorautoregressive model.

Appendix A: Proofs of martingale inequalities

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

$$\mathcal{P}_r(\kappa_r) = \mathsf{P}(\max_{1\leq\ell\leq L} A_{\ell,r} > \kappa_r), \qquad \mathcal{Q}_r(\kappa_r) = \mathsf{P}(\max_{1\leq\ell\leq L} |A_{\ell,r}| > \kappa_r),$$

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 $0 \leq r < \overline{r}$.* We first prove that, for any $\kappa_r, \kappa_{r+1} > 0$,

$$\mathcal{Q}_r(\kappa_r) \leq 2L\exp(-\frac{\kappa_r^2}{14\kappa_{r+1}}) + \mathcal{P}_{r+1}(\kappa_{r+1}) + \frac{\mathsf{E}D_{r+1}}{\kappa_{r+1}}. \tag{A.1}$$

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

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

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 $B_{\ell,r,i} = (z_{\ell,i}^{2^r} - \mathsf{E}_{i-1}z_{\ell,i}^{2^r})^2 + \mathsf{E}_{i-1}(z_{\ell,i}^{2^r} - \mathsf{E}_{i-1}z_{\ell,i}^{2^r})^2$. We then get

$$\mathcal{Q}_r(\kappa_r) \leq \mathsf{P}\{(\max_{1\leq\ell\leq L}|A_{\ell,r}| > \kappa_r)\cap(\max_{1\leq\ell\leq L} B_{\ell,r} \leq 7\kappa_{r+1})\} + \mathsf{P}(\max_{1\leq\ell\leq L} B_{\ell,r} > 7\kappa_{r+1}). \tag{A.2}$$

Consider the first term in (A.2), $\mathcal{S}_{1,r}$ say. By Boole's inequality this satisfies

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

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

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

Because $A_{\ell,r}$ is a martingale, the exponential inequality of Bercu and Touati (2008, Theorem 2.1) shows

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

Taken L times, this gives the first term in (A.1).

Consider the second term in (A.2), $\mathcal{S}_{2,r}$ say. Ignore the indices on $B_{\ell,r,i}$, E_{i-1} and $z_{\ell,i}^{2^r}$, and apply the inequality $(z-\mathsf{E}z)^2 \leq 2(z^2+\mathsf{E}^2 z)$ along with $\mathsf{E}^2 z \leq \mathsf{E}z^2$ and $\mathsf{E}(z-\mathsf{E}z)^2 \leq \mathsf{E}z^2$ to get that $B = (z-\mathsf{E}z)^2 + \mathsf{E}(z-\mathsf{E}z)^2 \leq 2z^2 + 3\mathsf{E}z^2 = 2(z^2 - \mathsf{E}z^2) + 5\mathsf{E}z^2$. Thus,

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

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

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

which are the last two terms of (A.1).

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

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

Apply iterated expectations and interchange maximum and expectation to get

$$\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}}) \leq \frac{L}{\kappa_{\bar{r}}} \max_{1\leq\ell\leq L} \mathsf{E} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell,i}^{2^{\bar{r}}} \leq \frac{L}{\kappa_{\bar{r}}} \mathsf{E} \max_{1\leq\ell\leq L} \textstyle\sum_{i=1}^n \mathsf{E}_{i-1} z_{\ell,i}^{2^{\bar{r}}} = \frac{L}{\kappa_{\bar{r}}} \mathsf{E} D_{\bar{r}}. \tag{A.3}$$

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

$$\mathcal{Q}_r(\kappa_r) \leq 2L\exp(-\frac{\kappa_r^2}{14\kappa_{r+1}}) + \mathcal{Q}_{r+1}(\kappa_{r+1}) + \frac{\mathsf{E}D_{r+1}}{\kappa_{r+1}} \qquad \text{for } r = 0, \ldots, \bar{r}-2, \tag{A.4}$$

$$\mathcal{Q}_r(\kappa_r) \leq 2L\exp(-\frac{\kappa_r^2}{14\kappa_{r+1}}) + \mathcal{P}_{r+1}(\kappa_{r+1}) + \frac{\mathsf{E}D_{r+1}}{\kappa_{r+1}} \qquad \text{for } r = \bar{r}-1. \tag{A.5}$$

Then sum from $r = 0$ to $\bar{r} - 2$ to get

$$\mathcal{Q}_0(\kappa_0) = \mathcal{Q}_{\bar{r}-1}(\kappa_{\bar{r}-1}) + \textstyle\sum_{r=0}^{\bar{r}-2} \{\mathcal{Q}_r(\kappa_r) - \mathcal{Q}_{r+1}(\kappa_{r+1})\}$$

and insert the bounds (A.4), (A.5) and $\mathcal{P}_{\bar{r}}(\kappa_{\bar{r}}) \leq \kappa_{\bar{r}}^{-1} L \mathsf{E} D_{\bar{r}}$ from (A.3). □

Appendix B: 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. Throughout the rest of Appendix we denote by C a constant which need not be the same in different expressions.

Introduce the function

$$J_{i,p}(x,y) = (\varepsilon_i/\sigma)^p\{1_{(\varepsilon_i \le \sigma y)} - 1_{(\varepsilon_i \le \sigma x)}\}, \tag{B.1}$$

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

$$\mathsf{H}_r(x) = \int_{-\infty}^{x} (1 + u^{2^r p})\mathsf{f}(u)du,$$

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

$$H_r = \mathsf{H}_r(\infty) = \int_{-\infty}^{\infty} (1 + u^{2^r p})\mathsf{f}(u)du < \infty.$$

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

$$0 \le |\mathsf{E}\{J_{i,p}(x,y)\}^{2^s}| \le \mathsf{E}\{|J_{i,p}(x,y)|^{2^s}\} \le \mathsf{H}_r(y) - \mathsf{H}_r(x), \tag{B.2}$$

noting that, for $q \ge p \ge 0$ and $\varepsilon \in \mathbb{R}$, $|\varepsilon^p| < 1 + |\varepsilon|^q$. We denote $\mathsf{H}_r(y) - \mathsf{H}_r(x)$ the H_r-distance between y and x.

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

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \tag{B.3}$$

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

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

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

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

$$\dot{\mathsf{H}}_r(c_-) = \dot{\mathsf{H}}_r(c_+) = H_r/(C_{\mathsf{H}} K^{1/2}). \tag{B.4}$$

This can be done for sufficiently large K since f is continuous and since the function $\dot{\mathsf{H}}_r(c) = (1 + c^{2^r p})\mathsf{f}(c)$ is integrable by Assumption 4.1(ia).

The first inequality concerns the H_r-distance of additive perturbations of the $]c_{k-1}, c_k]$ intervals. It is used in the proof of the inequality in Lemma B.2.

Lemma B.1. *Suppose Assumption 4.1(i) only holds for $\nu = 1$. Then a constant $C > 0$ exists so that for all K satisfying (B.4)*

$$\sup_{1 \le k \le K} \sup_{|d| \le K^{-1/2}} \{\mathsf{H}_r(c_k + d) - \mathsf{H}_r(c_{k-1} + d)\} \le C H_r/K.$$

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

$$\underline{\dot{\mathsf{H}}}_r(c) = \inf_{0 \le d \le c} \dot{\mathsf{H}}_r(d), \qquad \overline{\dot{\mathsf{H}}}_r(c) = \sup_{d \ge c} \dot{\mathsf{H}}_r(d),$$

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

$$C_{\mathsf{H}}^{-1}\overline{\dot{\mathsf{H}}}_r(c) \le \underline{\dot{\mathsf{H}}}_r(c) \le \dot{\mathsf{H}}_r(c) \le \overline{\dot{\mathsf{H}}}_r(c) \le C_{\mathsf{H}}\underline{\dot{\mathsf{H}}}_r(c). \tag{B.5}$$

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

$$\sup_{c \in \mathbb{R}} |\ddot{\mathsf{H}}_r(c)| < \infty. \tag{B.6}$$

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

$$H_r/K = \mathsf{H}_r(c_\ell) - \mathsf{H}_r(c_{\ell-1}) = (c_\ell - c_{\ell-1})\dot{\mathsf{H}}_r(c_\ell^*). \tag{B.7}$$

Two inequalities for $\dot{\mathsf{H}}_r(c)$ arise from (B.5) and condition (B.4). These are

$$\dot{\mathsf{H}}_r(c) \le \overline{\dot{\mathsf{H}}}_r(c) \le \overline{\dot{\mathsf{H}}}_r(c_+) \le C_{\mathsf{H}}\dot{\mathsf{H}}_r(c_+) = H_r/K^{1/2} \text{ for } \ c \ge c_+, \tag{B.8}$$

$$\dot{\mathsf{H}}_r(c) \ge \underline{\dot{\mathsf{H}}}_r(c) \ge \underline{\dot{\mathsf{H}}}_r(c_+) \ge \overline{\dot{\mathsf{H}}}_r(c_+)/C_{\mathsf{H}} \ge \dot{\mathsf{H}}_r(c_+)/C_{\mathsf{H}} = H_r/(C_{\mathsf{H}}^2 K^{1/2}) \text{ for } \ 0 \le c \le c_+. \tag{B.9}$$

In parallel to (B.9), which is derived for positive c, it holds for negative c that

$$\dot{\mathsf{H}}_r(c) \ge H_r/(C_{\mathsf{H}}^2 K^{1/2}) \qquad \text{for } 0 \ge c \ge c_-. \tag{B.10}$$

3. *Small arguments* $c_- \le c_k^* \le c_+$. Combine (B.7), (B.9) and (B.10) to get

$$c_k - c_{k-1} = H_r/\{K\dot{\mathsf{H}}_r(c_k^*)\} \le C_{\mathsf{H}}^2/K^{1/2}. \tag{B.11}$$

Two second order Taylor expansions give

$$\mathsf{H}_r(c_k + d) - \mathsf{H}_r(c_k) = d\dot{\mathsf{H}}_r(c_k) + (d^2/2)\ddot{\mathsf{H}}_r(c_k^{**}),$$
$$\mathsf{H}_r(c_{k-1} + d) - \mathsf{H}_r(c_{k-1}) = d\dot{\mathsf{H}}_r(c_{k-1}) + (d^2/2)\ddot{\mathsf{H}}_r(c_{k-1}^{**}),$$

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

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

The left hand side is $\mathcal{H} - H_r/K$. The mean-value theorem gives that for some $\tilde{c}_k$, $c_{k-1} \le \tilde{c}_k \le c_k$, $\dot{\mathsf{H}}_r(c_k) - \dot{\mathsf{H}}_r(c_{k-1}) = (c_k - c_{k-1})\ddot{\mathsf{H}}_r(\tilde{c}_k)$. Insert this and rearrange to get

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

Using the bound $c_k - c_{k-1} \leq C_{\mathsf{H}}^2/K^{1/2}$ from (B.11), and the bound $|d| \leq K^{-1/2}$, it follows that $0 \leq \mathcal{H} \leq C/K$, where $C = H_r + (C_{\mathsf{H}}^2 + 1) \sup_{c\in\mathbb{R}} |\ddot{\mathsf{H}}_r(c)|$ does not depend on K.

4. *Inequalities on tail grid point intervals.* Suppose $c_k^* \geq c_+$. This includes the situation where c_k^* and c_+ are in the same grid interval. Expansion (B.7) and inequality (B.8) imply

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

5. *Large arguments* $c_k^* \geq c_+$ so either $k \geq k_+ + 2$ or $k = k_+ + 1$ with $c_{k-1}^* \geq c_+$. In this case $c_{k-1}^* \geq c_+$ so that Item 4 shows that $c_{k+1} - c_k$, $c_k - c_{k-1}$ and $c_{k-1} - c_{k-2}$ are all larger than $|d|$. Therefore

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

It then holds that $0 \leq \mathcal{H} \leq \mathsf{H}_r(c_{k+1}) - \mathsf{H}_r(c_{k-2}) = 3H_r/K$.

6. *Intermediate arguments* $c_k^* \geq c_+$ so that $k = k_+$. Item 4 shows $c_k - c_{k-1} \geq |d|$ and $c_{k+1} - c_k \geq |d|$. Therefore $0 \leq \mathcal{H} \leq \mathsf{H}_r(c_{k+1}) - \mathsf{H}_r(c_{k-1}) = 2H_r/K$.

7. *Intermediate arguments* $c_k^* \geq c_+$ so that $k = k_+ + 1$ with $c_{k-1}^* < c_+$ and $c_{k-1} + d \geq c_+$. Decompose $0 \leq \mathcal{H} \leq \mathcal{H}_1 + \mathcal{H}_2$ where

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

Consider $\mathcal{H}_1$. Argue $\mathcal{H}_1 \leq 2H_r/K$ as in item 5.
Consider $\mathcal{H}_2$. Argue $c_{k-1} - c_+ \geq |d|$ as in item 4 and in turn $\mathcal{H}_2 \leq H_r/K$ as in item 5.

8. *Intermediate arguments* $c_k^* \geq c_+$ so $k = k_+ + 1$ with $c_{k-1}^* < c_+$ and $c_{k-1} + d < c_+$. Decompose $0 \leq \mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ were defined and analyzed in item 6, while

$$\mathcal{H}_3 = \mathsf{H}_r(c_+) - \mathsf{H}_r(c_{k-1} + d).$$

Since $c_+ \leq c_{k_+} = c_{k-1}$ and $c_{k-1} + d < c_+$ then $c_{k-1} - c_+ \leq |d|$. The mean-value theorem shows

$$\mathcal{H}_3 = \delta_{k,d}\dot{\mathsf{H}}_r(c_+) + (\delta_{k,d}^2/2)\ddot{\mathsf{H}}_r(c^{**}),$$

where $\delta_{k,d} = c_{k-1} + d - c_+$ while c^{**} satisfies $|c^{**} - c_+| \leq |\delta_{k,d}|$. Here $|\delta_{k,d}| \leq c_{k-1} - c_+ + |d| \leq 2|d| \leq 2K^{-1/2}$. Because (B.4) shows $\dot{\mathsf{H}}_r(c_+) = H_r/(C_{\mathsf{H}}K^{1/2})$, while $\ddot{\mathsf{H}}_r(c^{**})$ is bounded by (B.6), it follows that $\mathcal{H}_3 \leq C/K$. □

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 results are used in the proofs of Theorems 4.1, 4.2, that are concerned with estimation error b in the empirical process $\mathbb{F}_n^{q,p}(b,c)$. The proof uses Lemma B.1.

Lemma B.2. *Suppose Assumption 4.1(i) only holds for $\nu = 1$. For any $c \leq c_{K-1}$ we choose grid points, see (B.3), $c_{k-1} < c \leq c_k (\leq c_{K-1})$. For $c > c_{K-1}$ we consider $c_{K-1} < c < c_K (= \infty)$. Then an integer $k_J > 0$ exists such that, for all K satisfying*

(B.4) and all $c, d, d_m \in \mathbb{R}$ for which $|d| \leq K^{-1/2}$ and $|d - d_m| \leq K^{-1}$, integers $k^\dagger, k^\ddagger$ exist for which

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

Proof of Lemma B.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 $|\mathcal{I}_2| \leq 1_{(c^\dagger - K^{-1} \leq \varepsilon_i \leq c^\dagger + K^{-1})}$. Using first this inequality and then the mean-value theorem, it follows that

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

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

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

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

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

The next inequality gives a tightness type result for the function H_r. This lemma is used in the proof of the tightness result for the empirical process $\mathbb{F}_n^{g,p}(0, c)$ in Theorem 4.4.

Lemma B.3. *Let $c_\psi = \mathsf{F}^{-1}(\psi)$. For all densities satisfying Assumption 4.1(ia) for some $\nu < 1$, there exist $C_\nu, \phi_0 > 0$ such that for all $0 \leq \phi \leq \phi_0$ it follows that*

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

Proof of Lemma B.3. Let $\psi_0 = \mathsf{F}(0)$. Note that $2^r p$ is even for $r \in \mathbb{N}$, $p \in \mathbb{N}_0$.

1. *Let $\psi \geq \psi_0$.* Then $\mathsf{H}_r(c_{\psi+\phi}) - \mathsf{H}_r(c_\psi)$ is increasing in ψ since, with $\dot{c}_\psi = 1/\mathsf{f}(c_\psi)$,

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

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

$$\mathsf{H}_r(\infty)-\mathsf{H}_r(c_{1-\phi})=\int_{c_{1-\phi}}^{\infty}(1+u^{p2^r})\mathsf{f}(u)du=\phi+\int_{c_{1-\phi}}^{\infty}u^{p2^r}\mathsf{f}(u)du.$$

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

$$\mathsf{H}_r(\infty)-\mathsf{H}_r(c_{1-\phi})\leq\phi+C^\nu\int_{c_{1-\phi}}^{\infty}\{1-\mathsf{F}(u)\}^{-\nu}\mathsf{f}(u)du.$$

Substituting $x=\mathsf{F}(u)$, so that $dx=\mathsf{f}(u)du$ gives

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

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

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

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

Using the mean-value theorem there exists a ψ^*, in the interval $\psi_0-\phi\leq\psi^*\leq\psi_0+\phi$, for which

$$\mathcal{H}\leq\frac{\dot{\mathsf{H}}_r\{\dot{\mathsf{H}}_r(c_{\psi^*})\}}{\mathsf{f}(c_{\psi^*})}2\phi=2(1+c_{\psi^*}^{2^rp})\phi\leq 2\{1+\max(c_{\psi_0-\phi_0}^{2^rp},c_{\psi_0+\phi_0}^{2^rp})\}\leq C\phi,$$

for some $C>0$, because ϕ_0 can be chosen so that the two quantiles are finite.

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

Appendix C: Proofs of auxiliary 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 the axis as laid out in (B.3) with $K=\text{int}(H_rn^{1/2}/\delta)$ using Assumption 4.1(ia) with $\nu=1$ only.

2. *Assign c_ψ to the partitioned support.* Consider $0\leq\psi\leq 1$. 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 balls of radius $n^{-\zeta}$ with centers b_m, that is $M=\mathrm{O}\{n^{(1/4-\eta+\zeta)\dim x}\}$. Thus, for any b there exists a b_m so that $|b-b_m|<n^{-\zeta}$.

4. *Apply chaining.* For $k < K$ where $c_\psi \le c_k \le c_{K-1}$, we compare c_ψ to the nearest right grid point, c_k, using $\tilde{R}(b, c_\psi) = \tilde{R}(b_m, c_k) + \{\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_k)\}$, whereas for $k = K$, we use the nearest left grid point, c_{K-1}, and get $\tilde{R}(b, c_\psi) = \tilde{R}(b_m, c_{K-1}) + \{\tilde{R}(b, c_\psi) - \tilde{R}(b_m, c_{K-1})\}$. Therefore $R_n \le R_{n,1} + R_{n,2}$, where

$$\mathcal{R}_{n,1} = \max_{1\le k<K} \max_{1\le m\le M} |\tilde{R}(b_m, c_k)|,$$

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

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

5. *The term* $\mathcal{R}_{n,1}$. Use Theorem 5.2 to see that $\mathcal{R}_{n,1} = \mathsf{o}_\mathsf{P}(1)$. To see this, let $\upsilon = 1/2$ and let g_{in} have coordinates g^*_{in}. Then, for $z_{\ell i} = g^*_{in} J_{i,p}(c_k, c_k + \sigma^{-1} x'_{in} b_m)$ we write the coordinates of $\tilde{R}(b_m, c_k)$ as $n^{-1/2}\sum_{i=1}^n (z_{\ell i} - \mathsf{E}_{i-1} z_{\ell i})$, see definition in (B.1), and where ℓ represents the indices k, m. The conditions of Theorem 5.2 need to be verified.

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

The parameter ς. Because $|1_{(\varepsilon_i \le c_k + x'_{in} b_m)} - 1_{(\varepsilon_i \le c_k)}| \le 1_{(c_k - |x_{in}||b_m| < \varepsilon_i \le c_k + |x_{in}||b_m|)}$ we find for $1 \le q \le r$, that

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

using the mean-value theorem. Because $|b_m| \le n^{1/4-\eta}B$, while $\sup_{v\in\mathbb{R}} \dot{\mathsf{H}}_r(v) < \infty$ by Assumption 4.1(*ib*), we find

$$D_q = \max_{1\le\ell\le L} \sum_{i=1}^n \mathsf{E}_{i-1}(z_{\ell i})^{2^q} \le C_1 (n^{-1} \sum_{i=1}^n |g^*_{in}|^{2^q} |n^{1/2} x_{in}|) n^{3/4-\eta}. \tag{C.1}$$

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

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

Condition (*ii*) is that $\varsigma + \lambda < \upsilon 2^{\overline{r}}$ with $\overline{r} = r$. If $\zeta > \kappa$ is chosen sufficiently small, then

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

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

6. *Decompose* $\mathcal{R}_{n,2}$. It will be argued that $\mathcal{R}_{n,2} \le 3(\widetilde{\mathcal{R}}_{n,2} + 2\overline{\mathcal{R}}_{n,2}) + \mathsf{o}_\mathsf{P}(1)$, where

$$\widetilde{\mathcal{R}}_{n,2} = \max_{1\le k\le K} n^{-1/2} \sum_{i=1}^n |g_{in}| \{|J_{i,p}(c_{k-k_J}, c_k)| - \mathsf{E}_{i-1}|J_{i,p}(c_{k-k_J}, c_k)|\}, \tag{C.2}$$

$$\overline{\mathcal{R}}_{n,2} = \max_{1\le k\le K} n^{-1/2} \sum_{i=1}^n |g_{in}| \mathsf{E}_{i-1}|J_{i,p}(c_{k-k_J}, c_k)|. \tag{C.3}$$

To see this, let c_k denote the nearest right grid point for $c_\psi \le c_{K-1}$ while $c_k = c_{K-1}$ for $c_\psi > c_{K-1}$. Note first that $\tilde{R}^p_\mathsf{F}(b, c_\psi) - \tilde{R}^p_\mathsf{F}(b_m, c_k)$ 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 $\max_{1\leq i\leq n}|x_{in}| = \mathsf{O}_\mathsf{P}(n^{\kappa-1/2})$. Thus, for all $\epsilon>0$ a $C_x>0$ 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, using $d=x_{in}'b$ and $d_m=x_{in}'b_m$, $|d|=\mathrm{O}(n^{-1/4+\kappa-\eta})=\mathrm{o}(K^{-1/2})$ for $\eta-\kappa>0$ and $|d-d_m| = \mathrm{O}(n^{-1/2+\kappa-\zeta}) = \mathrm{o}(K^{-1})$ for $\zeta-\kappa>0$. Thus, for sufficiently large $n, |d|<K^{-1/2}$ and $|d-d_m|<K^{-1}$. Lemma B.2 using Assumption 4.1(i) then shows that a k_J exists so that, for all c,d,d_m, there exist $k^\dagger,k^\ddagger$ for which

$$|\mathcal{J}_i| \leq |J_{i,p}(c_{k-k_J},c_k)| + |J_{i,p}(c_{k^\dagger-k_J},c_{k^\dagger})| + |J_{i,p}(c_{k^\ddagger-k_J},c_{k^\ddagger})|, \tag{C.4}$$

As a consequence it holds, as desired, that $\mathcal{R}_{n,2}\leq 3(\widetilde{\mathcal{R}}_{n,2}+2\overline{\mathcal{R}}_{n,2})+\mathsf{o}_\mathsf{P}(1)$.

7. *The term* $\widetilde{\mathcal{R}}_{n,2}$ is $\mathsf{o}_\mathsf{P}(1)$ by Lemma 5.2. Let $\upsilon=1/2$. To see this, note that $\widetilde{\mathcal{R}}_{n,2}$ is the maximum of a family of martingales of the required form with $\ell=k$ so that $L=K$ 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 $K=\mathrm{int}(H_r n^{1/2}/\delta)$.

Condition (ii) holds with $\varsigma=1/2$ since $\mathsf{E}_{i-1}(J_{i,p})^{2^{\bar{r}}}\leq \mathsf{H}_r(c_k)-\mathsf{H}_r(c_{k-k_J})=k_JH_r/K$ for $r\geq\bar{r}=2$. Thus $\sum_{i=1}^n\mathsf{E}_{i-1}(J_{i,p})^{2^{\bar{r}}}=\mathrm{O}(n^{1-1/2})$, uniformly in ℓ,i.

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

8. *Bounding* $\bar{\mathcal{R}}_{n,2}$. Note $\mathsf{E}_{i-1}|J_{i,p}(c_{k-k_J},c_k)|\leq k_JH_r/K\leq 2k_J\delta n^{-1/2}$ uniformly in i,k by the same argument as in item 7 and since $K=\mathrm{int}(H_rn^{1/2}/\delta)$. It follows that $\bar{\mathcal{R}}_{n,2}\leq 2k_J\delta n^{-1}\sum_{i=1}^n|g_{in}|$. Here $n^{-1}\sum_{i=1}^n|g_{in}|$ $=\mathsf{O}_\mathsf{P}(1)$ by Markov's inequality and Assumption 4.1$(iiia)$, so that $\bar{\mathcal{R}}_{n,2}=\mathsf{O}_\mathsf{P}(\delta)$. Thus, choosing δ sufficiently small, $\bar{\mathcal{R}}_{n,2}$ 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

$$\mathcal{S}_1=\sup_{0\leq\psi\leq1}\sup_{|b|\leq n^{1/4+\kappa-\eta}B}\sup_{d\in\mathbb{R}}|\mathbb{F}_n^{1,0}(b,c_\psi+n^{\kappa-1/2}d)-\mathbb{F}_n^{1,0}(0,c_\psi+n^{\kappa-1/2}d)|=\mathsf{o}_\mathsf{P}(n^{-\omega}),$$

$$\mathcal{S}_2=\sup_{0\leq\psi\leq1}\sup_{|d|\leq n^{1/4+\kappa-\eta}B}|\mathbb{F}_n^{1,0}(0,c_\psi+n^{\kappa-1/2}d)-\mathbb{F}_n^{1,0}(0,c_\psi)|=\mathsf{o}_\mathsf{P}(n^{-\omega}).$$

For each term the proof of Theorem 4.1 is used with minor modifications. Since $p=0$ then $2^rp=0$ for all r, which simplifies the assumptions, see Remark 4.1(a). Moreover, when using Theorem 5.2, $z_{\ell,i}^{2^r}=z_{\ell,i}$ for all $r\geq1$. Thus, it suffices to check $\varsigma<2\upsilon$ and $\lambda<\infty$.

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

1. Choose $K=\mathrm{int}(H_rn^{1/2+1/8+\omega/2}/\delta)$ where $\omega<\eta-\kappa\leq1/4$.

2. For each $c_\psi+n^{\kappa-1/2}d$ there exist c_{k-1},c_k depending on n so that $c_{k-1}<c_\psi+n^{\kappa-1/2}d\leq c_k$.

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 $M=\mathrm{O}\{n^{(1/4+\kappa-\eta+\zeta)\dim x}\}$.

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 $\mathcal{R}_{n,1}$ is $\mathsf{o}_\mathsf{P}(n^{-1/8-\omega/2})$. Use Theorem 5.2 with $\upsilon = 3/8-\omega/2 > 1/2+\kappa-\eta$. Define $z_{\ell i}$ as before. Since $p=0$, $g_i n = 1$ then $|J_{i,p}(x,y)|^{2^r} = |J_{i,p}(x,y)|$ and $|z_{\ell i}| = |{z_{\ell i}}^{2^r}|$ for any $r \in \mathbb{N}_0$. The inequality (C.1) for D_q holds as before, uniformly in $q \in \mathbb{N}$ so $\varsigma = 3/4+\kappa-\eta$. Thus, *condition* (i) holds since $\varsigma = 3/4+\kappa-\eta < 3/4-\omega = 2\upsilon$. Moreover, $\lambda = 1/2+\omega+(1/4+\kappa-\eta+\zeta)\dim x$ is finite so *condition* (ii) holds for some r.

6. Lemma B.2 is an analytic result holding in finite samples. So the argument is not affected by the dependence of c_k on n through $c_\psi + n^{\kappa-1/2}d$. In particular, (C.4) holds as stated and therefore the decomposition of $\mathcal{R}_{n,2}$ holds, noting that K is now chosen differently.

7. The term $\tilde{\mathcal{R}}_{n,2}$ is $\mathsf{o}_\mathsf{P}(n^{-1/4})$. Use Theorem 5.2 with some $\upsilon > 3/16-\omega/4$. Here $\lambda = 5/8+\omega/2 < \infty$ by the definition of K, while $\varsigma = 3/8-\omega/2$ since $\mathsf{E}_{i-1}(J_{i,p})^4 = \mathsf{E}_{i-1}(J_{i,p}) \le \mathsf{H}_r(c_k) - \mathsf{H}_r(c_{k-k_J}) = k_J H_r/K$ so that $\sum_{i=1}^n \mathsf{E}_{i-1}(J_{i,p})^4 = \mathrm{O}(n^{1-5/8-\omega/2})$, uniformly in ℓ, i. Thus, *condition* (i) holds with $\varsigma = 3/8-\omega/2 \le 2\upsilon$ while *condition* (ii) holds for some r.

8. Note $\mathsf{E}_{i-1}|J_{i,p}(c_{k-k_J}, c_k)| \le 2k_J\delta n^{-5/8-\omega}$ uniformly in i,k by the same argument as in item 7. Since $g_{in}=1$ then $\bar{\mathcal{R}}_{n,2} = \mathrm{O}_\mathsf{P}(n^{-5/8-\omega}) = \mathsf{o}_\mathsf{P}(n^{-1/4})$.

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

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

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

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

Recalling the definition of $\overline{\mathsf{F}}_n^{g,p}$ 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

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

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

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

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

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

Thus, by the triangular inequality

$$|R(b, c_\psi)| \le n^{-1/2}\textstyle\sum_{i=1}^n |g_{in}||S_i(b, c_\psi)| \le \mathrm{O}(n^{-2\eta}) n^{-1}\textstyle\sum_{i=1}^n |g_{in}||n^{1/2}x_{in}|^2 \sup_{c\in\mathbb{R}} |\dot{\mathsf{g}}(c)|.$$

Due to Assumption 4.1($ib, iiib$), this expression is of order $\mathrm{O}_\mathsf{P}(n^{-2\eta})$ 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 $\epsilon < 1$. Take $0 < \epsilon$ and n as well as $0 < \phi^{(1-\nu)/4} \le \epsilon^2$ as given. Throughout, $C > 0$ denotes as usual a constant not depending on ϕ, n, ϵ, which may have a different value in different expressions. Let $r = 2$. Since $\psi^\dagger - \psi \le \phi$, Lemma B.3 with Assumption 4.1(ia) shows that $0 < \nu < 1$ and $C_v, \phi_0 > 0$ exist such that $\mathsf{H}_r(c_{\psi^\dagger}) - \mathsf{H}_r(c_\psi) \le C\phi^{1-\nu}$ for $0 \le \phi \le \phi_0$. The proof will use a dyadic argument. Given ϵ, ϕ, n we will choose numbers $\overline{m}, \underline{m}$ and derive a bound to the probability not depending on $\overline{m}, \underline{m}$.

2. *Fine grid.* Let $\overline{m}$ satisfy $2^{-\overline{m}} \le n^{-1/2}\epsilon\phi^{(1-\nu)/4} \le 2^{1-\overline{m}}$.

3. *Coarse grid.* Let $\underline{m}$ satisfy $2^{-\underline{m}-1}H_r < C\phi^{1-\nu} \le 2^{-\underline{m}}H_r$. For large n, $\overline{m} > \underline{m}$.

4. *Partition support.* For each of $m = \underline{m}, \ldots, \overline{m}$ partition axis as laid out in (B.3) with $K_m = 2^m$ points. For each m, points $c_{k_m,m}$ and $c_{k^\dagger,m}$ exist so that $\underline{c}_m = c_{k_m-1,m} < c_\psi \le c_{k_m,m} = \overline{c}_m$ and $\underline{c}^\dagger_m = c_{k^\dagger_m-1,m} < c_{\psi^\dagger} \le c_{k^\dagger_m,m} = \overline{c}^\dagger_m$. Then $\overline{c}_{m-1} = c_{k_{m-1},m-1}$ equals either $\overline{c}_m = c_{k_m,m}$ or $c_{k_m+1,m}$ so that $\overline{c}_{m-1} \ge \overline{c}_m$ and $\mathsf{H}(\overline{c}_{m-1}) - \mathsf{H}(\overline{c}_m)$ is either zero or $2^{-m}H_r$. There is at most one $\underline{m}$-grid point in the interval $c_\psi, c_{\psi^\dagger}$.

5. *Decompose* $J_{i,p}(c_\psi, c_{\psi^\dagger})$, see definition in (B.1). Split the $c_\psi, c_{\psi^\dagger}$ interval into three intervals where the partitioning points are $\overline{c}_{\overline{m}}$ and $\underline{c}^\dagger_{\overline{m}}$ which are the fine grid points to the right of c_ψ and to the left of $c_{\psi^\dagger}$, respectively. Note, that if $c_\psi, c_{\psi^\dagger}$ are in the same $\overline{m}$-interval then $\overline{c}_{\overline{m}} > \underline{c}^\dagger_{\overline{m}}$ and if they are in neighbouring $\overline{m}$-interval then $\overline{c}_{\overline{m}} = \underline{c}^\dagger_{\overline{m}}$. Thus,

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

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

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

If $\overline{c}_{\overline{m}-1} = \underline{c}^\dagger_{\overline{m}-1}$ then $J_{i,p}(\overline{c}_{\overline{m}-1}, \underline{c}^\dagger_{\overline{m}-1}) = 0$ and the iteration stops, noting that for $m < \overline{m}-1$ the m-grid points cross over so that $\overline{c}_m \ge \overline{c}_{\overline{m}-1} = \underline{c}^\dagger_{\overline{m}-1} \ge \underline{c}^\dagger_m$. If $\overline{c}_{\overline{m}-1} < \underline{c}^\dagger_{\overline{m}-1}$, the argument can be made again for $J_{i,p}(\overline{c}_{\overline{m}-1}, \underline{c}^\dagger_{\overline{m}-1})$. In the m-th step, the iteration continues if $\overline{c}_m < \underline{c}^\dagger_m$, so that if there are no other m-grid points between $\overline{c}_{\overline{m}}$ and $\underline{c}^\dagger_{\overline{m}}$, the contribution from the $(m-1)$-step is zero. Because there is at most one $\underline{m}$-point in the interval $c_\psi, c_{\psi^\dagger}$, 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

$$1_{(\overline{c}_{\overline{m}} < \underline{c}^\dagger_{\overline{m}})} J_{i,p}(\overline{c}_{\overline{m}}, \underline{c}^\dagger_{\overline{m}}) = \textstyle\sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m < \underline{c}^\dagger_m)} \{J_{i,p}(\overline{c}_m, \overline{c}_{m-1}) + J_{i,p}(\underline{c}^\dagger_{m-1}, \underline{c}^\dagger_m)\}.$$

6. *Decompose* $\mathcal{S} = n^{1/2}\{\mathbb{F}_n^{g,p}(0,c_{\psi^\dagger}) - \mathbb{F}_n^{g,p}(0,c_\psi)\}$. 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

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

7. *The term Z_1 : Finding martingale.* Bound $|J_{i,p}(c_\psi, \overline{c}_{\overline{m}})| \leq |J_{i,p}(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})|$ where the points $\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}}$ are two neighbouring points on the $\overline{m}$-grid, but their location depends on ψ. It follows that

$$\sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |Z_1|$$
$$\leq \frac{1}{\sqrt{n}}\sum_{i=1}^n |g_{in}|\{|J_{i,p}(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})| + \mathsf{E}_{i-1}|J_{i,p}(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})|\}$$
$$\leq \max_{1\leq\ell\leq 2^{\overline{m}}} \frac{1}{\sqrt{n}}\sum_{i=1}^n |g_{in}|\{|J_{i,p}(c_{\ell-1,m}, c_{\ell,m})| + \mathsf{E}_{i-1}|J_{i,p}(c_{\ell-1,m}, c_{\ell,m})|\}.$$

Thus, a martingale decomposition gives

$$\sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |Z_1| \leq \max_{1\leq\ell\leq 2^{\overline{m}}} |\widetilde{V}_{1,\ell,\overline{m}}| + 2\max_{1\leq\ell\leq 2^{\overline{m}}} \overline{V}_{1,\ell,\overline{m}},$$

where

$$\widetilde{V}_{1,\ell,\overline{m}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n |g_{in}|[|J_{i,p}(c_{\ell-1,\overline{m}}, c_{\ell,\overline{m}})| - \mathsf{E}_{i-1}\{|J_{i,p}(c_{\ell-1,\overline{m}}, c_{\ell,\overline{m}})|\}], \tag{C.5}$$
$$\overline{V}_{1,\ell,\overline{m}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n |g_{in}|\mathsf{E}_{i-1}\{|J_{i,p}(c_{\ell-1,\overline{m}}, c_{\ell,\overline{m}})|\}. \tag{C.6}$$

8. *The term Z_1 : The compensator $\overline{V}$.* Since

$$\mathsf{E}_{i-1}\{|J_{i,p}(c_{\ell-1,\overline{m}}, c_{\ell,\overline{m}})|^{2^r}\} \leq \mathsf{H}_r(c_{\ell,\overline{m}}) - \mathsf{H}_r(c_{\ell-1,\overline{m}}) = 2^{-\overline{m}}H_r,$$

Assumption 4.1$(ia, iiia)$ implies

$$\mathsf{E}\max_{1\leq\ell\leq 2^{\overline{m}}} \sum_{i=1}^n |g_{in}|^{2^r}\mathsf{E}_{i-1}\{|J_{i,p}(c_{\ell-1,\overline{m}}, c_{\ell,\overline{m}})|^{2^r}\} \leq nC2^{-\overline{m}}H_r. \tag{C.7}$$

Item 2 shows $2^{-\overline{m}} \leq n^{-1/2}\epsilon\phi^{(1-\nu)/4}$. Thus, the Markov inequality implies

$$\mathsf{P}(\max_{1\leq\ell\leq 2^{\overline{m}}} \overline{V}_{1,\ell,\overline{m}} > \epsilon) \leq \frac{1}{\epsilon}\mathsf{E}\max_{1\leq\ell\leq 2^{\overline{m}}} \overline{V}_{1,\ell,\overline{m}} \leq n^{1/2}\frac{1}{\epsilon}C2^{-\overline{m}}H_r = C\phi^{(1-\nu)/4}.$$

9. *The term* Z_1 *: The martingale* $\widetilde{V}$. Apply Theorem 5.3 with $z_{\ell,i} = g^*_{in}|J_{i,p}(c_{\ell,\overline{m}}, c_{\ell+1,\overline{m}})|$ where g^*_{in} is a coordinate of $|g_{in}|$, and with $L = 2^{\overline{m}}$ and $\kappa = \epsilon$ while $D = C2^{-\overline{m}}$ by the inequality (C.7), to get

$$\mathsf{P}(\max_{1\leq\ell\leq 2^{\overline{m}}} |\widetilde{V}_{1,\ell,\overline{m}}| > \epsilon) \leq C2^{-\overline{m}}\frac{\theta}{\epsilon} + C\frac{\theta^3}{n\epsilon} + C2^{\overline{m}}\exp(-\frac{\epsilon\theta}{14}), \tag{C.8}$$

where we can choose $\theta = 14\epsilon^{-1}(\log 2^{2\overline{m}} + \log\phi^{-1})$. First term in (C.8) satisfies

$$C2^{-\overline{m}}\frac{\theta}{\epsilon} \leq C\frac{1}{\epsilon^2}2^{-\overline{m}/2}\{\overline{m}2^{-\overline{m}/2} + \phi^{-(1-\nu)/2}2^{-\overline{m}/2}\phi^{(1-\nu)/2}\log\phi^{-1}\} \leq C\phi^{(1-\nu)/4},$$

since the bounds in items 1,3 imply $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $2^{-\overline{m}/2} \leq 2^{-\underline{m}/2} < C\phi^{(1-\nu)/2}$, while the functions $m2^{-m/2}$ and $\phi^{(1-\nu)/2}\log\phi^{-1}$ are bounded for $m \geq 1$ and $0 < \phi < 1$. Second term in (C.8): Use first the definition of θ with the inequality $(x+y)^3 \leq C(x^3+y^3)$ and then that the bounds in items 1, 2 imply $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $n^{-1}\epsilon^2 \leq C\phi^{-(1-\nu)/2}2^{-2\overline{m}}$ so that

$$\mathcal{P}_1 = C\frac{\theta^3}{n\epsilon} \leq C\frac{1}{n\epsilon^4}(\overline{m}^3 + \log^3\frac{1}{\phi}) \leq C\phi^{-(1-\nu)/2}2^{-2\overline{m}}\phi^{-3(1-\nu)/4}(\overline{m}^3 + \log^3\frac{1}{\phi}).$$

Rewrite this bound and argue as for the first term, to get that

$$\leq C\{2^{-\overline{m}/2}\phi^{-(1-\nu)/2}\}^3\{\overline{m}^3 2^{-\overline{m}/2} + \phi^{-(1-\nu)/2}2^{-\overline{m}/2}\phi^{(1-\nu)/2}\log^3\frac{1}{\phi}\}\phi^{(1-\nu)/4} \leq C\phi^{(1-\nu)/4}.$$

Third term in (C.8) satisfies

$$C2^{\overline{m}}\exp(-\frac{\epsilon\theta}{14}) = C2^{-\overline{m}}\phi \leq C\phi^{(1-\nu)/4},$$

since $\phi \leq \phi^{(1-\nu)/4}$ for $0 < \phi < 1$. In summary, $\mathsf{P}(\max_{1\leq\ell\leq 2^{\overline{m}}} |\widetilde{V}_{1,\ell,\overline{m}}| > \epsilon) \leq C\phi^{(1-\nu)/4}$.

10. *The terms* Z_2 *and* Z_3. Apply the same argument as in item 7-9.

11. *The term* Z_4*: finding martingale.* Recall that for instance $\overline{c}_m = c_{k_m,m}$ while $\overline{c}_{m-1}$ either equals $c_{k_m,m}$ or $c_{k_m+1,m}$, so that $\overline{c}_m, \overline{c}_{m-1}$ are at most 1 step apart in the m-grid. Let

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

It then holds that

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

Note that $\sum_{m=\underline{m}+1}^{\overline{m}} 2^{(\underline{m}-m)/4} \leq \sum_{j=1}^{\infty} 2^{-j/4} = (2^{1/4}-1)^{-1} < 6$, and that the right hand side does not depend on ψ. It therefore holds

$$\mathcal{P}_4 = \mathsf{P}(\sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |Z_4| > \epsilon) \leq \mathsf{P}\bigcup_{m=\underline{m}+1}^{\overline{m}} \left\{ \max_{1\leq\ell\leq 2^m} |M_{\ell,m,n}| > \frac{2^{(\underline{m}-m)/4}\epsilon}{6} \right\}.$$

Using Boole's inequality then

$$\mathcal{P}_4 \leq \sum_{m=\underline{m}+1}^{\overline{m}} \mathsf{P}\left\{ \max_{1\leq\ell\leq 2^m} |M_{\ell,m,n}| > \frac{2^{(\underline{m}-m)/4}\epsilon}{6} \right\}.$$

12. *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 coordinate of g_{in} and with $L = 2^m$ while $\kappa = 2^{(\underline{m}-m)/4}\epsilon/6$ and $D = C2^{-m}$, due to the inequality (C.7) with $\overline{m}$ replaced by m. Thus

$$\mathcal{P}_4 \leq C\sum_{m=\underline{m}+1}^{\overline{m}} \left\{ 2^{-m}\frac{\theta_m}{2^{(\underline{m}-m)/4}\epsilon} + \frac{\theta_m^3}{n2^{(\underline{m}-m)/4}\epsilon} + 2^m \exp(-\frac{2^{(\underline{m}-m)/4}\epsilon\theta_m}{84}) \right\}, \quad \text{(C.9)}$$

where we choose $2^{(\underline{m}-m)/4}\epsilon\theta_m/84 = \log(4^{m-\underline{m}}) + \log\phi^{-1}$. First term in (C.9) satisfies

$$\mathcal{P}_{41} = \sum_{m=\underline{m}+1}^{\overline{m}} 2^{-m}\frac{\theta_m}{2^{(\underline{m}-m)/4}\epsilon} \leq C\sum_{m=\underline{m}+1}^{\overline{m}} \frac{1}{2^{(\underline{m}+m)/2}\epsilon^2}\{(m-\underline{m}) + \log\frac{1}{\phi}\}.$$

Note $2^{-(\underline{m}+m)/2} = 2^{-(m-\underline{m})/2}2^{-\underline{m}}$. Items 1,3 imply $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $2^{-\overline{m}/2} \leq 2^{-\underline{m}/2} < C\phi^{(1-\nu)/2}$. Next, use that geometric sums are finite and argue as in item 9 to see that

$$\mathcal{P}_{41} \leq C\sum_{m=\underline{m}+1}^{\overline{m}} 2^{-(m-\underline{m})/2}\{2^{-\underline{m}}(m-\underline{m}) + \phi^{(1-\nu)}\log\frac{1}{\phi}\}\phi^{(1-\nu)/4} \leq C\phi^{(1-\nu)/4}.$$

The second term in (C.9) satisfies

$$\mathcal{P}_{42} = \sum_{m=\underline{m}+1}^{\overline{m}} \frac{\theta_m^3}{n2^{(\underline{m}-m)/4}\epsilon} \leq C\sum_{m=\underline{m}+1}^{\overline{m}} \frac{1}{n2^{(\underline{m}-m)}\epsilon^4}\{(m-\underline{m})^3 + \log^3\frac{1}{\phi}\}.$$

Items 1,2 imply $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $n^{-1}\epsilon^2 \leq \phi^{-(1-\nu)/2}2^{2-2\overline{m}}$ so that

$$\mathcal{P}_{42} \leq C\sum_{m=\underline{m}+1}^{\overline{m}} \phi^{-5(1-\nu)/4}2^{-2\overline{m}-\underline{m}+m}\{(m-\underline{m})^3 + \log^3\frac{1}{\phi}\}.$$

Rewrite $2^{-2\overline{m}-\underline{m}+m} = 2^{3(m-\overline{m})/2-(m-\underline{m})/2-\overline{m}/2-3\underline{m}/2}$ to get that $\mathcal{P}_{42}$ is bounded by

$$C\sum_{m=\underline{m}+1}^{\overline{m}} \{\phi^{-(1-\nu)/2}2^{-\underline{m}/2}\}^3 2^{3(m-\overline{m})/2}2^{-(m-\underline{m})/2}\{2^{-\overline{m}/2}(m-\underline{m})^3 + 2^{-\overline{m}/2}\log^3\frac{1}{\phi}\}\phi^{(1-\nu)/4}.$$

Argue as for first term using $2^{-\overline{m}/2} \leq 2^{-\underline{m}/2} < C\phi^{(1-\nu)/2}$ from item 3 to get $\mathcal{P}_{42} \leq C\phi^{(1-\nu)/4}$.

The third term in (C.9) satisfies, noting $2^{\underline{m}} \leq C\phi^{\nu-1}$

$$\mathcal{P}_{43} = \sum_{m=\underline{m}+1}^{\overline{m}} 2^m \exp(-\frac{2^{(\underline{m}-m)/4}\epsilon\theta_m}{84}) = \sum_{m=\underline{m}+1}^{\overline{m}} 2^{-(m-\underline{m})} 2^{\underline{m}}\phi.$$

Noting that $2^{\underline{m}} \leq C\phi^{\nu-1}$ then $\mathcal{P}_{43} \leq C\sum_{m=\underline{m}+1}^{\overline{m}} 2^{-(m-\underline{m})}\phi^{\nu} = C\phi^{\nu}$.

13. *The terms* Z_5. Apply the same argument as for Z_4.
14. *Combine* the bounds from items 8,9,10,12,13 to get

$$\mathsf{P}(\sup_{0\leq\psi\leq\psi^{\dagger}\leq 1:\psi^{\dagger}-\psi\leq\phi} |\mathcal{S}| > \epsilon) \leq \sum_{j=1}^{5} \mathsf{P}(\sup_{0\leq\psi\leq\psi^{\dagger}\leq 1:\psi^{\dagger}-\psi\leq\phi} |Z_j| > \epsilon) \leq C\phi^{(1-\nu)/4} + 2C\phi^{\nu}.$$

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 $\phi \downarrow 0$. □

Appendix D: Proofs of main Theorems 3.1-3.7

The main results for the forward search are proved in a series of steps. Theorem 3.1 shows that asymptotically the forward residuals behave like the quantiles of the absolute errors $|\varepsilon_i|$. It is therefore useful to start by reviewing some known results from the theory of quantile processes. Second, the forward search problem is reformulated in terms of a weighted and marked absolute empirical distribution function $\widehat{\mathsf{G}}_n$. At this point we work with absolute errors and it is natural to move from the general densities of Assumption 4.1 to the symmetric densities of Assumption 3.1. Third, this empirical distribution function is analysed using the results from Section 4. Fourth, the corresponding quantile processes are analysed. Fifth, a single step of the Forward Search is analysed using these results. Sixth, the iteration of the Forward Search is analysed.

D.1. Some known results from the theory of quantile processes

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

$$\widehat{\mathsf{G}}_n(c) = \frac{1}{n}\sum_{i=1}^{n} 1_{(|\varepsilon_i|\leq\sigma c)}. \tag{D.1}$$

The first result gives the asymptotic distribution of the empirical process

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

Lemma D.1. *Billingsley (1999, Theorem 16.15). Let $\mathbb{B}$ be a Brownian bridge so that $\mathbb{B}(\psi)$ is $\mathsf{N}\{0,\psi(1-\psi)\}$-distributed. Then, it holds $\mathbb{G}_n \overset{\mathsf{D}}{\to} \mathbb{B}$ on $D[0,1]$.*

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

$$\hat{c}_\psi = \widehat{\mathsf{G}}_n^{-1}(\psi) = \inf\{c : \widehat{\mathsf{G}}_n(c) \geq \psi\}. \tag{D.2}$$

Empirical quantiles and empirical distribution functions are linked as follows.

Lemma D.2. *Csörgő (1983, Corollaries 6.2.1, 6.2.2). Suppose that* f *is symmetric, differentiable, positive for* $\mathsf{F}^{-1}(0) < c < \mathsf{F}^{-1}(1)$*, decreasing for large* c*, and satisfying* $\gamma = \sup_{c>0} \mathsf{F}(c)\{1-\mathsf{F}(c)\}|\dot{\mathsf{f}}(c)|/\{\mathsf{f}(c)\}^2 < \infty$.
Then, for all $\zeta > 0$,
(a) $\sup_{0\le\psi\le 1} |2\mathsf{f}(c_\psi) n^{1/2}(\hat{c}_\psi - c_\psi) + n^{1/2}\{\widehat{\mathsf{G}}_n(c_\psi) - \psi\}| = \mathrm{o}_{\mathsf{P}}(n^{\zeta-1/4})$;
(b) $\sup_{0\le\psi\le 1} |2\mathsf{f}(c_\psi) n^{1/2}(\hat{c}_\psi - c_\psi) - n^{1/2}\{\mathsf{G}(\hat{c}_\psi) - \psi\}| = \mathrm{o}_{\mathsf{P}}(n^{\zeta-1/2})$;
(c) $\sup_{0\le\psi\le 1} |n^{1/2}\{\mathsf{G}(\hat{c}_\psi) - \psi\} + n^{1/2}\{\widehat{\mathsf{G}}_n(c_\psi) - \psi\}| = \mathrm{o}_{\mathsf{P}}(n^{\zeta-1/4})$.

The result in Lemma D.2(a) shows that the empirical quantile $\hat{c}_\psi$ satisfies, for $0 < \psi < 1$,

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

This is known as the Bahadur (1966) representation. Kiefer (1967, eq. 1.8,1.9) studied parts (b, c), which combine to (a). More details can be found in Csörgő (1983) who also gives almost sure, logarithmic rates.

Some weighted versions of the above results are also needed.

Lemma D.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* q_ψ *is increasing and continuous on* $0 \le \psi \le 1/2$ *and satisfies* $q_\psi = \{\psi \log\log(1/\psi)\}^{1/2} g_\psi$ *for a function* g_ψ *so* $\lim_{\psi\to 0} g_\psi = \infty$*. Then, a probability space exists on which one can define a Brownian bridge* $\mathbb{B}_n$ *for each* n*, so that*
(a) $\sup_{0\le\psi\le 1} |\{\mathbb{G}_n(c_\psi) - \mathbb{B}_n(\psi)\}/q_\psi| = \mathrm{o}_{\mathsf{P}}(1)$;
(b) $\sup_{1/(n+1)\le\psi\le n/(n+1)} |\{\mathsf{f}(c_\psi) n^{1/2}(\hat{c}_\psi - c_\psi) - \mathbb{B}_n(\psi)\}/q_\psi| = \mathrm{o}_{\mathsf{P}}(1)$ *provided the assumptions of Lemma D.2 hold.*

In Lemma D.3 a possible choice of q_w is $\{\psi(1-\psi)\}^\alpha$ for $\alpha < 1/2$, which will be used in the proof of Theorem 3.2. Finally, a continuity property of the Brownian bridge is needed.

Lemma D.4. *Revuz and Yor (1998, Theorem I.2.2 versions) A Brownian motion* $\mathbb{W}$ *is locally Hölder continuous of order* α *for all* $\alpha < 1/2$*. That is,*

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

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

D.2. Absolute empirical process representation

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

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

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 $nNx_ix_i'N$, so that the sum $n^{-1}\sum_{i=1}^n |g_{in}|$ 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*

$$\widehat{\mathsf{G}}_n^{g,p}(b,c)=\frac{1}{n}\sum_{i=1}^n g_{in}\varepsilon_i^p 1_{(|\varepsilon_i-x_{in}'b|\le\sigma c)}, \tag{D.3}$$

for $b\in\mathbb{R}^{\dim x}$ and $c\ge 0$. Here the weights are g_{in} and the 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 $g_{in}=nN'x_ix_i'N$, $p=0$. The variance estimator involves the terms mentioned as well as $g_{in}=1$, $p=2$. When $p=0$, the marks are $\varepsilon_i^0=1$ so that $\widehat{\mathsf{G}}_n^{g,0}$ is a weighted absolute empirical distribution function, similar to that studied by Koul and Ossiander (1994). When also $b=0$,then $\widehat{\mathsf{G}}_n^{1,0}$ equals the empirical distribution function $\widehat{\mathsf{G}}_n$ of (D.1).

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

$$\hat z^{(m)}=\sigma\inf\{c:\widehat{\mathsf{G}}_n^{1,0}(\hat b^{(m)},c)\ge\frac{m+1}{n}\}, \tag{D.4}$$

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

$$\frac{m+1}{n}=\widehat{\mathsf{G}}_n^{1,0}(\hat b^{(m)},\frac{\hat z^{(m)}}{\sigma})=\frac{1}{n}\sum_{i=1}^n 1_{(|\varepsilon_i-x_{in}'\hat b^{(m)}|\le\hat z^{(m)})}=\frac{1}{n}\sum_{i\in S^{(m+1)}}1. \tag{D.5}$$

The least squares estimator has estimation error

$$\hat b^{(m+1)}=N^{-1}(\hat\beta^{(m)}-\beta)=\{\widehat{\mathsf{G}}_n^{xx,0}(\hat b^{(m)},\frac{\hat z^{(m)}}{\sigma})\}^{-1}\{n^{1/2}\widehat{\mathsf{G}}_n^{x,1}(\hat b^{(m)},\frac{\hat z^{(m)}}{\sigma})\}, \tag{D.6}$$

while the asymptotically bias corrected least squares variance estimator satisfies

$$n^{1/2}\{(\hat\sigma_{cor}^{(m+1)})^2-\sigma^2\}=\frac{n^{1/2}}{\tau_{m/n}}[\widehat{\mathsf{G}}_n^{1,2}(\hat b^{(m)},\frac{\hat z^{(m)}}{\sigma})-\{\hat b^{(m+1)}\}'\widehat{\mathsf{G}}_n^{xx,0}(\hat b^{(m)},\frac{\hat z^{(m)}}{\sigma})\{\hat b^{(m+1)}\}]. \tag{D.7}$$

D.3. The absolute empirical distribution

The process $\widehat{\mathsf{G}}_n^{g,p}$ is now analysed using the auxiliary Theorems 4.1-4.4 for the process $\widehat{\mathsf{F}}_n^{g,p}$. Only the four combinations of g_{in},p are now considered as outlined in Section D.2. When checking Assumption 4.1 it suffices to check the conditions for the hybrid case where $g_{in}=nN'x_ix_i'N$ and $p=2$. The process $\widehat{\mathsf{G}}_n^{g,p}$ can be expressed in terms of $\widehat{\mathsf{F}}_n^{g,p}$ by

$$\widehat{\mathsf{G}}_n^{g,p}(b,c)=\widehat{\mathsf{F}}_n^{g,p}(b,c)-\lim_{c^+\downarrow c}\widehat{\mathsf{F}}_n^{g,p}(b,-c^+). \tag{D.8}$$

The asymptotic arguments are made on the probability scale $\psi = \mathsf{G}(c_\psi)$. When f is symmetric, the probability scales of G and F are related in a simple linear fashion, see (2.2), so that (D.8) translates into

$$\widehat{\mathsf{G}}_n^{g,p}\{b, \mathsf{G}^{-1}(\psi)\} = \widehat{\mathsf{F}}_n^{g,p}\{b, \mathsf{F}^{-1}(\frac{1+\psi}{2})\} - \lim_{\psi^+ \downarrow \psi} \widehat{\mathsf{F}}_n^{g,p}\{b, \mathsf{F}^{-1}(\frac{1-\psi^+}{2})\}. \tag{D.9}$$

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

$$\overline{\mathsf{G}}_n^{g,p}(b,c) = \frac{1}{n}\sum_{i=1}^n g_{in} \mathsf{E}_{i-1}\{\varepsilon_i^p 1_{(|\varepsilon_i - x_{in}' b| \le \sigma c)}\}, \qquad p = 0,1,2. \tag{D.10}$$

Form also the empirical process

$$\mathbb{G}_n^{g,p}(b,c) = n^{1/2}\{\widehat{\mathsf{G}}_n^{g,p}(b,c) - \overline{\mathsf{G}}_n^{g,p}(b,c)\}. \tag{D.11}$$

For later use note $\mathbb{G}_n^{1,0}(0,c) = \mathbb{G}_n(c)$. Note also that $\mathsf{E}_{i-1}\{\varepsilon_i^p 1_{(|\varepsilon_i| \le \sigma c)}\} = 0$ for odd p since 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 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 D.5. *For each ψ let $c_\psi = \mathsf{G}^{-1}(\psi)$. Suppose Assumption 3.1(ia, iib, iic) holds for some $0 \le \kappa < \eta \le 1/4$, but with $q_0 = 1 + 2^{r+1}$ only. Then, for all $B, \epsilon > 0$ and all $\omega < \eta - \kappa \le 1/4$,*

(a) $\sup_{0\le\psi\le1} \sup_{|b|,|d| \le n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{G}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) - \overline{\mathsf{G}}_n^{g,p}(0, c_\psi)\}$
$\quad -2\sigma^{p-1} c_\psi^p \mathsf{f}(c_\psi) n^{-1/2}\sum_{i=1}^n g_{in}\{1_{(p\ odd)} x_{in}'b + 1_{(p\ even)} n^{\kappa-1/2}\sigma d\}| = \mathrm{O}_{\mathsf{P}}\{n^{2(\kappa-\eta)}\}$*;*

(b) $\sup_{0\le\psi\le1} \sup_{|b|,|d| \le n^{1/4-\eta}B} |\mathbb{G}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{G}_n^{g,p}(0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(1)$*;*

(b') $\sup_{0\le\psi\le1} \sup_{|b|,|d| \le n^{1/4-\eta}B} |\mathbb{G}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \mathbb{G}_n^{1,0}(0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(n^{-1/8-\omega/2})$*;*

(c) $\lim_{\phi\downarrow0} \limsup_{n\to\infty} \mathsf{P}\{\sup_{0\le\psi\le\psi^\dagger\le1:\psi^\dagger-\psi\le\phi} |\mathbb{G}_n^{g,p}(0, c_{\psi^\dagger}) - \mathbb{G}_n^{g,p}(0, c_\psi)| > \epsilon\} \to 0.$

Proof of Lemma D.5. (a) Assumption 3.1(ia, iic) implies Assumption 4.1$(ib, iiib)$ with $r = 0$, $p \le 2$ 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 $\overline{\mathsf{F}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d)$. Thus, rewrite

$$\begin{aligned}\overline{\mathsf{F}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) &= n^{-1}\sum_{i=1}^n g_{in}\mathsf{E}_{i-1}\varepsilon_i^p 1_{\{\varepsilon_i - x_{in}'b \le \sigma(c_\psi + n^{\kappa-1/2}d)\}} \\ &= n^{-1}\sum_{i=1}^n g_{in}\mathsf{E}_{i-1}\varepsilon_i^p 1_{(\varepsilon_i - \overline{x}_{in}'\overline{b} \le \sigma c_\psi)},\end{aligned}$$

for $\overline{b} = (b', n^\kappa d)'$ and $\overline{x}_{in} = (x_{in}', n^{-1/2}\sigma)'$, where $|\overline{b}| \le 2n^{1/4+\kappa-\eta}B$ while $\overline{x}_{in}$ satisfies Assumption 4.1$(iiib)$ because $|\overline{x}_{in}|^2 = |x_{in}|^2 + n^{-1}\sigma^2$. Therefore we find, using that $\overline{\mathsf{G}}_n^{g,p}$ can be expressed in terms of $\overline{\mathsf{F}}_n^{g,p}$ as in (D.8), that $\sigma^{1-p}n^{1/2}\{\overline{\mathsf{G}}_n^{g,p}(b, c_\psi + n^{\kappa-1/2}d) -$

$\overline{\mathsf{G}}_n^{g,p}(0,c_\psi)\}$ has correction term

$$c_\psi^p \mathsf{f}(c_\psi) n^{-1} \textstyle\sum_{i=1}^n g_{in} n^{1/2}(x_{in}'b + n^{\kappa-1/2}\sigma d) \\ - (-c_\psi)^p \mathsf{f}(-c_\psi) n^{-1} \textstyle\sum_{i=1}^n g_{in} n^{1/2}(x_{in}'b - n^{\kappa-1/2}\sigma d) \\ = c_\psi^p \mathsf{f}(c_\psi) n^{-1/2} \textstyle\sum_{i=1}^n g_{in}[\{1-(-1)^p\}x_{in}'b + \{1+(-1)^p\} n^{\kappa-1/2}\sigma d],$$

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

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

The term $\mathcal{G}_1$ is $\mathrm{o}_\mathsf{P}(1)$ uniformly in $|b| \le n^{1/4-\eta}B$, $0 \le \psi \le 1$. To see this, expand $\mathbb{G}_n^{g,p}$ in a similar fashion to (D.8). Apply Theorem 4.1, noting that Assumption 3.1(ia, iib, iic) implies Assumption 4.1($i, ii, iiia$) with $p \le 2$, $g_{in} = 1, n^{1/2}x_{in}$ or $nx_{in}x_{in}'$ and the chosen r.

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

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

(c) Assumption 3.1(ia, iic) implies Assumption 4.1($ia, iiia$). Apply Theorem 4.4. □

D.4. A first analysis of the order statistics

The Forward Search is defined in terms of order statistics $\hat{z}^{(m)}$, see (D.4). A process version gives quantiles

$$\hat{c}_\psi^b = \inf\{c : \widehat{\mathsf{G}}_n^{1,0}(b,c) \ge \psi\}. \tag{D.12}$$

Setting $b = 0$ gives $\hat{c}_\psi^0 = \widehat{\mathsf{G}}_n^{-1}(\psi)$ as defined in (D.2) and studied in Lemma D.2. Evaluating the empirical distribution function at the quantile gives

$$\widehat{\mathsf{G}}_n^{1,0}(b, \hat{c}_\psi^b) = \frac{1}{n}\inf(x \in \mathbb{N}_0 : x \ge \psi n). \tag{D.13}$$

The first result gives an algebraic bound to the distance between $\hat{c}_\psi^b$ and $\hat{c}_\psi^0$. Probabilistic bounds follow.

Lemma D.6. *For all b, ψ, the quantiles $\hat{c}_\psi^b$ and $\hat{c}_\psi^0$ satisfy* $\sigma|\hat{c}_\psi^b - \hat{c}_\psi^0| < 2|b|\max_{1\le i\le n}|x_{in}|$.

Proof of Lemma D.6. 1. *A property of* $\widehat{\mathsf{G}}_n$. The quantile $\sigma\hat{c}_\psi^0$ is the left-continuous inverse of the right-continuous function $\widehat{\mathsf{G}}_n^{1,0}(0,c) = \widehat{\mathsf{G}}_n(c)$ in (D.2). Thus,

$$\widehat{\mathsf{G}}_n(y) < \widehat{\mathsf{G}}_n(\hat{c}_\psi^0) \le \widehat{\mathsf{G}}_n(z) \qquad \Rightarrow \qquad y < \hat{c}_\psi^0 \le z. \tag{D.14}$$

2. *A lower bound.* Let $x_{\max} = \max_{1\le i\le n}|x_{in}|$. Then it follows 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 $0 \le \psi \le 1$ and $z = \hat{c}_\psi^b + \sigma^{-1} x_{\max} |b|$,

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

Using (D.13) we find, for all b, ψ, that

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

which implies that $\sigma z = \sigma \hat{c}_\psi^b + x_{\max} |b| \ge \sigma \hat{c}_\psi^0$ by inequality (D.14).

3. *An upper bound.* For $y = \hat{c}_\psi^b - \sigma^{-1} 2 x_{\max} |b|$, we find

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

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

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

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

$$0 = \widehat{\mathsf{G}}_n^{1,0}(b, \hat{c}_\psi^b) - \widehat{\mathsf{G}}_n^{1,0}(0, \hat{c}_\psi^0) > \widehat{\mathsf{G}}_n(y) - \widehat{\mathsf{G}}_n(\hat{c}_\psi^0), \tag{D.15}$$

which implies that $y = \hat{c}_\psi^b - \sigma^{-1} 2 x_{\max} |b| < \hat{c}_\psi^0$ by inequality (D.14). □

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

Lemma D.7. *Suppose Assumption 3.1(ia, iib, iic) holds, but with $q_0 = 1 + 2^{r+1}$ only. Then, for all $\omega < \eta - \kappa$,*

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

Proof of Lemma D.7. If we combine Lemma D.6 with Assumption 3.1 (*iib*) we find that $\max_{1 \le i \le n} |x_{in}| = \mathrm{O}_{\mathsf{P}}(n^{\kappa - 1/2})$ to get that $\hat{c}_\psi^b - \hat{c}_\psi^0 = \mathrm{O}_{\mathsf{P}}(n^{-1/4+\kappa-\eta})$ for $|b| \le n^{1/4-\eta} B$. Thus, for any $\epsilon > 0$ a $C > 0$ exists so that the set $\mathcal{C}_n = \{|n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)| \le n^{1/4-\eta} C\}$ has probability $\mathsf{P}(\mathcal{C}_n) > 1 - \epsilon$. On this set it holds, using (D.13) and with $d = n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)$, that

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

Lemma D.5(a), using Assumption 3.1(ia, iic), shows that

$$n^{1/2}\{\overline{\mathsf{G}}_n^{1,0}(b, c_\psi + n^{\kappa-1/2}d) - \overline{\mathsf{G}}_n^{1,0}(0, c_\psi)\} - 2\sigma^{-1}\mathsf{f}(c_\psi)n^\kappa \sigma d = \mathrm{O}_\mathsf{P}(n^{2\kappa-2\eta}) = \mathrm{o}_\mathsf{P}(n^{-\omega}),$$

uniformly in $0 \leq \psi \leq 1$ and $|b|, |d| \leq n^{1/4-\eta}B$, for all $\omega < \eta - \kappa < 2(\eta - \kappa)$. Lemma D.5$(b')$ using Assumption 3.1(ia, iib, iic) shows that, uniformly in $0 \leq \psi \leq 1$ and $|b|, |d| \leq n^1/4 - \eta B$,

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

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

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

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, Section 15).

Lemma D.8. *Let* $c_\psi = \mathsf{G}^{-1}(\psi)$. *Suppose* f *is symmetric and decreasing for large* c *and that Assumption 3.1*(ib) *holds, but with* $q_0 = 1 + 2^{r+1}$ *only. Let* $|\psi^* - \psi| \leq |\mathsf{G}(\hat{c}_\psi^0) - \psi|$, *then*
(a) $\sup_{0\leq\psi\leq 1-c_n} |1-\mathsf{f}(c_\psi)/\mathsf{f}(c_{\psi^*})| = \mathrm{o}_\mathsf{P}(1)$, *for any sequence* $c_n \to 0$ *for which* $nc_n \to \infty$;
(b) $\sup_{0\leq\psi\leq n/(n+1)} |1 - \mathsf{f}(c_\psi)/\mathsf{f}(c_{\psi^*})| = \mathrm{O}_\mathsf{P}(1)$.

Proof of Lemma D.8. (a) By (2.2) $\mathsf{G}^{-1}(\psi) = \mathsf{F}^{-1}(y)$ for $y = (1 + \psi)/2$ varying in $1/2 \leq y \leq 1 - (2n + 2)^{-1}$. Let $\gamma = \sup_{c\in\mathbb{R}} \mathsf{F}(c)\{1 - \mathsf{F}(c)\}|\dot{\mathsf{f}}(c)|/\{\mathsf{f}(c)\}^2$ which is finite by Assumption 3.1(ib). It is first argued that for all $\epsilon > 0$ and $0 < c < 1$ and all n

$$\mathsf{P}\{\sup_{1/2+c\leq y\leq 1-c} |\frac{\mathsf{f}\{\mathsf{F}^{-1}(y)\}}{\mathsf{f}\{\mathsf{F}^{-1}(y^*)\}} - 1| > \epsilon\} \leq 4\{1 + \mathrm{int}(\gamma)\}\{\exp(-nch_1) + \exp(-nch_2)\}, \tag{D.16}$$

where, with $h(\lambda) = \lambda + \log(1/\lambda) - 1$,

$$h_1 = h[(1 + \epsilon)^{\{1+\mathrm{int}(\gamma)\}/2}],$$
$$h_2 = h[1/(1 + \epsilon)^{\{1+\mathrm{int}(\gamma)\}/2}].$$

This is nearly the statement of Theorem 1.5.1 of Csörgő (1983), which, however, has the denominator $\mathsf{f}(\theta_{y,n})$ instead of $\mathsf{f}\{\widehat{\mathsf{F}}_n^{-1}(y^*)\}$ where $\theta_{y,n}$ is a particular intermediate point between $\widehat{\mathsf{F}}_n^{-1}(y)$ and $\mathsf{F}^{-1}(y)$ rather than any intermediate point. Csörgő 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 $\widehat{\mathsf{F}}_n^{-1}(y)$ and $\mathsf{F}^{-1}(y)$ and not on the particular intermediate point $\theta_{y,n}$. This proves (D.16).

The inequality (D.16) implies that for any sequence $c_n \to 0$ for which $nc_n \to \infty$,

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

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 $\mathsf{G}(\hat{c}_\psi^0) - \psi = \mathrm{O}_\mathsf{P}(n^{-1/2})$ by Lemmas D.1, D.2, 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

$$\lim_{\epsilon\to\infty} \limsup_{n\to\infty} \mathsf{P}\left\{\sup_{1-c_n \le y \le 1-(2n+2)^{-1}} |\frac{\mathsf{f}\{\mathsf{F}^{-1}(y)\}}{\mathsf{f}\{\widehat{\mathsf{F}}_n^{-1}(y^*)\}} - 1| > \epsilon\right\} = 0. \tag{D.17}$$

Apply the inequality (D.16) with $c = (2n+2)^{-1}$ so that $nc \sim 1/2$. Then use that $h_1, h_2 \to \infty$ for $\epsilon \to \infty$ since $h(\lambda) \to \infty$ for $\lambda \to \infty$. □

The next result relates $\hat{c}_\psi^0$ to c_ψ.

Lemma D.9. *Suppose Assumption 3.1(ia, ib) holds with $q = 1$ only. Then*

$$\sup_{0\le\psi\le1} |(\hat{c}_\psi^0)^k \mathsf{f}(\hat{c}_\psi^0) - (c_\psi)^k \mathsf{f}(c_\psi)| = \mathrm{o}_\mathsf{P}(1) \qquad \textit{for } k = 0, 1.$$

Proof of Lemma D.9. 1. *Consider ψ so that $0 \le \psi \le 1 - 1/z_n$ for any sequence* $0 < z_n < \mathrm{o}(n^{1/2})$. Rewrite the process of interest as

$$(\hat{c}_\psi^0)^k \mathsf{f}(\hat{c}_\psi^0) - (c_\psi)^k \mathsf{f}(c_\psi) = \{(\hat{c}_\psi^0)^k - (c_\psi)^k\}\mathsf{f}(c_\psi) + (\hat{c}_\psi^0)^k \mathsf{f}(\hat{c}_\psi^0)\{1 - \frac{\mathsf{f}(c_\psi)}{\mathsf{f}(\hat{c}_\psi^0)}\}. \tag{D.18}$$

The first term is zero for $k = 0$. For $k = 1$, $n^{1/2}(\hat{c}_\psi^0 - c_\psi)\mathsf{f}(c_\psi) = -\widehat{\mathbb{G}}_n^{1,0}(0, c_\psi) + \mathrm{o}_\mathsf{P}(1)$ by Lemma D.2(a) using Assumption 3.1(ib). This in turn is tight due to Lemma D.1. Overall, the first term is $\mathrm{O}_\mathsf{P}(n^{-1/2})$. For the second term, note that $(\hat{c}_\psi^0)^k \mathsf{f}(\hat{c}_\psi^0)$ is bounded uniformly in $0 \le \psi \le 1$ due to Assumption 3.1(ia) with $q = 1$, while $1 - \mathsf{f}(c_\psi)/\mathsf{f}(\hat{c}_\psi^0)$ vanishes by Lemma D.8(a) using Assumption 3.1(ib).

2. *Consider ψ so that $\psi_n \le \psi \le 1$ for any sequence $\psi_n \to 1$.* Assumption 3.1(ia) and the continuity of f implies that $(c_\psi)^k \mathsf{f}(c_\psi)$ is continuous and convergent for $\psi \to 1$, and hence for $c_\psi \to G^{-1}(1)$. Rewrite

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

where $g_n = \sup_{0\le\psi\le1}\{\mathsf{G}(\hat{c}_\psi^0) - \psi\} = \mathrm{O}_\mathsf{P}(n^{-1/2})$ due to Lemmas D.1, D.2(c) using Assumption 3.1(ib). By the continuity of G^{-1}, $\hat{c}_{\psi_n}^0 \to \mathsf{G}^{-1}(1)$ in probability and therefore $(\hat{c}_\psi^0)^k \mathsf{f}(\hat{c}_\psi^0)$ and $(c_\psi)^k \mathsf{f}(c_\psi)$ converge to the same limit in probability, and their difference vanishes. □

D.5. A one-step result for the least squares estimator

A one-step result for the least squares estimator now follows. Equation (D.6) represents the one-step least squares estimator $\hat{\beta}^{(m+1)}$ in terms of $\widehat{\mathsf{G}}_n^{g,p}$. 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 $\hat{c}_\psi^b$ defined in (D.12) gives the following asymptotic uniform linearization result.

Lemma D.10. *Let* $c_\psi = \mathsf{G}^{-1}(\psi)$ *and*

$$\rho_\psi = 2c_\psi \mathsf{f}(c_\psi)/\psi. \tag{D.19}$$

Suppose Assumption 3.1(ia, ib, ii) holds, while $\psi_0 > 0$ *and* $\eta \leq 1/4$*, but with* $q_0 = 1+2^{r+1}$ *only. Then*

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

(b) $\sup_{0\leq\psi\leq 1, |b|\leq n^{1/4-\eta}B} |n^{1/2}\{\widehat{\mathsf{G}}_n^{xx,0}(b,\hat{c}_\psi^b) - \Sigma_n\psi\}| = \mathrm{O}_\mathsf{P}(1)$;

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

Proof of Lemma D.10. (a) The inequality of Lemma D.6 implies that

$$\sup_{0\leq\psi\leq 1} \sup_{|b|\leq n^{1/4-\eta}B} n^{1/2-\kappa}|\hat{c}_\psi^b - \hat{c}_\psi^0| = \mathrm{O}_\mathsf{P}(n^{1/4-\eta}), \tag{D.20}$$

since $\max_{1\leq i\leq n}|x_{in}| = \mathrm{O}_\mathsf{P}(n^{\kappa-1/2})$ by Assumption 3.1 (iib), where $0\leq\kappa<\eta\leq 1/4$. By definition

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

Lemma D.5(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 $0\leq\psi\leq 1$,

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

Note that $\overline{\mathsf{G}}_n^{x,1}(0,c_\psi) = 0$ due to the symmetry of f. Replace c_ψ by $\hat{c}_\psi^0$ and d by $n^{1/2-\kappa}(\hat{c}_\psi^b - \hat{c}_\psi^0)$, which is $\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$ due to (D.20). Thus, it holds on a set with large probability that

$$n^{1/2}\widehat{\mathsf{G}}_n^{x,1}(b,\hat{c}_\psi^b) = \mathbb{G}_n^{x,1}(0,\hat{c}_\psi^0) + 2\hat{c}_\psi^0\mathsf{f}(\hat{c}_\psi^0)\Sigma_n b + \mathrm{o}_\mathsf{P}(1), \tag{D.21}$$

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

First term. Let $a_\psi = n^{1/2}\{\mathsf{G}(\hat{c}_\psi^0) - \psi\}$. Theorem D.2$(c)$ using Assumption 3.1(ib) shows that $a_\psi = -\mathbb{G}_n(c_\psi) + \mathrm{o}_\mathsf{P}(1)$ uniformly in $0\leq\psi\leq 1$, which in turn is tight due to Lemma D.1. Expand

$$\hat{c}_\psi^0 = \mathsf{G}^{-1}\{\mathsf{G}(\hat{c}_\psi^0)\} = c_{\mathsf{G}(\hat{c}_\psi^0)} = c_{\psi+n^{-1/2}a_\psi}. \tag{D.22}$$

Lemma D.5(c) using Assumption 3.1(ia, iib, iic) shows $\mathbb{G}_n^{x,1}(0, \hat{c}_\psi^0) = \mathbb{G}_n^{x,1}(0, c_\psi) + \mathrm{o}_\mathsf{P}(1)$.

Second term. Use that $\hat{c}_\psi^0 \mathsf{f}(\hat{c}_\psi^0) = c_\psi \mathsf{f}(c_\psi) + \mathrm{o}_\mathsf{P}(1)$ uniformly in ψ by Lemma D.9 using Assumption 3.1(ia, ib).

(b) An expansion as in (D.21) gives

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

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

First term. This is $n^{-1/2}\mathbb{G}_n^{xx,0}(0, \hat{c}_\psi^0) = n^{-1/2}\mathbb{G}_n^{xx,0}(0, c_\psi) + \mathrm{o}_\mathsf{P}(n^{-1/2})$ by an argument as for the first term of (D.21).

Second term. Use that $\Sigma_n = \sum_{i=1}^n x_{in} x_{in}'$ is tight by Assumption 3.1(iia), while $\mathsf{G}(\hat{c}_\psi^0) = \psi + \mathrm{O}_\mathsf{P}(n^{-1/2})$ uniformly in ψ by Lemma D.1, D.2(c) using Assumption 3.1(ib). Thus,

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

Third term. This is $\mathrm{o}_\mathsf{P}(n^{-1/2})$ since $\mathsf{f}(\hat{c}_\psi^0)(\hat{c}_\psi^b - \hat{c}_\psi^0) = \mathrm{o}_\mathsf{P}(n^{-1/2})$ uniformly in ψ, b by Lemma D.7 using Assumption 3.1(ia, iia, iib), while Σ_n is tight by Assumption 3.1(iia).

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

$$\widehat{\mathsf{G}}_n^{xx,0}(b, \hat{c}_\psi^b) = \psi \Sigma_n \{1 + \mathrm{o}_\mathsf{P}(1)\},$$

for $\psi \ge \psi_0 > 0$ since $\Sigma_n \to \Sigma$ in distribution where $\Sigma > 0$ *a.s.* by Assumption 3.1(iia). Combine with the expression for the numerator in (a). □

For the variance estimator, expansions of the same type are needed.

Lemma D.11. *Suppose Assumption 3.1*(ia, ib, id, ii) *holds while* $\psi_0 > 0$ *and* $\eta \le 1/4$. *Then*
(a) $\sup_{\psi_0 \le \psi \le 1, |b| \le n^{1/4-\eta}B} |\{\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b)\}'\{\widehat{\mathsf{G}}_n^{xx,0}(b, \hat{c}_\psi^b)\}^{-1}\{\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b)\}| = \mathrm{O}_\mathsf{P}(n^{-1/2-2\eta})$;
(b) $\sup_{\psi_0 \le \psi \le n/(n+1), |b| \le n^{1/4-\eta}B} |n^{1/2}\{\widehat{\mathsf{G}}_n^{1,2}(b, \hat{c}_\psi^b) - \tau_\psi \sigma^2\} - \mathbb{G}_n^{1,2}(0, c_\psi) + \sigma^2 c_\psi^2 \mathbb{G}_n^{1,0}(c_\psi)| = \mathrm{o}_\mathsf{P}(1)$.

Proof of Lemma D.11. (a) Lemma D.10(a, c) using Assumption 3.1(ia, ib, ii) shows

$$n^{1/2}\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b) = \mathbb{G}_n^{x,1}(0, c_\psi) + 2c_\psi \mathsf{f}(c_\psi)\Sigma_n b + \mathrm{o}_\mathsf{P}(1), \tag{D.23}$$

$$\{\widehat{\mathsf{G}}_n^{xx,0}(b, \hat{c}_\psi^b)\}^{-1} n^{1/2}\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b) = (\Sigma_n \psi)^{-1}\mathbb{G}_n^{x,1}(0, c_\psi) - \rho_\psi b + \mathrm{o}_\mathsf{P}(1), \tag{D.24}$$

uniformly in $|b| \le n^{1/4-\eta}B$, $\psi_0 \le \psi \le 1$ for $\eta \le 1/4$. Because $\mathbb{G}_n^{x,1}(0, c_\psi)$ is tight by Lemma D.5(c) using Assumption 3.1(ia, iib, iic), since $\Sigma_n \to \Sigma$ in distribution where $\Sigma > 0$ *a.s.* by Assumption 3.1(iia) and since $|b| \le n^{1/4-\eta}B$, then both $\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b)$,see

(D.23), and $\{\widehat{\mathsf{G}}_n^{xx,0}(b,\hat{c}_\psi^b)\}^{-1}\widehat{\mathsf{G}}_n^{x,1}(b,\hat{c}_\psi^b)$, see (D.24), are $\mathrm{O}_\mathsf{P}(n^{-1/4-\eta})$. Thus, their product is $\mathrm{O}_\mathsf{P}(n^{-1/2-2\eta})$ as desired.

(b) The argument relates to that of the proof of Lemma D.10.

1. *Expansion.* By definition

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

Apply Lemma D.5(a,b) using Assumption 3.1(ia,iib,iic) to get

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

uniformly in $|b|,|d|\le n^{1/4-\eta}B$, $0\le\psi\le1$. Combine the first two terms to get

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

Replace c_ψ by $\hat{c}_\psi^0$. Since $n^{1/2-\kappa}(\hat{c}_\psi^b-\hat{c}_\psi^0)=\mathrm{O}_\mathsf{P}(n^{1/4-\eta})$ uniformly in $0\le\psi\le1$, $|b|\le n^{1/4-\eta}B$ by (D.20), we can replace $n^\kappa d$ by $n^{1/2}(\hat{c}_\psi^b-\hat{c}_\psi^0)$ on a set with large probability. Subtract $n^{1/2}\tau_\psi\sigma^2$ on both sides. Add and subtract $n^{1/2}\tau_{\mathsf{G}(\hat{c}_\psi^0)}\sigma^2$ on the right. Altogether we get

$$\begin{aligned}n^{1/2}\{\widehat{\mathsf{G}}_n^{1,2}(b,\hat{c}_\psi^b)-\tau_\psi\sigma^2\}&=n^{1/2}\{\widehat{\mathsf{G}}_n^{1,2}(0,\hat{c}_\psi^0)-\sigma^2\tau_{\mathsf{G}(\hat{c}_\psi^0)}\}\\&\quad+2\sigma^2(\hat{c}_\psi^0)^2\mathsf{f}(\hat{c}_\psi^0)n^{1/2}(\hat{c}_\psi^b-\hat{c}_\psi^0)+\sigma^2n^{1/2}\{\tau_{\mathsf{G}(\hat{c}_\psi^0)}-\tau_\psi\}+\mathrm{o}_\mathsf{P}(1),\end{aligned}\qquad\text{(D.25)}$$

uniformly in $0\le\psi\le1$, $|b|\le n^{1/4-\eta}B$. The three terms are analysed in turn.

2. *First term of* (D.25). Since $\overline{\mathsf{G}}_n^{1,2}(0,c)=\sigma^2\tau_{\mathsf{G}(c)}$, the first term equals $\mathbb{G}_n^{1,2}(0,\hat{c}_\psi^0)$. Lemmas D.1, D.2($c$) show that $\hat{c}_\psi^0=c_{\psi+n^{-1/2}\phi}$ where $\phi=n^{1/2}\{\mathsf{G}(\hat{c}_\psi^0)-\psi\}=\mathbb{G}_n(c_\psi)+\mathrm{o}_\mathsf{P}(1)$ is tight. Tightness of $\mathbb{G}_n^{1,2}$ was established in Lemma D.5(c) under the Assumption 3.1(ia,iib,iic), then implies that the first term equals $\mathbb{G}_n^{1,2}(0,c_\psi)+\mathrm{o}_\mathsf{P}(1)$ uniformly in $0\le\psi\le1$.

3. *The order of* $\hat{c}_\psi^0$ is $\mathrm{o}_\mathsf{P}(n^{\nu/2})$ for some $\nu<\eta-\kappa\le1/4$. The reason is that $\hat{c}_\psi^0\le\max_{i\le n}|\varepsilon_i|$, that $\mathsf{E}|\varepsilon_i|^q<\infty$ for some $q>2/(\eta-\kappa)$ by Assumption 3.1(ia), so that $q(\eta-\kappa)/2>1+\epsilon$ for some $\epsilon>0$. Thus, Boole's and Markov's inequalities imply that $\mathsf{P}(\max_i|\varepsilon_i|>Cn^{\nu/2})\le\sum_{i=1}^n\mathsf{P}(|\varepsilon_i|>Cn^{\nu/2})\le n(Cn^{\nu/2})^{-q}\mathsf{E}|\varepsilon_i|^q$ vanishes if $\nu=(\eta-\kappa)/(1+\epsilon)$.

4. *The order of* c_ψ^2 is $\mathrm{o}(n^{1/4-2\lambda})$ for some $\lambda>0$ when $\psi\le1-n^{-1}$. Because $\mathsf{E}|\varepsilon_i|^q<\infty$ for some $q>8$ by Assumption 3.1(ia), $\mathsf{P}(|\varepsilon_i|>\sigma c_\psi)\le c_\psi^{-q}\mathsf{E}(|\varepsilon_i/\sigma|^q)$ by the Markov inequality. Thus, $c_\psi^2=\mathrm{O}\{(1-\psi)^{-2/q}\}$. In particular, for $\psi\le1-n^{-1}$, $c_\psi^2=\mathrm{O}(n^{2/q})=\mathrm{o}(n^{1/4-2\lambda})$ for $1/4-2\lambda>2/q$ so that $\lambda<(q-8)/(8q)$.

5. *Second term of* (D.25) vanishes. Indeed, $\mathsf{f}(\hat{c}_\psi^0)n^{1/2}(\hat{c}_\psi^b-\hat{c}_\psi^0)=\mathrm{o}_\mathsf{P}(n^{-\omega})$ for all $\omega<\eta-\kappa$ uniformly in $0\le\psi\le1$, $b\le n^{1/4-\eta}B$ by Lemma D.7 using Assumption 3.1(ia,iib,iic). By item 3 then $(\hat{c}_\psi^0)^2=\mathrm{o}_\mathsf{P}(n^\nu)$ for some $\nu<\eta-\kappa$ and an ω exists so $\nu<\omega$.

6. *Third term of* (D.25). We will argue that

$$n^{1/2}(\tau_{\psi+n^{-1/2}\hat{\phi}} - \tau_\psi) - c_\psi^2\hat{\phi} = \mathsf{o_P}(1), \tag{D.26}$$

for $\psi_0 \leq \psi \leq n/(n+1)$ and $\hat{\phi} = -n^{1/2}\{\mathsf{G}(\hat{c}_\psi^0) - \psi\}$. This suffices since Lemmas D.1, D.2(c) using Assumption 3.1(c) show

$$\hat{\phi} = -\mathbb{G}_n(c_\psi) + \mathsf{o_P}(n^{\zeta-1/4}), \tag{D.27}$$

for all $\zeta > 0$ while item 4 shows $c_\psi^2 = \mathrm{o}(n^{1/4-2\lambda})$ for some $\lambda > 0$. This implies

$$n^{1/2}\{\tau_{\mathsf{G}(\hat{c}_\psi^0)} - \tau_\psi\} + c_\psi^2\mathbb{G}_n(c_\psi) = \mathrm{o}(n^{\zeta-2\lambda}) + \mathsf{o_P}(1) = \mathsf{o_P}(1),$$

as desired. To prove (D.26) write

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

Change variable $y = \mathsf{G}(u)$, $dy = 2\mathsf{f}(u)du$, and Taylor expand to get

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

for some ψ^* so $|\psi^* - \psi| \leq \phi$. Rewrite this, for some $\upsilon > 0$ yet to be chosen,

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

The first component is

$$\{\psi(1-\psi)\}^{-2\upsilon} = \mathrm{O}(n^{2\upsilon}), \tag{D.28}$$

for $\psi_0 \leq \psi \leq n/(n+1)$. The second component is $\psi(1-\psi)/\mathsf{f}(c_\psi) = \mathrm{O}(c_\psi)$ by Assumption 3.1(id). Since $c_\psi = \mathrm{o}(n^{1/8-\lambda})$ for some $\lambda > 0$ due to item 4, then

$$\mathcal{S}_3 = (c_{\psi^*} + c_\psi)\left[\frac{\phi}{\{\psi(1-\psi)\}^{1/2-\upsilon}}\right]\left[\frac{\mathsf{f}(c_\psi)n^{1/2}(c_{\psi^*} - c_\psi)}{\{\psi(1-\psi)\}^{1/2-\upsilon}}\right]\mathrm{O}_\mathsf{P}(n^{2\upsilon+1/8-\lambda-1/2}).$$

Evaluate this expression for ϕ replaced by $\hat{\phi}$. The first term is $c_\psi^* + c_\psi \leq \hat{c}_\psi^0 + 2c_\psi = \mathsf{o_P}(n^{1/8})$ due to items 3,4 so that

$$\mathcal{S}_3 = \left[\frac{\hat{\phi}}{\{\psi(1-\psi)\}^{1/2-\upsilon}}\right]\left[\frac{\mathsf{f}(c_\psi)n^{1/2}(c_{\psi^*} - c_\psi)}{\{\psi(1-\psi)\}^{1/2-\upsilon}}\right]\mathsf{o_P}(n^{2\upsilon-\lambda-1/4}).$$

The first component is $\{\mathbb{G}_n(c_\psi) + \mathsf{o_P}(n^{\zeta-1/4})\}/\{\psi(1-\psi)\}^{1/2-\upsilon}$ by (D.27). The first normalised summand is $\mathrm{O}_\mathsf{P}(1)$ by the Hölder continuity of Lemma D.4. The second summand is $\mathsf{o_P}(n^{\zeta-1/4})\mathsf{o_P}(n^{1/2-\upsilon})$ for $\psi_0 \leq \psi \leq n/(n+1)$ as in (D.28). Thus,

$$\mathcal{S}_3 = \frac{\mathsf{f}(c_\psi)n^{1/2}(c_{\psi^*} - c_\psi)}{\{\psi(1-\psi)\}^{1/2-\upsilon}}\mathsf{o_P}(n^{\upsilon+\zeta-\lambda}).$$

For the first component note $|c^*_\psi - c_\psi| \leq |c_\psi + n^{-1/2}\phi - c_\psi| = |c_{\hat{\psi}} - c_\psi|$. Lemma D.3($b$) using Assumption 3.1($ib$) then implies that a sequence of Brownian bridges $\mathbb{B}_n$ exists so that the first component is bounded by $\mathsf{o}_\mathsf{P}(1) + |\mathbb{B}_n(\psi)|/\{\psi(1-\psi)\}^{1/2-\upsilon}$ uniformly in $\psi_0 \leq \psi \leq n/(n+1)$. This in turn is $\mathsf{O}_\mathsf{P}(1)$ by the Hölder continuity of Lemma D.4. Overall it follows that $\mathcal{S}_3 = \mathsf{o}_\mathsf{P}(n^{\upsilon+\zeta-\lambda}) = \mathsf{o}_\mathsf{P}(1)$ since we can choose $\upsilon + \zeta < \lambda$. □

D.6. The forward plot of least squares estimators

The Forward Plot of least squares estimators is now considered. The one-step result in Lemma D.10 implies that the Forward Search iteration can be viewed as a fixed point problem. Indeed, the one-step result in Lemma D.10 implies an autoregressive relation between the one-step updated estimation error $\hat{b}^{(m+1)}$ and the previous estimation error $\hat{b}^{(m)}$. That is,

$$\hat{b}^{(m+1)} = \rho_\psi \hat{b}^{(m)} + (\psi\Sigma_n)^{-1}\mathbb{G}^{x,1}_n(0, c_\psi) + e_\psi(\hat{b}^{(m)}), \tag{D.29}$$

for $\psi = m/n + \mathrm{o}(1)$, an 'autoregressive coefficient' ρ_ψ defined in (D.19) 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, $\hat{b}^{(m)} = \mathrm{O}(1)$.

It is first established that ρ_ψ has nice properties for unimodal densities f.

Lemma D.12. *Suppose Assumption 3.1*(ia, ic) *holds. Then* $\rho_\psi = 2c_\psi \mathsf{f}(c_\psi)/\psi$ *satisfies*
(a) $\rho_\psi > 0$ *for all* $\psi > 0$*;*
(b) $\sup_{\psi_0 \leq \psi < 1} \rho_\psi < 1$ *for all* $\psi_0 > 0$.

Proof of Lemma D.12. (a) holds because $\mathsf{f}(c_\psi) > 0$ for $0 < \psi < 1$.

(b) If the conclusion were incorrect, there would exist a sequence ψ_n so that $\rho_{\psi_n} \to 1$ for $n \to \infty$. Let $\psi^\dagger$ be a limit point. We consider the cases where $\psi^\dagger < 1$ and $\psi^\dagger = 1$.

Suppose $\psi^\dagger < 1$. Then $\rho_{\psi^\dagger} = 1$, which implies $2c_{\psi^\dagger}\mathsf{f}(c_{\psi^\dagger}) = \psi^\dagger$. Since $\psi^\dagger = 2\int_0^{c_{\psi^\dagger}} \mathsf{f}(x)dx$ it holds that $\int_0^{c_{\psi^\dagger}} \{\mathsf{f}(x) - \mathsf{f}(c_{\psi^\dagger})\}dx = 0$. This contradicts Assumption 3.1(ic).

Suppose $\psi^\dagger = 1$. Because $\psi_n \to 1$, it must hold in this case that $c_{\psi_n}\mathsf{f}(c_{\psi_n}) \to 1$ for $n \to 0$. This contradicts that $c\mathsf{f}(c) \to 0$ for $c \to \infty$ by Assumption 3.1(ia). □

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 second, 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 (2013, Theorem 3.3).

Theorem D.13. *Suppose Assumption 3.1*($ia - ic, ii, iii$) *holds, but with* $q_0 = 1 + 2^{r+1}$ *only. Then, for all* $\psi_1 > \psi_0 > 0$ *and* $m_0/n = \psi_0 + \mathrm{o}(1)$*, the estimator* $\hat{\beta}_\psi$ *satisfies*

(a) $\sup_{\psi_0\le\psi\le1}|N^{-1}(\hat\beta_\psi-\beta)|=\mathsf{O}_\mathsf{P}(n^{1/4-\eta})$;
(b) $\sup_{\psi_1\le\psi\le1}|N^{-1}(\hat\beta_\psi-\beta)|=\mathsf{O}_\mathsf{P}(1)$.

Proof of Lemma D.13. Due to the embedding (3.1), 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 (D.29) recursively to get

$$\hat b^{(m+1)}=\textstyle\sum_{k=m_0}^m(\prod_{\ell=k+1}^m\rho_{\ell/n})\{\dfrac{n}{k}K_{k/n}^n+e_{k/n}(\hat b^{(k)})\}+(\prod_{k=m_0}^m\rho_{k/n})\hat b^{(m_0)},$$

with the convention that an empty product equals unity. Lemma D.12 using Assumption 3.1(ia,ic) shows that $\rho_\psi\le\rho_0$ for some $\rho_0<1$ for $\psi\ge\psi_0$, and therefore $\sum_{k=m_0}^m\rho_0^{m-k}\le\sum_{k=0}^\infty\rho_0^k=C$. This gives the bound

$$|\hat b^{(m+1)}|\le C\{\sup_{\psi_0\le\psi\le1}|\psi^{-1}K_\psi^n|+\max_{m_0\le k\le m}|e_{k/n}(\hat b^{(k)})|\}+\rho_0^{m-m_0+1}|\hat b^{(m_0)}|. \tag{D.30}$$

In this expression, the process $\psi^{-1}K_\psi^n$ in $D[\psi_0,1]$ for $\psi_0>0$, is tight by Lemma D.5(c) using Assumption 3.1(ia,iib,iic). Therefore, for any $\epsilon>0$ we first choose B so large that $\mathsf{P}(C\sup_{\psi_0\le\psi\le1}|\psi^{-1}K_\psi^n|\ge B)\le\epsilon/3$ for all n. The initial estimator is $\hat b^{(m_0)}=\mathsf{O}_\mathsf{P}(n^{1/4-\eta})$ by Assumption 3.1(iii), and we next choose B so large that $\mathsf{P}(|\hat b^{(m_0)}|\ge Bn^{1/4-\eta})\le\epsilon/3$ for all n. Finally, by Lemma D.10(c), $\sup_{\psi_0\le\psi\le1}\sup_{|b|\le3n^{1/4-\eta}B}|e_\psi(b)|=\mathsf{o}_\mathsf{P}(1)$, using Assumption 3.1$(ia,ib,ii)$. Thus, there is an n_0 such that

$$\mathsf{P}(C\sup_{\psi_0\le\psi\le1}\sup_{|b|\le3n^{1/4-\eta}B}|e_\psi(b)|\ge B)\le\epsilon/3,$$

for $n\ge n_0$. This implies that the set

$$\mathcal{A}_n=(C\sup_{\psi_0\le\psi\le1}|\psi^{-1}K_\psi^n|\le B)\cap(C\sup_{\psi_0\le\psi\le1}\sup_{|b|\le3n^{1/4-\eta}B}|e_\psi(b)|\le B)\cap(|\hat b^{(m_0)}|\le n^{1/4-\eta}B)$$

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

$$\max_{m_0\le k\le m}|\hat b^{(k)}|\le3n^{1/4-\eta}B\qquad\text{for } m=m_0,\dots,n,$$

on the set $\mathcal{A}_n$, which implies the desired result. For $m=m_0$, the initial estimator satisfies $|\hat b^{(m_0)}|\le n^{1/4-\eta}B$ on the set $\mathcal{A}_n$. Suppose the result holds for some m. This implies that

$$C\sup_{\psi_0\le\psi\le1}\max_{m_0\le k\le m}|e_\psi(\hat b^{(k)})|\le B \tag{D.31}$$

on the set $\mathcal{A}_n$. Thus, the bound (D.30) becomes

$$|\hat b^{(m+1)}|\le B+B+n^{1/4-\eta}B\le3n^{1/4-\eta}B,$$

because $n^{1/4-\eta}\ge1$ for $\eta\le1/4$. Thus, the result holds for $m+1$, which completes the induction.

(b) Consider next (D.30) for $\psi_1 n \leq m \leq n$. Here, $\sum_{k=0}^{n}\rho_0^k \leq C$, the first term is $\sup_{\psi_1\leq\psi\leq 1}|\psi^{-1}K_\psi^n| = \mathsf{O}_\mathsf{P}(1)$ due to tightness, while the second, as remarked above, is $\sup_{\psi_1\leq\psi\leq 1}\max_{m_0\leq k<n}|e_\psi(\hat{b}^{(k)})| = \mathsf{o}_\mathsf{P}(1)$. Because $\rho_0^{m-m_0} \leq \rho_0^{\mathrm{int}(\psi_1 n)-\mathrm{int}(\psi_0 n)}$ declines exponentially, $\rho_0^{m-m_0} < n^{-1/4}$ for large n and therefore the last term is $\max_{m\geq m_1}\rho_0^{m-m_0}|\hat{b}^{(m_0)}| = \mathsf{O}_\mathsf{P}(n^{-1/4+1/4-\eta}) = \mathsf{o}_\mathsf{P}(1)$, which proves ($b$). □

D.7. Proofs of main Theorems 3.1-3.7

Lemmas D.2, D.6 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 stated with slightly weaker conditions.

Remark D.1. *The proof below of Theorem 3.1 only requires Assumption 3.1($ia-ic, ii, iii$) with $q_0 = 1+2^{r+1}$.*

Proof of Theorem 3.1. 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 $\mathsf{P}(\mathcal{C}_n) \geq 1-\epsilon$. This follows from Lemma D.13 using Assumption 3.1($ia-ic, ii, iii$). Now, on $\mathcal{C}_n$ it holds that $\sigma^{-1}\hat{z}_\psi = \hat{c}_\psi^b$, see (D.4), for some $|b| \leq n^{1/4-\eta}B$. Thus it suffices to show that

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

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

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

The first term is $\mathsf{o}_\mathsf{P}(n^{\zeta-1/4})$ for all $\zeta > 0$ uniformly in $0 \leq \psi \leq 1$ by Lemma D.2(a) using Assumption 3.1(ib). In the second term, the ratio $\mathsf{f}(c_\psi)/\mathsf{f}(\hat{c}_\psi^0)$ is $\mathsf{O}_\mathsf{P}(1)$ uniformly in $0 \leq \psi \leq n/(n+1)$ by Lemma D.8 using Assumption 3.1(ia, ib), while $n^{1/2}\mathsf{f}(\hat{c}_\psi^0)(\hat{c}_\psi^b - \hat{c}_\psi^0) = \mathsf{o}_\mathsf{P}(n^{-\omega})$ uniformly in $0 \leq \psi \leq 1$ by Lemma D.7 using Assumption 3.1(ia, iib, iic). Combining the first statement with Lemma D.2(a) gives the second statement. □

Remark D.2. *The proof below of Theorem 3.2 only requires Assumption 3.1(ia, ib, id, ii).*

Proof of Theorem 3.2. The above theory for $\sigma^{-1}\hat{z}_\psi$ involves the population variance σ^2. The result gives an asymptotic expansion for $\hat{\sigma}^2_{\psi,cor}$, recalling, from (2.7), (2.8), (D.3) that

$$\begin{aligned} & n^{1/2}(\hat{\sigma}^2_{\psi,cor} - \sigma^2) \\ &= \frac{1}{\tau_\psi} n^{1/2}[\{\widehat{\mathsf{G}}^{1,2}_n(\hat{b}, \hat{c}^{\hat{b}}_\psi) - \tau_\psi\sigma^2\} - \{\widehat{\mathsf{G}}^{x,1}_n(\hat{b}, \hat{c}^{\hat{b}}_\psi)\}'\{\widehat{\mathsf{G}}^{xx,0}_n(\hat{b}, \hat{c}^{\hat{b}}_\psi)\}^{-1}\{\widehat{\mathsf{G}}^{x,1}_n(\hat{b}, \hat{c}^{\hat{b}}_\psi)\}]. \end{aligned} \tag{D.32}$$

Compare also the definitions in (3.2), (3.3) with (D.11) to see

$$\mathbb{G}_n(c_\psi) = \mathbb{G}^{1,0}_n(0, c_\psi), \qquad \tau_\psi\mathbb{L}_n(c_\psi) = \sigma^{-2}\mathbb{G}^{1,2}_n(0, c_\psi) - c^2_\psi\mathbb{G}^{1,0}_n(0, c_\psi). \tag{D.33}$$

Lemma D.11 using Assumption 3.1(ia, ib, id, ii) shows the first term in (D.32) equals the leading term $\mathbb{L}_n(c_\psi) + \mathsf{o}_\mathsf{P}(1)$ uniformly in $\psi_0 \leq \psi \leq n/(n+1)$ while the second term in (D.32) vanishes. □

Proof of Theorem 3.3. Note the identity

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

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

Proof of Theorem 3.4. We first show that $\widehat{d}^{(m)} \leq \widehat{z}^{(m)}$ and then we find an upper bound for $\widehat{z}^{(m)} - \widehat{d}^{(m)}$, and finally show that the difference is small.

1. *Inequality* $\widehat{d}^{(m)} \leq \widehat{z}^{(m)}$. Indeed, if $S^{(m)}$ is the ranks of $\widehat{\xi}^{(m)}_{(1)}, \ldots, \widehat{\xi}^{(m)}_{(m)}$ then $\widehat{d}^{(m)} = \widehat{z}^{(m)}$. If $S^{(m)}$ does not have this form, then its complement must include one of the ranks of $\widehat{\xi}^{(m)}_{(1)}, \ldots, \widehat{\xi}^{(m)}_{(m)}$, for instance that of $i^\dagger$. In that situation $\widehat{d}^{(m)} \leq \widehat{\xi}^{(m)}_{i^\dagger} \leq \widehat{\xi}^{(m)}_{(m)} \leq \widehat{\xi}^{(m)}_{(m+1)} = \widehat{z}^{(m)}$.

2. *The set* $S^{(m)}$ *consists* of the ranks of $\widehat{\xi}^{(m-1)}_{(1)}, \ldots, \widehat{\xi}^{(m-1)}_{(m)}$. It follows that for all $i \notin S^{(m)}$ then $\widehat{\xi}^{(m-1)}_i \geq \widehat{\xi}^{(m-1)}_{(m+1)} \geq \widehat{\xi}^{(m-1)}_{(m)} = \widehat{z}^{(m-1)}$.

3. *Inequality for deletion residual.* The absolute residual for observation i based on the set $S^{(m)}$, $\xi^{(m-1)}_i$ in step $m-1$, satisfies

$$\begin{aligned} \hat{\xi}^{(m-1)}_i = |y_i - x'_i\widehat{\beta}^{(m-1)}| &\leq |y_i - x'_i\widehat{\beta}^{(m)}| + |x'_i(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})| \\ &\leq \hat{\xi}^{(m)}_i + \max_{1\leq i\leq n} |N'x_i||N^{-1}(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})|. \end{aligned}$$

For $i \notin S^{(m)}$ we have from item 2 that $\xi_i^{(m-1)} \geq \hat{\xi}_{(m)}^{(m-1)} = \hat{z}^{(m-1)}$ and $\widehat{d}^{(m)} = \min_{i \notin S^{(m)}} \widehat{\xi}_i^{(m)}$ giving

$$\hat{z}^{(m-1)} \leq \widehat{d}^{(m)} + \max_{1 \leq i \leq n} |N' x_i| |N^{-1}(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})|,$$

and therefore, using $\widehat{d}^{(m)} \leq \widehat{z}^{(m)}$ we find

$$0 \leq \widehat{z}^{(m)} - \widehat{d}^{(m)} \leq \widehat{z}^{(m)} - \widehat{z}^{(m-1)} + |N^{-1}(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})| \max_i |N x_i|. \tag{D.34}$$

4. *Embed in the interval* $[0,1]$ *using* $\psi = m/n$. The asymptotic expansion for $\widehat{z}^{(m)}$ in Theorem 3.1 combined with the tightness of $\mathbb{G}_n$ in Lemma D.13 shows

$$\sup_{\psi_0 \leq \psi \leq n/(n+1)} |2\mathsf{f}(c_\psi)(\widehat{z}_\psi - \widehat{z}_{\psi - 1/n})| = \mathsf{o}_\mathsf{P}(n^{-1/2}),$$

while the asymptotic result for $\hat{\beta}^{(m)}$ in Lemma D.13 shows

$$\sup_{\psi_1 \leq \psi \leq n/(n+1)} |N^{-1}(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})| = \mathsf{o}_\mathsf{P}(n^{-1/2}).$$

5. *Combine.* The bound (D.34) and the triangle inequality give

$$\begin{aligned} 0 &\leq 2\mathsf{f}(c_{m/n})(\widehat{z}^{(m)} - \widehat{d}^{(m)}) \\ &\leq 2\mathsf{f}(c_{m/n})|\widehat{z}^{(m)} - \widehat{z}^{(m-1)}| + 2\mathsf{f}(c_{m/n})|N^{-1}(\widehat{\beta}^{(m)} - \widehat{\beta}^{(m-1)})| \max_i |N x_i|. \end{aligned}$$

The bounds in item 4, combined with the condition $\max_i |N x_i| = \mathsf{O}_\mathsf{P}(n^{\kappa - 1/2})$ for some $\kappa < \eta \leq 1/4$ by Assumption 3.1(iib), give a further bound

$$0 \leq 2\mathsf{f}(c_{m/n})(\widehat{z}^{(m)} - \widehat{d}^{(m)}) \leq \mathsf{o}_\mathsf{P}(n^{-1/2}) + \mathsf{o}_\mathsf{P}(n^{-1/2} n^{\kappa - 1/2}) = \mathsf{o}_\mathsf{P}(n^{-1/2}),$$

as desired. □

Proof of Theorem 3.5. Lemma D.10(c) using Assumption 3.1(ia, ib, ii) shows

$$b^\dagger = \{\widehat{\mathsf{G}}_n^{xx,0}(b, \hat{c}_\psi^b)\}^{-1}\{n^{1/2}\widehat{\mathsf{G}}_n^{x,1}(b, \hat{c}_\psi^b)\} = (\Sigma_n \psi)^{-1}\mathbb{G}_n^{x,1}(0, c_\psi) + \rho_\psi b + \mathsf{o}_\mathsf{P}(1),$$

uniformly in $|b| \leq n^{1/4-\eta} B$, $\psi_0 \leq \psi \leq 1$. Lemma D.13(b) using Assumption 3.1($ia - ic, ii, iii$) shows that $N^{-1}(\hat{\beta}_\psi - \beta)$ is uniformly bounded for $\psi \geq \psi_1$. Thus, on a set with large probability both $b^\dagger$ and b can be replaced by $N^{-1}(\hat{\beta}_\psi - \beta) + \mathsf{o}_\mathsf{P}(1)$. Lemma D.12 using Assumption 3.1(ia, ic) shows that $\rho_\psi \leq \rho_0 < 1$ for $\psi \geq \psi_0$. Thus, it holds

$$N^{-1}(\hat{\beta}_\psi - \beta) = \frac{1}{1 - \rho_\psi}(\Sigma_n \psi)^{-1}\mathbb{G}_n^{x,1}(0, c_\psi) + \mathsf{o}_\mathsf{P}(1).$$

Insert $\rho_\psi = 2c_\psi \mathsf{f}(c_\psi)/\psi$ and $\mathbb{K}_n(c_\psi) = \mathbb{G}_n^{x,1}(0, c_\psi)$ to get the desired expansion. □

Proof of Theorems 3.6 and 3.7. Tightness follows from Lemma D.5(c), and convergence of finite dimensional distributions follows from the central limit theorem for martingale differences, see Helland (1982, Theorem 3.2b) (Theorem 3.2b) using Assumption 3.1(iic). □

Appendix E: A result on order statistics of t-distributed variables

Theorem E.1. *Let $v_1, \ldots, v_n$ be independent absolute $\mathsf{t}_{m-\dim x}$ distributed. Consider the $(m+1)'$st smallest order statistic $\hat{v}^{(m)}_{(m+1)}$. Suppose $\dim x$ is fixed while $m \sim \psi n$ for some $0 < \psi < 1$. Let φ be the standard normal density. Then, as $n \to \infty$,*

$$2\varphi(c_{m/n}) n^{1/2} (\hat{v}^{(m)}_{(m+1)} - c_{m/n}) \stackrel{\mathsf{D}}{\to} \mathsf{N}\{0, \psi(1-\psi)\}.$$

Sketch of the proof of Theorem E.1. Let $\hat{v}^{(m)}_{(m+1)}$ be the $(m+1)'$st quantile of a sample of n scaled, absolute $\mathsf{t}_{m-\dim x}$ variables. To get a handle on the asymptotic distribution of $\hat{v}^{(m)}_{(m+1)}$ consider first the $(m+1)'$st smallest order statistic, $\hat{w}_{(m+1)}$ say, from n draws of absolute standard normal variables with distribution function $2\Phi(y) - 1$. This satisfies

$$2\varphi(c_{m/n}) n^{1/2} (\hat{w}_{(m+1)} - c_{m/n}) \stackrel{\mathsf{D}}{\to} \mathsf{N}\{0, \psi(1-\psi)\},$$

for $m \sim \psi n$ and $c_\psi = \mathsf{G}^{-1}(\psi)$ due to Lemmas D.1, D.2(a). For the $\mathsf{t}_{m-\dim x}$ order statistic $\hat{v}^{(m)}_{(m+1)}$ it is useful to Edgeworth expand $\mathsf{P}(\mathsf{t}_{m-\dim x} \leq y) = 2\{\Phi(y) + \mathrm{O}(n^{-1})\} - 1$, for $m \sim \psi n$, 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 $\mathsf{P}(\hat{v}^{(m)}_{(m+1)} \leq y)$ 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 result. □

References

Aroian, L.A. (1941) A study of R.A. Fisher's z distribution and the related F distribution. *Ann. Math. Statist.,* **12**, 429–448.

Atkinson, A.C. (1994) Fast very robust methods for detection of multiple outliers. *J. Amer. Stat. Assoc.,* **89**, 1329–1339.

Atkinson, A.C. and Riani, M. (2000) *Robust Diagnostic Regression Analysis.* New York: Springer.

Atkinson, A.C. and Riani. M. (2006) Distribution theory and simulations for tests of outliers in regression. *J. Comput. Graph. Statist.*, **15**, 460–476.

Atkinson, A.C., Riani, M. and Cerioli, A. (2010a) The forward search: Theory and data analysis (with discussion). *J. Korean Stat. Soc.,* **39**, 117–134.

Atkinson, A.C., Riani, M. and Cerioli, A. (2010b) Rejoinder: The forward search: Theory and data analysis. *J. Korean Stat. Soc.,* **39**, 161–163.

Bahadur, R.R. (1966) A note on quantiles in large samples. *Ann. Math. Statist.,* **37**, 577–580.

Bellini, T. (2012) Interactive Outlier Detection in Cointegration Analysis. Discussion paper, Universita di Parma.

Bercu, B. and Touati, A. (2008) Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.*, **18**, 1848–1869.

Bickel, P.J. (1975) One-step Huber estimates in the linear model. *J. Amer. Stat. Assoc.,* **70**, 428–434.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edn. New York: Wiley. MR1700749

Cavaliere, G. and Georgiev, I. (2013) Exploiting infinite variance through dummy variables in nonstationary autoregressions. *Econometric Theory,* **29**, 1162–1195.

Cerioli, A., Farcomini, A. and Riani M. (2014) Strong consistency and robustness of the forward search estimator of multivariate location and scatter. *J. Multivariate Anal.,* **125**, 167-183.

Csörgő, M. (1983) *Quantile Processes with Statistical Applications.* CBMS-NFS Regional Conference Series in Applied Mathematics 42, Society for Industrial and Applied Mathematics.

Dollinger, M.B. and Staudte, R.G. (1991) Influence functions of iteratively reweighted least squares estimators. *J. Amer. Stat. Assoc.,* **86**, 709–716.

Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econom. J.,* **12**, 367–381.

Guenther, W.C. (1977) An easy method for obtaining percentage points of order statistics. *Technometrics,* **19**, 319–321

Hadi, A.S. (1992) Identifying multiple outliers in multivariate data. *J. R. Stat. Soc. B Ser. Methodol.*, **54**, 761–771.

Hadi, A.S. and Simonoff, J.S. (1993) Procedures for the identification of multiple outliers in linear models. *J. Amer. Stat. Assoc.,* **88**, 1264-1272.

Hawkins, D.M. and Olive, D.J. (2002) Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algorithm (with discussion). *J. Amer. Stat. Assoc.,* **97**, 136–159.

Helland, I. (1982) Central limit theorems for martingales with discrete or continuous time. *Scand. J. Stat.,* **9**, 79–94.

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. *J. Korean Stat. Soc.,* **39**, 137–145.

Johansen, S. and Nielsen, B. (2013) Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator. *Econometrics,* **1**, 53–70.

Johansen, S. and Nielsen, B. (2014a) Asymptotic theory of M-estimators for multiple linear regression in time series. Discussion paper Nuffield College Discussion Paper.

Johansen, S. and Nielsen, B. (2014b) Outlier detection algorithms for least squares time series regression. Discussion paper Nuffield College Discussion Paper.

Kiefer, J. (1967) On Bahadur's representation of sample quantiles. *Ann. Math. Statist.,* **38**, 1323–1342.

Koul, H.L. and Ossiander, M. (1994) Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.,* **22**, 540–582.

Lee, S. and Wei, C.Z. (1999) On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.,* **27**, 237–61.

Nielsen, B. (2014) ForwardSearch. R package version 1. http://www.R-project.org.

R Development Core Team (2011). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

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. *Adv. Data Anal. Classif.,* **1**, 123–141.

Riani, M., Atkinson, A.C. and Cerioli, A. (2009) Finding an unknown number of multivariate outliers. *J. R. Stat. Soc. B Ser. Methodol.*, **71,** 447–466.

Ronchetti, E and Welsh, A.H. (2002) A journey in single steps: robust one-step M-estimation in linear regression *J. Stat. Plan. Inference,* **103**, 287–310.

Rousseeuw, P.J. (1984) Least median of squares regression. *J. Amer. Stat. Assoc.,* **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. *J. Amer. Stat. Assoc.,* **75**, 828–838.

Sampford, M.R. (1953) Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.,* **24**, 130–132.

Simpson, D.G., Ruppert, D. and Carroll, R.J. (1992) On one-step GM estimates and stability of inferences in linear regression. *J.Amer. Stat. Assoc.,* **87**, 439–450.

Shorack, G.R. (1979) Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-constructions. *Stochastic Process. Appl.,* **9**, 95–98.

Soms, A.P. (1976) An asymptotic expansion for the tail area of the t-distribution. *J. Amer. Stat. Assoc.,* **71**, 728-730.

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

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

Víšek, J.Á. (2006c) The least trimmed squares. Part III: Asymptotic normality. *Kybernetika,(Prague)* **42**, 203–224.